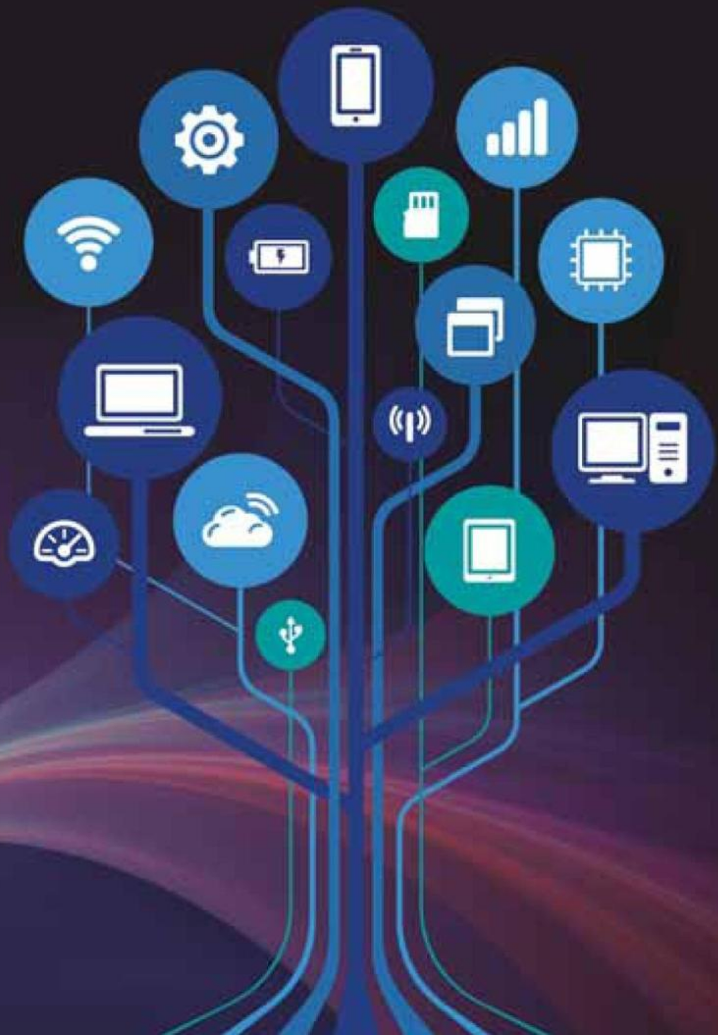
Fundamentals of MIMO Wireless Communications

Rakhesh Singh Kshetrimayum



Fundamentals of MIMO Wireless Communications

Since the first Multiple Input Multiple Output (MIMO) approach was patented in 1994 by T. Kailath and A. J. Paulraj, MIMO has been a well-researched topic in wireless communications. Presenting state-of-the-art MIMO techniques, this book builds a clear and coherent understanding of the essential concepts of MIMO wireless communications. Dr Kshetrimayum discusses all important techniques and tools of MIMO wireless communications which include but does not limit to MIMO channel models, power allocations and channel capacity, space-time codes, MIMO detection and antenna selection. An important feature of the book is to provide practical insights into the world of modern telecommunication systems. Some recent techniques developed in MIMO wireless in the last decade including spatial modulation, MIMO based cooperative communications, Large-scale MIMO systems, massive MIMO and STBC-SM are also covered. The concepts are supported by numerous solved examples, review questions, simulation figures and exercises. This course book is intended for graduate students in electronics, electrical and computer engineering.

Rakhesh Singh Kshetrimayum is a Professor at the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati. He is Life Fellow of IETE, Antenna Test and Measurement Society (ATMS) and a senior member of IEEE, USA. His research areas include printed monopole antennas, passive microwave devices and wireless communications.

Fundamentals of
MIMO Wireless Communications

Rakhesh Singh Kshetrimayum

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, vic 3207, Australia
4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108415699

First published 2017

Printed in India

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Kshetrimayum, Rakhesh Singh, author.
Title: Fundamentals of MIMO wireless communications / Rakhesh Singh Kshetrimayum.
Description: Delhi, India : Cambridge University Press, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2017000107 | ISBN 9781108415699 (hardback)
Subjects: LCSH: MIMO systems.
Classification: LCC TK5103.4836 .K75 2017 | DDC 004.6--dc23 LC record available at https://lccn.loc.gov/2017000107

ISBN 978-1-108-41569-9 Hardback

Additional resources for this publication at www.cambridge.org/9781108415699

Dedicated to
my beloved wife G. D. Kshetrimayum

Contents

Figures

Tables

Preface

This book aims to give its readers an opportunity to build a strong foundation in the subject of MIMO wireless communications. It is an ideal book for students pursuing senior undergraduate and junior postgraduate courses in MIMO wireless communications. The necessary background details of wireless communications have been put in Appendix A–G.

Chapter 1 gives a brief introduction to multiple-input multiple-output (MIMO) (pronounced 'My-Moe') systems. The receiver diversity techniques in single-input multiple-output (SIMO) systems are briefly discussed. Following that, MIMO transmit diversity scheme viz., Open Loop, Closed Loop and Blind MIMO systems are also talked about. Rate and diversity gain is defined and a discussion about the diversity multiplexing trade-offs is also provided.

Fading distributions, the precursors to MIMO channel models, are treated in chapter 2. Fading distributions could be divided into two kinds: (a) classical fading distributions and (b) generalized fading distributions. In classical fading distributions, the probability density function (pdf), the cumulative distribution function (cdf) and the moment generating function (mgf) of Gaussian, Rayleigh, Rice, Chi-squared and Nakagami-m fading distributions are provided. Among the many generalized fading distributions, k-μ, α-μ and η-μ fading distributions are investigated and in particular, classical fading distributions are also discussed.

Chapter 3 is devoted to the analytical MIMO channel models. Analytical MIMO channel models can be divided into four types: (a) independent and identically distributed (uncorrelated) MIMO channel model, (b) Kronecker (separately correlated) MIMO fading channel model, (c) fully correlated MIMO channel model, and (d) keyhole (rank deficit) MIMO channel model. Parallel decomposition of MIMO channel is discussed at the end of the chapter.

The capacity of a MIMO channel for uniform and adaptive power allocation scheme is treated in chapter 4. Uniform power allocation is employed when the channel state information (CSI) is available at the receiver but not at the transmitter (open loop MIMO system). Adaptive power allocation based on Waterfilling algorithm can be used when CSI is available at the receiver as well as at the transmitter (closed loop MIMO system). Near optimal power allocation for high and low SNR cases is described in the last section of the chapter.

In chapter 5, the capacity of simplified MIMO channels viz., (a) SISO channels, (b) SIMO channels, (c) MISO channels, (d) unity MIMO channel, and (e) identity MIMO channel, is investigated. The ergodic capacity and outage probability for some of the above fading channels are also found out.

The ergodic capacity and outage probability for i.i.d. fading MIMO channels are described in chapter 6. Then, the effect of antenna correlation on the MIMO channel capacity is observed. Finally, it is shown that if we have keyhole propagation for a highly scattered environment, the capacity is very low.

In chapter 7, a discussion on why we need space-time codes is presented at the very first. Then, the three code design criteria, viz., rank, determinant and trace, is provided. A study on the first and the most powerful space-time codes, also known as Alamouti space-time codes, is carried out. The performance comparison of Alamouti space-time code with maximal ratio combining (MRC) is described. The coding gain, diversity gain and code rate of Alamouti space-time code are presented. A very important concept in performance analysis of wireless communication over fading channel is that in order to find the average probability of error, we need to find the average conditional probability of error (CEP) over the received SNR. A channel is in outage whenever we transmit a message at a higher rate than the channel capacity. In the last section of the chapter, the outage probability and average probability of error for single input single output (SISO) system over fading channels, and an extension of this analysis for Alamouti space-time codes, are provided.

An extension of Alamouti space-time code for N_T number of transmitting antennas, also known as orthogonal space-time block codes (OSTBC), is introduced. But OSTBC does not provide any coding gain. There is another type of space-time code, termed as space-time trellis code (STTC), which provides both coding and diversity gains. An exploration of both of these space-time codes is carried out in chapter 8. The symbol error rate (SER) of OSTBC over spatially correlated Rayleigh fading, as well as the i.i.d. Rayleigh fading of MIMO channels is calculated. Pairwise error probability (PEP) calculation of space-time codes over correlated as well as i.i.d. Rayleigh fading is also carried out. A brief introduction to space-time Turbo codes is presented. Some sections are devoted to differential OSTBC and algebraic space-time codes, which include Perfect space time codes and Golden codes.

In MIMO detection, one needs to detect signals jointly, since many signals are transmitted from the transmitter to the receiver. Among the available detection techniques, maximum likelihood (ML) detection is the most optimal technique, but its complexity grows exponentially with the number of antennas. There are other sub-optimal techniques like Zero Forcing (ZF) and Minimum Mean Square Error (MMSE) which are less complex. Chapter 9 is devoted to such techniques. A comparison of the noise amplification in ZF and MMSE is presented and then, the performance of these techniques in terms of probability of error and outage probability is evaluated. Sphere decoding is also discussed in this chapter.

Chapter 10 contains Diagonal-Bell Laboratories Layered Space-Time (D-BLAST) and Vertical BLAST (V-BLAST), which are spatial multiplexed MIMO systems which give high spectral efficiencies. BLAST detection scheme is basically based on the following three steps: interference nulling, ordering to select the sub-stream with the largest signal-to-noise ratio (SNR) or other criteria, and successive interference cancellation.

MIMO systems increase the capacity and minimize the error rate as compared to SISO systems. But, they have a higher fabrication cost and energy consumption due to multiple RF chains. Selection of suitable antenna minimizes this by using lesser number of RF chains and switches. The best set of antennas should be selected at the transmitter or the receiver end, so as to maximize the channel capacity or the received SNR. Transmit antenna selection over η–μ fading channels is described in chapter 11 where soft antenna selection for closely spaced antennas is also introduced. Spatial Modulation (SM) is a multiple input multiple output (MIMO) wireless communication technique that gives better spectral efficiency for a fixed bandwidth and same signal constellation size. Symbol error rate (SER) performance of an SM system in generalized η-μ fading channels for several modulation schemes is evaluated. It is also shown that spatial modulation with antenna selection has a huge advantage in the outage probability over SM MIMO systems.

STBC-SM, MIMO based cooperative communication, Large scale MIMO systems and MIMO cognitive radios are some of the latest developments in MIMO wireless communications. A discussion on hybrid STBC and SM is provided also in this chapter. In chapter 12, it is shown that there can be a significant improvement in the performance of cooperative communication if one employs MIMO techniques. After this, Large scale MIMO systems are investigated. Three scenarios are discussed viz., Single User, Multi-user and Multi-cell Large scale MIMO systems. Finally MIMO Cognitive Radios are discussed in some detail.

Finally, there is a suggestion for all the students: hone your fundamentals! Technologies change every now and then, however, the fundamentals which are the building blocks for these new technologies remain the same.

Acknowledgments

I am grateful to all those students who took the EE642 MIMO Wireless Communications and EE635 Advanced Topics in Communications Systems courses, for their thought provoking questions and fruitful discussions. I am also grateful to the students who have worked on projects related to MIMO Wireless Communications, viz., A. K. Saxena (*presently with University of California, Irvine*), A. Polkam (*presently with State University of New York, Buffalo*), B. Kumbhani, G. Joseph, J. Yadav, K. Keerthi Vishal, L. Chaudhary, L. N. B. Reddy, M. Choudhary, M. Kulkarni (*presently with University of Texas, Austin*), N. Abhilash, N. Kumar, N. K. Meena, R. P. Singh (*presently with University of California, San Diego*), S. Alam, S. Kabra (*presently with Virginia Tech*), S. Sunani, S. A. Kumbhar, S. Goyal, S. Gupta, V. K. Mohandas and V. Viswanath.

I am thankful to all my colleagues and staff at the department of EEE, IIT Guwahati, including Dr A. Chatterjee, Professor A. K. Gogoi, Professor A. Mahanta, Dr A. Rajesh, Dr B. K. Rai, Professor H. B. Nemade, Dr H. S. Shekhawat, Dr K. Dhaka, Dr M. K. Bhuyan, Dr N. Nallam, Professor P. K. Bora, Dr P. R. Sahu (*presently with IIT Bhubaneswar*), Professor R. Bhattacharjee, Dr R. K. Sonkar, Dr S. Chouhan, Dr S. Das, Professor S. K. Bose, Dr S. K. Nayak, Dr S. R. Ahamed, Dr T. Jacob and my peers who have contributed to the MIMO literature – Dr C. Yuen (*Singapore*), Professor Y. Chen (*China*), Professor Y. Zhang (*Norway*), Dr C. Ling (*UK*) and Dr C. Sun (*China*). I have tried to refer to all the books, journals and conference papers which I had read over the years regarding MIMO technologies. Without the help of these scientific documents, writing this book may not have been possible. My heartiest thanks go to the authors of all those scientific documents for their contribution in MIMO and I hope to see many more interesting MIMO related papers in the future as well. I am really grateful to Professor Y. Zhang (*Norway*), Professor M. Jo (*South Korea*), Dr C. Ling (*UK*), Dr X. Zhang (*USA*) and Professor V. K. Bharghava (*Canada*) for sparing some time to review my book and for providing the excellent feedbacks. I am very grateful to M. Choudhary at CUP for helping me at the various stages of publication of this book. I am also very thankful to S. Bhattacharjee for meticulously editing even minor corrections.

Last but not the least, I am thankful to my wife, children, brothers, parents and relatives for their consistent support and encouragement.

Abbreviations

AF	Amplify and Forward
ASTBC	Algebraic Space Time Block Codes
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BEP	Bit Error Probability
BLAST	Bell Laboratories Layered Space Time
BPSK	Binary Phase Shift Keying
BS	Base Station
CDF	Cumulative Distribution Function
CGD	Coding Gain Distance
CoMP	Coordinated Multipoint Transmission
CPE	Conditional Probability of Error
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the Transmitter
CVP	Closest Vector Problem
DAST	Diagonal Algebraic Space Time
D-BLAST	Diagonal Bell Laboratories Layered Space Time
DBPSK	Differential Binary Phase Shift Keying
DF	Decode and Forward
DMUX	Demultiplexing
EP	Error Propagation
EVM	Error Vector Magnitude
FFT	Fast Fourier Transform
GSM	Global System for Mobile Communication
HAS	Hard Antenna Selection
H-BLAST	Horizontal Bell Laboratories Layered Space Time
ICI	Inter Cell Interference
iid	Independent and Identically Distributed
I-O	Input-Output
IUI	Inter User Interference
LLL	Lenstra, Lenstra & Lovasz
LLR	Log Likelihood Ratio
LOS	Line of Sight

LR	Lattice Reduction
LS MIMO	Large Scale Multiple-input Multiple-output
LTE	Long Term Evolution
MAC	Multiple Access Channel
MGF	Moment Generating Function
MIMO	Multiple-input Multiple-output
MISO	Multiple-input Single-output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MPAM	M-ary Pulse Amplitude Modulation
M-PSK	Multiple Phase Shift Keying
MRC	Maximal Ratio Combining
MSK	Minimum Shift Keying
MU MIMO	Multi User Multiple-input Multiple-output
NLOS	Non Line of Sight
NVD	Non Vanishing Determinant
Od	Orthogonal Defect
OOPSIC	Optimal Ordered Perfect Successive Interference Cancellation
OSIC	Ordered Successive Interference Cancellation
OSTBC	Orthogonal Space Time Block Code
PEP	Pairwise Error Probability
PSIC	Perfect Successive Interference Cancellation
PSD	Power Spectral Density
PSTBC	Perfect Space Time Block Code
PUs	Primary Users
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RV	Random Variable
RZF	Regularized Zero Forcing
SAS	Soft Antenna Selection
SC	Selection Combining
SER	Symbol Error Rate
SIC	Successive Interference Cancellation
SINR	Signal-plus-interference-plus-noise Ratio
SIMO	Single-input Multiple-output
SISO	Single-input Single-output
SM	Spatial Modulation
S/P	Serial to Parallel Conversion
SNR	Signal to Noise Ratio
SRCC	Systematic Recursive Convolutional Codes
STBC	Space Time Block Code
STTC	Space Time Trellis Code
SU MIMO	Single User Multiple-input Multiple-output

SUs	Secondary Users
SVD	Singular Value Decomposition
TAS	Transmit Antenna Selection
TAST	Threaded Algebraic Space Time
V-BLAST	Vertical Bell Laboratories Layered Space Time
ZF	Zero Forcing
ZMCSCG	Zero Mean Circular Symmetric Complex Gaussian

Introduction to MIMO Systems

CHAPTER 1

1.1 Introduction

In this chapter, we will give introduction to MIMO (pronounced "My-Moe", J. G. Andrews et al., 2007) systems. First we will summarize some of the background in wireless communications which are required for better understanding of MIMO systems.

We will start with three diversity, viz. frequency, time and space diversity to combat detrimental effects of wireless fading channels.

Then we will discuss about the fading channel characteristics. We will define two important terms, viz. coherence bandwidth and coherence time. Then we will define what are frequency-flat and selective fading and slow and fast fading.

Then we will discuss about multi-antenna systems like Multiple-Input Multiple-Output (MIMO), Multiple-Input Single-Output (MISO) and Single-Input Multiple-Output (SIMO) systems. In SIMO systems, we will briefly discuss about the receiver diversity techniques like Equal-Gain Combining (EGC), Selection Combining (SC) and Maximal Ratio Combining (MRC).

In MIMO transmit diversity schemes we will define Open Loop, Close Loop and Blind MIMO systems.

MIMO systems have rate and diversity gain over Single-Input Single-Output (SISO) systems. We will define rate and diversity gain and discuss concisely about the diversity multiplexing trade-offs.

Finally we will mention some of the applications of MIMO systems.

1.2 Diversity in wireless communications

Wireless communications, which allow movement while communicating, is a very attractive feature for the mobile users. But it is a challenge to the wireless engineers because of channel fading due to random signal attenuation and phase distortions from the Multipath Components (MPCs). There are three diversity techniques (S. Haykin et al., 2005) to mitigate fading, viz.

(a) *Frequency diversity*: In frequency diversity techniques, we will send information bearing signals by carriers whose frequency gap is greater than coherence bandwidth of the channel; for instance, frequency hopped spread spectrum system.

(b) *Time diversity*: In time diversity techniques, we will send information bearing signals in different time slots which is greater than coherence time of the channel: for example, channel coding with interleaving and

(c) *Space diversity*: In space diversity techniques, we employ multiple antennas, which are placed amply far away, at the transmitter and receiver. Space diversity was left aside for many years in the literature due to the problem of spatial interference. We will explore this diversity in this book.

1.3 Wireless fading channel characteristics

The characteristics of wireless communication channels between the transmitter and receiver decide the performance of the wireless systems. Let us try to understand some terminologies of fading channel first. The time and frequency variations of the channel are quantified in terms of channel coherence time (D. Tse et al., 2005) and coherence bandwidth.

The *coherence bandwidth,* (B_c), is the frequency range, (Δf), over which the channel frequency response, $h(f)$, is flat and is inversely proportional to delay spread of the channel (R. Janaswamy, 2001 and A. K. Jagannatham, 2016). Assume a signal, $x(t)$, is sent over a channel with impulse response, $h(t)$, then the output of the wireless channel, $y(t)$, can be obtained as the convolution of $x(t)$ and $h(t)$ for a linear time invariant system. In the frequency domain, the convolution of $x(t)$ and $h(t)$ is transformed to multiplication of $X(f)$ and $h(f)$. If the bandwidth, B_s, of the signal is less than the coherence bandwidth, B_c, of the channel, then the output will be undistorted. On the contrary, if the bandwidth, B_s, of the signal is greater than the coherence bandwidth, B_c, of the channel, then the output will be distorted.

The motion of users in wireless communications gives rise to Doppler shift which in turn converts the wireless channel coefficient into time-varying and introduces time-selectivity in wireless channels. The *coherence time,* (T_c), is approximate duration of the time for which the wireless channel can be assumed constant, and is inversely proportional to the maximum Doppler frequency shift or spread, (B_d). Let us summarize.

(a) When coherence bandwidth is greater than signal bandwidth, then all frequency components of the signal will experience the similar kind of fading (*frequency-flat fading*).

(b) If coherence bandwidth is smaller than signal bandwidth, then all frequency components will not experience same fading (*frequency-selective fading*).

(c) When coherence time is greater than symbol time duration, it means channel variation is slower than the signal variation (*slow fading*).

(d) If coherence time is smaller than symbol time duration, it means channel variation is faster than the signal variation (*fast fading*).

Review question 1.1 *What is the coherence bandwidth of channel?*

Review question 1.2 *What is coherence time of channel?*

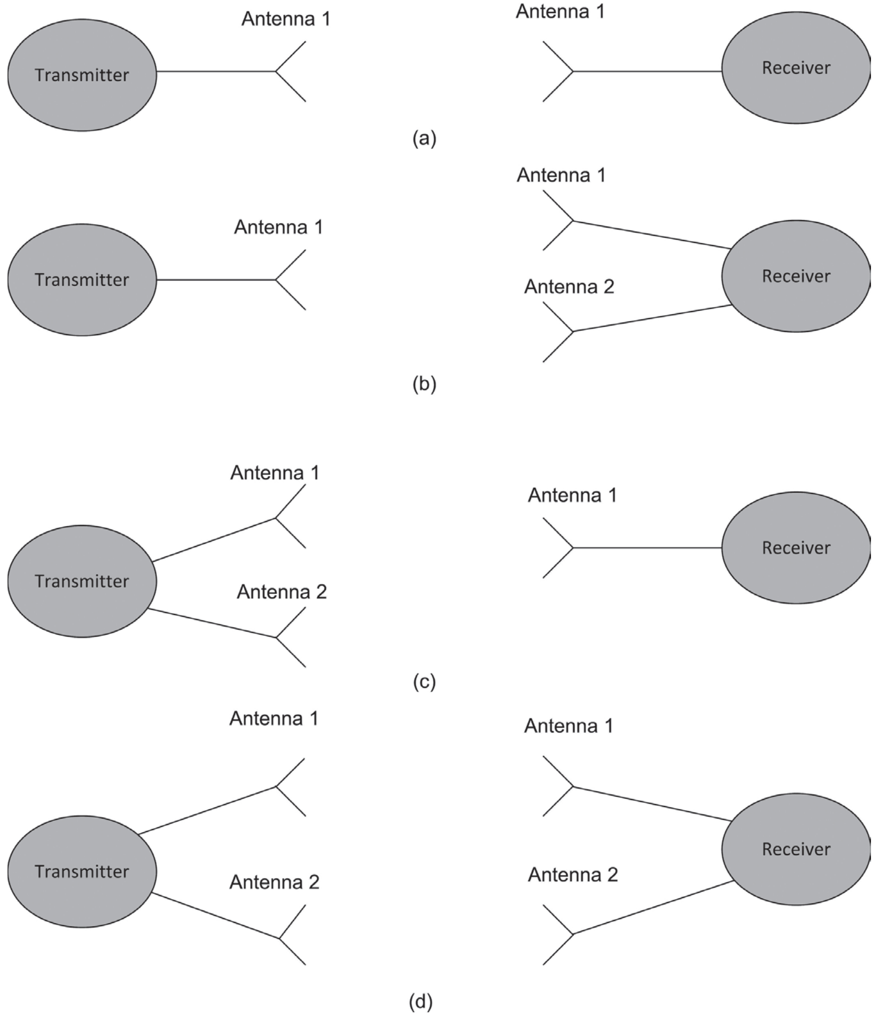


Fig. 1.1 (a) 1×1 SISO (b) 1×2 SIMO (c) 2×1 MISO and (d) 2×2 MIMO systems

1.4 What are MIMO systems?

Space diversity employing multiple antennas at the transmitter and receiver, also popularly known as Multiple-Input Multiple-Output (MIMO)/multi-antenna systems, are capacity boosters for wireless channels without penalty in bandwidth and power. The capacity of the channel increases linearly with the minimum of N_R or N_T for a $N_R \times N_T$ MIMO system in a rich Rayleigh scattering environment. N_T and N_R are number of transmitting and receiving antennas, respectively.

A particular case for $N_T = N_R = 1$ is referred to as *Single-Input, Single-Output (SISO) system,* which is the single link communication as depicted in Fig. 1.1 (a). Low-Density Parity-Check Code (LDPC) and Turbo codes with iterative decoding algorithms are the capacity booster for SISO systems.

Another particular case is for $N_T = 1$ and $N_R \geq 2$; such a system is referred to as *Single-Input, Multiple-Output (SIMO) system* (receive diversity). A 1×2 SIMO system is shown in Fig. 1.1 (b). Receiver diversity techniques like Equal Gain Combining (EGC), Selection Combining (SC) and Maximal Ratio Combining (MRC) can be employed at the receiver to combat multipath fading phenomenon. For SIMO systems, there are N_R channel links. SC selects the signal branch with the highest signal-to-noise ratio (SNR) among the N_R channel links. Each channel link will have different channel gain coefficients, path delays and phases. EGC co-phases signal on each branch and then combine them with equal weight. MRC outputs the weighted sum of all the branches. Weights are chosen as the complex conjugate of the channel gain coefficients. Branches with high SNR should be weighted more than with branches with low SNR. MRC is optimal in terms of SNR but complex to implement from the other two combining schemes (A. Molisch, 2005 and A. Goldsmith, 2005). We will discuss MRC briefly when we study Alamouti space-time codes.

Another particular case is for $N_T \geq 2$ and $N_R = 1$; such a system is referred to as *Multiple-Input, Single-Output (MISO) system* (transmit diversity). A 2×1 MISO system is depicted in Fig. 1.2 (c). Mobile station (MS) is generally small and receive diversity is not cost-impressive. Instead transmit diversity at the base station (BS) is a better choice.

But in 5G wireless communications, both receive and transmit diversities are envisaged. A *MIMO* system employing N_T transmitting antennas and N_R receiving antennas has both transmit and receive diversities (J. G. Proakis et al., 2007). A 2×2 MIMO system is also shown in Fig. 1.1 (d).

Review question 1.3 *What is the main advantage of MIMO system?*

Review question 1.4 *What is Equal gain combining?*

Review question 1.5 *What is Maximum ratio combining?*

Review question 1.6 *What is Selection combining?*

1.5 Which are the three cases of MIMO transmit diversity schemes?

Let us consider three types of transmit diversity (C. Yuen et al., 2007):

(a) *Closed loop MIMO system*: In closed loop MIMO system, feedback of channel gain and phase from the receiver is given to the transmitter. Hence Channel State Information (CSI) is available at both the transmitter and receiver. CSI could be of two types (T. Brown et al., 2012): instantaneous channel and statistical average of the channel (distribution of the channel).

(b) *Open loop MIMO system*: In open loop MIMO system, the receiver estimates the channel using the feed forward pilot signals but no feedback given to the transmitter. Hence CSI is available at the receiver and not at the transmitter. It is usually difficult to obtain the instantaneous CSI at the transmitter, but it is fairly possible to obtain CSI at the receiver by sending a training sequence or separate pilot signals.

(c) *Blind MIMO system*: In blind MIMO system, no channel state information is available at both the transmitter and receiver.

Pilot signals are sent orthogonal to the message signal either in frequency or time. For N_T transmit antenna an optimal pilot signal set comprises N_T mutually orthogonal signals with equal power, each assigned for a transmitting antenna (H. Huang et al., 2012). For slow fading, channel stability enables the receiver to acquire the CSI required for coherent detection of the transmitted code-word. In fast fading case, the channel coefficients vary fast and reliable channel estimation may not possible. In order to acquire CSI properly, we require more pilot signals resources than that of the slow varying channel. Hence, CSI is not available, and the receiver should operate in a non-coherent mode.

Review question 1.7 *What are non-coherent and coherent systems?*

1.6 Why MIMO systems?

In SISO system, data rate can be increased by either increasing the transmission bandwidth and power (a direct implication of Shannon's channel capacity formula, $C = BW \log_2 (1 + SNR)$, T. M. Cover et al. 1999). But frequency spectrum is a valuable resource and sometimes restricted for use, increasing transmission bandwidth is not acceptable solution for increasing data rate. Similarly, increasing transmission power is also not a proper solution for increasing data rate since we need highly expensive radio frequency (RF) amplifier. It will also reduce battery life time of mobile unit. Besides, there is the problem of higher interference with higher power. It may be considered unlawful by transmission regulations (H. Huang et al., 2012). MIMO increases the spectral efficiency without increasing the transmission power and bandwidth. Note that if the bandwidth is fixed, data rate and spectral efficiency could be used interchangeably. Basically two fundamental gains are achieved from MIMO systems (E. Biglieri et al., 2004):

(a) *Rate gain:* For parallel MIMO channels, there is at the most minimum $\{N_R, N_T\}$ rate gain from that of a SISO system. It is also known as multiplexing gain. In spatial multiplexing MIMO systems different data are sent through the parallel channels with the help of a serial to parallel converter as shown in Fig. 1.2 (a). It is a highly simplified model for easier understanding. It gives higher transmission rate.

(b) *Diversity gain:* The maximum number of independent paths travelled by each signals can be at the most $N_R \times N_T$. It is highly possible that not all the paths are highly faded. In this case, same data is sent through all the multiple antennas at the transmitter. If some of the paths are completely down, some paths will still be working. The receiver tries to make an efficient use of this to decode the data accurately. It gives higher link reliability.

For instance, the maximum data rate of a 3×3 MIMO system and diversity gains for a 2×2 MIMO system depicted in Fig. 1.2 (b) and (c) are 3 and 4, respectively. For 3×3 MIMO system of Fig. 1.2 (a) and (b), we may be sending bits in three parallel streams. Hence the data rate will be tripled. For 2×2 MIMO system of Fig. 1.2 (c), we may be sending same data stream over the four independent paths. So the data rate is the same with that of a SISO system. In this case, if any of the links/paths is down, the other link may be working. The receiver can decode the data stream correctly (with less probability of error) using the working link/path. This is possible because of the diversity gain. Assumption made is that the rank of the 3×3 MIMO channel is 3 and all the paths are independent for the 2×2 MIMO system. Other gains may be array gain and interference reduction gain which we will not discuss here.

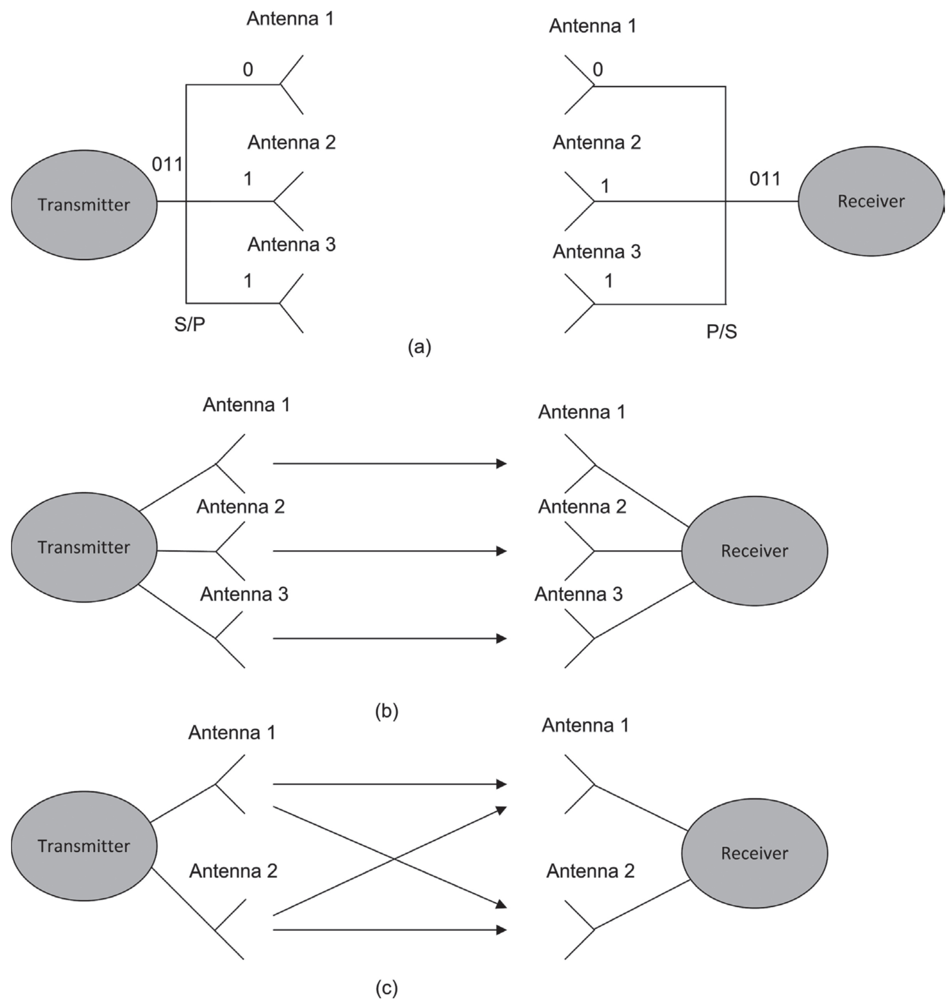


Fig. 1.2 (a) Spatial MUX 3×3 MIMO system (simplified case) (b) Rate gain of a 3×3 MIMO system, and (c) Diversity gain of a 2×2 MIMO system {S/P: Serial to parallel conversion; P/S: Parallel to serial conversion}

In fact there is trade-off, popularly known as *diversity-multiplexing trade-off,* between these two fundamental gains (diversity and rate gains). It is called *diversity-multiplexing trade-off* because if we increase diversity gain, rate gain automatically reduces and vice versa. MIMO system must be designed considering this trade-off. A simple characterization of this trade-off is given for block fading channel in the limit of asymptotically high SNR.

The rate gain is associated with the data rate of transmission. What is the exact relationship (mathematically)? Note that a transmission scheme is said to achieve *multiplexing gain r* if the data rate (bps) per unit Hertz, R (SNR) which is a function of SNR satisfy

$$r = \lim_{SNR \to \infty} \frac{R(SNR)}{\log_2(SNR)} \tag{1.1}$$

Hence the multiplexing gain is given by slope of the data rate for fixed frame error rate plotted as function of the SNR on a linear-log scale. The diversity gain is associated with the probability of error in detection. What is the exact relationship (mathematically)? A transmission scheme is said to achieve *diversity gain, d,* if the probability of error, P_e (*SNR*), as functions of SNR satisfies

$$d = -\lim_{SNR \to \infty} \frac{\log_2\{P_e(SNR)\}}{\log_2(SNR)} \tag{1.2}$$

Hence diversity gain is given by the negative of slope of frame error rate for a fixed transmission rate plotted as a function of SNR on a log–log scale. For a given *r*, the *optimal diversity gain,* d_{opt} *(r),* is the supreme diversity gain that can be accomplished by any MIMO system. It is shown (L. Zheng et al., 2003) that if the fading block length, $T \geq N_T + N_R - 1$, then, the optimal diversity gain can be calculated as

$$d_{opt} = (N_T - r)(N_R - r), 0 \leq r \leq \min(N_T, N_R) \tag{1.3}$$

The maximum value of rate gain *r* is always the minimum of (N_T, N_R) since we can have that many parallel data streams only. Note that if one employ entire transmit and receive antennas for enhancing diversity then one may achieve full diversity gain $N_T N_R$ (r = 0, it means we are not using any antenna for rate gain). Instead one may also employ a few antennas to augment data rate sacrificing the diversity gain.

For instance, we may consider the following case study for diversity-multiplexing trade-off.

Example 1.1

Assume that the multiplexing gain, *(r),* and diversity gain, *(d),* satisfy the diversity-multiplexing trade-off $d_{opt} = (N_T - r)(N_R - r)$ for SNR→∞. Assume $N_T = N_R = 7$ MIMO system with an SNR of 10 dB, one needs a spectral efficiency of R = 16 bps per Hertz. Find the supreme diversity gain such MIMO system can achieve.

Solution

Note that Shannon's channel capacity in bits/sec/Hz for a SISO link is $\log_2(1 + SNR)$. For high SNR case, it is approximately $\log_2(SNR)$. Our spatial multiplexing MIMO system here is equivalent to *r* parallel SISO channels (*r* parallel Gaussian channels) and its capacity is $r \log_2(SNR)$. With SNR = 10 dB, to get R = 16 bps, we require $r \log_2(SNR) = R$ which implies that $r \log_2(10^{1.0}) = 16$. Hence, r = 4.8165. Therefore five antennas may be used for multiplexing and remaining (7-2) two antennas may be used for diversity. The maximum diversity gain can be calculated as, $d_{opt} = (N_T - r)(N_R - r)$ = (7 – 5) (7 – 5) = 4. This means we are sending data over five parallel data streams only and we are utilizing four paths/links for decreasing the probability of error in detection.

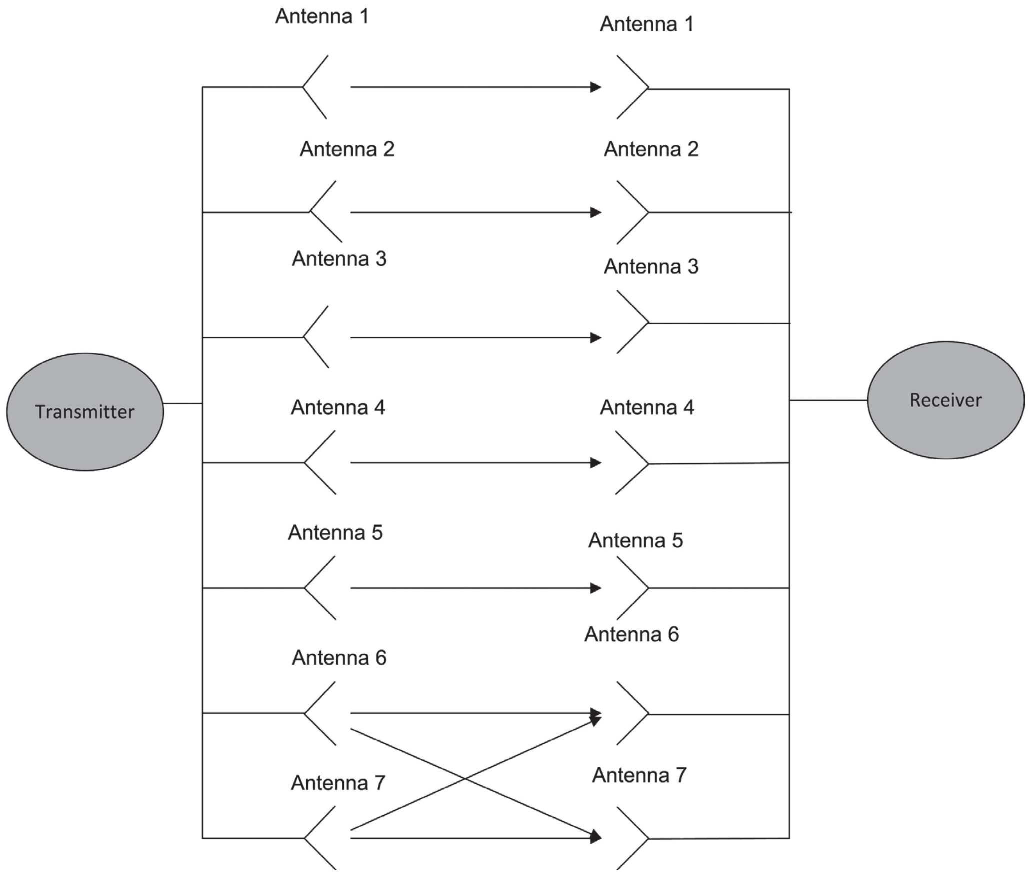


Fig. 1.3 Illustration of diversity-multiplexing trade-off of Example 1.1

Table 1.1
Diversity-multiplexing trade-offs for a 7×7 MIMO system

Serial No.	*Rate gain (r)*	*Diversity gain (d)*
1.	2	25
2.	3	16
3.	4	9
4.	5	4
5.	6	1

Similarly, there are other possible MIMO system designs as shown in Table 1.1.

1.7 Applications of MIMO systems

MIMO has been accepted in numerous wireless standards such as IEEE 802.11n (K.-L. Du et al., 2010 and N. Costa et al., 2010). Note that IEEE 802.11 standard is for wireless local area network (WLAN) applications, IEEE 802.15 standard is for wireless personal area networks (WPAN) and IEEE 802.16 standard is for wireless metropolitan area networks (WMAN). MIMO–OFDM may be employed for next generation WLANs. OFDM stands for Orthogonal Frequency Division Multiplexing. OFDM splits the information stream into N parallel sub-streams which are then transmitted by modulating onto N distinct orthogonal sub-carriers. Basically, it converts a high data stream into a number of low-rate sub-streams that are transmitted over parallel, narrowband channels that can be easily equalized.

MIMO is also envisaged for use in Fifth generation (5G) cellular downlink and uplink. MIMO–OFDM is used in wideband code division multiple access (WCDMA), CDMA 2000 (3 G mobile technology standards), IEEE 802.11n, IEEE 802.16m and 4G Long-term evolution (LTE). Spectral efficiency of 2–3 bits/sec/Hz is available in present cellular and WLAN systems. Bell Labs layered space time (BLAST) coding can achieve spectral efficiency of 42 bits/sec/Hz (B. Vucetic et al., 2003). MIMO-based Wi-Fi and Wireless Interoperability of Microwave Access (WiMAX) (IEEE 802.16 standard) systems are available in the market whereas MIMO-based High speed packet access (HSPA+) and Long term evolution (LTE) are in offing (A. Sibille et al., 2010). Nowadays, single user MIMO (base station to single subscriber and vice versa) is expanding into multiuser MIMO (base station to multiple users), network MIMO (multi-base station to single user), large-scale MIMO (with hundreds or thousands of transmitting and receiving antennas) and MIMO based cooperative communication and cognitive radios.

Review question 1.8 *What are the two gains achieved from MIMO systems?*

Review question 1.9 *What is diversity-multiplexing trade-off?*

Review question 1.10 *List some applications of MIMO systems.*

1.8 Summary

Figure 1.4 shows the chapter in a nutshell. In this chapter, we have discussed about wireless fading channels and diversity techniques. In fading channels, we have defined frequency-flat, frequency-selective, fast and slow fading channels. We have discussed briefly time, frequency and space diversity. We have defined SISO, SIMO, MISO and MIMO systems. In receiver diversity, we have mentioned about SC, EGC and MRC schemes. In transmit diversity, we have defined close loop, open loop and blind MIMO systems. We have also studied about rate and diversity in MIMO systems and their trade-off. Finally, we have mentioned some applications of MIMO systems.

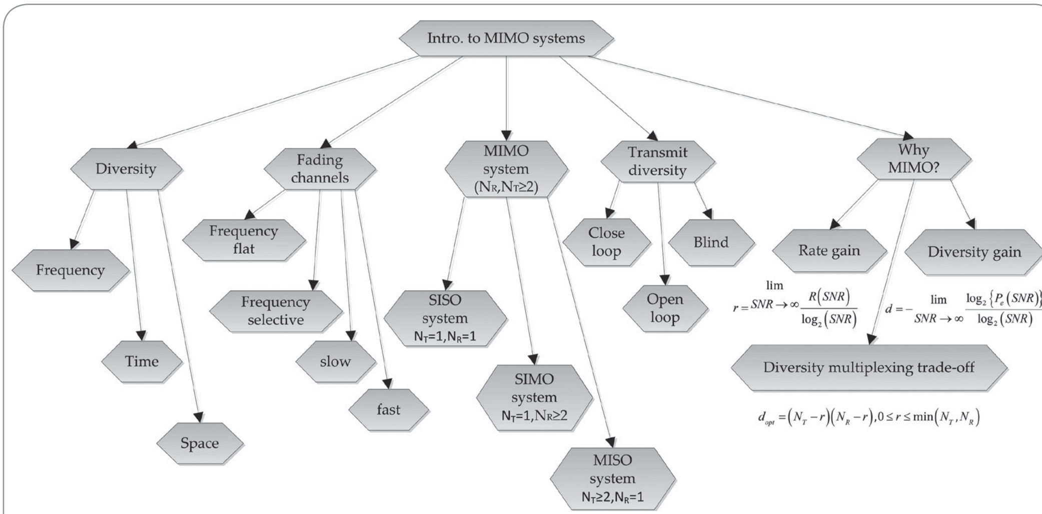


Fig. 1.4 Chapter 1 in a nutshell

Exercises

Exercise 1.1

Explain the two gains we can achieve from MIMO systems taking example for 4×4 MIMO system. Assume rich Rayleigh scattering environment.

Exercise 1.2

Define diversity and rate gain of a MIMO system.

Exercise 1.3

What are close loop, open loop and blind MIMO systems?

Exercise 1.4

Which diversity was left aside for many years? Why?

Exercise 1.5

What are frequency-flat, frequency-selective, fast and slow fading channels?

References

1. Andrews, J. G., A. Ghosh, and R. Muhamed. 2007. *Fundamentals of WIMAX*. Upper Saddle River: Prentice Hall.
2. Biglieri, E. and G. Taricco. 2004. *Transmission and Reception with Multiple Antennas: Theoretical Foundations*. Boston: Now Publishers.
3. Brown, T., E. D. Carvalho, and P. Kyritsi. 2012. *Practical Guide to the MIMO Radio Channel*. Chichester: John Wiley & Sons.
4. Costa, N. and S. Haykin. 2010. *Multiple-input Multiple-output Channel Models*. New Jersey: John Wiley & Sons.
5. Cover, T. M. and J. A. Thomas. 1999. *Elements of Information Theory*. New Jersey: Wiley.
6. Du, K.-L. and M. N. S. Swamy. 2010. *Wireless Communications*. New Delhi: Cambridge University Press.
7. Goldsmith, A. 2005. *Wireless Communications*. New Delhi: Cambridge University Press.
8. Haykin, S. and M. Moher. 2005. *Modern Wireless Communications*. New Delhi: Pearson.
9. Huang, H., C. B. Papadias, and S. Venkatesan. 2012. *MIMO Communication for Cellular Networks*. New York: Springer.
10. Jagannatham, A. K. 2016. *Principles of Modern Wireless Communications*. New Delhi: McGraw Hill Education (India) Private Limited.
11. Janaswamy, R. 2001. *Radiowave Propagation and Smart Antennas for Wireless Communications*. New York: Kluwer Academic Publishers.
12. Molisch, A. *Wireless Communications*. 2005. New Delhi: John Wiley & Sons.
13. Proakis, J. G. and M. Salehi. 2007. *Digital Communications*, New York: McGraw Hill.
14. Sibille, A., C. Oestges, and A. Zanella. 2011. *MIMO from Theory to Implementation*. Oxford: Elsevier Academic Press.
15. Tse, D. and P. Viswanath. 2005. *Fundamentals of Wireless Communications*. Cambridge: Cambridge University Press.
16. Vucetic, B. and J. Yuan. 2003. *Space-time Coding*. Chichester: John Wiley & Sons.
17. Yuen, C., Y. L. Guan, and T. Tjhung. 2007. *Quasi-orthogonal Space-time Block Codes*. London Imperial College Press.
18. Zheng, L. and D. N. Tse. May 2003. 'Diversity and multiplexing: A fundamental trade-off in multiple antenna channels'. *IEEE Trans. Information Theory* 1073–96.

Classical and Generalized Fading Distributions

CHAPTER 2

2.1 Introduction

In the next chapter, we will discuss about MIMO channel models. We will first discuss about fading distributions, a precursor to MIMO channel models. Fading distributions could be divided into two kinds:

(a) classical fading distributions
(b) generalized fading distributions

In classical fading distributions, we will find the probability density function (pdf), cumulative distribution function (cdf) and moment generating function (mgf) of Gaussian, Rayleigh, Rice, Chi-squared and Nakagami-m fading distributions. Among the many generalized fading distributions, we will investigate k-μ, α-μ and η-μ fading distributions and the above classical fading distributions are particular cases of these generalized fading distributions.

2.2 Introduction to fading distributions

Let us suppose a MIMO system consisting of N_T transmit antennas and N_R receive antennas. The received signal vector, $\mathbf{y}(l)$, at discrete time l may be expressed in terms of the transmitted signal vector, $\mathbf{x}(l)$, by

$$\mathbf{y}(l) = \mathbf{H}(l) * \mathbf{x}(l) + \mathbf{n}(l) \tag{2.1}$$

Here, * is the convolution operation, $\mathbf{y}(l)$ is an $N_R \times 1$ vector and $\mathbf{x}(l)$ is an $N_T \times 1$ vector. $\mathbf{n}$(l) is an $N_R \times 1$ additive white Gaussian noise (AWGN) vector and $\mathbf{H}(l)$ is the channel matrix, representing the channel impulse response at any discrete time l. Note that for a $N_T \times N_R$ MIMO system, $\mathbf{H}(l)$ becomes an $N_R \times N_T$ dimensional matrix. For a frequency-flat fading channel, above equation may be expressed as (see section 2.2 of N. Costa et al., 2010)

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \tag{2.2}$$

It is also popularly known as narrowband MIMO channel model. For narrowband MIMO channel, the channel coefficient between each transmit antenna and each receive antenna is a complex RV. For a frequency-flat channel (B. S. Paul et al., 2008), the components of the channel matrix are expressed as:

$$h_{ij} = \alpha_{ij} \exp\left(j\varphi_{ij}\right) \tag{2.3}$$

where, h_{ij} represents the channel gain for the signal path from the j^{th} transmit antenna to the i^{th} receive antenna.

α_{ij} and φ_{ij} represent the channel gains and phase shifts for the signal link from j^{th} transmit antenna to the i^{th} receive antenna, respectively. The phase is usually modelled as uniform random variable which may not be always true. The distribution of α_{ij} depends on the environment. We will consider various fading distributions for $\alpha_{ij.}$

Cumulative distribution function (cdf):

The cdf of a random variable (RV) X is defined as

$$P_X(x) = P\left[X \leq x\right] = \int_{-\infty}^{x} p_U\left(u\right) du$$

Note that $P_X(x)$ and $p_X(x)$ denotes cdf and pdf of RV X.

Moment generating functions (mgf):

The mgf of a RV X is defined as $M_X(s) = E[\exp(sX)]$ which can be expressed as

$$M_X(s) = \int_{-\infty}^{\infty} \exp\left(sx\right) p_X\left(x\right) dx$$

It can be shown that the n^{th} derivative of mgf for $x = 0$ is equal to the n^{th} moment of the RV X. Hence mgf generates moments of a RV.

$$M_X^n\left(x\right)\Big|_{x=0} = E\left[X^n\right]; n \geq 1$$

Characteristic function (cf):

The cf of a RV X is defined as $C_X(\omega) = E[\exp(j\omega X)]$, which can be expressed as

$$C_X(\omega) = \int_{-\infty}^{\infty} \exp\left(j\omega x\right) p_X\left(x\right) dx$$

From mgf we can obtain cf by putting $s = j\omega$. Similarly from cf, we can obtain mgf by putting $j\omega = s$. From cf, we can also find the pdf by taking inverse transform (A. Papoulis et al., 2002) as follows.

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-j\omega x\right) C_X\left(\omega\right) dx$$

Review question 2.1 *What is frequency-flat fading channel?*

Review question 2.2 *The channel gain coefficients are complex and defined as $h_{ij} = \alpha_{ij} \exp(j\varphi_{ij})$. What is the most commonly accepted distribution for the phase φ_{ij}?*

Review question 2.3 *Define mgf and cf of a RV X. How are they related?*

2.3 Classical fading distributions

In classical fading distributions, we will discuss the following fading distributions which are widely used for various fading scenarios (G. L. Stuber et al., 2001, M. K. Simon et al., 2005 and J. G. Proakis et al., 2007).

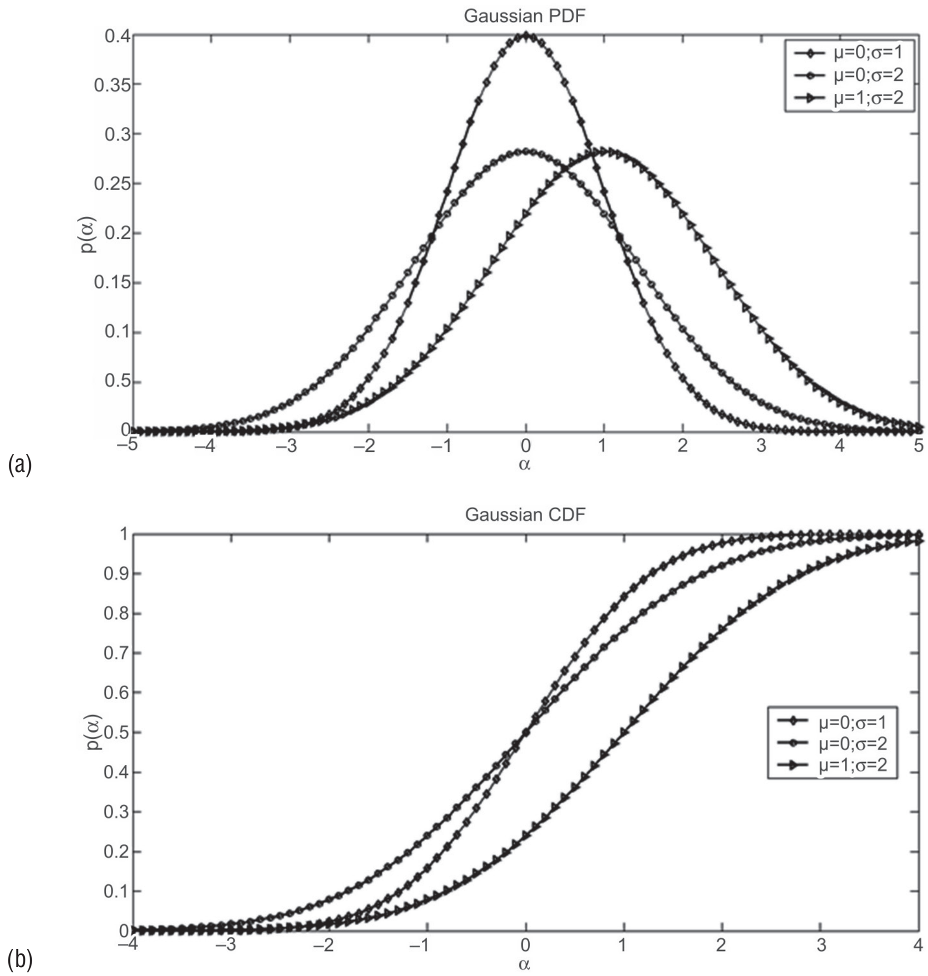


Fig. 2.1 Gaussian (a) Probability density function (pdf) and (b) Cumulative distribution function (cdf)

Review question 2.4 *What is the central limit theorem?*

Central limit theorem (M. Capinski et al., 2001): Let $\{X_n\}$ be a sequence of i.i.d. RVs with finite expectation $\mu = E(X_n)$ and finite variance $\sigma^2 > 0$. For $S_n = X_1 + X_2 + \ldots + X_n$, then

$$\underset{n\to\infty}{Lim}\, P\left[\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) \le x\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

In other words, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges in a distribution to a RV having the normal distribution with mean 0 and variance 1.

2.3.1 Gaussian fading distribution

Gaussian pdf and cdf are depicted in Fig. 2.1 for various values of means and variances. MATLAB commands are "normpdf" and "normcdf", respectively. It is one of most widely used distributions in wireless communications. According to the Central Limit theorem, the sum of numerous continuous random variables as the number increases, tends toward Gaussian distribution. A random variable X is said to have Gaussian distributed $X_G \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2 if it has pdf as

$$p_{X_G}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.4a}$$

Its cdf is given by $$p_{X_G}(x) = 1 - Q\left(\frac{x-\mu}{\sigma}\right) \tag{2.4b}$$

The cdf can be derived from pdf as follows.

$$\begin{aligned}
P_{X_G}(x) &= \int_{-\infty}^{x} p_{X_G}(u)du \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)du \\
&= 1 - \int_{x}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)du \\
&= 1 - \int_{\frac{x-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(v)^2}{2}\right)dv \\
&= 1 - Q\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}$$

Normal distribution is a particular case of Gaussian distribution for which $X_N \sim N(\mu = 0, \sigma^2 = 1)$ and its pdf and cdf are $p_{X_N}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ and $P_{X_N}(x) = 1 - Q(x)$, respectively. Note that, Q(−∞) = 1.

The mgf of normal distribution is

$$M_{X_N}(s) = E\left[e^{sX_N}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{sx} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2 - 2sx}{2}} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-s)^2}{2} + \frac{s^2}{2}} dx$$

$$= e^{\frac{s^2}{2}} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-s)^2}{2}} dx = e^{\frac{s^2}{2}}$$

Hence cf is $C_{X_N}(\omega) = E\left[e^{j\omega X_N}\right] = e^{-\frac{\omega^2}{2}}$.

Example 2.1

Find the mgf and cf of Gaussian RV $\left(X_G \sim N\left(\mu, \sigma^2\right)\right)$.

Solution

We can obtain a Gaussian RV $\left(X_G \sim N\left(\mu, \sigma^2\right)\right)$ from the normal RV $\left(X_N \sim N\left(\mu = 0, \sigma^2 = 1\right)\right)$ as $X_G = \mu + \sigma X_N$. Hence the mgf of Gaussian distribution can be obtained as

$$M_{X_G}(s) = E\left[e^{sX_G}\right] = E\left[e^{s(\mu + \sigma X_N)}\right] = e^{s\mu} E\left[e^{s\sigma X_N}\right]$$

$$= e^{s\mu} M_{X_N}(s\sigma) = e^{s\mu} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{s\sigma x} e^{-\frac{x^2}{2}} dx$$

$$= e^{s\mu} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2 - 2s\sigma x}{2}} dx$$

$$= e^{s\mu} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x - s\sigma)^2}{2} + \frac{(s\sigma)^2}{2}} dx$$

$$= e^{\frac{2s\mu + (s\sigma)^2}{2}} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x - s\sigma)^2}{2}} dx = e^{\frac{2s\mu + (s\sigma)^2}{2}}$$

Therefore, mgf and cf of Gaussian distribution are

$$M_{X_G}(s) = e^{\frac{2s\mu+(s\sigma)^2}{2}} \tag{2.4c}$$

$$C_{X_G}(\omega) = E\left[e^{j\omega X_G}\right] = e^{\frac{2j\omega\mu-(\omega\sigma)^2}{2}} = e^{j\omega\mu}e^{-\frac{\omega^2\sigma^2}{2}} \tag{2.4d}$$

A note on Additive White Gaussian Noise (AWGN):

Thermal noise is caused by thermal motion of electrons in all dissipative components- resistors, wires, and so on, which is completely random. Even though individual noise mechanisms might have different distributions, the aggregate of many such noise mechanisms will tend toward Gaussian distribution from Central limit theorem. It is additive means that the noise is added to the received signal and there is no multiplicative term here.

Define *X(t)* to be Gaussian random process (S. G. Wilson, 1996 and B. Sklar et al., 2009) with zero mean *E(X(t))=0* and autocorrelation function (acf) $E\left(X(t)X(t+\tau)\right) = \frac{N_0}{2}\delta(\tau)$.

- It is Gaussian means its pdf of n-samples is jointly n-order Gaussian.
- Each RV (sample) has zero mean.
- RVs (samples) at different time instants are uncorrelated since acf is a delta function.
- Since these RVs are jointly Gaussian, they are independent (for jointly Gaussian pdf, uncorrelated means independence).
- Note that power spectral density (psd) which is given by Fourier transform of the acf is equal to $\frac{N_0}{2}$ which implies that this process has equal power over all frequency spectrum.
- White light is composed of equal amount of all visible colours; here, spectrum is white because it constitutes of equal amount of power for all frequencies.
- Over the whole frequency spectrum noise power is infinite.
- But we are interested in finding the noise power within the bandwidth *(f, f+B)* which gives noise power as N_0B Watts = *kTB* Watts (*k* is the Boltzman's constant which is equal to 1.38 × 10^{-23} J/K and *T* is the temperature in Kelvin) by integrating over both the negative and positive frequency regions.
- The reason for ½ factor in the acf is conventional in communication engineering because it gets cancelled when we integrate the psd (also called as two-sided psd) over both the negative and positive frequency regions within the bandwidth.
- Note that one may find Gaussian process with non-white spectrum and a stochastic process with white spectrum but no Gaussian density functions.
- One should not use white and Gaussian process inter changeably.

Review question 2.5 *Explain about AWGN.*

2.3.2 Chi-squared distribution

If $X_1, \ldots, X_{2n}$ are i.i.d. zero mean and same variance σ^2 Gaussian RVs, then $X_{C_c} = \sum_{i=1}^{2n} X_i^2$ is a central Chi-squared RV with $2n$ degrees of freedom. Its pdf is given by

$$p_{X_{C_c}}(x) = \begin{cases} \dfrac{x^{n-1} e^{-\frac{x}{2\sigma^2}}}{2^n \Gamma(n)\, \sigma^{\frac{n}{2}}}, & x > 0 \\ 0, & otherwise \end{cases} \tag{2.5a}$$

The mgf of central Chi-squared RV can be obtained as follows:

$$M_{X_{C_c}}(s) = \left[E\left(e^{sX^2}\right)\right]^{2n}$$

where, $$E\left(e^{sX^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{sx^2} e^{-\frac{x^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{\left(1-2\sigma^2 s\right)x^2}{2\sigma^2}} dx$$

Substituting $\sqrt{\dfrac{1-2s\sigma^2}{2\sigma^2}}\, x = \dfrac{y}{\sqrt{2}}; dx = \dfrac{dy}{\sqrt{\dfrac{1-2s\sigma^2}{\sigma^2}}}$, we have,

$$E\left(e^{sX^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \frac{1}{\sqrt{\dfrac{1-2s\sigma^2}{\sigma^2}}} dy = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-2s\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{1-2s\sigma^2}}$$

Therefore,

$$M_{X_{C_c}}(s) = \left[\frac{1}{\sqrt{1-2s\sigma^2}}\right]^{2n} \tag{2.5b}$$

If $X_1, \ldots, X_{2n}$ are i.i.d. with different means μ_i and same variance σ^2 Gaussian RVs, then $X_{C_n} = \sum_{i=1}^{2n} X_i^2$ is a non-central Chi-squared RV with $2n$ degrees of freedom. Its pdf is given by

$$p_{X_{C_n}}(x) = \begin{cases} \dfrac{\left(\dfrac{x}{m^2}\right)^{\frac{2n-2}{4}} e^{-\frac{m^2+x}{2\sigma^2}} I_{n-1}\left(\dfrac{m}{\sigma^2}\sqrt{x}\right)}{2\sigma^2}, & x > 0 \\ 0, & otherwise \end{cases} \tag{2.5c}$$

where, $m = \sqrt{\sum_{i=1}^{2n} \mu_i^2}$.

Example 2.2

Find the mgf and cf of non-central Chi-squared RV.

Solution

The mgf of non-central Chi-squared RV can be obtained as follows:

$$M_{X_{Cn}}(s) = \left[E\left(e^{sX^2}\right)\right]^{2n}$$

$$E\left(e^{sX^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{sx^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\left\{\frac{\left(1-2s\sigma^2\right)x^2+\mu^2-2\mu x}{2\sigma^2}\right\}}dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\left\{\frac{\sqrt{1-2s\sigma^2}\,x}{\sqrt{2\sigma^2}}-\frac{\mu}{\sqrt{2\sigma^2}\sqrt{1-2s\sigma^2}}\right\}^2+\left(\frac{\mu^2}{2\sigma^2}-\left(\frac{\mu}{\sqrt{2\sigma^2}\sqrt{1-2s\sigma^2}}\right)^2\right)}dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{s\mu^2}{1-2s\sigma^2}}\int_{-\infty}^{\infty} e^{-\left\{\frac{\sqrt{1-2s\sigma^2}\,x}{\sqrt{2\sigma^2}}-\frac{\mu}{\sqrt{2\sigma^2}\sqrt{1-2s\sigma^2}}\right\}^2}dx$$

Substituting $\sqrt{1-2s\sigma^2}\,x - \frac{\mu}{\sqrt{1-2s\sigma^2}} = y;\, dx = \frac{dy}{\sqrt{1-2s\sigma^2}}$, we have,

$$E\left(e^{sX^2}\right) = \frac{1}{\sqrt{1-2s\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{s\mu^2}{1-2s\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}}dy$$

$$= \frac{1}{\sqrt{1-2s\sigma^2}}e^{\frac{s\mu^2}{1-2s\sigma^2}}$$

Therefore,

$$M_{X_{C_n}}(s) = \left(\frac{1}{1-2s\sigma^2}\right)^n e^{\frac{sm^2}{1-2s\sigma^2}} \tag{2.5d}$$

From mgf, cf can be obtained as

$$C_{X_{C_n}}(\omega) = \left(\frac{1}{1-2j\omega\sigma^2}\right)^n e^{\frac{j\omega m^2}{1-2j\omega\sigma^2}} \tag{2.5e}$$

2.3.3 Rayleigh fading distribution

For non-line-of-sight (NLOS) propagation between the transmitter and receiver, the transmitted signal reaches the receiver after being reflected (smooth and large surface), diffracted (sharp edges), refracted (from one medium to another) and scattered (rough surfaces) from diverse obstacles (e.g., building, forests, hills, etc.) adjacent to the receiver. Hence numerous copies of the transmitted signal arrived at the receiver from multitude directions, with varied delays and phase shifts. Generally it has been assumed to be a complex Gaussian random process. The amplitude distribution for this random process is modelled as Rayleigh distribution. The pdf of Rayleigh fading distribution is given by

$$p_{\alpha_{Ra}}(\alpha) = \frac{\alpha}{\sigma^2} e^{-\frac{\alpha^2}{2\sigma^2}}; \alpha \geq 0 \tag{2.6a}$$

The cdf of Rayleigh fading distribution is given by

$$\begin{aligned} P_{\alpha_{Ra}}(\alpha) &= \int_{-\infty}^{\alpha} p_{\beta_{Ra}}(\beta) d\beta \\ &= \int_{0}^{\alpha} \frac{\beta}{\sigma^2} e^{-\frac{\beta^2}{2\sigma^2}} d\beta \\ &= \begin{cases} 1 - e^{-\frac{\alpha^2}{2\sigma^2}}, & \alpha > 0 \\ 0, & otherwise \end{cases} \end{aligned} \tag{2.6b}$$

The Rayleigh distribution is the envelope of two i.i.d. Gaussian RVs (see Exercise 2.1).

$$\alpha = \sqrt{X_1^2 + X_2^2}; X_1 \sim N(0, \sigma^2); X_2 \sim N(0, \sigma^2)$$

The mean and variance are given by

$$E(\alpha) = \sigma\sqrt{\frac{\pi}{2}}; Var(\alpha) = \left(2 - \frac{\pi}{2}\right)\sigma^2$$

The mgf of Rayleigh distribution is given by

$$\begin{aligned} M_{X_{Ra}}(s) &= {}_1F_1\left(1, \frac{1}{2}; \frac{1}{2}s^2\sigma^2\right) + \sqrt{\frac{\pi}{2}} s\sigma e^{\frac{s^2\sigma^2}{2}} \\ &= -e^{\frac{1}{2}s^2\sigma^2} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}s^2\sigma^2\right)^k}{(2k-1)k!} + \sqrt{\frac{\pi}{2}} s\sigma e^{\frac{s^2\sigma^2}{2}} \end{aligned}$$

In the above equation, ${}_1F_1(a, b; x)$ is the confluent hypergeometric function (G. B. Arfken et al., 2005) which is defined in

(a) infinite series form $${}_1F_1(a,b;x) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b)x^k}{\Gamma(a)\Gamma(b+k)k!}, b \neq 0,-1,-2,\cdots$$

(b) definite integral form $${}_1F_1(a,b;x) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)}\int_0^1 e^{xt}t^{a-1}(1-t)^{b-a-1}\,dt$$

Rayleigh pdf and cdf are depicted in Fig. 2.2 for various values of means and variances. MATLAB commands are “raylpdf” and “raylcdf”, respectively.

Review question 2.6 *Write down the infinite series and definite integral form of confluent hypergeometric function.*

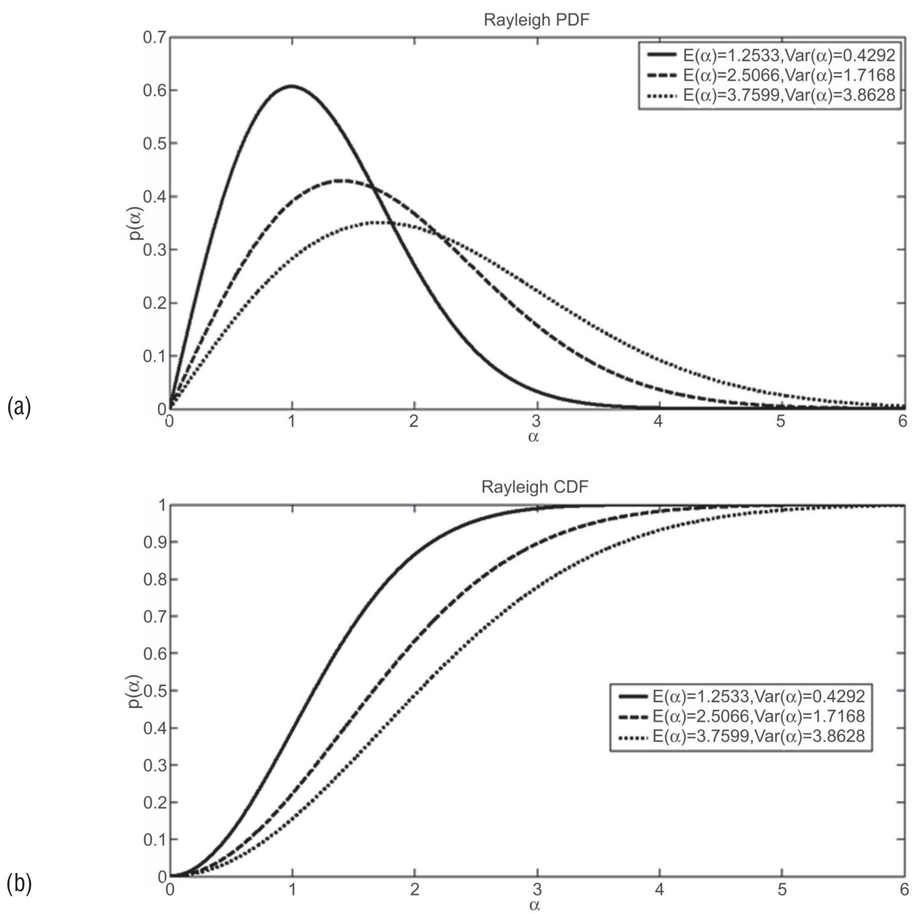


Fig. 2.2 Rayleigh (a) pdf and (b) cdf

2.3.4 Rice fading distribution

If a strong specular or line-of-sight (LOS) path is present in a highly scattered Rayleigh fading environment, the fading behaves like Ricean fading. The pdf of Rice fading distribution is given by

$$p_{\alpha_{Ri}}(\alpha) = \frac{2(1+K)e^{-K}\alpha}{\Omega} e^{-\frac{(1+K)\alpha^2}{\Omega}} I_0\left(2\sqrt{K}\alpha\sqrt{\frac{1+K}{\Omega}}\right); \alpha \geq 0 \qquad (2.7a)$$

It can be shown that the Rice distribution is the envelope $\sqrt{X_1^2 + X_2^2}$ of two independent and identically distributed (i.i.d.) Gaussian random variables (RVs) $X_1 \sim N(m_1, \sigma^2)$ and $X_2 \sim N(m_2, \sigma^2)$ where, $K = \frac{m_1^2 + m_2^2}{2\sigma^2}; \Omega = m_1^2 + m_2^2 + 2\sigma^2$ (see Exercise 2.1). The notation $N(m_1, \sigma^2)$ means Gaussian distribution with mean, m_1, and variance, σ^2. Rice pdf and cdf are depicted in Fig. 2.3 for various values of Ricean parameter K. K = 0 looks like that of Rayleigh distribution. MATLAB commands are "ricepdf" and "ricecdf", respectively.

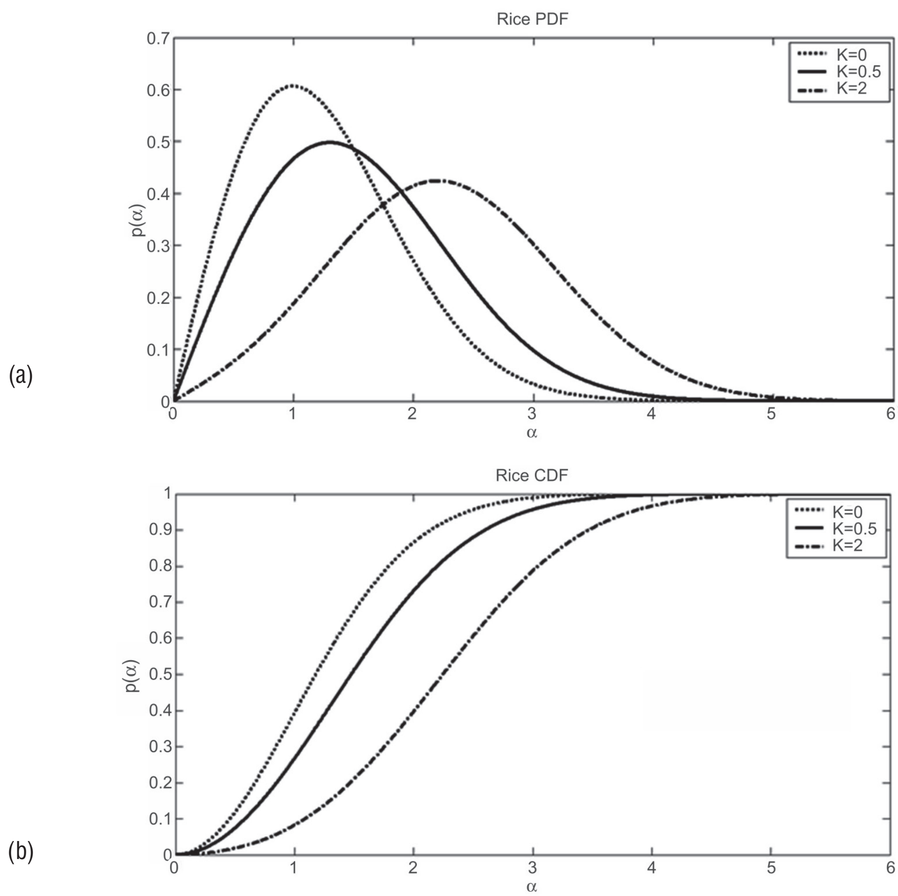


Fig. 2.3 Rice (a) pdf and (b) cdf ($m_1 = 0,1,2$; $m_2 = 0$; $\sigma^2 = 1$)

Example 2.3

Write a short MATLAB function to estimate Rice random variable (RV) pdf, cdf, mean, variance and standard deviations.

Solution

Note that we could model Rice MIMO channel as

$$\mathbf{H}_{Rice} = \sqrt{\frac{K}{1+K}}\mathbf{H}_{LOS} + \sqrt{\frac{1}{1+K}}\mathbf{H}_{Rayleigh} \tag{2.7b}$$

Some possible examples of LOS channel matrix for a 2×2 MIMO system are

$$\mathbf{H}_{LOS} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{H}_{LOS} = \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}$$

$$\mathbf{H}_{LOS} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

We will assume all LOS components give channel coefficients of 1.

Rice pdf, cdf, mean, variance and standard deviation:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clc;
close all;
clear all;
n=100000;%number of samples
x=randn(1,n);
y=randn(1,n);
r1=sqrt(x.^2+y.^2)/sqrt(2);
K_dB=10;
K = 10^(K_dB/10);
r = sqrt(K/(K+1)) + sqrt(1/(K+1))*r1;
N=floor(n/100);%number of bins
A=min(r); B=max(r);%range of sample values
Delta=(B-A)/N;% bin width
t=A-Delta/2+[1:N]*Delta;% horizontal axis of bin midpoints
f=hist(r,t)/(Delta*n);%Histogram (vertical axis of density estimates)
subplot(2,1,1);
bar (t,f);%Bar graph of estimated pdf
get(gca);%returns the handle of the current selected plot
set(gca,'fontsize',14);%sets the font size
set(get(gca,'children'),'linewidth',2);%sets the line thickness
xlabel('{\alpha}'); ylabel('p({\alpha})')
```

```
title('Estimated Rice pdf')
p=Delta*f;
CDF=cumsum(p);
subplot(2,1,2);
plot(t,CDF);
get(gca);%returns the handle of the current selected plot
set(gca,'fontsize',14);%sets the font size
set(get(gca,'children'),'linewidth',2);%sets the line thickness
xlabel('{\alpha}'); ylabel('P({\alpha})')
title('Estimated Rice CDF')
disp('The mean of the Rice RV');
mean(r)
disp('The variance of the Rice RV');
var(r)
disp('The standard deviation of the Rice RV');
std(r)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
The mean of the Rice RV
ans =
    1.2203
The variance of the Rice RV
ans =
    0.0194
The standard deviation of the Rice RV
ans =
    0.1391
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

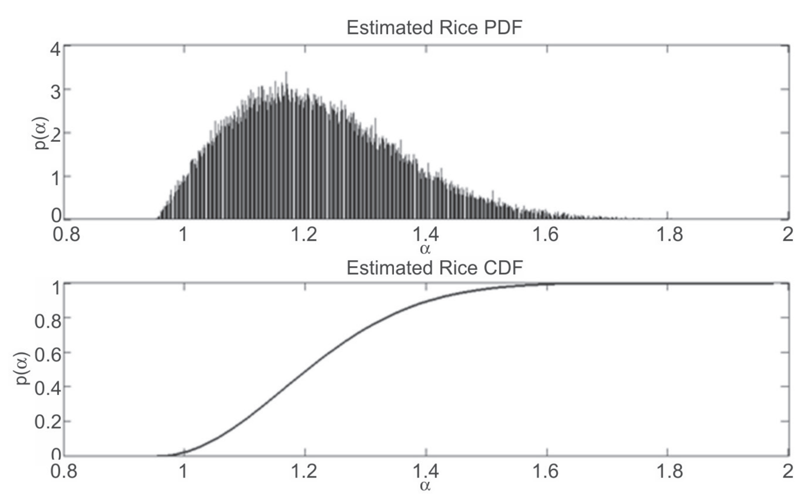


Fig. 2.4 Estimated Rice pdf and cdf

The above example shows how to estimate pdf and cdf from randomly generated samples of data. This may be of use in channel simulations where we generate random samples and check what we are generating is appropriate for a particular fading distribution like k-μ, η-μ and α-μ by estimating their pdf and cdf. We will discuss k-μ, η-μ and α-μ distributions in the next section.

2.3.5 Nakagami-m fading distribution

The Nakagami-m fading is a general model obtained from experimental data fitting. The advantage of Nakagami-m fading is that it can model a wide range of fading statistics by adjusting its fading parameter m. The probability density function (pdf) of Nakagami-m fading distribution is given by

$$p_\alpha(\alpha) = \frac{2m^m \alpha^{2m-1}}{\Omega^m \Gamma(m)} e^{-\frac{m\alpha^2}{\Omega}}; \alpha \geq 0; m \geq \frac{1}{2} \tag{2.8}$$

where, $\Omega = E(\alpha^2)$

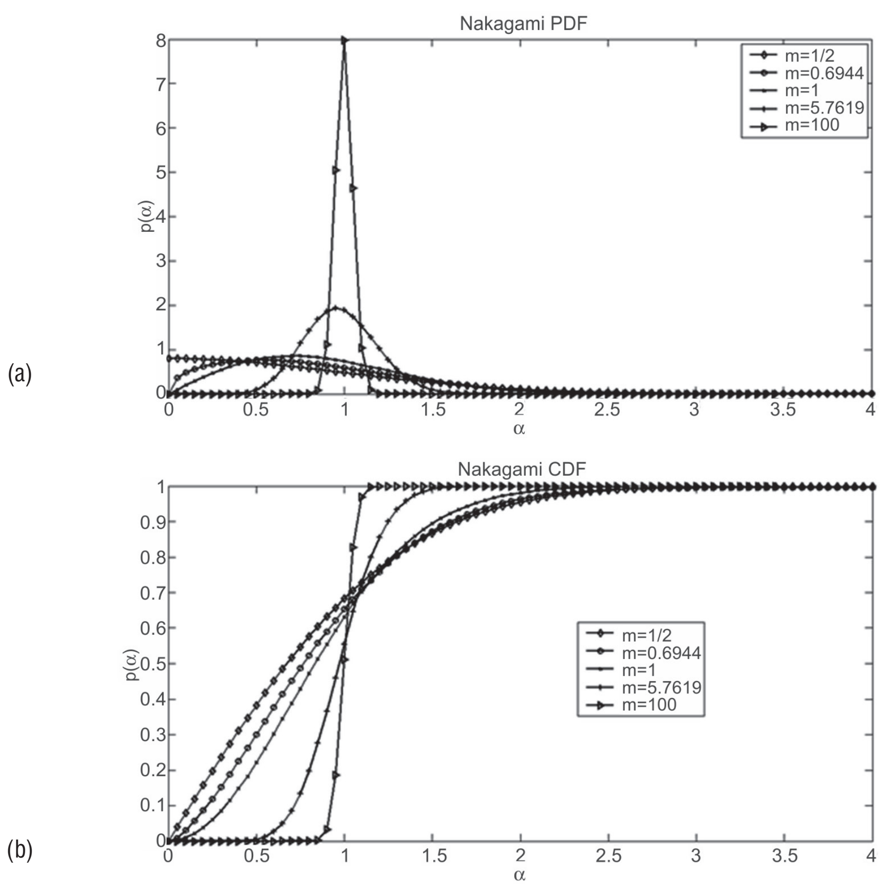


Fig. 2.5 Nakagami (a) pdf and (b) cdf for $\Omega = 1$

Usually the channel gains are modelled to be Nakagami-m distributed, since it encompasses both Rayleigh and Rician distributions by varying the m parameter. The smaller the m, the more severe is the fading. m=1 corresponds to Rayleigh fading, large value of m looks like log–normal fading and $m \to \infty$ is for non-fading cases. The fading parameter m varies from ½ to ∞ with m=½ (Gaussian), m=1 (Rayleigh) and $m \to \infty$ (non-faded AWGN).

From Fig. 2.5, it can be observed that the pdf and cdf for (a) m=1/2 looks like that of the Gaussian distribution (b) m=1 is similar to that of Rayleigh distribution.

Review question 2.7 *What is the mgf for central and non-central Chi-squared distribution?*

Review question 2.8 *When do we use Rice fading model?*

Review question 2.9 *When do we use Rayleigh fading model?*

Review question 2.10 *What are two parameters which describe Nakagami fading?*

Recently three generalized fading distributions have come up which are superset of the existing classical fading distributions. In other words, existing classical distributions are subset or particular or special cases of these three generalized fading distributions. *In these fading distributions, fading is generally considered as composed of many clusters of multipaths or rays.* Usually classical fading distributions consider only one cluster of multipaths or rays. Note that multipath waves within any one cluster the phases of scattered waves are random and have similar delay times with delay-time spreads of different clusters being relatively large.

2.4 Generalized fading distributions

These are three recent generalized fading distributions (M. D. Yacoub, 2007) viz.,

1. k-μ,
2. α-μ and
3. η-μ fading distributions.

They include the classical fading distributions as special cases.

All three generalized distributions will consider clusters of multipath waves. Within any one cluster, the phases of the scattered waves are random and have similar delay times with delay-time spreads of different clusters being relatively large. Depending on which environment in which it is propagating, they may be divided into two groups. Two kinds of environments will be considered:

(i) homogeneous environment (k-μ)
(ii) non-homogenous environment (α-μ and η-μ)

2.4.1 k-μ fading distributions

The clusters of multipath waves are assumed to have the scattered waves with identical powers *but within each cluster a dominant component is found that presents an arbitrary power.* Hence, it is better suited for line-of-sight (LOS) signal propagation.

The pdf of k-μ distributed random variable (RV) is given by

$$p_{\alpha_{k-\mu}}(\alpha_l) = \frac{2\mu(1+k)^{\frac{\mu+1}{2}}\alpha_l^{\mu} e^{-\frac{\mu(1+k)}{\Omega_l}\alpha_l^2}}{k^{\frac{\mu-1}{2}} e^{\mu k}\Omega_l^{\frac{\mu+1}{2}}} I_{\mu-1}\left(2\mu\sqrt{\frac{k(1+k)}{\Omega_l}}\alpha_l\right) \tag{2.9}$$

In the above equation,

(a) $\Omega_l = E(\alpha_l^2)$, k > 0 and μ > 0 are the main parameters of the distribution,
(b) $I_\upsilon(\cdot)$ represents the υ^{th} order modified Bessel function of the first kind,
(c) k is the ratio of the total power due to dominant components to the total power due to scattered waves and
(d) μ represents the number of clusters.

This fading distribution includes following classical fading distribution as special cases:

(i) Rice (μ = 1 and k = K),

$$p_{\alpha_{k=K,\mu=1}}(\alpha_l) = \frac{2(1+K)e^{-K}\alpha_l e^{-\frac{(1+K)}{\Omega_l}\alpha_l^2}}{\Omega_l} I_0\left(2\sqrt{\frac{K(1+K)}{\Omega_l}}\alpha_l\right)$$

which is the pdf of Rice distribution (Eq. 2.7a)
μ = 1 means the number of clusters is one.

(ii) Nakagami-m (k → 0 and μ = m),
For small arguments,

$$I_{\mu-1}(y) \cong \frac{\left(\frac{y}{2}\right)^{\mu-1}}{\Gamma(\mu)}$$

Hence,

$$p_{\alpha_{k-\mu}}(\alpha_l) = \frac{2\mu(1+k)^{\frac{\mu+1}{2}}\alpha_l^{\mu} e^{-\frac{\mu(1+k)}{\Omega_l}\alpha_l^2}}{k^{\frac{\mu-1}{2}} e^{\mu k}\Omega_l^{\frac{\mu+1}{2}}\Gamma(\mu)} \left(\mu\sqrt{\frac{k(1+k)}{\Omega_l}}\alpha_l\right)^{\mu-1}$$

$$\Rightarrow \quad p_{\alpha_{k\to 0,\mu=m}}(\alpha_l) = \lim_{k\to 0} \frac{2m^m(1+k)^m \alpha_l^{2m-1} e^{-\frac{m(1+k)}{\Omega_l}\alpha_l^2}}{e^{mk}\Omega_l^m\Gamma(m)}$$

$$= \frac{2m^m \alpha_l^{2m-1} e^{-\frac{m}{\Omega_l}\alpha_l^2}}{\Omega_l^m\Gamma(m)}$$

which is the pdf of Nakagami-m distribution (equation 2.8).

(iii) Rayleigh (μ = 1 and k → 0)

For $m = \mu = 1$, in the above pdf we have the Rayleigh distribution.

$$p_{\alpha_{k\to 0,m=1}}(\alpha_l) = \frac{2\alpha_l e^{-\frac{\alpha_l^2}{\Omega_l}}}{\Omega_l} = \frac{\alpha_l e^{-\frac{\alpha_l^2}{2\sigma^2}}}{\sigma^2}; \Omega_l = 2\sigma^2$$

(iv) one sided Gaussian distribution (μ = 0.5 and k → 0).

$$p_{\alpha_{k\to 0,m=\frac{1}{2}}}(\alpha_l) = \sqrt{\frac{2}{\pi\Omega_l}} e^{-\frac{\alpha_l^2}{2\Omega_l}}$$

where, $\Omega_l = \sigma^2$

The mgf is usually found out for the instantaneous signal-to-noise ratio (SNR) $\gamma = \alpha_l^2$ (M. K. Simon et al., 2005). For the k-μ distribution, the pdf of instantaneous SNR is

$$p_{\gamma_{k-\mu}}(x) = \frac{\mu(1+k)^{\frac{\mu+1}{2}} x^{\frac{\mu-1}{2}} e^{-\frac{\mu(1+k)x}{\overline{\gamma}}}}{k^{\frac{\mu-1}{2}} e^{\mu k} \overline{\gamma}^{\frac{\mu+1}{2}}} I_{\mu-1}\left(2\mu\sqrt{\frac{k(1+k)x}{\overline{\gamma}}}\right); \overline{\gamma} = E(\gamma)$$

The mgf of instantaneous SNR for k -μ fading distribution (N. Ermolova, 2008) is given by

$$M_{\gamma_{k-\mu}}(s) = E\left(e^{-s\gamma}\right)$$

$$= \int_0^\infty e^{-s\gamma} p_{\gamma_{k-\mu}}(\gamma) d\gamma$$

$$= \left(\frac{\mu(1+k)}{\mu(1+k)+s\overline{\gamma}}\right)^{\mu} e^{\frac{\mu^2 k(1+k)}{\mu(1+k)+s\overline{\gamma}} - \mu k}$$

We can also deduce the mgf of classical fading distributions from the above mgf.

2.4.2 α-μ fading distributions

For this distribution,

(i) the clusters of multipath waves are assumed to have the scattered waves with identical powers.

(ii) the resulting envelope is obtained as a nonlinear function of the modulus of the sum of the multipath components.

The assumption made is that at a certain given point, the received signal comprises of an arbitrary n number of multipath components (MPCs), and the propagation scenario is such that the envelope of received signal is perceived as a nonlinear function of the modulus of the sum of these MPCs. Suppose that such a non-linearity is expressed by a power parameter $\alpha > 0$ thereby the emerging envelope R is

$$R^{\alpha} = \sum_{i=1}^{n}\left(X_i^2 + Y_i^2\right)$$

where, X_i and Y_i are mutually independent Gaussian processes, $[E\{X_i\} = E\{Y_i\} = 0;\ E(X_i^2) = E(Y_i^2) = \frac{\hat{r}^{\alpha}}{2n}]$ and α-root mean value, $\hat{r} = \sqrt[\alpha]{E\left(R^{\alpha}\right)}$.

Applying the accepted method of transformation of variables in probability theory, the pdf $f_R(r)$ of R is obtained as

$$f_R(r) = \frac{\alpha n^n r^{\alpha n-1}}{\hat{r}^{\alpha n}\Gamma(n)} e^{-n\frac{r^{\alpha}}{\hat{r}^{\alpha}}}$$

The probability density function of the α-μ signal (replacing $n = \mu$) is obtained as follows.

$$f_R(r) = \frac{\alpha \mu^{\mu} r^{\alpha\mu-1}}{\hat{r}^{\alpha\mu}\Gamma(\mu)} e^{-\mu\frac{r^{\alpha}}{\hat{r}^{\alpha}}} \tag{2.10}$$

One special case of α-μ distribution is

(a) Weibull distribution for $\mu = 1$ (one cluster).

$$f_R(r) = \frac{\alpha r^{\alpha-1}}{\hat{r}^{\alpha}} e^{-\frac{r^{\alpha}}{\hat{r}^{\alpha}}} = \alpha\beta r^{\alpha-1} e^{-\beta r^{\alpha}};\ \beta = \frac{1}{\hat{r}^{\alpha}} \tag{2.11}$$

Weibull fading statistics is best fit for mobile radio systems operating in 800/900 MHz. Note that in this case,

(i) Weibull distribution with $\alpha = 2$ is like Rayleigh distribution.

$$f_R(r) = \frac{2r}{\hat{r}^2} e^{-\frac{r^2}{\hat{r}^2}} = \frac{r}{\gamma^2} e^{-\frac{r^2}{2\gamma^2}};\ \gamma^2 = \frac{\hat{r}^2}{2}$$

(ii) Weibull distribution with $\alpha = 1$ behaves like negative exponential distribution.

$$f_R(r) = \frac{e^{-\frac{r}{\hat{r}}}}{\hat{r}} = \chi \exp\left(-\chi r\right);\ \chi = \frac{1}{\hat{r}}$$

(b) Another special case of α-μ distribution is Nakagami-m distribution for $\alpha = 2$.

$$f_R(r) = \frac{2\mu^{\mu} r^{2\mu-1}}{\hat{r}^{2\mu}\Gamma(\mu)} e^{-\mu\frac{r^2}{\hat{r}^2}} = \frac{2\mu^{\mu} r^{2\mu-1}}{\Omega^{\mu}\Gamma(\mu)} e^{-\mu\frac{r^2}{\Omega}};\ \Omega = \hat{r}^2$$

(i) Note that in this case, Nakagami-m distribution with μ = 1 is like Rayleigh distribution.

$$f_R(r) = \frac{2r}{\hat{r}^2} e^{-\frac{r^2}{\hat{r}^2}} = \frac{r}{\gamma^2} e^{-\frac{r^2}{2\gamma^2}};\ \gamma^2 = \frac{\hat{r}^2}{2}$$

(ii) Nakagami-m distribution with μ =1/2 behaves like one-sided Gaussian distribution.

$$f_R(r) = \sqrt{\frac{2}{\pi \hat{r}^2}} e^{-\frac{r^2}{2\hat{r}^2}}$$

2.4.3 η-μ fading distributions

The probability density function (pdf) of η-μ distributed RV is given by

$$p_{\alpha_{\eta-\mu}}(\alpha_l) = \frac{4\sqrt{\pi}\mu^{\mu+\frac{1}{2}} h^{\mu} \alpha_l^{2\mu} e^{-\frac{2\mu h}{\Omega_l}\alpha_l^2}}{\Gamma(\mu) H^{\mu-\frac{1}{2}} \Omega_l^{\mu+\frac{1}{2}}} I_{\mu-\frac{1}{2}}\left(\frac{2\mu H}{\Omega_l}\alpha_l^2\right) \tag{2.12}$$

The parameter μ denotes the number of clusters, h and H is dependent on the parameter η. The parameter η is characterized for two cases:

In *Case 1*, it is assumed *that the in-phase and quadrature phase components of the MPCs within each cluster are* ***independent*** *of each other and have* ***disparate average powers***. The parameter $\eta \in$ *(0, ∞)* is the ratio of these above mentioned two powers, and

$$h = \frac{(1+\eta)^2}{4\eta}; H = \frac{1-\eta^2}{4\eta}; \frac{H}{h} = \frac{1-\eta}{1+\eta}.$$

For $0 < \eta \leq 1$, we have $H \geq 0$ whereas for $0 < \eta^{-1} \leq 1$, we have $H \leq 0$. Since $I_v(-z) = (-1)^v I_v(z)$, the distribution is symmetric around $\eta = 1$; therefore, power distribution may be considered only within one of the regions.

In *Case 2,* assumption made is *that the in-phase and quadrature phase components of MPCs within each cluster are* ***correlated*** *and have* ***same powers***. The parameter $\eta \in (-1, 1)$ is the correlation coefficient between these components and

$$h = \frac{1}{1-\eta^2};\ H = \frac{\eta}{1-\eta^2};\ \frac{H}{h} = \eta.$$

Similar to format 1, for $0 < \eta \leq 1$, we have $H \geq 0$ whereas for $-1 < \eta \leq 0$, we have $H \leq 0$. Since $I_v(-z) = (-1)^v I_v(z)$, the distribution is symmetric around $\eta = 0$; therefore, power distribution may be considered only within one of the regions.

The pdf of the instantaneous SNR $\gamma = \alpha_l^2$ of the η-μ distribution is

$$p_{\gamma_{\eta-\mu}}(x) = \frac{2\sqrt{\pi}\mu^{\mu+\frac{1}{2}} h^{\mu} x^{\mu-\frac{1}{2}} e^{-\frac{2\mu x h}{\overline{\gamma}}}}{\Gamma(\mu) H^{\mu-\frac{1}{2}} \overline{\gamma}^{\mu+\frac{1}{2}}} I_{\mu-\frac{1}{2}}\left(\frac{2\mu H x}{\overline{\gamma}}\right); \overline{\gamma} = E(\gamma)$$

The mgf of instantaneous SNR for η-μ fading distribution (N. Ermolova, 2008) is given by

$$M_{\gamma_{\eta-\mu}}(s) = E\left(e^{-s\gamma}\right)$$

$$= \int_0^\infty e^{-s\gamma} p_{\gamma_{\eta-\mu}}(\gamma)\, d\gamma$$

$$= \left(\frac{4\mu^2 h}{\left(2(h-H)\mu + s\overline{\gamma}\right)\left(2(h+H)\mu + s\overline{\gamma}\right)} \right)^{\mu}$$

We can also deduce the mgf of classical fading distributions from the above mgf as follows.

1. Rayleigh fading $(\eta = 1, \mu = 0.5)$

$$M_{\gamma_{\eta=1,\mu=0.5}}(s) = \frac{1}{1 + s\overline{\gamma}}$$

2. Nakagami-m $\left(\eta = 1, \mu = \frac{m}{2}\right)$

For format 1, $\eta \to 1$, implies

$$h = \frac{(1+\eta)^2}{4\eta} = 1; H = \frac{1-\eta^2}{4\eta} = 0$$

For format 2,
$\eta \to 0$ implies

$$h = \frac{1}{1-\eta^2} = 1; H = \frac{\eta}{1-\eta^2} = 0$$

$$M_{\gamma_{\eta=1,\mu=\frac{m}{2}}}(s) = \left(\frac{m}{m + s\overline{\gamma}} \right)^{m}$$

3. Nakagami-q $\left(\eta = q^2, \mu = \frac{1}{2}\right)$

$$M_{\gamma_{\eta,\mu=0.5}}(s) = \left(\frac{h}{\left((h-H) + s\overline{\gamma}\right)\left((h+H) + s\overline{\gamma}\right)} \right)^{0.5}$$

For format 1,

$$h = \frac{(1+\eta)^2}{4\eta} = \frac{\left(1+q^2\right)^2}{4q^2}; H = \frac{1-\eta^2}{4\eta} = \frac{1-q^4}{4q^2}; h - H = \frac{q^2+1}{2}; h + H = \frac{1+q^2}{2q^2}$$

Therefore,

$$M_{\gamma_{\eta,\mu=0.5}}(s) = \left(\frac{\frac{\left(1+q^2\right)^2}{4q^2}}{\left(\frac{q^2+1}{2} + s\overline{\gamma}\right)\left(\frac{1+q^2}{2q^2} + s\overline{\gamma}\right)} \right)^{0.5}$$

$$= \left(\frac{\left(1+q^2\right)^2}{\left(1+q^2+2s\overline{\gamma}\right)\left(1+q^2+2q^2 s\overline{\gamma}\right)} \right)^{0.5}$$

$$\therefore \qquad M_{\gamma_{\eta=q^2,\mu=0.5}}(s) = \frac{1+q^2}{\left[\left(1+q^2\right)^2\left(1+2s\overline{\gamma}\right)+4q^2s^2\overline{\gamma}^2\right]^{0.5}}$$

It is more suitable for NLOS signal propagation. Hoyt fading (or Nakagami-q) is best fit for satellite links subject to strong atmospheric scintillation.

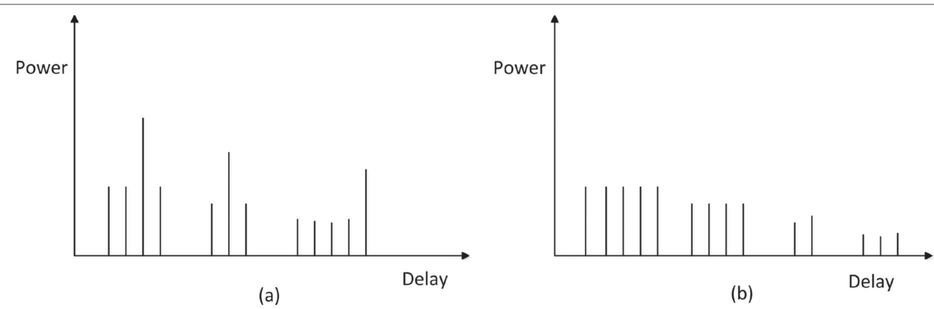


Fig. 2.6 Power fading profile

Review question 2.11 *What do you mean by generalized fading distributions?*

Review question 2.12 *Which generalized fading model is appropriate for the fading profile depicted in Fig. 2.6 (a) and (b) approximately?*

2.4.4 Simulating generalized fading distributions

Finding an explicit formula for $F^{-1}(y)$ for the cdf of a RV Y (Inverse cdf method) is not always possible. Besides, there may be more efficient method for generating the RV. The acceptance/rejection method is an accurate method for simulation of generalized fading RVs.

Example 2.4

Explain the basic idea of acceptance/rejection method with the help of an example.

Solution

Using the following procedure, we may generate a uniformly distributed random numbers Y between ¼ and 1.

Step 1: Generate a random number W.

Step 2a: If W≥ ¼, accept Y=W, go to step 3.

Step 2b: If W< ¼, reject W, return to step 1.

Step 3: If another uniform random variate in [1/4,1] is required, go to step 1.

Similar kind of things we also do in acceptance/rejection method for generating an arbitrary RV Y whose known pdf is $f(\rho)$. The basic idea is to find an alternate RV W with pdf $h(\rho)$ for which we already have an efficient algorithm to generate the RV and $h(\rho)$ is close to $f(\rho)$. What do you mean by close? It means that the ratio of $\frac{f(\rho)}{h(\rho)}$ is bounded by a constant C>0, i.e., $\sup_{\rho}\left\{\frac{f(\rho)}{h(\rho)}\right\} \le C$.

In practice, we want C to be close to 1 as possible. What is acceptance/rejection method? Suppose we wish to obtain samples of RV Y whose pdf is $f(\rho)$. We can do so by the following method:

Step 1: Sample another RV W with pdf $h(\rho)$ which can be generated easily and whose pdf is near to pdf of RV Y. Also sample a uniform RV U(0,1).

Step 2: Accept Y = W if $U \le \frac{f(\rho)}{Ch(\rho)}$ otherwise return to step 1. The constant C satisfies $\frac{f(\rho)}{h(\rho)} \le C, \forall \rho$. The probability of acceptance is 1/C.

Acceptance/Rejection method algorithm (S. Ross, 2002, A. Papoulis, 2002 and J. E. Gentle, 2005):
Suppose that we have a known method for simulating a RV W with pdf, $h(\rho)$. One may use this as a basis for simulating another RV Y with pdf $f(\rho)$. Suppose it is true that for a constant C, $\frac{f(\rho)}{h(\rho)} \le C, \forall \rho$, then one may use the two step procedures given below for generation of RV Y with pdf $f(\rho)$.

- Simulate W having pdf $h(\rho)$ and simulate a random number U.
- If $U \le \frac{f(\rho)}{Ch(\rho)}$, then set Y=W, otherwise go back to the previous step.

Review question 2.13 *What is acceptance/rejection method?*

Example 2.5

Explain the acceptance/rejection method for generating α-µ, η-µ and k-µ RVs.

Solution

Table 2.1

Values of a, b and ρ_0 for α-μ, k-μ and η-μ fading channels

Fading distributions	*k-μ*	*α-μ*	*η-μ*
a	$\mu(\kappa+1)$	$\frac{9}{2}\mu\sqrt[3]{\frac{3\mu-1}{3\mu}}$	$\frac{\mu(\eta+1)}{\eta}$
ρ_0	Solve $\frac{dp_{X_{\kappa-\mu}(\rho)}}{d\rho}=0$	$\sqrt[3]{\frac{3\mu-1}{3\mu}}$	Solve $\frac{dp_{X_{\eta-\mu}(\rho)}}{d\rho}=0$
b	$p_{X_{\kappa-\mu}(\rho_0)}$	$\frac{3\mu^{\mu}}{\Gamma(\mu)}\left(\frac{3\mu-1}{3\mu}\right)^{\frac{3\mu-1}{3}}e^{-\frac{3\mu-1}{3}}$	$p_{X_{\eta-\mu}(\rho_0)}$

To generate RV *X-Y* ($\alpha-\mu$, $\kappa-\mu$, $\eta-\mu$), acceptance/rejection method makes use of another RV *W* having pdf $h(\rho)$ similar to the pdf of *X-Y*, $f_P^{x-y}(\rho)$. To generate a deviate from a distribution with pdf $f_P^{x-y}(\rho)$, a RV *W* is chosen so that we can easily generate realizations of it and so that its pdf $h(\rho)$ can be scaled to majorize $f_P^{x-y}(\rho)$ using some constant *C*; that is, so that

$$g(\rho) = Ch(\rho) \geq f_P^{x-y}(\rho)\forall\rho$$

This process is also known as majorizing the density. The pdf $h(\rho)$ is called the majorizing density, and $g(\rho) = Ch(\rho)$ is called the majorizing or hat function. The majorizing hat function $g(\rho)$ for $\alpha-\mu$, $\kappa-\mu$, $\eta-\mu$ is given as (R. Cogliatti et al., 2012 and Q. M. Zhu et al., 2011)

$$g(\rho) = Ch(\rho) = be^{-a(\rho-\rho_0)^2} \geq f_P^{x-y}(\rho)$$

where, a, b and ρ_0 are tabulated in Table 2.1 and $f_P^{x-y}(\rho)$ is the desired pdf for $x-y=\alpha-\mu$, $\kappa-\mu$, $\eta-\mu$.

1/C is the probability of acceptance and C could be obtained as

$$C = \frac{b}{2}\sqrt{\frac{\pi}{a}}\left(1 + erf\left(\sqrt{a}\,\rho_0\right)\right)$$

where, erf is the error function.

2.5 Summary

Figure 2.7 shows the chapter in a nutshell. In this chapter, we have discussed about two types of fading distributions, viz. classical and generalized fading. In classical fading, we have considered Rayleigh, Rician, Weibull, Hoyt and Nakagami-m. In generalized fading, we have studied *k-μ*, *α-μ* and *η-μ*. Classical fading are usually a particular case of generalized fading. In order to characterize such fading distributions, we need to know their pdf, cdf, mgf and cf. In order to calculate the performance analysis of wireless communications over such fading channels, we are more interested in the pdf and mgf of the instantaneous SNR, which will be clearer from the ensuing chapters. Since classical fading are usually particular cases of generalized fading, one may evaluate the performance analysis of wireless communications over a generalized fading and deduce the performance over classical fading as particular cases. It unifies the performance analysis of wireless communications over fading channels significantly.

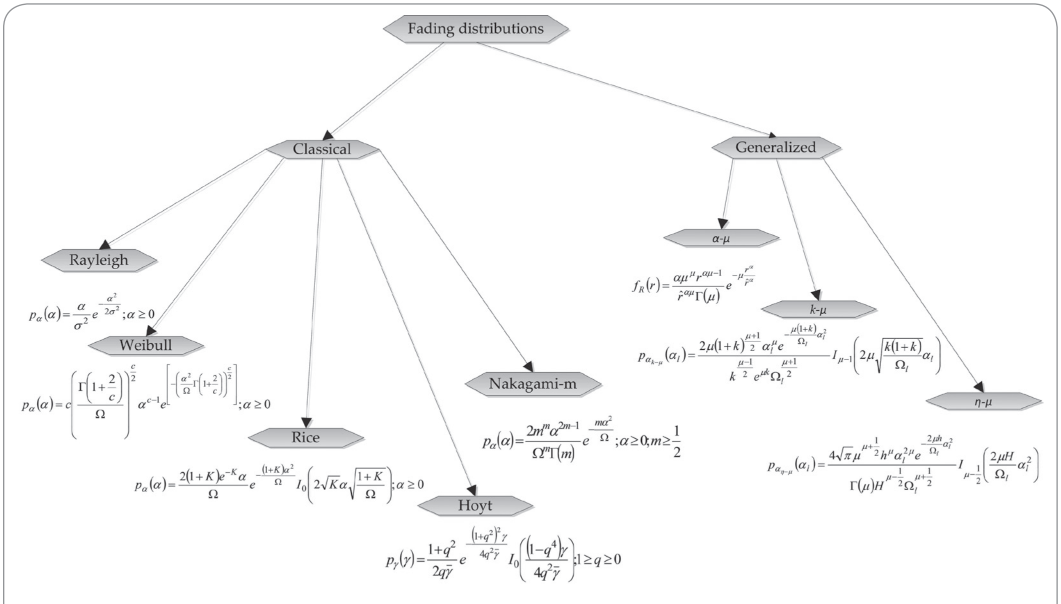


Fig. 2.7 Chapter 2 in a nutshell

Exercises

Exercise 2.1

Show that the root-sum-square value $\left(\alpha = \sqrt{X_1^2 + X_2^2}\right)$ of two independent Gaussian RVs (X_1, X_2) with mean $\mu = 0$ and variance σ^2 is Rayleigh distributed whose pdf is given by $p_{\alpha_{Ra}}(\alpha) = \frac{\alpha}{\sigma^2} e^{-\frac{\alpha^2}{2\sigma^2}}; \alpha \geq 0$. (Also note that if we allow X_1 to have mean $\mu \neq 0$ then we get the Rician distribution whose pdf is given by $p_{\alpha_{Ri}}(\alpha) = \frac{\alpha}{\sigma^2} I_0\left(\frac{\mu x}{\sigma^2}\right) e^{-\frac{\alpha^2+\mu^2}{2\sigma^2}}; \alpha \geq 0$)

Exercise 2.2

Find the values of (a) $Q(0)$ (b) $Q(\infty)$ and (c) $Q(-x)$

Exercise 2.3

What are the assumptions in Format I and II of η-μ fading distributions?
Following are MATLAB based exercises.

Exercise 2.4

Note that complementary error function is defined as $erfc(x)=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-z^2}dz$. Show that $Q(x)=\frac{1}{2}erfc\left(\frac{x}{\sqrt{2}}\right)$.

In MATLAB, one can plot Q function using 'erfc' command. Also plot the two bounds of Q function: (a) Chenoff's bound $Q(x) \le \frac{1}{2}e^{-\frac{x^2}{2}}$ and (b) Chiani's bound (M. Chiani et al., 2003) $Q(x) \le \frac{1}{12}e^{-\frac{x^2}{2}} + \frac{1}{4}e^{-\frac{2x^2}{3}}; x > 0.5$ and see how tight they are for different values of x.

Exercise 2.5

Write a short MATLAB function to estimate Rayleigh RV pdf, cdf, mean, variance and standard deviations.

Exercise 2.6

Write a short MATLAB program to generate α-μ RV.

Exercise 2.7

Write a short MATLAB program to generate k-μ RV.

Exercise 2.8

Write a short MATLAB program to generate η-μ RV.

Exercise 2.9

What is majorizing hat function $g(\rho)$ for $\alpha-\mu$, $\kappa-\mu$, $\eta-\mu$?

Exercise 2.10

What is the probability of acceptance in acceptance/rejection method?

References

1. Arfken, G. B. and H. J. Weber. 2005. *Mathematical Methods for Physicists*. New Delhi: Academic Press.
2. Capinski, M. and T. Zastawniak. 2001. *Probability through Problems*. New Delhi: Springer Verlag.

3. Chiani, M., D. Dardari, and M. K. Simon. 2003. 'New exponential bound and approximations for the computation of error probability in fading channels'. *IEEE Trans. Wireless Communications* 2(4): 840–845.
4. Cogliatti, R., R. A. A. de Souza, and M. D. Yacoub Nov. 2012. 'Practical, highly efficient algorithm for generating k-μ and η-μ variates and a near-100% efficient algorithm for generating α-μ variates'. *IEEE Communications Letters* 16(11): 1089–7798.
5. Costa, N. and S. Haykin. 2010. *Multiple-input Multiple-output Channel Models*. New Jersey: John Wiley & Sons.
6. Ermolova, N. 2008. 'Moment generating functions of the generalized η-μ and k-μ distributions and their applications to performance evaluations of communication systems'. *IEEE Communications Letters* 12(7): 502–504.
7. Gentle, J. E. 2005. *Random Number Generation and Monte Carlo Methods*. 2nd edition. New York: Springer.
8. Magableh, A. M. and M. M. Matalgah. June 2009. 'Moment generating function of the generalized distribution with applications'. *IEEE Comm. Lett.* 13(6): 411–413.
9. Papoulis, A. and S. U. Pillai. 2002. *Probability, Random Variables and Stochastic Processes*. New Delhi: Tata McGraw Hill.
10. Paul, B. S. and R. Bhattacharjee. Nov.–Dec. 2008. 'MIMO Channel Modelling: A Review'. *IETE Technical Review* 25(6): 315–319.
11. Proakis, J. G. and M. Salehi. 2007. *Digital Communications*. New York: McGraw Hill.
12. Ross, S. 2002. *A First Course in Probability*. New Delhi: Pearson.
13. Simon, M. K. and M.-S. Alouini. 2005. *Digital Communications over Fading Channels*. New Jersey: Wiley.
14. Sklar, B. and P. K. Ray. 2009. *Digital Communications: Fundamentals and Applications*. New Delhi: Pearson.
15. Stuber, G. L. 2001. *Principles of Mobile Communication*. Dordrecht: Kluwer Academic Publishers.
16. Wilson, S. G. 1996. *Digital Modulation and Coding*. New Delhi: Pearson.
17. Yacoub, M. D. Feb. 2007. 'The k-μ distribution and the η-μ distribution'. *IEEE Antennas Propagat. Mag.* 49(1): 68–81.
18. Zhu, Q. M., X. Y. Dang, D. Z. Xu, and X. M. Chen. Sep. 2011. 'Highly efficient rejection method for generating Nakagami-m sequences'. *Electron. Lett.* 47(19): 1100–1101.

Analytical MIMO Channel Models

CHAPTER 3

3.1 Introduction

In this chapter, we will discuss about analytical MIMO channel models. Analytical MIMO channel models can be divided into four types.

(a) independent and identically distributed (uncorrelated) MIMO channel model
(b) fully correlated MIMO channel model
(c) Kronecker (separately correlated) MIMO channel model
(d) keyhole (rank deficit) MIMO channel model

We will investigate independent and identically distributed (i.i.d.) MIMO channel model. In order to understand i.i.d. MIMO channel model we need to review some topics in complex random variable, complex random vector and complex random matrix. Here we will find the pdf of random matrix with i.i.d. elements. We will investigate joint distribution of eigenvalues of complex central Wishart random matrix and marginal distribution of an eigenvalue of complex central Wishart random matrix.

Practical MIMO channels are never uncorrelated. One needs to analyze MIMO system performance for correlated MIMO channel models. In fully correlated MIMO fading model, we need to consider all the co- and cross-correlations between all the channel coefficients for various channel paths between the transmitting antennas and receiving antennas. This covariance matrix (fourth order tensor) size and consequently the number of elements of the covariance matrix ($(N_R N_T)^2$) becomes prohibitively large as the number of transmitting and receiving antennas increase.

We will discuss about Kronecker MIMO channel model which have separated the antenna correlation at the transmitter and receiver sides. The underlying assumption is that the correlation matrix of the MIMO channel matrix is equal to the Kronecker product of the antenna correlation at the transmitter and receiver sides. This assumption has reduced the number of correlation elements calculation for the covariance matrix construction to just $N_R^2 + N_T^2$ (N. Costa et al., 2010). We will mention about three correlation types, viz. constant, exponential and circular. We will also find the mgf of quadratic form of a Hermitian matrix $\boldsymbol{A}$ in complex Gaussian variates, $\mathbf{v}$.

We will study about the rank reduced keyhole MIMO channel model and find its pdf.

Finally we will discuss about the parallel decomposition of MIMO channel. We can convert the MIMO channel into R_H parallel Gaussian channels where R_H is the rank of the channel matrix $\boldsymbol{H}$.

3.2 Independent and identically distributed (uncorrelated) MIMO fading model

Uncorrelated MIMO fading model is very popular with the theoreticians because it gives an easier path of analysis of the MIMO system analysis. But in practical scenario, we need to consider correlated MIMO channel models. Let us start our discussion of MIMO channels with uncorrelated MIMO channel models first and discuss correlated MIMO channel models next. We have discussed about various RVs and its fading distributions in the previous chapter. In MIMO analysis, the channel is a complex random matrix and hence we need to find the pdf, cdf and mgf of complex random matrix. We will study complex RV (elements of a complex random vector), complex random vector (rows or columns of a complex random matrix are complex random vectors) and complex random matrix. After all, a random matrix could be converted into a random vector by stacking all the columns of the random matrix in a vector.

3.2.1 Complex random variable

A complex RV, Z = X+jY, can be considered as a pair of real RVs, X and Y. The pdf of a complex RV is defined to be the joint pdf of its real and complex parts. If X and Y are Gaussian, the pdf of Z is bivariate Gaussian distributed.

$$p_{X,Y}\left(x,y\right) = \frac{1}{k_1}\exp\left\{-\frac{1}{k_2}\left[\frac{\left(x-m_X\right)^2}{\sigma_x^2} + \frac{\left(y-m_Y\right)^2}{\sigma_y^2} - 2\rho\frac{\left(x-m_X\right)\left(y-m_Y\right)}{\sigma_x\sigma_y}\right]\right\} \tag{3.1}$$

where, ρ is the correlation coefficient between RVs, X and Y, m_X, m_Y, $\sigma_X{}^2$, $\sigma_Y{}^2$ are the mean and variance of the RVs, X and Y, respectively, and $k_1 = 2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}, k_2 = 2\left(1-\rho^2\right)$.

Example 3.1

Find the pdf of a complex normal RV.

Solution

A complex normal *RV,* (Z=X+jY), is a complex RV, whose real (X) and imaginary (Y) parts are i.i.d. Gaussian with zero mean and variance $\frac{1}{2}$. For independent X and Y ($\rho = 0$), zero mean ($m_X = m_Y = 0$) and $\frac{1}{2}$ variance ($\sigma_x^2 = \sigma_y^2 = \frac{1}{2}$), we have,

$$p_Z\left(z\right) = p_{X,Y}\left(x,y\right) = \frac{1}{\sqrt{2\pi\frac{1}{2}}\sqrt{2\pi\frac{1}{2}}}\exp\left\{-\frac{1}{2}\left[\frac{x^2}{\frac{1}{2}} + \frac{y^2}{\frac{1}{2}}\right]\right\} = \frac{1}{\pi}\exp\left\{-\left(x^2+y^2\right)\right\} = \frac{1}{\pi}\exp\left\{-|z|^2\right\} \tag{3.2}$$

3.2.2 Complex random vector

A complex random vector is defined as

$$\mathbf{z} = \mathbf{x} + j\mathbf{y} \tag{3.3}$$

where, **x** and **y** are real-valued random vectors of size n.

Real-valued covariance and cross-covariance matrices of random vectors x and y:

One may define the following real-valued matrices (covariance and cross-covariance matrices) for a complex random vector, **z**.

$$\mathbf{C_x} = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^T\right];$$

$$\mathbf{C_y} = \mathrm{E}\left[(\mathbf{y} - \mathrm{E}[\mathbf{y}])(\mathbf{y} - \mathrm{E}[\mathbf{y}])^T\right];$$

$$\mathbf{C_{xy}} = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{y} - \mathrm{E}[\mathbf{y}])^T\right];$$

$$\mathbf{C_{yx}} = \mathrm{E}\left[(\mathbf{y} - \mathrm{E}[\mathbf{y}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^T\right];$$

Matrices $\mathbf{C}_x$ and $\mathbf{C}_y$ are the vector covariance matrices of real random vectors, **x** and **y**, respectively, and hence they are symmetric and non-negative definite. Also note that the vector cross-covariance matrices are related as

$$\mathbf{C_{yx}} = \mathbf{C_{xy}}^T$$

Vector correlation matrices are defined as

$$\mathbf{R_x} = E(\mathbf{x}\mathbf{x}^T)$$

Hence for a real random vector **x** with correlation matrix $\mathbf{R_x}$, covariance matrix $\mathbf{C_x}$ and vector expected mean value $\boldsymbol{\mu}_\mathbf{x}$, $\mathbf{C_x} = \mathbf{R_x} - \boldsymbol{\mu}_\mathbf{x}\boldsymbol{\mu}^T_x$.

Example 3.2

Find the correlation and covariance matrices for real random vector, $\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$.

Solution

The correlation and covariance matrices are

$$\mathbf{R_x} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & E[X_1X_3] \\ E[X_2X_1] & E[X_2^2] & E[X_2X_3] \\ E[X_3X_1] & E[X_3X_2] & E[X_3^2] \end{bmatrix}; \mathbf{C_x} = \begin{bmatrix} \mathrm{var}[X_1^2] & \mathrm{cov}[X_1, X_2] & \mathrm{cov}[X_1, X_3] \\ \mathrm{cov}[X_2, X_1] & \mathrm{var}[X_2^2] & \mathrm{cov}[X_2, X_3] \\ \mathrm{cov}[X_3, X_1] & \mathrm{cov}[X_3, X_2] & \mathrm{var}[X_3^2] \end{bmatrix}.$$

It can be proved that a Gaussian random vector **x** (R. D. Yates et al., 2005) has independent components iff $\mathbf{C_x}$ is a diagonal matrix. An n-dimensional standard normal random vector **x** is the n-dimensional Gaussian random vector with $\boldsymbol{\mu}_\mathbf{x} = \mathbf{0}$, $\mathbf{C_x} = \mathbf{I}$. For a Gaussian random vector, $\mathbf{x} \sim N(\boldsymbol{\mu}_\mathbf{x}, \mathbf{C_x})$, if **A** is an n×n matrix with the property $\mathbf{AA}^T = \mathbf{C_x}$. The random vector $\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})$ is a standard normal random vector.

If we define a 2n-dimensional real vector $\tilde{\mathbf{z}}$, whose pdf will be the same with that of the complex random vector **z**, it is clear that the covariance matrix of this real vector can be expressed as

$$\tilde{\mathbf{z}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; \mathbf{C}_{\tilde{\mathbf{z}}} = \begin{bmatrix} \mathbf{C_x} & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & \mathbf{C_y} \end{bmatrix}$$

Complex-valued covariance and pseudo-covariance matrices of random vectors z:

We can also define the following two complex-valued matrices

$$\mathbf{C_z} = \mathrm{E}\left[(\mathbf{z} - \mathrm{E}[\mathbf{z}])(\mathbf{z} - \mathrm{E}[\mathbf{z}])^{\mathbf{H}}\right]; \tilde{\mathbf{C}}_\mathbf{z} = \mathrm{E}\left[(\mathbf{z} - \mathrm{E}[\mathbf{z}])(\mathbf{z} - \mathrm{E}[\mathbf{z}])^T\right] \quad (3.4)$$

where, $\mathbf{C_z}$ and $\tilde{\mathbf{C}}_\mathbf{z}$ are called the covariance and the pseudo-covariance of the complex random vector **z**.

It could be verified that for any **z**, the covariance matrix is Hermitian and positive definite and pseudo-covariance is skew-Hermitian. From these definitions, it can be verified that

$$\mathbf{C_z} = \mathbf{C_x} + \mathbf{C_y} + j(\mathbf{C_{yx}} - \mathbf{C_{xy}}); \tilde{\mathbf{C}}_\mathbf{z} = \mathbf{C_x} - \mathbf{C_y} + j(\mathbf{C_{yx}} + \mathbf{C_{xy}})$$

Vector correlation matrices are defined as

$$\mathbf{R_Z} = E(\mathbf{zz}^H)$$

Hence for zero mean case, $\mathbf{C_z} = \mathbf{R_z}$.

Multivariate joint cdf:

The joint cdf of RVs $X_1, \ldots, X_n$ is defined as $P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P[X_1 \le x_1,\cdots,X_n \le x_n]$.

Multivariate joint pdf:

The joint pdf of continuous RVs $X_1, \ldots, X_n$ is defined as

$$p_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \frac{\partial^n P_{X_1,\cdots,X_n}(x_1,\cdots,x_n)}{\partial x_1,\cdots,x_n}$$

A Gaussian random vector (H. Wymeersch, 2007) is completely characterized by its mean ($\mathbf{m} = E(\mathbf{z})$) and covariance matrix, ($\boldsymbol{\Phi} = E[(\mathbf{z} - \mathbf{m})(\mathbf{z} - \mathbf{m})^H]$). When **x** is real we write $\mathbf{x} \sim N_R^n(\mathbf{m}, \boldsymbol{\Phi})$ and its pdf (G. A. F. Seber, 2003) is given by

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Phi})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \boldsymbol{\Phi}^{-1}(\mathbf{x} - \mathbf{m})\right) \quad (3.5a)$$

For a complex n-multivariate Gaussian distribution with mean $\mathbf{m} \in C^n$ and covariance matrix $\mathbf{\Phi} \in C^{n\times n}$ is denoted by $\mathbf{z} \sim N_C^n(\mathbf{m}, \mathbf{\Phi})$. It is basically a complex **z** with independent imaginary and real parts with same covariance matrix, $\frac{1}{2}\mathbf{\Phi}$. Note that the subscript c denotes that it is a complex distribution and superscript n means that it is an n-multivariate distribution and N means it is normal or Gaussian distribution. Its pdf (A. van den Bos, 1995) is given by

$$p(\mathbf{z}) = \frac{1}{(\pi)^n \det(\mathbf{\Phi})} \exp\left(-(\mathbf{z}-\mathbf{m})^H \mathbf{\Phi}^{-1}(\mathbf{z}-\mathbf{m})\right) \tag{3.5b}$$

In multivariate statistical analysis, we are often interested in a random real n-vector **x** that is distributed with mean vector **m** and covariance matrix **Φ** where,

$$\mathbf{m} = E(\mathbf{x}), \mathbf{\Phi} = E\left[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\right]$$

We may sometimes assume that **x** is also Gaussian distributed. Typically **m** and **Φ** are unknown; they can be estimated as

$$\hat{\mathbf{\Phi}} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mathbf{m}})(x_i - \hat{\mathbf{m}})^T ; \hat{\mathbf{m}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Check the notation used: **x** and **y** for real random vector and **z** for complex random vector.

Example 3.3

Describe the following terms:
(a) Proper random vector
(b) Circularly symmetric random vector
(c) $N_c(0, N_0)$

Solution

(a) Proper random vectors
A complex random vector is called proper if its pseudo-covariance is zero $\tilde{\mathbf{C}}_\mathbf{z} = 0$ which implies that, $\mathbf{C}_\mathbf{x} = \mathbf{C}_\mathbf{y}; \mathbf{C}_{\mathbf{yx}} = -\mathbf{C}_{\mathbf{xy}}$.

For the special case, when random vector size n = 1 (complex random variable Z = X+jY say), the conditions for being proper become VAR[X]=VAR[Y]; COV[X,Y]=–COV[Y,X] which means that Z is proper if X and Y have equal variances and are uncorrelated. In this case, VAR[Z] = 2VAR[X]. For jointly Gaussian random variable, uncorrelated is equivalent to independent. We can conclude that a complex Gaussian random variable Z is proper *iff* its real and complex parts are independent with equal variance.

(b) For circular symmetric complex random vector **z**, $e^{j\varphi}\mathbf{z}$ has the same distribution as **z** for any φ

$$E[\mathbf{z}] = E\left[e^{j\varphi}\mathbf{z}\right] = e^{j\varphi}E[\mathbf{z}] = 0$$

$$E[\mathbf{z}\mathbf{z}^T] = E\left[e^{j\varphi}\mathbf{z}\left[e^{j\varphi}\mathbf{z}\right]^T\right] = e^{2j\varphi}E\left[\mathbf{z}\mathbf{z}^T\right] = 0$$

Since the mean and pseudo-covariance matrices are zero, the covariance matrix fully specifies first- and second-order statistics of a circular symmetric random vector.

(c) So a circular symmetric Gaussian random vector with covariance matrix $\mathbf{C}_z$ is denoted as N_c (0, $\mathbf{C}_z$). Here N means normal or Gaussian distributed and subscript c means it is complex. For additive white Gaussian noise, we shall write N_c (0, N_0). A complex normal random vector $\mathbf{z}$ whose real and imaginary parts are independent and identically distributed (i.i.d.) satisfies a circular symmetry property: $e^{j\varphi}\mathbf{z}$ has the same distribution as $\mathbf{z}$ for any φ. We call such a random vector circular symmetric complex Gaussian, denoted by, N_c (0, σ^2), where, σ^2 is the variance of each complex vector components.

3.2.3 Complex random matrix

A random matrix is a matrix whose elements are random variables. If the elements of random matrix are complex random variables, then it is a complex random matrix. A random matrix like random vectors can have joint pdf of its elements (T. W. Anderson, 2003).

A random matrix $\mathbf{H} \in C^{N_R \times N_T}$ is said to have a matrix-variate Gaussian distribution with mean $\mathbf{M} \in C^{N_R \times N_T}$ and covariance matrix $\mathbf{\Phi} \otimes \mathbf{\Psi}$ (where $\mathbf{\Phi} \in C^{N_R \times N_R} > 0$ and $\mathbf{\Psi} \in C^{N_T \times N_T} > 0$ are Hermitian) if $vect\left(\mathbf{H}^H\right) \sim N_C^{N_R,N_T}\left[vect\left(\mathbf{M}^H\right), \mathbf{\Phi} \otimes \mathbf{\Psi}\right]$. We denote $\mathbf{H} \sim N_C^{N_R,N_T}\left[\mathbf{M}, \mathbf{\Phi} \otimes \mathbf{\Psi}\right]$ and its pdf is given by

$$p_{\mathbf{H}}\left(\mathbf{H}\right) = \pi^{-N_R N_T} \det\left(\mathbf{\Phi}\right)^{-N_T} \det\left(\mathbf{\Psi}\right)^{-N_R} \exp\left[-trace\left\{\mathbf{\Phi}^{-1}\left(\mathbf{H}-\mathbf{M}\right)\mathbf{\Psi}^{-1}\left(\mathbf{H}-\mathbf{M}\right)^H\right\}\right]$$

which could be expressed as

$$p_{\mathbf{H}}\left(\mathbf{H}\right) = \pi^{-N_R N_T} \det\left(\mathbf{\Phi}\right)^{-N_T} \det\left(\mathbf{\Psi}\right)^{-N_R} etr\left\{-\mathbf{\Phi}^{-1}\left(\mathbf{H}-\mathbf{M}\right)\mathbf{\Psi}^{-1}\left(\mathbf{H}-\mathbf{M}\right)^H\right\} \tag{3.6}$$

In the above, "etr" is the abbreviation for "exponential trace". For instance, a $N_R \times N_T$ complex random matrix $\mathbf{H}$ with i.i.d. $N_c(0,1)$ entries is denoted as $\mathbf{H} \sim N_C^{N_R,N_T}\left[\mathbf{0}, \mathbf{I}_{N_R} \otimes \mathbf{I}_{N_T}\right]$. Its pdf is given by

$$p_{\mathbf{H}}\left(\mathbf{H}\right) = \pi^{-N_R N_T} etr\left\{-\mathbf{H}\mathbf{H}^H\right\} \tag{3.7}$$

It is easy to prove the equation (3.7).

Example 3.4

Prove equation (3.7).

Solution

A complex normal matrix **H** is a random matrix whose elements are complex normal RV whose pdf from equation (3.2) is given by

$$h_{i,j} \sim \frac{1}{\pi}\exp\left\{-\left|h_{i,j}\right|^2\right\}$$

For i.i.d. complex normal matrix, the joint pdf of elements of **H** is equal to the product pdf of each complex normal RV. Hence,

$$p_{\mathbf{H}}(\mathbf{H}) = \prod_{i,j=1}^{N_R,N_T}\frac{1}{\pi}\exp\left\{-\left|h_{i,j}\right|^2\right\} = \frac{1}{\pi^{N_RN_T}}\exp\left(-\sum_{i,j=1}^{N_R,N_T}\left|h_{i,j}\right|^2\right)$$

Note that trace for a square matrix is equal to the sum total of its diagonal elements and hence

$$trace\left(\mathbf{HH}^H\right)$$

$$= trace\left(\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \ddots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T}\end{bmatrix}\begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N_R1} \\ h_{12} & h_{22} & \cdots & h_{N_R2} \\ \vdots & \ddots & \ddots & \vdots \\ h_{1N_T} & h_{2N_T} & \cdots & h_{N_RN_T}\end{bmatrix}^*\right)$$

$$= trace\left(\begin{bmatrix} \left|h_{11}\right|^2+\left|h_{12}\right|^2+\cdots+\left|h_{1N_T}\right|^2 & \cdots & \cdots & \cdots \\ \cdots & \left|h_{21}\right|^2+\left|h_{22}\right|^2+\cdots+\left|h_{2N_T}\right|^2 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \left|h_{N_R1}\right|^2+\left|h_{N_R2}\right|^2+\cdots+\left|h_{N_RN_T}\right|^2 \end{bmatrix}\right)$$

$$= \sum_{i,j=1}^{N_R,N_T}\left|h_{i,j}\right|^2$$

Hence,
$$p_{\mathbf{H}}(\mathbf{H}) = \frac{1}{\pi^{N_RN_T}}\exp\left(-Trace\left(\mathbf{HH}^H\right)\right)$$

This joint distribution of entries of **H** matrix will be used in the calculation of pairwise error probability of space-time codes in the later part of the book.

Review question 3.1 *What is the pdf of a complex normal matrix* **H**?

Note that for each and every wireless communication, the spectral efficiency is vulnerable on the channel propagation conditions, which is reliant on the channel environment. The elements of the channel matrix **H** are usually modelled to be i.i.d. (uncorrelated). This is the most elementary model of all the analytical MIMO channel models. The covariance matrix of i.i.d. MIMO channel **H** is given as

$$\mathbf{R}_H = \sigma_h^2\mathbf{I} \tag{3.8}$$

Hence for an uncorrelated channel, the covariance matrix is a diagonal matrix, with each diagonal term identical to σ_h^2 (usually taken as 1). The non-diagonal terms of the covariance matrix give the cross covariance between the independent channel components and consequently zero. This is the case of highly scattered environment. σ_h^2 is the variance of the elements for MIMO channel matrix **H**. For an i.i.d. MIMO channel **H**, its elements h_{ij} are i.i.d. complex RV with zero mean and unit variance. For i.i.d. Rayleigh fading MIMO channel, $h_{ij} \sim N_C(0,1)$. In other words, $h_{ij} = h_{ij}^{real} + h_{ij}^{imag}$ and $h_{ij}^{real}, h_{ij}^{imag} \sim N_R(0,1/2)$. Initial analysis for capacity calculation for MIMO systems employed i.i.d. channels.

Review question 3.2 *What is the covariance matrix of an i.i.d. MIMO channel* **H**?

The covariance and correlation (R. D. Yates et al., 2005) of two RVs X and Y is defined as $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_X)]$ and $Cor(X, Y) = E[XY]$ respectively. Hence, $Cov(X, Y) = Cor(X, Y) - \mu_X\mu_Y$. For zero mean RVs X and Y, $Cov(X, Y) = Cor(X, Y)$.

The two RVs X and Y are
(a) orthogonal if $Cor(X, Y) = 0$ and
(b) uncorrelated if $Cov(X, Y) = 0$

Note that uncorrelated means covariance is zero.

If X and Y are independent, then $Cor(X, Y) = \mu_X\mu_Y$ which implies that $Cov(X, Y) = Cor(X, Y) - \mu_X\mu_Y = 0$.

If X = Y, then $Cov(X, Y) = Var(X) = Var(Y)$.

Further, a Gaussian random vector **X** has independent components iff covariance matrix is diagonal matrix. A normal Gaussian random vector **X** has independent components iff covariance matrix is an identity matrix.

Independent and non-identically distributed (i.i.n.d.):

RVs $X_1, \ldots, X_N$ are i.i.n.d. if for all $x_1, \ldots, x_N$, $p_{X_1,\cdots,X_N}(x_1,\cdots,x_N) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_N}(x_N)$.

Independent and identically distributed (i.i.d.):

RVs $X_1, \ldots, X_N$ are i.i.d. if for all $x_1, \ldots, x_N$, $p_{X_1,\cdots,X_N}\left(x_1,\cdots,x_N\right) = p_X\left(x_1\right)p_X\left(x_2\right)\cdots p_X\left(x_N\right)$.

Example 3.5

Explain the uncorrelated (i.i.d.) fading MIMO channel model for a 2×2 MIMO system. Show that the covariance matrix is $\mathbf{I}_4$.

Solution

Let us consider a 2×2 MIMO system (for illustration purpose) whose channel matrix is

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

Here h_{ij} means the channel gain coefficients for signal transmitted from transmitting antenna j to receiving antenna i.

Assuming that $E\{h_{ij}\} = 0$; $i, j = 1, 2$ we can find the covariance matrix $\mathbf{R_H}$ as

$$E[vect(\mathbf{H})\{vect(\mathbf{H})\}^H] = E\left[\begin{bmatrix} h_{11} \\ h_{21} \\ h_{12} \\ h_{22} \end{bmatrix}\begin{bmatrix} h_{11} & h_{21} & h_{12} & h_{22} \end{bmatrix}^*\right]$$

$$= \begin{bmatrix} E\left[|h_{11}|^2\right] & E\left[h_{11}h_{21}^*\right] & E\left[h_{11}h_{12}^*\right] & E\left[h_{11}h_{22}^*\right] \\ E\left[h_{21}h_{11}^*\right] & E\left[|h_{21}|^2\right] & E\left[h_{21}h_{12}^*\right] & E\left[h_{21}h_{22}^*\right] \\ E\left[h_{12}h_{11}^*\right] & E\left[h_{12}h_{21}^*\right] & E\left[|h_{12}|^2\right] & E\left[h_{12}h_{22}^*\right] \\ E\left[h_{22}h_{11}^*\right] & E\left[h_{22}h_{21}^*\right] & E\left[h_{22}h_{12}^*\right] & E\left[|h_{22}|^2\right] \end{bmatrix}$$

Note that h_{11}, h_{12}, h_{21} and h_{22} are all mutually independent and identically distributed (uncorrelated) RVs with zero mean. Hence,

$$E[vect(\mathbf{H})\{vect(\mathbf{H})\}^H] = \begin{bmatrix} E\left[|h_{11}|^2\right] & 0 & 0 & 0 \\ 0 & E\left[|h_{21}|^2\right] & 0 & 0 \\ 0 & 0 & E\left[|h_{12}|^2\right] & 0 \\ 0 & 0 & 0 & E\left[|h_{22}|^2\right] \end{bmatrix}$$

If we choose

$$E\left[|h_{11}|^2\right] = E\left[|h_{12}|^2\right] = E\left[|h_{21}|^2\right] = E\left[|h_{22}|^2\right] = 1$$

then

$$E[vect(\mathbf{H})\{vect(\mathbf{H})\}^H] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This analysis can be easily done for any arbitrary $N_R \times N_T$ MIMO system.

Uncorrelated (i.i.d.) channels are usually taken as Rayleigh for most of the MIMO analysis. It may be possible to extend this MIMO channel model to any one of the following classical or generalized fading distributions.

(a) Nakagami-m
(b) Rice
(c) η-μ
(d) k-μ
(e) α-μ

Review question 3.3 *What do you mean by i.i.d. and i.n.i.d.?*

An important matrix we will frequently use in MIMO capacity analysis is the random complex Wishart matrix which is defined as

$$\mathbf{Q} = \begin{cases} \mathbf{H}\mathbf{H}^H, & N_R < N_T \\ \mathbf{H}^H\mathbf{H}, & N_R \geq N_T \end{cases}$$

where, $\mathbf{H}$ is the random MIMO channel matrix.

Review question 3.4 *What is complex Wishart matrix?*

3.2.4 Complex non-central Wishart matrix

Let us define m = min(N_R, N_T) and n = max{N_R, N_T}. If $\mathbf{H} \sim N_C^{m,n}[\mathbf{M}, \mathbf{I}_n \otimes \mathbf{\Sigma}]$ (where $E(\mathbf{H}) = \mathbf{M}$; cov ($\mathbf{H}$) = $\mathbf{I}_n \otimes \mathbf{\Sigma}$; $m \leq n$) then $\mathbf{Q}$ is complex non-central Wishart distributed and is denoted by, $\mathbf{Q} \sim W_C^m(n, \mathbf{\Sigma}, \mathbf{\Theta} = \mathbf{\Sigma}^{-1}\mathbf{M}^H\mathbf{M})$. The joint pdf of $\mathbf{Q}$ is given by

$$p_{\mathbf{Q}}(\mathbf{Q}) = \frac{1}{\tilde{\Gamma}_m(n)} \det(\mathbf{\Sigma})^{-n} \det(\mathbf{Q})^{n-m}\, etr\left\{-\mathbf{\Sigma}^{-1}\mathbf{Q}\right\} \times {}_0F_1\left(n; \mathbf{\Theta}\mathbf{\Sigma}^{-1}\mathbf{Q}\right)$$

where, ${}_pF_q$ is hypergeometric function defined in section 11.2.4. The matrix $\mathbf{\Theta}$ is called the non-central parameter matrix. When $\mathbf{\Theta} = \mathbf{0}$, the non-central Wishart distribution reduces to the central Wishart distribution.

3.2.5 Complex central Wishart matrix

If $\mathbf{H} \sim N_C^{m,n}[\mathbf{M} = 0, \mathbf{I}_n \otimes \mathbf{\Sigma}]$ (where, $E(\mathbf{H}) = \mathbf{M} = \mathbf{0}$; cov ($\mathbf{H}$) = $\mathbf{I}_n \otimes \mathbf{\Sigma}$; $m \leq n$) then $\mathbf{Q}$ is central complex Wishart distributed and is denoted by $\mathbf{Q} \sim W_C^m(n, \mathbf{\Sigma})$. Then, $\mathbf{Q}$ is a $m \times m$ positive semi-definite matrix $\left(\mathbf{x}^H\mathbf{Q}\mathbf{x} = \left\|\mathbf{H}^H\mathbf{x}\right\|^2 \geq 0 \forall \mathbf{x} \in C^m\right)$. A random Hermitian positive-definite matrix $\mathbf{Q} \in C^{m\times m}$ is said to have a complex Wishart distribution with parameters m, n and $\mathbf{\Sigma} \in C^{m\times m} > 0$ denoted by $\mathbf{Q} \sim W_C^m(n, \mathbf{\Sigma})$, $m \leq n$ if its pdf is given by

$$p_{\mathbf{Q}}(\mathbf{Q}) = \frac{1}{\tilde{\Gamma}_m(n)} \det(\mathbf{\Sigma})^{-n} \det(\mathbf{Q})^{n-m}\, etr\left\{-\mathbf{\Sigma}^{-1}\mathbf{Q}\right\} \quad (3.9)$$

In the above equation, $\tilde{\Gamma}_m(\alpha)$ is the complex multivariate gamma function.

$$\tilde{\Gamma}_m(\alpha) = \pi^{\frac{m(m-1)}{2}} \prod_{i=0}^{m-1} \Gamma(\alpha - i), \mathrm{Re}(\alpha) > m-1, \mathbf{Q} > 0$$

where, Γ is the gamma function.

If the covariance matrix $\mathbf{\Sigma}$ is an identity matrix usually in case of MIMO analysis, the above equation can be simplified as

$$p_{\mathbf{Q}}(\mathbf{Q}) = \frac{1}{\tilde{\Gamma}_m(n)} \det(\mathbf{Q})^{n-m}\, etr\{-\mathbf{Q}\} \tag{3.10}$$

Review question 3.5 *What is complex central and non-central Wishart matrix?*

From spectral theorem, there exists an $m \times m$ unitary matrix $\mathbf{U}$ and diagonal eigenmatrix $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$; $\lambda_i > 0\, \forall\, i$ such that $\mathbf{Q} = \mathbf{U}\, \boldsymbol{\lambda}\, \mathbf{U}^H$. Since $\mathbf{Q}$ (complex Wishart matrix) is random, λ (eigenvalue and correspondingly eigenvalue matrix) is also random. We are interested in finding the joint distribution of eigenvalues of the central complex Wishart matrix $\mathbf{Q}$ and its marginal distributions.

Review question 3.6 *What is spectral theorem?*

3.2.6 Joint distribution of eigenvalues of complex Wishart matrix

If $\mathbf{Q} \sim W_C^m(n, \mathbf{I}_m)$ i.e., $\mathbf{\Sigma} = \mathbf{I}_m$ then the pdf of the ordered eigenvalues of $\mathbf{Q}$ (A. Edelman, 1989) $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$; $0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_m$ is given by

$$p_{\lambda}^{ordered}(\lambda_1, \lambda_2, \cdots, \lambda_m) = \frac{1}{\Gamma_m(n)\Gamma_m(m)} \exp\left(-\sum_{i=1}^{m} \lambda_i\right) \prod_{i=1}^{m} \lambda_i^{n-m} \prod_{j=i+1}^{m} \left(\lambda_i - \lambda_j\right)^2 \tag{3.11}$$

where, $\Gamma_m(a) = \prod_{i=1}^{m-1} \Gamma(a-i)$

The joint pdf of the unordered eigenvalues can be obtained by dividing by m! in the above expression.

$$p_{\lambda}^{unordered}(\lambda_1, \lambda_2, \cdots, \lambda_m) = \frac{1}{m!\,\Gamma_m(n)\Gamma_m(m)} \exp\left(-\sum_{i=1}^{m} \lambda_i\right) \prod_{i=1}^{m} \lambda_i^{n-m} \prod_{j=i+1}^{m} \left(\lambda_i - \lambda_j\right)^2 \tag{3.12}$$

Review question 3.7 *How to obtain the joint pdf of the unordered eigenvalues from the pdf of the ordered eigenvalues of* $\mathbf{Q}$*?*

3.2.7 Marginal distribution of an eigenvalue of complex Wishart matrix

In order to find the marginal pdf of λ_1, we need to integrate the unordered joint pdf above with respect to $\lambda_2, \ldots, \lambda_m$, and hence we get (E. Biglieri, 2005)

$$p(\lambda_1) = \frac{1}{m}\sum_{i=1}^{m} \zeta_i^2\left(\lambda_1\right)\lambda_1^{n-m}e^{-\lambda_1}; \zeta_i^2\left(\lambda_1\right)$$
$$= \frac{(i-1)!}{(i-1+n-m)!}\left(L_{i-1}^{n-m}\left(\lambda_1\right)\right)^2$$

which can be further simplified as

$$p(\lambda_1) = \frac{1}{m}\sum_{i=0}^{m-1}\frac{(i)!}{(i+n-m)!}\left(L_i^{n-m}\left(\lambda_1\right)\right)^2 \lambda_1^{n-m}e^{-\lambda_1}$$

In the above equation, L_i^{n-m} are the associated Laguerre polynomials of order i (Eq. (8.970.1) of I. S. Gradshteyn et al., 2000) given by

$$L_i^{n-m}\left(\lambda_1\right) = \frac{1}{i!}e^x x^{m-n}\frac{d^i}{dx^i}\left(e^{-x}\left(\lambda_1\right)^{n-m+i}\right)$$

$$= \sum_{l=0}^{i}(-1)^l\binom{i+n-m}{i-l}\frac{\left(\lambda_1\right)^l}{l!}$$

From the identity given in Eq. (8.976.3) of I. S. Gradshteyn et al., 2000, we have

$$\left(L_i^{n-m}\left(\lambda_1\right)\right)^2 = \frac{\Gamma(n-m+i+1)}{2^{2i}i!}\sum_{j=0}^{i}\frac{\binom{2i-2j}{i-j}(2j)!}{j!\Gamma(n-m+j+1)}L_{2j}^{2(n-m)}\left(2\lambda_1\right)$$

Hence the marginal pdf of $\lambda_1 > 0$ could be expressed as

$$p(\lambda_1) = \frac{1}{m}\sum_{i=0}^{m-1}\frac{(i)!}{2^{2i}i!}\sum_{j=0}^{i}\frac{\binom{2i-2j}{i-j}(2j)!}{j!\Gamma(n-m+j+1)}L_{2j}^{2(n-m)}\left(2\lambda_1\right)\lambda_1^{n-m}e^{-\lambda_1}$$

Substituting the associated Laguerre polynomials $L_{2j}^{2(n-m)}\left(2\lambda_1\right)$ into a finite sum given above, we have,

$$p(\lambda_1) = \frac{1}{m}\sum_{i=0}^{m-1}\frac{(i)!}{2^{2i}i!}\sum_{j=0}^{i}\frac{\binom{2i-2j}{i-j}(2j)!}{j!\Gamma(n-m+j+1)}\sum_{l=0}^{2j}(-1)^l\binom{2j+2n-2m}{2j-l}\frac{\left(2\lambda_1\right)^l}{l!}\lambda_1^{n-m}e^{-\lambda_1}$$

which can be further simplified as

$$p(\lambda_1) = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{i}\sum_{l=0}^{2j}\frac{(-1)^l(2j)!}{2^{2i-l}j!\,l!(n-m+j)!}\binom{2i-2j}{i-j}\binom{2j+2n-2m}{2j-l}(\lambda_1)^{l+n-m}e^{-\lambda_1} \tag{3.13}$$

The above pdf is required for capacity calculation of MIMO systems. But in real life applications, the components of the channel matrix are correlated for closely spaced antennas. The less separated are the antennas in space, the higher is the correlation coefficient. Werner Weichselberger et al. (2006) presented a novel stochastic model for MIMO radio channels that is based on the joint correlation properties of both link ends. There are two types of correlation: temporal and spatial. Temporal correlation is used for modelling channels for moving terminals. Spatial correlation is used due to employment of multiple antennas. For MISO, the spatial correlation is for multiple transmit antennas and for SIMO, the spatial correlation is for multiple antennas at the receiver. For MIMO, spatial correlation exists for both at transmitter and receiver since both links are equipped with multiple antennas (A. Sibille et al., 2011).

3.3 Fully correlated MIMO channel model

In finding $E[h_{ij}h_{kl}^*]$, we need to take the expectation over the joint distributions of RVs, h_{ij} and h_{kl}, where h_{ij} and h_{kl} are the channel coefficients. It will be more appropriate to represent the elements of the covariance matrix as a tensor (A. B. Gershman et al., 2005) since four indices (i and j combine to form row indices of covariance matrix $\mathbf{R_H}$ whereas k and l combine to form column indices of covariance matrix $\mathbf{R_H}$) are used. Therefore,

$$\mathbf{R}_{\mathbf{H}}^{ij,kl} = E\left[h_{ij}h_{kl}^*\right]$$

In fully correlated MIMO channel model, we need to consider all the co- and cross-correlations between all the channel coefficients for various channel paths between the transmitting antennas and receiving antennas. This covariance matrix size and consequently the number of elements of the covariance matrix becomes prohibitively large as the number of transmitting and receiving antennas increases. For a given $\mathbf{H}$ matrix, the correlation matrix (N. Costa et al., 2012) can be defined as

$$\mathbf{R}_H = E\left\{vec(\mathbf{H})\left[vec(\mathbf{H})\right]^H\right\}$$

$$= E\left\{\begin{bmatrix} h_{11} \\ \vdots \\ h_{N_R1} \\ h_{12} \\ \vdots \\ h_{N_R2} \\ \vdots \\ h_{N_RN_T}\end{bmatrix}\begin{bmatrix} h_{11}^* & \cdots & h_{N_R1}^* & h_{12}^* & \cdots & h_{N_R2}^* & \cdots & h_{N_RN_T}^*\end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} h_{11}h_{11}^* & \cdots & h_{11}h_{N_R1}^* & h_{11}h_{12}^* & \cdots & h_{11}h_{N_R2}^* & \cdots & h_{11}h_{N_RN_T}^* \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{N_R1}h_{11}^* & \cdots & h_{N_R1}h_{N_R1}^* & h_{N_R1}h_{12}^* & \cdots & h_{N_R1}h_{N_R2}^* & \cdots & h_{N_R1}h_{N_RN_T}^* \\ h_{12}h_{11}^* & \cdots & h_{12}h_{N_R1}^* & h_{12}h_{12}^* & \cdots & h_{12}h_{N_R2}^* & \cdots & h_{12}h_{N_RN_T}^* \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{N_R2}h_{11}^* & \cdots & h_{N_R2}h_{N_R1}^* & h_{N_R2}h_{12}^* & \cdots & h_{N_R2}h_{N_R2}^* & \cdots & h_{N_R2}h_{N_RN_T}^* \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{N_RN_T}h_{11}^* & \cdots & h_{N_RN_T}h_{N_R1}^* & h_{N_RN_T}h_{12}^* & \cdots & h_{N_RN_T}h_{N_R2}^* & \cdots & h_{N_RN_T}h_{N_RN_T}^* \end{bmatrix}\right\}$$

where, E is the expectation operator and *vec* is the vectorization operation. Obviously the above correlation matrix will have $(N_RN_T)^2$ components. Note that the above matrix is a fourth-order tensor and each element of the correlation matrix requires four indices (quadruplet) as shown below.

$$\mathbf{R}_{\mathbf{H}}^{ij,kl} = E\left[h_{ij}h_{kl}^*\right]; i,k \in \{1, \ 2, \ \cdots, \ N_R\}; j,l \in \{1, \ 2, \ \cdots, \ N_T\}$$

It is highly complex model for analysis. Hence we can assume that the antenna correlation at the transmitter and receiver side is separable as in the next section.

Review question 3.8 *How many components will be there for correlation matrix for fully correlated MIMO channel?*

3.4 Separately correlated (Kronecker) MIMO channel model

The Separately correlated (Kronecker) MIMO channel model, on the other hand, was based on the key hypothesis that the transmitter and the receiver correlations are detachable. This model implies that the correlation matrix at the receiver is independent of the transmission and vice versa. In other words, the multipath components arriving at the receiver is independent of which antennas have been used for transmission at the transmitter side (A. Sibille et al., 2011). It could also be represented by two ring model (N. Costa et al., 2012), in which one ring of scatterers surrounds the transmitter and another ring of scatterers surrounds the receiver. There is no coupling between these two rings of scatterers at the receiver and transmitter side. This model though has taken the correlation into consideration turns out to be less accurate than Weischselberger model (W. Weischselberger, 2003) which has considered the coupling between the transmitter and receiver correlation.

In this case, assumption is that we can write the fully correlated covariance matrix of the previous section as a Kronecker product of the correlation matrix at the receiver ($\mathbf{R}_{\mathbf{T_X}}$) and correlation matrix at the transmitter ($\mathbf{R}_{\mathbf{R_X}}$). It is reasonable with some accuracy that since the transmitter and receiver is usually at a very far distance than the distance between the antennas at the transmitter (or receiver) itself. It has been experimentally verified that two antennas at the BS (MS) should be at least ten (three) wavelengths apart to have sufficient decorrelation (B. Vucetic et al., 2003). A simple stochastic MIMO channel model has been developed by J. P. Kermoal et al. (2002). The Kronecker model introduces some structure to the correlation matrix and assumes that the channel correlation (J. P. Kermoal et al., 2002, E. G. Larsson et al., 2003 and T. Brown et al., 2012) can be expressed as the Kronecker product of correlation matrix at the transmitter ($\mathbf{R}_{\mathbf{T_X}}$) and correlation matrix at the receiver ($\mathbf{R}_{\mathbf{R_X}}$)

$$\mathbf{R}_{\mathbf{H}} = \mathbf{R}_{T_X}^{T} \otimes \mathbf{R}_{R_X} \tag{3.14}$$

where, $\otimes$ denotes the Kronecker product.

It will be shown shortly that the channel matrix $\mathbf{H}$ for Kronecker channel model can be written as

$$\mathbf{H} = \mathbf{R}_{\mathbf{R_X}}^{1/2} \mathbf{H}_w \mathbf{R}_{\mathbf{T_X}}^{1/2} \tag{3.15}$$

where, $\mathbf{H}_w$ is the channel matrix for uncorrelated or i.i.d. (spatially white that's why the subscript w in $\mathbf{H}$) MIMO fading channel and $()^{1/2}$ denotes the Hermitian square root of matrix. This equation is extensively utilized for theoretical calculations and MIMO channel simulations. It is the adopted MIMO channel model (K.-L. Du et al., 2010) for IEEE 802.11n and IEEE 802.20 (Mobile Broadband Wireless Access).

From the identity on vectorization of multiplication of three matrices, we have,

$$\because \quad vect(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) vect\left(\mathbf{B}\right)$$

$$\therefore \quad vect(\mathbf{H}) = \left(\mathbf{R}_{\mathbf{T_X}}^{T/2} \otimes \mathbf{R}_{\mathbf{R_X}}^{1/2}\right) vect\left(\mathbf{H}_w\right)$$

Hence,
$$\mathbf{h} = N_C\left(0, \mathbf{R}_{T_X}^{T} \otimes \mathbf{R}_{R_X}\right) \tag{3.16}$$

where, $\mathbf{h} = vect(\mathbf{H})$.

Note that the correlation matrix $\mathbf{R}_H$ for any random channel matrix is defined as

$$\mathbf{R}_H = E\left[\mathbf{h}\mathbf{h}^H\right]; \mathbf{h} = vect\left(\mathbf{H}\right)$$

The correlation matrices at the transmitter and the receiver are calculated as

$$\mathbf{R}_{\mathbf{T_X}} = E\left\{\left(\mathbf{H}^H\mathbf{H}\right)^T\right\};\ \mathbf{R}_{\mathbf{R_X}} = E\left(\mathbf{H}\mathbf{H}^H\right)$$

How to obtain the correlated channel matrix $\mathbf{H}$ for a given i.i.d. or spatially white channel matrix $\boldsymbol{H}_w$? How to introduce correlation in an otherwise i.i.d. or spatially white channel matrix? This is what we are going to do exactly in the next discussion. Assume that the correlation matrix $\boldsymbol{R}_H$ is known to us.

Another way of looking at the Kronecker MIMO channel model is

$$\mathbf{H} = unvec\left(\sqrt{\mathbf{R}_H}\mathbf{h}_w\right) = unvec\left(\sqrt{\mathbf{R}_H}\, vect\left(\mathbf{H}_w\right)\right)$$

Note that "unvec" is the reverse process of vectorization. "vect" (vectorization) stacks all the columns of a matrix into a column vector. "Unvec" converts back a vectorized matrix into its corresponding matrix. We will show that when we introduce the correlation matrix $\boldsymbol{R}_H$ into spatially white channel matrix $\boldsymbol{H}_w$, the correlation matrix of the correlated channel matrix $\mathbf{H}$ is indeed $\boldsymbol{R}_H$.

Therefore,

$$\mathbf{R}_H = E\left[\mathbf{hh}^H\right] = E\left[vect(\mathbf{H})\{vect(\mathbf{H})\}^H\right] = \sqrt{\mathbf{R}_H}\mathbf{I}_{N_R N_T}\sqrt{\mathbf{R}_H}^H = \mathbf{R}_H^{1/2}\mathbf{R}_H^{H/2} = \mathbf{R}_H$$

That's why the name Hermitian square root for the square root of the matrix $\boldsymbol{R}_H$ due to the last operation in the above equation.

For Kronecker model, we have assumed that

$$\mathbf{R}_H = E\left[\mathbf{hh}^H\right] = \mathbf{R}_{T_X}^T \otimes \mathbf{R}_{R_X}$$

For this particular case of correlation matrix $\boldsymbol{R}_H$ as given above, let us find the expression for correlated channel matrix **H** in terms of receiver correlation, transmitter correlation and spatially white channel matrix.

Since, $\quad (\mathbf{A} \otimes \mathbf{B})\, vect(\mathbf{C}) = vect\left(\mathbf{BCA}^T\right)$

Then, we have,

$$\mathbf{H} = unvec\left(\sqrt{\mathbf{R}_H}\mathbf{h}_w\right) = unvec\left(\sqrt{\mathbf{R}_{T_x}^T \otimes \mathbf{R}_{R_X}}\mathbf{h}_w\right)$$

$$\Rightarrow \qquad \mathbf{H} = unvec\left(\mathbf{R}_{T_x}^{T/2} \otimes \mathbf{R}_{R_x}^{1/2} vect(\mathbf{H}_w)\right)$$

$$= unvec\left\{vect\left(\mathbf{R}_{R_x}^{1/2}\mathbf{H}_w\mathbf{R}_{T_x}^{1/2}\right)\right\} = \mathbf{R}_{R_x}^{1/2}\mathbf{H}_w\mathbf{R}_{T_x}^{1/2}$$

An alternate component-wise expression for transmitter correlation (correlation between the columns of **H**, independent of the receive antennas) is

$$\mathbf{R}_{T_x,nq} = \sum_{m=1}^{N_R}\mathbf{R}_{mn,mq}$$

For example, $\mathbf{R}_{T_x,11} = \sum_{m=1}^{N_R}\mathbf{R}_{m1,m1} = \mathbf{R}_{11,11} + \mathbf{R}_{21,21} + \cdots + \mathbf{R}_{N_R 1, N_R 1}$ is the correlation between the first column with first column of **H**.

Component-wise expression for receiver correlation (correlation between the rows of **H**, independent of the transmit antennas) is

$$\mathbf{R}_{R_x,mp} = \sum_{n=1}^{N_T}\mathbf{R}_{mn,pn}$$

For example, $\mathbf{R}_{R_x,22} = \sum_{n=1}^{N_T}\mathbf{R}_{2n,2n} = \mathbf{R}_{21,21} + \mathbf{R}_{22,22} + \cdots + \mathbf{R}_{2N_T,2N_T}$ is the correlation of second row with second row of **H**.

Let us consider a 2×2 MIMO system for illustration purpose whose channel matrix is

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

The transmitter correlation matrix is given by

$$\begin{aligned}
\mathbf{R_{T_X}} &= E[\mathbf{H}^H\mathbf{H}]^T \\
&= E\left[\left[\begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix}^* \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}\right]^T\right] \\
&= \begin{bmatrix} E\left[|h_{11}|^2 + |h_{21}|^2\right] & E\left[h_{11}^* h_{12} + h_{21}^* h_{22}\right] \\ E\left[h_{12}^* h_{11} + h_{22}^* h_{21}\right] & E\left[|h_{12}|^2 + |h_{22}|^2\right] \end{bmatrix}^T \\
&= \begin{bmatrix} E\left[|h_{11}|^2 + |h_{21}|^2\right] & E\left[h_{12}^* h_{11} + h_{22}^* h_{21}\right] \\ E\left[h_{11}^* h_{12} + h_{21}^* h_{22}\right] & E\left[|h_{12}|^2 + |h_{22}|^2\right] \end{bmatrix}
\end{aligned}$$

The receiver correlation matrix is given by

$$\begin{aligned}
\mathbf{R_{R_X}} &= E[\mathbf{H}\mathbf{H}^H] \\
&= E\left[\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix}^*\right] \\
&= \begin{bmatrix} E\left[|h_{11}|^2 + |h_{12}|^2\right] & E\left[h_{11} h_{21}^* + h_{12} h_{22}^*\right] \\ E\left[h_{21} h_{11}^* + h_{22} h_{12}^*\right] & E\left[|h_{21}|^2 + |h_{22}|^2\right] \end{bmatrix}
\end{aligned}$$

Now the diagonal and off-diagonal terms are all non-zero. Note that $\mathbf{R}_{R_X}$ will have N_R^2 elements and $\mathbf{R}_{T_X}$ will have N_T^2 elements. In total in order to find $\mathbf{R_H} = \mathbf{R}^T{}_{T_X} \otimes \mathbf{R}_{R_X}$ we require to find $(N_R^2 + N_T^2)$ elements only. This is a significant reduction in comparison to fully correlated MIMO channel matrix of previous case, where we need to find $(N_R N_T)^2$ elements of the correlation matrix $\mathbf{R}_H = E\{\mathbf{h}\mathbf{h}^H\}$.

For instance, consider a 2×2 MIMO system with receiver correlation and transmitter correlation as follows.

$$\mathbf{R}_{R_X} = \begin{bmatrix} 1 & \rho_R \\ \rho_R & 1 \end{bmatrix}; \mathbf{R}_{T_X} = \begin{bmatrix} 1 & \rho_T \\ \rho_T & 1 \end{bmatrix}$$

Then for Kronecker channel model, the correlation matrix becomes

$$\begin{aligned}
\mathbf{R}_H &= \mathbf{R}_{T_X}^T \otimes \mathbf{R}_{R_X} \\
&= \begin{bmatrix} \mathbf{R}_{R_X} & \rho_T \mathbf{R}_{R_X} \\ \rho_T \mathbf{R}_{R_X} & \mathbf{R}_{R_X} \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} 1 & \rho_R & \rho_T & \rho_R\rho_T \\ \rho_R & 1 & \rho_R\rho_T & \rho_T \\ \rho_T & \rho_R\rho_T & 1 & \rho_R \\ \rho_R\rho_T & \rho_T & \rho_R & 1 \end{bmatrix}$$

where, ρ_R and ρ_T are the receiver correlation coefficient and transmitter correlation coefficient, respectively.

Based on the antenna array geometry, correlation could be of various types. We will consider three commonly used correlation types in the following (K. S. Ahn et al., 2007, M. K. Simon et al., 2005 and V. A. Aalo, 1995):

(a) Constant
(b) Circular and
(c) Exponential.

Constant correlation model is the worst case scenario. It is suitable for antenna array of three antennas placed on an equilateral triangle or for closely spaced antennas other than linear arrays. The correlation matrix **R** for this case is given by

$$\mathbf{R}_{\text{constant}} = \begin{bmatrix} 1 & x & \cdots & x \\ x & 1 & \cdots & x \\ \vdots & \vdots & \vdots & \vdots \\ x & x & \cdots & 1 \end{bmatrix} \tag{3.17}$$

This can be expressed simply as

$$\mathbf{R}_{\text{constant}}\Big|_{j,k} = \begin{cases} 1 & if \quad j = k \\ x & if \quad j \neq k \end{cases}$$

This model assumes that the power correlation coefficient is the same between all the channel coefficients/paths/links pairs.

Circular correlation model is appropriate for antennas lying on a circle, or four antennas placed on a square. The correlation matrix **R** for this case is given by

$$\mathbf{R}_{\text{circular}} = \begin{bmatrix} 1 & x_1 & \cdots & x_{n-1} \\ x_{n-1}^* & 1 & \cdots & x_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^* & x_2^* & \cdots & 1 \end{bmatrix} \tag{3.18}$$

This can be expressed simply as

$$\mathbf{R}_{\text{circular}}\Big|_{j,k} = \begin{cases} 1 & if \quad j = k \\ x_{k-j} & if \quad j < k \\ \left(x^*\right)_{n-(j-k)} & if \quad j > k \end{cases}$$

Exponential correlation model is suitable model for equally spaced linear antenna array. The correlation matrix **R** for this case is given by

$$\mathbf{R}_{\text{exponential}} = \begin{bmatrix} 1 & x & \cdots & x^{n-1} \\ x^* & 1 & \cdots & x^{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \left(x^*\right)^{n-1} & \left(x^*\right)^{n-2} & \cdots & 1 \end{bmatrix} \tag{3.19}$$

Joint mgf:

For any n RVs $X_1, \ldots, X_n$, the joint mgf (S. Ross, 2002) is defined for all real values of $s_1, \ldots, s_n$, by

$$M_{X_1,\cdots,X_n}\left(s_1,\cdots,s_n\right) = E\left[e^{s_1X_1+s_2X_2+\cdots+s_nX_n}\right]$$

If RVs $X_1, \ldots, X_n$, are i.i.n.d. then

$$M_{X_1,\cdots,X_n}\left(s_1,\cdots,s_n\right) = M_{X_1}\left(s_1\right)M_{X_2}\left(s_2\right)\cdots M_{X_n}\left(s_n\right)$$

If RVs are i.i.d. then

$$M_{X_1,\cdots,X_n}\left(s\right) = \left[M_X\left(s\right)\right]^n$$

This can be expressed simply as

$$\mathbf{R}_{\text{exponential}}\Big|_{j,k} = \begin{cases} 1 & \textit{if} \quad j = k \\ x^{k-j} & \textit{if} \quad j < k \\ \left(x^*\right)^{j-k} & \textit{if} \quad j > k \end{cases}$$

This model assumes exponential power correlation coefficient between any pair of channel coefficients/paths/links. It is suitable for equi-spaced linear antenna arrays whose correlation decreases with antenna spacing.

Review question 3.9 *Why can we separate the correlation at the transmitter and receiver?*

Let us find the mgf of quadratic form of matrix **A** in complex Gaussian variates **v** (R. Janaswamy et al., 2001). The derivations are given in appendix B.

$$M_y(s) = \frac{\exp\left(s\boldsymbol{\mu}_v^H\left(\mathbf{I} - s\mathbf{R}_\mathbf{v}\mathbf{A}\right)^{-1}\mathbf{A}\boldsymbol{\mu}_v\right)}{\left|\mathbf{I} - s\mathbf{R}_\mathbf{v}\mathbf{A}\right|}$$

A particular case, when $\boldsymbol{\mu}_v = 0$, we have,

$$M_y(s) = \left|\mathbf{I} - s\mathbf{R}_\mathbf{v}\mathbf{A}\right|^{-1}$$

We require this mgf for performance analysis of space-time codes over correlated channels.

3.5 Uncorrelated (keyhole) MIMO channel model

In few situations, keyhole effect can be seen. In such situations, scattering exists both at the transmitter and the receiver edges. But the entire rays from the transmitter tunnel pass through a tiny opening, in order to arrive at the receiver.

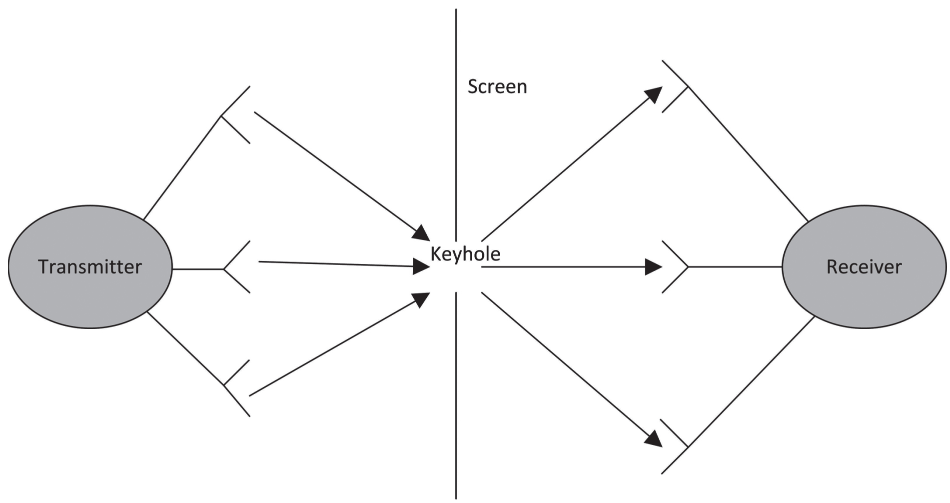


Fig. 3.1 A keyhole 3 × 3 MIMO channel

The scenario is similar to MISO system cascaded with a SIMO system. Hence it minimizes the degrees of freedom and consequently a scanty rank channel matrix emerges (rank of the channel matrix equals 1). Therefore, even if there are many scatterers and widely spaced antenna elements at the transmitter and receiver, the spectral efficiency of a pinhole channel gets diminished because of the rank deficit channel matrix. Such model is appropriate for indoor wireless communication through corridor or underpass or subway. Cooperative communication employing the amplify-and-forward protocol may be considered as pinhole channels.

Example 3.6

Show that the rank of Keyhole channel for a 3 × 3 MIMO system is 1.

Solution

Let us consider a 3 × 3 MIMO system with three transmit and receive antennas in highly scattered environment similar to what has been shown in Fig. 3.1. This MIMO system would under normal propagation conditions produce a rank 3 channel matrix and multiplexing gain of 3. Note that the rank of the channel matrix equals the rate/multiplexing gain which we will discuss in the next section. However if these three sets of antennas (transmit and receive antennas) are separated by a screen with a small hole, we get a keyhole propagation in which the rank of the channel matrix equals 1 and hence there are no rate or multiplexing gain. Let us assume that the transmitted signal vector is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

where, x_1, x_2 and x_3 are the signals transmitted from the first, second and third antennas, respectively.

The signal incident at the keyhole is given by

$$y = \mathbf{h}_{left}\mathbf{x} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

where, h_1, h_2 and h_3 are the channel coefficients in the left hand side of the keyhole for the transmitted signals x_1, x_2 and x_3.

The signal at the other side of the keyhole is given by

$$y_1 = \alpha y$$

where, α is the keyhole attenuation.

The signal vector at the receive antennas on the right side of the keyhole is given by

$$\mathbf{r} = \mathbf{h}_{right} y_1 = \alpha \mathbf{h}_{right}\mathbf{h}_{left}\mathbf{x} = \alpha \begin{bmatrix} h_4 \\ h_5 \\ h_6 \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \alpha \begin{bmatrix} h_4h_1 & h_4h_2 & h_4h_3 \\ h_5h_1 & h_5h_2 & h_5h_3 \\ h_6h_1 & h_6h_2 & h_6h_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{Hx}$$

where, $\mathbf{h}_{\text{right}}$ is the channel matrix describing the propagation on the right hand side of the keyhole and h_4, h_5 and h_6 are the channel coefficients corresponding to the first, second and third receive antennas, respectively.

The equivalent channel matrix can be represented as

$$\mathbf{H} = \alpha \begin{bmatrix} h_4h_1 & h_4h_2 & h_4h_3 \\ h_5h_1 & h_5h_2 & h_5h_3 \\ h_6h_1 & h_6h_2 & h_6h_3 \end{bmatrix}$$

The rank of this channel matrix is one which implies that there is no multiplexing gain. Note that rank of a matrix is the number of linearly independent rows or columns of the matrix. In the above matrix, we can obtain the second and third rows from the first row by multiplying a constant factor. Hence we have rank of the above matrix as 1.

Review question 3.10 *Give an example of a keyhole MIMO channel.*

In keyhole MIMO channel there will be diversity gain but no multiplexing gain. The channel matrix $\mathbf{H}$ for keyhole MIMO channels is given by

$$\mathbf{H} = \boldsymbol{\beta}\boldsymbol{\alpha}^T = \begin{pmatrix} \alpha_1\beta_1 & \alpha_2\beta_1 & \cdots & \alpha_{N_T}\beta_1 \\ \alpha_1\beta_2 & \alpha_2\beta_2 & \cdots & \alpha_{N_T}\beta_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1\beta_{N_R} & \alpha_2\beta_{N_R} & \cdots & \alpha_{N_T}\beta_{N_R} \end{pmatrix} \tag{3.20a}$$

where, $\boldsymbol{\alpha}$ is for the equivalent MISO channel, which is distributed as $\boldsymbol{\alpha} \sim N_C^{N_T}\left(0, \mathbf{I}_{N_T}\right)$ and $\boldsymbol{\beta}$ is for the equivalent SIMO channel which is distributed as $\boldsymbol{\beta} \sim N_C^{N_R}\left(0, \mathbf{I}_{N_R}\right)$. Now we can calculate the Z as

$$Z = \mathbf{H}\mathbf{H}^H = \|\boldsymbol{\alpha}\|^2 \|\boldsymbol{\beta}\|^2 \tag{3.20b}$$

The pdf of Z is required for calculation of capacity of MIMO systems for a keyhole MIMO channel.

Example 3.7

Find the pdf of Z given in Eq. (3.22).

Solution

Note that $U = \|\boldsymbol{\alpha}\|^2$ and $V = \|\boldsymbol{\beta}\|^2$ are the sums of N_T and N_R independent exponential RVs, respectively, and hence they are central Chi-square distributed with $2N_T$ and $2N_R$ degrees of freedom. We are considering $2N_T$ and $2N_R$ degrees of freedom because we are considering $\alpha_i; i = 1, \ldots, N_T$ and $\beta_i; i = 1, \ldots, N_R$ as complex RVs and it has real and imaginary parts which are real RVs.

Note that pdf of Chi-square RV (sum of squares of i.i.d. zero mean Gaussian RVs with common variance σ^2) with $2N$ degrees of freedom is given by

$$p_\chi(\chi) = \frac{1}{(N-1)!}\chi^{N-1} e^{-\frac{\chi}{2\sigma^2}}; \chi > 0$$

Hence, for our case $\sigma^2 = 1/2$, $N = N_T$ for RV U and $N = N_R$ for RV V (A. Paulraj et al., 2003), we have,

$$p_U(u) = \frac{1}{(N_T - 1)!} u^{N_T - 1} e^{-u}; u > 0$$

$$p_V(v) = \frac{1}{(N_R - 1)!} v^{N_R - 1} e^{-v}; v > 0$$

We can utilize the following theorem on transformations for functions of two RVs (NPTEL course by V. V. Rao as well as P. K. Bora).

Given two continuous RVs X and Y, define two new RVs which are defined using the transformation $Z = g(X,Y)$ and $W = h(X,Y)$.

Theorem: Solve the system of equations, $g(X,Y) = z_1$ and $h(X,Y) = w_1$. Then,

$$p_{Z,W}(z,w) = \frac{p_{X,Y}(x_1, y_1)}{\left|J\left(\frac{z,w}{x_1,y_1}\right)\right|}; J\left(\frac{z,w}{x_1,y_1}\right) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \frac{\partial z}{\partial x}\frac{\partial w}{\partial y} - \frac{\partial z}{\partial y}\frac{\partial w}{\partial x}$$

For instance, let us consider the transformation of functions of two RVs X and Y as $Z=XY$ and $W=Y$, then Jacobian for this transformation is

$$J\left(\frac{z,w}{x_1,y_1}\right) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\partial(xy)}{\partial x} & \frac{\partial(xy)}{\partial y} \\ \frac{\partial(y)}{\partial x} & \frac{\partial(y)}{\partial y} \end{vmatrix} = \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix} = y = w$$

Therefore,

$$p_{Z,W}(z,w) = \frac{p_{X,Y}\left(\frac{z}{w},w\right)}{|w|}$$

For the marginal density with respect to z, we have,

$$p_Z(z) = \int_{-\infty}^{\infty} \left|\frac{1}{w}\right| p_{X,Y}\left(\frac{z}{w},w\right) dw \tag{3.21}$$

For our case, the variables are $X = U$, $Y = V$, $u = w$ and $z = z$. Hence, the pdf of Z when U and V are independent RVs is given (H. Shin et al. 2003) by

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{|u|} p_U(u)\, p_V\left(\frac{z}{u}\right) du \\ &= \int_0^{\infty} \frac{1}{|u|} \frac{1}{(N_T-1)!} u^{N_T-1} e^{-u} \frac{1}{(N_R-1)!} \left(\frac{z}{u}\right)^{N_R-1} e^{-\frac{z}{u}} du \\ &= \frac{1}{\Gamma(N_T)\Gamma(N_R)} \int_0^{\infty} \frac{1}{u^{N_R-N_T+1}} (z)^{N_R-1} e^{-u-\frac{z}{u}} du \end{aligned}$$

If we assume that $t = u, x = 2\sqrt{z}$, we have,

$$p_Z(z) = \frac{1}{\Gamma(N_T)\Gamma(N_R)} \int_0^{\infty} \frac{1}{(t)^{N_R-N_T+1}} \left(\frac{x}{2}\right)^{2(N_R-N_T+N_T)-2} e^{-t-\frac{x^2}{4t}} dt$$

$$\Rightarrow \qquad p_Z(z) = \frac{2}{\Gamma(N_T)\Gamma(N_R)} \left(\frac{x}{2}\right)^{N_R-N_T+2N_T-2} \frac{1}{2}\left(\frac{x}{2}\right)^{N_R-N_T} \int_0^{\infty} \frac{1}{(t)^{N_R-N_T+1}} e^{-t-\frac{x^2}{4t}} dt$$

Hence,

$$p_Z(z) = \frac{2z^{\frac{N_T+N_R}{2}-1}}{\Gamma(N_T)\Gamma(N_R)} K_{N_R-N_T}\left(2\sqrt{z}\right); z \geq 0 \tag{3.22}$$

The n^{th} order modified Hankel function is expressed as

$$K_n(x) = \frac{1}{2}\left(\frac{x}{2}\right)^n \int_0^{\infty} \frac{1}{t^{n+1}} \exp\left(-t - \frac{x^2}{4t}\right) dt; |\arg x| < \frac{\pi}{2}, \mathrm{Re}\left(x^2\right) > 0$$

D. Gesbert et al. (2002) introduced a double-scattering MIMO channel model that includes both fading correlation and rank deficiency. It encompasses both Kronecker and Keyhole MIMO channel models.

3.6 MIMO channel parallel decomposition

3.6.1 What is MIMO channel parallel decomposition?

When both transmitter and receiver have multiple antennas, multiplexing gain improves (A. Goldsmith, 2005). The multiplexing gain of a MIMO system stems from the fact that a MIMO channel can be decomposed into a number of R_H parallel, independent channels. This process of decomposing the MIMO channel into R_H parallel, independent Gaussian channels is also referred to as MIMO channel parallel decomposition.

3.6.2 What is multiplexing (MUX) gain?

By sending independent data onto these separate channels, one can obtain an R_H-times gain in the data rate with respect to a system with a single antenna at the transmitter and receiver (SISO). This augmented data rate is called multiplexing (MUX) gain. Because R_H (the rank of matrix **H**) cannot exceed the number of columns or rows of **H**, it follows that $R_H \leq \min\left(N_R, N_T\right)$. If **H** is full rank for rich scattering environment like that of uncorrelated Rayleigh MIMO channel, then $R_H = \min\left(N_R, N_T\right)$. Other environments may lead to a low rank **H**: a channel with high correlation among the channel gain coefficients in **H** is rank deficit or a keyhole channel will have rank 1.

3.6.3 How do we parallel decompose a MIMO channel?

By precoding the channel input **x** and shaping the output **y**, we can decompose the MIMO channel in parallel. These precoding and combining matrices are obtained from the eigen-decomposition of **Q** and hence the resulting parallel sub-channels are referred to as eigenmodes. In precoding, the input **x** to the antennas is linearly transformed into the input vector

$$\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$$

In receiver shaping, we multiply the channel output **y** by $\mathbf{U}^H$ as shown in Fig. 3.2.

$$\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y} = \mathbf{U}^H\left(\mathbf{Hx} + \mathbf{n}\right)$$

The combined effect of transmit precoding and receiver shaping converts the MIMO channel into R_H parallel SISO channels with input $\tilde{\mathbf{x}}$ and output $\tilde{\mathbf{y}}$.

3.6.4 How do we obtain the U and V matrices?

The columns of **U** are the eigenvectors of $\mathbf{HH}^H$ and the columns of **V** are the eigenvectors of $\mathbf{H}^H\mathbf{H}$. **U** and **V** are orthonormal or unitary matrices. That means $\mathbf{U}^H\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^H\mathbf{V} = \mathbf{I}$. In order to obtain the

U and **V** matrices, we need the channel matrix, **H**. Hence in order to do transmit precoding we need channel state information at the transmitter (CSIT); and for receiver shaping we need the channel state information at the receiver (CSIR). This means channel state information (CSI) is required both at the transmitter and receiver, and therefore, such MIMO system is referred to as closed loop MIMO system.

3.6.5 How to obtain the SVD of channel matrix?

From Singular Value Decomposition (SVD) of the channel matrix, **H**, we have,

$$\mathbf{H} = \mathbf{U} \quad \mathbf{\Sigma} \quad \mathbf{V}^H$$

where, $\mathbf{\Sigma} = diag\left(\sigma_i\right)$ is a diagonal matrix of singular values of **H** (σ_i).

The singular values (σ_i) can be calculated as $\sigma_i = \sqrt{\lambda_i}$ where λ_i is the ith largest eigenvalue of the R_H non-zero eigenvalues of $\mathbf{HH}^H$.

Example 3.8

Find the SVD of a MIMO channel given as

$$\mathbf{H} = \begin{bmatrix} 1+2i & 2+3i \\ 3+4i & 4+5i \end{bmatrix}$$

Solution

Let us calculate the singular values by the above described method i.e., let us find the eigenvalues of $\mathbf{HH}^H$ and take the square root of the eigenvalues. Note that the **H** matrix is complex. Hence it may have complex eigenvalues. MATLAB command "[V D] = eig(H)" will give the diagonal matrix **D** with eigenvalues and **V** matrix whose columns are the eigenvectors.

For our case,

$$\mathbf{V} = \begin{bmatrix} 0.8070 & 0.4642 + 0.0482i \\ -0.5899 - 0.0283i & 0.8844 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} -0.3567 - 0.2631i & 0 \\ 0 & 5.3567 + 7.2631i \end{bmatrix}$$

We can see that the eigenvalues and eigenvectors of a complex **H** matrix may be also complex. For our case, eigenvalues of $\mathbf{HH}^H$ can be obtained as

$$\mathbf{V} = \begin{bmatrix} 0.8811 + 0.1037\text{i} & 0.4583 + 0.0539\text{i} \\ -0.4615 & 0.8871 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.1909 & 0 \\ 0 & 83.8091 \end{bmatrix}$$

Taking the square root and keeping in descending order of eigenvalues, we get,

$$\boldsymbol{\Sigma} = \begin{bmatrix} 9.1547 & 0 \\ 0 & 0.4369 \end{bmatrix}$$

We can also find the SVD decomposition $\mathbf{H} = \mathbf{U} \quad \boldsymbol{\Sigma} \quad \mathbf{V}^H$ directly which will be shown in the next example. The SVD of $\mathbf{H}$ for our example is

$$\mathbf{H} = \begin{bmatrix} -0.2271 - 0.4017i & 0.6889 + 0.5589i \\ -0.5238 - 0.7160i & -0.3899 - 0.2469i \end{bmatrix} \begin{bmatrix} 9.1547 & 0 \\ 0 & 0.4369 \end{bmatrix} \begin{bmatrix} -0.5971 & -0.8012 + 0.0401i \\ -0.8022 & 0.5963 - 0.0298i \end{bmatrix}$$

$\mathbf{U}$ and $\mathbf{V}$ are unitary matrices.

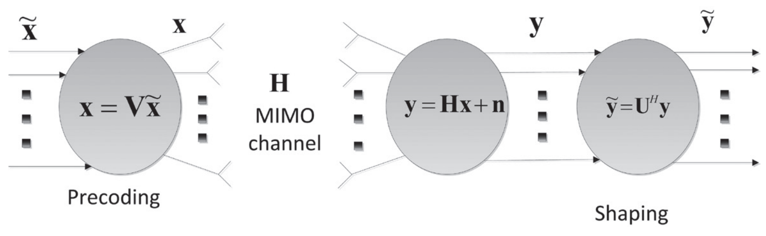


Fig. 3.2 Parallel decomposition of a MIMO channel using precoding and shaping

Hence, the combined effect of precoding at the transmit and shaping at the receiver is

$$\tilde{\mathbf{y}} = \mathbf{U}^H \left(\mathbf{U} \quad \boldsymbol{\Sigma} \quad \mathbf{V}^H \, \mathbf{x} + \mathbf{n} \right)$$

$$\Rightarrow \qquad \tilde{\mathbf{y}} = \mathbf{U}^H \left(\mathbf{U} \quad \boldsymbol{\Sigma} \quad \mathbf{V}^H \, \mathbf{V}\tilde{\mathbf{x}} + \mathbf{n} \right)$$

$$\Rightarrow \qquad \tilde{\mathbf{y}} = \boldsymbol{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{3.23}$$

It may be noted that the product of a unitary matrix with the noise vector does not modify the noise distribution. It may be good to write the above matrices component-wise in order to interpret clearly.

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_{R_H} \\ \tilde{y}_{R_H+1} \\ \vdots \\ \tilde{y}_{N_R} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{R_H} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_{R_H} \\ \tilde{x}_{R_H+1} \\ \vdots \\ \tilde{x}_{N_T} \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \\ \vdots \\ \tilde{n}_{R_H} \\ \tilde{n}_{R_H+1} \\ \vdots \\ \tilde{n}_{N_R} \end{bmatrix}$$

which can be further simplified as

$$\begin{aligned}
\tilde{y}_1 &= \sigma_1 \tilde{x}_1 + \tilde{n}_1 \\
\tilde{y}_2 &= \sigma_2 \tilde{x}_2 + \tilde{n}_2 \\
\vdots \quad & \ddots \quad \vdots \\
\tilde{y}_{R_H} &= \sigma_{R_H} \tilde{x}_{R_H} + \tilde{n}_{R_H} \\
\tilde{y}_{R_H+1} &= \tilde{n}_{R_H+1} \\
\vdots \quad & \ddots \quad \vdots \\
\tilde{y}_{N_R} &= \tilde{n}_{N_R}
\end{aligned}$$

Hence, the received signal for MIMO systems employing transmit precoding and receiver shaping will consist of transmitted signal multiplied by the channel gain with additive noise for R_H parallel, independent channels (R_H is the rank of the channel matrix **H**). The remaining $N_R - R_H$ parallel independent channels will receive only noise and eventually will be disposed of. Therefore, the precoding at the transmitter and shaping at the receiver converts the MIMO channel into R_H parallel, independent channels where i^{th} channel has input $\tilde{x}_i$, output $\tilde{y}_i$, noise $\tilde{n}_i$ and channel gain σ_i. For these non-interfering separate channels, the maximum likelihood (ML) decoding may be employed at the receiver. Besides, by sending independent data for each of the R_H parallel channels, the MIMO channel can support R_H times the data rate of a SISO system. However, the performance for each channel will be determined by its gain σ_i. Note that the channel **H** matrix is complex usually. We have considered real **H** matrix for easier illustration purpose in the following example.

Example 3.9

Find the parallel decomposition for the given MIMO channel.

$$\mathbf{H} = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \end{bmatrix}$$

Solution

Let us find the parallel decomposition for the given MIMO channel.

$$\mathbf{H} = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \end{bmatrix}.$$

The SVD of $\mathbf{H} = \mathbf{U} \ \mathbf{\Sigma} \ \mathbf{V}^H$ is given by

$$\mathbf{H} = \begin{bmatrix} -0.2148 & 0.8872 & 0.4082 \\ -0.5206 & 0.2496 & -0.8165 \\ -0.8263 & -0.3879 & 0.4082 \end{bmatrix} \begin{bmatrix} 1.6848 & 0 & 0 \\ 0 & 0.1068 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.4797 & -0.5724 & -0.6651 \\ -0.7767 & -0.0757 & 0.6253 \\ -0.4082 & 0.8165 & -0.4082 \end{bmatrix}$$

Note that the MATLAB command "[U S V] = svd(H)" could be used for obtaining the above three matrices. The diagonal entries of $\mathbf{\Sigma}$ are the square roots of the positive eigenvalues of $\mathbf{HH}^H$. Since there are two non-zero singular values and so $R_H = 2$, hence two parallel channels are there. The channel gains for the two channels are $\sigma_1 = 1.6848$ and $\sigma_2 = 0.1068$. We can notice that the second channel has a diminutive gain. Hence, this particular channel will give large detection error and inferior performance in terms of spectral efficiency.

Note that there are two types of eigenvalues and eigenvectors (left and right) of the matrix $\mathbf{H}$ denoted by λ_L, λ_R and $\mathbf{y}_L$, $\mathbf{y}_R$, respectively. They satisfy the following relations

$$(\mathbf{y}_L)^H \mathbf{H} = \lambda_L (\mathbf{y}_L)^H$$

$$\mathbf{H}\mathbf{y}_R = \lambda_R \mathbf{y}_R$$

Example 3.10

Show that the $\mathbf{HH}^H$ is a positive definite matrix if $\mathbf{H}$ is non-singular.

Solution

A non-zero vector $\mathbf{y} \in C^n$ is an eigenvector (left) of matrix $\mathbf{H} \in C^{n\times n}$ with eigenvalue λ_L if

$$\mathbf{y}^H\mathbf{H} = \lambda_L \mathbf{y}^H$$

Hence, $$\left(\mathbf{y}^H\mathbf{H}\right)^H = \left(\lambda_L \mathbf{y}^H\right)^H$$

$\Rightarrow$ $$\mathbf{H}^H\mathbf{y} = (\lambda_L)^* \mathbf{y}$$

Multiplying the above two equations, we have,

$$\mathbf{y}^H\mathbf{HH}^H\mathbf{y} = \lambda_L \mathbf{y}^H (\lambda_L)^* \mathbf{y} = |\lambda_L|^2 \mathbf{y}^H\mathbf{y} = |\lambda_L|^2 \|\mathbf{y}\|^2 > 0$$

Hence $\mathbf{HH}^H$ is always positive definite matrix.

Note that for complex $\mathbf{H}$ matrices, $\mathbf{\Sigma}$ remains real. Even for the negative eigenvalues of $\mathbf{H}$ matrix, the eigenvalues of $\mathbf{HH}^H$ become positive and hence $\mathbf{\Sigma}$ will have all positive singular values. Also note that every real symmetric ($\mathbf{H} = \mathbf{H}^T$) or complex Hermitian matrix ($\mathbf{H} = \mathbf{H}^H$) has real eigenvalues. Its eigenvectors can be chosen to be orthonormal. In MIMO analysis, since $\mathbf{HH}^H$ is always Hermitian, its eigenvalues are positive and their square roots give the singular values. You may refer to G. Strang, 2006 for further information on SVD.

For a detailed analysis of MIMO systems using MATLAB, readers may refer to L. Bai et al., 2012 and Y. S. Cho et al., 2010.

Review question 3.11 *How do we select the antenna with best channels?*

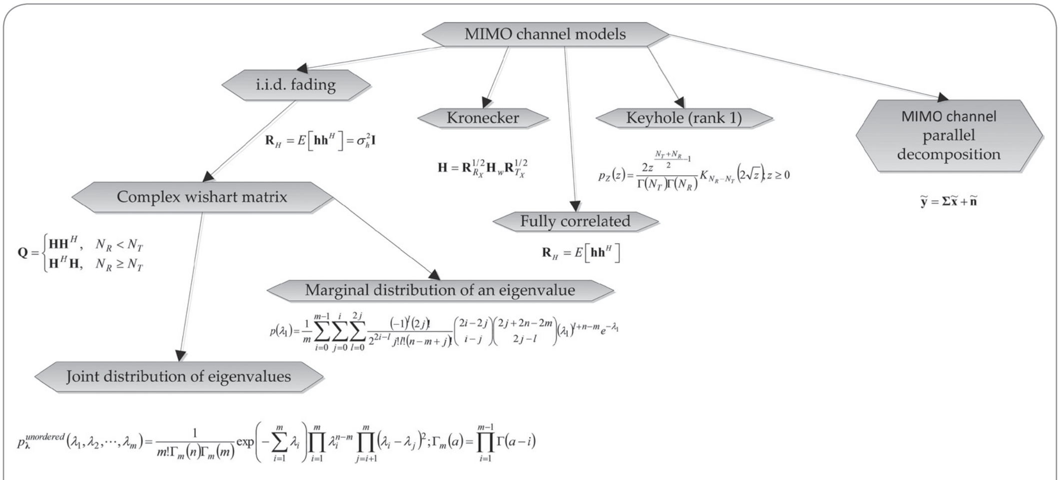


Fig. 3.3 Chapter 3 in a nutshell

3.7 Summary

Figure 3.3 shows the chapter in a nutshell. In MIMO channel models, we have studied i.i.d. fading, Kronecker, Keyhole and fully correlated channels. A sound understanding of these channel models is required for performance analysis of MIMO wireless communications. We have also studied about the parallel decomposition of MIMO channels.

Exercises

Exercise 3.1

Show that Kronecker model of MIMO channel could be expressed as $\mathbf{h} = N_C\left(0, \mathbf{R}_{T_X}^T \otimes \mathbf{R}_{R_X}\right)$ where, $\mathbf{h}$ is the vectorization of the channel matrix $\mathbf{H}$.

Exercise 3.2

Explain parallel decomposition of the deterministic MIMO channel (spatial multiplexed MIMO system) with representative equations and block diagram.

Exercise 3.3

Show that for i.i.d. Rayleigh fading MIMO channel, the correlation matrix is a diagonal matrix.

Exercise 3.4

Give an example of correlation matrix for (a) constant (b) circular and (c) exponential correlation models.
Following are MATLAB based exercises.

Exercise 3.5

Given that the channel **H** matrix for a 5×5 MIMO system as $\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.2 & 0.3 & 0.4 & 0.5 & 0.1 \\ 0.3 & 0.4 & 0.5 & 0.1 & 0.2 \\ 0.4 & 0.5 & 0.1 & 0.2 & 0.3 \\ 0.5 & 0.1 & 0.2 & 0.3 & 0.4 \end{bmatrix}$, find the multiplexing gain.

Exercise 3.6

Write short MATLAB program for calculating received signal for a 2×2 MIMO system employing QSPK over i.i.d. Rayleigh MIMO channel. Assume that noise is AWGN.

Exercise 3.7

Write short MATLAB program for calculating received signal for a 4×4 MIMO system employing 4-QAM over i.i.d. Rice MIMO channel. Assume that noise is AWGN.

Exercise 3.8

Assume that the MIMO channel is i.i.d. Rayleigh fading. Write short MATLAB program for calculating received signal for a 3×3 MIMO system employing 16-QAM.

Exercise 3.9

Assume that the MIMO channel is i.i.d. k-μ fading. Write short MATLAB program for calculating received signal for a 2×2 MIMO system employing 4-QAM.

Exercise 3.10

Assume that the MIMO channel is i.i.d. η-μ fading. Write short MATLAB program for calculating received signal for a 2×2 MIMO system employing 4-QAM.

References

1. Aalo, V. A. Aug. 1995. 'Performance of maximal-ratio diversity systems in a correlated Nakagami-fading environment'. *IEEE Trans. Commun.* COM-43. 2360–2369.

2. Ahn, K. S. and H. K. Baik. Sep. 2007. 'Asymptotic performance and exact symbol error probability analysis of orthogonal STBC in spatially correlated Rayleigh MIMO channel'. *IEICE Trans. Fundamentals*. E-90A(9).
3. Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd edition. New Delhi: John Wiley & Sons.
4. Bai, L. and J. Choi. 2012. *Low Complexity MIMO Detection*. London: Springer.
5. Biglieri, E. 2005. *Coding for Wireless Channels*. New York: Springer.
6. Bora, P. K. *Probability and Random Processes*, NPTEL (http://nptel.iitm.ac.in/).
7. Brown, T., E. D. Carvalho, and P. Kyritsi. 2012. *Practical Guide to the MIMO Radio Channel*. Chichester: John Wiley & Sons.
8. Capinski, M. and T. Zastawniak. 2001. *Probability through Problems*. New York: Springer Verlag.
9. Cho, Y. S., J. Kim, W. Y. Yang, and C.-G. Kang. 2010. *MIMO–OFDM Wireless Communications using MATLAB*. Singapore: Wiley.
10. Costa, N. and S. Haykin. 2010. *Multiple-input Multiple-output Channel Models*. New Jersey: John Wiley & Sons.
11. Du, K.-L. and M. N. S. Swamy. 2010. *Wireless Communications*. New Delhi: Cambridge University Press.
12. Duman, T. M. and A. Ghrayeb. 2007. *Coding for MIMO Communication Systems*. Chichester: John Wiley & Sons.
13. Edelman, E. 1989. *Eigenvalues and Condition Numbers of Random Matrices*. PhD Thesis, MIT.
14. Gershman, A. B. and N. D. Sidiropoulos. 2005. *Space-time Processing for MIMO Communications*. Chichester: John Wiley & Sons.
15. Gesbert, D., H. Bolcskei, D. A. Gore, and A. J. Paulraj. Dec. 2002. 'Outdoor MIMO wireless channels: models and performance prediction'. *IEEE Trans. Commun*. 50. 1926–1934.
16. Goldsmith, A. 2005. *Wireless Communications*. New Delhi: Cambridge University Press.
17. Gradshteyn, I. S. and I. M. Rhyzik. 2000. *Table of Integrals, Series and Products*. Oxford: Academic Press.
18. Janaswamy, R. 2001. *Radiowave Propagation and Smart Antennas for Wireless Communications*. New York: Kluwer Academic Publishers.
19. Kermoal, J. P., L. Schumacher, K. I. Pederson, P. E. Mogensen, and F. Frederiksen. Aug. 2002. 'A Stochastic MIMO Radio Channel Model With Experimental Validation'. *IEEE Journal on Selected Areas in Communications*. 20(6). 1211–1226.
20. Larsson, E. G. and P. Stoica. 2003. *Space-time Block Coding for Wireless Communications*. Cambridge: Cambridge University Press.
21. Ostges, C. and B. Clerckx. *2013. MIMO Wireless Networks*. Oxford: Elsevier Academic Press.
22. Papoulis, A. and S. U. Pillai. 2002. *Probability, Random Variables and Stochastic Processes*. New Delhi: Tata McGraw Hill.
23. Paul, B. S. and R. Bhattacharjee. Nov.–Dec. 2008. 'MIMO Channel Modelling: A Review'. *IETE Technical Review*. 25(6). 315–319.
24. Paulraj, A., R. Nabar, and D. Gore. 2003. *Introduction to Space-time Wireless Communications*. Cambridge: Cambridge University Press.
25. Proakis, J. G. and M. Salehi. 2007. *Digital Communications*. New York: McGraw Hill.
26. Rao, V. V. *Principles of Communications*. NPTEL (http://nptel.iitm.ac.in/).
27. Ross, S. 2002. *A First Course in Probability*. New Delhi: Pearson.
28. Seber, G. A. F. and A. J. Lee. 2003. *Linear Regression Analysis*. New Jersey: John Wiley & Sons.
29. Shin, H. and J. H. Lee. Jan. 2003. 'Effect of keyholes on the symbol error rate of space-time block codes'. *IEEE Comm. Lett*. 7. 27–29.
30. Sibille, A., C. Oestges, and A. Zanella. 2011. *MIMO from Theory to Implementation*. Oxford: Elsevier Academic Press.

31. Simon, M. K. and M.-S. Alouini. 2005. *Digital Communications over Fading Channels*. New Jersey: Wiley.
32. Strang, G. 2006. *Linear Algebra and its Applications*. New Delhi: Cengage Learning India.
33. Stuber, G. L. 2001. *Principles of Mobile Communication*. Dordrecht: Kluwer Academic Publishers.
34. van den Bos, A. Mar. 1995. 'The Multivariate Complex Normal Distribution- A Generalization'. *IEEE Trans. Inform. Theory*. 41(2). 537–539.
35. Vucetic, B. and J. Yuan. 2003. *Space-time Coding*. Chichester: John Wiley & Sons.
36. Weichselberger, W. 2003. *Spatial Structure of Multiple Antenna Radio Channels*. Institut fur Nachrichtentechnik und Hochfrequenztechnik, Technische Universitat Wien, PhD thesis.
37. Weichselberger, W., M. Herdin, H. Özcelik, and E. Bonek. Jan. 2006. 'A stochastic MIMO channel model with joint correlation of both link ends.' *IEEE Trans. Wireless Comm.* 5(1). 90–100.
38. Wymeersch, H. 2007. *Iterative Receiver Design*. Cambridge: Cambridge University Press.
39. Yacoub, M. D. Feb. 2007. 'The k-µ distribution and the η-µ distribution'. *IEEE Antennas Propagat. Mag.* 49(1). 68–81.
40. Yates, R. D. and D. J. Goodman. 2005. *Probability and Stochastic Processes*. New Delhi: John Wiley & Sons.

Power Allocation in MIMO Systems

CHAPTER 4

4.1 Introduction

Since using SVD, we can decompose a MIMO channel into R_H parallel Gaussian channels, where R_H is the rank of the MIMO channel matrix, we will use the knowledge on the capacity of the parallel Gaussian channel (see Appendix C) to find the capacity of a MIMO channel for uniform and adaptive power allocation scheme. Uniform power allocation is employed when the channel state information (CSI) is available at the receiver but not at the transmitter (open loop MIMO system). We can use adaptive power allocation based on Water-filling algorithm when CSI is available at the receiver as well as the transmitter (closed loop MIMO system). We will also discuss near optimal power allocation for high and low SNR cases.

Note that power allocation plays a significant role in deciding MIMO capacity. Power allocation was not an important issue in SISO since only single antenna was employed at the transmitter and receiver. Usually we allocate all the power to the single transmit antenna. But for MIMO it is one of the most important factors for increasing capacity. We have numerous antennas at the transmitter and receiver for MIMO case. The fundamental question is how much power we allocate to each transmit antennas. Hence if we allot power equally to all transmit antennas or unequally to each transmit antenna, capacity of the MIMO channel will be definitely different. If this is the case, then how we optimally allocate power to MIMO channels can be considered as an optimization problem to maximize capacity. To allocate power adaptively we need the CSI at the transmitter, also since power allocation is done at the transmitter. Intuitively we will allocate more power to better channels than the bad channels. We may not allocate any power at all to some of the worst channels. We will discuss these in detail in the following sections. In practical scenarios, we can allocate power near optimally for MIMO channels for two cases: high and low SNR regimes.

4.2 Uniform power allocation

The capacity indicates the best viable transmission data rate over the channel for miniscule probability of error. Shanon provided the expression of the achievable communication rate of a channel with noise (C. E. Shanon, 1948). If the transmission rate is greater than the capacity, the system is in outage and the receiver makes decoding errors with a non-negligible probability. In this section, let us derive the capacity of MIMO channels for uniform power allocation. Usually the channel state information is available at the receiver (CSIR) but not available at the transmitter (no CSIT). Such a MIMO system is referred to as Open loop MIMO system. Assume that input signal vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_T} \end{bmatrix}$. $x_1, x_2, \ldots, x_i, \ldots$ and x_{N_T} are mutually independent (uncorrelated). Assume that the overall power of the transmitted signal is P and equal power is given to individual transmit antenna.

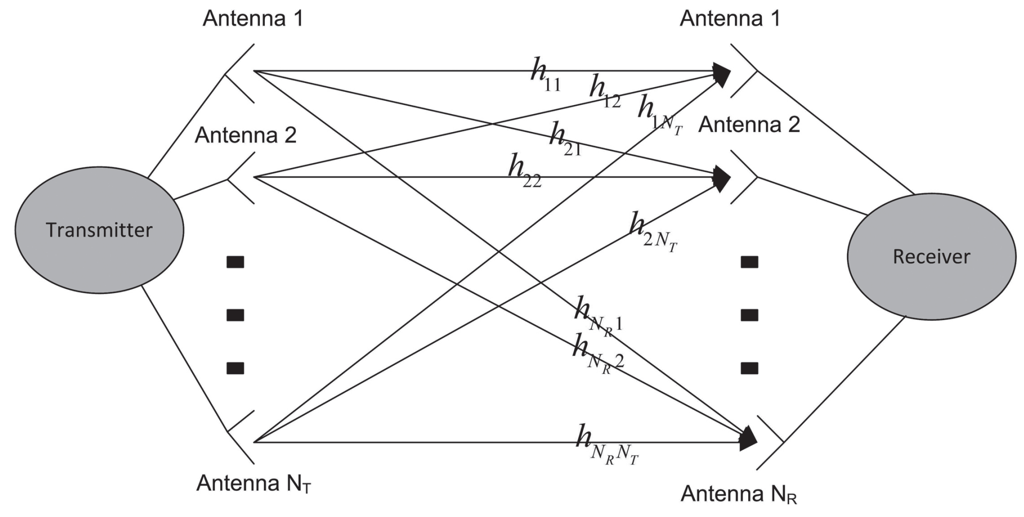


Fig. 4.1 $N_R \times N_T$ MIMO channel

Now consider the covariance matrix of **x**

$$\mathbf{R_{xx}} = \begin{bmatrix} Var\left(|x_1|^2\right) & Cov\left(x_1, x_2^*\right) & \cdots & Cov\left(x_1, x_{N_T}^*\right) \\ Cov\left(x_2, x_1^*\right) & Var\left(|x_2|^2\right) & \cdots & Cov\left(x_2, x_{N_T}^*\right) \\ \vdots & \vdots & \ddots & \vdots \\ Cov\left(x_{N_T}, x_1^*\right) & Cov\left(x_{N_T}, x_2^*\right) & \cdots & Var\left(|x_{N_T}|^2\right) \end{bmatrix} = \begin{bmatrix} \frac{P}{N_T} & 0 & \cdots & 0 \\ 0 & \frac{P}{N_T} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{P}{N_T} \end{bmatrix} = \frac{P}{N_T} \mathbf{I}_{N_T} \quad (4.1)$$

Hence the total power is given by

$$P = trace\ (\mathbf{R_{xx}})$$

Trace equals the sum of diagonal elements of a matrix.

Let us assume that the channel gain h_{ij} are available at the receiver (CSIR) but not accessible at the transmitter (no CSIT). Besides let us presume that the channel gain is normalized which means

$$\sum_{j=1}^{N_T} \left|h_{ij}\right|^2 = N_T \text{ for deterministic channel}$$

$$\sum_{j=1}^{N_T} E\left(\left|h_{ij}\right|^2\right) = N_T \text{ for random channel}$$

Then we can write the received signal for frequency flat channel (T. Brown et al., 2012),

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \tag{4.2}$$

where covariance of the noise vector is $\mathbf{R_{nn}} = \sigma^2 \mathbf{I}_{N_R}$.

Now the average SNR at each receive antenna is

$$\gamma_i = \frac{\sum_{j=1}^{N_T} \left|h_{ij}\right|^2 \frac{P}{N_T}}{\sigma^2} \tag{4.3}$$

where, σ^2 is the noise variance.

Note that h_{ij} may be deterministic or random. For random case, h_{ij} and γ_i is a RV. For instance, for Rayleigh fading case, h_{ij} is complex Gaussian and γ_i is central Chi-square random variable with $2N_T$ degrees of freedom.

Review question 4.1 *What is the average SNR at each receive antenna?*

We can write the $N_R \times N_T$ MIMO channel matrix using SVD (A. Goldsmith, 2005) as follows:

$$\mathbf{H} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H$$

where, $\boldsymbol{\Sigma}$ is $N_R \times N_T$ positive diagonal matrix, $\mathbf{U}$ is a $N_R \times N_R$ unitary matrix and $\mathbf{V}$ is a $N_T \times N_T$ unitary matrix.

The diagonal elements of $\boldsymbol{\Sigma}$ are the positive square roots of eigenvalues of matrix $\mathbf{HH}^H$.

Using section 3.6 on parallel decomposition of MIMO channels, we have,

$$\tilde{y}_i = \sigma_i \tilde{x}_i + \tilde{n}_i; i = 1, 2, \cdots, R_H$$

Using the Shannon capacity formula for R_H parallel Gaussian channels, the channel capacity for equal power allocation is

$$C = W \sum_{i=1}^{R_H} \log_2\left(1 + \frac{P_{r_i}}{\sigma^2}\right) = W \sum_{i=1}^{R_H} \log_2\left(1 + \frac{\lambda_i P}{N_T \sigma^2}\right) = W \log_2 \prod_{i=1}^{R_H}\left(1 + \frac{\lambda_i P}{N_T \sigma^2}\right) \tag{4.4a}$$

where, W is the bandwidth of the channel.

Review question 4.2 *What is the capacity of R_H parallel Gaussian channels?*

Q is the Wishart matrix defined as

$$\mathbf{Q} = \begin{cases} \mathbf{H}\mathbf{H}^H, & N_R < N_T \\ \mathbf{H}^H\mathbf{H}, & N_R \geq N_T \end{cases}$$

If $\lambda_i = \sigma_i^2; i = 1, 2, \cdots, R_H$ are eigenvalues of **Q**, then

$$\prod_{i=1}^{R_H} \left(1 + \frac{P\lambda_i}{N_T\sigma^2}\right) = \det\left(\mathbf{I}_{R_H} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)$$

The above relation is proved in example 4.1. Therefore the capacity formula (B. Vucetic et al., 2003) becomes

$$C = W\log_2\left\{\det\left(\mathbf{I}_{R_H} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)\right\} \tag{4.4b}$$

Note that the above capacity formula is a function of **Q** which is dependent on **H**. Hence, we need CSIR to calculate capacity for equal power allocation. The non-zero eigenvalues of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{H}^H\mathbf{H}$ are the identical. Hence the spectral efficiencies of the channels with the matrices **H** and $\mathbf{H}^H$ are equal (reciprocity). Since the channel elements are RVs, this formula represents instantaneous capacities or mutual information. The mean channel capacity can be obtained by averaging over all realizations of the channel coefficients.

Review question 4.3 *What is the SVD of a MIMO channel matrix* **H***?*

Review question 4.4 *What is instantaneous capacity of a MIMO channel for equal power allocation?*

Example 4.1

Prove that, $$\prod_{i=1}^{R_H} \left(1 + \frac{P\lambda_i}{N_T\sigma^2}\right) = \det\left(\mathbf{I}_{R_H} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)$$

Solution

Since λ_i are eigenvalues of **Q** matrix,

we have, $$\mathbf{Q}\mathbf{x}_i = \lambda_i\mathbf{x}_i; i = 1, 2, \cdots, R_H$$

where, $\mathbf{x}_i$ are the eigenvectors for λ_i and **Q** has R_H non-zero eigenvalues (rank of **Q** matrix is R_H).

Therefore,

$$\left(\frac{P}{N_T\sigma^2}\mathbf{Q}\right)\mathbf{x}_i = \left(\frac{P}{N_T\sigma^2}\lambda_i\right)\mathbf{x}_i; i = 1, 2, \cdots, R_H$$

Since the identity matrix has all its eigenvalues equal as 1. Note that for a diagonal matrix, the diagonal elements equal the eigenvalues.

Hence,

$$\left(\mathbf{I}_{R_H} + \frac{P}{N_T\sigma^2}\mathbf{Q}\right)\mathbf{x}_i = \left(1 + \frac{P}{N_T\sigma^2}\lambda_i\right)\mathbf{x}_i; i = 1,2,\cdots,R_H$$

We also know that determinant of a matrix equals the multiplication of its eigenvalues. Therefore,

$$\prod_{i=1}^{R_H}\left(1 + \frac{P\lambda_i}{N_T\sigma^2}\right) = \det\left(\mathbf{I}_{R_H} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)$$

This is a very crude way of proving but it serves our purpose well.

4.3 Adaptive power allocation

Let us presume that CSI is available at the transmitter. Usually the channel state information is available at the receiver (CSIR). If the receiver sends the CSI to the transmitter through a feedback channel, then, the channel state information is also available at the transmitter (CSIT). Such a MIMO system is referred to as Closed loop MIMO system. Now we may distribute power adaptively to individual transmit antenna to boost the spectral efficiency. Hence, the channel capacity may be expressed as

$$C = W\sum_{i=1}^{R_H}\log_2\left(1 + \frac{\lambda_i P_i}{\sigma^2}\right) \tag{4.5a}$$

where, P_i is the transmit power at the i^{th} transmit antenna.

Example 4.2

Use Water-filling algorithm to maximize the channel capacity for adaptive power allocation.

Solution

We need to maximize C by choosing P_i properly. Water-filling algorithm can be utilized in obtaining the capacity under the ensuing power constraint

$$\sum_{i=1}^{N_T} P_i = P$$

where, P_i is the power allotted to the i^{th} transmit antenna and P is the overall power which shall be kept fixed.

Hence the capacity can be written as

$$C = W\sum_{i=1}^{R_H}\log_2\left(1 + \frac{P\lambda_i P_i}{P\sigma^2}\right) = W\sum_{i=1}^{R_H}\log_2\left(1 + \frac{\gamma_i P_i}{P}\right); \gamma_i = \frac{P\lambda_i}{\sigma^2} \tag{4.5b}$$

Using the method of Lagrange multipliers (G. B. Arfken et al., 2005), let us introduce the cost or objective function as

$$F = \sum_{i=1}^{R_H} \log_2\left(1 + \frac{\gamma_i P_i}{P}\right) + \zeta\left(P - \sum_{i=1}^{R_H} P_i\right) \tag{4.5c}$$

where, ζ is the Lagrange multiplier.

The unknown transmit power P_i are determined by setting the partial derivative of the cost or objective function F to zero.

$$\frac{dF}{dP_i} = 0$$

$$\Rightarrow \quad \frac{d\left\{\log_2\left(1 + \frac{\gamma_i P_i}{P}\right) - \zeta P_i\right\}}{dP_i} = 0$$

$$\Rightarrow \quad \frac{1}{\ln(2)} \frac{1}{1 + \frac{\gamma_i P_i}{P}} \frac{\gamma_i}{P} - \zeta = 0$$

$$\Rightarrow \quad \frac{1}{\frac{P}{\gamma_i} + P_i} - \zeta \ln(2) = 0$$

$$\Rightarrow \quad P_i = \frac{1}{\zeta \ln(2)} - \frac{P}{\gamma_i}$$

$$\Rightarrow \quad \frac{P_i}{P} = \frac{1}{\zeta \ln(2)} - \frac{1}{\gamma_i}$$

$$\Rightarrow \quad \frac{P_i}{P} = \frac{1}{\gamma_0} - \frac{1}{\gamma_i}$$

Since power allocated should be greater than or equal to zero ($P_i \geq 0$), we have,

$$\Rightarrow \quad \frac{P_i}{P} = \left[\frac{1}{\gamma_0} - \frac{1}{\gamma_i}\right]^+$$

where the notation $[k]^+ = \begin{cases} k, & k > 0 \\ 0, & k \leq 0 \end{cases}$

So the power constraint is

$$\sum_{i=1}^{R_H} \frac{P_i}{P} = \sum_{i=1}^{R_H} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_i}\right)^+ = 1$$

The MIMO channel capacity may be rewritten as follows.

$$C = W \sum_{i=1}^{R_H} \log_2\left(1 + \frac{\gamma_i P_i}{P}\right) = W \sum_{i=1}^{R_H} \log_2\left(1 + \gamma_i \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_i}\right)^+\right) = W \sum_{i:\gamma_i \geq \gamma_0}^{R_H} \log_2\left(\frac{\gamma_i}{\gamma_0}\right)^+ \tag{4.5d}$$

Example 4.3

Find the spectral efficiency and optimal power distribution for the MIMO channel $\mathbf{H} = \begin{bmatrix} 1+2i & 2+3i \\ 3+4i & 4+5i \end{bmatrix}$, assuming $\gamma = \frac{P}{\sigma^2} = 5dB$ and BW=1 Hz.

Solution

The SVD of $\mathbf{H} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^H$ is given by

$$\mathbf{H} = \begin{bmatrix} -0.2271 - 0.4017i & 0.6889 + 0.5589i \\ -0.5238 - 0.7160i & -0.3899 - 0.2469i \end{bmatrix} \begin{bmatrix} 9.1547 & 0 \\ 0 & 0.4369 \end{bmatrix} \begin{bmatrix} -0.5971 & -0.8012 + 0.0401i \\ -0.8022 & 0.5963 - 0.0298i \end{bmatrix}$$

The singular values of the channel are $\sqrt{\lambda_1} = 9.1547$ and $\sqrt{\lambda_2} = 0.4369$.

Note that $\gamma_i = \gamma\lambda_i = \frac{P}{\sigma^2}\lambda_i = 3.1623\lambda_i$. Hence, $\gamma_1 = 265.0276$ and $\gamma_2 = 0.6037$.

Considering that power is distributed to the two parallel channels, the power constraint becomes

$$\sum_{i=1}^{2}\left(\frac{1}{\gamma_0} - \frac{1}{\gamma_i}\right) = 1$$

$$\Rightarrow \qquad \frac{2}{\gamma_0} = 1 + \sum_{i=1}^{2}\frac{1}{\gamma_i} = 2.6602$$

Note that $\gamma_2 < \gamma_0 = 0.751$. It means that $\frac{P_2}{P} = \left[\frac{1}{\gamma_0} - \frac{1}{\gamma_2}\right]$ is a negative number and hence $P_2 = 0$. Therefore, the second channel is not allocated any power. Then the power constraint yields

$$\frac{1}{\gamma_0} - \frac{1}{\gamma_1} = 1$$

$$\Rightarrow \qquad \frac{1}{\gamma_0} = 1 + \frac{1}{\gamma_1} = 1.0038$$

For this case, $\gamma_1 > \gamma_0 = 0.99624$. The capacity is given by

$$C = \log_2\left(\frac{\gamma_1}{\gamma_0}\right) = \log_2\left(\frac{265.0276}{0.99624}\right) = 8.0554 \text{ bits/sec/Hz}$$

Interpretation on $\log_2$ (1+SNR) curve

Let us try to analyze the $\log_2(1 + SNR)$ curve for low and high SNR regions (D. Tse et al., 2005). $\log_2(1 + SNR)$ is a concave function. It implies that augmenting the SNR experience decree of diminishing marginal returns. In other words, the greater the SNR, the lesser is the effect on spectral efficiency. To be precise, let us study the approximations for high and low SNR regions:

$$\log_2(1 + SNR) \approx SNR \log_2 e \text{ when } SNR \approx 0 \tag{4.6a}$$

and
$$\log_2(1 + SNR) \approx \log_2 SNR \text{ when } SNR >> 1 \tag{4.6b}$$

When the power is low, the capacity increases linearly with the received power P (equation 4.6a). When the SNR is high, the capacity increases logarithmically with the received power P (equation 4.6b).

Review question 4.5 *What is the MIMO channel capacity for adaptive power allocation?*

Review question 4.6 *How does $\log_2 (1 + SNR)$ increase with SNR for high and low SNR regions?*

4.4 Near optimal power allocation

We can find near optimal power allocation for the high and low SNR regions. Let us find it for the high SNR region first. Assume R_H is the rank of the channel matrix $\mathbf{H}$.

4.4.1 High SNR

For large SNR, the water level is deep, it is advantageous to distribute equal power to all sub-channel with the non-zero eigenvalues. Hence, at high SNR,

$$C = W\sum_{i=1}^{R_H} \log_2\left(1 + \frac{\lambda_i P_i}{\sigma^2}\right) \approx W\sum_{i=1}^{R_H} \log_2\left(\frac{\lambda_i P_i}{\sigma^2}\right)$$

$$\Rightarrow \qquad C \approx W\sum_{i=1}^{R_H} \log_2\left(\frac{\lambda_i P}{\sigma^2 R_H}\right) = WR_H \log_2\left(\frac{P}{\sigma^2}\right) + W\sum_{i=1}^{R_H} \log_2\left(\frac{\lambda_i}{R_H}\right) \tag{4.7a}$$

Hence for large SNR the spectral efficiency of the MIMO channel is amply greater than the spectral efficiency of the SISO channel. The capacity increases linearly with the rank of the MIMO channel matrix (R_H). The rank gives a first order insight of the channel capacity. For a detailed investigation, we need to see how large non-zero singular values are. Note that among the channels with the same total power gain, the one that has the highest capacity is the one with all the singular values equal. More generally, the less spread out the singular values, the larger the capacity in the high SNR regime. In numerical analysis, $\kappa(\mathbf{H}) = \frac{\max\left(\sqrt{\lambda_i}\right)}{\min\left(\sqrt{\lambda_i}\right)}$ where λ_i are the eigenvalues of $\mathbf{Q}$ is defined as condition number of the matrix. The matrix is accustomed as well-conditioned if the condition number is adjacent to 1. Hence, one may conclude that a well-conditioned channel matrix expedites communication for large SNR region.

4.4.2 Low SNR

For low SNR, it is advantageous to supply power to the strongest eigenmode exclusively. We need to fill water of the deepest vessel. This means the sender opportunistically transmits exclusively to the finest channel. The resulting capacity is given below.

$$C = W \sum_{i=1}^{R_H} \log_2\left(1 + \frac{\lambda_i P_i}{\sigma^2}\right) \approx W \sum_{i=1}^{R_H} \frac{\lambda_i P_i}{\sigma^2} \log_2(e)$$

$$\Rightarrow \qquad C \approx W \frac{\lambda_{\max} P}{\sigma^2} \log_2(e) \qquad (4.7b)$$

The MIMO channel provides a power gain of max $\{\lambda_i\}$ in comparison to that of the SISO case. For such channel, the rank or condition number of the channel matrix gives no sense. The best power distribution scheme for low SNR is to supply the entire power to the finest sub-channel (the one with the largest eigenvalue) which is also known as opportunistic communication.

Review question 4.7 *How to allocate power to MIMO channels for low and high SNR regions?*

4.5 Summary

Figure 4.2 shows the chapter in a nutshell. In this chapter, we have found out the MIMO channel capacity for uniform and adaptive power allocation. In adaptive power allocation, Water-filling algorithm has been employed. CSIR is required for uniform power allocation whereas both CSIR and CSIT are required for adaptive power allocation. Near optimal power allocation of MIMO channels for high and low SNR cases are also discussed.

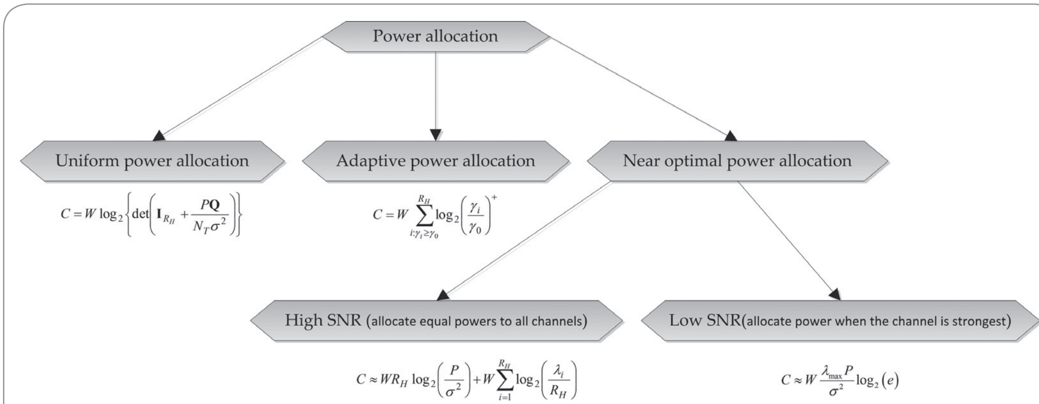


Fig. 4.2 Chapter 4 in a nutshell

Exercises

Exercise 4.1

Find the spectral efficiency and best power allocation for the MIMO channel whose $\mathbf{H} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.9 \end{bmatrix}$, assuming $\gamma = \frac{P}{\sigma^2} = 10dB$ and BW = 1 Hz.

Exercise 4.2

What are the key performance-deciding parameters of MIMO channel capacity for high and low SNR cases?

Following are MATLAB based exercises.

Exercise 4.3

Find the spectral efficiency and best power allocation for the MIMO channel $\mathbf{H} = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \end{bmatrix}$, assuming $\gamma = \frac{P}{\sigma^2} = 3dB$ and BW=1 Hz.

References

1. Arfken, G. B. and H. J. Weber. 2005. *Mathematical Methods for Physicists*. New Delhi: Academic Press.
2. Brown, T., E. D. Carvalho, and P. Kyritsi. 2012. *Practical Guide to the MIMO Radio Channel*. Indianapolis: John Wiley & Sons.
3. Shannon, C. E. 1948. 'A Mathematical Theory of Communication'. *Bell Labs Tech. J.* Vol. 27, page numbers, 379–423.
4. Tse, D. and P. Viswanath. 2005. *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press.
5. Vucetic, B. and J. Yuan. 2003. *Space-time Coding*. Chichester: John Wiley & Sons.

Channel Capacity of Simplified MIMO Channels

CHAPTER 5

5.1 Introduction

Capacity of a MIMO channel equals the maximum data rate that can be transmitted over the channel with arbitrarily small probability of error and it is the subject of focus for this chapter. We will first find the capacity of some cases of MIMO channels with fixed coefficients. We will take up some simplified cases: (a) SISO channels (b) SIMO channels (c) MISO channels (d) unity MIMO channel and (e) identity MIMO channel. Real life or practical MIMO channels are not deterministic but random. We will find the ergodic capacity and outage probability for some of the above fading channels. Parallel Gaussian channels are discussed in Appendix C. Appendix C also reviews information theory which may be of use to readers who are not familiar with the information theory. For further studies on this subject, one may also refer to T. M. Cover et al., 2006.

5.2 Capacity for deterministic MIMO channel

It is good to start with a discussion on MIMO capacity for simple and deterministic MIMO channels. In this section, let us find the maximum transmission rate for MIMO channel with fixed or constant channel coefficients. Generally channel state information (CSI) is available at the receiver and not at the transmitter.

5.2.1 SISO channel

Let us consider a SISO channel with $N_T = N_R = 1$ and $h = 1$. The Shannon formula for AWGN channel capacity is

$$C = W \log_2 \left(1 + \frac{P}{\sigma^2}\right) \tag{5.1}$$

where, P is the signal power and σ^2 is the variance of the noise.

5.2.2 SIMO channel

For MIMO we have, $C = W\sum_{i=1}^{R_H}\log_2\left(1 + \frac{\lambda_i P}{N_T\sigma^2}\right)$. Since SIMO channel is a vector ($R_H = 1$), its SVD will have a single singular value equals to the Frobenius norm of the vector. There is only one transmitting antenna and therefore SNR is $\frac{P}{\sigma^2}$. Hence,

$$C = W\log_2\left(1 + \frac{P}{\sigma^2}\|\mathbf{h}\|^2\right) \tag{5.2a}$$

If the channel coefficients are equal and normalized $|h_1|^2 = |h_2|^2 = \cdots = |h_{N_R}|^2 = 1$, the capacity becomes

$$C = W\log_2\left(1 + \sum_{i=1}^{N_R}|h_i|^2\frac{P}{\sigma^2}\right) = W\log_2\left(1 + \frac{N_R P}{\sigma^2}\right) \tag{5.2b}$$

If we now compare this capacity with the capacity of a single antenna channel, we see that SIMO increases the effective SNR and provides a power gain but no MUX gain.

How does one achieve this capacity practically?

Consider SVD of the channel vector

$$\mathbf{h} = \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} = \sqrt{\frac{1}{N_R}}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\bullet\sqrt{N_R}\bullet 1 = \mathbf{U}\quad\Sigma\quad\mathbf{V}^H$$

Hence all power P goes to the single antenna, no transmitter precoding.

After receiver shaping, the total received signal voltage is

$$\sqrt{\frac{1}{N_R}}\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\sqrt{P} = \sqrt{N_R P}$$

Hence the signal power is $N_R P$.

5.2.3 MISO channel

In this system, there are N_T transmit antennas and only one receive antenna ($N_R = 1$). The channel is represented by the vector: $\mathbf{h} = \begin{bmatrix}h_1 & h_2 & \cdots & h_{N_T}\end{bmatrix}$. We have, for MIMO channel,

$$C = W\log_2\left\{\det\left(\mathbf{I}_{R_H} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)\right\}$$

where $\mathbf{Q} = \begin{cases} \mathbf{H}\mathbf{H}^H, & N_R < N_T \\ \mathbf{H}^H\mathbf{H}, & N_R \geq N_T \end{cases}$

Since $N_R < N_T$, we have for MISO channel, $\mathbf{Q} = \mathbf{h}\mathbf{h}^H = \sum_{j=1}^{N_T} \left|h_j\right|^2$ and also $\mathbf{I}_{R_H} = 1$. Hence, we get the capacity for MISO channel for equal power allocation as

$$C = W\log_2\left(1 + \frac{P\mathbf{h}\mathbf{h}^H}{N_T\sigma^2}\right) = W\log_2\left(1 + \sum_{j=1}^{N_T}\left|h_j\right|^2 \frac{P}{N_T\sigma^2}\right)$$

If the channel coefficients are equal and normalized $\left|h_1\right|^2 = \left|h_2\right|^2 = \ldots = \left|h_{N_T}\right|^2 = 1$, the capacity becomes

$$C_{\text{equal}} = W\log_2\left(1 + \frac{P}{\sigma^2}\right) \tag{5.3a}$$

The capacity doesn't increase with the number of transmit antennas. This is the case when we allot the power equally for all transmitting antennas. If we assume CSI is available at the transmitter, we can apply water-filling algorithm. Since the rank of the vector channel matrix is one, there is only one nonzero eigenvalue of $\mathbf{hh}^{\mathrm{H}}$ and is given by $\lambda = \sum_{j=1}^{N_T}\left|h_j\right|^2$. So we get the capacity for equal and normalized channel coefficients as

$$C_{\text{waterfilling}} = W\log_2\left(1 + \frac{\lambda P}{\sigma^2}\right) = W\log_2\left(1 + \sum_{j=1}^{N_T}\left|h_j\right|^2 \frac{P}{\sigma^2}\right) = W\log_2\left(1 + \frac{N_T P}{\sigma^2}\right) \tag{5.3b}$$

Here we see that there is a power gain for MISO channel when the power is allotted using the water-filling algorithm but no MUX gain.

How does one achieve this capacity practically?

Consider SVD of the channel vector

$$\mathbf{h} = \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix} = 1 \bullet \sqrt{N_T} \bullet \sqrt{\frac{1}{N_T}}\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix} = \mathbf{U} \quad \Sigma \quad \mathbf{V}^H$$

Hence after transmitter precoding, each antenna sends equal signal power and voltage.

$$\tilde{\mathbf{x}} = \sqrt{\frac{1}{N_T}}\begin{bmatrix}1 \\ 1 \\ \vdots \\ 1\end{bmatrix} x$$

No receiver shaping, the total received signal voltage is $N_T\sqrt{\frac{P}{N_T}}$

Hence the signal power is $N_T P$.

5.2.4 MIMO channel with unity channel matrix

Consider **H** with all-1 matrix, a special case of spatial interference. Its SVD is

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{1/N_R} \\ \vdots \\ \sqrt{1/N_R} \\ \sqrt{1/N_R} \end{bmatrix} \left[\sqrt{N_R N_T}\right] \left[\sqrt{1/N_T} \quad \sqrt{1/N_T} \quad \cdots \quad \sqrt{1/N_T}\right]$$

Since there is only one singular value, $R_H = 1$, $\sqrt{\lambda_1} = \sqrt{N_R N_T}$, and hence $\lambda_1 = N_R N_T$. We can allot all the power P to single channel with non-zero eigenvalue yielding the channel capacity

$$C = W \log_2 \left(1 + N_R N_T \frac{P}{\sigma^2}\right) \tag{5.4}$$

In this case, we see that there is diversity gain from proper combination of the received signals but no rate or MUX gain.

How does one achieve this capacity practically?

After transmitter precoding, each antenna sends equal signal power and voltage.

$$\tilde{\mathbf{x}} = \sqrt{\frac{1}{N_T}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x$$

After receiver shaping, the total received signal voltage is

$$\sqrt{\frac{1}{N_R}} [1 \quad 1 \quad \cdots \quad 1] \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{P}{N_T}} \\ \sqrt{\frac{P}{N_T}} \\ \vdots \\ \sqrt{\frac{P}{N_T}} \end{bmatrix} = \sqrt{\frac{1}{N_R}} [1 \quad 1 \quad \cdots \quad 1] \begin{bmatrix} \sqrt{N_T P} \\ \sqrt{N_T P} \\ \vdots \\ \sqrt{N_T P} \end{bmatrix} = \sqrt{N_T N_R P}$$

Hence the signal power is $N_T N_R P$.

5.2.5 MIMO channel with identity channel matrix

For identity matrix, SVD of **H** gives **U** and **V** matrix as identity matrices only. Hence this MIMO system needs no transmit precoding and receiver shaping. Assume that **H** matrix size is $R_H \times R_H$. Due to the structure of **H**, there is no spatial interference here and transmission occurs over R_H

parallel AWGN channels, each with SNR $\left[\frac{P}{R_H \sigma^2}\right]$ for equal power allocation and hence with capacity $W \log_2\left[1+\frac{P}{R_H \sigma^2}\right]$ bit/dimension pair for each channel. Since singular values and eigenvalues of identity matrix equals 1. Channel gains for each path also equal 1. Thus for parallel R_H Gaussian channels,

$$C = R_H W \log_2\left[1+\frac{P}{R_H \sigma^2}\right] \tag{5.5}$$

We see here that we have a rate gain, since the capacity is proportional to the number of transmit antennas. We can observe that as $R_H \to \infty$, the capacity tends to the limiting value (asymptotic capacity) $C = \frac{P}{\sigma^2} W \log_2 e$. Note that $\underset{\frac{R_H \sigma^2}{P}\to\infty}{Lim} \log_2\left(1+\frac{P}{R_H \sigma^2}\right)^{\frac{R_H \sigma^2}{P}} = \log_2 e$.

Review question 5.1 *Can we have MUX gain for SIMO/MISO channels?*

Review question 5.2 *What are the capacities of MIMO channel with unit and identity channel matrices?*

5.3 Capacity of random MIMO channel

MIMO channel are usually random. Hence we need to find the capacity of random MIMO channel rather than the deterministic MIMO channels. For fading channels, the channel is random and hence capacity which is a function of channel is also random. There are two types of capacity for fading channels:

(a) Ergodic capacity
(b) Outage capacity

For a frequency non-selective MIMO channels, there are two classifications which will be widely investigated for calculation of MIMO channel capacity. They are

(a) Ergodic channels
For frequency non-selective channels

$$\mathbf{y}(l) = \mathbf{H}(l)\mathbf{x}(l) + \mathbf{n}(l)$$

with H(l), $-\infty < l < \infty$ as an i.i.d. random process. This fading model is ergodic. During transmission, a long enough code word experiences all states of the channel.

(b) Non-ergodic channels
   Each code word however long experiences only one state of the channel.
   The average capacity is denoted by $\langle C \rangle$. For ergodic channel, the channel coefficients vary with time and it is possible to experience all states of the channel over the entire frame of data. Hence we can average or take expectation of the random channel capacity over the PDF of the channel. The channel capacity which is a function of random channel is also random and it will have CDF and PDF.

Another capacity of importance is the outage capacity which is the appropriate capacity to describe for non-ergodic channels. For non-ergodic channel, it is possible to experience only limited realizations of the channels for entire frame of data. The outage probability denoted as P_{out} is the probability that the channel capacity C drops below a certain threshold information rate R. For any given data rate R, there is a finite probability that for any coding scheme will not be supported reliably over the channel. In other words, reliable transmission rate R for the channel is possible with the probability $1 - P_{out}$. P_{out} is calculated from the CDF of the channel.

Review question 5.3 *What do you mean by ergodic and non-ergodic channels?*

Review question 5.4 *Define average capacity and outage probability.*

5.3.1 SISO fading channel

For a SISO channel, the I–O relationship can be expressed as

$$y = hx + n$$

where, y is the received signal, x is the transmitted signal and n is the AWGN.

We will consider two performance parameters for random channel. They are ergodic capacity and outage probability.

(a) Ergodic capacity

Ergodic capacity is the average of the instantaneous capacity of the random channel. It is found out by taking the expectation of the instantaneous capacity over the probability density function (PDF) of $\alpha = |h|^2$ where, h is the random channel gain coefficient as given below.

$$\langle C \rangle = E\left\{W \log_2 \left(1 + \alpha SNR\right)\right\} = \int_0^{\infty} W \log_2 \left(1 + \alpha SNR\right) p_{\alpha}(\alpha) d\alpha; \ \alpha = |h|^2 \tag{5.6a}$$

where W is the bandwidth.

The above expression for average channel capacity could be obtained analytically or numerically.

(b) Outage probability

In the case of random channels, it is better to characterize the system performance by computing the outage probability i.e., the probability that the rate R is greater than the channel capacity $C(h)$. In such a case, the outage probability

$$P_{out} = \Pr ob\left(C(h) < R\right) = \Pr ob\left(W \log_2 (1 + \alpha SNR) < R\right); \alpha = |h|^2 \tag{5.6b}$$

where, W is the bandwidth.

The $q(0 \le q \le 1)$ outage capacity is defined as the information rate (R) that guarantees no outage for $(1 - q)$ of the channel realizations. For instance, information rate R for which $q = 0.1$ or $q = 0.2$ outage capacities may be calculated.

Example 5.1

Show that if we have the PDF of $|h|$ as $p_{|h|}(|h|)$, we can obtain the PDF of effective SNR $\alpha = |h|^2 \frac{P}{\sigma^2}$ with $\alpha_0 = \Omega \frac{P}{\sigma^2}$ by making a change of variable for the fading PDF as follows:

$$p_\alpha(\alpha) = \frac{p_{|h|}\left(\sqrt{\frac{\Omega\alpha}{\alpha_0}}\right)}{2\sqrt{\frac{\alpha\alpha_0}{\Omega}}} \tag{5.7a}$$

Solution

This can be proved easily using the following procedure and theorem of probability.

Assume X is a continuous random variable (RV) with PDF $f_X(x)$. Let Y be a new RV obtained from X by the transformation Y=g(X). If we wish to determine the PDF of Y=g(X) ($f_Y(y)$) in terms of x. We can employ the following theorem on transformations for functions of one RV (A. Papoulis et al., 2002).

Theorem: To find $f_Y(y)$ for a specific y, we solve the equation y=g(x). Denoting its real roots by $x_n = g^{-1}(y_n)$, we can show that

$$f_Y(y) = \frac{f_X(x_1)}{g'(x_1)} + \ldots + \frac{f_X(x_n)}{g'(x_n)} + \ldots \tag{5.7b}$$

where, $g'(x)$ is the derivative of $g(x)$.

For equation (5.7a), $y = \alpha$ and $x = |h|$. From $\alpha = |h|^2 \frac{P}{\sigma^2}$ and $\alpha_0 = \Omega \frac{P}{\sigma^2}$, we have, $|h| = \sqrt{\frac{\sigma^2}{P}\alpha} = \sqrt{\frac{\Omega}{\alpha_0}\alpha}$. This is equivalent to $x = g^{-1}(y)$. Since $y = g(x)$ for this case is $\alpha = |h|^2 \frac{P}{\sigma^2}$.

Hence, $g'(x) = 2|h|\frac{P}{\sigma^2} = 2\sqrt{\frac{\sigma^2}{P}\alpha}\frac{P}{\sigma^2} = 2\sqrt{\frac{P}{\sigma^2}\alpha} = 2\sqrt{\frac{\alpha_0}{\Omega}\alpha}$. Now we have the final equation (5.7a) by substituting above values of $x = g^{-1}(y)$ and $g'(x)$ in the equation (5.7b).

The outage probability P_{out} is the probability that the channel capacity C drops below a certain rate R. This is the cumulative distribution function (CDF) of the random variable C and threshold is R. The outage probability for data rate R can be obtained as

$$\begin{aligned}
P_{\text{out}}(R) &= \Pr ob\left(C < R\right) \\
&= \Pr ob\left(W\log_2\left(1 + \alpha SNR\right) < R\right) \\
&= \Pr ob\left(\alpha < \frac{2^{\frac{R}{W}} - 1}{SNR}\right)
\end{aligned}$$

$$\Rightarrow \qquad P_{\text{out}}(R) = \Pr ob\left(\alpha < \frac{e^{\frac{R}{W\ln 2}} - 1}{SNR}\right) \tag{5.8}$$

The CDF of a random variable (RV) α is defined as

$$P_\alpha(x) = \int_0^x P_\alpha(\alpha)\,d\alpha$$

In outage setting, threshold x should be chosen as $x = \frac{e^{\frac{R}{W\ln 2}} - 1}{SNR}$.

Classical fading distributions

Rayleigh fading

Let us consider the widely used Rayleigh fading model, where h is a zero mean, circularly symmetric complex Gaussian (ZMCSCG) RV. Since h is ZMCSCG RV, $\alpha = |h|^2$ is exponential i.e., it has PDF $p_\alpha(\alpha) = \frac{1}{\alpha_0}\exp\left(-\frac{\alpha}{\alpha_0}\right)u(\alpha)$ where, α_0 is the mean value of α and $u(\alpha)$ is the unit step function.

The ergodic capacity is given by

$$\langle C\rangle = \int_0^\infty W\log_2\left(1 + \alpha(SNR)\right)\frac{1}{\alpha_0}\exp\left(-\frac{\alpha}{\alpha_0}\right)d\alpha \tag{5.9a}$$

The outage probability can be obtained from

$$P_{\text{out}}(R) = \Pr ob\left(W\log_2\left(1 + \alpha SNR\right) < R\right)$$

$$= \Pr ob\left(\alpha < \frac{2^{\frac{R}{W}} - 1}{SNR} \right)$$

$$= \Pr ob\left(\alpha < \frac{e^{\frac{R}{W \ln 2}} - 1}{SNR} \right)$$

Also we know that, CDF of exponential RV is given by

$$P_\alpha(x) = \int_{-\infty}^{x} P_\alpha(\alpha)\, d\alpha = 1 - e^{-\frac{x}{\alpha_0}}$$

Hence, $$P_{out}(R) = 1 - e^{\frac{e^{\frac{R}{W \ln 2}} - 1}{SNR\alpha_0}}$$

Example 5.2

Show that there cannot be any reliable transmission at any rate guaranteeing a zero outage probability regardless of the value of the bandwidth (BW) and transmit power for a Rayleigh fading channel.

Solution

We can express the rate R in terms of outage probability from the previous equation as follows.

$$1 - SNR\alpha_0 \ln\left(1 - P_{out}(R)\right) = e^{\frac{R}{W \ln 2}}$$

$$\Rightarrow \quad (W \ln 2) \ln\left(1 - SNR\alpha_0 \ln\left(1 - P_{out}(R)\right)\right) = R$$

$$\Rightarrow \quad \frac{R}{W} = \log_2\left(1 - SNR\alpha_0 \ln\left(1 - P_{out}(R)\right)\right) \tag{5.9b}$$

From the above equation it turns out that if we want to have a zero outage probability ($P_{out} = 0$), we obtain $R = 0$ (S. Barbarossa, 2003). Hence there cannot be any reliable transmission at any rate guaranteeing a zero outage probability regardless of the value of the bandwidth (BW) and transmit power.

For $SNR = 5dB$, $\alpha_0 = 0.7$ and 0.2 outage capacity can be calculated as 0.5791 bits/sec/Hz. Below we will try to find the outage capacity for various classical fading distributions.

Rice fading

If $|h|$ is Rice distributed, $\alpha = |h|^2$ is non-central Chi square distributed with two degrees of freedom and non-centrality parameter, υ^2. The CDF of non-central Chi square distributed with two degrees of freedom and non-centrality parameter υ^2 is given by

$$P_\alpha(x) = 1 - Q_1\left(\upsilon, \sqrt{x}\right)$$

where, Q_1 is the Marcum Q-function.

Hence,

$$P_{out}(R) = 1 - Q_1\left(\upsilon, \sqrt{\frac{e^{\frac{R}{W \ln 2}} - 1}{SNR}}\right) \tag{5.10}$$

Marcum Q-function:

In the above equation, $Q_1(a,b) = \int_b^\infty x e^{-\frac{a^2+x^2}{2}} I_0(ax)\,dx; I_0(x) = \sum_{k=0}^\infty \left(\frac{x^k}{2^k k!}\right)^2$

Note that a and b are nonnegative real numbers and I_0 is the zeroth order modified Bessel function of the first kind. Another alternate expression for $Q_1(a,b)$ is

$$Q_1(a,b) = e^{-\frac{a^2+b^2}{2}} \sum_{k=0}^\infty \left(\frac{a}{b}\right)^k I_1(ab)$$

where $I_\mu(x) = \sum_{s=0}^\infty \frac{1}{s!(s+\mu)!}\left(\frac{x}{2}\right)^{2s+\mu}$.

MATLAB command to calculate Marcum Q-function is Q = marcumq(a,b).

Since $Q_1(\upsilon,0) = 1$, we have again for zero outage probability for $R=0$.

Nakagami-m fading

If $|h|$ is Nakagami-m distributed, $\alpha = |h|^2$ is gamma distributed.

The CDF of gamma distribution is given by

$$P_\alpha(x) = \frac{\gamma\left(m, \frac{m}{\alpha_0} x\right)}{\Gamma(m)}$$

Hence the outage probability is given by

$$P_{out}(R) = \frac{\gamma\left(m, \frac{m}{\alpha_0} \frac{e^{\frac{R}{W \ln 2}} - 1}{SNR}\right)}{\Gamma(m)} \tag{5.11}$$

In the above equation, gamma function is given by

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

It is also frequently called as Euler definite integral.

The MATLAB command to calculate gamma function is y=gamma(x) where x must be real.

For positive integer m, it reduces to

$$\Gamma(m) = (m-1)!$$

The incomplete gamma function is the generalization of the gamma function by the variable limit integrals as follows

$$\gamma(x, a) = \int_0^{a} t^{x-1} e^{-t} dt, \mathrm{Re}(x) > 0$$

For positive integer m, it may be expressed as

$$\gamma(m, a) = (m-1)!\left(1 - e^{-a}\sum_{s=0}^{m-1} \frac{a^s}{s!}\right).$$

Generalized fading distributions

Let us try to find the outage capacity of various generalized fading distributions. For CDF of generalized fading distributions, one may refer to M. D. Yacoub (2007).

η-µ fading

The CDF of α can be obtained from the given formula.

$$P_\alpha(x) = 1 - Y_\mu\left(\frac{H}{h}, \sqrt{\frac{2h\mu x}{\alpha_0}}\right)$$

Yacoub's integral could be obtained as

$$Y_\mu(x, y) = \frac{\sqrt{\pi}\, 2^{\frac{3}{2}-\mu}\left(1-x^2\right)^{\mu}}{\Gamma(\mu)\, x^{\mu-\frac{1}{2}}} \int_y^{\infty} e^{-t^2} t^{2\mu} I_{\mu-\frac{1}{2}}\left(t^2 x\right) dt$$

where $-1 < x < 1$, $y \geq 0$

$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{\alpha+2k}}{k!\,\Gamma(\alpha+k+1)}, x \geq 0$ is the modified Bessel function of first kind and order α.

The MATLAB command to calculate this is I = besseli(nu,Z,1) where nu is equal to the variable α and it must be real. The argument Z can be complex.

Hence the outage probability is given by

$$P_{\text{out}}(R) = 1 - Y_{\mu}\left(\frac{H}{h}, \sqrt{\frac{2h\mu\frac{e^{\frac{R}{W\ln 2}}-1}{SNR}}{\alpha_0}}\right) \tag{5.12}$$

k-μ fading

From the CDF of α, we can obtain the outage probability as

$$P_{\text{out}}(R) = 1 - Q_{\mu}\left(\sqrt{2k\mu}, \sqrt{\frac{2(1+k)\mu\frac{e^{\frac{R}{W\ln 2}}-1}{SNR}}{\alpha_0}}\right) \tag{5.13}$$

The generalized Marcum Q function is calculated as

$$Q_{\mu}(a,b) = \int_b^{\infty} x\left(\frac{x}{a}\right)^{\mu-1} e^{-\frac{a^2+x^2}{2}} I_{\mu-1}(ax)\,dx = Q_1(a,b) + e^{-\frac{a^2+b^2}{2}} \sum_{k=1}^{\mu-1}\left(\frac{b}{a}\right)^k I_k(ab).$$

The MATLAB command to calculate generalized Marcum Q function is Q = marcumq(a,b,m) where m is the μ variable in the above expression for Marcum Q-function.

We can also explore for α-μ fading cases. It is left as an exercise for the readers.

5.3.2 SIMO fading channel

Consider a SIMO system with one transmit antenna and N_R receive antennas. Noise corrupts the transmitted signal at the receive antennas and it is distributed as $N_C^{N_R}(0, \sigma^2\mathbf{I})$. The transmit signal power constraint is P. The channel for the $1 \times N_R$ MIMO system is assumed to be frequency flat i.i.d. Rayleigh fading and CSIR is available. Let us compute the ergodic capacity and outage probability of the channel.

For a time slot m, the received signal (D. Tse et al., 2005) can be written as

$$\mathbf{y}(m) = \mathbf{h}(m)x(m) + \mathbf{n}(m)$$

where, $\boldsymbol{h}(m) \sim N_C^{N_R}(0, \mathbf{I})$ and $x(m) \sim N_C(0, P)$.

Let us compute the capacity of the SIMO channel. Dropping the time index m, we can rewrite the I–O relation of SIMO system as

$$\boldsymbol{y} = \boldsymbol{h}x + \boldsymbol{n}$$

where, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_R} \end{bmatrix}; \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_R} \end{bmatrix}; \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{N_R} \end{bmatrix}$

Assume that $E\left[|x|^2\right] = P$. Let us assume that at time slot m, the channel is fixed for finding the instantaneous capacity at time slot m. The covariance of the received signal vector can be calculated as

$$\mathbf{R_{yy}} = E[\mathbf{yy}^H] = E[(\mathbf{h}x + \mathbf{n})\,(\mathbf{h}x + \mathbf{n})^H] = \mathrm{P}\mathbf{hh}^H + \sigma^2\mathbf{I}_{N_R}$$

Note that we have assumed that x and $\boldsymbol{n}$ are independent. Then, mutual information

$$I(x;\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}\,|\,x) = h(\mathbf{y}) - h(\mathbf{n})$$

∵ $h(\mathbf{y}\,|\,\mathbf{x}) = h((\mathbf{h}\mathrm{x} + \mathbf{n})/x) = h(\mathbf{n}/x) = h(\mathbf{n})$ due to translation invariance of the entropy (h) and independence. Note that a brief discussion on Shannon information content (SIC) of an outcome, entropy and mutual information is given in Appendix C. Since jointly proper Gaussian random vectors maximize the differential entropy (refer to Appendix C).

For a given covariance matrix $\boldsymbol{R}$, the complex multivariate Gaussian distribution maximize entropy on $(-\infty, \infty)^N$.

Example 5.3

Find the entropy of complex multivariate Gaussian distribution.

Solution

The entropy of complex multivariate Gaussian distribution could be obtained from its pdf as follows. A zero mean multivariate complex Gaussian distribution has the following pdf.

$$\varphi(\mathbf{x}) = \frac{1}{\pi^N|\mathbf{R}|}\exp\left(-\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x}\right) = |\pi\mathbf{R}|^{-1}\exp\left(-\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x}\right)$$

We can find the entropy as follows.

$$\begin{aligned}
h_f(\mathbf{x}) &= E_f\left(-\log_2\left(\varphi(\mathbf{x})\right)\right) = -\left(\log_2 e\right)E_f\left(-\ln|\pi\mathbf{R}| - \mathbf{x}^H\mathbf{R}^{-1}\mathbf{x}\right) \\
&= \left(\log_2 e\right)\left(\ln|\pi\mathbf{R}| + E_f\left(\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x}\right)\right) = \left(\log_2 e\right)\left[\ln|\pi\mathbf{R}| + E_f\left(\sum_{i,j} x_i\left(\mathbf{R}^{-1}\right)_{ij}x_j\right)\right] \\
&= \left(\log_2 e\right)\left[\ln|\pi\mathbf{R}| + E_f\left(\sum_{i,j} x_i x_j\left(\mathbf{R}^{-1}\right)_{ij}\right)\right] = \left(\log_2 e\right)\left[\ln|\pi\mathbf{R}| + \left(\sum_{i,j} E_f\left(x_j x_i\right)\left(\mathbf{R}^{-1}\right)_{ij}\right)\right]
\end{aligned}$$

$$= (\log_2 e)\left[\ln|\pi\mathbf{R}| + \left(\sum_{i,j}(\mathbf{R})_{ji}\left(\mathbf{R}^{-1}\right)_{ij}\right)\right] = (\log_2 e)\left[\ln|\pi\mathbf{R}| + \left(\sum_{i,j}\left(\mathbf{R}\mathbf{R}^{-1}\right)_{jj}\right)\right]$$

$$= (\log_2 e)\left[\ln|\pi\mathbf{R}| + \left(\sum_{i,j}(\mathbf{I})_{jj}\right)\right] = (\log_2 e)\left[\left(\ln|\pi\mathbf{R}|\right) + N\right] = (\log_2 e)\left[\left(\ln|\pi e\mathbf{R}|\right)\right]$$

$$= \left[\left(\log_2|\pi e\mathbf{R}|\right)\right]$$

Hence, $h(\mathbf{y}) = \log_2\left|\pi e\mathbf{R}_{\mathbf{yy}}\right|; h(\mathbf{n}) = \log_2\left|\pi e\mathbf{R}_{\mathbf{nn}}\right| = \log_2\left|\pi e\sigma^2\mathbf{I}_{N_R}\right| = \log_2\left(\pi e\sigma^2\right)^{N_R}$.

We next use the upper bound on the mutual information by rewriting the capacity of the channel as

$$I(x,\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{n}) \le \log_2\left(\det\left(\pi e\mathbf{R}_{yy}\right)\right) - \log_2\left[\left(\pi e\sigma^2\right)^{N_R}\right]$$

$$= \log_2\left(\frac{\det\left(P\mathbf{h}\mathbf{h}^H + \sigma^2\mathbf{I}_{N_R}\right)}{\left(\sigma^2\right)^{N_R}}\right) = \log_2\det\left(\mathbf{I}_{N_R} + \frac{P}{\sigma^2}\mathbf{h}\mathbf{h}^H\right)$$

For any two matrices M×N matrix **A** and N×M matrix **B**, we have,

$$\det\left(\mathbf{I}_M + \mathbf{AB}\right) = \det\left(\mathbf{I}_N + \mathbf{BA}\right)$$

Hence, for SIMO system, using the above identity, we have,

$$I(x,\mathbf{y}) \le \log_2\det\left(1 + \frac{P}{\sigma^2}|\mathbf{h}|^2\right)$$

Therefore,

$$C = W\log_2\left(1 + \frac{P}{\sigma^2}|\mathbf{h}|^2\right) \tag{5.14}$$

Rayleigh fading channels

Ergodic capacity:

The ergodic capacity of this channel is given by

$$\langle C\rangle = E\left(\log_2\left(1 + \frac{P}{\sigma^2}\|\mathbf{h}\|^2\right)\right)$$

The RV $\|\mathbf{h}\|^2 = \mathbf{h}^H\mathbf{h} = \sum_{i=1}^{N_R}|h_i|^2$ is a sum of the square of $2N_R$ independent Gaussian RVs and hence it is Chi-square distributed with $2N_R$ degrees of freedom. Therefore, the PDF of this RV is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{1}{2^{N_R}(N_R-1)!}x^{N_R-1}e^{-\frac{x}{2}}$$

Therefore, the ergodic capacity of this channel is given by

$$\langle C\rangle = \frac{1}{2^{N_R}(N_R-1)!}\int_0^{\infty}\left(1+\frac{P}{\sigma^2}x\right)x^{N_R-1}e^{-\frac{x}{2}}dx \tag{5.15}$$

The close form expression for ergodic capacity of SIMO Rayleigh fading channel is given in example 6.1. We can compare this channel capacity with that of SISO case. We will find here the ergodic capacity for high SNR case. We can rewrite the ergodic capacity as follows:

$$\langle C\rangle = E\left[\log_2\left(1+\frac{N_R P}{N_R\sigma^2}\|\mathbf{h}\|^2\right)\right]$$

For high SNR case,

$$\langle C\rangle \approx E\left(\log_2\left(\frac{N_R P}{N_R\sigma^2}\|\mathbf{h}\|^2\right)\right)$$

$$= E\left(\log_2\left(N_R\frac{P}{\sigma^2}\right)\right) + E\left(\log_2\left(\frac{\|\mathbf{h}\|^2}{N_R}\right)\right)$$

$$\Rightarrow \qquad \langle C\rangle \approx \log_2\left(\frac{N_R P}{\sigma^2}\right) + E\left\{\log_2\left(\frac{\|h\|^2}{N_R}\right)\right\} \tag{5.16}$$

Note that in high SNR region, the ergodic capacity of the i.i.d. Rayleigh channel is equal to that of AWGN having an effective SNR of $\frac{N_R P}{\sigma^2}$ with an additional term which reduces capacity. The second tends to zero as $N_R \to \infty$ since the PDF of $\frac{\|\mathbf{h}\|^2}{N_R}$ approaches a Dirac delta function centred at 1.

Outage capacity:

For a threshold or target rate of R (bits/s/Hz), outage probability is given by

$$\Pr ob(R) = \Pr ob\left\{\log_2\left(1+\frac{P}{\sigma^2}\|\mathbf{h}\|^2\right) < R\right\} = \Pr ob\left\{\log_2\left(\|\mathbf{h}\|^2 < \frac{2^R-1}{\frac{P}{\sigma^2}}\right)\right\}$$

Hence the corresponding threshold on $\|\mathbf{h}\|^2$ is $\frac{2^R-1}{P/\sigma^2}$. The RV $\|\mathbf{h}\|^2$ is a sum of the square of $2N_R$ independent Gaussian RVs and hence it is Chi-square distributed with $2N_R$ degrees of freedom. Therefore, the PDF of this RV is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{1}{2^{N_R}(N_R - 1)!} x^{N_R - 1} e^{-\frac{x}{2}}$$

Example 5.4

Compute the outage probability and diversity gain of Rayleigh SIMO channel.

Solution

Let us compute the outage probability as follows.

$$P_{out}(R) = \int_0^{\frac{2^R - 1}{\frac{P}{\sigma^2}}} \frac{1}{2^{N_R}(N_R - 1)!} x^{N_R - 1} e^{-\frac{x}{2}} dx$$

Substituting $y = \frac{x}{2}$, we have,

$$P_{out}(R) = \int_0^{\frac{1}{2}\frac{2^R - 1}{\frac{P}{\sigma^2}}} \frac{1}{(N_R - 1)!} y^{N_R - 1} e^{-y} dy$$

For a high SNR $\left(\frac{P}{\sigma^2}\right)$, since $y < \frac{2^R - 1}{2\frac{P}{\sigma^2}}$ and $\frac{2^R - 1}{P/\sigma^2}$ tends to zero for high SNR, we have, $e^{-y} \approx 1$. Hence,

$$P_{out}(R) \approx \int_0^{\frac{1}{2}\frac{2^R - 1}{2\frac{P}{\sigma^2}}} \frac{1}{(N_R - 1)!} y^{N_R - 1} dx = \frac{1}{(N_R)!} y^{N_R} \Bigg]_0^{\frac{1}{2}\frac{2^R - 1}{\frac{P}{\sigma^2}}} = \frac{1}{2^{N_R}} \frac{\left(2^R - 1\right)^{N_R}}{\left(\frac{P}{\sigma^2}\right)^{N_R} (N_R)!} \tag{5.17}$$

One can see that outage probability for high SNR case is proportional to $\frac{1}{\left(\frac{P}{\sigma^2}\right)^{N_R}}$ where $\left(\frac{P}{\sigma^2}\right)$ is SNR. Hence there is diversity gain of N_R with respect to (w.r.t.) SISO case.

Let us try to find the exact outage capacity of a SIMO system with i.i.d. Rayleigh fading.

Example 5.5

Find the $P_{out}(R,P)$ for an i.i.d. Rayleigh fading channel with $N_T = 1$ (number of transmit antennas) and N_R receive antennas, where γ is the SNR and R is rate in bits/sec/Hz.

Solution

Consider $N_T = 1$ (only one antenna). In this case, it is clear that the outage probability is

$$\Pr ob\left(\log_2 \left| \mathbf{I}_{N_R} + \frac{\gamma}{N_T} \mathbf{h}\mathbf{h}^H \right| < R \right) = \Pr ob\left(\log_2 \left| 1 + \gamma \mathbf{h}^H \mathbf{h} \right| < R \right)$$

Since $\mathbf{h}^H\mathbf{h}$ is a Chi-square random variable with $2N_R$ degrees of freedom (N_R is the number of receive antennas) and mean N_R, we can compute the outage probability as

$$\begin{aligned} \Pr ob\left(\log_2 \left| 1 + \gamma \mathbf{h}^H \mathbf{h} \right| < R \right) &= \Pr ob\left(\left| 1 + \gamma \mathbf{h}^H \mathbf{h} \right| < 2^R \right) \\ &= \Pr ob\left(\gamma \mathbf{h}^H \mathbf{h} < 2^R - 1 \right) \\ &= \Pr ob\left(\mathbf{h}^H \mathbf{h} < \frac{2^R - 1}{\gamma} \right) \end{aligned}$$

$$\therefore \qquad P_{\text{out}}(R) = \frac{\gamma\left(N_R, \frac{2^R - 1}{\gamma} \right)}{\Gamma\left(N_R \right)}$$

where, $\gamma(x, a)$ is the incomplete gamma function.

η-µ fading channels

Outage capacity:
Outage probability can be obtained from

$$\Pr ob\left(R \right) = \Pr ob\left\{ \log_2 \left(1 + \frac{P}{\sigma^2} \|\mathbf{h}\|^2 \right) < R \right\}$$

It could be also obtained from the CDF from $\eta - N_R\mu$ square variate distribution.

$$P_x\left(y \right) = \int_0^y \frac{2\sqrt{\pi}\left(N_R\mu \right)^{N_R\mu+\frac{1}{2}} h^{N_R\mu} x^{N_R\mu-\frac{1}{2}} e^{-\frac{2N_R\mu h}{\Omega}x}}{\Gamma\left(N_R\mu \right) H^{N_R\mu-\frac{1}{2}} \Omega^{N_R\mu+\frac{1}{2}}} I_{N_R\mu-\frac{1}{2}}\left(\frac{2N_R\mu H}{\Omega} x \right) dx \qquad (5.18a)$$

where threshold y is $\dfrac{2^R - 1}{\frac{P}{\sigma^2}}$.

Ergodic capacity:
The ergodic capacity of this channel is given by

$$\langle C \rangle = E\left(\log_2 \left(1 + \frac{P}{\sigma^2} \|\mathbf{h}\|^2 \right) \right)$$

The RV $\|\mathbf{h}\|^2$ is a sum of the square of N_R independent $\eta-\mu$ RVs (M. D. Yacoub, 2007) and hence it is $\eta-N_R\mu$ square variate distributed. Therefore, the PDF of this RV (D. Dixit et al., 2012) is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{2\sqrt{\pi}\left(N_R\mu\right)^{N_R\mu+\frac{1}{2}} h^{N_R\mu} x^{N_R\mu-\frac{1}{2}} e^{-\frac{2N_R\mu h}{\Omega}x}}{\Gamma\left(N_R\mu\right) H^{N_R\mu-\frac{1}{2}} \Omega^{N_R\mu+\frac{1}{2}}} I_{N_R\mu-\frac{1}{2}}\left(\frac{2N_R\mu H}{\Omega}x\right); E(x) = \Omega$$

Therefore, the ergodic capacity of this channel is given by

$$\langle C\rangle = \int_0^\infty \log_2\left(1+\frac{P}{\sigma^2}x\right) \frac{2\sqrt{\pi}\left(N_R\mu\right)^{N_R\mu+\frac{1}{2}} h^{N_R\mu} x^{N_R\mu-\frac{1}{2}} e^{-\frac{2N_R\mu h}{\Omega}x}}{\Gamma\left(N_R\mu\right) H^{N_R\mu-\frac{1}{2}} \Omega^{N_R\mu+\frac{1}{2}}} I_{N_R\mu-\frac{1}{2}}\left(\frac{2N_R\mu H}{\Omega}x\right)dx \quad (5.18b)$$

k-μ fading channels

Outage capacity:

Outage probability can be obtained from

$$\Pr ob(R) = \Pr ob\left\{\log_2\left(1+\frac{P}{\sigma^2}\|\mathbf{h}\|^2\right) < R\right\}$$

It could be also obtained from the CDF from $k-N_R\mu$ square variate distribution.

$$P_x(y) = \int_0^y \frac{N_R\mu(1+k)^{\frac{N_R\mu+1}{2}} x^{\frac{N_R\mu}{2}} e^{-\frac{N_R\mu(1+k)}{\Omega}x}}{k^{\frac{N_R\mu-1}{2}} e^{N_R\mu k} \Omega^{\frac{N_R\mu+1}{2}}} I_{N_R\mu-1}\left(2N_R\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right)dx \quad (5.19a)$$

where threshold y is $\dfrac{2^R-1}{\frac{P}{\sigma^2}}$.

Ergodic capacity:

The ergodic capacity of this channel is given by

$$\langle C\rangle = E\left(\log_2\left(1+\frac{P}{\sigma^2}\|\mathbf{h}\|^2\right)\right)$$

The RV $\|\mathbf{h}\|^2$ is a sum of the square of N_R independent $k-\mu$ RVs and hence it is $k-N_R\mu$ square variate distributed. Therefore, the PDF of this RV is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{N_R\mu(1+k)^{\frac{N_R\mu+1}{2}} x^{\frac{N_R\mu}{2}} e^{-\frac{N_R\mu(1+k)}{\Omega}x}}{k^{\frac{N_R\mu-1}{2}} e^{N_R\mu k} \Omega^{\frac{N_R\mu+1}{2}}} I_{N_R\mu-1}\left(2N_R\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right); E(x) = \Omega$$

Therefore, the ergodic capacity of this channel is given by

$$\langle C \rangle = \int_0^{\infty} \log_2\left(1 + \frac{P}{\sigma^2}x\right) \frac{N_R \mu (1+k)^{\frac{N_R\mu+1}{2}} x^{\frac{N_R\mu}{2}} e^{-\frac{N_R\mu(1+k)}{\Omega}x}}{k^{\frac{N_R\mu-1}{2}} e^{N_R\mu k} \Omega^{\frac{N_R\mu+1}{2}}} I_{N_R\mu-1}\left(2N_R\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right) dx \quad (5.19b)$$

5.3.3 MISO fading channel

Consider now a MISO system with one receive antenna and N_T transmit antennas. Noise corrupts the transmitted signal at the receive antenna. The transmit signal power constraint is $\frac{P}{\sigma^2}$. The channel for the $N_T \times 1$ MIMO system is assumed to be frequency flat i.i.d. Rayleigh fading and CSIR is available. Let us compute the ergodic capacity and outage probability of the channel. For a time slot m, the received signal can be written as

$$y(m) = \mathbf{h}(m)\mathbf{x}(m) + n(m)$$

where, $\boldsymbol{h}(m) \sim N_C^{N_T}(0, \mathbf{I})$, $\boldsymbol{x}(m) \sim N_C^{N_T}\left(0, \frac{P}{N_T}\mathbf{I}\right)$, $N_C(0, \sigma^2)$ and $E\left(|x(m)|^2\right) \le P$.

Since channel is a vector ($\mathbf{h} = \begin{bmatrix} h_1 & h_2 & \cdots & h_{N_T} \end{bmatrix}$), its SVD will have a single singular value equals to the Frobenius norm of the vector. If we assume that channel is not known at the transmitter, we can have equal power allocation and therefore, each transmitting antenna will send signal with power of $\frac{P}{N_T}$. Hence the effective SNR for each path is $\frac{P}{N_T\sigma^2}$. The instantaneous capacity for uniform power allocation is given as

$$C_{uniform} = \log_2\left[1 + \frac{P}{N_T\sigma^2}\|\mathbf{h}\|^2\right]$$

Rayleigh fading channels

Ergodic capacity:
The ergodic capacity of this channel is given by

$$\langle C \rangle = E_{\mathbf{h}} \log_2\left[1 + \frac{P}{\sigma^2 N_T}\|\mathbf{h}\|^2\right]$$

The RV $\|\mathbf{h}\|^2 = \sum_{j=1}^{N_T} |h_j|^2$ is a sum of the square of $2N_T$ independent Gaussian RVs and hence it is Chi-square distributed with $2N_T$ degrees of freedom. Therefore, the PDF of this RV is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{1}{2^{N_T}(N_T-1)!} x^{N_T-1} e^{-\frac{x}{2}}$$

Therefore, the approximate ergodic capacity of this channel for high SNR case is given by

$$\langle C \rangle \approx \log_2\left(\frac{P}{\sigma^2}\right) + E\left\{\log_2\left(\frac{\|\mathbf{h}\|^2}{N_T}\right)\right\} \tag{5.20a}$$

The derivation is quite similar to that of SIMO case and hence will not be discussed again. One point to be noticed is that there is no power or array gain in the first term of the ergodic capacity of equation (5.20a) w.r.t. SISO case. The close form expression for ergodic capacity of MISO Rayleigh fading channel is also given in example 6.1.

Outage capacity:

Let's now move to the outage setting. The outage probability is given by

$$P_{out}(R) = \Pr ob\left\{\log_2\left[1 + \frac{P}{\sigma^2 N_T}\|\mathbf{h}\|^2\right] < R\right\}$$

$$= \Pr ob\left\{\|\mathbf{h}\|^2 < \frac{N_T\left(2^R - 1\right)}{\frac{P}{\sigma^2}}\right\}$$

$$\Rightarrow \qquad P_{out}(R) = \Pr ob\left\{\|\mathbf{h}\|^2 < \frac{2^R - 1}{\frac{P}{\sigma^2 N_T}}\right\}$$

Hence, at high SNR (similar analysis with SIMO case above), we get,

$$\Rightarrow \qquad P_{out}(R) \approx \frac{\left(2^R - 1\right)^{N_T}}{2^{N_T}\left(N_T\right)!\left(\frac{P}{\sigma^2 N_T}\right)^{N_T}} \tag{5.20b}$$

One can see that outage probability for high SNR case is proportional to $\frac{1}{\left(\frac{P}{\sigma^2}\right)^{N_T}}$ where $\left(\frac{P}{\sigma^2}\right)$ is SNR. Hence there is diversity gain of N_T w.r.t. SISO case.

η-μ fading channels

Ergodic capacity:

The ergodic capacity of this channel is given by

$$\langle C \rangle = E_{\mathbf{h}} \log_2\left[1 + \frac{P}{\sigma^2 N_T}\|\mathbf{h}\|^2\right]$$

The RV $\|\mathbf{h}\|^2$ is a sum of the square of N_T independent $\eta-\mu$ RVs (M. D. Yacoub, 2007) and hence it is $\eta-N_T\mu$ square variate distributed. Therefore, the PDF of this RV (D. Dixit et al., 2012) is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{2\sqrt{\pi}(N_T\mu)^{N_T\mu+\frac{1}{2}} h^{N_T\mu} x^{N_T\mu-\frac{1}{2}} e^{-\frac{2N_T\mu h}{\Omega}x}}{\Gamma(N_T\mu) H^{N_T\mu-\frac{1}{2}} \Omega^{N_T\mu+\frac{1}{2}}} I_{N_T\mu-\frac{1}{2}}\left(\frac{2N_T\mu H}{\Omega}x\right); E(x) = \Omega$$

Therefore, the ergodic capacity of this channel is given by

$$\langle C\rangle = \int_0^\infty \log_2\left(1+\frac{P}{N_T\sigma^2}x\right)\frac{2\sqrt{\pi}(N_T\mu)^{N_T\mu+\frac{1}{2}} h^{N_T\mu} x^{N_T\mu-\frac{1}{2}} e^{-\frac{2N_T\mu h}{\Omega}x}}{\Gamma(N_T\mu) H^{N_T\mu-\frac{1}{2}} \Omega^{N_T\mu+\frac{1}{2}}} I_{N_T\mu-\frac{1}{2}}\left(\frac{2N_T\mu H}{\Omega}x\right)dx \quad (5.21a)$$

Outage capacity:

Outage probability can be obtained from

$$\Pr ob(R) = \Pr ob\left\{\log_2\left(1+\frac{P}{N_T\sigma^2}\|\mathbf{h}\|^2\right) < R\right\}$$

It could be also obtained from the CDF from $\eta - N_T\mu$ square variate distribution.

$$P_x(y) = \int_0^y \frac{2\sqrt{\pi}(N_T\mu)^{N_T\mu+\frac{1}{2}} h^{N_T\mu} x^{N_T\mu-\frac{1}{2}} e^{-\frac{2N_T\mu h}{\Omega}x}}{\Gamma(N_T\mu) H^{N_T\mu-\frac{1}{2}} \Omega^{N_T\mu+\frac{1}{2}}} I_{N_T\mu-\frac{1}{2}}\left(\frac{2N_T\mu H}{\Omega}x\right)dx \quad (5.21b)$$

where threshold y is $\dfrac{2^R-1}{\frac{P}{\sigma^2 N_T}}$.

k-μ fading channels

Ergodic capacity:

The ergodic capacity of this channel is given by

$$\langle C\rangle = E\left(\log_2\left(1+\frac{P}{N_T\sigma^2}\|\mathbf{h}\|^2\right)\right)$$

The RV $\|\mathbf{h}\|^2$ is a sum of the square of N_T independent $k-\mu$ RVs and hence it is $k-N_T\mu$ square variate distributed. Therefore, the PDF of this RV is

$$f_{\|\mathbf{h}\|^2}(x) = \frac{N_T\mu(1+k)^{\frac{N_T\mu+1}{2}} x^{\frac{N_T\mu}{2}} e^{-\frac{N_T\mu(1+k)}{\Omega}x}}{k^{\frac{N_T\mu-1}{2}} e^{N_T\mu k} \Omega^{\frac{N_T\mu+1}{2}}} I_{N_T\mu-1}\left(2N_T\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right); E(x) = \Omega$$

Therefore, the ergodic capacity of this channel is given by

$$\langle C \rangle = \int_0^{\infty} \log_2\left(1 + \frac{P}{N_T\sigma^2}x\right) \frac{N_T\mu(1+k)^{\frac{N_T\mu+1}{2}} x^{\frac{N_T\mu}{2}} e^{-\frac{N_T\mu(1+k)}{\Omega}x}}{k^{\frac{N_T\mu-1}{2}} e^{N_T\mu k} \Omega^{\frac{N_T\mu+1}{2}}} I_{N_T\mu-1}\left(2N_T\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right)dx \tag{5.22a}$$

Outage capacity:
Outage probability can be obtained from

$$\Pr ob(R) = \Pr ob\left\{\log_2\left(1 + \frac{P}{N_T\sigma^2}\|\mathbf{h}\|^2\right) < R\right\}$$

It could be also obtained from the CDF from $k - N_T\mu$ square variate distribution.

$$P_x(y) = \int_0^{y} \frac{N_T\mu(1+k)^{\frac{N_T\mu+1}{2}} x^{\frac{N_T\mu}{2}} e^{-\frac{N_T\mu(1+k)}{\Omega}x}}{k^{\frac{N_T\mu-1}{2}} e^{N_T\mu k} \Omega^{\frac{N_T\mu+1}{2}}} I_{N_T\mu-1}\left(2N_T\mu\sqrt{\frac{k(1+k)x}{\Omega}}\right)dx \tag{5.22b}$$

where threshold y is $\dfrac{2^R - 1}{\dfrac{P}{\sigma^2 N_T}}$.

Review question 5.5 *For SISO channel,*

(a) If $|h|$ is Rayleigh distributed, then $\alpha = |h|^2$ is .... distributed.

(b) If $|h|$ is Rice distributed, then $\alpha = |h|^2$ is .... distributed.

(c) If $|h|$ is Nakagami-m distributed, then $\alpha = |h|^2$ is .... distributed.

Review question 5.6 *What is the close-form expression for Yacoub's integral?*

Review question 5.7 *What is the instantaneous capacity of (a) SIMO and (b) MISO i.i.d. Rayleigh fading channel?*

Review question 5.8 *What is the outage and ergodic capacity of (a) SIMO and (b) MISO i.i.d. Rayleigh fading channel?*

Review question 5.9 *What is the diversity order of a (a) SIMO and (b) MISO i.i.d. Rayleigh fading channel?*

Review question 5.10 *What is the outage and ergodic capacity of (a) SIMO and (b) MISO for i.i.d. η-μ and k-μ fading channels?*

5.4 Summary

Figure 5.1 shows the chapter in a nutshell. We start with the determination of MIMO channel capacity for fixed channels like SISO, SIMO, MISO, unity and identity channel matrices. Then move onto the random channels. We found out the ergodic capacity and outage probability for SISO, i.i.d. SIMO and MISO channels. About the fading distributions, we have considered both classical and generalized fading distributions.

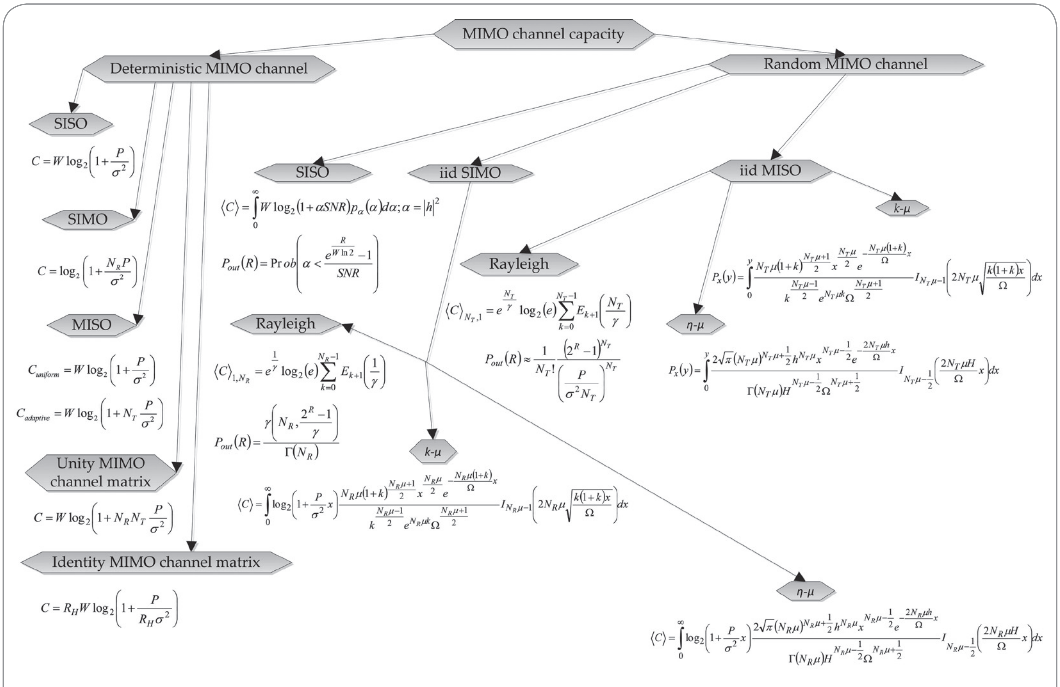


Fig. 5.1 Chapter 5 in a nutshell

Exercises

The following exercises may not have close-form formulae, but may be explored.

Exercise 5.1

Find the ergodic and outage capacity for i.i.d. uncorrelated classical fading SIMO channel whose fading distribution is

(a) Nakagami-m
(b) Hoyt
(c) Weibull

Exercise 5.2

Find the ergodic and outage capacity for i.i.d. uncorrelated classical fading MISO channel whose fading distribution is
(a) Nakagami-m
(b) Hoyt
(c) Weibull

Exercise 5.3

Find the ergodic and outage capacity for i.i.d. uncorrelated generalized fading SIMO channel whose fading distribution is α-μ distributed.

Exercise 5.4

Find the ergodic and outage capacity for i.i.d. uncorrelated generalized fading MISO channel whose fading distribution is α-μ distributed.

Exercise 5.5

How would one find the ergodic and outage capacity for separately correlated Rayleigh MISO/SIMO channel?

References

1. Barbarossa, S. 2003. *Multiantenna Wireless Communication Systems*. Norwood: Artech House.
2. Cover, T. M. and J. A. Thomas. 2006. *Elements of Information Theory*. Indianapolis: Wiley.
3. Dixit, D. and P. R. Sahu. Aug. 2012. 'Performance of L-Branch MRC receiver in η-μ *and* k-μ fading channels for QAM signals'. *IEEE Wireless Communications Letters*. 1(4). 316–319.
4. Tse, D. and P. Viswanath. 2005. *Fundamentals of Wireless Communications*. Cambridge: Cambridge University Press.
5. Yacoub, M. D. Feb. 2007. 'The k-μ distribution and the η-μ distribution'. *IEEE Antennas Propagat. Mag.* 49(1). 68–81.

MIMO Channel Capacity

CHAPTER 6

6.1 Introduction

Compared to conventional single antenna system, the channel capacity of a MIMO system with N_T transmit and N_R receive antennas can be increased by a factor that is at the most the value of min (N_T, N_R), without additional transmit power or spectral bandwidth for an i.i.d. Rayleigh fading channel. MIMO techniques can be broadly divided into two: (a) diversity techniques and (b) spatial multiplexing (MUX). The diversity technique intends to receive the same information-bearing signals in the multiple antennas or transmit them from the multiple antennas, and hence improves the transmission reliability. When the spatial MUX techniques are used, the maximum transmission rate can be same as that of the MIMO channel capacity. But when diversity techniques are used, the achievable transmission speed or rate can much lower than the capacity of the channel. We will find the capacity of random MIMO channel for three cases: (a) i.i.d. Rayleigh fading channels, (b) Separately correlated Rayleigh fading MIMO channels, and (c) Keyhole Rayleigh fading MIMO channels. We will find the ergodic capacity and outage probability for i.i.d. fading MIMO channels. Then we will see the effect of antenna correlation on the MIMO channel capacity. How antenna correlation reduces the capacity of the channel? Finally we will find the capacity for a keyhole MIMO channel. We will show that for a highly scattered environment, the capacity is very low, if we have keyhole propagation.

6.2 Capacity of i.i.d. Rayleigh fading MIMO channels

In capacity analysis for MIMO channels, we will find the ergodic capacity and outage capacity from the instantaneous capacity. Similar analysis could be carried out for i.i.d. uncorrelated fading for any of the classical fading distributions viz., Rice, Nakagami-m, Hoyt and Weibull.

6.2.1 Ergodic capacity

By using the SVD, the MIMO fading channel with the channel matrix **H** can be represented by an equivalent channel consisting of R_H (rank of **H**) decoupled parallel Gaussian sub-channels. Thus the capacities of sub-channels add up, giving for the overall instantaneous capacity for uniform or equal power allocation as $C = W\sum_{i=1}^{R_H}\log_2\left(1+\frac{\lambda_i P}{N_T\sigma^2}\right)$. For ergodic case, we need to average the instantaneous capacity over the PDF of the channel matrix **H** and hence the average or ergodic capacity for uniform power allocation is given by

$$\langle C\rangle = E\left\{W\sum_{i=1}^{R_H}\log_2\left(1+\frac{\lambda_i P}{N_T\sigma^2}\right)\right\}$$

where, $\sqrt{\lambda_i}$ are the singular values of the channel matrix.

Alternatively, we could also write the mean MIMO capacity for ergodic fading channels as

$$\langle C\rangle = E\left\{W\log_2\det\left(\mathbf{I}_{N_R}+\frac{P\mathbf{Q}}{N_T\sigma^2}\right)\right\}$$

where, **Q** is the Wishart matrix defined as $\mathbf{Q} = \begin{cases}\mathbf{H}\mathbf{H}^H, & N_R < N_T\\ \mathbf{H}^H\mathbf{H}, & N_R \geq N_T\end{cases}$

The above expectation is taken over the statistics of the random matrix **H**. The exact computation of ergodic capacity was carried out by E. Teletar (1999). The capacity of the channel with N_T transmit antennas and N_R receive antennas under power constraint P equals

$$\langle C\rangle = \int_0^\infty \log_2\left(1+\frac{Px}{N_T}\right)\sum_{k=0}^{m-1}\frac{k!}{(k+n-m)}\left[L_k^{n-m}(x)\right]^2 x^{n-m}e^{-x}dx \tag{6.1a}$$

where, $m = \min\{N_T, N_R\}$ and $n = \max\{N_T, N_R\}$.

L_k^{n-m} are the associated Laguerre polynomials of order k.

$$L_k^{n-m}(x) = \frac{1}{k!}e^x x^{m-n}\frac{d^k}{dx^k}\left(e^{-x}x^{n-m+k}\right)$$

See example 6.2 as well for an alternate close-form expression. We will derive a close-form expression (H. Shin et al., 2003) here as given below. Assume W=1 for brevity of the analysis.

$$\langle C\rangle = E\left\{W\log_2\det\left(\mathbf{I}_{N_R}+\frac{P\mathbf{Q}}{N_T\sigma^2}\right)\right\}$$

$$\langle C\rangle = (\log_2 e)E\left[\sum_{k=1}^{m}\left(1+\frac{\gamma}{N_T}\lambda_k\right)\right] = m(\log_2 e)E\left\{\ln\left(1+\frac{\gamma}{N_T}\lambda\right)\right\}$$

It is equivalent to m times (m is the rank of the full rank matrix **H**) finding the expectation of an arbitrary and unordered eigenvalue λ. From section 3.2.7, we have the marginal PDF of an unordered λ is given by

$$p(\lambda)=\frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{i}\sum_{l=0}^{2j}\frac{(-1)^l(2j)!}{2^{2i-l}j!l!(n-m+j)!}\binom{2i-2j}{i-j}\binom{2j+2n-2m}{2j-l}(\lambda)^{l+n-m}e^{-\lambda}$$

Hence,
$$\langle C\rangle = m(\log_2 e)\int_0^\infty \ln\left(1+\frac{\gamma}{N_T}\lambda\right)p(\lambda)\,d\lambda$$

$$=(\log_2 e)\sum_{i=0}^{m-1}\sum_{j=0}^{i}\sum_{l=0}^{2j}\frac{(-1)^l(2j)!}{2^{2i-l}j!l!(n-m+j)!}\binom{2i-2j}{i-j}\binom{2j+2n-2m}{2j-l}\int_0^\infty \ln\left(1+\frac{\gamma}{N_T}\lambda\right)(\lambda)^{l+n-m}e^{-\lambda}d\lambda$$

In order to calculate

$$I=\int_0^\infty \ln\left(1+\frac{\gamma}{N_T}\lambda\right)(\lambda)^{l+n-m}e^{-\lambda}d\lambda$$

$$\because\qquad I_q(\nu)=\int_0^\infty \ln(1+x)\,x^{q-1}e^{-\upsilon x}dx=(q-1)!e^{\upsilon}\sum_{k=1}^{q}\frac{\Gamma(-q+k,\upsilon)}{\upsilon^k}$$

The complementary incomplete gamma function is given by

$$\Gamma(-q+k,\upsilon)=\int_\nu^\infty e^{-x}x^{-q+k-1}dx;-q+k>0$$

If we assume that $x=\frac{\gamma}{N_T}\lambda; q-1=l+n-m; \upsilon=\frac{N_T}{\gamma}$, we have,

$$\therefore\qquad I=(\upsilon)^{q-2}\int_0^\infty \ln(1+x)(x)^{q-1}e^{-\upsilon x}dx$$

$$=(q-1)!e^{\upsilon}(\upsilon)^{q-2}\sum_{k=1}^{q}\frac{\Gamma(-q+k,\upsilon)}{\upsilon^k}$$

$$=(q-1)!e^{\upsilon}\sum_{k=1}^{q}\upsilon^{-k+q-2}\Gamma(-q+k,\upsilon)$$

$$=(q-1)!e^{\upsilon}\sum_{k=0}^{q-1}\upsilon^{-k+q-1}\Gamma(-q+k+1,\upsilon)$$

The exponential integral function of order υ could be expressed as

$$E_r(\upsilon)=\int_1^\infty e^{-\upsilon y}y^{-r}dy;\ \upsilon>0, r=0,1,\ldots$$

The exponential integral function is a particular case of the complementary incomplete gamma function.

$$\Gamma(1-r,\upsilon)=\int_\upsilon^\infty e^{-x}x^{-r}dx;1-r>0$$

Substituting $x = y\upsilon$, $dx = dy\upsilon$, we have,

$$\Gamma(1-r,\upsilon) = \int_1^\infty e^{-\nu y}(\nu y)^{-r}\,\nu dy = \nu^{-r+1}\int_1^\infty e^{-\nu y}y^{-r}dy$$

Therefore,

$$E_r(\upsilon) = \upsilon^{r-1}\Gamma(1-r,\upsilon)$$

If we assume that $r - 1 = q - k - 1$, then $1 - r = -q + k + 1$. Hence,

$$\therefore \qquad I = (q-1)!\,e^{\upsilon}\sum_{k=0}^{q-1}E_r(\upsilon)$$

Putting back, $r = q - k$; $q - 1 = l + n - m$; $\upsilon = \dfrac{N_T}{\gamma}$, we have,

$$I = (l+n-m)!\,e^{\frac{N_T}{\gamma}}\sum_{k=0}^{l+n-m}E_{l+n-m+1-k}\left(\frac{N_T}{\gamma}\right)$$

Note that $k = 0$, $l + n - m + 1 - k = l + n - m + 1$ and $k = l + n - m$, $l + n - m + 1 - k = 1$. Hence it is similar to k going from 1 to $l + n - m + 1$. Hence, it can be further expressed as

$$I = (l+n-m)!\,e^{\frac{N_T}{\gamma}}\sum_{k=0}^{l+n-m}E_{k+1}\left(\frac{N_T}{\gamma}\right)$$

Therefore,

$$\langle C\rangle = (\log_2 e)\,e^{\frac{N_T}{\gamma}}\sum_{i=0}^{m-1}\sum_{j=0}^{i}\sum_{l=0}^{2j}\frac{(-1)^l(2j)!(l+n-m)}{2^{2i-l}\,j!\,l!\,(n-m+j)!}\binom{2i-2j}{i-j}\binom{2j+2n-2m}{2j-l}\sum_{k=0}^{l+n-m}E_{k+1}\left(\frac{N_T}{\gamma}\right) \qquad (6.1b)$$

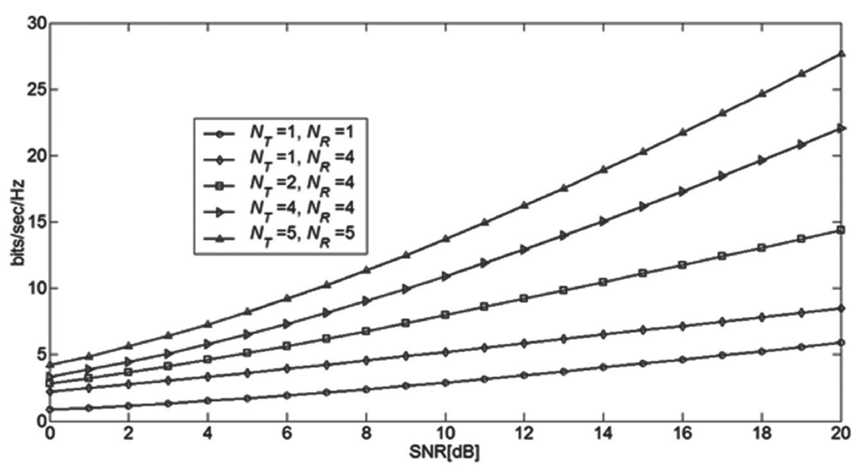


Fig. 6.1 Ergodic capacity vs *SNR (dB)* of open loop $N_T \times N_R$ MIMO system

Example 6.1

The ergodic capacity for i.i.d. Rayleigh fading $N_T \times N_R$ MIMO channel (H. Shin et al., 2003) can be calculated as

$$\langle C \rangle_{N_T, N_R} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{i=0}^{m-1} \sum_{j=0}^{i} \sum_{l=0}^{2j} \left\{ \frac{(-1)^l (2j)!(n-m+l)!}{2^{2i-l} j! l! (n-m+j)!} \binom{2i-2j}{i-j} \binom{2j+2n-2m}{2j-l} \right\} \sum_{k=0}^{n-m+l} E_{k+1}\left(\frac{N_T}{\gamma}\right)$$

where, $E_k(x) = \int_1^{\infty} e^{-xy} y^{-k} dy$ (exponential integral function of order k), $m = \min\{N_T, N_R\}$ and $n = \max\{N_T, N_R\}$, find the ergodic capacity of

(a) MIMO channel with $N_R = N_T = N$ antennas at the transmitter and receiver.
(b) MISO channel with N_T antennas at the transmitter and 1 antenna at the receiver.
(c) SIMO channel with one antenna at the transmitter and N_R antennas at the receiver.

Solution:

(a) Given that $N_R = N_T = N$ and hence, $n = \max\{N_T, N_R\} = N$ and $m = \min\{N_T, N_R\} = N$.

$$\langle C \rangle_{N,N} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{i=0}^{N-1} \sum_{j=0}^{i} \sum_{l=0}^{2j} \left\{ \frac{(-1)^l (2j)!(l)!}{2^{2i-l} j! l! (j)!} \binom{2i-2j}{i-j} \binom{2j}{2j-l} \right\} \sum_{k=0}^{l} E_{k+1}\left(\frac{N_T}{\gamma}\right)$$

$$\Rightarrow \quad \langle C \rangle_{N,N} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{i=0}^{N-1} \sum_{j=0}^{i} \sum_{l=0}^{2j} \left\{ \frac{(-1)^l (2j)!(l)!}{2^{2i-l} j! l! (j)!} \binom{2i-2j}{i-j} \frac{(2j)!}{(2j-l)!l!} \sum_{k=0}^{l} E_{k+1}\left(\frac{N_T}{\gamma}\right) \right\}$$

$$\Rightarrow \quad \langle C \rangle_{N,N} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{i=0}^{N-1} \sum_{j=0}^{i} \sum_{l=0}^{2j} \left\{ \frac{(-1)^l (2j)!}{2^{2i-l} j! (j)!} \binom{2i-2j}{i-j} \binom{2j}{l} \sum_{k=0}^{l} E_{k+1}\left(\frac{N_T}{\gamma}\right) \right\}$$

$$\Rightarrow \quad \langle C \rangle_{N,N} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{i=0}^{N-1} \sum_{j=0}^{i} \sum_{l=0}^{2j} \left\{ \frac{(-1)^l}{2^{2i-l}} \binom{2j}{j} \binom{2i-2j}{i-j} \binom{2j}{l} \sum_{k=0}^{l} E_{k+1}\left(\frac{N_T}{\gamma}\right) \right\}$$

(b) Given that $N_R = 1$ and hence, $n = \max\{N_T, 1\} = N_T$ and $m = \min\{N_T, 1\} = 1$.
Therefore, i = j = l = 0, we have,

$$\langle C \rangle_{N_T,1} = e^{\frac{N_T}{\gamma}} \log_2(e) \sum_{k=0}^{N_T-1} E_{k+1}\left(\frac{N_T}{\gamma}\right)$$

(c) Given that $N_T = 1$ and hence, $n = \max\{N_R, 1\} = N_R$ and $m = \min\{N_R, 1\} = 1$.
Therefore, i = j = l = 0, we have,

$$\langle C\rangle_{1,N_R} = e^{\frac{1}{\gamma}} \log_2(e) \sum_{k=0}^{N_R-1} E_{k+1}\left(\frac{1}{\gamma}\right)$$

Example 6.2

Show that a simple upper bound on the capacity of ergodic Rayleigh fading MIMO channel is given as

$$\langle C\rangle \le \min\left\{N_R \log_2(1+\gamma), N_T \log_2\left(1+\frac{N_R\gamma}{N_T}\right)\right\}.$$

Solution

The exact computation of capacity is rather tedious as we have seen. We can instead calculate a simple upper bound of capacity. Note that the log–det function is concave over the set of nonnegative matrices. Therefore, applying Jensen's inequality (refer to Appendix C), we have

$$\langle C\rangle = E\left[\log_2\left|\mathbf{I}_{N_R} + \frac{\gamma}{N_T}\mathbf{H}\mathbf{H}^H\right|\right] \le \log_2\left|\mathbf{I}_{N_R} + \frac{\gamma}{N_T}E\left(\mathbf{H}\mathbf{H}^H\right)\right| = N_R \log_2(1+\gamma)$$

In the above we have used the relation $E\left(\mathbf{H}\mathbf{H}^H\right) = N_T\mathbf{I}_{N_R}$. Note that for i.i.d. MIMO channel,

$$\mathrm{E}(\mathbf{H}\mathbf{H}^H) = E\left(\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \ddots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T}\end{bmatrix}\begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N_R1} \\ h_{12} & h_{22} & \cdots & h_{N_R2} \\ \vdots & \ddots & \ddots & \vdots \\ h_{1N_T} & h_{2N_T} & \cdots & h_{N_RN_T}\end{bmatrix}^*\right)$$

$$= \begin{bmatrix} E\left(|h_{11}|^2+|h_{12}|^2+\cdots+|h_{1N_T}|^2\right) & 0 & \cdots & 0 \\ 0 & E\left(|h_{21}|^2+|h_{22}|^2+\cdots+|h_{2N_T}|^2\right) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & E\left(|h_{N_R1}|^2+|h_{N_R2}|^2+\cdots+|h_{N_RN_T}|^2\right)\end{bmatrix}$$

In retrospect, the matrices $\mathbf{H}\mathbf{H}^H$ and $\mathbf{H}^H\mathbf{H}$ have identical nonzero eigenvalues; therefore

$$\langle C\rangle = E\left\{\log_2\left|\mathbf{I}_{N_T} + \frac{\gamma}{N_T}\mathbf{H}^H\mathbf{H}\right|\right\} \le \log_2\left|\mathbf{I}_{N_T} + \frac{\gamma}{N_T}E\left(\mathbf{H}^H\mathbf{H}\right)\right| = N_T \log_2\left(1+\frac{N_R\gamma}{N_T}\right)$$

In the above we have used the relation $E\left(\mathbf{H}^H\mathbf{H}\right) = N_R\mathbf{I}_{N_T}$. By combining the above two cases, we can obtain the upper bound as

$$\langle C\rangle \le \min\left\{N_R \log_2(1+\gamma), N_T \log_2\left(1+\frac{N_R\gamma}{N_T}\right)\right\}.$$

6.2.2 Outage capacity

Outage probability is the probability that the transmission rate R beats the capacity of the channel (E. Biglieri et al. 2004 and Y. W. Liang, 2005). The mutual information (instantaneous capacity) is a RV given by

$$C = W \log_2 \left\{ \det \left(\mathrm{I}_{N_R} + \frac{P\mathbf{Q}}{N_T \sigma^2} \right) \right\}$$

Outage probability can be obtained from

$$\Pr ob\,(R) = \Pr ob \left\{ W \log_2 \det \left(\mathbf{I}_{N_R} + \frac{P}{N_T \sigma^2} \mathbf{HH}^H \right) < R \right\}$$

We may adopt the asymptotic results which implies that as N_T and N_R tend to infinitude, the instantaneous capacity C leads to a Gaussian RV. It has been observed by many researchers that this result is appropriate for Gaussian RV even for a diminutive number of transmitting and receiving antennas. Using the Replica method (A. M. Tulino et al., 2004) from statistical mechanics, we can prove that C is asymptotically Gaussian RV. Hence by calculating the mean and variance, we can describe its asymptotic behavior. Therefore the outage probability may be nearly approximated for all combination of N_T and N_R antennas as

$$P_{\text{out}}(R) \approx Q\left(\frac{\mu_C - R}{\sigma_C} \right) \tag{6.1c}$$

where

$$\mu_C = -N_T \left\{ (1+\beta)\log(w) + q_0 r_0 \log e + \log r_0 + \beta \log\left(\frac{q_0}{\beta} \right) \right\} (\text{bit/dimension pair})$$

$$\sigma_C^2 = -\log e \bullet \log\left(1 - \frac{q_0^2 r_0^2}{\beta} \right) (\text{bit/dimension pair})^2$$

$$q_0 = \frac{\beta - 1 - w^2 + \sqrt{\left(\beta - 1 - w^2\right)^2 + 4w^2\beta}}{2w}$$

$$r_0 = \frac{1 - \beta - w^2 + \sqrt{\left(1 - \beta - w^2\right)^2 + 4w^2}}{2w}$$

where, $\beta = \dfrac{N_R}{N_T}$ is the ratio of the number of receiving antennas and transmitting antennas and $w = \sqrt{\dfrac{\sigma^2}{P}}$ is the square root of the ratio of noise variance and signal power.

Now we can calculate R as

$$R = \mu_C - \sigma_C Q^{-1}(P_{\text{out}}(R))$$

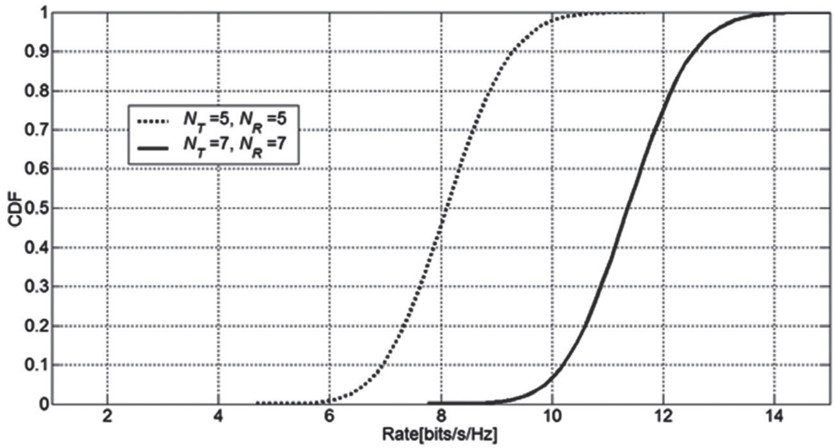


Fig. 6.2 CDF of open loop $N_T \times N_R$ MIMO channel capacity for $SNR = 5dB$

It can be read from the graph generated from Monte Carlo simulations in Fig. 6.2 that for a 5×5 MIMO channel, the 0.2 outage capacity is approximately 7.5 bits/sec/Hz for SNR of 5 dB; whereas, for a 7 ×7 MIMO channel, the 0.2 outage capacity is approximately 10.5 bits/sec/Hz for SNR of 5 dB.

6.3 Capacity of separately correlated Rayleigh fading MIMO channel

Let us calculate the instantaneous capacity of the MIMO channel when the gains between the transmitted and received antennas are correlated (Kronecker model). The channel gain matrix for the separately correlated fading case is

$$\mathbf{H} = \mathbf{R}_{R_X}^{1/2}\mathbf{H}_w\mathbf{R}_{T_X}^{1/2}$$

Hence the capacity is given by

$$C = W\log_2\left\{\det\left(\mathbf{I}_{N_R} + \frac{P\mathbf{Q}}{N_T\sigma^2}\right)\right\} = W\log_2\left\{\det\left(\mathbf{I}_{N_R} + \frac{P\mathbf{H}\mathbf{H}^H}{N_T\sigma^2}\right)\right\}$$

$$\Rightarrow \qquad \frac{C}{W} = \log_2\left\{\det\left(\mathbf{I}_{N_R} + \frac{P}{N_T\sigma^2}\mathbf{R}_{R_X}^{1/2}\mathbf{H}_w\mathbf{R}_{T_X}\mathbf{H}_w^H\mathbf{R}_{R_X}^{H/2}\right)\right\}$$

For $N_T = N_R = N$, and assuming that the matrices $\mathbf{R}_{R_X}$ and $\mathbf{R}_{T_X}$ are full rank, we have,

$$\Rightarrow \qquad \frac{C}{W} \approx \log_2\left\{\det\left(\frac{P}{N_T\sigma^2}\mathbf{R}_{R_X}^{1/2}\mathbf{H}_w\mathbf{R}_{T_X}\mathbf{H}_w^H\mathbf{R}_{R_X}^{H/2}\right)\right\}$$

$$= \log_2\left\{\det\left(\frac{P}{N_T\sigma^2}\mathbf{H}_w\mathbf{H}_w^H\right)\right\} + \log_2\left\{\det\left(\mathbf{R}_{R_X}\right)\right\} + \log_2\left\{\det\left(\mathbf{R}_{T_X}\right)\right\} \quad (6.2a)$$

In the above equation, we have used, det(**I** + **AB**) = det(**I** + **BA**) and $\log_2(1 + \text{SNR}) \approx \log_2(SNR)$ when $SNR >> 1$. Hence the MIMO channel capacity has been reduced and the amount of reduction in the capacity is given by $\log_2\left\{\det\left(\mathbf{R}_{R_X}\right)\right\} + \log_2\left\{\det\left(\mathbf{R}_{T_X}\right)\right\}$. Note the above two terms are always negative since the $\det(\mathbf{R}_{R_X})$ and $\det(\mathbf{R}_{T_X})$ have values which are less than or equal to 1.

Example 6.3

Show that $\log_2\left\{\det\left(\mathbf{R}_{R_X}\right)\right\} + \log_2\left\{\det\left(\mathbf{R}_{T_X}\right)\right\}$ is always negative.

Solution

Note that $\mathbf{R}_{T_X} = E[\mathbf{H}^H\mathbf{H}]^T$ and $\mathbf{R}_{R_X} = E[\mathbf{H}\mathbf{H}^H]$. The diagonal elements are 1 and off-diagonal elements hold a value between 0 and 1. Hence trace($\mathbf{R}_{T_X}$) = N_T (number of transmitting antennas) and trace trace($\mathbf{R}_{R_X}$) = N_R (number of receiving antennas). Note that product of all eigenvalues of a matrix is equal to the determinant of the matrix. And the geometric mean is bounded by the arithmetic mean.

$$\left(\prod_{i=1}^{N_R}\lambda_i\right)^{\frac{1}{N_R}} \le \frac{1}{N_R}\sum_{i=1}^{N_R}\lambda_i = \frac{1}{N_R}trace\left(\mathbf{R}_{R_X}\right) = 1$$

Therefore,

$$\det\left(\mathbf{R}_{R_X}\right) = \prod_{i=1}^{N_R}\lambda_i \le 1$$

In the similar way we can show that

$$\det\left(\mathbf{R}_{T_X}\right) = \prod_{i=1}^{N_T}\lambda_i \le 1$$

Hence, $\log_2\left\{\det\left(\mathbf{R}_{R_X}\right)\right\} + \log_2\left\{\det\left(\mathbf{R}_{T_X}\right)\right\}$ is always negative.

6.3.1 Ergodic capacity of equi-correlated Rayleigh fading MIMO channels

We have considered equi-correlation at the transmitter and receiver

$$\mathbf{R}_{R_X} = \mathbf{R}_{T_X} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

where correlation coefficient ρ = 0.3, 0.5, 0.7, 0.9.

It can be observed from Fig. 6.3 that the correlation reduces the capacity. Ergodic capacity is highest for uncorrelated case when $\rho = 0$ and it keeps on decreasing for higher values of correlation coefficient, ρ. The exact computation of average capacity for doubly correlated MIMO channel is reported in H. Shin et al., 2006. It is a lengthy analysis. We will do a simplified analysis for the case when the number of antennas tends to infinity also known as asymptotic analysis.

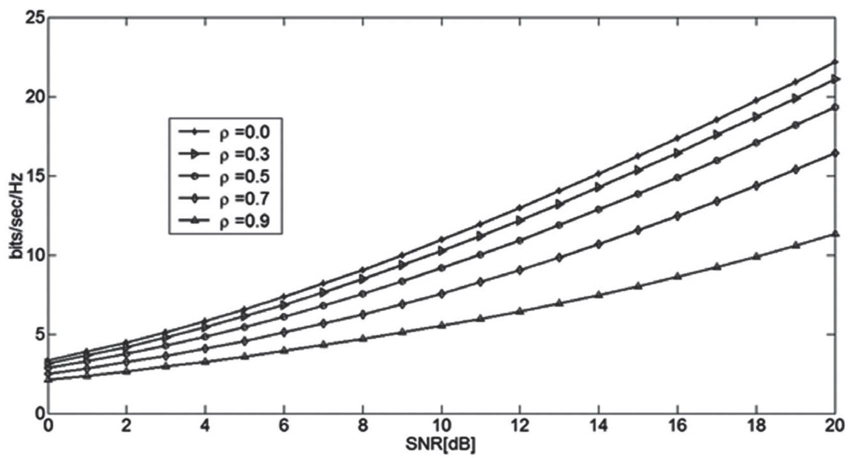


Fig. 6.3 Ergodic capacity of 4 × 4 open loop MIMO system for equi-correlated Rayleigh fading MIMO channel

Asymptotic analysis

If we take the expectation of Eq. (6.2a) for $N_R = N_T = N$ case, we have,

$$\langle C \rangle = N \log_2\left(\frac{P}{\sigma^2 N}\right) + E\left(\log_2 \left|\mathbf{H}_w \mathbf{H}_w^H\right|\right) + \log_2 \left|\mathbf{R}_{T_X} \mathbf{R}_{R_X}\right|$$

$$= N \log_2\left(\frac{\gamma}{N}\right) + E\left(\log_2 \left|\mathbf{H}_w \mathbf{H}_w^H\right|\right) + \log_2 \left|\mathbf{R}_{T_X} \mathbf{R}_{R_X}\right|$$

The strong law of large numbers (S. Ross, 2002):

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs, each having a finite mean $E(X_i) = \mu$. Then, with probability 1, $\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu$ as $n \to \infty$.

Example 6.4

Show that $E\left(\left|\mathbf{H}_w \mathbf{H}_w^H\right|\right) \to \left|N\mathbf{I}_N\right| = N^N$ as $n \to \infty$.

Solution

Note that

$$E\left(\mathbf{H}_w\mathbf{H}_w^H\right) = E\left(\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \ddots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{bmatrix}\begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N_R1} \\ h_{12} & h_{22} & \cdots & h_{N_R2} \\ \vdots & \ddots & \ddots & \vdots \\ h_{1N_T} & h_{2N_T} & \cdots & h_{N_RN_T} \end{bmatrix}^*\right)$$

$$= E\left(\begin{bmatrix} |h_{11}|^2 + |h_{12}|^2 + \cdots + |h_{1N_T}|^2 & 0 & \cdots & 0 \\ 0 & |h_{21}|^2 + |h_{22}|^2 + \cdots + |h_{2N_T}|^2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & |h_{N_R1}|^2 + |h_{N_R2}|^2 + \cdots + |h_{N_RN_T}|^2 \end{bmatrix}\right)$$

Consider each element of the above matrix $\mathbf{H}_w\mathbf{H}_w^H$ for large N. The mean will be N from strong law of large number for diagonal elements and zero for off-diagonal elements (note that all elements of $\mathbf{H}_w$ are independent and identically distributed). Hence, we have,

$$E\left(\mathbf{H}_w\mathbf{H}_w^H\right) \to N\mathbf{I}_N$$

Therefore, $E\left(\left|\mathbf{H}_w\mathbf{H}_w^H\right|\right) \to \left|N\mathbf{I}_N\right| = N^N$

The asymptotic ergodic capacity ($N \to \infty$) for a correlated $N \times N$ MIMO Rayleigh fading channel at high SNR is approximately given by

$$\left\langle C^{asymptotic} \right\rangle \approx N\log_2\left(\frac{\gamma}{N}\right) + N\log_2(N) + \log_2\left|\mathbf{R}_{T_X}\mathbf{R}_{R_X}\right| = N\log_2(\gamma) + \log_2\left|\mathbf{R}_{T_X}\mathbf{R}_{R_X}\right| \quad (6.2b)$$

Hence the capacity increases linearly with the number of antennas with a term which reduces the capacity, i.e., $\log_2\left|\mathbf{R}_{T_X}\mathbf{R}_{R_X}\right|$ which is always negative. Let us explore this second term in more detail in example 6.5.

Review question 6.1 *What is strong law of large numbers?*

Example 6.5

Assume a constant and separately correlated $N \times N$ MIMO channel model with receiver and transmitter correlation matrices as given below. It is worst case analysis of correlation as mentioned in section 3.4.

$$\mathbf{R}_{R_X} = \begin{bmatrix} 1 & \rho_R & \cdots & \rho_R \\ \rho_R & 1 & \cdots & \rho_R \\ \vdots & \vdots & \ddots & \vdots \\ \rho_R & \rho_R & \cdots & 1 \end{bmatrix}_{N\times N}$$

$$\mathbf{R}_{T_X} = \begin{bmatrix} 1 & \rho_T & \cdots & \rho_T \\ \rho_T & 1 & \cdots & \rho_T \\ \vdots & \vdots & \ddots & \vdots \\ \rho_T & \rho_T & \cdots & 1 \end{bmatrix}_{N \times N}$$

Find the approximate asymptotic ergodic capacity of such channel. Note that ρ_T, $\rho_R \in [0,1]$ and the ergodic capacity for a correlated fading channel at high SNR is approximately given by $\langle C^{asymptotic} \rangle \approx N \log_2(\gamma) + \log_2 \left| \mathbf{R}_{T_X} \mathbf{R}_{R_X} \right|$.

Solution

A $N \times N$ correlation matrix is called N^{th}-order (positive definite) constant correlation matrix with correlation coefficient $\rho \in [0,1]$, denoted by $\mathbf{R}(\rho)$, if it has the following structure:

$$\mathbf{R}_{R_X} = \begin{bmatrix} 1 & \rho_R & \cdots & \rho_R \\ \rho_R & 1 & \cdots & \rho_R \\ \vdots & \vdots & \ddots & \vdots \\ \rho_R & \rho_R & \cdots & 1 \end{bmatrix}_{N \times N}$$

$$\mathbf{R}_{T_X} = \begin{bmatrix} 1 & \rho_T & \cdots & \rho_T \\ \rho_T & 1 & \cdots & \rho_T \\ \vdots & \vdots & \ddots & \vdots \\ \rho_T & \rho_T & \cdots & 1 \end{bmatrix}_{N \times N}$$

This correlation model may approximate closely spaced antennas and may be used for the worst case analysis. Since eigenvalues of $\mathbf{R}_{T_X}$ are $1 + (N-1)\rho_T$ and $1 - \rho_T$ with N-1 multiplicities, its determinant can be written as

$$\left| \mathbf{R}_{T_X} \right| = (1 - \rho_T)^{N-1} (1 - \rho_T + N\rho_T)$$

Similarly, $$\left| \mathbf{R}_{R_X} \right| = (1 - \rho_R)^{N-1} (1 - \rho_R + N\rho_R)$$

Since determinant of a matrix is equal to the product of all its eigenvalues, and eigenvalue λ_i with multiplicity k would contribute $(\lambda_i)^k$ to the product, hence,

$$\langle C^{\text{asymptotic}} \rangle \approx N \log_2(\gamma) + \log_2 \left| \mathbf{R}_{T_X} \mathbf{R}_{R_X} \right| = N \log_2(\gamma) + \log_2 \left| \mathbf{R}_{T_X} \right| + \log_2 \left| \mathbf{R}_{R_X} \right|$$

$$= N \log_2(\gamma) + \log_2 \left\{ (1 - \rho_T)^{N-1} (1 - \rho_T + N\rho_T) \right\} + \log_2 \left\{ (1 - \rho_R)^{N-1} (1 - \rho_R + N\rho_R) \right\}$$

$$= N \log_2(\gamma) + (N-1)\log_2(1 - \rho_T) + \log_2(1 - \rho_T + N\rho_T) + (N-1)\log_2(1 - \rho_R) + \log_2(1 - \rho_R + N\rho_R)$$

6.3.2 Outage capacity of separately correlated Rayleigh fading MIMO channels

Outage capacity of doubly correlated MIMO channel is provided in H. Shin et al., 2006. We have generated outage capacity of equi-correlated Rayleigh fading MIMO channels using Monte Carlo simulations. It can be read from the graphs in Fig. 6.4, for a 5 × 5 MIMO channel, the 0.2 outage capacity is approximately 7.5 bits/sec/Hz for SNR of 5 dB when $\rho = 0$ (uncorrelated), whereas, the 0.2 outage capacity is approximately 6.5 bits/sec/Hz, 5.7 bits/sec/Hz, 4.5 bits/sec/Hz and 3 bits/sec/Hz for $\rho = 0.3, 0.5, 0.7, 0.9$, respectively. Hence the 0.2 outage capacity decreases with increase in the correlation coefficient, ρ.

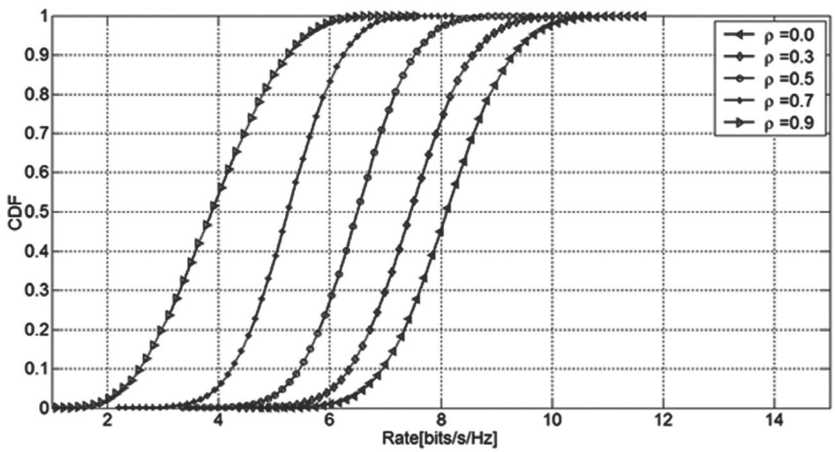


Fig. 6.4 CDF of open loop 5 × 5 MIMO channel capacity for $SNR = 5dB$

6.4 Capacity of keyhole Rayleigh fading MIMO channel

Let us consider the keyhole channel which was described in the previous chapters. The rank of the keyhole channel matrix is one and thus there is no MUX gain whereas diversity gain is found in the channel. The capacity of the channel is given by

$$C = W \log_2\left(1 + \frac{\lambda P}{\sigma^2}\right)$$

where, $\sqrt{\lambda}$ is the singular value of the channel matrix $\mathbf{H}$.

6.4.1 Ergodic capacity of keyhole Rayleigh fading MIMO channel

The ergodic capacity of keyhole MIMO is obtained by taking expectation of the instantaneous capacity expression of the above equation over the pdf of keyhole Rayleigh channel model of section 3.5.

$$p_Z(z) = \frac{2z^{\frac{N_T+N_R}{2}-1}}{\Gamma(N_T)\Gamma(N_R)} K_{N_R-N_T}\left(2\sqrt{z}\right); z \geq 0$$

One may refer to (H. Shin et al., 2003) for proof of the above equation.

$$\langle C \rangle = \int_0^\infty \log_2\left(1 + \frac{\gamma z}{N_T}\right) p_Z(z)\,dz = \int_0^\infty \log_2\left(\frac{\gamma z}{N_T}\right) p_Z(z)\,dz + \int_0^\infty \log_2\left(1 + \frac{N_T}{\gamma z}\right) p_Z(z)\,dz = I_1 + I_2$$

where, $I_1 = \log_2\left(\frac{\gamma}{N_T}\right) + \log_2(e)\left\{\Psi(N_T) + \Psi(N_R)\right\}; I_2 = \frac{\log_2(e)}{\Gamma(N_T)\Gamma(N_R)} G_{2,4}^{3,2}\left(\frac{N_T}{\gamma}\Big|_{N_R,N_T,1,0}^{1,1}\right)$

Hence,

$$\langle C \rangle = \log_2\left(\frac{P}{\sigma^2 N_T}\right) + \log_2(e)\left\{\Psi(N_T) + \Psi(N_R)\right\} + \frac{\log_2(e)}{\Gamma(N_T)\Gamma(N_R)} G_{2,4}^{3,2}\left(\frac{P}{\sigma^2 N_T}\Big|_{N_R,N_T,1,0}^{1,1}\right) \tag{6.3}$$

$\Psi(z) = \frac{d}{dx}\left(\ln\Gamma(z)\right) = \frac{\Gamma'(z)}{\Gamma(z)} = \ln(x) - \sum_{k=0}^{\infty}\left[\frac{1}{x+k} - \ln\left(1 + \frac{1}{x+k}\right)\right]$ is the Euler's digamma function. $G_{p,q}^{m,n}(\bullet)$ is the Meijer's G-function. The expression of Meijer's G function (I. S. Gradshteyn et al., 2000) is given by

$$G_{p,q}^{m,n}\left(x\Big|\begin{matrix} a_1, & \ldots, & a_p \\ b_1 & \ldots, & b_q \end{matrix}\right) = \frac{1}{2\pi i}\int \frac{\prod_{j=1}^{m}\Gamma(b_j - s)\prod_{j=1}^{n}\Gamma(1 - a_j - s)}{\prod_{j=m+1}^{q}\Gamma(1 - b_j + s)\prod_{j=n+1}^{p}\Gamma(a_j - s)} x^s\,ds; 0 \leq m \leq q, 0 \leq n \leq p$$

The poles of $\Gamma(b_j - s)$ must not coincide with the poles of $\Gamma(1 - a_k + s)$ for any j and k ($j = 1, \ldots, m$; $k = 1, \ldots, n$).

6.4.2 Outage capacity of keyhole Rayleigh fading MIMO channel

It can be observed from the graph in Fig. 6.5, for a 5 ×5 MIMO channel, the 0.2 outage capacity is approximately 3 bits/sec/Hz for SNR of 5 dB, whereas, for a 7 ×7 MIMO channel, the 0.2 outage capacity is approximately 3.5 bits/sec/Hz for SNR of 5 dB. There is significant reduction in the 0.2 outage capacity of Fig. 6.2 for the 5 × 5 and 7 × 7 MIMO channel due to the keyhole propagation.

Review question 6.2 *What is the outage and ergodic capacity of i.i.d. Rayleigh fading MIMO channel?*

Review question 6.3 *What is the simple bound on the ergodic capacity of i.i.d. Rayleigh fading MIMO channel?*

Review question 6.4 *What is the asymptotic ergodic capacity of separately correlated Rayleigh fading MIMO channel for high SNR case?*

Review question 6.5 *What is the ergodic capacity of keyhole Rayleigh fading MIMO channel?*

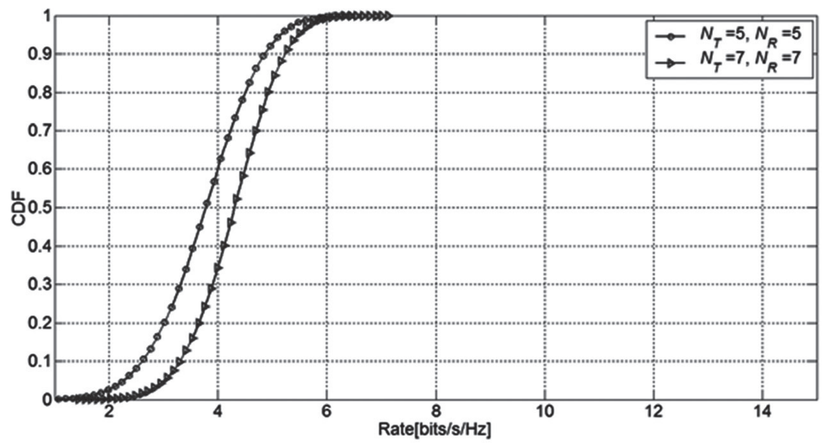


Fig. 6.5 CDF of open loop $N_T \times N_R$ MIMO channel capacity for *SNR = 5dB* for keyhole propagation

6.5 Summary

Figure 6.6 shows the chapter in a nutshell. In this chapter, we have found out the ergodic and outage capacity of random MIMO channels, viz. i.i.d. MIMO channel, Kronecker MIMO channel and keyhole MIMO channel.

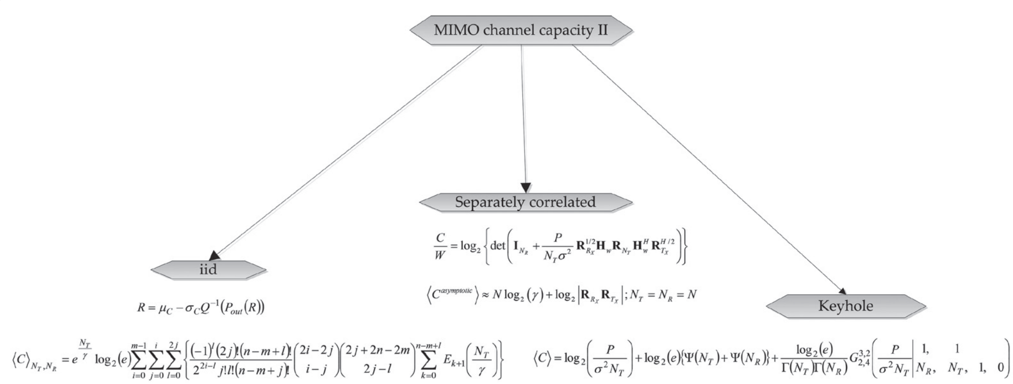


Fig. 6.6 Chapter 6 in a nutshell

Exercises

The following exercises may not have close-form formulae, but may be explored.

Exercise 6.1

Find the ergodic and outage capacity for separately correlated MIMO channel whose fading distribution is
(a) Nakagami-m
(b) Rice
(c) Hoyt
(d) Weibull

Exercise 6.2

Find the ergodic and outage capacity for keyhole MIMO channel whose fading distribution is
(a) Nakagami-m
(b) Rice
(c) Hoyt
(d) Weibull

Exercise 6.3

Find the ergodic and outage capacity for i.i.d. (uncorrelated) MIMO channel whose fading distribution is
(a) α-μ
(b) k-μ
(c) η-μ

Exercise 6.4

Find the ergodic and outage capacity for separately correlated MIMO channel whose fading distribution is
(a) α-μ
(b) k-μ
(c) η-μ

Exercise 6.5

Find the ergodic and outage capacity for Keyhole MIMO channel whose fading distribution is
(a) α-μ
(b) k-μ
(c) η-μ

References

1. Biglieri, E. and G. Taricco. 2004. *Transmission and Reception with Multiple Antennas: Theoretical Foundations*. Boston: Now Publishers.
2. Goldsmith, A. 2005. *Wireless Communications*. Cambridge: Cambridge University Press.
3. Gradshteyn, I. S. and I. M. Ryzhik. 2000. *Tables of Integrals, Series and Products*. London: Academic Press.
4. Haykin, S. and M. Moher. 2005. *Modern Wireless Communications*. New Delhi: Pearson.

5. Liang, Y. W. April 19th, 2005. 'Ergodic and Outage Capacity of Narrowband MIMO Gaussian Channels'. Department of Electrical and Computer Engineering, University of British Columbia. {www.ece.ubc.ca/~yangl/mimo_cap/mimo.pdf}.
6. Papoulis, A. and S. U. Pillai. 2002. *Probability, Random Variables and Stochastic Processes*. New Delhi: McGraw Hill.
7. Ross, S. 2002. *A First Course in Probability*. New Delhi: Pearson.
8. Shin, H. and J. H. Lee. 2003. 'Capacity of multiple-antenna fading channels: Spatial fading correlation, Double scattering and Keyhole'. *IEEE Trans. Information Theory*. 2636–2647.
9. Shin, H., M. Z. Win, J. H Lee, and M. Chiani. Aug. 2006. 'On the capacity of doubly correlated MIMO channels'. *IEEE Transaction on Wireless Communications*. 5(8). 2253–2265.
10. Telatar, I. E. Nov./ Dec. 1999. 'Capacity of multi-antenna Gaussian channels'. *Europ. Trans. Telecommun*. 10(6). 586–595.
11. Tulino, A. M. and S. Verdu. 2004. *Random Matrix Theory and Wireless Communications*. Boston: Now Publishers.

Introduction to Space-Time Codes

CHAPTER 7

7.1 Introduction

In this chapter, we will first discuss why we need space-time codes. Then we will discuss about the three code design criteria viz., rank, determinant and trace. We will study the first and most powerful space-time codes also known as Alamouti space-time codes, named after the inventor S. M. Alamouti. We will see the performance comparison of Alamouti space-time code with diversity combining scheme viz. maximal ratio combining (MRC) and find equivalence in the two systems. We will study the coding gain, diversity gain and code rate of the two antennas transmit diversity technique (Alamouti space-time code). We will see that due to orthogonality of transmitted signals, we can decouple the signals transmitted from antenna 1 and 2 at the receiver. This concept will be extended for any number of antennas which is called orthogonal space-time block codes (OSTBC). It will be discussed in the next chapter. A very important concept in performance analysis of wireless communication over fading channel is to find the average probability of error we need to average the conditional probability of error (CEP) over the received SNR. A channel is in outage whenever we are transmitting message at a rate higher than the channel capacity. In the last section, we will find the outage probability and average probability of error for single input single output (SISO) system over fading channels and extend this analysis further for Alamouti space-time codes.

7.2 Why space-time codes?

There are basically two types of space-time codes:

(a) Space-time block codes which is an extension of block codes
(b) Space-time trellis codes which is an extension of convolutional codes

Space-time block codes give diversity gain but no coding gain. Whereas, space-time trellis codes give both coding and diversity gain. We will discuss both these space-time codes in detail later. What do we mean by the terms diversity and coding gain in the context of space-time coding?

In space-time coded systems (R. Bose, 2008), the approximate symbol error rate (SER) for the system may be expressed as

$$\bar{P}_e \approx \frac{c}{(G_c S)^{G_d}} \tag{7.1a}$$

where, S represents the SNR, c is some constant, $G_c \geq 1$ is the coding gain and G_d represents the diversity order of the system.

If we take log to the base 2 of the equation (7.1a), we have,

$$\log_2(\bar{P}_e) \approx \log_2 c - G_d \log_2 G_c - G_d \log_2 S \tag{7.1b}$$

Recall the definition of diversity order/gain of Eq. (1.2). The diversity gain/order determines the negative slope of an error rate curve plotted vs SNR on a log-log scale (refer to Fig. 7.1). In other words, a space-time coded scheme with diversity order G_d has an error probability at high SNR behaving as $\bar{P}_e \approx (S)^{-G_d}$ (E. Biglieri et al., 2007). If there is some coding gain, then average probability of error will be of the form $\bar{P}_e \approx \frac{1}{(G_c S)^{G_d}}$. If there were no array or power gain then the probability of error expression will be of the form $\bar{P}_e \approx \frac{1}{G_c (S)^{G_d}}$. Note that the diversity gain is maximal for i.i.d. channel and it usually decreases for the correlated MIMO channels (B. Clerckx et al., 2013). The coding gain gives the horizontal shifting of the uncoded system error rate curve to the space-time coded error rate curve plotted on a log-log scale obtained for the same diversity order (refer to Fig. 7.1).

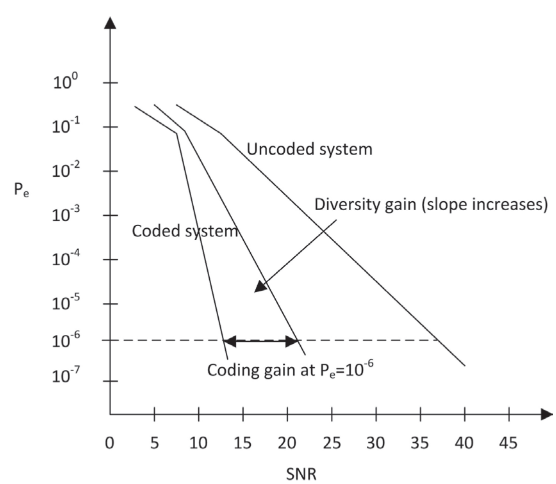


Fig. 7.1 Typical probability of error vs SNR curve for coded and uncoded system showing coding and diversity gain (BER curves are usually waterfall type but we have shown straight lines for illustration purpose only)

In yesteryears, spatially multiplexed MIMO systems (chapter 10) and space-time code designs were done separately. Spatially multiplexed MIMO systems were designed to increase the rate gain whereas space-time codes were designed to increase the diversity gain. In the recent developments of space-time codes, one tries to construct codes which follows the diversity multiplexing trade-offs. In the light of these new progresses, the following two definitions (M. O. Sinnokrot et al., 2010) will be also used for classifying space-time codes (chapter 8).

- A space-time code is fully diverse if the codeword difference matrix $\mathbf{D}(\mathbf{C}^i, \mathbf{C}^j) = \mathbf{C}^i - \mathbf{C}^j$ has rank N_T for every two distinct code words $\mathbf{C}^i$ and $\mathbf{C}^j$.
- A space-time code has the non-vanishing determinant (NVD) property if the coding gain for a fully diverse code does not tends to zero as the constellation size gets bigger.

Review question 7.1 *Define diversity and coding gain.*

Review question 7.2 *Which space-time code has both coding and diversity gain?*

Review question 7.3 *What is fully diverse space-time code?*

Review question 7.4 *What is non-vanishing determinant (NVD) property of space-time code?*

7.3 Code design criteria

If we consider transmission over a binary symmetric channel using a linear block channel code, then the bit error rate (BER) of the system depends on the Hamming distances of the codeword pairs (D. G. Hoffman et al., 1991). If we denote the minimum Hamming distance between the group of every possible codeword pairs by $d_{\min}$, then such a code can correct every error patterns of less than or equal to the largest integer less than or equal to $\frac{d_{\min}-1}{2}$. The design policy for such a code is to maximize the minimum Hamming distance among the codeword pairs.

Furthermore, for an AWGN channel, a good code design policy is to maximize the minimum Euclidean distance among every possible codeword pairs. The fundamental question is: what are the code design criteria for space-time codes?

7.3.1 What are the design criteria for space-time codes?

There are three criteria for designing space-time codes (H. Jafarkhani, 2005):

- Rank criterion
- Determinant criterion
- Trace criterion

The probability that the decoder decides in favour of codeword matrix $\mathbf{C}^2$ instead of the codeword matrix $\mathbf{C}^1$ which was transmitted is given by Pairwise error probability. In other words, the rank of the codeword difference matrix $\mathbf{D}(\mathbf{C}^1, \mathbf{C}^2)$ multiplied by the number of receive antennas (N_R) gives the diversity order of the space-time code.

Example 7.1

Show that the diversity gain of space-time codes is rN_R.

Solution

The diversity gain (Eq. 1.2) was defined as

$$G_d = -\lim_{SNR\to\infty} \frac{\log_2(P_e)}{\log_2(\gamma)}$$

Hence diversity gain of space-time codes can be obtained from the Eq. (8.12b) as follows.

$$G_d = \lim_{\gamma\to\infty} - \frac{\log_2\left\{\frac{1}{2}\frac{4^{rN_R}}{\left(\prod_{n=1}^{r}\lambda_n\right)^{N_R}\gamma^{rN_R}}\right\}}{\log_2(\gamma)}$$

$$= \lim_{\gamma\to\infty} \frac{\log_2(2) + \log_2\left(\prod_{n=1}^{r}\lambda_n\right)^{N_R} - \log_2\left(4^{rN_R}\right) + rN_R\log_2\{\gamma\}}{\log_2(\gamma)}$$

$$\approx \lim_{\gamma\to\infty} \frac{rN_R\log_2\{\gamma\}}{\log_2(\gamma)} = rN_R \tag{7.2}$$

Similarly the coding gain ($\left(\prod_{n=1}^{r}\lambda_n\right)^{\frac{1}{r}} = \left|\mathbf{A}\left(\mathbf{C}^1,\mathbf{C}^2\right)\right|^{\frac{1}{r}}$ E. Biglieri et al., 2007), where, $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{D}^H(\mathbf{C}^1, \mathbf{C}^2) \bullet \mathbf{D}(\mathbf{C}^1, \mathbf{C}^2) = (\mathbf{C}^1 - \mathbf{C}^2)^H \bullet (\mathbf{C}^1 - \mathbf{C}^2)$ relates to the multiplication of the non-zero eigenvalues of the codeword distance matrix $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2)$. A full diversity of $N_R N_T$ is feasible if the matrix $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2)$ is full rank. In this case, the coding gain relates to the products of eigenvalues (λ_n) or the determinant of the matrix, $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2)$. We can define coding gain distance (CGD) between two codewords $\mathbf{C}^1$ and $\mathbf{C}^2$: $CGD(\mathbf{C}^1, \mathbf{C}^2) = \det(\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2))$ and coding gain is defined as $G_c = \left[CGD\left(\mathbf{C}^1,\mathbf{C}^2\right)\right]^{\frac{1}{r}}$ where, r is that rank of $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2)$.

Example 7.2

Find the CGD and coding gain of two codeword matrices $\mathbf{C}^1$ and $\mathbf{C}^2$ for space-time trellis code given by

$$\mathbf{C}^1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{C}^2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

Also find the rank and trace of the codeword distance matrix.

Solution

The codeword difference matrix is

$$\mathbf{D}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{C^1} - \mathbf{C^2} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$$

The codeword distance matrix is

$$\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{D}\left(\mathbf{C^1},\mathbf{C^2}\right)\mathbf{D}\left(\mathbf{C^1},\mathbf{C^2}\right)^H = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

The rank and trace of the codeword distance matrix is 2 and 8 obviously.

The coding gain distance

$$CGD = \left(\mathbf{C^1},\mathbf{C^2}\right) = \left|\mathbf{A}\left(\mathbf{C^1},\mathbf{C^2}\right)\right| = 16$$

The coding gain is $G_C = \left(CGD\right)^{\frac{1}{2}} = \sqrt{16} = 4$

The diversity order of the system is twice the number of receiving antennas.

How do I get this codeword matrix? It will be explained later when we discuss space-time trellis codes (STTC). If we have a matrix with rank less than $r < N_T$, then we can define CGD as the product of the non-zero eigenvalues of $\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2)$. The three code design criteria are summarized below.

(a) A sound design policy to assure full diversity is to ensure that for every possible codeword pairs $\mathbf{C}^i$ and $\mathbf{C}^j$, $i \neq j$, the $\mathbf{A}(\mathbf{C}^i, \mathbf{C}^j)$ *is full rank* (RANK CRITERION).

(b) To augment the coding gain of a full diversity code, further good design policy would be to maximize the minimum determinant of the matrices $\mathbf{A}(\mathbf{C}^i, \mathbf{C}^j)$ $\forall i \neq j$ (DETERMINANT CRITERION).

We can also express average PEP upper bound of space-time codes over flat fading i.i.d. Rayleigh channel as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) \leq \frac{1}{4}\exp\left(-N_R \left\|\mathbf{D}\left(\mathbf{C}^1,\mathbf{C}^2\right)\right\|_F^2 \frac{\gamma}{4}\right) \tag{7.3}$$

A norm gives a real-valued function which is a measure of the size or length of multi-component mathematical entities such as vectors and matrices. The Euclidean norm of a vector $\vec{F} = a\hat{x} + b\hat{y} + c\hat{z}$ or $\vec{F} = \left(a,b,c\right)$ is defined as $\left\|\vec{F}\right\|_e = \sqrt{a^2 + b^2 + c^2}$ and it is the distance of this vector $\vec{F}$ from the origin. For n-dimensional space, Euclidean norm of a vector $\vec{X} = \left(x_1, x_2, \ldots, x_n\right)$ can be computed as $\left\|\vec{X}\right\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$. This concept can be extended to a matrix and we can find Frobenius norm of an $n \times n$ matrix $\mathbf{A}$ as $\left\|\mathbf{A}\right\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}^2}$. As with remaining vector norms, it quantifies a specific value for size of the matrix. Note that Frobenius norm of a matrix $\mathbf{A}$ could be also defined as $\left\|\mathbf{A}\right\|_F = \sqrt{trace\left(\mathbf{A}\mathbf{A}^H\right)} = \sqrt{trace\left(\mathbf{A}^H\mathbf{A}\right)}$.

(c) Another good design policy is to maximize the minimum squared Frobenius norm of codeword difference matrix $\left\|\mathbf{D}\left(\mathbf{C}^i,\mathbf{C}^j\right)\right\|_F^2 = Trace\left(\mathbf{D}^H\left(\mathbf{C}^i,\mathbf{C}^j\right)\bullet\mathbf{D}\left(\mathbf{C}^i,\mathbf{C}^j\right)\right) = Trace\left(\mathbf{A}\left(\mathbf{C}^i,\mathbf{C}^j\right)\right)$ among all possible $i \neq j$ (TRACE CRITERION). If we maximize the minimum trace of codeword distance matrix between all pairs of codeword matrices then we can minimize the probability of error bound.

Review question 7.5 *What are the three code design policies for space-time codes?*

7.4 Alamouti space-time codes

We have discussed the space-time code design criteria. What is this space-time coding? How to do coding in both space and time? Let us try to find answers to these questions in the next sub-section by taking the simple example of Alamouti space-time codes. Alamouti space-time code was developed by S. M. Alamouti in 1998. For coherent space-time code (B. Ahmed et al., 2015) channel state information (CSI) is available at the receiver. If CSI is not available at the receiver (for non-coherent space-time code), we can design differential space-time code which can be decoded at the receiver without the knowledge of CSI (H. Jafarkhani, 2005). We will consider coherent space-time code in this section.

7.4.1 What is space-time coding?

We will answer this question with reference to Alamouti space-time codes. The information bit streams are first modulated using any M-ary modulator which takes $\log_2(M)$ bits at a time and outputs symbol based on the M-ary modulation scheme. The Alamouti space-time encoder (shown in Fig. 7.2) then takes a block of two modulated symbols s_1 and s_2 in each encoding process and generates the codeword matrix (**S**).

$$\mathbf{S} = \begin{bmatrix} \mathbf{s^1} \\ \mathbf{s^2} \end{bmatrix} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}$$

The first column of **S** matrix serves for the first transmission period and the second column for the second transmission period. The first row gives the symbols transmitted for the first antenna and second row provides to the symbols transmitted from the second antenna. In the course of first symbol period, the first antenna sends s_1 and second antenna sends s_2. In the course of second symbol period, the first antenna sends $-s_2^*$ and the second antenna sends s_1^* (* denotes complex conjugate). Hence we are sending symbols both in space (over two antennas in space) and time (two transmission intervals over time). This is known as space-time coding (M. Jankiraman, 2004). We will denote first and second rows of the Alamouti space-time codeword by $\mathbf{s}^1 = [s_1 - s_2^*]^T$ and $\mathbf{s}^2 = [s_2 - s_1^*]^T$ respectively. Note that $\mathbf{s^1}$ and $\mathbf{s^2}$ are orthogonal (i.e., the inner product of $\mathbf{s^1}$ and $\mathbf{s^2}$, $\langle \mathbf{s}^1, \mathbf{s}^2 \rangle = 0$).

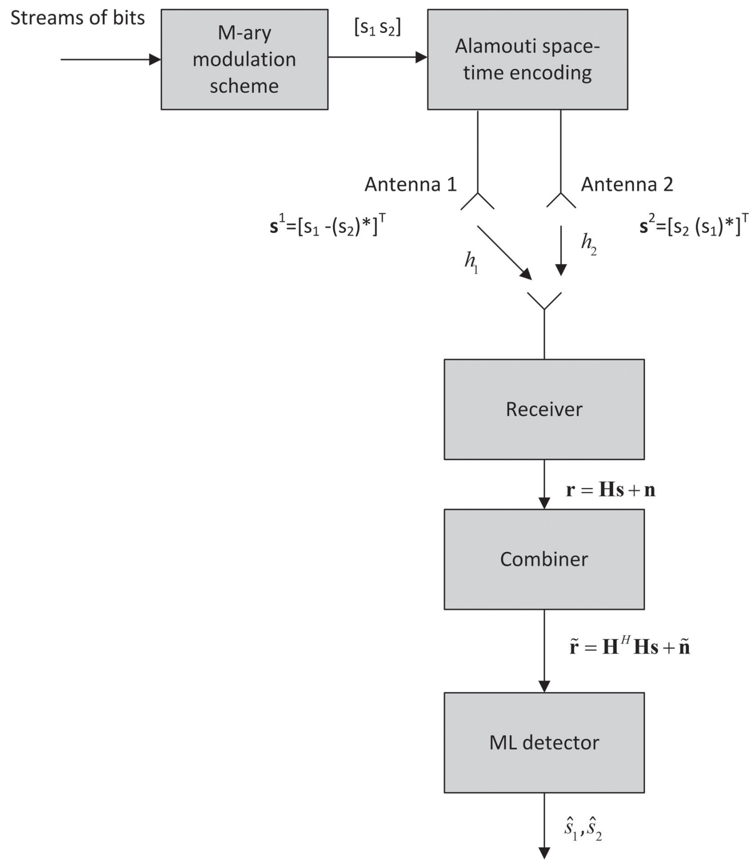


Fig. 7.2 A block diagram of Alamouti space-time encoder and decoder

7.4.2 What are Alamouti space-time codes?

S. M. Alamouti (1998) proposed space-time code for two antennas transmit diversity scheme and hence it is known as Alamouti space-time code. We have already discussed the encoder structure of such space-time codes. We will discuss about its detection process in this section. Let us assume single antenna at the receiver and double antennas at the transmitter as depicted in Fig. 7.2. The channel gain coefficients from antennas 1 and 2 are represented by $h_1(t)$ and $h_2(t)$, respectively, at time t. If we allow these coefficients to remain equal for two successive symbol periods, we have,

$$h_1(t) = h_1(t+T) = h_1 = |h_1|e^{j\theta_1}$$

$$h_2(t) = h_2(t+T) = h_2 = |h_2|e^{j\theta_2}$$

where, $|h_i|$ and θ_i, $i = 1,2$ are the channel gain's amplitude and phase shift for the link from transmit antenna i to the receive antenna and T is the symbol duration.

At the receiver the signals after passing through the additive white Gaussian noise (AWGN) channel may be expressed as

$$r_1 = h_1 s_1 + h_2 s_2 + n_1$$

$$r_2 = -h_1 s_2^* + h_2 s_1^* + n_2$$

where, n_1 and n_2 are independent ZMCCCG additive white noise samples at time t and $t+T$, respectively.

Note that r_1 and r_2 are received signals in two time intervals. In matrix form, the above equations could be expressed as

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}^T \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

An equivalent form of the above equation,

$$\begin{bmatrix} r_1 \\ r_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}$$

In compact matrix form, the received signal vector may be written as

$$\mathbf{r} = \mathbf{Hs} + \mathbf{n}$$

Note that we have taken the complex conjugate of the r_2 in the previous matrix representation.

The 2×1 vector signal output of the combiner is given as:

$$\begin{bmatrix} \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2^* \end{bmatrix} = \begin{bmatrix} h_1^* r_1 + h_2 r_2^* \\ h_2^* r_1 - h_1 r_2^* \end{bmatrix}$$

Similarly the 2×1 vector of additive complex noise in the combiner output is

$$\begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} = \begin{bmatrix} h_1^* n_1 + h_2 n_2^* \\ h_2^* n_1 - h_1 n_2^* \end{bmatrix}$$

Therefore,

$$\begin{bmatrix} \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix}$$

In matrix form, the output of the combiner can be written as

$$\tilde{\mathbf{r}} = \mathbf{H}^H \mathbf{Hs} + \tilde{\mathbf{n}}$$

which can be simplified as

$$\begin{bmatrix} \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix} = \begin{bmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_1|^2 + |h_2|^2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix}$$

$$= \left(|h_1|^2 + |h_2|^2 \right) \mathbf{I}_2 \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix}$$

$$= \left(|h_1|^2 + |h_2|^2 \right) \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix}$$

In order to perform this operation, the combiner needs the CSI. The above matrix equation can be written as

$$\tilde{r}_1 = \left(|h_1|^2 + |h_2|^2\right)s_1 + \tilde{n}_1; \tilde{r}_2 = \left(|h_1|^2 + |h_2|^2\right)s_2 + \tilde{n}_2 \tag{7.4}$$

Thus s_1 is separated from s_2. The two signals are completely decoupled after the combining operation. This greatly simplifies the detection strategy as we will see shortly.

7.4.3 ML detection

Combiner output is given to the Maximum likelihood (ML) decoder which minimizes the below decision metric. Optimal ML decision criterion is the minimum Euclidean distance criterion. For $0 \le t \le T$, we may write

$$\hat{s}_1 = \underset{\min\{m\}}{\arg} \left\| \tilde{r}_1 - \left(|h_1|^2 + |h_2|^2\right)s_m \right\|$$

where, $s_m \in \{s_k\}_{k=1}^{M}$ is one of the M-ary symbols. We may express above decision criterion as

$$\hat{s}_1 = \underset{\min\{m\}}{\arg} \left\| \frac{\tilde{r}_1}{\left(|h_1|^2 + |h_2|^2\right)} - s_m \right\| \tag{7.5a}$$

We have detected and estimated symbol 1 in the first time interval.

For $T \le t \le 2T$, we have,

$$\hat{s}_2 = \underset{\min\{m\}}{\arg} \left\| \tilde{r}_2 - \left(|h_1|^2 + |h_2|^2\right)s_m \right\|$$

We may express the above decision criterion as

$$\hat{s}_2 = \underset{\min\{m\}}{\arg} \left\| \frac{\tilde{r}_2}{\left(|h_1|^2 + |h_2|^2\right)} - s_m \right\| \tag{7.5b}$$

We have detected and estimated symbol 2 in the second time interval.

In order to perform the ML detection, the detector requires CSI.

7.4.4 What is the equivalent MRC receiver diversity?

In the case of maximum ratio combining (MRC), we will have two receive antennas at the receiver and one transmit antenna at the transmitter. It is similar to 1×2 SIMO system. Hence the received signals are

$$r_1 = h_1 s_0 + n_1$$

$$r_2 = h_2 s_0 + n_2$$

and the signal after the combining operation is given by

$$\tilde{r}_0 = h_1^* r_1 + h_2^* r_2 = \left(|h_1|^2 + |h_2|^2\right) s_0 + h_1^* n_1 + h_2^* n_2 \tag{7.6}$$

Note that in MRC combiner, we multiply the received signal on every branch with the complex conjugate of the channel gain coefficients of the corresponding branch and sum them. Note that the MRC signal is equivalent to the resulting combined signals of the transmit diversity scheme (Alamouti space-time code) above, except for a phase difference in the noise components. Hence, the two schemes have the same effective SNR. This shows that the diversity order from Alamouti space-time code (with one receive antenna) is the same as that of the two branch MRC. This equivalence could be generalized to higher values of N_R. For instance, the performance of this $N_T = 2$ and $N_R = 2$ Alamouti scheme is equivalent to that of $N_T = 1$ and $N_R = 4$ MRC (provided that each transmit antenna transmits the same power as with $N_T = 1$). In general, Alamouti space-time code with $N_T = 2$ and N_R number of receiving antennas has the same performance of a MRC with $2N_R$ receive antennas (M. Jankiraman, 2004).

Example 7.3

Discuss performance analysis of maximal-ratio combining in brief. Find the diversity order.

Solution

By selecting the antenna weight as $w_{MRC} = \alpha_n^*$ where α_n is the instantaneous attenuation of the n^th diversity branch, the combined output SNR is the sum of the branch SNRs

$$\gamma_{MRC} = \sum_{n=1}^{N_R} \gamma_n$$

The average combined SNR is given by

$$\bar{\gamma}_{MRC} = N_R \bar{\gamma}$$

where, $\bar{\gamma}$ is the average SNR on each branch.

Assuming each branch to have average equal $SNR\,\bar{\gamma}$ and independent Rayleigh faded branches, then γ_{MRC} is distributed as Chi-square distribution with $2N_R$ degrees of freedom with mean $N_R\bar{\gamma}$.

$$p_\gamma(\gamma) = \frac{\gamma^{N_R-1} e^{-\frac{\gamma}{\bar{\gamma}}}}{\bar{\gamma}^{N_R}(N_R - 1)!};\ \gamma \geq 0$$

The outage probability for given threshold γ_0 is

$$P_{\text{out}} = \Pr(\gamma < \gamma_0) = \int_0^{\gamma_0} p_\gamma(\gamma)\, d\gamma = 1 - e^{-\frac{\gamma_0}{\bar{\gamma}}} \sum_{k=1}^{N_R} \frac{\left(\frac{\gamma_0}{\bar{\gamma}}\right)^{k-1}}{(k-1)!}$$

It can be inferred that the above expression decays with $\left(\frac{1}{\bar{\gamma}}\right)^{N_R}$ at high SNR. Hence the diversity order is N_R.

Average error rates if BPSK is employed for MRC over Rayleigh fading channel is expressed as

$$P_{e,MRC} = \int_0^\infty Q\left(\sqrt{2\gamma}\right)p_\gamma(\gamma)\,d\gamma$$

Using Chernoff's bound on Q-function $Q(\gamma) \le \frac{1}{2}e^{-\frac{\gamma^2}{2}}$, we have,

$$P_{e,MRC} \le \frac{1}{2}\int_0^\infty e^{-\gamma}p_\gamma(\gamma)\,d\gamma = \frac{1}{2}\int_0^\infty e^{-\gamma}\frac{\gamma^{N_R-1}e^{-\frac{\gamma}{\bar{\gamma}}}}{\bar{\gamma}^{N_R}(N_R-1)!}\,d\gamma = \frac{1}{2}\frac{1}{(1+\bar{\gamma})^{N_R}} \le \frac{1}{2}\frac{1}{(\bar{\gamma})^{N_R}}$$

For further exploration on MRC, one may refer to T. M. Duman et al., 2007 and K. L. Du et al., 2007.

7.4.5 Diversity gain, coding gain and code rate

In the Alamouti scheme, there are orthogonal transmissions. This means that the receiver observes two entirely orthogonal streams. For two distinct code sequences, $\mathbf{S}_A$ and $\mathbf{S}_B$ produced by the inputs $(s_{A1} \quad s_{A2})$ and $(s_{B1} \quad s_{B2})$, where $(s_{A1} \quad s_{A2}) \neq (s_{B1} \quad s_{B2})$.

$$\mathbf{S}_A = \begin{bmatrix} s_{A1} & -s_{A2}^* \\ s_{A2} & s_{A1}^* \end{bmatrix}; \mathbf{S}_B = \begin{bmatrix} s_{B1} & -s_{B2}^* \\ s_{B2} & s_{B1}^* \end{bmatrix}$$

The rows of the codeword matrix are orthogonal, and hence the rows of the code word difference matrix are orthogonal too. The codeword difference matrix is given by

$$\mathbf{D}(\mathbf{S}_A, \mathbf{S}_B) = \begin{bmatrix} s_{A1} - s_{B1} & -s_{A2}^* + s_{B2}^* \\ s_{A2} - s_{B2} & s_{A1}^* - s_{B1}^* \end{bmatrix}$$

Since $(s_{A1} \quad s_{A2}) \neq (s_{B1} \quad s_{B2})$, obviously the distance matrices of every two unequal code words have a full rank of two. Hence, Alamouti scheme provides full transmit diversity of two. The code word distance matrix $\mathbf{A}(\mathbf{S}_A, \mathbf{S}_B) = \mathbf{D}(\mathbf{S}_A, \mathbf{S}_B)\,\mathbf{D}^H(\mathbf{S}_A, \mathbf{S}_B)$ is given below.

$$\mathbf{A}(\mathbf{S}_A, \mathbf{S}_B) = \begin{bmatrix} |s_{A1} - s_{B1}|^2 + |s_{A2} - s_{B2}|^2 & 0 \\ 0 & |s_{A1} - s_{B1}|^2 + |s_{A2} - s_{B2}|^2 \end{bmatrix}$$

7.4.5.1 *How about the diversity gain?*

We know that the diversity gain is dependent on the rank of the codeword distance matrix. The code word distance matrix has two equal eigenvalues. Two eigenvalues, hence the codeword distance matrix has achieved the full diversity.

7.4.5.2 *How about the coding gain?*

Coding gain distance (CGD) decides the coding gain. Let us find the CGD for our case. It can be calculated from the determinant of the codeword distance matrix. We have,

$$\begin{aligned}|\mathbf{A}(\mathbf{S}_A, \mathbf{S}_B)| &= \det\left(\begin{bmatrix} |s_{A1}-s_{B1}|^2+|s_{A2}-s_{B2}|^2 & 0 \\ 0 & |s_{A1}-s_{B1}|^2+|s_{A2}-s_{B2}|^2 \end{bmatrix}\right) \\ &= \left(|s_{A1}-s_{B1}|^2+|s_{A2}-s_{B2}|^2\right)^2\end{aligned}$$

The determinant of a matrix is also same as that of the multiplication of its eigenvalues. The coding gain can be calculated as

$$\begin{aligned}CGD^{\frac{1}{r}} &= \left\{\left(|s_{A1}-s_{B1}|^2+|s_{A2}-s_{B2}|^2\right)^2\right\}^{1/2} \\ &= \left(|s_{A1}-s_{B1}|^2+|s_{A2}-s_{B2}|^2\right)\end{aligned}$$

We can observe that G_c is equal to that of the uncoded system. It is identical to the squared Euclidean distance in the signal constellation (uncoded case). This implies that the coding gain is 1. This is the disadvantage of Alamouti scheme. Unlike the space-time trellis codes, this scheme achieves the transmit diversity gain without CSI at the transmitter but has no coding gain.

7.4.5.3 *How about the code rate?*

The code rate for Alamouti space-time code is 1 since we transmit two symbols over two time periods.

Review question 7.6 *What are coherent and non-coherent space-time codes?*

Review question 7.7 *Do we have coding gain with Alamouti space-time codes?*

Review question 7.8 *What is the equivalent MRC diversity for $2 \times N_R$ MIMO system employing Alamouti space-time code?*

7.5 SER analysis for Alamouti space-time code over fading channels

Since we know that symbol error rate (SER) is a function of the received SNR. For our case, the received signal for $0 \le t \le T$ (see Eq. 7.4) is

$$\tilde{r}_1 = \left(|h_1|^2+|h_2|^2\right)s_1 + h_1^* n_1 + h_2 n_2^*$$

Hence the received SNR can be expressed as

$$
\begin{aligned}
SNR_{Alamouti} &= \frac{\left(|h_1|^2 + |h_2|^2\right)^2 E_s}{\left(|h_1|^2 + |h_2|^2\right)\sigma^2} \\
&= \frac{\left(|h_1|^2 + |h_2|^2\right) E_s}{\sigma^2} \\
&= \frac{\left(|h_1|^2\right) E_s}{\sigma^2} + \frac{\left(|h_2|^2\right) E_s}{\sigma^2} \\
&= \gamma_1 + \gamma_2
\end{aligned}
$$

Transmit power is constrained as P, which implies that the overall transmit power should be the same immaterial of the number of transmit antennas. If channel state information (CSI) is not available at the transmitter, then we will allot equal power to each channel. Hence, we should modify the above equation as follows

$$
\begin{aligned}
SNR_{Alamouti} &= \frac{\left(|h_1|^2\right) \frac{E_s}{2}}{\sigma^2} + \frac{\left(|h_2|^2\right) \frac{E_s}{2}}{\sigma^2} \\
&= \frac{\left(|h_1|^2\right) \frac{PT}{2}}{\sigma^2} + \frac{\left(|h_2|^2\right) \frac{PT}{2}}{\sigma^2} = \frac{\gamma_1 + \gamma_2}{2} \qquad (7.7a)
\end{aligned}
$$

where, T is the symbol period.

For SISO case, the I–O relation is given as

$$y = hs + n$$

where, y is the received signal, n is AWGN, s is the symbol transmitted and h is the channel gain coefficient.

For single antenna case, total transmit power is P. Hence,

$$SNR_{SISO} = \frac{\left(|h|^2\right) E_s}{\sigma^2} = \frac{\left(|h|^2\right) PT}{\sigma^2} = \gamma \qquad (7.7b)$$

In this section, we will try to find the SER of various modulation schemes for single-input single-output (SISO) system over various fading channels and extend it for Alamouti space-time code. For SER analysis, there are two basic steps:

(a) First, find the conditional error probability (CEP) for the specific modulation scheme.
(b) Second, average it over the pdf of the received SNR to obtain average symbol error rate (SER).

In the moment generating function (MGF) based approach; we may express the SER as function of the MGF of the particular fading channel. The MGF for various fading channels is given in the book on digital communications over fading channels by M. K. Simon et al. (2005). The CEP for various modulation schemes is listed in wireless communications book by A. Goldsmith (2005). We will try to get the close-form equations for SER of various modulation schemes for SISO and Alamouti space-time code over various fading channels.

7.5.1 BER for BPSK over Rayleigh fading channel

Bit error rate (BER) of binary phase shift keying (BPSK) for single-input single-output (SISO) over Rayleigh fading channel has been widely explored (for example textbook for wireless communications by D. Tse et al., 2005). For the sake of completeness, we will start from BER analysis of BPSK for SISO over Rayleigh fading channel.

SISO:

For single antenna case, conditional error probability (CEP) for BPSK is given by

$$P_b(E \mid \gamma) = Q\left(\sqrt{2SNR}\right) = Q\left(\sqrt{2\gamma}\right)$$

Then average bit error rate (BER) can be obtained by averaging over the pdf of received SNR γ.

$$P_b(E) = E\left[P_b(E \mid \gamma)\right] = \int_0^\infty Q\left(\sqrt{2\gamma}\right) p_\gamma(\gamma) d\gamma$$

where, E is the expectation operator.

$$\because \qquad Q(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta, x \geq 0$$

Using this alternate form of Q(.) function (M. K. Simon et al., 2005) given above, we can obtain the average BER as

$$P_b(E) = \int_0^\infty \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{2\gamma}{2\sin^2\theta}\right) d\theta p_\gamma(\gamma) d\gamma$$

Integrating with respect to γ first, we have, the average BER of BPSK for SISO over any fading channel as

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\left(\int_0^\infty \exp\left(-\frac{\gamma}{\sin^2\theta}\right) p_\gamma(\gamma) d\gamma\right) d\theta = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \mathrm{M}_\gamma\left(-\frac{1}{\sin^2\theta}\right) d\theta$$

where, $\mathrm{M}_\gamma(s)$ is the moment generating function (MGF) of SNR γ.

For Rayleigh fading, the SNR (γ) is exponentially distributed.

Hence, MGF of γ is given by

$$\mathrm{M}_\gamma(s) = \int_0^\infty \exp(s\gamma)\frac{1}{\bar{\gamma}}\exp\left(-\frac{\gamma}{\bar{\gamma}}\right) d\gamma = \frac{1}{\bar{\gamma}}\int_0^\infty \exp\left(s\gamma - \frac{\gamma}{\bar{\gamma}}\right) d\gamma = \frac{1}{\bar{\gamma}} \left.\frac{\exp\left(s\gamma - \frac{\gamma}{\bar{\gamma}}\right)}{s - \frac{1}{\bar{\gamma}}}\right|_0^\infty = \frac{1}{1 - s\bar{\gamma}}$$

where, $\bar{\gamma} = E(\gamma)$

Hence, the BER of BPSK for SISO case over Rayleigh fading channel is given by

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{1}{1 + \frac{1}{\sin^2\theta}\bar{\gamma}} d\theta = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\sin^2\theta + \bar{\gamma}} d\theta$$

$$\because \qquad I_n(\bar{\gamma}) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left(\frac{\sin^2\theta}{\sin^2\theta + \bar{\gamma}}\right)^n d\theta$$

$$= \frac{1}{2} - \left[\frac{1}{2} - A(\bar{\gamma})\right]\sum_{i=0}^{n-1}\binom{2i}{i}A(\bar{\gamma})^i\left[1 - A(\bar{\gamma})\right]^i ; A(\bar{\gamma}) = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}\right]$$

where, n is an integer.

For n = 1, we have,

$$\therefore \qquad I_1(\bar{\gamma}) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left(\frac{\sin^2\theta}{\sin^2\theta + \bar{\gamma}}\right) d\theta$$

$$= \frac{1}{2} - \left[\frac{1}{2} - \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}\right]\right]$$

$$= \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}\right]$$

Hence, for Rayleigh, the average probability of error for BPSK for SISO is

$$P_b(E) = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}\right] \tag{7.8a}$$

Alamouti space-time code:

For Alamouti space-time code, conditional error probability (CEP) for BPSK is given by

$$P_b(E \mid \gamma_1, \gamma_2) = Q\left(\sqrt{2SNR}\right) = Q\left(\sqrt{2\frac{\gamma_1 + \gamma_2}{2}}\right) = Q\left(\sqrt{\gamma_1 + \gamma_2}\right)$$

Then average bit error rate (BER) of BPSK for Alamouti space-time code over a fading channel can be calculated as

$$P_b(E) = \int_0^{\infty}\int_0^{\infty} Q\left(\sqrt{\gamma_1 + \gamma_2}\right) p_{\gamma_1,\gamma_2}(\gamma_1, \gamma_2) d\gamma_1 d\gamma_2$$

For independent channel, we have,

$$P_b(E) = \int_0^{\infty}\int_0^{\infty} Q\left(\sqrt{\gamma_1 + \gamma_2}\right) p_{\gamma_1}(\gamma_1)\, p_{\gamma_2}(\gamma_2) d\gamma_1 d\gamma_2$$

$$\because \qquad Q\left(\sqrt{\gamma_1+\gamma_2}\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{\gamma_1+\gamma_2}{2\sin^2\theta}\right)d\theta$$

$$= \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{\gamma_1}{2\sin^2\theta}\right)\exp\left(-\frac{\gamma_2}{2\sin^2\theta}\right)d\theta$$

Hence,

$$P_b(E) = \int_0^{\infty}\int_0^{\infty}\frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{\gamma_1}{2\sin^2\theta}\right)\exp\left(-\frac{\gamma_2}{2\sin^2\theta}\right)d\theta p_{\gamma_1}(\gamma_1)p_{\gamma_2}(\gamma_2)d\gamma_1 d\gamma_2$$

$$= \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\int_0^{\infty}\int_0^{\infty} \exp\left(-\frac{\gamma_1}{2\sin^2\theta}\right)\exp\left(-\frac{\gamma_2}{2\sin^2\theta}\right)p_{\gamma_1}(\gamma_1)p_{\gamma_2}(\gamma_2)d\gamma_1 d\gamma_2 d\theta$$

Expressing in terms of MGF, we have,

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \mathrm{M}_{\gamma_1}\left(-\frac{1}{2\sin^2\theta}\right)\mathrm{M}_{\gamma_2}\left(-\frac{1}{2\sin^2\theta}\right)d\theta$$

For identical channels, we have the BER of BPSK for Alamouti space-time code over any fading channel, given as

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left(\mathrm{M}_{\gamma}\left(-\frac{1}{2\sin^2\theta}\right)\right)^2 d\theta$$

For Rayleigh fading case, the average BER for Alamouti space-time code employing BPSK is given by

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left(\frac{2\sin^2\theta}{2\sin^2\theta+\bar{\gamma}}\right)^2 d\theta \tag{7.8b}$$

7.5.2 BER for DBPSK over Nakagami fading channel

Note that in BPSK and MPSK information is carried in signal phase. In MQAM, information is carried in both phase and amplitude of the signal. Hence we need to have coherent demodulation at the receiver, for that we need to match the transmitted signal carrier phase and phase of the receiver phase. In differential modulation, we can utilize the previous symbol's phase as phase reference for the current symbol. Hence, we do not have the necessity of the coherent phase reference at the receiver (A. Goldsmith, 2005).

SISO:

For differential binary phase shift keying (DBPSK), the CEP for SISO case is given as $P_b(E \mid \gamma) = \frac{1}{2}e^{-\gamma}$

Hence the average BER of DBPSK for SISO over any fading channel is calculated as

$$P_b(E) = \int_0^\infty \frac{1}{2} e^{-\gamma} p_\gamma(\gamma) d\gamma$$

Hence, the average BER of DBPSK for SISO over any fading channel is given as

$$P_b(E) = \frac{1}{2} \mathrm{M}_\gamma(-1)$$

The MGF of Nakagami-m fading is given by

$$\mathrm{M}_\gamma(s) = \left(1 - \frac{s\overline{\gamma}}{m}\right)^{-m}$$

Hence, the average BER for DBPSK SISO case over a Nakagami-m fading channel,

$$P_b(E) = \frac{1}{2} \mathrm{M}_\gamma(-1) = \frac{1}{2}\left(1 + \frac{\overline{\gamma}}{m}\right)^{-m} \tag{7.9a}$$

Alamouti space-time code:

For Alamouti space-time code, we have average BER for DBPSK over any fading channel, given as

$$P_b(E) = \int_0^\infty \int_0^\infty \frac{1}{2} e^{-\frac{\gamma_1+\gamma_2}{2}} p_{\gamma_1}(\gamma_1) p_{\gamma_2}(\gamma_2) d\gamma_1 d\gamma_2$$

$$\Rightarrow \qquad P_b(E) = \int_0^\infty \int_0^\infty \frac{1}{2} e^{-\frac{\gamma_1}{2}} e^{-\frac{\gamma_2}{2}} p_{\gamma_1}(\gamma_1) p_{\gamma_2}(\gamma_2) d\gamma_1 d\gamma_2$$

Expressing in terms of MGF, we have the average BER of DBPSK over any fading channel as

$$P_b(E) = \frac{1}{2} \mathrm{M}_{\gamma_1}\left(-\frac{1}{2}\right) \mathrm{M}_{\gamma_2}\left(-\frac{1}{2}\right) = \frac{1}{2}\left(\mathrm{M}_{\gamma_1}\left(-\frac{1}{2}\right)\right)^2$$

Hence for Nakagami fading, the average BER for Alamouti space-time code employing DBPSK is obtained as

$$P_b(E) = \frac{1}{2}\left(1 + \frac{\overline{\gamma}}{2m}\right)^{-2m} \tag{7.9b}$$

7.5.3 SER for MPSK over Hoyt fading channel

SISO:

For M-ary phase shift keying (M-PSK), CEP is given by (J. Craig, 1991)

$$P_b(E \mid \gamma) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \exp\left(-\frac{g_{PSK}\gamma}{\sin^2\theta}\right) d\theta, \; g_{PSK} = \sin^2\left(\frac{\pi}{M}\right)$$

Then the average SER of M-PSK for SISO over a fading channel

$$P_b(E) = \int_0^{\infty} \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \exp\left(-\frac{g_{PSK}\gamma}{\sin^2\theta}\right) d\theta p_\gamma(\gamma) d\gamma$$

Integrating with respect to γ first, we have,

$$P_b(E) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \int_0^{\infty} \exp\left(-\frac{g_{PSK}\gamma}{\sin^2\theta}\right) p_\gamma(\gamma) d\gamma d\theta$$

Expressing in terms of MGF, we have the average SER of MPSK for SISO over any fading channel as

$$P_b(E) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \mathrm{M}_\gamma\left(-\frac{g_{PSK}}{\sin^2\theta}\right) d\theta$$

$$= \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \mathrm{M}_\gamma\left(-\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}\right) d\theta$$

For Hoyt fading MGF is given by

$$\mathrm{M}_\gamma(s) = \left((1 - 2s\overline{\gamma}) + \frac{(2s\overline{\gamma})^2 q^2}{(1+q^2)^2}\right)^{-\frac{1}{2}}$$

where, q ranges from 0 to 1.

Hence the average SER of M-PSK for SISO over a Hoyt fading channel

$$P_b(E) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \left(\left(1 + 2\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}\overline{\gamma}\right) + \frac{\left(2\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}\overline{\gamma}\right)^2 q^2}{(1+q^2)^2}\right)^{-\frac{1}{2}} d\theta \quad (7.10a)$$

Alamouti space-time code:

For Alamouti space-time code, MPSK over fading channel, we have SER as

$$P_b(E) = \int_0^{\infty}\int_0^{\infty} \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \exp\left(-\frac{g_{PSK}\frac{\gamma_1+\gamma_2}{2}}{\sin^2\theta}\right) d\theta p_{\gamma_1}(\gamma_1) p_{\gamma_2}(\gamma_2) d\gamma_1 d\gamma_2$$

Expressing in terms of MGF, we have the average SER of MPSK for Alamouti space-time code over any fading channel as

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\left(\mathrm{M}_\gamma\left(-\frac{g_{PSK}}{2\sin^2\theta}\right)\right)^2 d\theta$$

For Hoyt fading, the average SER for Alamouti space-time code employing M-PSK is

$$P_b(E) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\left(\left(1+\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}\bar{\gamma}\right)+\frac{\left(\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}\bar{\gamma}\right)^2 q^2}{\left(1+q^2\right)^2}\right)^{-1} d\theta \quad (7.10b)$$

7.5.4 SER for M-QAM over Rice fading channel

SISO:

For M-ary quadrature amplitude modulation (M-QAM), the CEP for SISO is given as

$$P_b(E\mid\gamma) = 4\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{2g_{QAM}\gamma}\right)-4\left(1-\frac{1}{\sqrt{M}}\right)^2 Q^2\left(\sqrt{2g_{QAM}\gamma}\right); g_{QAM} = \frac{3}{2(M-1)}$$

$$\because \qquad Q^2(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{4}}\exp\left(-\frac{x^2}{2\sin^2\theta}\right)d\theta, x\geq 0$$

Then the average SER of M-QAM for SISO over any fading channel

$$P_b(E) = \int_0^\infty 4\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{2g_{QAM}\gamma}\right)-4\left(1-\frac{1}{\sqrt{M}}\right)^2 Q^2\left(\sqrt{2g_{QAM}\gamma}\right)p_\gamma(\gamma)\,d\gamma$$

Integrating with respect to γ first, we have,

$$P_b(E) =$$

$$4\left(1-\frac{1}{\sqrt{M}}\right)\frac{1}{\pi}\int_0^{\frac{\pi}{2}}\int_0^\infty\exp\left(-\frac{g_{QAM}\gamma}{\sin^2\theta}\right)p_\gamma(\gamma)\,d\gamma d\theta - 4\left(1-\frac{1}{\sqrt{M}}\right)^2\frac{1}{\pi}\int_0^{\frac{\pi}{4}}\int_0^\infty\exp\left(-\frac{g_{QAM}\gamma}{\sin^2\theta}\right)p_\gamma(\gamma)\,d\gamma d\theta$$

Expressing in terms of MGF, we have the average SER of M-QAM for SISO over any fading channel as

$$P_b(E) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\mathrm{M}_\gamma\left(-\frac{g_{QAM}}{\sin^2\theta}\right)d\theta - \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\mathrm{M}_\gamma\left(-\frac{g_{QAM}}{\sin^2\theta}\right)d\theta$$

For Rice fading MGF is given by

$$\mathrm{M}_\gamma(s) = \frac{1+K}{1+K-s\bar{\gamma}}\exp\left(\frac{Ks\bar{\gamma}}{1+K-s\bar{\gamma}}\right)$$

For Rice fading, SER of MQAM for SISO is given by

$$P_b(E) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\frac{1+K}{1+K+\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}\exp\left(\frac{-K\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}{1+K+\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}\right)d\theta$$
$$-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\frac{1+K}{1+K+\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}\exp\left(\frac{-K\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}{1+K+\frac{g_{QAM}}{\sin^2\theta}\bar{\gamma}}\right)d\theta \quad (7.11a)$$

Alamouti space-time code:

For Alamouti space-time code, M-QAM over fading channel, we have SER as

$$P_b(E) =$$
$$\int_0^\infty\int_0^\infty\left[4\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{2g_{QAM}\frac{\gamma_1+\gamma_2}{2}}\right)-4\left(1-\frac{1}{\sqrt{M}}\right)^2 Q^2\left(\sqrt{2g_{QAM}\frac{\gamma_1+\gamma_2}{2}}\right)\right]p_{\gamma_1}(\gamma_1)\,p_{\gamma_2}(\gamma_2)d\gamma_1 d\gamma_2$$

Integrating with respect to γ_1 and γ_2 first, we have,

$$P_b(E) = 4\left(1-\frac{1}{\sqrt{M}}\right)\frac{1}{\pi}\int_0^{\frac{\pi}{2}}\int_0^\infty\int_0^\infty\exp\left(-\frac{g_{QAM}\frac{\gamma_1+\gamma_2}{2}}{\sin^2\theta}\right)p_{\gamma_1}(\gamma_1)\,p_{\gamma_2}(\gamma_2)\,d\gamma_1 d\gamma_2 d\theta$$
$$-4\left(1-\frac{1}{\sqrt{M}}\right)^2\frac{1}{\pi}\int_0^{\frac{\pi}{4}}\int_0^\infty\int_0^\infty\exp\left(-\frac{g_{QAM}\frac{\gamma_1+\gamma_2}{2}}{\sin^2\theta}\right)p_{\gamma_1}(\gamma_1)\,p_{\gamma_2}(\gamma_2)\,d\gamma_1 d\gamma_2 d\theta$$

Expressing in terms of MGF, we have,

$$P_b(E) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\mathrm{M}_{\gamma_1}\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)\mathrm{M}_{\gamma_2}\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)d\theta$$
$$-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\mathrm{M}_{\gamma_1}\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)\mathrm{M}_{\gamma_2}\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)d\theta$$

For equal channels, we have the SER of M-QAM for Alamouti space-time code over any fading channel as

$$P_b(E) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\left(\mathrm{M}_\gamma\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)\right)^2 d\theta - \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\left(\mathrm{M}_\gamma\left(-\frac{g_{QAM}}{2\sin^2\theta}\right)\right)^2 d\theta$$

For Rice fading, the SER of M-QAM for Alamouti space-time code is given by

$$P_b(E) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\left[\frac{1+K}{1+K+\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}\exp\left(\frac{-K\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}{1+K+\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}\right)\right]^2 d\theta$$

$$-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\left[\frac{1+K}{1+K+\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}\exp\left(\frac{-K\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}{1+K+\frac{g_{QAM}}{2\sin^2\theta}\bar{\gamma}}\right)\right]^2 d\theta \quad (7.11b)$$

Review question 7.9 *List the CEP for popular modulation schemes viz. BPSK, DBPSK, MPSK, MQAM.*

Review question 7.10 *What are Craig's alternate forms of Q(x) and $Q^2(x)$?*

7.6 Summary

Figure 7.3 shows the chapter in a nutshell. We have defined coding and diversity gains. In this chapter, we have discussed about the space-time code design criteria viz., rank, determinant and trace criteria. Then we have studied about the Alamouti space-time code and its equivalence with the MRC. Finally we have calculated the SER for Alamouti space-time code over various i.i.d. fading channels.

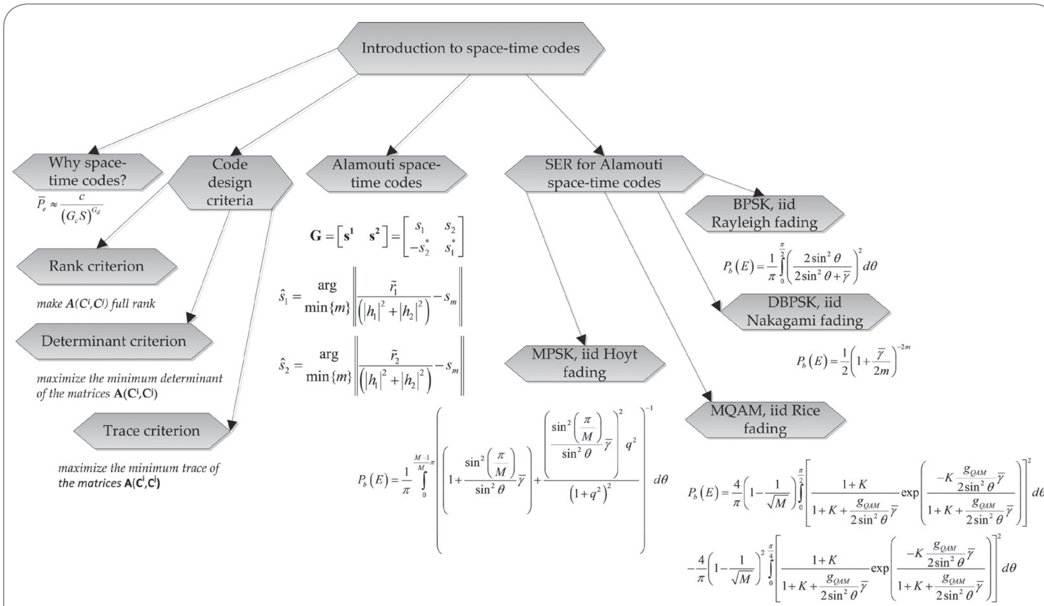


Fig. 7.3 Chapter 7 in a nutshell

Exercises

Exercise 7.1

Find the MGF of the received SNR γ of SISO system for various fading distributions:

(a) Rice
(b) Nakagami
(c) η-μ
(d) κ-μ

Exercise 7.2

Find the SER of Alamouti space-time code over i.i.d. κ-μ fading channel for various modulation schemes

(a) BPSK
(b) DBPSK
(c) MPSK
(d) MQAM

Exercise 7.3

Find the SER of Alamouti space-time code over i.i.d. η-μ fading channel for various modulation schemes

(a) BPSK
(b) DBPSK
(c) MPSK
(d) MQAM

Exercise 7.4

Explain the three code design criteria.

Exercise 7.5

How can one design codes which will maximize both rate and diversity gains?

References

1. Ahmed, B. and M. A. Matin. 2015. *Coding for MIMO OFDM in Future Wireless Systems*. London: Springer.
2. Alamouti, S. M. 1998. 'A simple transmit diversity technique for wireless communications'. *IEEE Journal on Selected Areas in Communications*. 16(8). 1451–1458.
3. Biglieri, E., R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor. 2007. *MIMO Wireless Communications*. Cambridge: Cambridge University Press.
4. Bose, R. 2008. *Information Theory, Coding and Cryptography*. New Delhi: Tata McGraw Hill.
5. Clerckx, B. and C. Oestges. 2013. *MIMO Wireless Networks*. Oxford: Elsevier.
6. Craig, J. 1991. 'A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations'. Proc. *IEEE MILCOM*. 25.5.1–25.5.5. Boston, MA.
7. Du, K. L. and M. N. S. Swamy. 2007. *Wireless Communication Systems*. New Delhi: Cambridge University Press.
8. Duman, T. M. and A. Ghrayeb. 2007. *Coding for MIMO Communication Systems*. Chichester: John Wiley and Sons.

9. Goldsmith, A. 2005. *Wireless Communications*. New Delhi: Cambridge University Press.
10. Hoffman, D. G., D. A. Leonard, C. C. Lindner, K. T. Phelps, C. A. Rodger, and J. R. Wall. 1991. *Coding Theory The Essentials*. New York: Marcel Dekker.
11. Jafarkhani, H. 2005. *Space-time Coding: Theory and Practice*. Cambridge: Cambridge University Press.
12. Jankiraman, M. 2004. *Space-time Codes and MIMO Systems*. Boston: Artech House.
13. Simon, M. K. and M.-S. Alouini. 2005. *Digital Communications over Fading Channels*. New Jersey: Wiley.
14. Sinnokrot, M. O. and V. K. Madisetti. 2010. 'Space-time block coding', in *The Digital Signal Processing Handbook*. V. K. Madisetti, Ed., Boca Raton, FL: CRC Press. 27-1–27-16.
15. Tse, D. and P. Viswanath. 2005. *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press.

Space-Time Block and Trellis Codes

CHAPTER 8

8.1 Introduction

We have discussed Alamouti space-time code in the last chapter. An extension of Alamouti space-time code for N_T number of transmitting antennas is also known as orthogonal space-time block codes (OSTBC). But OSTBC code rate tends to $\frac{1}{2}$ as N_T increases, which is a major concern for coding communities. There is another type of space-time code termed as space-time trellis code (STTC), which provides both coding and diversity gains. We will explore both of these space-time codes in this chapter. We will find the symbol error rate (SER) of coherent OSTBC over spatially correlated Rayleigh fading as well as i.i.d. Rayleigh fading MIMO channels. We will also consider the non-coherent space-time codes like Differential space-time codes in which CSIR is not available at the receiver. Pairwise error probability (PEP) calculation of space-time codes over correlated as well as i.i.d. Rayleigh fading will be also carried out. A brief introduction to space-time Turbo codes will be presented. In the recent past, researchers have concentrated more on STBC than STTC due to its high decoding complexity. In the last part of this chapter, we will consider some of the latest developments in STBC which are also known as Algebraic space-time codes which can achieve full rate and full diversity.

8.2 Space-time block codes

The Alamouti space-time code was for two transmitting antennas only. How do we generalize it for any number of transmitting antennas? V. Tarokh et al., (1999) extended the Alamouti space-time code to orthogonal space-time block codes (OSTBC) for any number of transmitting antennas based on the theory of orthogonal designs. The orthogonal property makes the decoding complexity minimal and we can decouple the symbols at the decoder. Let us now consider the STBC encoder as illustrated in Fig. 8.1. This is an extension of the Alamouti space-time code (which has two transmit antennas) to N_T transmit antennas.

Much of the characteristics of space-time block code (STBC) is specified by the generator matrix $\mathbf{G}$ (J. G. Proakis et al., 2005), having N_L rows and N_T columns, of the form.

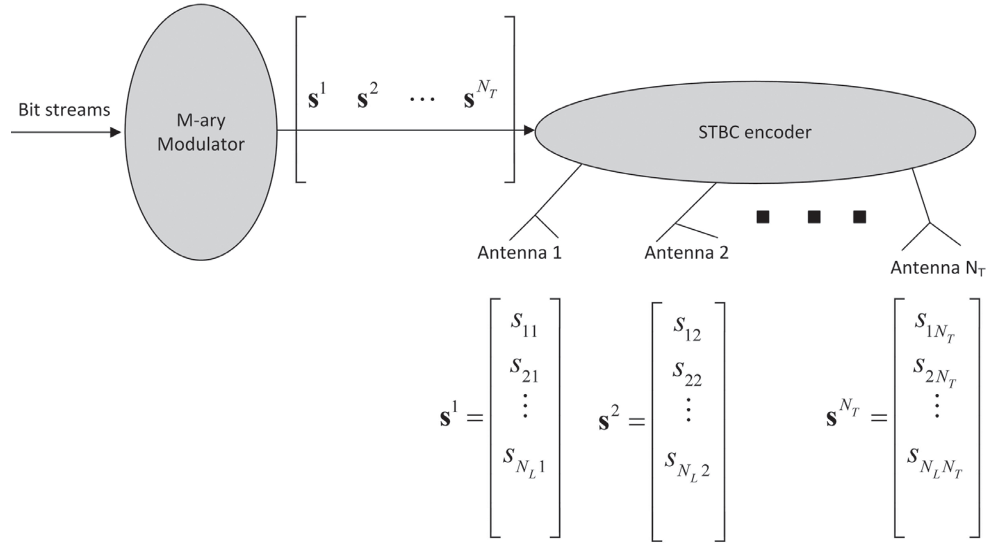


Fig. 8.1 Space-time block code encoder

$$\mathbf{G} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1N_T} \\ s_{21} & s_{22} & \dots & s_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N_L 1} & s_{N_L 2} & \dots & s_{N_L N_T} \end{bmatrix} \tag{8.1}$$

where the components $\{s_{ij}\}$ are M-ary symbols.

A space-time block code is described by the relationship between k-tuple input signal $s_1, s_2, \dots, s_k$ (usually k is taken as N_T) and set of signals to be sent from N_T antennas over N_L time slots/periods which is given by generator matrix. Note $\{s_{ij}\}$ are functions of k-tuple input signal $s_1, s_2, \dots, s_k$ and their complex conjugates. At time slot i, s_{ij} is sent from j antenna. Since k information bits are sent over N_L time intervals, the spatial rate of the code is $R = \frac{k}{N_L}$.

By using N_T transmit antennas, each row of **G** consisting of N_T signal points (symbols) is sent on the N_T antennas in a time slot. Therefore the first row of N_T symbols is sent on the N_T antennas in the first time slot, second row of N_T symbols is sent on the N_T antennas in the second time slot, and the N_L^{th} row of N_T symbols is sent on the N_T antennas in the N_L^{th} time slot. Hence N_L time slots are employed to send the symbols in the N_L rows of the generator matrix **G.** At the receiver one may employ any number of receive antennas and the design is immaterial of the number of receive antennas N_R. For example, S. M. Alamouti in 1998, proposed a STBC for $N_T = 2$ transmit antennas and $N_R = 1$ receive antenna. The generator matrix for Alamouti space-time code is

$$\mathbf{G} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$$

Thus the two symbols are sent on the double antennas in the two time slots. Hence the spatial code rate is R = 1 for the Alamouti space-time code. This is the maximum possible rate of an OSTBC for complex signals. We have observed that Alamouti space-time code achieves the maximum diversity and can be detected easily using maximum likelihood (ML) detector. These two desirable characteristics were achieved as a result of the orthogonality characteristic of the generator matrix **G** for the Alamouti code, which we may express as:

(a) The column vectors $\mathbf{s}^1 = \begin{bmatrix} s_1 \\ -s_2^* \end{bmatrix}$ and $\mathbf{s}^2 = \begin{bmatrix} s_2 \\ s_1^* \end{bmatrix}$ are orthogonal; i.e., $\left(\mathbf{s}^1\right)^H \bullet \mathbf{s}^2 = 0$ and

(b) $\mathbf{G}^H\mathbf{G} = \left[|s_1|^2 + |s_2|^2\right]\mathbf{I}_2$ where $\mathbf{I}_2$ is a 2×2 identity matrix

Thus, full diversity and low decoding complexity are achieved as a consequence of the orthogonality property of **G** given in the equation above. The Alamouti space-time code provides full diversity of two without channel state information (CSI) at the transmitter and an uncomplicated ML decoding system at the receiver. Such a system provides a guaranteed overall diversity gain of $2N_R$, without CSI at the transmitter as we have discussed before. Because of these, the scheme was extended to any number of transmit antennas by using the theory of orthogonal designs. The generalized schemes are known as orthogonal space-time block codes (OSTBC). These codes achieve full transmit diversity of $N_T N_R$ while allowing an uncomplicated ML decoding algorithm. The entries of the **G** are chosen such that they are linear combinations of $s_1, s_2, \ldots, s_{N_T}$ and their conjugates. The matrix itself is constructed based on orthogonal designs such that, $\mathbf{G}^H\mathbf{G} = \left[|s_1|^2 + |s_2|^2 + \cdots + |s_{N_T}|^2\right]\mathbf{I}_{N_T}$,

where, N_T is the number of transmit antennas, $\mathbf{G}^H$ is the Hermitian of **G**, and $\mathbf{I}_{N_T}$ is an $N_T \times N_T$ identity matrix.

An STBC is called OSTBC if

$$\mathbf{G}^H\mathbf{G} = \sum_{n=1}^{N_T} |s_n|^2 \, \mathbf{I} \tag{8.2}$$

OSTBC presumes that the channel coefficients remain the same over a period of N_T symbols, i.e., $h_{ij}(t) = h_{ij}$; $t = 1, 2, \ldots, N_T$. This block fading assumption is needed for uncomplicated linear decoding of OSTBC. We will also assume that the channel is frequency non-selective.

This approach yields full diversity. These code transmission matrices are cleverly constructed such that the rows and columns of each matrix are orthogonal to each other (i.e., the dot product of each column with another column is zero). If this condition is satisfied, the above equation will be satisfied, yielding the full transmit diversity. The orthogonality allows us to achieve full transmit diversity and enables receiver to decouple the signals transmitted from different antennas (T. M. Duman et al., 2007). Dependent on the type of signal constellation used, space-time block codes can be divided into OSTBC with real signals or OSTBC with complex signals.

Review question 8.1 *What are the two desirable characteristics of OSTBC?*

Review question 8.2 *What is code rate?*

8.2.1 OSTBC for real signals

Let us consider the generation of real transmission matrices. At the outset, it should be noted that orthogonality condition is crucial for our design, $\mathbf{G}^T\mathbf{G} = \left[|s_1|^2 + |s_2|^2 + \cdots + |s_{N_T}|^2\right]\mathbf{I}_{N_T}$. Let us consider square transmission matrices. It is difficult to find full rate and full diversity space-time codes for square generator matrices. Such matrices exist if the number of transmit antennas $N_T = 2$, 4, 8 (H. Jafarkhani, 2005) for real signals. The matrix being square, these codes are full rate ($R=1$), and also achieves full transmit diversity. Real orthogonal generator matrix designs provide a diversity order of $N_T N_R$. The transmission matrices for $N_T = 2$, 4, 8 are given by

$$\mathbf{G}_2 = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix} \tag{8.3a}$$

$$\mathbf{G}_4 = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 \\ s_2 & s_1 & s_4 & -s_3 \\ s_3 & -s_4 & s_1 & s_2 \\ s_4 & s_3 & -s_2 & s_1 \end{bmatrix} \tag{8.3b}$$

$$\mathbf{G}_8 = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 & -s_5 & -s_6 & -s_7 & -s_8 \\ s_2 & s_1 & -s_4 & s_3 & -s_6 & s_5 & s_8 & -s_7 \\ s_3 & s_4 & s_1 & -s_2 & -s_7 & -s_8 & s_5 & s_6 \\ s_4 & -s_3 & s_2 & s_1 & -s_8 & s_7 & -s_6 & s_5 \\ s_5 & s_6 & s_7 & s_8 & s_1 & -s_2 & -s_3 & -s_4 \\ s_6 & -s_5 & s_8 & -s_7 & s_2 & s_1 & s_4 & -s_3 \\ s_7 & -s_8 & -s_5 & s_6 & s_3 & -s_4 & s_1 & s_2 \\ s_8 & s_7 & -s_6 & -s_5 & s_4 & s_3 & -s_2 & s_1 \end{bmatrix} \tag{8.3c}$$

Hence, each column of an orthogonal design is a permutation with sign change of the first column. Just a simple check, see all the columns are orthogonal or not. This result in simple ML decoding by decoupling the decision for each transmitted symbol. The code rate for all these matrices is equal to one. We can also find the diversity gain for the given generator matrices (Y. S. Cho et al., 2010). For instance, for 2×2 generator matrix of

$$\mathbf{G}_2 = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix}$$

For two different space-time codewords, let us find the codeword difference matrix and codeword distance matrix, which are used for calculation of pairwise error probability between two codewords. Assume that the transmitted codeword is $\mathbf{S}_B$ but receiver decides in favour of $\mathbf{S}_A$(erroneous codeword). Then codeword difference matrix of $\mathbf{S}_A$ and $\mathbf{S}_B$ is

$$\mathbf{D}(\mathbf{S}_A, \mathbf{S}_B) = \begin{bmatrix} s_{A1} - s_{B1} & -s_{A2} + s_{B2} \\ s_{A2} - s_{B2} & s_{A1} - s_{B1} \end{bmatrix}$$

The codeword distance matrix of $\mathbf{S}_A$ and $\mathbf{S}_B$ is

$$\mathbf{A}(\mathbf{S}_A, \mathbf{S}_B) = \mathbf{D}(\mathbf{S}_A, \mathbf{S}_B)\left(\mathbf{D}(\mathbf{S}_A, \mathbf{S}_B)\right)^T$$

$$= \begin{bmatrix} (s_{A1} - s_{B1})^2 + (s_{A2} - s_{B2})^2 & 0 \\ 0 & (s_{A1} - s_{B1})^2 + (s_{A2} - s_{B2})^2 \end{bmatrix}$$

Rank of this matrix is 2. Hence it achieves diversity gain of $2N_R$. Similarly, we can show that generator matrices of $\mathbf{G}_4$ and $\mathbf{G}_8$ also achieve diversity gain of $4N_R$ and $8N_R$, respectively. But there is no coding gain. $G_d = (CGD)^{\frac{1}{r}} = \left(\left|\mathbf{A}(\mathbf{S}_A, \mathbf{S}_B)\right|\right)^{\frac{1}{2}} = \left((s_{A1} - s_{B1})^2 + (s_{A2} - s_{B2})^2\right)$, which is the same as the squared Euclidean distance of the symbols for uncoded system. Hence there is no coding gain. Assuming a single receiving antenna (two transmitting antennas) and we can check the decoding of this real OSTBC. For instance, let us consider generator matrix of $\mathbf{G}_2$. For this case, the received signal is

$$[y_1 \quad y_2] = \sqrt{\frac{E_S}{2N_0}}\begin{bmatrix} h_1 & h_2 \end{bmatrix}\begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix} + \begin{bmatrix} n_1 & n_2 \end{bmatrix}$$

In the above equation, y_1 and y_2 are the signals received by the receiving antenna over the first and second time intervals.

Note that we have normalized the power by introducing a term of the form $\sqrt{\frac{E_s}{N_T N_0}}$ explicitly. If this term is not included, then we have to consider this power division separately when we calculate the received SNR. The above equation can be expressed in equivalent form as

$$\begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = \sqrt{\frac{E_s}{2N_0}}\begin{vmatrix} h_1 & h_2 \\ h_2 & -h_1 \end{vmatrix}\begin{vmatrix} s_1 \\ s_2 \end{vmatrix} + \begin{vmatrix} n_1 \\ n_2 \end{vmatrix}$$

$$\Rightarrow \qquad \mathbf{y} = \sqrt{\frac{E_s}{2N_0}}\mathbf{Hs} + \mathbf{n}$$

Note that columns of $\mathbf{H}$ matrix are orthogonal; therefore,

$$\tilde{\mathbf{y}} = \mathbf{H}^T\mathbf{Y} = \sqrt{\frac{E_s}{2N_0}}\mathbf{H}^T\mathbf{Hs} + \mathbf{H}^T\mathbf{n}$$

$$= \sqrt{\frac{E_s}{2N_0}}\begin{bmatrix} h_1^2 + h_2^2 & 0 \\ 0 & h_1^2 + h_2^2 \end{bmatrix}\mathbf{s} + \tilde{\mathbf{n}} \tag{8.4a}$$

Hence, the ML detection is

$$\hat{s}_j = \underset{\min\{m\}}{\arg}\left\|\frac{\tilde{y}_j}{\sqrt{\frac{E_S}{2N_0}}\left(h_1^2 + h_2^2\right)} - s_m\right\|; j = 1, 2 \tag{8.4b}$$

where, $s_m \in \{s_k\}_{k=1}^{M}$ is one of the M-ary pulse amplitude modulation (MPAM) symbols.

Note that in the above equation we are estimating symbols s_1 and s_2 which we have transmitted from the two transmitting antennas over two time intervals. It could be any one of the MPAM signals. Similar detection process could be carried out for other OSTBC as well both for real and complex signals.

Review question 8.3 *Can OSTBC for real signal constellations achieve unity code rate for $N_T > 2$?*

8.2.2 OSTBC for complex signal constellations

The Alamouti scheme is one such matrix with complex entries for double-transmit antennas. This scheme gives the full diversity of $2N_R$ with a full code rate of 1. It has been shown in the literature (H. Jafarkhani, 2005 and V. Tarokh et al., 1999) that orthogonal complex designs with R = 1 do not exist for $N_T > 2$ transmit antennas. However by reducing the code rate, it is possible to devise complex orthogonal designs for 2-D signal constellations. For example, an orthogonal generator matrix for a STBC that transmits four complex-valued Phase Shift Keying (PSK) or Quadrature Amplitude Modulation (QAM) symbols on $N_T = 4$ transmit antennas is given below. For this code generator, the four complex-valued symbols are transmitted in eight consecutive time slots.

$$\mathbf{G}_4 = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_2 & s_1 & -s_4 & s_3 \\ -s_3 & s_4 & s_1 & -s_2 \\ -s_4 & -s_3 & s_2 & s_1 \\ s_1^* & s_2^* & s_3^* & s_4^* \\ -s_2^* & s_1^* & -s_4^* & s_3^* \\ -s_3^* & s_4^* & s_1^* & -s_2^* \\ -s_4^* & -s_3^* & s_2^* & s_1^* \end{bmatrix} \tag{8.5}$$

Hence the spatial rate for this code is $R = \frac{1}{2}$. We also observe that $\mathbf{G}^H\mathbf{G} = c\left[|s_1|^2 + |s_2|^2 + |s_3|^2 + |s_4|^2\right]\mathbf{I}_4$,

where, c is a constant. So this code provides fourth order diversity in the case of one receive antenna and $4N_R$ diversity with N_R receive antennas. Diversity and coding gain could be obtained from the rank and determinant of codeword distance matrix as we have done for Alamouti space-time code and OSTBC for real signals for $N_T = 2$. The detection analysis could be carried out similar to the case of OSTBC for real signals.

8.2.3 Symbol error rate (SER) for OSTBC over spatially correlated Rayleigh fading MIMO channel

Consider the I–O relation of a coherent OSTBC based MIMO system as

$$\mathbf{Y} = \mathbf{HC}_{\text{OSTBC}} + \mathbf{N}$$

where, $\mathbf{Y}$ is an $N_R \times N_L$ matrix, $\mathbf{H}$ is an $N_R \times N_T$ channel matrix and $\mathbf{C}_{OSTBC}$ is an $N_T \times N_L$ OSTBC encoded transmission matrix and $\mathbf{N}$ is $N_R \times N_L$ matrix with i.i.d. complex circular Gaussian random variables (RVs) with each components distributed as $N_c(0, \sigma^2)$.

Assume L symbols are transmitted over T symbol periods with code rate given by $R = \frac{L}{T}$. Since the transmission matrix of OSTBC $\mathbf{C}_{OSTBC}$ is orthogonal $(\mathbf{CC}^H = c\left(|s_1|^2 + |s_2|^2 + \ldots + |s_L|^2\right)\mathbf{I}_{N_T})$ where, c is a constant dependent on the transmission matrix of OSTBC $\mathbf{C}_{OSTBC}$. For instance, for Alamouti space-time code, $c = 1$. Due to orthogonality of the OSTBC encoded transmission matrix, all the signals are decoupled at the receiver. Hence, one can write

$$r_k = \left[\sum_{i=1}^{N_T}\sum_{j=1}^{N_R}|h_i|^2\right]s_k + \tilde{n}_k$$

$$= \left[c\|\mathbf{H}\|^2\right]s_k + \tilde{n}_k; k = 1,2,\ldots,N_L \quad (8.6a)$$

We can estimate the transmitted symbol by using maximum likelihood (ML) detection as usual.

$$\hat{s}_k = \underset{\min s \in S}{\arg}\left|r_k - \left[c\|\mathbf{H}\|^2\right]s\right|^2 \quad (8.6b)$$

where, s is a symbol and S is the symbol alphabet.

Like in Alamouti space-time code, we can calculate the effective signal to noise ratio (SNR) as

$$\gamma = \frac{E_s\|\mathbf{H}\|^2}{N_T\sigma^2} = \frac{P_0 T\|\mathbf{H}\|^2}{N_T\sigma^2} = \frac{P_0\|\mathbf{H}\|^2 T}{N_T\sigma^2} = \frac{\overline{\gamma}\|\mathbf{H}\|^2 T}{N_T} = \frac{\overline{\gamma}\|\mathbf{H}\|^2}{N_T R} \quad (8.6c)$$

The Moment Generating Function (MGF) of effective SNR is derived in (K. S. Ahn et al., 2007)

$$\mathrm{M}_\gamma(s) = \prod_{i=1}^{R_D}\left(\frac{1}{1 - \lambda_i\frac{\overline{\gamma}sT}{N_T}}\right)^{n_i} = \frac{1}{\left(1 - \lambda_1\frac{\overline{\gamma}sT}{N_T}\right)^{n_1}\left(1 - \lambda_2\frac{\overline{\gamma}sT}{N_T}\right)^{n_2}\cdots\left(1 - \lambda_{R_D}\frac{\overline{\gamma}sT}{N_T}\right)^{n_{R_D}}} \quad (8.6d)$$

where, R_D is the number of distinct eigenvalues λ_i of correlation matrix $\mathbf{R}(\mathbf{R} = E(vec(\mathbf{H})vec(\mathbf{H})^H) = \mathbf{R}_T^T \otimes \mathbf{R}_R$ with SVD $\mathbf{R} = \mathbf{U\Sigma V^H}$).

Note that n_i is the set n_k indices corresponding to λ_k. Therefore, $\sum_{j=1}^{R_D} n_i = R_H$ where R_H is the rank of $\mathbf{R}$ matrix.

We can rewrite MGF as

$$\mathrm{M}_\gamma(s) = \prod_{i=1}^{R_D}\left(\frac{1}{1 - s\Psi_i}\right)^{n_i}; \lambda i\frac{\overline{\gamma}T}{N_T} = \Psi_i$$

By partial fraction expansion (E. B. Saff, 2003), MGF can be expressed as

$$\mathrm{M}_\gamma(s) = \sum_{i=1}^{R_D}\sum_{j=1}^{n_i}\frac{A_{ij}}{\left(1 - s\Psi_1\right)^j}$$

$$= \sum_{j=1}^{n_1}\frac{A_{1j}}{\left(1 - s\Psi_1\right)^j} + \sum_{j=1}^{n_2}\frac{A_{2j}}{\left(1 - s\Psi_2\right)^j} + \ldots + \sum_{j=1}^{n_{R_D}}\frac{A_{R_D j}}{\left(1 - s\Psi_{n_{RD}}\right)^j}$$

In the above equation we can find the coefficients A_{ij} using the Heaviside formulae (E. Kreyszig, 1999)

$$A_{ij} = \left(\Psi_i\right)^i \frac{1}{\left(n_i - j\right)!} \lim_{s \to 0} \frac{d^{n_i - j}}{ds^{n_i - j}}\left[(s)^{n_i} M_\gamma\left(s - \frac{1}{\Psi_i}\right)\right]$$

The inverse Laplace transform of the MGF gives probability density function (pdf) of γ as

$$P_\gamma(\gamma) = \sum_{i=1}^{R_D}\sum_{j=1}^{n_i}\frac{A_{ij}}{\Gamma(j)}\left(\frac{N_T R}{\overline{\gamma}\lambda_i}\right)^j \gamma^{j-1} e^{-\frac{N_T R}{\overline{\gamma}\lambda_i}\gamma}$$

Here scalar R is equal to $\frac{1}{T}$ and $\Gamma(j)$ is the gamma function. The conditional error probability (CEP) for MPAM is given by

$$P_b(E|\gamma) = a_{PAM} Q\left(\sqrt{g_{PAM}\gamma}\right); a_{PAM} = 2\left(1 - \frac{1}{M}\right); g_{PAM} = \frac{6}{M^2 - 1}$$

For M-ary phase shift keying (M-PSK) (J. G. Proakis et al., 2005), CEP is given by

$$P_b(E|\gamma) = a_{PSK} Q\left(\sqrt{g_{PSK}\gamma}\right); a_{PSK} = 2, g_{PSK} = 2\sin^2\left(\frac{\pi}{M}\right)$$

For M-ary quadrature amplitude modulation (M-QAM), the CEP is given as

$$P_b(E|\gamma) = 2a_{QAM} Q\left(\sqrt{g_{QAM}\gamma}\right) - a_{QAM}^2 Q^2\left(\sqrt{g_{QAM}\gamma}\right);$$

$$a_{\mathrm{QAM}} = 2\left(1 - \frac{1}{\sqrt{M}}\right); g_{QAM} = \frac{3}{M - 1}$$

The symbol error rate (SER) could be obtained by averaging the CEP over the pdf of γ. For MPAM, SER is

$$P_b(E) = \int_0^\infty P_b\left(E|\gamma\right) p_\gamma\left(\gamma\right) d\gamma = \int_0^\infty a_{PAM} Q\left(\sqrt{g_{PAM}\gamma}\right) p_\gamma\left(\gamma\right) d\gamma$$

Taking the constant a_{PAM} outside the integral, we have,

$$P_b(E) = a_{PAM}\int_0^\infty Q\left(\sqrt{g_{PAM}\gamma}\right) p_\gamma\left(\gamma\right) d\gamma$$

Integrating over the pdf of γ, we have,

$$P_b(E) = a_{PAM}\int_0^{\infty} Q\left(\sqrt{g_{PAM}\gamma}\right)\sum_{i=1}^{R_D}\sum_{j=1}^{n_i}\frac{A_{ij}}{\Gamma(j)}\left(\frac{N_T R}{\overline{\gamma}\lambda_i}\right)^j \gamma^{j-1} e^{-\frac{N_T R}{\overline{\gamma}\lambda_i}\gamma} d\gamma$$

Using new and compact notation, $q_i = \dfrac{N_T R}{\overline{\gamma}\lambda_i} = \dfrac{1}{\psi_i}$, we have,

$$P_b(E) = a_{PAM}\int_0^{\infty} Q\left(\sqrt{g_{PAM}\gamma}\right)\sum_{i=1}^{R_D}\sum_{j=1}^{n_i}\frac{A_{ij}}{\Gamma(j)}(q_i)^j \gamma^{j-1} e^{-q_i\gamma} d\gamma$$

Taking out the two summations outside the integral, keeping only those terms which are dependent on γ inside the integral, we have,

$$P_b(E) = a_{PAM}\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_{ij}\frac{(q_i)^j}{\Gamma(j)}\int_0^{\infty} Q\left(\sqrt{g_{PAM}\gamma}\right)\gamma^{j-1} e^{-q_i\gamma} d\gamma$$

Applying Chernoff's bound,

$$Q\left(\sqrt{g_{PAM}\gamma}\right) \le \frac{1}{2}\exp\left(-\frac{g_{PAM}\gamma}{2}\right)$$

$$\Rightarrow \quad P_b(E) \le a_{PAM}\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_{ij} I_1\left(g_{PAM}, q_i, j\right); a_{PAM} = 2\left(1-\frac{1}{M}\right), g_{PAM} = \frac{6}{M^2-1}; q_i = \frac{N_T R}{\overline{\gamma}\lambda_i} \quad (8.6e)$$

where, $I_1\left(g_{PAM}, q_i, j\right) = \dfrac{1}{2}\displaystyle\int_0^{\infty}\exp\left(-\frac{g_{PAM}+2q_i}{2}\gamma\right)\gamma^{j-1}d\gamma$

Similar expression for Chernoff bound of SER for M-PSK

$$P_b(E) \le a_{PSK}\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_{ij} I_1\left(g_{PSK}, q_i, j\right); a_{PSK} = 2, g_{PSK} = 2\sin^2\left(\frac{\pi}{M}\right); q_i = \frac{N_T R}{\overline{\gamma}\lambda_i} \quad (8.6f)$$

where, $I_1\left(g_{PSK}, q_i, j\right) = \dfrac{1}{2}\displaystyle\int_0^{\infty}\exp\left(-\frac{g_{PSK}+2q_i}{2}\gamma\right)\gamma^{j-1}d\gamma$

The approximate BER can be obtained as $\dfrac{P_b(E)}{\log_2 M}$.

Similarly, for M-QAM, the SER is given as

$$\begin{aligned} P_b(E) &= \int_0^{\infty} P_b\left(E|\gamma\right) p_\gamma(\gamma)\, d\gamma \\ &= \int_0^{\infty}\left(2a_{QAM}Q\left(\sqrt{g_{QAM}\gamma}\right) - a_{QAM}^2 Q^2\left(\sqrt{g_{QAM}\gamma}\right)\right) p_\gamma(\gamma)\, d\gamma \end{aligned}$$

Taking the constant a_{QAM} outside the integral, we have,

$$P_b(E) = 2a_{QAM}\int_0^{\infty} Q\left(\sqrt{g_{QAM}\gamma}\right) p_\gamma(\gamma)\, d\gamma - a_{QAM}^2\int_0^{\infty} Q^2\left(\sqrt{g_{QAM}\gamma}\right) p_\gamma(\gamma)\, d\gamma$$

Integrating over the pdf of γ, we have,

$$P_b(E) = 2a_{QAM}\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_j \frac{(q_i)^j}{\Gamma(j)}\int_0^\infty Q\left(\sqrt{g_{QAM}\gamma}\right)\gamma^{j-1}e^{-q_i\gamma}d\gamma - a_{QAM}^2\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_j \frac{(q_i)^j}{\Gamma(j)}\int_0^\infty Q^2\left(\sqrt{g_{QAM}\gamma}\right)\gamma^{j-1}e^{-q_i\gamma}d\gamma$$

Applying Chernoff's bound,

$$Q\left(\sqrt{g_{QAM}\gamma}\right) \le \frac{1}{2}\exp\left(-\frac{g_{QAM}\gamma}{2}\right); Q^2\left(\sqrt{g_{QAM}\gamma}\right) \le \frac{1}{4}\exp\left(-\frac{g_{QAM}\gamma}{2}\right)$$

$$\Rightarrow \qquad P_b(E) \le \left(2a_{QAM} - \frac{a_{QAM}^2}{2}\right)\sum_{i=1}^{R_D}\sum_{j=1}^{n_i} A_{ij} I_1\left(g_{QAM}, q_i, j\right) \tag{8.6g}$$

where, $a_{QAM} = 2\left(1 - \frac{1}{\sqrt{M}}\right), g_{QAM} = \frac{3}{(M-1)}; q_i = \frac{N_T R}{\bar{\gamma}\lambda_i}$

$$I_1\left(g_{QAM}, q_i, j\right) = \frac{1}{2}\int_0^\infty \exp\left(-\frac{g_{QAM} + 2q_i}{2}\gamma\right)\gamma^{j-1}d\gamma$$

M. Kulkarni et al., (2014) also evaluated the performance of OSTBC in equi-correlated Rayleigh fading MIMO channel.

Review question 8.4 *Write down the SER bound of OSTBC over correlated Rayleigh fading MIMO channel.*

8.2.4 Differential OSTBC

For non-coherent STBC (chapter 7), we can apply differential STBC so that we can estimate the transmitted symbol without CSIR. Consider I–O model of a $N_T \times N_R$ MIMO system. The received signal at time t can be expressed as

$$\mathbf{Y}_t = \mathbf{H}\mathbf{X}_t + \mathbf{N}_t$$

where, $\mathbf{Y}_t$ is an $N_R \times N_L$ matrix, $\mathbf{H}$ is an $N_R \times N_T$ channel matrix (for frequency flat Rayleigh fading, elements of $\mathbf{H}$ are i.i.d. CSCG distributed as $N_C(0,1)$) and $\mathbf{X}_t$ is an $N_T \times N_L$ OSTBC encoded transmission matrix at time t and $\mathbf{N}_t$ is $N_R \times N_L$ matrix with i.i.d. complex circular Gaussian random variables (RVs) with each components distributed as $N_c(0, \sigma^2)$.

Unitary constellations

Let S denote the symbol alphabet from a unitary constellation ($\forall s_j \in S, |s_j|^2 = 1$), for instance, BPSK, QPSK, M-PSK. Assume $\{s_j\}_{j=1}^{p}$ be a block of symbols to be sent at a time t. Then define

$$\mathbf{U}_t = \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\left(\mathbf{V}_j s_j^{real} + i\mathbf{W}_j s_j^{imag}\right), i = \sqrt{-1}$$

where, $\mathbf{V}_j$, $\mathbf{W}_j$ satisfy the amicable orthogonal designs as mentioned in G. Ganesan et al. (2002). Then we can show that $\mathbf{U}_t\mathbf{U}_t^H = \mathbf{I}_{N_T \times N_T}$ which means that $\mathbf{U}_t$ is a unitary matrix.

Differential modulation and detection

Assume we initially transmit $\mathbf{X}_0 = \mathbf{I}_{N_T \times N_T}$. At time t, we may transmit $\mathbf{X}_t = \mathbf{X}_{t-1}\mathbf{U}_t$. The received signal matrix at time t can be expressed as

$$\mathbf{Y}_t = \mathbf{H}\mathbf{X}_t + \mathbf{N}_t = \mathbf{H}\mathbf{X}_{t-1}\mathbf{U}_t + \mathbf{N}_t$$

Note CSIR is not available, hence $\mathbf{H}\mathbf{X}_{t-1}$ is not known at the receiver. But the received signal at time *t-1* was

$$\mathbf{Y}_{t-1} = \mathbf{H}\mathbf{X}_{t-1} + \mathbf{N}_{t-1}$$

Since $\mathbf{N}_{t-1}$ is a Gaussian white noise, $\mathbf{Y}_{t-1}$ can be taken as the ML estimate of $\mathbf{H}\mathbf{X}_{t-1}$. Substituting this, we have,

$$\mathbf{Y}_t = \mathbf{H}\mathbf{X}_{t-1}\mathbf{U}_t + \mathbf{N}_t \cong \mathbf{Y}_{t-1}\mathbf{U}_t + \mathbf{N}_t$$

Hence the ML detection could be carried out without CSIR

$$\left\{\hat{s}_j\right\}_{j=1}^{p} = \arg \max_{\{s_j\},\, s_j \in S} \; trace\left\{\left(\mathbf{Y}_t - \mathbf{Y}_{t-1}\mathbf{U}_t\right)^H \left(\mathbf{Y}_t - \mathbf{Y}_{t-1}\mathbf{U}_t\right)\right\}$$

which can be further simplified as

$$\hat{s}_j = \arg \max_{s_j \in S} \; \mathrm{Re}\,al\left\{trace\left\{\left(\mathbf{Y}_t^H\mathbf{Y}_{t-1}\mathbf{V}_j\right)s_j^{real}\right\}\right\} + \mathrm{Re}\,al\left\{trace\left\{\left(\mathbf{Y}_t^H\mathbf{Y}_{t-1}\mathbf{W}_j\right)s_j^{imag}\right\}\right\}$$

Review question 8.5 *Explain the ML decoding of Differential STBC.*

8.2.5 Other STBCs

We have considered here OSTBC only. There are many other STBCs as well. A book which is completely devoted to STBC only is of E. G. Larsson et al., 2003. Unitary space-time code is discussed in T. M. Duman et al., 2007. Non-orthogonal STBC are explained in H. Jafarkhani (2005). Many quasi-orthogonal STBCs are proposed in C. Yuen et al., 2007. The maximal rate of a complex orthogonal STBC with N_T transmit antennas (X. B. Liang, 2003) is given by

$$r_{\max} = \frac{\left\lceil \frac{N_T}{2} \right\rceil + 1}{2\left\lceil \frac{N_T}{2} \right\rceil}$$

Hence, r_{max} tends to be half as the number of transmit antennas increases, which is a major concern with coding communities. Many STBCs have come up to overcome this. Quasi-orthogonal STBC compromise the orthogonality constraint to empower rate-one transmission, sacrificing the increment in decoding complexity. We will consider full diversity and full rate codes in the last section of this chapter. One of the recent developments is the distributed space-time coding (Y. Jing, 2013) for cooperative relay networks.

Review question 8.6 *What is the maximal code rate for OSTBC for complex signal constellations for $N_T > 2$?*

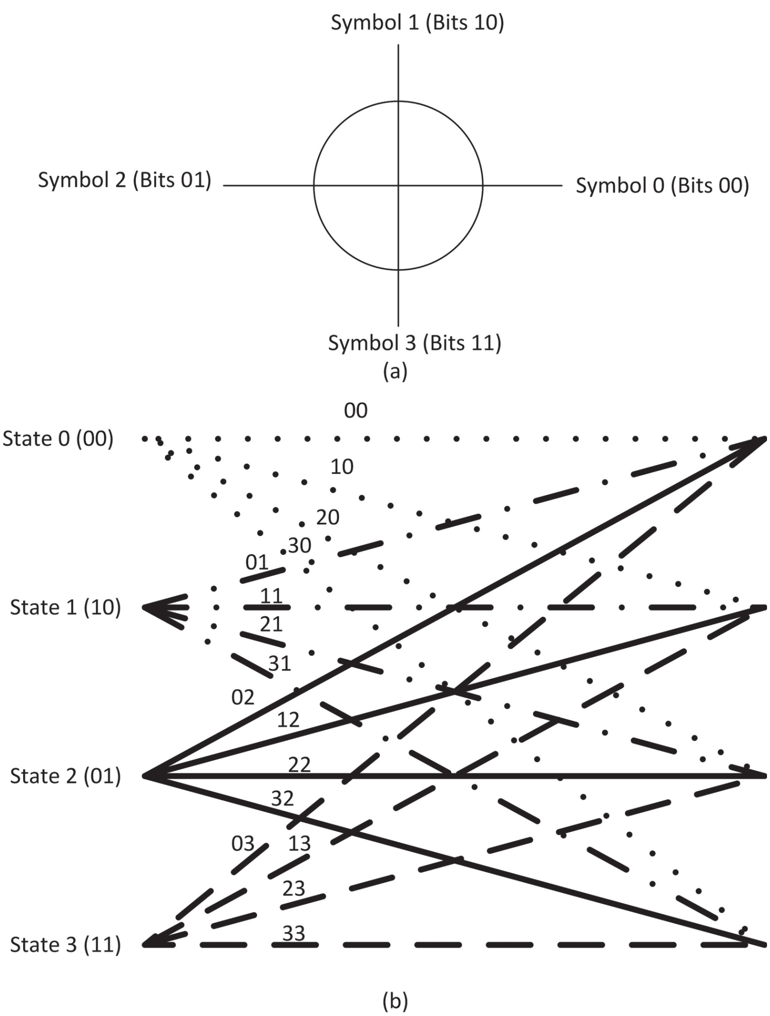


Fig. 8.2 (a) Quadrature phase shift keying (QPSK) constellation diagram ($\Xi(0)\rightarrow e^{0}=1$; $\Xi(1)\rightarrow e^{j\Pi/2}=j$; $\Xi(2)\rightarrow e^{2j\Pi/2}=-1$; $\Xi(3)\rightarrow e^{3j\Pi/2}=-j$ where Ξ is the M-ary mapping function) (one could also do Gray coding for QPSK) (b) Trellis diagram of a QPSK, four-state trellis code for $N_T = 2$ with a rate of 2bps/Hz

8.3 Space-time trellis codes

8.3.1 STTC trellis diagram

STTC is an extension of conventional trellis codes to multi-antenna systems. These codes may be designed to extract the diversity gain and coding gain (V. Kuhn, 2006). Each STTC can be described using a trellis. The literal meaning of "trellis" is "a light frame made of long narrow pipes of wood that cross each other, used to support climbing plants". Note the similarity of trellis with the trellis diagram such as depicted in Fig. 8.2. The number of nodes in the trellis diagram corresponds to the number of states in the trellis. The N_T outputs of each branch correspond to the symbols to be transmitted from N_T antennas. Figure 8.2 shows trellis diagram for a simple quadrature phase shift keying (QPSK), four state trellis code for $N_T = 2$ with rate 2bps/Hz.

Table 8.1

Output symbols for different input symbols and states

Input States	0	1	2	3
0	00	10	20	30
1	01	11	21	31
2	02	12	22	32
3	03	13	23	33

Input: 1 2 3 1 0 0 1
Antenna 1: 1 2 3 1 0 0 1
Antenna 2: 0 1 2 3 1 0 0

Let us find out how to interpret the trellis of this code (C. Oestges et al., 2007).

On the numero uno node (i.e., the number one state) the achievable outputs are 00, 10, 20 and 30. It implies that if the input symbol are 0, 1, 2 or 3, the output symbols are correspondingly 00 (0 on antenna 2 and 0 on antenna 1), 10 (1 on antenna 2 and 0 on antenna 1), 20 (2 on antenna 2 and 0 on antenna 1) or 30 (3 on antenna 2 and 0 on antenna 1). Furthermore, the ensuing state will be respectively, 0, 1, 2 or 3. The trellis has four nodes corresponding to four states**.** The states are denoted as, $S_t = 0, 1, 2, 3$. The input to the encoder is a pair of bits (00, 10, 01, 11) which are mapped to the corresponding phases that are numbered (0, 1, 2, 3), respectively. The indices 0, 1, 2, 3 correspond to the four phases, which are called symbols. Initially, the encoder is in state $S_t = 0$. Then for each pair of input bits, which are mapped into a corresponding symbol, the encoder generates a pair of symbol, the first of which is transmitted on the second antenna and the second symbol is sent concurrently on the first antenna.

For example, when the encoder is in state $S_t = 0$ and the input bits are 11, the symbol is a 3. The STTC outputs the pair of symbols (3,0), corresponding to the phases $3\pi/2$ and 0 (corresponding signals $\Xi(3)\rightarrow e^{3j\Pi/2}=-j$ and $\Xi(0)\rightarrow e^{0}=1$ where Ξ is the M-ary mapping function). The signal -j is transmitted in the second antenna and 1 signal is transmitted on the first antenna. At this point encoder goes to state $S_t = 3$. If the next two input bits are 01, the encoder outputs the symbols (2,3) which are transmitted on the two antennas (second antenna transmits -1 and first antenna transmits -j). Thus the encoder goes to state $S_t = 2$, and this process continues. At the end of a block of input bits, say a frame of data, zeros are inserted in the data stream to return the encoder to the state $S_t = 0$. Thus the STTC transmits at a bit rate of 2bps/Hz.

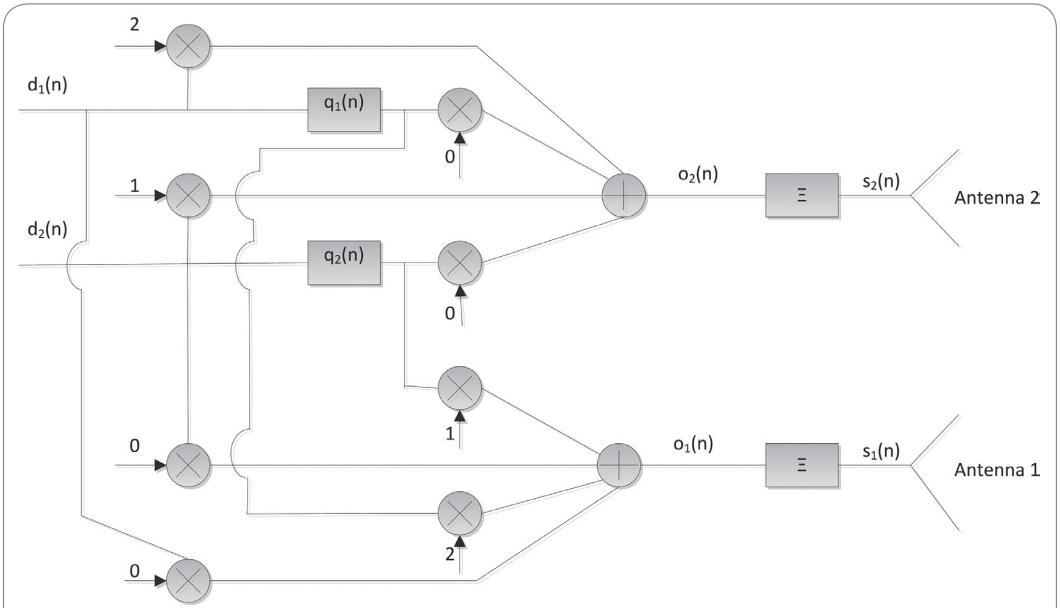


Fig. 8.3 Space-time trellis encoder (m=s=2) for trellis diagram of Fig. 8.2

8.3.2 STTC encoder

The delay diversity scheme was started by A. Wittneben (1993). It is the simplest STTC and illustrates the concept of STTC well. We will denote the generator matrix for the above STTC by $\mathbf{W}(M, S, N_T)$ where, M signifies M-ary modulation scheme, S is the number of states in the trellis diagram and N_T is the number of transmitting antennas. Each STTC will have a unique generator matrix. The generator matrix will have N_T columns and $m + s$ rows ($m = \log_2 M$ and s is the number of memory elements in the encoder). Each entry is being a number between 0 to M-1. The generator matrix for Wittneben Delay diversity is given by

$$\mathbf{W}(4,4,2)^T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} \tag{8.7a}$$

The transmit antennas send delayed version of the message bits. We are considering QPSK, trellis with four states and number of transmitting antennas equal to 2 as depicted in Fig. 8.3. From the input bits and present state of the trellis, we can find the transmitted symbols from the transmitting antennas. Note that $\log_2(M)$ bits (m = 2 bits for QPSK) are fed into the shift register (SR) at a time.

$$\Xi\left[\left(\begin{bmatrix} Input & bits & State & bits \\ d_2(n) & d_1(n) & q_2(n) & q_1(n) \end{bmatrix}[\mathbf{G}]^T\right)\bmod(M)\right]$$

$$= \Xi\begin{bmatrix} Output & bits \\ o_2(n) & o_1(n) \end{bmatrix} = \begin{bmatrix} s_2(n) & s_1(n) \end{bmatrix}; \Xi \quad M-ary \quad mapping \quad function \tag{8.7b}$$

G is the generator matrix. The output bits will be mapped to M-ary symbols using the M-ary mapping function Ξ which maps integer values to the M-ary symbols. For instance for M-PSK constellation, Ξ (i) = exp(2πji/M) where j is the √(-1) and i is an integer between 0 and M-1. It will be clear from the following example.

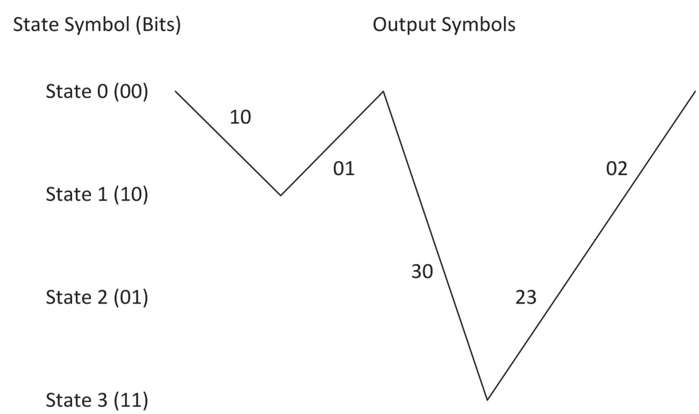


Fig. 8.4 Trellis path corresponding to this input bits stream 01110010

Example 8.1

Let us revisit the space-time trellis code of Fig. 8.2. Let us assume that the input bit stream for this code is 01110010. Figure 8.4 above shows the trellis path corresponding to this input bits stream. Note that we have to add 00 at the end to guarantee that the state-machine return to state zero. Using the generator matrix $\mathbf{G}^T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$, find the set of transmitted symbols.

Solution

First two inputs are 10 and state is 00 (State 0). Hence, the output can be obtained as

$$\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

Therefore, $\Xi(1) = e^{j\Pi/2} = j$ and $\Xi(0) = e^{j0\Pi/2} = 1$ are sent at time t = 1 from the second and first antennas, respectively.

Now the next state is 10 (State 1). Next two input bits are 00, hence, the outputs are

$$\begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}\begin{bmatrix}1 & 0\\ 2 & 0\\ 0 & 1\\ 0 & 2\end{bmatrix} = (0 \quad 1)$$

Therefore, $\Xi(0) = e^{j0\Pi/2} = 1$ and $\Xi(1) = e^{j1\Pi/2} = \mathrm{j}$ are sent at time $t = 2$ from the second and first antennas, respectively. Now the next state is 00 (State 0). Next two input bits are 11, hence, the outputs are

$$\begin{pmatrix}1 & 1 & 0 & 0\end{pmatrix}\begin{bmatrix}1 & 0\\ 2 & 0\\ 0 & 1\\ 0 & 2\end{bmatrix} = (3 \quad 0)$$

Therefore, $\Xi(3) = e^{j3\Pi/2} = -j$ and $\Xi(0) = e^{j0\Pi/2} = 1$ are sent at time $t = 3$ from the second and first antennas, respectively. Now the next state is 11 (State 3). Next two input bits are 01, hence, the outputs are

$$\begin{pmatrix}0 & 1 & 1 & 1\end{pmatrix}\begin{bmatrix}1 & 0\\ 2 & 0\\ 0 & 1\\ 0 & 2\end{bmatrix} = (2 \quad 3)$$

Therefore, $\Xi(2) = e^{j2\Pi/2} = -1$ and $\Xi(3) = e^{j3\Pi/2} = -j$ are sent at time $t = 4$ from the second and first antennas respectively. Now the next state is 01 (State 2). Next two input bits are 00, hence, the outputs are

$$\begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}\begin{bmatrix}1 & 0\\ 2 & 0\\ 0 & 1\\ 0 & 2\end{bmatrix} = (0 \quad 2)$$

Therefore, $\Xi(0) = e^{j0\Pi/2} = 1$ and $\Xi(2) = e^{j2\Pi/2} = -1$ are sent at time $t = 5$ from the second and first antennas, respectively.

8.3.3 Rank and coding gain distance (CGD) calculations

The minimum value of CGD among all possible pairs of codewords is an indication of the performance of the code. In the case of STTC, any valid codeword starts from state zero and ends at state zero. Since the common branches between $\mathbf{C}^1$ and $\mathbf{C}^2$ do not contribute to the CGD, the minimum CGD may correspond to the determinant for any pair of paths diverging from a state and merging to the same state after P transitions. For two different space-time codewords, let us find the codeword difference matrix and codeword distance matrix which are used for calculation of pairwise error probability between two codewords.

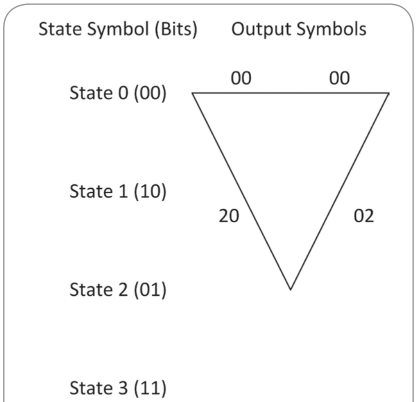


Fig. 8.5 Example 8.2 trellis path

Example 8.2

Figure 8.5 shows an example of P = 2 transitions for the STTC of Fig. 8.2. Note that the generator matrix for this STTC is $\mathbf{G}^T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$. The first path stays in state zero during both transitions that is 000. The second path goes to state 2 in the first transition and merges to state zero in the second transition that is 020. Find the rank and CGD for these two trellis paths.

Solution

The corresponding codewords for the two different trellis paths after doing M-ary mapping Ξ for QPSK shown in Fig. 8.5 are

$$\mathbf{C}^1 = \begin{pmatrix} \Xi(0) = e^{j0} = 1 & \Xi(0) = e^{j0} = 1 \\ \Xi(0) = e^{j0} = 1 & \Xi(0) = e^{j0} = 1 \end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix} \Xi(2) = e^{j2\frac{\pi}{2}} = -1 & \Xi(0) = e^{j0} = 1 \\ \Xi(0) = e^{j0} = 1 & \Xi(2) = e^{j2\frac{\pi}{2}} = -1 \end{pmatrix}$$

Assume that the transmitted codeword is $\mathbf{C}^2$ but receiver decides in favour of $\mathbf{C}^1$ (erroneous codeword). Then codeword difference and distance matrix of $\mathbf{C}^1$ and $\mathbf{C}^2$ and CGD can be calculated as

$$\mathbf{D}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{C}^1 - \mathbf{C}^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right)\left(\mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right)\right)^T = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

$$CGD = \left|\mathbf{A}\left(\mathbf{C}^1,\mathbf{C}^2\right)\right| = 16$$

Note that the codeword distance matrix $\mathbf{A}\left(\mathbf{C}^1,\mathbf{C}^2\right) = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$ has rank 2. If we assume this is the minimum determinant codeword pairs, then the coding gain is

$$G_d = \left(\left|\mathbf{A}\left(\mathbf{C}^1,\mathbf{C}^2\right)\right|\right)^{\frac{1}{r}} = \left(\left|\mathbf{A}\left(\mathbf{C}^1,\mathbf{C}^2\right)\right|\right)^{\frac{1}{2}} = \sqrt{16} = 4\,.$$

Example 8.3

A two states and r = 1bit/s/Hz using BPSK STTC is depicted in Fig. 8.6 (a). The minimum CGD for this code is 16. Find an example of two codeword that provide a CGD of 16 which corresponds to different trellis paths.

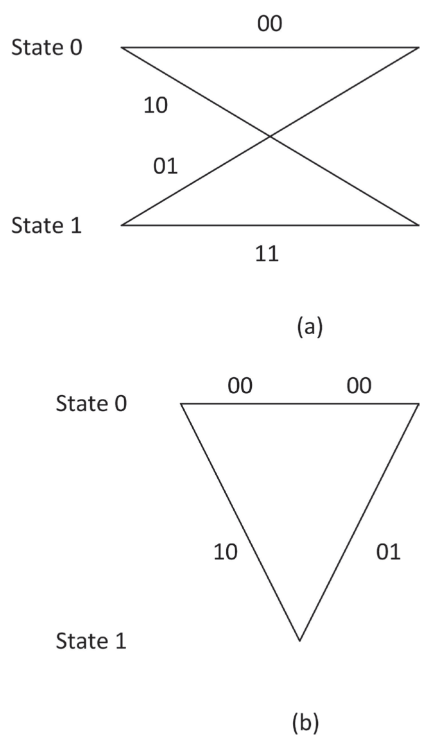


Fig. 8.6 Example 8.3 (a) 2-state STTC (b) trellis path

Solution

We can consider the two different trellis paths as depicted in Fig. 8.6 (b). The first path stays in state zero during two transitions, that is, 000. The second path goes to state one in the 1st transition and merges to state zero in the 2nd transition i.e., 010.

The corresponding codewords for the two different trellis paths after doing M-ary mapping Ξ for BPSK shown in Fig. 8.6(b) are

$$\mathbf{C}^1 = \begin{pmatrix} \Xi(0) = e^{j0} = 1 & \Xi(0) = e^{j0} = 1 \\ \Xi(0) = e^{j0} = 1 & \Xi(0) = e^{j0} = 1 \end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix} \Xi(1) = e^{j\pi} = -1 & \Xi(0) = e^{j0} = 1 \\ \Xi(0) = e^{j0} = 1 & \Xi(1) = e^{j\pi} = -1 \end{pmatrix}$$

Assume that the transmitted codeword is $\mathbf{C}^2$ but receiver decides in favour of $\mathbf{C}^1$ (erroneous codeword). Then codeword difference and distance matrix of $\mathbf{C}^1$ and $\mathbf{C}^2$ and CGD can be calculated as

$$\mathbf{D}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{C}^1 - \mathbf{C}^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{A}(\mathbf{C}^1, \mathbf{C}^2) = \mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right)\left(\mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right)\right)^T = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

$$CGD = \left|\mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right)\right| = 16$$

Note that the codeword distance matrix $\mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$ has rank 2. If we assume this is the minimum determinant codeword pairs, then the coding gain is

$$G_d = \left(\left|\mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right)\right|\right)^{\frac{1}{r}} = \left(\left|\mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right)\right|\right)^{\frac{1}{2}} = \sqrt{16} = 4$$

Example 8.4

Assume that for an STTC with $N_T = 2$, the transmitted codeword is 220313 and the decoder decides in favour of the codeword 330122. The symbols transmitted are from QPSK scheme. Find the coding gain and diversity gain for this case.

Solution

$$\mathbf{C}^1 = \mathbf{C}^1 = \begin{pmatrix} j & -j \\ 1 & -j \\ -1 & -1 \end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix} -1 & -1 \\ 1 & j \\ -j & -j \end{pmatrix}$$

$$\mathbf{D}(\mathbf{C}^2, \mathbf{C}^1) = \mathbf{C}^2 - \mathbf{C}^1 = \begin{pmatrix} -1-j & -1+j \\ 0 & 2j \\ -j+1 & -j+1 \end{pmatrix}$$

$$\mathbf{A}(\mathbf{C}^2, \mathbf{C}^1) = \mathbf{D}\left(\mathbf{C}^2, \mathbf{C}^1\right)\left(\mathbf{D}\left(\mathbf{C}^2, \mathbf{C}^1\right)\right)^H = \begin{bmatrix} 4 & 2+2j & -2-2j \\ 2-2j & 4 & -2+2j \\ -2+2j & -2-2j & 4 \end{bmatrix}$$

The codeword distance matrix has only two eignevalues 2.5359 and 9.4641. Therefore, rank of codeword distance matrix is 2 (diversity gain is $2N_R$) and coding gain is

$$G_d = (CGD)^{1/r} = (CGD)^{1/2} = (2.5359 \times 9.4641)^{1/2} = \sqrt{24}$$

Table 8.2

STTC employing QPSK for two transmit antennas designed based on rank and determinant criteria

STTC	*Generator sequence*	*Rank*	*Coding gain*
T(4,4,2)	$\mathbf{T}(4,4,2) = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}$	2	2
Y(4,4,2)	$\mathbf{Y}(4,4,2) = \begin{bmatrix} 2 & 0 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{bmatrix}$	2	$\sqrt{8}$
B(4,4,2)	$\mathbf{B}(4,4,2) = \begin{bmatrix} 0 & 2 & 3 & 1 \\ 2 & 2 & 1 & 0 \end{bmatrix}$	2	$\sqrt{8}$
T(4,8,2)	$\mathbf{T}(4,8,2) = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 \\ 1 & 2 & 0 & 0 & 2 \end{bmatrix}$	2	$\sqrt{12}$
Y(4,8,2)	$\mathbf{Y}(4,8,2) = \begin{bmatrix} 0 & 2 & 1 & 0 & 2 \\ 2 & 1 & 0 & 2 & 2 \end{bmatrix}$	2	4
B(4,8,2)	$\mathbf{B}(4,8,2) = \begin{bmatrix} 0 & 2 & 1 & 2 & 2 \\ 1 & 2 & 0 & 0 & 2 \end{bmatrix}$	2	$\sqrt{12}$
T(4,16,2)	$\mathbf{T}(4,16,2) = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 & 0 \\ 1 & 2 & 2 & 0 & 0 & 2 \end{bmatrix}$	2	$\sqrt{12}$
Y(4,16,2)	$\mathbf{Y}(4,16,2) = \begin{bmatrix} 0 & 2 & 1 & 1 & 2 & 0 \\ 2 & 2 & 1 & 2 & 0 & 2 \end{bmatrix}$	2	$\sqrt{32}$
B(4,16,2)	$\mathbf{B}(4,16,2) = \begin{bmatrix} 2 & 1 & 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 1 & 0 & 2 \end{bmatrix}$	2	$\sqrt{20}$

8.3.4 STTC design using rank and determinant criteria

In the following, we will discuss some STTC started by V. Tarokh et al., (1998), Q. Yan et al., (2000) and S. Baro et al., (2000). We will denote the generator matrix for the above STTC by $\mathbf{T}(M, S, N_T)$, $\mathbf{Y}(M, S, N_T)$ and $\mathbf{B}(M, S, N_T)$, respectively. M signifies M-ary modulation scheme, S is the number of states in the trellis diagram and N_T is the number of transmitting antennas. All the above STTC achieves full diversity of $N_T N_R$ and hence we will compare their coding gains only. STTCs could be designed using rank and determinant criteria. Optimal STTC could be generated using computer based search. Some of the common STTC using QPSK modulation scheme is listed in Table 8.2 along with their rank and coding gains for two transmit antenna case for different number of states. From the determinant criteria, we can design STTC with higher coding gains. We can observe earlier STTC by Tarokh has the lowest coding gains whereas STTC by Yan has the highest coding gains. One can explore STTC design using trace criterion in B. Vucetic et al., (2003). Super orthogonal space-time trellis code is discussed in M. K. Simon et al., (2005) and H. Jafarkhani (2005).

Review question 8.7 *Does STTC have both coding and diversity gain?*

8.3.5 STTC maximum likelihood (ML) decoding

Due to similarity of convolutional codes and STTC, we can use Viterbi algorithm for decoding STTC (V. Kuhn, 2006). In STTC, the N_T symbols from N_T transmitting antennas at one particular time slot interfere incoherently at the every receive antenna. We will consider the simplified case for $N_T = 2$ transmit antennas and N_R receive antennas. N_L symbols are transmitted over $N_L + Q$ time slots where $Q = N - 1$ is the memory of the convolutional code representing the state-machine which has N shift registers. The ML decoding finds the most likely valid path that starts from state zero and merges to the state zero after $N_L + Q$ time slots. If we assume that the ith antenna receives $r_{1,i}$, $r_{2,i}$, …, $r_{N_L+Q,i}$, at time slots $t = 1, 2, \ldots, N_L + Q$. For a branch of the trellis which transmits symbols s_1 and s_2 from transmitting antennas 1 and 2 the corresponding branch metric is given as,

$$\sum_{i=1}^{N_R} \left| r_{t,i} - h_{i,1}s_1 - h_{i,2}s_2 \right|^2 \tag{8.8a}$$

The path metric is the tally of the branch metrics for the branches that form the path which kicks off from state zero and merges to state zero after $N_L + Q$ time slots. The most likely path is the one which has the minimum path metric. Hence, the ML decoder finds the set of symbols that construct a valid path (which kicks off from state zero and merges to state zero after $N_L + Q$ time slots) by solving the following minimization problem (H. Jafarkhani, 2005).

$$\min_{s_{1,1}, s_{2,1}, \cdots, s_{N_L+Q,1}, s_{1,2}, s_{2,2}, \cdots, s_{N_L+Q,2}} \sum_{t=1}^{N_L+Q} \sum_{i=1}^{N_R} \left| r_{t,i} - h_{i,1}s_{t,1} - h_{i,2}s_{t,2} \right|^2 \tag{8.8b}$$

Review question 8.8 *Explain the ML decoding of STTC for* $N_T = 2$.

8.4 Performance analysis of space-time codes over separately correlated MIMO channel

Let us denote the codeword difference matrix between two codewords, $\mathbf{C}^1$ and $\mathbf{C}^2$ by $\boldsymbol{\Delta} = \mathbf{C}^1 - \mathbf{C}^2$. Assume that we are sending the space-time codeword $\mathbf{C}^1$ and the decoder decides in favour of $\mathbf{C}^2$. Hence the conditional pairwise error probability (PEP) is given as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2 \middle| \mathbf{H}\right) = Q\left(\sqrt{\frac{E_s}{2N_0}\|\mathbf{H}\boldsymbol{\Delta}\|^2}\right) = Q\left(\sqrt{\frac{\gamma}{2}\|\mathbf{H}\boldsymbol{\Delta}\|^2}\right)$$

where, N_0 is the one-sided noise power spectral density.

How do we obtain the above conditional PEP?

Using the alternate form of Q-function (J. Craig, 1991), we have,

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2 \middle| \mathbf{H}\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{\gamma\|\mathbf{H}\boldsymbol{\Delta}\|^2}{4\sin^2\theta}\right)d\theta$$

Hence the unconditional pairwise error probability (PEP) is given as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\int_0^{\infty} \exp\left(-\frac{\gamma\|\mathbf{H}\boldsymbol{\Delta}\|^2}{4\sin^2\theta}\right)p_{\|\mathbf{H}\boldsymbol{\Delta}\|^2}(\zeta)d\zeta d\theta = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} M_{\|\mathbf{H}\boldsymbol{\Delta}\|^2}\left(-\frac{\gamma}{4\sin^2\theta}\right)d\theta \quad (8.9a)$$

In the above equation, assuming $s = -\dfrac{\gamma}{4\sin^2\theta}$, we can denote $M_{\|\mathbf{H}\boldsymbol{\Delta}\|^2}(s)$ as the moment generating function (MGF) of $\|\mathbf{H}\boldsymbol{\Delta}\|^2$.

Note that

$$\|\mathbf{H}\boldsymbol{\Delta}\|^2 = trace\left(\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Delta}^H\mathbf{H}^H\right) = \left\{vect\left(\mathbf{H}^H\right)\right\}^H\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta}\boldsymbol{\Delta}^H\right)vect\left(\mathbf{H}^H\right)$$

One can verify the above expression for 2×2 MIMO system. For separately correlated MIMO channel, we have,

$$\mathbf{H} = \mathbf{R}_{\mathbf{R_X}}^{1/2}\mathbf{H}_w\mathbf{R}_{\mathbf{T_X}}^{1/2}$$

where, $\mathbf{H}_w$ is the channel matrix for uncorrelated or i.i.d. (spatially white that's why the subscript w in **H**) MIMO fading channel.

Also, $$vect(\mathbf{H}) = \sqrt{\mathbf{R}_H}\,vect\left(\mathbf{H}_w\right)$$

Therefore, $$\|\mathbf{H}\boldsymbol{\Delta}\|^2 = \left\{vect\left(\mathbf{H}_w^H\right)\right\}^H\left(\mathbf{R}_H\right)^{H/2}\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta}\boldsymbol{\Delta}^H\right)\left(\mathbf{R}_H\right)^{1/2} vect\left(\mathbf{H}_w^H\right)$$

where, $\mathbf{R_H} = \mathbf{R}_{R_X} \otimes \mathbf{R}_{T_X}$ is the spatial correlation matrix.

Theorem I

Consider the random quadratic form of a Hermitian matrix **A** in complex Gaussian multivariate $\mathbf{v} \sim N_c^N(\boldsymbol{\mu}_\mathbf{v}, \mathbf{R}_\mathbf{v})$ as

$$y = \text{Quad}_\mathbf{A}(\mathbf{v}) = \mathbf{v}\mathbf{A}\mathbf{v}^H$$

The MGF of the y (see appendix B) is given as

$$M_y^{(s)} = \int_0^\infty e^{sy}p_Y(y)\,dy = \frac{\exp\left[s\left(\boldsymbol{\mu}_\mathbf{V}\right)\mathbf{A}\left(\mathbf{I} - s\mathbf{R}_\mathbf{V}\mathbf{A}\right)^{-1}\left(\boldsymbol{\mu}_\mathbf{V}\right)^H\right]}{\left|\mathbf{I} - s\mathbf{R}_\mathbf{V}\mathbf{A}\right|} \quad (8.9b)$$

where, **I** is the identity matrix with appropriate size.

8.4.1 Rayleigh fading MIMO channel

Assuming $\mathbf{v} = \left\{vect\left(\mathbf{H}_w^H\right)\right\}^H$ is a zero mean ($\boldsymbol{\mu}_\mathbf{v} = \mathbf{0}$) Gaussian vector with covariance matrix as an identity matrix $\mathbf{R}_\mathbf{v} = \mathbf{I}_{N_R N_T}$ and $\mathbf{A} = \left(\mathbf{R}_H\right)^{H/2}\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta\Delta}^H\right)\left(\mathbf{R}_H\right)^{1/2}$ is a Hermitian matrix. Note that for ($\boldsymbol{\mu}_\mathrm{v} = \mathbf{0}$), the exponential term in the MGF expression of Eq. (8.9b) will become 1. Hence, the MGF for random quadratic form of a Hermitian matrix $\mathbf{A}$ for correlated Rayleigh fading channel is

$$M_y(s) = \left|\mathbf{I}_{N_R N_T} - s\mathbf{A}\right|^{-1} \tag{8.10a}$$

Then, the MGF of y becomes

$$M_y(s) = \left|\mathbf{I}_{N_R N_T} - s\left(\mathbf{R}_H\right)^{H/2}\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta\Delta}^H\right)\left(\mathbf{R}_H\right)^{1/2}\right|^{-1}$$

Since $\left|\mathbf{I} + \mathbf{AB}\right| = \left|\mathbf{I} + \mathbf{BA}\right|$, we have,

$$M_y(s) = \left|\mathbf{I}_{N_R N_T} - s\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta\Delta}^H\right)\mathbf{R}_H\right|^{-1}$$

Since $\mathbf{R}_H = \mathbf{R}_{R_X} \otimes \mathbf{R}_{T_X}$, we have,

$$M_y(s) = \left|\mathbf{I}_{N_R N_T} - s\left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Delta\Delta}^H\right)\left(\mathbf{R}_{R_X} \otimes \mathbf{R}_{T_X}\right)\right|^{-1}$$

Since $\left(\mathbf{A} \otimes \mathbf{B}\right)\left(\mathbf{C} \otimes \mathbf{D}\right) = \mathbf{AC} \otimes \mathbf{BD}$, we have,

$$M_y(s) = \left|\mathbf{I}_{N_R N_T} - s\left(\mathbf{R}_{R_X} \otimes \boldsymbol{\Delta\Delta}^H\mathbf{R}_{T_X}\right)\right|^{-1}$$

If $\lambda_\mathbf{A}$ is an eigenvalue of $\mathbf{A}$ and $\lambda_\mathbf{B}$ is an eigenvalue of $\mathbf{B}$, then $\lambda_\mathbf{A}\lambda_\mathbf{B}$ is an eigenvalue of $\mathbf{A} \otimes \mathbf{B}$. Hence,

$$M_y(s) = \prod_{n=1}^{r}\prod_{m=1}^{\hat{r}}\left(1 - s\lambda_n\mu_m\right)^{-1}$$

where, r is the rank of $\boldsymbol{\Delta\Delta}^H\mathbf{R}_{T_X}$; $\hat{r}$ is the rank of $\mathbf{R}_{R_X}$; λ_n is the eignvalue of $\boldsymbol{\Delta\Delta}^H\mathbf{R}_{T_X}$; μ_m is the eignvalue of $\mathbf{R}_{R_X}$.

In our case, $s = -\dfrac{\gamma}{4\sin^2\theta}$; hence, the exact PEP becomes

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\prod_{n=1}^{r}\prod_{m=1}^{\hat{r}}\left(1 + \frac{\gamma\lambda_n\mu_m}{4\sin^2\theta}\right)^{-1}d\theta \tag{8.10b}$$

Example 8.5

Find the PEP for the following two space-time codewords:

(a) $\mathbf{C}^1 = \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$

(b) $\mathbf{C}^1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

(c) $\mathbf{C}^1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}; \mathbf{C}^2 = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}$

Assume i.i.d. Rayleigh fading MIMO channel.

Solution

For i.i.d. Rayeligh fading MIMO channel,

$$\mathbf{R}_{T_X} = \mathbf{I}_{N_T} \quad \Rightarrow \quad \mathbf{\Delta\Delta}^H \mathbf{R}_{T_X} = \mathbf{\Delta\Delta}^H$$

$$\mathbf{R}_{R_X} = \mathbf{I}_{N_R} \quad \Rightarrow \quad \mu_m = 1, \hat{r} = N_R$$

Hence, Eq. (8.10b) could be simplified as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \prod_{n=1}^{r} \left(1 + \frac{\gamma\lambda_n}{4\sin^2\theta}\right)^{-N_R} d\theta$$

$$= \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \prod_{n=1}^{r} \left(\frac{4\sin^2\theta}{4\sin^2\theta + \gamma\lambda_n}\right)^{N_R} d\theta$$

Note that $\mathbf{\Delta}$ is the codeword difference matrix also denoted by $\mathbf{D}(\mathbf{C}^1, \mathbf{C}^2)$ for two codewords $\mathbf{C}^1$ and $\mathbf{C}^2$. All we need to find is the eigenvalues of codeword distance matrix.

(a) $\mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \Rightarrow \mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$

Hence the eigenvalues are 4 and 4.

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\frac{4\sin^2\theta}{4\sin^2\theta + \gamma 4}\right]^{2N_R} d\theta = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\frac{\sin^2\theta}{\sin^2\theta + \gamma}\right]^{2N_R} d\theta$$

(b) $\mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \Rightarrow A\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$

Hence the eigenvalue is 4.

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \int_0^{\frac{\pi}{2}} \left[\frac{4\sin^2\theta}{4\sin^2\theta + \gamma 4}\right]^{N_R} d\theta = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\frac{\sin^2\theta}{\sin^2\theta + \gamma}\right]^{N_R} d\theta$$

(c) $\mathbf{D}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 0 & 2 \\ 0 & 2 \\ 2 & 0 \end{pmatrix}; \mathbf{A}\left(\mathbf{C}^1, \mathbf{C}^2\right) = \begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix}$

Hence the eigenvalues are 4 and 8.

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\left[\frac{4\sin^2\theta}{4\sin^2\theta+\gamma 4}\right]^{N_R}\left[\frac{4\sin^2\theta}{\sin^2\theta+\gamma 8}\right]d\theta$$

$$= \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\left[\frac{\sin^2\theta}{\sin^2\theta+\gamma}\right]^{N_R}\left[\frac{\sin^2\theta}{\sin^2\theta+2\gamma}\right]^{N_R}d\theta$$

The close form expressions of the above PEP can be obtained using the following integral.

$$\because \qquad I_n\left(\bar{\gamma}\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\left(\frac{\sin^2\theta}{\sin^2\theta+\bar{\gamma}}\right)^n d\theta$$

$$= \frac{1}{2} - \left[\frac{1}{2} - A\left(\bar{\gamma}\right)\right]\sum_{i=0}^{n-1}\binom{2i}{i}A\left(\bar{\gamma}\right)^i\left[1 - A\left(\bar{\gamma}\right)\right]^i;$$

$$A\left(\bar{\gamma}\right) = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1+\bar{\gamma}}}\right]$$

where, n is an integer.

Once we have the PEP between all codewords, we can find the union bound for error probability as

$$P\left(e\right) \le \frac{1}{M}\sum_{\mathbf{C}}\sum_{\mathbf{C}\ne\hat{\mathbf{C}}} P\left(\mathbf{C} \to \hat{\mathbf{C}}\right) \tag{8.11a}$$

where, M is the total number of codewords.

For Chernoff bound, put $\theta = \frac{\pi}{2}$ or $\sin^2\left(\theta\right) = 1$ in equation 8.10 (b) for PEP, then

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \le \frac{1}{2}\prod_{n=1}^{r}\prod_{m=1}^{\hat{r}}\left(1 + \frac{\gamma\lambda_n\mu_m}{4}\right)^{-1} \tag{8.11b}$$

For high SNR case, PEP is bounded as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \le \frac{1}{2}\prod_{n=1}^{r}\prod_{m=1}^{\hat{r}}\left(\frac{\gamma\lambda_n\mu_m}{4}\right)^{-1} \tag{8.12a}$$

For i.i.d. Rayeligh fading MIMO channel,

$$\boldsymbol{\Delta\Delta}^H\mathbf{R}_{T_X} = \boldsymbol{\Delta\Delta}^H$$

$$\mathbf{R}_{R_X} = \mathbf{I}_{N_R} \qquad \Rightarrow \qquad \mu_m = 1,\ \hat{r} = N_R$$

$$\Rightarrow P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \le \frac{1}{2}\prod_{n=1}^{r}\left(\frac{\gamma\lambda_n\mu_m}{4}\right)^{-N_R} = \frac{1}{2}\frac{4^{rN_R}}{\left(\prod_{n=1}^{r}\lambda_n\right)^{N_R}\gamma^{rN_R}} \tag{8.12b}$$

8.4.2 Rician fading MIMO channel

Assuming $\mathbf{v} = \left\{vect\left(\mathbf{H}_w^H\right)\right\}^H$ is a non-zero mean $\boldsymbol{\mu}_v \neq \mathbf{0}$ Gaussian vector with covariance matrix as an identity matrix $\mathbf{R}_\mathbf{v} = \left(\frac{1}{1+K}\right)\mathbf{I}_{N_R N_T}$ where, K is the Rice parameter and $\mathbf{A} = (\mathbf{R}^H)^{H/2}\left(\mathbf{I}_{N_R} \otimes \Delta\Delta^H\right)\left(\mathbf{R}_H\right)^{1/2}$ is a Hermitian matrix. The MGF for random quadratic form of a Hermitian matrix $\boldsymbol{A}$ (Appendix B) for correlated Rayleigh fading channel is

$$M_y(s) = \frac{\exp\left[s\boldsymbol{\mu}_\mathbf{v}\mathbf{A}\left\{\mathbf{I}_{N_R N_T} - \frac{s\mathbf{A}}{(K+1)}\right\}^{-1}(\boldsymbol{\mu}_\mathbf{v})^H\right]}{\left|\mathbf{I}_{N_R N_T} - \frac{s\mathbf{A}}{(K+1)}\right|} \tag{8.13a}$$

Therefore, the exact PEP is given as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{e^{\left[-\frac{\gamma}{4\sin^2\theta}\boldsymbol{\mu}_\mathbf{v}\mathbf{A}\left\{\mathbf{I}_{N_R N_T} + \frac{\gamma\mathbf{A}}{4\sin^2\theta(K+1)}\right\}^{-1}\boldsymbol{\mu}_\mathbf{v}^H\right]}}{\left|\mathbf{I}_{n_R n_T} + \frac{\gamma\mathbf{A}}{4\sin^2\theta(K+1)}\right|}d\theta \tag{8.13b}$$

Rayleigh fading is a particular case of Rician fading where $\boldsymbol{\mu}_v \neq \mathbf{0}$ and $K=0$. Once we have the PEP between all codewords, we can find the union bound for error probability as

$$P(e) \leq \frac{1}{M}\sum_{\mathbf{C}}\sum_{\mathbf{C}\neq\hat{\mathbf{C}}} P\left(\mathbf{C} \to \hat{\mathbf{C}}\right)$$

where, M is the total number of codewords.

For Chernoff bound, put $\theta = \frac{\pi}{2}$ or $\sin^2(\theta) = 1$, then

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \leq \frac{1}{2}\frac{e^{\left[-\frac{\gamma}{4}\boldsymbol{\mu}_\mathbf{v}\mathbf{A}\left\{\mathbf{I}_{N_R N_T} + \frac{\gamma\mathbf{A}}{4(K+1)}\right\}^{-1}\boldsymbol{\mu}_\mathbf{v}^H\right]}}{\left|\mathbf{I}_{n_R n_T} + \frac{\gamma\mathbf{A}}{4(K+1)}\right|} \tag{8.14a}$$

For high SNR case, the exponential term could be simplified as

$$e^{\left[-\frac{\gamma}{4}\boldsymbol{\mu}_\mathbf{v}\left\{\frac{4(K+1)}{\gamma}\right\}\boldsymbol{\mu}_\mathbf{v}^H\right]} = e^{\left[-\boldsymbol{\mu}_\mathbf{v}(K+1)\boldsymbol{\mu}_\mathbf{v}^H\right]}$$

Note that we could model Rice MIMO channel as (see Eq. 2.7b)

$$\mathbf{H}_{Rice} = \sqrt{\frac{K}{1+K}}\mathbf{H}_{LOS} + \sqrt{\frac{1}{1+K}}\mathbf{H}_{Rayleigh}$$

Therefore, assuming $\mathbf{H}_{LOS}$ is an all 1 $N_R \times N_T$ matrix.

$$\boldsymbol{\mu}_{\mathbf{v}}\,(K+1)\,\boldsymbol{\mu}_{\mathbf{v}}^H = N_R N_T \left(\frac{K}{K+1}\right) \times (K+1)$$

Note that

$$\left|\mathbf{I}_{n_R n_T} + \frac{\gamma \mathbf{A}}{4(K+1)}\right|^{-1} = \prod_{n=1}^{r} \prod_{m=1}^{\hat{r}} \left(1 + \frac{\gamma \lambda_n \mu_m}{4(K+1)}\right)^{-1}$$

For high SNR,

$$\left|\mathbf{I}_{n_R n_T} + \frac{\gamma \mathbf{A}}{4(K+1)}\right|^{-1} = \prod_{n=1}^{r} \prod_{m=1}^{\hat{r}} \left(\frac{\gamma \lambda_n \mu_m}{4(K+1)}\right)^{-1}$$

Hence, for high SNR case, PEP is bounded as

$$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \le \frac{1}{2}\left(\frac{\gamma}{4(K+1)}\right)^{-r\hat{r}} e^{[-N_R N_T K]} \prod_{n=1}^{r} \prod_{m=1}^{\hat{r}} \left(\frac{1}{\lambda_n \mu_m}\right) \tag{8.14b}$$

For i.i.d. Rician fading MIMO channel,

$$\mathbf{R}_{T_X} = \mathbf{I}_{N_T} \qquad \Rightarrow \qquad \Delta\Delta^H \mathbf{R}_{T_X} = \Delta\Delta^H$$

$$\mathbf{R}_{R_X} = \mathbf{I}_{N_R} \qquad \Rightarrow \qquad \boldsymbol{\mu}_m = 1,\ \hat{r} = N_R$$

Hence, $$P\left(\mathbf{C}^1 \to \mathbf{C}^2\right)_{bound} \le \frac{1}{2}\left(\frac{4}{\gamma}\right)^{rN_R} (K+1)^{rN_R}\, e^{[-N_R N_T K]} \left(\prod_{n=1}^{r} \frac{1}{\lambda_n}\right)^{N_R} \tag{8.15}$$

Rayleigh fading is a particular case of Rician fading where $\boldsymbol{\mu}_{\mathbf{v}} \neq \mathbf{0}$ and $K = 0$. Comparing the PEP bound for i.i.d. Rayleigh fading with Rician fading case, we can see that the extra term $(K+1)^{rN_R}\, e^{[-N_R N_T K]}$ is getting multiplied. Note that $e^{[-N_R N_T K]}$ decays very fast and it is a very small number hence the combined effect of $(K+1)^{rN_R}\, e^{[-N_R N_T K]}$ is a small number much less than 1. Hence PEP bound for Rician case is smaller than Rayleigh case. This is reasonable since Rician has line of sight (LOS) component and decrease the PEP.

Review question 8.9 *Write down the SER bound of STC over correlated Rayleigh fading MIMO channel.*

Review question 8.10 *Write down the SER bound of STC over correlated Rician fading MIMO channel.*

8.5 Introduction to space-time turbo encoders

One may refer to appendix E on basics of turbo codes. Note that the space-time turbo codes are extension of turbo codes for multiple antennas. We will consider the case of two transmit antennas for illustration purpose. Figure 8.7 depicts turbo space-time coded modulation scheme. It comprises two parallel concatenated systematic and recursive STTCs. One of the antennas is always linked to the systematic output of the systematic and recursive STTC 1, and the second antenna is joined to the parity symbols of the two systematic and recursive STTCs. With puncturing (full rate may be achieved), one parity symbol for each of the two systematic and recursive STTCs may be sent through the channel and ignore the other parity symbol. But then full diversity may not be achieved. This space-time turbo codes is very similar to the binary turbo codes we have discussed in Appendix E. But there are some differences. The interleavers operate on symbols rather than on bits. There are interleavers for the systematic and recursive STTC 2 and de-interleaving operation before sending over the channel. It makes sure that the systematic symbols for both the systematic and recursive STTCs are equal. Its decoders are very similar to the binary turbo decoders except that trellis diagram will be used for symbols rather than bits. Iterative decoders are employed in turbo codes. One may refer to T. M. Duman et al., (2007), S. J. Johnson (2010) and E. Biglieri (2005) for turbo codes decoders.

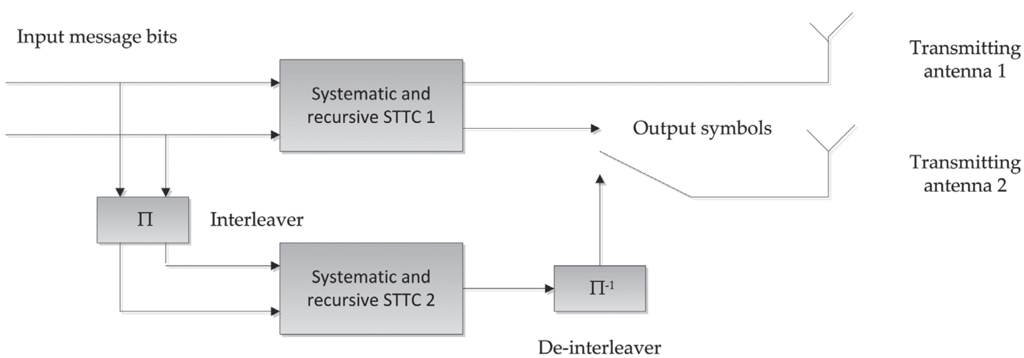


Fig. 8.7 Turbo space-time coded modulation scheme (STTC: space-time trellis code)

Review question 8.11 *Explain the space–time turbo encoder.*

8.6 Algebraic space-time codes

Alamouti space-time codes (see section 7.4), is the only full rate (rate r = 1) and full diversity OSTBC for complex signal constellations, which were designed for $N_T = 2$ MIMO system. There are no OSTBC with full rate and full diversity for $N_T > 2$ for complex signal constellations. Can we have a full rate and full diversity STBC for $N \times N$ MIMO system? This is possible with some class of space-time codes popularly known as Algebraic space-time codes. In this section we will discuss about Diagonal Algebraic Space-time Codes (full diversity and rate one code), Threaded Algebraic Space-time Codes (full diversity and rate r code where, $r \geq 1$) and Perfect Space-time Codes (full diversity and rate r code where $r \geq 1$). These codes are designed using Algebraic structures (an introduction to algebraic structures is given in Appendix F).

8.6.1 Diagonal algebraic space-time codes

Damen codes also popularly known as diagonal algebraic space-time codes (M. O. Damen et. al., 2002) can be constructed for $N_T = N$ equal to two and multiples of four. The code construction is a two-step process.

- First find an optimal unitary matrix of dimension N having the maximal diversity.
- Second use Hadamard matrix to multiplex information symbols over space and time.

Consider a number field $K = Q\left(\theta = e^{\frac{i\pi}{4}}\right)$ of degree two over base field $Q(i)$ (q-QAM constellations). The minimum polynomial of θ, $p_\theta(x) = x^2 - i$. Its root $\theta = e^{\frac{i\pi}{4}}$ is called an algebraic number and we will consider algebra of θ. Its conjugate $\bar{\theta} = -e^{\frac{i\pi}{4}}$. $\mathbf{b} = [1 \quad \theta]$ is the integral basis of $K = Q(\theta)$ and each element of $K = Q(\theta)$ can be expressed in polynomial form as $x = a + b\theta, a, b \in Q(i)$. Let the embedding $\sigma: \theta \mapsto -\theta$ be the generator of the Galois group of K. The lattice $L = \sigma(O_K)$ where O_K is the ring of integers ($O_K = \{a + b\theta, a, b \in Z[i]\}$) have a generator matrix as $\mathbf{G} = \begin{bmatrix} 1 & \theta \\ 1 & \bar{\theta} \end{bmatrix}$. The unitary matrix of dimension two can be constructed as $\mathbf{U}_2 = \frac{\mathbf{G}}{\sqrt{2}}$. For two QAM information symbols, $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$, we can rotate this input symbol vector by multiplying with the unitary matrix $\mathbf{U}$ as $\mathbf{x} = \mathbf{U}_2\mathbf{s} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \theta \\ 1 & -\theta \end{bmatrix}\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} s_1 + \theta s_2 \\ s_1 - \theta s_2 \end{bmatrix}$. This operation augment the algebraic dimension of constellation as $K = Q(\theta)$ is a vector space of dimension two over $Q(i)$. An STBC will not incur any information loss if the maximum instantaneous mutual information of the equivalent MIMO channel that includes the STBC codeword is equal to the maximum instantaneous mutual information of the MIMO channel without the STBC codeword. Any unitary transform will preserve the mutual information while changing the diversity and coding gain (B. S. Rajan, 2014). The above codeword could be rewritten in a diagonal matrix form as

$$\mathbf{X} = \frac{1}{\sqrt{2}}\begin{bmatrix} s_1 + \theta s_2 & 0 \\ 0 & s_1 - \theta s_2 \end{bmatrix}$$

Now applying the Hadamard transformation, we have the codeword matrix of DAST for N = 2.

$$\begin{aligned} \mathbf{C}_2^{DAST} &= \mathbf{H}_2^{Hadm}\mathbf{X} \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} s_1 + \theta s_2 & 0 \\ 0 & s_1 - \theta s_2 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix} s_1 + \theta s_2 & -(s_1 - \theta s_2) \\ s_1 + \theta s_2 & s_1 - \theta s_2 \end{bmatrix} \end{aligned}$$

For dimension N, we can construct codes similarly.

$$\mathbf{C}^{DAST} = \mathbf{H}_N^{Hadm} \bullet \mathbf{X} = \mathbf{H}_N^{Hadm} \bullet diag\left(\mathbf{U} \bullet \mathbf{s}\right)$$

Note that we need to construct an optimal $N \times N$ unitary matrix $\mathbf{U}$ in the above equation. $\mathbf{H}_N^{Hadm}$ is the Hadamard matrix of N dimension which could be obtained from $\frac{N}{2} \times \frac{N}{2}$ Hadamard matrix as follows $\mathbf{H}_N^{Hadm} = \begin{pmatrix} \mathbf{H}_{N/2}^{Hadm} & \mathbf{H}_{N/2}^{Hadm} \\ \mathbf{H}_{N/2}^{Hadm} & -\mathbf{H}_{N/2}^{Hadm} \end{pmatrix}$ and $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}$ is the symbol vector. The coding gain of N-dimension DAST codes are given by

$$\delta_{\min} = \begin{cases} \dfrac{2^{\frac{2}{N}}}{\sqrt{5}}, & N = 2,4 \\[2ex] \dfrac{1}{2^{\frac{N-1}{N}}}, & N \geq 8 \end{cases}$$

8.6.2 Threaded algebraic space-time codes

Gamal code, also popularly known as Threaded Algebraic Space-time (TAST) code (H. E. Gamal et al., 2003), are fully diverse and full rate (can achieve any rate $r \geq 1$) for any number of transmit antennas. The rate r TAST code for $N_T \times N_R$ MIMO system can be constructed as

$$\mathbf{C}_{N_T}^{TAST} = \sum_{k=1}^{r} \mathbf{D}_k\left(\mathbf{U}_{N_T}\mathbf{x}_k\right)\left(\gamma\Pi\right)^{k-1}$$

where,

- $\Pi = \left[\mathbf{e}_{N_T}, \mathbf{e}_1, \cdots, \mathbf{e}_{N_T-2}, \mathbf{e}_{N_T-1}\right]$ and $\mathbf{e}_i$ is the i$^{\text{th}}$ column of a $N_T \times N_T$ identity matrix
- γ is a unit-magnitude complex number which is dependent on the QAM alphabet size and N_T
- $\mathbf{D}_k\left(\mathbf{U}_{N_T}\mathbf{x}_k\right) = diag\left(\mathbf{U}_{N_T}\mathbf{x}_k\right)$ is the diagonal matrix with diagonal elements consisting of a rotated version of the k$^{\text{th}}$ symbol
- $\mathbf{x}_k = \begin{bmatrix} x_{k,0} \\ x_{k,1} \\ \vdots \\ x_{k,N_T-2} \\ x_{k,N_T-1} \end{bmatrix}$ is the vector consisting of N_T information symbols of the k$^{\text{th}}$ thread
- $\mathbf{U}_{N_T}$ is the $N_T \times N_T$ unitary rotation matrix

For example, for $N_T = 2$ and $r = 2$, TAST code can be generated as follows:

$$\mathbf{C}_2^{TAST} = \sum_{k=1}^{2} \mathbf{D}_k \left(\mathbf{U}_{N_T} \mathbf{x}_k\right) \left(\gamma \Pi\right)^{k-1} = \mathbf{D}_1 \left(\mathbf{U}_2 \mathbf{x}_1\right) + \mathbf{D}_2 \left(\mathbf{U}_2 \mathbf{x}_2\right) \gamma \Pi$$

For this case,

$$\mathbf{U}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & e^{\frac{i\pi}{4}} \\ 1 & -e^{\frac{i\pi}{4}} \end{bmatrix}; \gamma \mathbf{\Pi} = \gamma \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\therefore \quad \mathbf{C}_2^{TAST} = Diag\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & e^{\frac{i\pi}{4}} \\ 1 & -e^{\frac{i\pi}{4}} \end{bmatrix} \begin{bmatrix} x_{1,0} \\ x_{1,1} \end{bmatrix} \right) + Diag\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & e^{\frac{i\pi}{4}} \\ 1 & -e^{\frac{i\pi}{4}} \end{bmatrix} \begin{bmatrix} x_{2,0} \\ x_{2,1} \end{bmatrix} \right) \begin{bmatrix} 0 & \gamma \\ \gamma & 0 \end{bmatrix}$$

$$\Rightarrow \quad \mathbf{C}_2^{TAST} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_{1,0} + e^{\frac{i\pi}{4}} x_{1,1} & 0 \\ 0 & x_{1,0} - e^{\frac{i\pi}{4}} x_{1,1} \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & \gamma\left(x_{2,0} + e^{\frac{i\pi}{4}} x_{2,1} \right) \\ \gamma\left(x_{2,0} - e^{\frac{i\pi}{4}} x_{2,1} \right) & 0 \end{bmatrix}$$

$$\Rightarrow \quad \mathbf{C}_2^{TAST} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_{1,0} + e^{\frac{i\pi}{4}} x_{1,1} & \gamma\left(x_{2,0} + e^{\frac{i\pi}{4}} x_{2,1} \right) \\ \gamma\left(x_{2,0} - e^{\frac{i\pi}{4}} x_{2,1} \right) & x_{1,0} - e^{\frac{i\pi}{4}} x_{1,1} \end{bmatrix}$$

where, $\gamma = \sqrt{e^{\frac{i\pi}{6}}}$ for QPSK and $\gamma = \sqrt{e^{\frac{i\pi}{4}}}$ for 16-QAM constellations.

8.6.3 Perfect space-time codes

Perfect STBCs are linear space-time codes that were proposed by F. Oggier et al., (2006). These codes are called perfect since they achieve full diversity, a non-vanishing determinant (NVD) for increasing spectral efficiency, uniform average transmitted energy per antenna, and achieve rate, $r \geq 1$. A perfect STBC has the following properties:

- *Full rank*: the determinant of the difference of any two distinct codewords is not null.
- *Full rate*: all N^2 degrees of freedom of the system are utilized which allows to send N^2 symbols from either QAM or HEX.
- *NVD for increase data rate*: the minimum determinant of a PSTC is lower bounded by a constant different from zero which is independent of spectral efficiency.
- *Efficient shaping*: the energy needed to transmit the linear combination of the information symbols on each layer is same to the energy used for transmitting the symbols themselves since each layer is modelled from the rotated version of the Z^N or Λ_N where Λ_2 is the hexagonal lattice.
- *Uniform energy*: it uses uniform average energy per antennas in all $T = N$ time slots, i.e., every coded symbols in the code matrix have the identical average energy.

The construction of algebraic codes involves (B. Ahmed et al., 2015):

- *Choice of the base*: The base is the body on which the extension base is defined and to which the information symbols belong. We will deal with the base $L = Q(i)$ (q-QAM constellations) or $L = Q(j)$ (q-HEX constellations). Note that $i = \sqrt{-1}$ and j is a primitive 3rd root of unity ($j^3 = 1, j = e^{\frac{i2\pi}{3}}$).
- *Choice of the extension*: The expansion base is the body on which the cyclic division algebra is constructed. In case of PSTBC, we shall then choose an extension cyclic base *K* of *L* of degree *n*. For fields K and L, K/L denotes that K is an extension of L (hence, K is an algebra over L) and [K : L] = *n* shows that K is a finite extension of L of degree *n*.
- *Definition of the cyclic algebra*: We may represent σ as the Galois group generator of base *K*, *Gal(K/L)*. *Gal(*K/L*)* denotes the Galois group of K/L, i.e., the group of L-linear automorphisms of K. If σ is any L-linear automorphism of K, $\langle\sigma\rangle$ denotes the cyclic group generated by σ (also called as generator). Let $A = Gal(K/L, \sigma, \gamma)$ a cyclic algebra of *n* degree. The algebra *A* is a division algebra, which necessitates that $\gamma, \gamma^2, \ldots, \gamma^{n-1}$ are not norm in L^*.
- *Definition of Space-time code*: Elements of A have matrix representation. Non-zero elements of A have an inverse. A space-time code can be defined as a finite subset of A.

Let $K = Q(i, \theta)$ be a cyclic extension of the base field *L* (such as $Q(i)$ or $Q(j)$) of degree n with Galois group $G_{K/L} = \langle\sigma\rangle$. $A(K/L, \sigma, \gamma)$ is a cyclic algebra of degree n iff

$$A = 1 \bullet K \oplus e \bullet K \oplus \cdots \oplus e^{n-1}K$$

where, $e \in A, e^n = \gamma \in L^*$ and $ze = e\sigma(z), \forall z \in K$.

A is a cyclic division algebra (CDA) iff $\gamma, \gamma^2, \ldots, \gamma^{n-1}$ are not norm in L^* which is the set *L* excluding the zero element. We can associate a multiplication matrix to each element. For example, $A(K/L, \sigma, \gamma)$ is a cyclic algebra of degree 2 iff $A = 1 \bullet K \oplus e \bullet K$ where $e \in A$, $e^2 = \gamma \in L^*$ and $ze = e\sigma(z)$, $\forall z \in K$. Let us find the multiplication of two elements $a_1, a_2 \in A$.

$$\begin{aligned} a_1a_2 &= (z_1 + ez_2)(z_3 + ez_4) \\ &= z_1z_3 + e\sigma(z_1)z_4 + ez_2z_3 + \gamma\sigma(z_2)z_4 \\ &= z_1z_3 + \gamma\sigma(z_2)z_4 + e\left(\sigma(z_1)z_4 + z_2z_3\right) \end{aligned}$$

For the basis $\{1 \quad e\}$, we may express the above equation as

$$a_1a_2 = \begin{pmatrix} z_1 & \gamma\sigma(z_2) \\ z_2 & \sigma(z_1) \end{pmatrix}\begin{pmatrix} z_3 \\ z_4 \end{pmatrix}$$

Thus there is a corresponding matrix for every element

$$a_1 = (z_1 + ez_2) \in A \leftrightarrow \begin{pmatrix} z_1 & \gamma\sigma(z_2) \\ z_2 & \sigma(z_1) \end{pmatrix}$$

Two particular cases:

(a) $a_1 = z_1 \in A \leftrightarrow \begin{pmatrix} z_1 & 0 \\ 0 & \sigma(z_1) \end{pmatrix}$ which can be expressed for n-dimension as

$$z_1 \leftrightarrow \begin{pmatrix} z_1 & & & \\ & \sigma(z_1) & & \\ & & \ddots & \\ & & & \sigma^{n-1}(z_n) \end{pmatrix}.$$

(b) $a_1 = e \in A \leftrightarrow \begin{pmatrix} 0 & \gamma \\ 1 & 0 \end{pmatrix}$, which can be expressed for n-dimension as

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & \gamma \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

$$\Rightarrow a = z_0 + ez_1 + e^2 z_2 + \cdots + e^{n-1} z_{n-1} \leftrightarrow \begin{pmatrix} z_0 & \gamma\sigma(z_{n-1}) & \gamma\sigma^2(z_{n-2}) & \gamma\sigma^3(z_{n-3}) & \cdots & \gamma\sigma^{n-1}(z_1) \\ z_1 & \sigma(z_0) & \gamma\sigma^2(z_{n-1}) & \gamma\sigma^3(z_{n-2}) & \cdots & \gamma\sigma^{n-1}(z_2) \\ z_2 & \sigma(z_1) & \sigma^2(z_0) & \gamma\sigma^3(z_{n-1}) & \cdots & \gamma\sigma^{n-1}(z_3) \\ z_3 & \sigma(z_2) & \sigma^2(z_1) & \sigma^3(z_0) & \cdots & \gamma\sigma^{n-1}(z_4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n-1} & \sigma(z_{n-2}) & \sigma^2(z_{n-3}) & \sigma^3(z_{n-4}) & \cdots & \sigma^{n-1}(z_0) \end{pmatrix}$$

Elements in a cyclic algebra A of degree n can be described by matrices of the form

$$\begin{pmatrix} z_0 & z_1 & \cdots & z_{n-1} \\ \gamma\sigma(z_{n-1}) & \sigma(z_0) & \cdots & \sigma(z_{n-2}) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma\sigma^{n-1}(z_1) & \gamma\sigma^{n-1}(z_2) & \cdots & \sigma^{n-1}(z_0) \end{pmatrix}$$

The n^2 information symbols $s_{l,m}$ are encoded into codewords by $z_l = \sum_{m=0}^{n-1} s_{l,m} b_m, l = 0,1,\cdots,n-1$ where $\{b_m\}_{m=0}^{n-1}$ is a basis of K. We may restrict the elements of A to $Z_i \in I \subset K$ so that the signal constellation on each layer is a finite subset of the rotated versions of the lattices Z^{2n} or Λ_2^n. Hence,

$$C_I = \begin{pmatrix} z_0 & z_1 & \cdots & z_{n-1} \\ \gamma\sigma(z_{n-1}) & \sigma(z_0) & \cdots & \sigma(z_{n-2}) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma\sigma^{n-1}(z_1) & \gamma\sigma^{n-1}(z_2) & \cdots & \sigma^{n-1}(z_0) \end{pmatrix}; Z_i \in I \subseteq K, i = 0,1,\cdots,n-1$$

Choose γ such that none of its power is a norm in L^* to get a non-vanishing determinant (NVD). $\det\left(\mathbf{C}_i - \mathbf{C}_j\right) \neq 0; \mathbf{C}_i \neq \mathbf{C}_j \in \mathbf{C}_I$. For linear codes, it can further be simplified as

$\Rightarrow \det\left(\mathbf{C}\right) \neq 0; \mathbf{0} \neq \mathbf{C} \in \mathbf{C}_I$.

Also choose $|\gamma| = 1$ to guarantee the same average energy is sent from each antenna for each channel use. This limits the choice to

$\gamma \in \{\pm 1, \quad \pm i\} \in Z[i]$ or $\gamma \in \left\{1, \quad j, \quad j^2, \quad -1, \quad -j, \quad j^2\right\} \in Z[j]$.

Let us summarize:

- Cyclic algebra yields a full-rate linear code.
- Cyclic division algebra fulfils the rank criterion.
- The restriction for $Z_i \in I$ results in good shaping and produce energy efficient codes.
- The choice of $|\gamma| = 1$ makes sure that the same average energy is transmitted from each antenna.
- The property of NVD gives codes which are optimal in terms of the diversity-multiplexing trade-off.

The 2 × 2 PSTBC also known as Golden code is a finite subset of the CDA of degree 2

$$A\left(K = Q(i,\theta) / L = Q(i), \sigma, \gamma = i\right) \text{ for Galois group } G_{K/L} = \langle \sigma \rangle$$

where, $\sigma : \theta = \dfrac{1+\sqrt{5}}{2} \mapsto \overline{\theta} = \dfrac{1-\sqrt{5}}{2}$.

The ring of integers of K is defined as $O_K = \left\{a + b\theta \mid a, b \in Z(i)\right\}$

where, θ is root of polynomial $x^2 - x - 1 = 0$ and also called as Golden number.

Note that Z denotes the ring of rational integers. We can construct the cyclic algebra for any $z_1, z_2 \in K$ as $a_1 = z_1 + ez_2$ where $e^2 = \gamma$ and $ze = e\sigma(z) = e\sigma(a + b\theta) = e(a + b\overline{\theta})$. The corresponding matrix for $a_1 = z_1 + ez_2$ is

$$A = \left\{\mathbf{X}_a = \begin{bmatrix} z_1 & z_2 \\ \gamma\sigma(z_2) & \sigma(z_1) \end{bmatrix}\right\}; z_1, z_2 \in K$$

Let a be an element of a cyclic algebra A. Then the determinant of its corresponding matrix ($\mathbf{X}_a$) is called reduced norm of a. Hence the reduced norm is defined as $N_r(a) = \det(\mathbf{X}_a)$. An ideal I (F. Oggier et al., 2007) of a commutative ring A is an additive subgroup of A which is stable under multiplication by A, i.e., $aI \subseteq I \; \forall \; a \in A$. An ideal I is principal if it is of the form $I_\alpha = (\alpha)A = \{\alpha y, y \in A\}$, $a \in I$ usually written as $I_\alpha = (\alpha)$. Let $I_\alpha = (\alpha)$ be the principal ideal generated by $\alpha = 1 + i - i\theta$. Then, the Golden code (2 × 2 PSTBC) can be constructed as follows:

$$g = \frac{1}{\sqrt{5}}\left(z_1 + ez_2\right); z_1, z_2 \in I_\alpha$$

where, $e^2 = \gamma = i$ and $ze = e\sigma(z) = e\sigma(a + b\theta) = e\left(a + b\overline{\theta}\right)$

The matrix construction is

$$\mathbf{G} = \frac{1}{\sqrt{5}}\begin{bmatrix} z_1 & z_2 \\ i\sigma(z_2) & \sigma(z_1) \end{bmatrix}; z_i = \alpha(x_i + y_i\theta) \in K; x_i, y_i \in Z[i]$$

It can be observed that **G** is a linear code. In the above equation, x_1, x_2, y_1, y_2 are information symbols taken from M-QAM constellation carved from $Z[i]$. The decoding of algebraic space-time codes including perfect space-time codes can be carried out using sphere decoding, which will be discussed in the next chapter. Let us find the minimum determinant of the Golden code as follows.

$$\delta_{\min}(\mathbf{G}) = \min_{\substack{\mathbf{X} \in \mathbf{G} \\ \mathbf{X} \neq \mathbf{0}}} \left|\det(\mathbf{X})\right| = \frac{1}{25}\left|N_r(\alpha)\right|_{\alpha=1+i-i\theta} = \frac{1}{25}|2+i|^2 = \frac{1}{5} \neq 0$$

The rate, r PSTBC for a MIMO system with N_T transmit antennas (P. Elia et. al., 2007) can be constructed as

$$\mathbf{C}_{N_T}^{PSTBC} = \frac{1}{\sqrt{\lambda}}\sum_{k=1}^{r}\mathbf{D}_k\mathbf{\Gamma}^{k-1}$$

where, $\mathbf{D}_k = diag\left(z_k, \sigma(z_k), \sigma^2(z_k), \cdots, \sigma^{N_T-1}(z_k)\right)$

λ is a suitable real-valued scalar designed so that the STBC meets the energy constraint

γ is a unit magnitude complex number dependent on N_T

$$\mathbf{\Gamma} = \left(\gamma\mathbf{e}_{N_T}, \mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{N_T-2}, \mathbf{e}_{N_T-1}\right)$$

For rate $r = 2$ and, $N_T = 2$, $\lambda = 5$, we have

$$\mathbf{C}_2^{PSTBC} = \frac{1}{\sqrt{5}}[\mathbf{D}_1 + \mathbf{D}_2\mathbf{\Gamma}] = \frac{1}{\sqrt{5}}\left[\begin{pmatrix} z_1 & 0 \\ 0 & \sigma(z_1) \end{pmatrix} + \begin{pmatrix} z_2 & 0 \\ 0 & \sigma(z_2) \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \gamma & 0 \end{pmatrix}\right]$$

$$\Rightarrow \quad \mathbf{C}_2^{PSTBC} = \frac{1}{\sqrt{5}}\begin{bmatrix} z_1 & z_2 \\ \gamma\sigma(z_2) & \sigma(z_1) \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} z_1 & z_2 \\ i\sigma(z_2) & \sigma(z_1) \end{bmatrix}$$

which is our 2 × 2 Golden code.

Review question 8.12 *Are DAST, TAST and PSTBC fully diverse code?*

Review question 8.13 *What are the achievable rates for DAST, TAST and PSTBC?*

Review question 8.14 *What is NVD? Give some examples of STBCs having NVD characteristics.*

Review question 8.15 *What are properties of PSTBC which make it the most desirable STBC?*

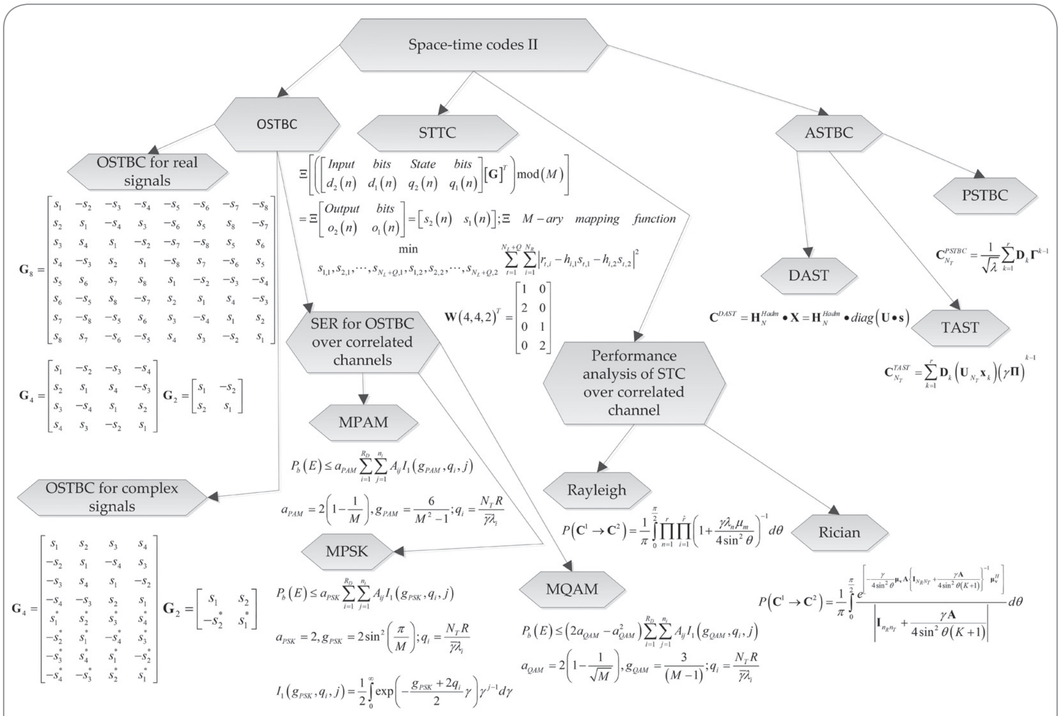


Fig. 8.8 Chapter 8 in a nutshell

8.8 Summary

Figure 8.8 shows the chapter in a nutshell. In this chapter, we have studied about the space-time block and trellis codes. Even though, STTC comes up in the literature before STBC, researchers are spending more time in finding better STBC in terms of rate and diversity. The main reason being STTC are very difficult to decode. The first STBC is Alamouti space-time code which was later extended to orthogonal STBC. OSTBC has the benefit of simple ML decoding complexity since the receiver can decouple all the transmitted symbols. But the main issue with OSTBC is that the code rate of OSTBC tends to $\frac{1}{2}$ as N_T increases for complex signal constellations. Many STBCs have come up in the literature to overcome this issue. First attempt was quasi-orthogonal STBC in which orthogonality is compromised to get unity rate code. Such codes have higher ML decoding complexity since they are quasi-orthogonal. We have done a detailed analysis for finding the SER of OSTBC and STC over correlated Rayleigh faded MIMO channel. We have also mentioned about the space-time Turbo codes. We have discussed briefly about the non-coherent space-time codes viz., differential space-time codes. One of the recent developments in STBC is the algebraic space-time codes. Algebraic space-time codes, as the name suggests, are designed using algebraic structures. Among the ASTBC, we have discussed about the Diagonal algebraic space-time codes (unity rate and full diversity codes), threaded algebraic space-time codes (rate $r \geq 1$ and full diversity codes) and finally perfect space-time codes (rate $r \geq 1$ and full diversity codes).

Exercises

The following exercises may not have close-form formulae, but may be explored.

Exercise 8.1

Find the PEP bound for space-time code (STC) over i.i.d. classical fading MIMO channels
(a) Nakagami-m (b) Rice
(c) Hoyt (d) Weibull

Exercise 8.2

Find the exact PEP for STC over i.i.d. classical fading MIMO channels
(a) Nakagami-m (b) Rice
(c) Hoyt (d) Weibull

Exercise 8.3

Find the PEP bound for STC over i.i.d. generalized fading MIMO channels
(a) α-μ (b) k-μ
(c) η-μ

Exercise 8.4

Find the exact PEP for STC over i.i.d. generalized fading MIMO channels
(a) α-μ (b) k-μ
(c) η-μ

Exercise 8.5

Find the SER of OSTBC for PAM/PSK/QAM for uncorrelated Rayleigh fading MIMO channel.

Exercise 8.6

Find the SER of OSTBC for PAM/PSK/QAM for spatially correlated Rice fading MIMO channel.

Exercise 8.7

Find the SER of OSTBC for PAM/PSK/QAM for uncorrelated η-μ fading MIMO channel.

Exercise 8.8

Find the SER of OSTBC for PAM/PSK/QAM for uncorrelated Nakagami-m fading MIMO channel.

Exercise 8.9

Find the SER of OSTBC for PAM/PSK/QAM for spatially correlated k-μ fading MIMO channel.

Exercise 8.10

Find the SER of OSTBC for BPSK/DBPSK/MPSK/MQAM/MSK of Alamouti space-time codes over i.i.d. generalized fading MIMO channels:

(a) α-μ (b) k-μ

(c) η-μ

The following are MATLAB based exercises.

Exercise 8.11

Write a MATLAB program for implementing Baro's STTC. It will have an encoder using the generator sequence (**B**(4,4,2)) listed in Table 8.2. Decoding could be done using Maximum likelihood (ML) based Viterbi algorithm. Plot frame error rate vs SNR (one frame could be considered for 130 symbols). Assume Rayleigh i.i.d. MIMO fading channel.

Exercise 8.12

Write a MATLAB program for implementing Yan's STTC. It will have an encoder using the generator sequence (**Y**(4,8,2)) listed in Table 8.2. Decoding could be done using ML based Viterbi algorithm. Plot frame error rate vs SNR (one frame could be considered for 130 symbols). Assume Rayleigh i.i.d. MIMO fading channel.

Exercise 8.13

Write a MATLAB program implementing Tarokh's STTC. It will have an encoder using the generator sequence (**T**(4,16,2)) listed in Table 8.2. Decoding could be done using ML based Viterbi algorithm. Plot frame error rate vs SNR (one frame could be considered be for 130 symbols). Assume Rayleigh i.i.d. MIMO fading channel.

References

1. Ahmed, B. and M. A. Matin. 2015. *Coding for MIMO-OFDM in Future Wireless Systems*. London: Springer.
2. Ahn, K. S. and H. K. Baik. Sep. 2007. 'Asymptotic performance and exact symbol error probability analysis of orthogonal STBC in spatially correlated Rayleigh MIMO channel'. *IEICE Trans. Fundamentals*. E-90A(9). 1965–1975.
3. Alamouti, S. M. 1998. 'A simple transmit diversity technique for wireless communications'. *IEEE Journal on Selected Areas in Communications*. 16(8). 1451–1458.
4. Baro, S., G. Baush, and A. Hansmann. 2000. 'Improved Codes for space-time trellis coded modulation'. *IEEE Communications Letters*. 4(1). 20–22.
5. Berrou, C., A. Glavieux, and P. Thitimajshima. 1993. 'Near Shannon limit error-correcting coding and decoding: turbo-codes'. In *Proc. International Conference on Communications*. Geneva, Switzerland. 1064–1070.

6. Biglieri, E. 2005. *Coding for Wireless Channels*. New York: Springer.
7. Bose, R. 2008. *Information Theory, Coding and Cryptography*. New Delhi: Tata McGraw Hill.
8. Cho, Y. S., J. Kim, W. Y. Yang, and C.-G. Kang. 2010. *MIMO-OFDM Wireless Communications using MATLAB*. Singapore: Wiley.
9. Craig, J. 1991. 'A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations'. In *Proc. IEEE MILCOM*. 25.5.1–25.5.5. Boston, MA.
10. Damen, M. O., K. Abeid-Meraim, and J.-C. Belfiore. March 2002. 'Diagonal algebraic space-time codes'. *IEEE Trans. Information Theory*. 48(3). 628–636.
11. Duman, T. M. and A. Ghrayeb. 2007. *Coding for MIMO Communication Systems*. Chichester: John Wiley & Sons.
12. Elia, P., B. A. Sethuraman, and P. Vijay Kumar. Nov. 2007. 'Perfect space-time codes with minimum and non-minimum delay for any number of transmit antennas'. *IEEE Trans. Inform. Theory*. 53(11). 3853–3868.
13. Gamal, H. E. and M. O. Damen. May 2003. 'Universal space-time coding'. *IEEE Trans. Inform. Theory*. 49(5). 1097–1119.
14. Ganesan, G. and P. Stoica. Feb. 2002. 'Differential modulation using space-time block codes'. *IEEE Signal Processing Letters*. 9(2). 57–60.
15. Hedayat, A., H. Shah, and A. Nostrania. Nov. 2005. 'Analysis of space-time coding in correlated fading channels'. *IEEE Trans. Wireless Comm*. 4(6). 2882–2891.
16. Jafarkhani, H. 2005. *Space-time Coding: Theory and Practice*. Cambridge: Cambridge University Press.
17. Jankiraman, M. 2004. *Space-time Codes and MIMO Systems*. Boston: Artech House.
18. Jing, Y. 2013. *Distributed Space-time Coding*. New York: Springer.
19. Johnson, S. J. 2010. *Iterative Error Corrections*. Cambridge: Cambridge University Press.
20. Kreyszig, E. 1999. *Advanced Engineering Mathematics*. New Delhi: John Wiley & Sons.
21. Kuhn, V. 2006. *Wireless Communications over MIMO Channels*. Chichester: Wiley.
22. Kulkarni, M., L. Choudhary, B. Kumbhani, and R. S. Kshetrimayum. July 2014. 'Performance analysis comparison of TAS/MRC and OSTBC in equicorrelated Rayleigh fading MIMO channels'. *IET Communications*. 8(10). 1850–1858.
23. Larsson, E. G. and P. Stoica. 2003. *Space-time Block Coding for Wireless Communications*. Cambridge: Cambridge University Press.
24. Lathi, B. P. 2009. *Modern Digital and Analog Communication Systems*. New Delhi: Oxford University Press.
25. Leonard, D. A., C. C. Lindner, K. T. Phelps, C. A. Rodger, and J. R. Wal. 1991. *Coding Theory The Essentials*. New York: Marcel Dekker.
26. Liang, X.-B. October 2003. 'Orthogonal designs with maximal rates'. *IEEE Trans. Inf. Theory*. 49(10). 2468–2503.
27. Moon, T. K. 2005. *Error Correction Coding*. New Delhi: John Wiley & Sons.
28. Oberg, T. 2001. *Modulation, Detection and Coding*. Chichester: John Wiley & Sons.
29. Oestges, C. and B. Clerckx. 2007. *MIMO Wireless Communications*. Oxford: Elsevier.
30. Oggier, F., J.-C. Belfiore, and E. Viterbo. 2007. *Cyclic Division Algebras: A Tool for Space-time Coding*. Boston: Now Publishers.
31. Proakis, J. G. and M. Salehi. 2008. *Digital Communications*. New York: McGraw-Hill.
32. Rajan, B. S. 2014. 'Space-time block codes'. In *Channel Coding: Theory, Algorithms and Applications*, D. Declerq, M. Fossorier and E. Biglieri, Eds., Oxford, UK: Elsevier Academic Press. 451–495.
33. Saff, E. B. and A. D. Snider. 2003. *Fundamentals of Complex Analysis*. Noida: Pearson.
34. Simon, M. K. and M.-S. Alouini. 2005. *Digital Communications over Fading Channels*. New Jersey: Wiley.

35. Simon, M. K. 'A moment generating function (MGF)-based approach for performance evaluation of space-time coded communication systems'. Nov. 2002. *Wireless Communication and Mobile Computing.* 2(7). 667–692.
36. Tarokh, V., H. Jafarkhani, and A. R. Calderbank. 1999. 'Space-time block coding for wireless communications: performance results'. *IEEE Journal on Selected Areas in Communications.* 17(3). 451–460.
37. Tarokh, V., N. Seshadri, and A. Calderbank. 1998. 'Space-time codes for high data rate wireless communication: performance criterion and code construction'. *IEEE Transactions on Information Theory.* 44(2). 744–765.
38. Vucetic, B. and J. Yuan. 2003. *Space-time Coding*. Chichester: John Wiley & Sons.
39. Wilson, S. G. 1996. *Digital Modulation and Coding.* Noida: Pearson.
40. Wittneben, A. 1993. 'A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation'. In *Proc. IEEE International Conference on Communications*. Geneva, Switzerland. 1630–1634.
41. Yan, Q. and R. S. Blum. 2000. 'Optimum space-time convolutional codes'. In *Proc. Wireless Communications and Networking Conference*. Chicago, USA.
42. Yuen, C., Y. L. Guan, and T. Tjhung. 2007. *Quasi-orthogonal Space-time Block Codes.* London: Imperial College Press.

Introduction to MIMO Detection

CHAPTER 9

9.1 Introduction

In MIMO detection, we need to detect signals jointly since many signals are transmitted from the transmitter to the receiver. For instance, consider a 2×1 MIMO system with two transmit antennas and one single receive antenna. Two antennas are transmitting two signals at the same time; hence the receiving antenna receives both signals. Hence, we need to detect both the signals jointly. Among the available detection techniques, Maximum Likelihood (ML) detection is the optimal technique, but its complexity grows exponentially with the number of antennas. There are other sub-optimal techniques like Zero Forcing (ZF) and Minimum Mean Square Error (MMSE) which are less complex. We will study first how to implement those techniques. We will compare the noise amplification in ZF and MMSE. Then we will find the performance of these techniques in terms of probability of error and outage probability. We also discuss about Sphere Decoding (SD) in the last section of this chapter. SD is less complex than ML but has similar performance with that of ML.

9.2 Maximum likelihood (ML) detector

Let us consider a $N_T \times N_R$ MIMO system whose I–O relation (matrix form) at any symbol time t for frequency flat fading is given by

$$\mathbf{r}_t = \mathbf{H}_t \mathbf{s}_t + \mathbf{n}_t \tag{9.1}$$

where symbol time slot $t = 1, 2, \ldots, N_L$ and N_L may be considered as frame or packet length.

In component form, it can be expressed as

$$\begin{bmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N_R,t} \end{bmatrix} = \begin{bmatrix} h_{11,t} & h_{12,t} & \cdots & h_{1N_T,t} \\ h_{21,t} & h_{22,t} & \cdots & h_{2N_T,t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1,t} & h_{N_R2,t} & \cdots & h_{N_RN_T,t} \end{bmatrix} \begin{bmatrix} s_{1,t} \\ s_{2,t} \\ \vdots \\ s_{N_R,t} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ \vdots \\ n_{N_R,t} \end{bmatrix} \tag{9.2}$$

Basically, we estimate the transmitted signal vector ***s*** from the known received vector ***r*** for the given channel matrix ***H***. ML is optimal in performance when the input symbol alphabet *S* consists of equi-probable symbols which are a reasonable assumption for real systems. But its complexity has exponential growth as we will see since it involves brute force search over all possible combinations. Hence it is not feasible to employ ML detectors at the Mobile Station (MS) of a cellular network since MS are small and computationally extensive signal processing is not possible. Let us denote $s_{i,t}$ is the data symbol transmitted from the ith transmit antenna at the symbol time *t* and $s_{i,t} \in S$, $i = 1, 2, \ldots, N_T$. Also $\arg \underset{s}{\min} f(s)$ means that among all possibilities of *s*, that particular *s* which will minimize the function *f*(*s*). ML detection outputs the vector which minimizes the Euclidean distance between the received vector and all possible combinations of the transmitted symbol vectors.

$$\hat{\mathbf{s}} = \arg \underset{\mathbf{x}}{\min} \left\| \mathbf{r} - \mathbf{Hs} \right\|^2 \tag{9.3}$$

Equivalently, ML detection is to find the best symbol vector that maximizes the likelihood function as

$$\hat{s}_{i,t} = \underset{s \in S^{N_T}}{\arg \max} \; f(\mathbf{r} \mid \mathbf{s}) \tag{9.4}$$

where, $f(\mathbf{r}|\mathbf{s})$ is the likelihood function for **s**, for the given received vector **r**. Note that $\mathbf{n}_t$ is circular symmetric complex Gaussian noise vector. Hence the likelihood function will be complex multivariate Gaussian distributed.

$$f(\mathbf{r} \mid \mathbf{s}) = \frac{1}{\pi \mathbf{R}_n} \exp\left(-(\mathbf{r} - \mathbf{Hs})^H \mathbf{R}_n^{-1} (\mathbf{r} - \mathbf{Hs})\right) \tag{9.5}$$

Maximizing a negative exponential function is equivalent to minimizing its argument as follows.

$$\hat{s}_{i,t} = \underset{s \in S^{N_T}}{\arg \min} \; (\mathbf{r} - \mathbf{Hs})^H \mathbf{R}_n^{-1} (\mathbf{r} - \mathbf{Hs}) \tag{9.6}$$

Example 9.1

Explain the ML detection for a 2 × 2 MIMO system.

Solution

Consider a 2 × 2 MIMO system at time instant t. We have the received signal, channel matrix, transmitted signal and noise vector as follows:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}; \mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}; \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}; \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

Now we can write the received signal vector in terms of the channel matrix, transmitted signal and noise vector for frequency flat fading as follows

$$\mathbf{r} = \mathbf{Hs} + \mathbf{n}$$

$$\Rightarrow \qquad \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

Hence, $r_1 = h_{11}s_1 + h_{12}s_2 + n_1; r_2 = h_{21}s_1 + h_{22}s_2 + n_2$

At the detector, we want to detect s_1 and s_2 at time t, but there exists interference between these two signals for both the receiving antennas. Optimal receiver for this case is the ML receiver. Assume that s_k are modulated in M-ary constellation i.e., $s_k \in \{s_1, s_2, \cdots, s_M\}$. We need to find the minimum metric of the Euclidean distance

$$\min_{i, j \in \{1,2,\cdots,M\}} \left[\left\| r_1 - \left(h_{11}s_i + h_{12}s_j\right)\right\|^2 + \left\| r_2 - \left(h_{21}s_i + h_{22}s_j\right)\right\|^2 \right]$$

For instance, 16-QAM, (s_1, s_2) are (1 of 16 symbols, 1 of 16 symbols) implies 16 × 16 pairs. Metric calculations of 256 are required. For 3 × 3 MIMO system, $N_T = N_R = 3$, metric calculations of 16^3 = 4096 are required. For 5 × 5 MIMO system, $N_T = N_R = 5$, metric calculations of 16^5 = 10,48,576 are required, which is obviously impractical. This gets more worst if we consider large MIMO systems where we consider hundreds of transmitter and receiver antennas.

The decoding complexity increases exponentially $|S|^{N_T} = M^{N_T}$ with the number of transmit antennas (N_T) and constellation size (M). A minor simplification of ML decoding is sphere decoding (SD). It tries to find the transmitted signal vector by comparing only signal vectors within the radius of a sphere. If there are no signal vectors within the sphere, it increases the sphere size. If there are many signal vectors within the radius of the sphere, it will reduce the sphere radius. One may explore further on sphere decoding in B. M. Hochwald et al., (2003) and it will be explored in the last section of this chapter.

Review question 9.1 *How does the complexity of ML increases?*

9.2.1 Performance analysis

Let us try to find the PEP for detecting $\mathbf{s}_2$ when the signal vector transmitted was $\mathbf{s}_1$.

$$P\left(\mathbf{s}_1 \to \mathbf{s}_2\right) = \Pr ob\left(\left\|\mathbf{y} - \mathbf{Hs}_2\right\|^2 \le \left\|\mathbf{y} - \mathbf{Hs}_2\right\|^2\right)$$

Define $$\mathbf{d} = \mathbf{s}_1 - \mathbf{s}_2$$

Note that $$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_R} \end{bmatrix}; \mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{bmatrix}; \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{N_T} \end{bmatrix}$$

The PEP can be calculated as

$$P\left(\mathbf{s}_1 \to \mathbf{s}_2\right) = Q\left(\sqrt{\frac{\|\mathbf{Hd}\|^2}{2N_0}}\right) \tag{9.7}$$

Using Chernoff's bound (neglecting the 1/2 factor), PEP is bounded as

$$P\left(\mathbf{s}_1 \to \mathbf{s}_2\right) \le \exp\left(-\frac{\|\mathbf{Hd}\|^2}{4N_0}\right) \tag{9.8}$$

We need to vectorize **Hd** matrix so that we can apply the above theorem.

$\because$ $\quad vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\, vec(\mathbf{B})$

This theorem could be proved (see Exercise 9.2). But we will use the following corollary of this theorem.

Corollary:

(a) When we assume that the last matrix is **I**, we have,

$$vec(\mathbf{AB}) = vec(\mathbf{ABI}) = (\mathbf{I}^T \otimes \mathbf{A})\, vec(\mathbf{B}) = (\mathbf{I} \otimes \mathbf{A})vec(\mathbf{B})$$

(b) When we assume that the first matrix is **I**, we have,

$$vec(\mathbf{AB}) = vec(\mathbf{IAB}) = (\mathbf{B}^T \otimes \mathbf{I})\, vec(\mathbf{A})$$

$\therefore$ $\quad vec(\mathbf{Hd}) = (\mathbf{d}^T \otimes \mathbf{I}_{N_R})\, vec(\mathbf{H})$

$\Rightarrow$ $\quad \mathbf{h} \sim N_c\,[0, (\mathbf{d}^T \otimes \mathbf{I}_{N_R})^H (\mathbf{d}^T \otimes \mathbf{I}_{N_R})]$

Therefore, the average PEP with respect to **h** is given by

$$\begin{aligned}
&< P\left(\mathbf{s}_1 \to \mathbf{s}_2\right) > \\
&\le E\left[\exp\left(-\frac{\|\mathbf{Hd}\|^2}{4N_0}\right)\right] \\
&\le E\left[\exp\left(-\frac{(\mathbf{Hd})^H\,\mathbf{Hd}}{4N_0}\right)\right] \\
&= E\left[\exp\left(-\frac{\left(vec(\mathbf{H})\right)^H \left(\mathbf{d}^T \otimes \mathbf{I}_{N_R}\right)^H \left(\mathbf{d}^T \otimes \mathbf{I}_{N_R}\right) vec(\mathbf{H})}{4N_0}\right)\right]
\end{aligned}$$

Also we know that,

$\because$ $\quad (\mathbf{A} \otimes \mathbf{B})\,(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$

$\therefore$ $\quad \mathbf{d}^*\mathbf{d}^T \otimes (\mathbf{I}_{N_R})^H\,\mathbf{I}_{N_R} = \mathbf{d}^*\mathbf{d}^T \otimes \mathbf{I}_{N_R}$

$$< P(\mathbf{s}_1 \to \mathbf{s}_2) >$$

$$\leq E\left[\exp\left(-\frac{\left(\left(vec(\mathbf{H})\right)^{H}\left(\mathbf{d}^{*}\mathbf{d}^{T}\otimes\mathbf{I}_{N_R}\right)vec(\mathbf{H})\right)}{4N_0}\right)\right] \tag{9.9}$$

We can show that for a symmetric and positive semi-definite matrix **A** (J. Choi, 2010) and

$$\mathbf{h} \sim N_c(0,\mathbf{R_h}),$$
$$E\left(\exp\left(-\mathbf{h}^{H}\mathbf{A}\mathbf{h}\right)\right) = \det\left(\mathbf{I}+\mathbf{A}\mathbf{R_h}\right)^{-1}$$

You may also see Appendix B. For,

$$\mathbf{A} = \frac{\mathbf{d}^{*}\mathbf{d}^{T}\otimes\mathbf{I}_{N_R}}{4N_0}$$

$$\therefore \quad \left\langle P\left(\mathbf{s}_1\rightarrow\mathbf{s}_2\right)\right\rangle \leq \det\left[\mathbf{I}+\frac{\left(\mathbf{d}^{*}\mathbf{d}^{T}\otimes\mathbf{I}_{N_R}\right)}{4N_0}\right]^{-1} \qquad \because \mathbf{I}\otimes\mathbf{I}=\mathbf{I}$$

$$\therefore \quad \left\langle P\left(\mathbf{s}_1\rightarrow\mathbf{s}_2\right)\right\rangle \leq \det\left[\left(\mathbf{I}+\frac{\mathbf{d}^{*}\mathbf{d}^{T}}{4N_0}\right)\otimes\mathbf{I}_{N_R}\right]^{-1}$$

$$\therefore \quad \left\langle P\left(\mathbf{s}_1\rightarrow\mathbf{s}_2\right)\right\rangle \leq \det\left(\mathbf{I}+\frac{\mathbf{d}^{*}\mathbf{d}^{T}}{4N_0}\right)^{-N_R} \tag{9.10}$$

This expression is very similar to PEP of space-time codes.

Review question 9.2 *Write the expression for Chernoff bound of Q function?*

Review question 9.3 *What is PEP bound for ML detection?*

9.2.2 Diversity gain

From the above equation on the upper bound on PEP we can say that the diversity gain of the ML detection is N_R for a zero mean circular symmetric complex Gaussian (ZMCSCG) channel. Another alternative MIMO detection technique is to employ simpler and easy to implement linear detectors but they have poorer performance.

Review question 9.4 *What is the diversity gain of ML detection?*

9.3 Linear sub-optimal detectors

ML detectors are optimal but impractical. Low complexity suboptimal detectors like zero forcing (ZF) and Minimum mean square error (MMSE) are preferable.

In linear detector, a linear pre-processor (**W**) is first applied to the received signal vector $\hat{\mathbf{s}} = \mathbf{W}^H\mathbf{r}$. Then each element of estimate ($\hat{\mathbf{s}}$) is considered as the received signal in the absence of other signals and from which the associated signal is independently detected. It is a two-step approach. It consists of a linear pre-processor (**W**) and N_R (single–signal) detectors. Hence the complexity grows linearly with N_R.

9.3.1 ZF detector

Consider the received signal in the i^{th} antenna given by

$$r_i = \begin{bmatrix} h_{i,1} & h_{i,2} & \cdots & h_{i,N_T} \end{bmatrix} \mathbf{s} + n_i \tag{9.11}$$

If we assume that k[th] stream is the desired signal, then, we can express the above received signal in the following way.

$$r_i = h_{i,k}s_k + \sum_{j=1, j\neq k}^{N_T} h_{i,j}s_j + n_i \tag{9.12}$$

If we want to suppress the interference then we need to project the received signal onto a subspace which is orthogonal to the interference. ZF detector will de-correlate the desired stream from the other streams. In ZF detector, the linear pre-processor suppresses the other signals completely. The pre-processor output is given by

$$\hat{\mathbf{s}} = \mathbf{W}_{ZF}^H\mathbf{r} = \mathbf{H}^{+}\mathbf{r} = \mathbf{s} + \mathbf{H}^{+}\mathbf{n} \tag{9.13}$$

where, $\mathbf{W}_{ZF}^H = \mathbf{H}^{+} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ is the Moore Penrose pseudo-inverse of **H**

Review question 9.5 *Explain ZF detection.*

Example 9.2

Show that for ZF $\mathbf{W}_{ZF}^H = \mathbf{H}^{+} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$.

Solution

Note that the ZF searches for unconstrained vector $\mathbf{s} \in C^{N_T}$ (not constrained to alphabet S) that minimizes the squared Euclidean to the received vector **r** as

$$\underset{\mathbf{s} \in C^{N_T}}{\arg\min} \; \left\|\mathbf{r} - \mathbf{H}\mathbf{s}\right\|^2$$

This can be done by taking partial derivative $\|\mathbf{r}-\mathbf{Hs}\|^2$ w.r.t. $\mathbf{s}^H$ and setting to $\mathbf{0}$ as follows.

$$\frac{\partial}{\partial \mathbf{s}^H}(\mathbf{r}-\mathbf{Hs})^H(\mathbf{r}-\mathbf{Hs}) = \frac{\partial}{\partial \mathbf{s}^H}\left(\mathbf{r}^H\mathbf{r}-\mathbf{r}^H\mathbf{Hs}-\mathbf{s}^H\mathbf{H}^H\mathbf{r}+\mathbf{s}^H\mathbf{H}^H\mathbf{Hs}\right)$$

$$= -\mathbf{H}^H\mathbf{r}+\mathbf{H}^H\mathbf{Hs}$$

Then we obtain

$$\mathbf{H}^H\mathbf{Hs} = \mathbf{H}^H\mathbf{r}$$

$$\Rightarrow \qquad \mathbf{s} = (\mathbf{H}^H\mathbf{H})^{-1}\,\mathbf{H}^H\mathbf{r}$$

$$\therefore \qquad \mathbf{W}_{ZF}^H = \mathbf{H}^+ = (\mathbf{H}^H\mathbf{H})^{-1}\,\mathbf{H}^H$$

It can be seen that in the ZF detector the spatial interference has been wiped out completely from the received signal and hence the name zero forcing. In order for the pseudo-inverse to exist, N_T must be less than or equal to N_R otherwise $\mathbf{H}^H\mathbf{H}$ is singular and its inverse does not exist (G. Strang, 2006).

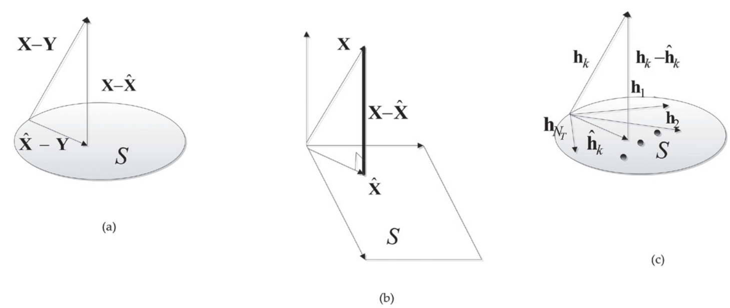


Fig. 9.1 (a) Projection in Hilbert space (b) Projection in 3-D case and (c) Projection in ZF

Let S be subspace of Hilbert space H. For a given element $\mathbf{X}\in$H, we are interested in finding an element of S that best matches $\mathbf{X}$ which we will call as projection of $\mathbf{X}$ and denote it as $\hat{\mathbf{X}} \in S$. The projection error is $\|\mathbf{X}-\hat{\mathbf{X}}\|$. The projection $\hat{\mathbf{X}} \in S$ is "closest" to $\mathbf{X}$ onto S which satisfies $\|\mathbf{X}-\hat{\mathbf{X}}\| = \min_{\mathbf{Y}\in S}\|\mathbf{X}-\mathbf{Y}\|$ iff $\mathbf{X}-\hat{\mathbf{X}} \perp \mathbf{Y}, \forall \mathbf{Y} \in \mathbf{S}$.

Note that (see Fig. 9.1 a)

$$\|\mathbf{X}-\mathbf{Y}\|^2 = \|\mathbf{X}-\hat{\mathbf{X}}+\hat{\mathbf{X}}-\mathbf{Y}\|^2 = \|\mathbf{X}-\hat{\mathbf{X}}\|^2+\|\hat{\mathbf{X}}-\mathbf{Y}\|^2+2\,\mathrm{Re}\left\{(\mathbf{X}-\hat{\mathbf{X}}),(\hat{\mathbf{X}}-\mathbf{Y})\right\}$$

Using projection theorem, $2\,\mathrm{Re}\left\{(\mathbf{X}-\hat{\mathbf{X}}),(\hat{\mathbf{X}}-\mathbf{Y})\right\} = 0$

Therefore, we have the Pythagoras theorem,

$$\|\mathbf{X}-\mathbf{Y}\|^2 = \|\mathbf{X}-\hat{\mathbf{X}}\|^2+\|\hat{\mathbf{X}}-\mathbf{Y}\|^2$$

Hence the RHS sum is minimized when $\mathbf{Y} = \hat{\mathbf{X}}$.

It will be clearer, if we consider the 3-D case. Assume that x- and y-axis form the 2-D subspace S. **X** is an arbitrary vector in 3-D plane, we need to find the projection of **X** onto the 2D xy-plane. It basically means dropping a perpendicular line from **X** to the 2-D xy-plane. This projection (thick line in Fig. 9.1b) is perpendicular to xy-plane and therefore orthogonal to every vector in S (xy-plane).

Geometrical interpretation of ZF:

Let us assume that the desired symbol index is k, then, this desired symbol will be modulated by $\mathbf{h}_k$ column of channel matrix **H**. The remaining columns will be modulated by the interfering symbols. One may define the interfering subspace span by the columns of **H** matrix $\mathbf{h}_l$ where $l \neq k$. In that case, we may write $\mathbf{h}_k$ in terms of a vector of interference subspace and one that is orthogonal to it.

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \left(\mathbf{h}_k - \hat{\mathbf{h}}_k\right)$$

where, $\hat{\mathbf{h}}_k$ denotes the projection of $\mathbf{h}_k$ onto the interference subspace and $\left(\mathbf{h}_k - \hat{\mathbf{h}}_k\right)$ is the projection error and perpendicular to the $\hat{\mathbf{h}}_k$ as illustrated in Fig. 9.1 (c).

ZF detector should discard the first term to null the interference. Then it retains the second term so that linear pre-processing vector is chosen as

$$\mathbf{w}_k = \frac{\mathbf{h}_k - \hat{\mathbf{h}}_k}{C}$$

Note that normalizing constant C can be taken as $C = \left\|\mathbf{h}_k - \hat{\mathbf{h}}_k\right\|$

Example 9.3

Let us consider a 2 × 2 MIMO system with $s \in S = \{-3, -1, 1, 3\}$ and $N_T = N_R = 2$. The channel matrix is given by $H = \begin{bmatrix} 2 & 0.5 \\ 1 & 2 \end{bmatrix}$. Suppose the received signal vector is $\mathbf{r} = \begin{bmatrix} 1 \\ 0.9 \end{bmatrix}$. What is the ZF detector's output?

Solution

The ZF detector's output is given by

$$\hat{\mathbf{s}} = \mathbf{H}^{+}\mathbf{r} = [0.5 \quad 0.2]^T$$

Thus the hard decision of for **s** becomes $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ which is different from the ML decision of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Example 9.4

What happens to noise power for ZF?

Solution

Let us denote noise after ZF as

$$\mathbf{H}^{+}\mathbf{n} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{n} = \mathbf{z}$$

The error performance of MIMO detection is directly related with the power of $\mathbf{H}^{+}\mathbf{n}$ or $\left\|\mathbf{H}^{+}\mathbf{n}\right\|_2^2$.

Note that for singular matrix, some of the singular values are zero. For ill-conditioned matrix, we can set some threshold t and select those eigenvalues whose value is greater than t and ignore other eigenvalues in calculation of inverse of a matrix.

Note that **U**, **V** and **Q** in the following expression are unitary and $\mathbf{H} = (\mathbf{U} \quad \boldsymbol{\Sigma} \quad \mathbf{V}^H)$. Using the SVD, the post-detected noise power is

$$
\begin{aligned}
\|\mathbf{z}\|_2^2 &= \left\|\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}\right\|_2^2 = \left\|\left(\mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^H\right)^{-1}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^H\mathbf{n}\right\|_2^2 \\
&= \left\|\mathbf{V}\boldsymbol{\Sigma}^{-2}\mathbf{V}^H\mathbf{V}\,\boldsymbol{\Sigma}\,\mathbf{U}^H\mathbf{n}\right\|^2 = \left\|\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\mathbf{n}\right\|^2 \\
\because \quad \|\mathbf{Q}\mathbf{x}\|_2^2 &= \mathbf{x}^H\mathbf{Q}^H\mathbf{Q}\mathbf{x} = \mathbf{x}^H\mathbf{x} = \|\mathbf{x}\|_2^2 \\
\therefore \quad E\left\{\|\mathbf{z}\|_2^2\right\} &= E\left\{\left\|\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\mathbf{n}\right\|_2^2\right\} = E\left\{tr\left(\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\mathbf{n}\mathbf{n}^H\mathbf{U}\boldsymbol{\Sigma}^{-1}\right)\right\} \\
&= tr\left\{\boldsymbol{\Sigma}^{-1}E\left(\mathbf{n}\mathbf{n}^H\right)\mathbf{U}^H\mathbf{U}\boldsymbol{\Sigma}^{-1}\right\} = tr\left\{\boldsymbol{\Sigma}^{-2}\sigma_n^2\right\} = \sigma_n^2 tr\left\{\boldsymbol{\Sigma}^{-2}\right\} \\
&= \sum_{i=1}^{N_T}\frac{\sigma_n^2}{\sigma_i^2} \approx \frac{\sigma_n^2}{\sigma_{\min}^2} \qquad (9.14)
\end{aligned}
$$

Looking at the above equation, one can infer that for not well behaved channel matrix, $\sigma_{\min}^2$ is very small and hence $\frac{\sigma_n^2}{\sigma_{\min}^2}$ will be a large number. The main hurdle with linear detector is that noise power is getting amplified due to application of the linear pre-processor (**W**) for ill-behaved channel matrix. In order to overcome such hurdles, one can employ techniques like lattice reduction (LR) which will be discussed in the next chapter.

9.3.1.1 *Outage probability and diversity gain*

SINR for ZF is given in Eq. (9.15). SINR is distributed Chi-square with $2(N_R - N_T + 1)$ degrees-of-freedom (Example 9.6). Outage probability and diversity gain are derived in Example 9.7 and 9.8, respectively.

Example 9.5

Find the SINR expression for ZF.

Solution

The post-detected noise $\mathbf{H}^{+}\mathbf{n} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n} = \mathbf{z}$ is a ZMCSCG with covariance matrix given by

$$\mathbf{R}_{zz} = E\left(\mathbf{z}\mathbf{z}^H\right) = E\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}\mathbf{n}^H\mathbf{H}\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)^H\right)$$

$$= \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H E\left(\mathbf{n}\mathbf{n}^H\right)\mathbf{H}\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)^H$$

$$= \sigma_n^2\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)^H = \sigma_n^2\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$$

If we construct a new matrix by removing kth row of the **H** matrix which we will denote as $\hat{\mathbf{H}}$, therefore, $\mathbf{R}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = \sigma_n^2\left(\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\right)$.

Now we can define the instantaneous signal to interference noise ratio (SINR) for the kth received symbol as

$$SINR_{ZF} = \rho\gamma_{ZF} = \frac{E_k}{\left(\mathbf{R}_{\mathbf{zz}}\right)_{k,k}} = \frac{E_k}{\sigma_n^2\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)_{kk}} \tag{9.15}$$

where, ρ is the mean SNR which has been shown (M. Rupp et al., 2003) to be a Chi-square RV (γ_{ZF} is distributed $\chi_{2(N_R-N_T+1)}$) with $2(N_R - N_T + 1)$ degrees-of-freedom.

Example 9.6

Find the distribution of SINR of ZF.

Solution

Let us try to find the distribution of SINR of ZF.

For simplicity, we will analyze for sub-channel 1. One can extend this analysis for any other sub-channel. We can partition the channel matrix **H** as $\mathbf{H} = \left[\mathbf{h}_1\hat{\mathbf{H}}\right]$ where $\mathbf{h}_1$ is the first column vector for the desired sub-channel 1 and $\hat{\mathbf{H}}$ is the matrix after removing the first column. Therefore,

$$\left(\mathbf{H}^H\mathbf{H}\right)^{-1} = \left(\begin{bmatrix}\mathbf{h}_1^H \\ \hat{\mathbf{H}}^H\end{bmatrix}\begin{bmatrix}\mathbf{h}_1 & \hat{\mathbf{H}}\end{bmatrix}\right)^{-1} = \begin{bmatrix}\mathbf{h}_1^H\mathbf{h}_1 & \mathbf{h}_1^H\hat{\mathbf{H}} \\ \hat{\mathbf{H}}^H\mathbf{h}_1 & \hat{\mathbf{H}}^H\hat{\mathbf{H}}\end{bmatrix}^{-1}$$

We have from block matrix inverse formula (H. Lütkepohl, 1996 and T. Kailath, 1980) that if we partition an arbitrary matrix **A** as

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}$$

Then, we can find its inverse as

$$\mathbf{A}^{-1} = \begin{bmatrix}\mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22}\end{bmatrix}$$

where, $$\mathbf{A}^{11} = \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}$$

Hence, $$\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)_{11} = \left(\mathbf{h}_1^H\mathbf{h}_1 - \mathbf{h}_1^H\hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H\mathbf{h}_1\right)^{-1}$$

$$= \left(\mathbf{h}_1^H\left(\mathbf{I} - \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H\right)\mathbf{h}_1\right)^{-1}$$

$$= \left(\mathbf{h}_1^H\left(\mathbf{I} - \hat{\mathbf{P}}_1\right)\mathbf{h}_1\right)^{-1}$$

where, $\hat{\mathbf{P}}_1$ is called the projection matrix for sub-channel 1.

Now $$SINR_{ZF} = \rho\gamma_{ZF} = \frac{E_1}{\left(\mathbf{R_{zz}}\right)_{1,1}} = \frac{E_1}{\sigma_n^2\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)_{11}} = \frac{E_1\left(\mathbf{h}_1^H\left(\mathbf{I} - \hat{\mathbf{P}}_1\right)\mathbf{h}_1\right)}{\sigma_n^2}$$

In order to find the distribution of the positive quadratic form of

$$\alpha = \mathbf{h_1^H}\left(\mathbf{I} - \hat{\mathbf{P}}_1\right)\mathbf{h_1},$$

where, $\left(\mathbf{I} - \hat{\mathbf{P}}_1\right)$ is Hermitian and non-negative.

We can diagonalize the inner matrix by a unitary transformation $\mathbf{Q}$ as

$$\mathbf{Q^H}\boldsymbol{\lambda}\mathbf{Q} = \mathbf{I} - \hat{\mathbf{P}}_1$$

Hence $$\alpha = \mathbf{h_1^H}\mathbf{Q^H}\boldsymbol{\lambda}\mathbf{Q}\mathbf{h_1} = \mathbf{g_1^H}\boldsymbol{\lambda}\mathbf{g_1} = \sum_{i=1}^{N_R}\lambda_i\left|g_i\right|^2$$

where, $\boldsymbol{\lambda}$ is a diagonal matrix $diag\left(\lambda_1 \cdots \lambda_{N_R}\right)$ and λ_i are eigenvalues of $\mathbf{I} - \hat{\mathbf{P}}_1$. Therefore, conditioned on the eigenvalues, the random variable α is a weighted sum-of-squares of Gaussian random variables and the probability distribution may be found out.

$\mathbf{h}_1^H\left(\mathbf{I} - \hat{\mathbf{P}}_1\right)\mathbf{h}_1$ and correspondingly SINR for the sub-channel will have Chi-square distribution with $2(N_R - N_T + 1)$ degrees of freedom.

Example 9.7

Find the outage probability of ZF.

Solution

Consider the separate spatial encoding case (A. Hedayat et al., 2007), the data is de-multiplexed (DMUX) to several sub-streams, each one of them separately encoded and feed to the corresponding transmitting antenna and sent through the channel. For this case, if any one of the data sub-stream is in outage (assume equal data rate for each sub-streams), the whole MIMO system is in outage. The mutual information between the kth transmitted symbol vector $\mathbf{s}_k$ and kth estimated symbol vector $\hat{\mathbf{s}}_k$ at the output of the ZF detector (J. Choi, 2010) could be obtained as

$$I\left(\mathbf{s}_k;\hat{\mathbf{s}}_k\right) = \log_2(1 + SINR_{ZF}) = \log_2(1 + \rho\gamma_{ZF})$$

Therefore, outage probability for a target data rate of R could be obtained as

$$\begin{aligned} P_{\text{out}} &= 1 - \Pr ob\left(\bigcap_{k=1}^{N_T}\left[I\left(\mathbf{s}_k;\hat{\mathbf{s}}_k\right) \geq \frac{R}{N_T}\right]\right) \\ &= 1 - \Pr ob\left(\bigcap_{k=1}^{N_T}\left[\log_2\left(1 + \rho\gamma_{ZF}\right) \geq \frac{R}{N_T}\right]\right) \end{aligned}$$

Assume that the sub-channel outage probabilities are independent and equal. Then, we have,

$$P_{\text{out}} = 1 - \Pr ob\left(\left[\log_2\left(1 + \rho\gamma_{ZF}\right) \geq \frac{R}{N_T}\right]\right)^{N_T}$$

For outage probabilities for sub-channels are small, we have,

$$\begin{aligned} P_{\text{out}} &\approx \Pr ob\left(\left[\log_2\left(1 + \rho\gamma_{ZF}\right) < \frac{R}{N_T}\right]\right)^{N_T} \\ &= N_T \Pr ob\left(\left[\log_2\left(1 + \rho\gamma_{ZF}\right) < \frac{R}{N_T}\right]\right) \end{aligned}$$

Since γ_{ZF} is distributed $\chi_{2(N_R - N_T + 1)}$, we can find outage probability from the CDF as follows.

$$\begin{aligned} P_{\text{out}} &\approx N_T \Pr ob\left(\gamma_{ZF} < \frac{2^{\frac{R}{N_T}} - 1}{\rho}\right) \\ &= N_T\left(1 - e^{-\frac{\left(2^{\frac{R}{N_T}}-1\right)}{\rho}}\sum_{i=1}^{N_R - N_T + 1}\frac{\left(\frac{2^{\frac{R}{N_T}} - 1}{\rho}\right)^{i-1}}{\left(i-1\right)!}\right) \end{aligned} \quad (9.16)$$

We have the CDF of γ_{ZF}; hence, the outage probability is given by the above equation. Therefore it is easy to see that the outage probability decays as $\frac{1}{\rho^{N_R - N_T + 1}}$. The diversity gain for ZF is $N_R - N_T + 1$. The system is underdetermined, if $N_T > N_R$.

Example 9.8

Show that the outage probability for ZF MIMO detection decays as $\frac{1}{\rho^{N_R - N_T + 1}}$.

Solution

Since $\rho\gamma_{ZF}$ is a Chi-square RV (γ_{ZF} is distributed $\chi_{2(N_R - N_T + 1)}$ with $2(N_R - N_T + 1)$ degrees-of-freedom, we have the CDF as in equation (9.16).

Let i goes from 0 to $N_R - N_T$.

Using the infinite series expansion of exponential function, we get,

$$P_{\text{out}} = N_T \left\{ 1 - e^{-\frac{\left(2^{\frac{R}{N_T}}-1\right)}{\rho}} \left\{ e^{\frac{\left(2^{\frac{R}{N_T}}-1\right)}{\rho}} - \sum_{i=N_R-N_T+1}^{\infty} \frac{\left(\frac{2^{\frac{R}{N_T}}-1}{\rho}\right)^i}{(i)!} \right\} \right\}$$

$$= N_T \left\{ e^{-\frac{\left(2^{\frac{R}{N_T}}-1\right)}{\rho}} \sum_{i=N_R-N_T+1}^{\infty} \frac{\left(\frac{2^{\frac{R}{N_T}}-1}{\rho}\right)^i}{(i)!} \right\}$$

For high SNR case ($\rho \to \infty$), we have,

$$\lim_{\rho \to \infty} P_{out} = N_T \left\{ \frac{\left(2^{\frac{R}{N_T}}-1\right)^{N_R-N_T+1}}{\rho^{N_R-N_T+1}\left(N_R - N_T + 1\right)!} \right\} \tag{9.17}$$

Hence the diversity gain is $N_R - N_T + 1$

9.3.1.2 *Performance analysis*

For the performance analysis (exact BER calculation of quadrature phase shift keying (QPSK) modulation) of MIMO systems employing ZF detector over independent and identical distributed (i.i.d.) Rice and Rayleigh fading channel have been carried out by R. Xu et al., (2006). C. Siriteanu et al., (2011) have derived average error probability expression for transmit-correlated Rician fading MIMO channel employing ZF detector. M. Kiessling et al., (2003) also tried to calculate the analytical performance of MIMO zero-forcing receivers in correlated Rayleigh fading environments. Let us find the exact BER of BPSK modulated MIMO systems employing ZF detector over an i.i.d. Rayleigh fading channel.

The post-detection SINR of ZF detector is given by

$$SINR_{ZF} = \rho\gamma_{ZF} = \frac{E_k}{\sigma_n^2\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)_{k,k}}; k = 1,2,\ldots,N_T$$

where, $\rho = \frac{E_k}{\sigma_n^2}$ is the mean SNR. Assume $\mathbf{h}_i$ is the i[th] row vector of $\mathbf{H}$, then, $\mathbf{h}_i$ has complex multivariate normal distribution.

$$\mathbf{h}_i \sim N_C^{N_T}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

where, $\boldsymbol{\mu}_i$ is the mean vector and $\boldsymbol{\Sigma}_i$ is the covariance matrix.

Suppose all the row vectors $\mathbf{h}_i$ have the complex multivariate normal distribution with the same covariance matrix, $\boldsymbol{\Sigma}$. Then $\mathbf{Z} = \mathbf{H}^H\mathbf{H}$ follows a complex Wishart distribution denoted by

$$\mathbf{Z} \sim W_C^{N_T}\left(N_R, \mathbf{M}, \boldsymbol{\Sigma}\right)$$

where, $\mathbf{M} = \left[\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{N_R}\right]^T$.

For $\mathbf{M} = \mathbf{0}$, we have central complex Wishart distribution and $\mathbf{M} \neq \mathbf{0}$, then we have non-central complex Wishart distribution. The non-central complex Wishart distribution can be approximated by central complex Wishart distribution as $Z \sim W_C^{N_T}\left(N_R, \hat{\boldsymbol{\Sigma}}\right); \hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \frac{1}{N_R}\mathbf{M}^H\mathbf{M}$. The pdf (D. Gore et al., 2002) of post-detected SINR for $\mathbf{Z}$ following complex Wishart distribution is given by

$$P(\gamma_k) = \frac{\exp\left(\dfrac{-\gamma_k}{\left\{\rho \big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk}\right\}}\right)}{\left\{\rho \big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk}\right\}\Gamma\left(N_R - N_T + 1\right)}\left(\frac{\gamma_k}{\left\{\rho \big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk}\right\}}\right)^{N_R - N_T} \quad ; k = 1, 2, \ldots, N_T$$

Hence CDF is given by

$$P(\gamma_k) = \frac{\gamma\left(N_R - N_T + 1, \dfrac{\gamma_k}{\left\{\rho \big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk}\right\}}\right)}{\Gamma\left(N_R - N_T + 1\right)}$$

The average BER for k^th symbol is given by

$$P_e(k) = \int_0^\infty Q\left(\sqrt{2\gamma_k}\right) p\left(\gamma_k\right) d\gamma_k$$

Let $\gamma_k = \gamma\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\}$, then

$$P_e(k) = \frac{1}{\Gamma\left(N_R - N_T + 1\right)} \int_0^\infty Q\left(\sqrt{2\gamma\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\}}\right) \exp\left(-\gamma\right)\left(\gamma\right)^{N_R - N_T} d\gamma \tag{9.18}$$

This is the same integration we have used in section 8.2.3.

$$I_1(p, q, m) = \frac{q^m}{\Gamma(m)} \int_0^\infty Q\left(\sqrt{p\gamma}\right) e^{-q\gamma} g^{m-1} d\gamma$$

$$= \frac{1}{2\sqrt{\pi}\left(1 + r\right)^m} \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma\left(m + 1\right)} \left(\frac{r}{r+1}\right)^{\frac{1}{2}} {}_2F_1\left(1, m + \frac{1}{2}; m + 1; \frac{1}{1 + r}\right); r = \frac{p}{2q}$$

For our case $p = 2\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\}, q = 1, m = N_R - N_T + 1$.

The above integration can be further simplified (for positive integer values of m) to

$$I_1(p, q, m) = \frac{1}{2}\left[1 - \zeta \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{1 - \zeta^2}{4}\right)^k\right]$$

where, $\zeta = \sqrt{\dfrac{p}{p + 2q}} = \sqrt{\dfrac{2\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\}}{2\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\} + 2}} = \sqrt{\dfrac{\rho}{\rho + \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk}}}$

Therefore, average BER for symbol k is simply

$$P_e(k) = I_1\left(2\left\{ \rho \Big/ \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{kk} \right\}, 1, N_R - N_T + 1\right) \tag{9.19}$$

Hence we need to find the $\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{kk}$ in order to solve the above integration. Let us consider the case of most widely used i.i.d. Rayleigh fading MIMO channel. In this case, we have,

$$\because \qquad \mathbf{M} = \mathbf{0}$$

$$\therefore \qquad \hat{\mathbf{\Sigma}} = \mathbf{\Sigma} = \mathbf{I}_{N_T}$$

$$\Rightarrow \qquad \left(\hat{\mathbf{\Sigma}}^{-1}\right)_{kk} = 1$$

Therefore, average BER for symbol k is

$$P_e(k) = I_1(2\rho, 1, N_R - N_T + 1); I_1(p,q,m) = \frac{1}{2}\left[1 - \zeta \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{1-\zeta^2}{4}\right)^k\right]; \zeta = \sqrt{\frac{p}{p+2q}}$$

9.3.2 MMSE detector

As we have seen for ZF, noise was getting enhanced even if the spatial interference was removed.

MMSE detector minimizes the mean-square value of the spatial interference plus noise. In this the detector tries to minimize the mean square error between the actual signal and detected signal (see Exercise 9.6).

MMSE detector is another detector whose processor output is given by

$$\hat{\mathbf{s}} = \mathbf{W}_{MMSE}^H \mathbf{r} = \left(\mathbf{H}^H \mathbf{H} + \frac{N_0}{E_s} \mathbf{I}\right)^{-1} \mathbf{H}^H \mathbf{r}$$

$$= \left(\mathbf{H}^H \mathbf{H} + \frac{N_0}{E_s} \mathbf{I}\right)^{-1} \mathbf{H}^H (\mathbf{Hs}) + \left(\mathbf{H}^H \mathbf{H} + \frac{N_0}{E_s} \mathbf{I}\right)^{-1} \mathbf{H}^H (\mathbf{n}) \qquad (9.20)$$

One can observe that the above expression for MMSE pre-processor matrix is very similar to that of ZF pre-processor matrix except for an extra term, $\frac{N_0}{E_S}\mathbf{I}$, which will reduce the noise enhancement as we will see latter.

Example 9.9

Show that $\mathbf{W}_{MMSE}^H = \sigma_s^2 \mathbf{H}^H \left(\sigma_s^2 \mathbf{H}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_{N_T}\right)^{-1}$

Solution

$$\underset{\mathbf{s} \in C^{N_T \times N_R}}{\arg\min} \; E\left\|\mathbf{W}^H \mathbf{r} - \mathbf{s}\right\|^2$$

This can be done by taking partial derivative $E\left\|\mathbf{W}^H\mathbf{r}-\mathbf{s}\right\|^2$ w.r.t. $\mathbf{W}^H$ and setting to $\mathbf{0}$ as follows.

$$\frac{\partial}{\partial \mathbf{W}}\left(E\left[tr\left\{\left(\mathbf{W}^H\mathbf{r}-\mathbf{s}\right)\left(\mathbf{W}^H\mathbf{r}-\mathbf{s}\right)^H\right\}\right]\right)$$

$$= \frac{\partial}{\partial \mathbf{W}}\left(E\left[tr\left\{\left(\mathbf{W}^H\mathbf{r}\mathbf{r}^H\mathbf{W}-\mathbf{W}^H\mathbf{r}\mathbf{s}^H-\mathbf{s}\mathbf{r}^H\mathbf{W}+\mathbf{s}\mathbf{s}^H\right)\right\}\right]\right)$$

$$= \frac{\partial}{\partial \mathbf{W}}\left(\mathbf{W}^H\mathbf{R_{rr}}\mathbf{W}-\mathbf{W}^H\mathbf{R_{rs}}-\mathbf{R_{sr}}\mathbf{W}+\mathbf{R_{ss}}\right)$$

$$= \mathbf{W}^H\mathbf{R_{rr}}-\mathbf{R_{sr}}$$

Then, $\mathbf{W}_{MMSE}^H = \mathbf{R_{sr}}\mathbf{R_{rr}}^{-1}$.

Assuming noise vector and signal vector are independent.

$$\mathbf{R_{rr}} = \mathbf{H}\mathbf{R_{ss}}\mathbf{H}^H + \mathbf{R_{nn}} = \sigma_s^2\mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}_{N_T};\mathbf{R_{ss}} = \sigma_s^2\mathbf{I}_{N_T},\mathbf{R_{nn}} = \sigma_n^2\mathbf{I}_{N_R}$$

$$\mathbf{R_{sr}} = \mathbf{R_{ss}}\mathbf{H}^H = \sigma_s^2\mathbf{H}^H$$

Therefore, $\mathbf{W}_{MMSE}^H = \sigma_s^2\mathbf{H}^H\left(\sigma_s^2\mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}_{N_T}\right)^{-1}$

It can be shown (see Exercise 9.7) that

$$\sigma_s^2\mathbf{H}^H\left(\sigma_s^2\mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}_{N_T}\right)^{-1} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}_{N_T}\right)^{-1}\mathbf{H}^H = \mathbf{W}_{MMSE}^H$$

It is easy to verify that when the MIMO system operates in low SNR region, the noise component in the MMSE pre-processor is the dominant term and hence the filter behaves like a matched filter. In the high SNR region, where the interference is the main source of error, the filter behaves like a ZF detector. To sum up, the MMSE detector provides a good trade-off between the noise reduction and interference suppression thus achieving the highest SINR value among all linear estimators.

Review question 9.6 *Explain MMSE detection.*

Example 9.10

What happens to noise power for MMSE?

Solution

Let us denote noise after MMSE as

$$\left(\mathbf{H}^H\mathbf{H} + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{H}^H(\mathbf{n}) = \mathbf{z}.$$

Using SVD of **H** (Y. S. Cho et al., 2010), we have,

$$\|\mathbf{z}\|_2^2 = \left\|\left(\mathbf{H}^H\mathbf{H} + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{H}^H(\mathbf{n})\right\|_2^2$$

$$= \left\|\left(\mathbf{V\Sigma}^2\mathbf{V^H} + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{V\Sigma U}^H(\mathbf{n})\right\|_2^2$$

$$\because \quad (\mathbf{\Sigma}^{-1}\mathbf{V}^H)^{-1} = \mathbf{V\Sigma}$$

$$\therefore \quad \|\mathbf{z}\|_2^2 = \left\|\left(\mathbf{\Sigma V}^H + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\mathbf{V}^H\right)^{-1}\mathbf{U}^H(\mathbf{n})\right\|_2^2$$

$$= \left\|\mathbf{V}\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{U}^H(\mathbf{n})\right\|_2^2$$

Since the multiplication of a unitary matrix does not change the Frobenius norm, we have,

$$\|\mathbf{z}\|_2^2 = \left\|\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{U}^H(\mathbf{n})\right\|$$

We also know that,

$$Tr\left(\mathbf{BB}^H\right) = Tr\left(\mathbf{B}^H\mathbf{B}\right) = \|\mathbf{B}\|_2^2$$

Hence,

$$E\left(\|\mathbf{z}\|_2^2\right) = E\left(\left\|\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{U}^H(\mathbf{n})\right\|\right)$$

$$= E\left(Tr\left[\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{U}^H(\mathbf{n})\right]\mathbf{n}^H(\mathbf{U})\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\right)$$

$$= Tr\left(\left[\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{U}^H\right]E\left(\mathbf{nn}^H\right)(\mathbf{U})\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-1}\right)$$

$$= Tr\left(\left(\mathbf{\Sigma} + \frac{N_0}{E_s}\mathbf{\Sigma}^{-1}\right)^{-2}N_0\right)$$

$$= \sum_{i=1}^{N_T} N_0\left(\sigma_i + \frac{N_0}{E_s\sigma_i}\right)^{-2}$$

$$= \sum_{i=1}^{N_T} N_0 \left(\frac{E_s \sigma_i^2 + N_0}{E_s \sigma_i} \right)^{-2}$$

$$= \sum_{i=1}^{N_T} N_0 \left(\frac{E_s \sigma_i}{E_s \sigma_i^2 + N_0} \right)^{2}$$

$$= \sum_{i=1}^{N_T} N_0 \frac{\sigma_i^2 E_s^2}{\left(E_s \sigma_i^2 + N_0\right)^2}$$

$$\approx \frac{\sigma_{\min}^2 E_s^2}{\left(E_s \sigma_{\min}^2 + N_0\right)^2} \tag{9.21}$$

When the channel matrix is not well-behaved, the condition number $\left(\frac{\max(\sigma_i)}{\min(\sigma_i)}\right)$ is very large and minimum singular value is very small. Hence there is noise enhancement in MMSE as well. But it is less pronounced than that of the ZF detector. Note that the term $\sigma_{\min}^2$ appeared in both the numerator and denominator, hence the noise enhancement has been reduced.

9.3.2.1 *Outage probability and diversity gain*

The SINR for the k[th] symbol of MMSE detector (E. K. Onggosanusi et al., 2002) has been shown as

$$SINR_{MMSE} = \rho \gamma_k^{MMSE} = \mathbf{h}_k^H \left(\hat{\mathbf{H}} \hat{\mathbf{H}}_k^H + \rho^{-1} \mathbf{I} \right)^{-1} \mathbf{h}_k$$

In the above equation, $\mathbf{h}_k$ is the k[th] column of $\mathbf{H}$ matrix and if we remove this column from $\mathbf{H}$ matrix, we get the matrix $\hat{\mathbf{H}}_k$. The quantity ρ is defined as $E\left(\mathbf{s}\mathbf{s}^H\right) = \frac{\rho_{total}}{N_T} \mathbf{I}_{N_L N_T} = \rho \mathbf{I}_{N_L N_T}$ where, N_L is the block length. The CDF of $\rho\gamma_k$ is given (A. Hedayat et al., 2007) as

$$P(y) = 1 - \exp(-y) \sum_{n=1}^{N_R - N_T + 1} \frac{A_n(\rho y)}{(n-1)!} (y)^{n-1} \; ; A_n(\rho y) = \begin{cases} 1, & N_R - N_T + 1 \geq n \\ \dfrac{1 + \sum_{i=1}^{N_R - n} C^i (\rho y)^i}{(1+\rho y)^{N_T - 1}}, & N_R - N_T + 1 < n \end{cases}$$

where, C_i is the coefficient of y_i in $(1 + y)^{N_T - 1}$.

Hence outage probability for separate spatial encoding case is given as

$$P_{out} \approx N_T P \left(\frac{2^{\frac{R}{N_T}} - 1}{\rho} \right) \tag{9.22}$$

It can be shown that MMSE MIMO detector has diversity gain of $N_R - N_T + 1$ (see Exercise 9.4) same as that of the ZF MIMO detector. But there may be some difference in the diversity order of ZF and MMSE MIMO detection which will be discussed in the next chapter (conservation theorem).

9.3.2.2 *Performance analysis*

In linear detector, a linear pre-processor ($\mathbf{W}$) is first applied to the received signal vector, $\hat{\mathbf{s}} = \mathbf{W}^H\mathbf{r}$. Then we can do individual detection of $\hat{\mathbf{s}}$.

Without loss of generality, let us assume that we are detecting $\hat{s}_k, \quad k = 1$.

Then,
$$\hat{s}_1 = \mathbf{w}_1^H\mathbf{r} = \mathbf{w}_1^H\mathbf{h}_1 s_1 + \mathbf{w}_1^H\mathbf{n}$$

One may show that, $\mathbf{w}_1^H\mathbf{n} \sim N_C\left(0, \sigma_n^2\right)$.

Then the conditional error probability (CEP) for sub-channel 1 is given by

$$CEP = Q\left(\sqrt{2\mathbf{w}_1^H\mathbf{h}_1}\right)$$

For sub-channel 1, the corresponding weight vector is proportional to

$$\mathbf{w}_1 \propto \left(\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{h}_1$$

$$CEP = Q\left(\sqrt{2\mathbf{h}_1^H\left(\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{h}_1}\right)$$

Using eigen-decomposition of $\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H$, we have,

$$\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H + \frac{N_0}{E_s}\mathbf{I} = \mathbf{U}\left(\lambda + \frac{N_0}{E_s}\mathbf{I}\right)\mathbf{U}^H$$

Assuming, $\mathbf{x} = \mathbf{U}^H\mathbf{h}_1$, we get,

$$\mathbf{h}_1^H\left(\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{h}_1 = \sum_{i=1}^{N_R}\left(\lambda_i + \frac{N_0}{E_s}\right)^{-1}|x_i|^2$$

Note that the rank of $\hat{\mathbf{H}}_1$ is $N_T - 1 \le N_R$, hence, $N_R - N_T + 1$ eigenvalues of $\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H$ are zero.

$$\mathbf{h}_1^H\left(\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{h}_1 = \frac{E_s}{N_0}\sum_{i=1}^{N_R-N_T+1}|x_i|^2 + \sum_{i=N_R-N_T+2}^{N_R}\left(\lambda_i + \frac{N_0}{E_s}\right)^{-1}|x_i|^2$$

Hence,

$$CEP \le \exp\left(\frac{-E_s}{N_0}\sum_{i=1}^{N_R-N_T+1}|x_i|^2\right)\exp\left(\sum_{i=N_R-N_T+2}^{N_R}\left(-\lambda_i - \frac{N_0}{E_s}\right)^{-1}|x_i|^2\right)$$

Since $\mathbf{x} = \mathbf{U}^H \mathbf{h}_1 \sim N_c\left(0, \sigma_h^2 \mathbf{I}\right)$, we know that all x_i are independent of each other. Therefore,

$$BER \leq E\left\{\exp\left(-\frac{E_s}{N_0}\sum_{i=1}^{N_R-N_T+1}\left|x_i\right|^2\right)\right\} E\left\{\exp\left(-\sum_{i=N_R-N_T+2}^{N_R}\left(\lambda_i + \frac{N_0}{E_s}\right)^{-1}\right)\left|x_i\right|^2\right\}$$

which can be approximated as

$$BER \cong \left(\frac{1}{1+\bar{\gamma}}\right)^{N_R-N_T+1}\left(\frac{N_R + \frac{1}{\bar{\gamma}}}{N_R + 1 + \frac{1}{\bar{\gamma}}}\right)^{N_T-1} ; \bar{\gamma} = \frac{E_b \sigma_h^2}{N_0} \tag{9.23}$$

The linear detectors have a linear growth in complexity with the number of antennas. But, the performance gap between the ML detector and linear estimators are huge. The diversity gain for ML detection was N_R whereas ZF and MMSE detectors have diversity gain of $N_R - N_T + 1$ (we will discuss this in more detail when we discuss conservation theorem later). So we need to have more sophisticated algorithms which may be non-linear techniques to bridge the gap in the performance of linear detectors with ML detection. One such approach is Successive Interference Cancellation (SIC) where in every step of decoding we subtract the decoded stream from the received vector. If the decisions are correct the received vector will have less interference which will increase the diversity order of the next stream. This non-linear detection could be considered in conjunction with linear techniques like ZF and MMSE detectors which results in ZF-SIC and MMSE-SIC, respectively.

Review question 9.7 *What is the diversity gain of ZF and MMSE detection?*

9.4 Sphere decoding

As we have seen in section 9.2, the complexity of ML detection grows exponentially. Is there a way to reduce this complexity without compromising the performance? That's what Sphere Decoding (SD) exactly does. How does SD achieve this? Simply stated, it tries to find the ML solution vector within a sphere instead of all possible transmitted signal vectors (ML detection). But there may be no vector at all or numerous vectors inside the chosen sphere. How to handle such situations? In the first case, one may increase the radius of the sphere. In the second case, one may decrease the radius of the sphere so that only one vector exists inside the sphere which will give us the ML solution. Hence, SD is an iterative decoding which converges to the ML solution when the number of iterations is unbounded.

First step is converting the complex I–O MIMO system model into an equivalent real system model.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

$$\Rightarrow \qquad \mathbf{y}_{equi}^{real} = \mathbf{H}_{equi}^{real}\mathbf{x}_{equi}^{real} + \mathbf{n}_{equi}^{real}$$

$$\Rightarrow \qquad \begin{bmatrix}\mathrm{Re}(\mathbf{y})\\ \mathrm{Im}(\mathbf{y})\end{bmatrix} = \begin{bmatrix}\mathrm{Re}(\mathbf{H}) & -\mathrm{Im}(\mathbf{H})\\ \mathrm{Im}(\mathbf{H}) & \mathrm{Re}(\mathbf{H})\end{bmatrix}\begin{bmatrix}\mathrm{Re}(\mathbf{x})\\ \mathrm{Im}(\mathbf{x})\end{bmatrix} + \begin{bmatrix}\mathrm{Re}(\mathbf{n})\\ \mathrm{Im}(\mathbf{n})\end{bmatrix}$$

Example 9.11

Convert a complex 2 × 2 MIMO I–O model to an equivalent real system model.

Solution

For a 2 × 2 MIMO system,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

Separating the imaginary and real parts, we have,

$$\Rightarrow \quad \begin{bmatrix} y_1^{real} + jy_1^{imag} \\ y_2^{real} + jy_2^{imag} \end{bmatrix} = \begin{bmatrix} h_{11}^{real} + jh_{11}^{imag} & h_{12}^{real} + jh_{12}^{imag} \\ h_{21}^{real} + jh_{21}^{imag} & h_{22}^{real} + jh_{22}^{imag} \end{bmatrix} \begin{bmatrix} x_1^{real} + jx_1^{imag} \\ x_2^{real} + jx_2^{imag} \end{bmatrix} + \begin{bmatrix} n_1^{real} + jn_1^{imag} \\ n_2^{real} + jn_2^{imag} \end{bmatrix}$$

Hence the real equivalent model is

$$\begin{bmatrix} y_1^{real} \\ y_2^{real} \\ y_1^{imag} \\ y_2^{imag} \end{bmatrix} = \begin{bmatrix} h_{11}^{real} & h_{12}^{real} & -h_{11}^{imag} & -h_{12}^{imag} \\ h_{21}^{real} & h_{22}^{real} & -h_{21}^{imag} & -h_{22}^{imag} \\ h_{11}^{imag} & h_{12}^{imag} & h_{11}^{real} & h_{12}^{real} \\ h_{21}^{imag} & h_{22}^{imag} & h_{21}^{real} & h_{22}^{real} \end{bmatrix} \begin{bmatrix} x_1^{real} \\ x_2^{real} \\ x_1^{imag} \\ x_2^{imag} \end{bmatrix} + \begin{bmatrix} n_1^{real} \\ n_2^{real} \\ n_1^{imag} \\ n_2^{imag} \end{bmatrix}$$

MLD for the real equivalent system can be expressed as

$$\underset{\mathbf{x}_{equi}^{real} \in \boldsymbol{\chi}_{equi}^{real}}{\arg \min} \left\| \mathbf{y}_{equi}^{real} - \mathbf{H}_{equi}^{real} \mathbf{x}_{equi}^{real} \right\|^2$$

MLD search for ML solution over the symbol alphabet, $\boldsymbol{\chi}_{equi}^{real}$. However, for SD, we will search the solution over a sphere of radius r_{SD} only. Hence,

$$\left\| \mathbf{y}_{equi}^{real} - \mathbf{H}_{equi}^{real} \mathbf{x}_{equi}^{real} \right\|^2 \leq (r_{SD})^2 \tag{9.24}$$

Let us consider the QR decomposition of the real equivalent channel matrix

$$\mathbf{H}_{equi}^{real} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0}_{(2N_R - 2N_T) \times 2N_T} \end{bmatrix}$$

Note that $\mathbf{H}_{equi}^{real}$ is a $2N_R \times 2N_T$ matrix. Multiplying by $\mathbf{Q}^H = \begin{bmatrix} \mathbf{Q}_1^H \\ \mathbf{Q}_2^H \end{bmatrix}$ and using the unitary property of the **Q** matrix, we have,

$$\left\| \begin{bmatrix} \mathbf{Q}_1^H \\ \mathbf{Q}_2^H \end{bmatrix} \mathbf{y}_{equi}^{real} - \begin{bmatrix} \mathbf{R} \\ \mathbf{0}_{(2N_R - 2N_T) \times 2N_T} \end{bmatrix} \mathbf{x}_{equi}^{real} \right\|^2 \leq (r_{SD})^2$$

$$\Rightarrow \qquad \left\|\mathbf{Q}_1^H \mathbf{y}_{equi}^{real} - \mathbf{R}\mathbf{x}_{equi}^{real}\right\|^2 \le \left(r_{SD}\right)^2 - \left\|\mathbf{Q}_2^H \mathbf{y}_{equi}^{real}\right\|^2$$

Substituting the new $\mathbf{y}^n = \mathbf{Q}_1^H \mathbf{y}_{equi}^{real}$, $r_n = \left(r_{SD}\right)^2 - \left\|\mathbf{Q}_2^H \mathbf{y}_{equi}^{real}\right\|^2$ and replacing with a shorter notation $\mathbf{x}^r = \mathbf{x}_{equi}^{real}$, we have,

$$\left\|\mathbf{y}^n - \mathbf{R}\mathbf{x}^r\right\|^2 \le \left(r_n\right)^2$$

Since **R** is upper triangular matrix, we can write the above inequality in component form as

$$\sum_{i=1}^{2N_T}\left(y_i^n - \sum_{j=1}^{2N_T} R_{ij}x_j^r\right)^2 \le \left(r_n\right)^2 \tag{9.25}$$

Example 9.12

Find the above SD metric for a 2 × 2 MIMO system.

Solution

$$\left\|\begin{bmatrix} y_1^n \\ y_2^n \\ y_3^n \\ y_4^n \end{bmatrix} - \begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ 0 & R_{22} & R_{23} & R_{24} \\ 0 & 0 & R_{33} & R_{34} \\ 0 & 0 & 0 & R_{44} \end{bmatrix}\begin{bmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{bmatrix}\right\|^2 \le \left(r_n\right)^2$$

$$\Rightarrow \quad \left|y_1^n - R_{11}x_1^r - R_{12}x_2^r - R_{13}x_3^r - R_{14}x_4^r\right|^2 + \left|y_2^n - R_{22}x_2^r - R_{23}x_3^r - R_{24}x_4^r\right|^2 + \left|y_3^n - R_{33}x_3^r - R_{34}x_4^r\right|^2 + \left|y_4^n - R_{44}x_4^r\right|^2 \le \left(r_n\right)^2$$

Reordering the terms in the LHS, we have,

$$\left|y_4^n - R_{44}x_4^r\right|^2 + \left|y_3^n - R_{34}x_4^r - R_{33}x_3^r\right|^2 + \left|y_2^n - R_{24}x_4^r - R_{23}x_3^r - R_{22}x_2^r\right|^2 + \left|y_1^n - R_{14}x_4^r - R_{13}x_3^r - R_{12}x_2^r - R_{11}x_1^r\right|^2 \le \left(r_n\right)^2$$

Expanding SD metric for a $N_T \times N_R$ MIMO system, we have,

$$\left|y_{2N_T}^n - R_{2N_T,2N_T}x_{2N_T}^r\right|^2 + \left|y_{2N_T-1}^n - R_{2N_T-1,2N_T}x_{2N_T}^r - R_{2N_T-1,2N_T-1}x_{2N_T-1}^r\right|^2 + \cdots + \left|y_1^n - R_{1,2N_T}x_{2N_T}^r - R_{1,2N_T-1}x_{2N_T-1}^r - \cdots - R_{1,1}x_1^r\right|^2 \le \left(r_n\right)^2$$

Note that the first term is dependent only on $x_{2N_T}^r$; therefore, we can have a necessary condition as follows.

$$\left|y_{2N_T}^n - R_{2N_T,2N_T}x_{2N_T}^r\right|^2 \le \left(r_n\right)^2$$

In other words, we can look for $x_{2N_T}^r$ in the interval

$$LB_{2N_T} = \left\lceil \frac{-r_n + y^n_{2N_T}}{R_{2N_T,2N_T}} \right\rceil \le x^r_{2N_T} \le \left\lfloor \frac{r_n + y^n_{2N_T}}{R_{2N_T,2N_T}} \right\rfloor = UB_{2N_T}$$

where, LB_{2N_T} is the lower bound for $x^r_{2N_T}$, UB_{2N_T} is the upper bound for $x^r_{2N_T}$, $\lceil a \rceil$ is the smallest integer greater than a and $\lfloor a \rfloor$ is the greatest integer smaller than a.

The second term depends only on $x^r_{2N_T}$ and $x^r_{2N_T-1}$. We can have second condition from the first and second term of the SD metric inequality as follows.

$$\left| y^n_{2N_T} - R_{2N_T,2N_T} x^r_{2N_T} \right|^2 + \left| y^n_{2N_T-1} - R_{2N_T-1,2N_T} x^r_{2N_T} - R_{2N_T-1,2N_T-1} x^r_{2N_T-1} \right|^2 \le (r_n)^2$$

Therefore we can look for $x^r_{2N_T-1}$ in the interval

$$LB_{2N_T-1} = \left\lceil \frac{-r_n^{2N_T-1} + y^n_{2N_T-1|2N_T}}{R_{2N_T-1,2N_T-1}} \right\rceil \le x^r_{2N_T-1} \le \left\lfloor \frac{r_n^{2N_T-1} + y^n_{2N_T-1|2N_T}}{R_{2N_T-1,2N_T-1}} \right\rfloor = UB_{2N_T-1}$$

where, $y^n_{2N_T-1|2N_T} = y^n_{2N_T-1} - R_{2N_T-1,2N_T} x^r_{2N_T}$ and $\left(r_n^{2N_T-1}\right)^2 = (r_n)^2 - \left| y^n_{2N_T} - R_{2N_T,2N_T} x^r_{2N_T} \right|^2$.

Following the same procedure, we can find the interval in which one can look for $x^r_{2N_T-2}, x^r_{2N_T-3}, \cdots, x^r_1$. In SD, the multi-dimensional search of MLD is transformed to multiple searches in one dimension. Now we can write SD algorithm as shown in Fig. 9.2.

The different steps and decisions in the above flow chart are given below.

Step 1: Find the QR factorization of $\mathbf{H}^{real}_{equi} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0}_{(2N_R-2N_T)\times 2N_T} \end{bmatrix}$ and $\mathbf{y}^n = \mathbf{Q}_1^H \mathbf{y}^{real}_{equi}$.

Step 2: Set $k = 2N_T$, $r_n = (r_{SD})^2 - \left\| \mathbf{Q}_2^H \mathbf{y}^{real}_{equi} \right\|^2$ and $y^n_{2N_T|2N_T+1} = y^n_{2N_T}$.

Note that the SD algorithm starts by detecting the last element in $\mathbf{x}^r$.

Step 3: Set the bounds $LB_k = \left\lceil \frac{-r_n^k + y^n_{k|k+1}}{R_{k,k}} \right\rceil \le x^r_k \le \left\lfloor \frac{r_n^k + y^n_{k|k+1}}{R_{k,k}} \right\rfloor = UB_k$ and $x^r_k = LB_k - 1$

Step 4: Increase $x_k = x_k + 1$

Step 5: Decrease $k = k - 1$

$$y^n_{k|k+1} = y^n_k - \sum_{j=k+1}^{2N_T} R_{k,j} x^r_j$$

Note that the subscript $k \mid k+1$ in $y^n_{k|k+1}$ above is used to denote that the received signal y^n_k in the k[th] layer, is adjusted with the already estimated symbol components $x^r_{k+1}, x^r_{k+2}, \cdots, x^r_{2N_T}$ (already detected layers).

$$\left(r_n^k\right)^2 = \left(r_n^{k+1}\right)^2 - \left(y_{k+1|k+2}^n - R_{k+1,k+1}x_{k+1}^r\right)^2$$

Step 6: $k = k + 1$
Step 7: Save $\mathbf{x}^r$ and find its distance from $\mathbf{y}_{equi}^{real}$.

Note that whenever the SD algorithm finds a vector $\mathbf{x}^r$ inside the sphere, it sets the new sphere radius $\left\|\mathbf{y}_{equi}^{real} - \mathbf{H}_{equi}^{real}\mathbf{x}^r\right\|^2 = \left(r_{SD}\right)^2$ and restart the algorithm until it finds a single vector inside the sphere which is declared as the ML solution.

Decision 1: $x_k^r \le UB_k$?
Decision 2: $k = 2N_T + 1$?
Decision 3: $k = 1$?

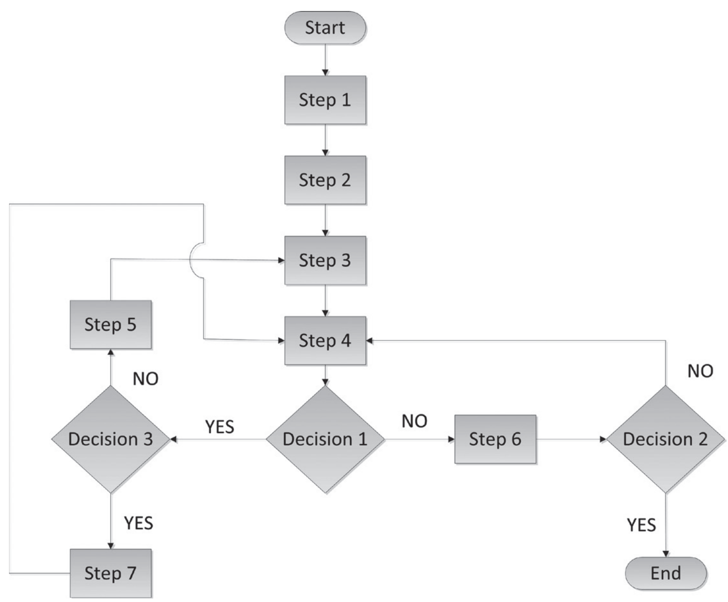


Fig. 9.2 Flowchart for SD algorithm

Example 9.13

Write the SD pseudo code for a simple 2 × 2 MIMO system.

Solution

Step 1:

Find the QR factorization of

$$\mathbf{H}_{equi}^{real} = \left[\mathbf{Q}_1 \quad \mathbf{Q}_2\right]\begin{bmatrix} \mathbf{R} \\ \mathbf{0}_{(2N_R - 2N_T)\times 2N_T} \end{bmatrix}$$

and $\mathbf{y}^n = \mathbf{Q}_1^H \mathbf{y}_{equi}^{real}$.

Step 2:

Set $k = 4$,

$r_n = \left(r_{SD}\right)^2 - \left\|\mathbf{Q}_2^H \mathbf{y}_{equi}^{real}\right\|^2$

and $y_{4|5}^n = y_4^n$.

Step 3:

Set the bounds $LB_k = \left\lceil \frac{-r_n^k + y_{k|k+1}^n}{R_{k,k}} \right\rceil \leq x_k^r \leq \left\lfloor \frac{r_n^k + y_{k|k+1}^n}{R_{k,k}} \right\rfloor = UB_k$

and $x_k^r = LB_k - 1$

Step 4:

Increase $x_k = x_k + 1$

Decision 1: $x_k^r \leq UB_k$?

If no then

Step 6:

$k = k + 1$

Decision 2: $k = 5$?

If yes the

stop.

If no then go to step 4

If yes then

Decision 3: $k = 1$?

If no then

Step 5: Decrease $k = k - 1$

$$y_{k|k+1}^n = y_k^n - \sum_{j=k+1}^{2N_T} R_{k,j} x_j^r$$

$$\left(r_n^k\right)^2 = \left(r_n^{k+1}\right)^2 - \left(y_{k+1|k+2}^n - R_{k+1,k+1} x_{k+1}^r\right)^2$$

If yes then

Step 7: Save $\mathbf{x}^r$ and find its distance from $\mathbf{y}_{equi}^{real}$

Go to Step 4: Increase $x_k = x_k + 1$

How do we decide the radius of the sphere? We can choose $\left(r_{SD}\right)^2 \propto N_R \sigma_n^2$ where, σ_n^2 is the noise variance. Hence for low SNR the radius of the sphere is large and SD is less efficient. But SD is highly efficient for high SNR since the sphere of the radius is small. One can apply lattice reduction techniques in combination with SD to improve its performance. More details about the SD algorithm are given in T. Kailath et al., (2005) and F. A. Monteiro et al., (2014).

Review question 9.8 *How does SD compare to ML in terms of performance and complexity?*

9.5 Summary

Figure 9.3 summarizes the MIMO detection techniques. In introduction to MIMO detection techniques, we have discussed about ML, ZF, MMSE and SD. The diversity gain for ML is N_R. The diversity gain for ZF is $N_R - N_T + 1$. MMSE has slightly higher diversity gain than ZF as we will discuss in the next chapter. But the gap in performance is still quite large between the ML and linear sub-optimal detectors like ZF and MMSE. SD has similar performance with that of ML with reduced complexity.

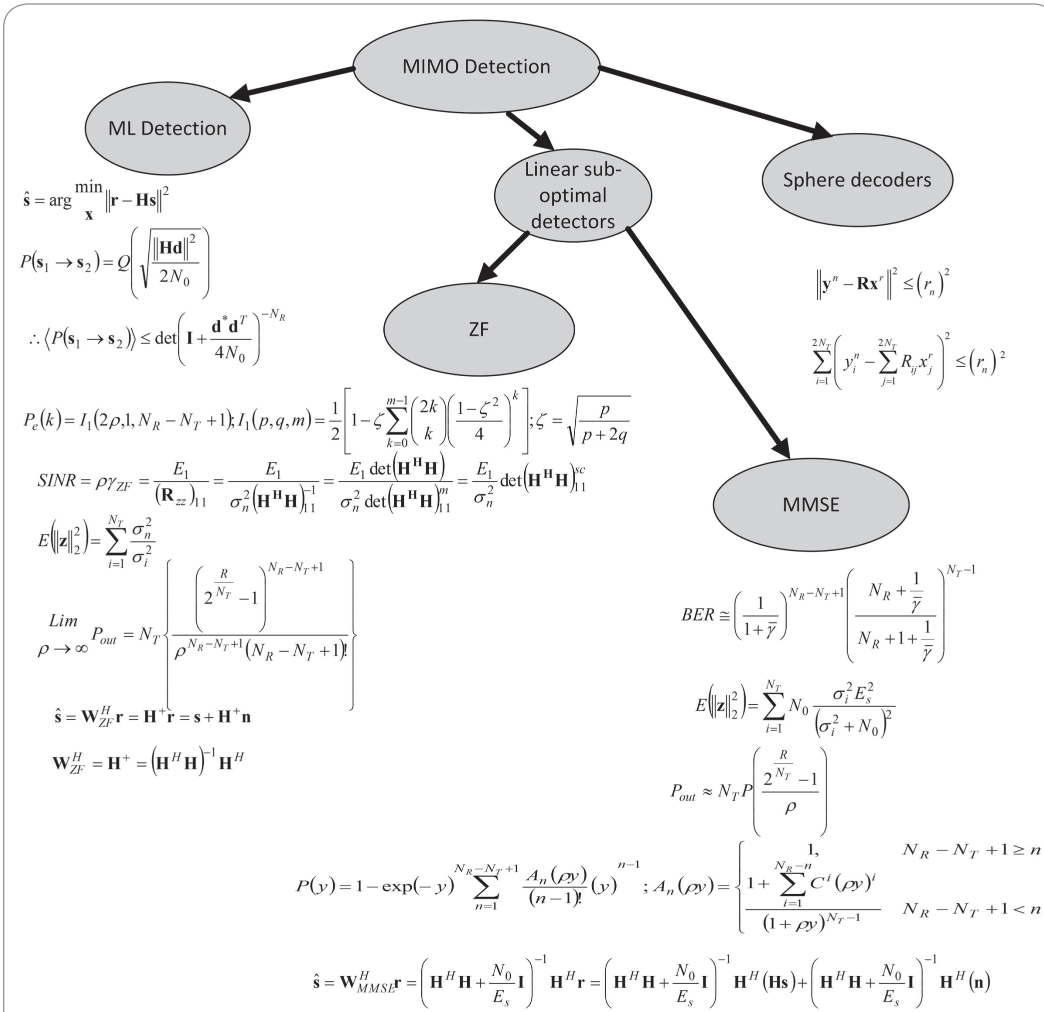


Fig. 9.3 Chapter 9 in a nutshell

Exercises

Exercise 9.1

Show that for a 2×2 MIMO system, covariance matrix of $\mathbf{n}^H\mathbf{Hd}$ equals $\frac{N_0}{2}\|\mathbf{Hd}\|^2$, where **H** is the channel matrix, $\mathbf{d} = \mathbf{s}_1 - \mathbf{s}_2$ and **n** is the CSCG.

Exercise 9.2

Show that $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$.

Exercise 9.3

Prove that $vec(\mathbf{AB}) = (\mathbf{B}^T \otimes \mathbf{I})vec(\mathbf{A})$.

Exercise 9.4

Show that the diversity order of MMSE detector for a i.i.d. Rayleigh fading channel is $N_R - N_T + 1$.

Exercise 9.5

Prove that the SINR for the kth symbol of MMSE detector can be expressed as

$$\rho\gamma_{MMSEk} = \mathbf{h}_k^H\left(\hat{\mathbf{H}}\hat{\mathbf{H}}_k^H + \rho^{-1}\mathbf{I}\right)^{-1}\mathbf{h}_k$$

Exercise 9.6

Find the $N_R \times N_T$ preprocessing matrix **W** which will give minimum mean square error $\min E\left[\left(\mathbf{s} - \mathbf{W}^H\mathbf{r}\right)^2\right]$.

Hint: The (Wiener) solution gives $\mathbf{W}_{MMSE}^H = \left(\mathbf{H}^H\mathbf{H} + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\mathbf{H}^H$.

Exercise 9.7

Show that $\sigma_s^2\mathbf{H}^H\left(\sigma_s^2\mathbf{H}\mathbf{H}^H + \sigma_s^2\mathbf{I}_{N_T}\right)^{-1} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}_{N_T}\right)^{-1}\mathbf{H}^H = \mathbf{W}_{MMSE}^H$.

Following are MATLAB based exercises.

Exercise 9.8

Implement the SD algorithm in MATLAB for a 2 × 2 MIMO system.

References

1. Choi, J. 2010. *Optimal Combining and Detection*. Cambridge: Cambridge University Press.
2. Gore, D., R. W. Heath, and A. Paulraj. 2002. 'On performance of the zero forcing receiver in presence of transmit correlation'. In *Proc. IEEE Int. Symp. on Information Theory*. Lausanne, Switzerland. 159.
3. Gore, D., R. W. Heath, and A. Paulraj. Nov. 2002. 'Transmit selection in spatial multiplexing systems'. *IEEE Communications Letters*. 6(11). 491–493.
4. Hedayat, A. and A. Nostrania. Dec. 2007. 'Outage and diversity of linear receivers in flat-fading MIMO channels'. *IEEE Trans. Signal Processing*. 55(12). 5868–5873.
5. Hochwald, B. M. and S. Brink. 2003. 'Achieving near-capacity on a multiple-antennas channel'. *IEEE Trans. Comm*. 51(3). 389–399.
6. Kailath, T. 1980. *Linear Systems*. New Jersey: Prentice-Hall, Inc.
7. Kailath, T., H. Vikalo, and B. Hassibi. 2006. 'MIMO receive Algorithms'. In *Space-time Wireless Systems: From Array Processing to MIMO Communications*. H. Bolcskei, D. Gesbert, C. Papadias, and A. J. van der Veen, Eds., Cambridge, UK: Cambridge University Press, 2006, pp. 302-321.
8. Kiessling, M. and J. Speidel. 2003. 'Analytical performance of MIMO zero-forcing receivers in correlated Rayleigh fading environments'. In *Proc. IEEE Signal Processing Advances in Wireless Communications*. Rome, Italy.
9. Lütkepohl, H. 1996. *Handbook of Matrices*. Chichester: John Wiley & Sons.
10. Monteiro, F. A., I. J. Wassell, and N. Souto. 2014. 'MIMO detection methods'. In *MIMO Processing for 4G and Beyond*, M. M. da Silva and F. A. Monteiro, Eds. Boca Raton: CRC Press. 47–117.
11. Onggosanusi, E. K., A. G. Dabak, T. Schmidl, and T. Muharemovic. 2002. 'Capacity analysis of frequency-selective MIMO channels with sub-optimal detectors'. In *Proc. IEEE ICASSP*. 2369–2372.
12. Proakis, J. G. and M. Salehi. 2008. *Digital Communications*. New York: McGraw-Hill.
13. Rupp, M., C. Mecklenbrauker, and G. Gritsch. 2003. 'High diversity with simple space-time block-codes and linear receivers'. In *Proc. IEEE GLOBECOM*. 302–306.
14. Siriteanu, C., X. Shi, and Y. Miyanaga. 2011. 'Analysis and simulation of MIMO zero-forcing detection performance for correlated and estimated Rician-fading channel'. In *Proc. AusCTW*. 182–187.
15. Strang, G. 2006. *Linear Algebra and its Applications*. New Delhi: Cengage Learning India.
16. Winters, J. H., J. Salz, and R. D. Gitlin. Feb./Mar./Apr., 1994. 'The impact of antenna diversity on the capacity of wireless communication systems'. *IEEE Trans. Comm*. 42(2/3/4). 1740–1751.
17. Xu, R. and F. C. M. Lau. Feb. 2006. 'Performance analysis for MIMO systems using zero forcing detector over fading channels'. *IEE Proc. Communications*. 153(1). 74–80.

Advanced MIMO Detection Techniques

CHAPTER 10

10.1 Introduction to spatially multiplexed MIMO systems

We have observed in chapter 2 that the capacity of wireless communication links is increased by using multiple antennas at the transmitter and the receiver. To achieve these capacities, a transmission scheme, called Diagonal-Bell Laboratories Layered Space-time (D-BLAST) has been proposed by G. J. Foschini (1996). In an i.i.d. Rayleigh scattering environment, this processing structure leads to theoretical rates which grow linearly with the number of antennas (for $N_R = N_T$) with these rates approaching 90% of Shannon's capacity. But this has large computational complexity required for implementation of this scheme.

A simplified version, called Vertical BLAST (V-BLAST) has been proposed by G. J. Foschini et al., (1999). They have demonstrated spectral efficiencies of 20–40 bps/Hz at average signal-to-noise ratio (SNR) ranging from 24 to 34 dB could be achieved in indoor environments. The essential difference between D-BLAST and V-BLAST lies in the vector encoding process. The D-BLAST code blocks are organized along diagonals in space-time. In V-BLAST, however, the vector encoding process is simply a demultiplexing (DMUX) operation followed by independent bit-to-symbol mapping of each sub-stream. No inter sub-stream coding, or coding of any kind, is required.

Note that BLAST detection scheme can be done in one of the following ways:

1. interference nulling to reduce the effect of the other (interfering) signals on the desired one is employed in successive interference cancellation (SIC) which will be discussed in section 10.4;
2. ordering to select the sub-stream with the largest signal-to-noise ratio (SNR) or other criteria along with SIC, which will be discussed later in ordered successive interference cancellation (OSIC) (section 10.5).

10.2 Vertical/horizontal layered space-time transmission

Vertical Bell Laboratories Layered Space-time Transmission (V-BLAST) suggests simultaneous transmission of independent uncoded data sub-streams. The input data is separated into several sub-streams with demultiplexing (DMUX) operation which are associated with each transmit antenna.

This process is followed by independent bit-to-symbol mapping (M-ary) of each sub-stream and may be interleaved (optional) as depicted in Fig. 10.1 (a). We consider a V-BLAST system with N_T transmitting antennas. At the transmitter the data is passed through a serial-to-parallel converter (S/P converter) and transformed into N_T sub-streams, where each sub-stream is sent through a different transmit antenna. As usual in any communication system, after the S/P converter, all sub-streams will be modulated and may be interleaved (refer to Appendix E for a brief description on interleaver) and sent through the transmitting antenna.

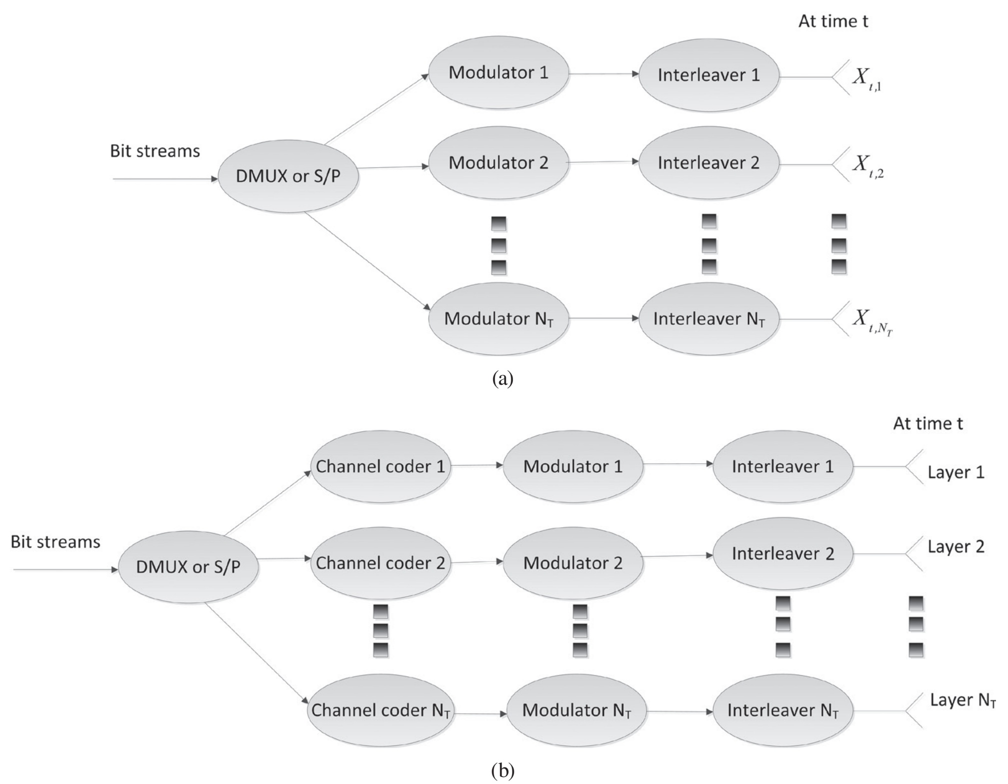


Fig. 10.1 (a) Vertical layered space-time transmission (b) Horizontal layered space-time transmission (Note that V-BLAST architecture can also include optional encoder in series after the message source. But the main difference between V-BLAST and H-BLAST is that in V-BLAST, the channel coding can be done over time whereas in H-BLAST the channel coding can be done over space and time.)

The transmission matrix (**X**) for V-BLAST can be represented as

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{2,1} & x_{3,1} & \cdots \\ x_{1,2} & x_{2,2} & x_{3,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,N_T} & x_{2,N_T} & x_{3,N_T} & \cdots \end{bmatrix} \tag{10.1}$$

where in $x_{t,j}$, t is the time index and j is the antenna index.

Hence the first row of **X** matrix is transmitted from the first antenna for time t = 1,2,3,… Second row of **X** matrix is transmitted from the second antenna for time t = 1,2,3,… and so on.

Example 10.1

For $N_T = 5$, write down the transmission matrix **X** for V-BLAST transmission. Assume that there are 35 sub-streams.

Solution

The V-BLAST demultiplexes the data stream into sub-streams referred to as layers and sent one sub-stream over one transmit antenna.

$$\mathbf{X} = \begin{bmatrix} 1 & 6 & 11 & 16 & 21 & 26 & 31 \\ 2 & 7 & 12 & 17 & 22 & 27 & 32 \\ 3 & 8 & 13 & 18 & 23 & 28 & 33 \\ 4 & 9 & 14 & 19 & 24 & 29 & 34 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 \end{bmatrix}$$

The numbers in the above matrix show the order in which the data symbols (sub-streams) in the original streams. It can be seen that the original stream is mapped vertically into the columns of the transmission matrix, hence the name Vertical BLAST. First row is transmitted from first antenna, second row is transmitted from second antenna and so on. First column is transmitted in the first time instant, second column in the second time instant and so on.

This transmission method will yield inter-stream interference. For instance, consider the first column of the **X** transmission matrix. All the N_T antennas are transmitting simultaneously at the time index t = 1, hence any antenna at the receiver will receive all the signals streams from transmitting antennas 1 to N_T. This interfering signals decrease the signal-to-interference-noise ratio (SINR) at the receiver dramatically. V-BLAST detection is done with zero forcing successive interference cancellation (ZF-SIC)/ minimum mean square error successive interference cancellation (MMSE-SIC) detectors which will cancel the interference from the previously detected signals. Such detectors will be discussed in later sections.

If we introduce channel coding for each data sub-streams before modulation in the V-BLAST then we have horizontal BLAST (Fig. 10.1b). Although V-BLAST can also employ channel coders like H-BLAST, in the literature (D.-S. Shiu and M. Kahn, 1999), the term has been used for H-BLAST and we have retained it. We can still employ the ZF-SIC or MMSE-SIC for detection. The only difference now will be to introduce a channel decoder at the receiver.

Review question 10.1 *What is the difference between H-BLAST and V-BLAST?*

10.3 Diagonal Bell labs layered space-time transmission

The information stream is DMUX into N_T sub-streams and each data sub-stream is transmitted by a different antenna through a diagonal interleaving scheme. Table 10.1 (a) shows four different data sub-streams a, b, c and d for instance. Each sub-stream may contain a block of coded symbols.

The sub-streams are cyclically shifted before sending it over the N_T antennas. It will ensure higher diversity order than the H-BLAST since same sub-streams are transmitted from different antennas. This results in diagonally layered signal in space and time. As we can see from Table 10.1 (a), for N_T = 4, there are four layers and each codeword is divided into four blocks (number of blocks should be equal to N_T). The decoder decodes layer (sub-stream) by layer. The first layer is detected without any error, since it is transmitted alone (see Table 10.1 a). After that, the second layer is demodulated and detected and it has only one interferer from the first layer. But the first layer is already decoded, it can be subtracted. The third will face two interferers. But the first and second layers are already detected and they can be subtracted. The process goes on. Note that the decoding in the previous layers should be error free, otherwise the whole process would suffer from error propagation. ZF-SIC and MMSE-SIC algorithms for D-BLAST are given in T. M. Duman et al., (2007). There are many unused time slots (some of the transmitting antennas are sitting idle) in D-BLAST, threaded D-BLAST can be employed to increase transmission rate by wrapping sub-streams as depicted in Table 10.1 (b).

Table 10.1

(a) Diagonal Bell labs layered space-time transmission (D-BLAST) (b) Threaded D-BLAST (Note that $a_{1,1}$, $a_{2,2}$, $a_{3,3}$ and $a_{4,4}$ are referring to the same sub-stream which has been transmitted from the 1st, 2nd, 3rd and 4th transmitting antenna at time slots 1, 2, 3 and 4, respectively)

(a)

Transmitting antenna 1	$a_{1,1}$	$b_{2,1}$	$c_{3,1}$	$d_{4,1}$			
Transmitting antenna 2		$a_{2,2}$	$b_{3,2}$	$c_{4,2}$	$d_{5,2}$		
Transmitting antenna 3			$a_{3,3}$	$b_{4,3}$	$c_{5,3}$	$d_{6,3}$	
Transmitting antenna 4				$a_{4,4}$	$b_{5,4}$	$c_{6,4}$	$d_{7,4}$
Time slots	1	2	3	4	5	6	7

(b)

Transmitting antenna 1	$a_{1,1}$	$b_{2,1}$	$c_{3,1}$	$d_{4,1}$	$a_{5,1}$	$b_{6,1}$	$c_{7,1}$
Transmitting antenna 2	$d_{1,2}$	$a_{2,2}$	$b_{3,2}$	$c_{4,2}$	$d_{5,2}$	$a_{6,2}$	$b_{7,2}$
Transmitting antenna 3	$c_{1,3}$	$d_{2,3}$	$a_{3,3}$	$b_{4,3}$	$c_{5,3}$	$d_{6,3}$	$a_{7,3}$
Transmitting antenna 4	$b_{1,4}$	$c_{2,4}$	$d_{3,4}$	$a_{4,4}$	$b_{5,4}$	$c_{6,4}$	$d_{7,4}$
Time slots	1	2	3	4	5	6	7

Example 10.2

For $N_T = 5$, write down the transmission matrix **X** for D-BLAST transmission.

Solution

$$\mathbf{X} = \begin{bmatrix} 1 & 5 & 4 & 3 & 2 & 1 & 5 \\ 2 & 1 & 5 & 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 & 5 & 4 & 3 \\ 5 & 4 & 3 & 2 & 1 & 5 & 4 \end{bmatrix}$$

There are as many sub-streams or layers as the number of transmit antennas. But the sub-streams are not transmitted as it is unlike V-BLAST. The sub-streams are cyclically reordered and are transmitted repeatedly.

Review question 10.2 *What is the difference between D-BLAST and threaded D-BLAST?*

10.4 Successive interference cancellation detection

We will discuss Zero Forcing Successive Interference Cancellation (ZF-SIC) and Minimum-mean Squared Successive Interference Cancellation (MMSE-SIC) for V-BLAST detection.

10.4.1 Zero forcing successive interference cancellation detection

As we have mentioned in section 10.2, BLAST transmission method will yield inter-stream interference since all antennas at the transmitter are transmitting simultaneously. Any receiver antenna will receive streams from all the transmitting antennas at any time. The duty of detector is to detect and decode these streams one by one. When the receiver wants to decode a stream from one transmitting antenna, all other streams from the remaining transmitting antennas are acting as an interferer. Is there any way of removing these inter-stream interferers? That's what we are going to explore in this sub-section. QR decomposition (see Appendix A) will be used in the ZF-SIC detectors.

Assume any $N_R \times N_T$ channel matrix $\mathbf{H}$ where $N_T \leq N_R$ which can be decomposed as

$$\mathbf{H} = \mathbf{QR} \tag{10.2}$$

where, $\mathbf{Q}$ is a $N_R \times N_T$ matrix with its orthonormal columns being the ZF nulling vectors.

$$\mathbf{Q}^H\mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}$$

$$\Rightarrow \qquad \mathbf{Q} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_{N_T} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$\mathbf{R}$ is $N_T \times N_T$ upper triangular matrix**.**

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N_{T-1}} & R_{1N_T} \\ 0 & R_{21} & \cdots & R_{2N_{T-1}} & R_{2N_T} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & R_{N_{T-1}N_{T-1}} & R_{N_{T-1}N_T} \\ 0 & \cdots & 0 & 0 & R_{N_T N_T} \end{bmatrix}$$

Note that for an $N_R \times N_T$ channel matrix $\mathbf{H}$ with rank $R_{\mathbf{H}}$, then $\mathbf{Q}$ is $N_R \times R_{\mathbf{H}}$ and $\mathbf{R}$ is $R_{\mathbf{H}} \times N_T$. Once $\mathbf{H}$ matrix is perfectly estimated, we can calculate $\mathbf{Q}$ and $\mathbf{R}$ from $\mathbf{H}$. Then the following signal processing is performed to find the transmitted symbols, $x_1, x_2, \ldots, x_{N_T}$. Assume $\mathbf{r}$ is the received vector for a MIMO system with channel matrix $\mathbf{H}$**,** $\mathbf{n}$ is the AWGN noise vector whose elements are complex normal Gaussian distributed with zero mean and variance $\left(N_c(0,\sigma_n^2)\right)$ and $\mathbf{x}$ is the transmitted signal vector. We can calculate the $\mathbf{y}$ vector by pre-multiplying the $\mathbf{r}$ vector with $\mathbf{Q}^H$, which is virtually the nulling step, as

$$\begin{aligned}\mathbf{y} &= \mathbf{Q}^H\mathbf{r} = \mathbf{Q}^H\left(\mathbf{Hx}+\mathbf{n}\right) = \mathbf{Q}^H\left(\mathbf{QRx}+\mathbf{n}\right)\\ &= \mathbf{Rx}+\mathbf{Q}^H\mathbf{n} = \mathbf{Rx}+\mathbf{z}\end{aligned} \tag{10.3}$$

$\mathbf{z}$ is another Gaussian noise vector with same mean and variance as $\mathbf{n}$.

The above equation can be written in element wise format of the matrix as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{N_T-1} \\ y_{N_T} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N_{T-1}} & R_{1N_T} \\ 0 & R_{22} & \cdots & R_{2N_{T-1}} & R_{2N_T} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & R_{N_{T-1}N_{T-1}} & R_{N_{T-1}N_T} \\ 0 & \cdots & 0 & 0 & R_{N_TN_T} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{N_{T-1}} \\ x_{N_T} \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_{N_{T-1}} \\ z_{N_T} \end{bmatrix}$$

which is basically

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{N_{T-1}} \\ y_{N_T} \end{bmatrix} = \begin{bmatrix} R_{11}x_1 + R_{12}x_2 + \ldots + R_{1N_{T-1}}x_{N_{T-1}} + R_{1N_T}x_{N_T} + z_1 \\ \vdots \\ R_{N_{T-1}N_{T-1}}x_{N_{T-1}} + R_{N_TN_T}x_{N_T} + z_{N_{T-1}} \\ R_{N_TN_T}x_{N_T} + z_{N_T} \end{bmatrix} \tag{10.4a}$$

Note that

$$y_i = \sum_{j=1}^{N_T} R_{ij}x_j + z_i; i = 1,2,\ldots,N_T$$

We will employ nearest neighbourhood rule in the detection of symbols.

What is the nearest neighbourhood rule?

In simple words, consider a test point x; assume that x' is the closest point to x out of the rest of the test points. Then the estimate of x can be assumed as x'. In the nearest neighbourhood rule, we estimate the transmitted symbol by finding the Euclidean distance with the noisy received signal with all possible symbols which may be transmitted. The symbol which gives the smallest Euclidean distance with the received signal is assumed to be transmitted. It is similar to finding the nearest neighbour of the received signal vector with possible symbols in a constellation diagram, hence the name nearest neighbourhood rule.

After the QR decomposition of the channel matrix as described above, we can do the MIMO detection in the following ways:

1. Detect for $i = N_T$, then estimate (x_{N_T}) using nearest neighborhood rule.
2. Then cancel estimate (x_{N_T}) from y_{N_T} to detect x_{N_T-1}.

$$y_{N_T-1} = R_{N_T-1N_T-1}x_{N_T-1} + R_{N_T-1N_T}x_{N_T} + z_{N_T-1}$$

$$\Rightarrow \quad y_{N_T-1} - R_{N_T-1N_T} x_{N_T} = R_{N_T-1N_T-1} x_{N_T-1} + z_{N_T-1}$$

If we have estimated x_{N_T} correctly, that means $\hat{x}_{N_T} = x_{N_T}$, then

$$\hat{x}_{N_T-1} = g\left(\frac{y_{N_T-1} - R_{N_T-1N_T}\hat{x}_{N_T}}{R_{N_T-1N_T-1}}\right)$$

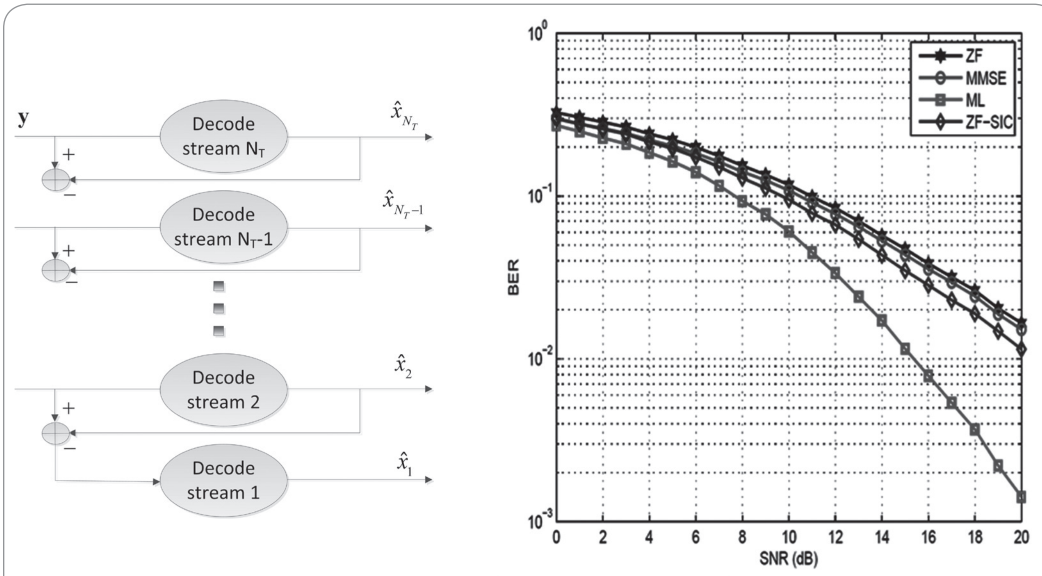
where, g is the slicing function.

3. Note that y_i can be expressed as

$$y_i = R_{ii}x_i + \sum_{j=i+1}^{N_T} R_{ij}x_j + z_i \tag{10.4b}$$

where, x_i is the current detected signal.
y_i contains a lower level of interference than the received signal **r** as the interference from x_l for $l < i$ are suppressed. $\sum_{j=i+1}^{N_T} R_{ij}x_j$ is the interference from $x_{i+1}, x_{i+2}, \cdots, x_{N_T}$ which can be cancelled by using the available estimates of $x_{i+1}, x_{i+2}, \cdots, x_{N_T}$ which are already detected. Hence the current signal x_i can be estimated as

$$\hat{x}_i = g\left(\frac{y_i - \sum_{j=i+1}^{N_T} R_{ij}\hat{x}_j}{R_{ii}}\right)$$



(a) (b)

Fig. 10.2 (a) Illustration of successive interference cancellation (SIC) (b) BER performance comparison of conventional detectors in 2 × 2 MIMO system using 64-QAM over i.i.d. Rayleigh fading MIMO channel

The basic idea of successive interference cancellation (SIC) is to cancel the interference from the previously detected symbols as depicted in Fig. 10.2 (a). This will reduce the interference and hence increases effective SINR. Figure 10.2 (b) depicts the performance of various conventional detectors viz. ZF, MMSE, ML and ZF-SIC over i.i.d. Rayleigh fading MIMO channel. As expected, ML has the best BER performance, then ZF-SIC, followed by MMSE and ZF respectively.

Example 10.3

Explain ZF-SIC for 3×3 MIMO system.

Solution

For $N_R = N_T = 3$, we have

$$y_1 = R_{11}x_1 + R_{12}x_2 + R_{13}x_3 + z_1$$
$$y_2 = R_{22}x_1 + R_{23}x_3 + z_1$$
$$y_3 = R_{33}x_3 + z_3$$

$$\Rightarrow \quad \hat{x}_3 = g\left(\frac{y_3}{R_{33}}\right)$$
$$\hat{x}_2 = g\left(\frac{y_2 - R_{23}\hat{x}_3}{R_{22}}\right)$$
$$\hat{x}_1 = g\left(\frac{y_1 - R_{12}\hat{x}_2 - R_{13}\hat{x}_3}{R_{11}}\right)$$

g denotes the slicing operation as the closest constellation point selection.

Review question 10.3 *Write down the steps of ZF-SIC.*

10.4.2 Minimum-mean squared successive interference cancellation detector

We can see the equivalent ZF detection of MMSE detection (R. Bohnke et al., 2003) by defining an extended channel matrix and received vector as follows:

$$\mathbf{H}_{ext} = \begin{bmatrix} \mathbf{H} \\ \sqrt{\frac{N_0}{E_s}}\mathbf{I} \end{bmatrix}; \mathbf{y}_{ext} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}; \mathbf{n}_{ext} = \begin{bmatrix} \mathbf{n} \\ \sqrt{\frac{N_0}{E_s}}\mathbf{I} \end{bmatrix} \tag{10.6}$$

where, $\frac{E_S}{N_0}$ is the signal-to-noise ratio (SNR).

Now we can see the equivalence for the new extended ZF detector (see section 9.3.1) by finding the

$$\hat{\mathbf{s}} = \mathbf{W}_{ZF}^{H}\mathbf{r} = \mathbf{H}^{+}\mathbf{r} = \mathbf{s} + \mathbf{H}^{+}\mathbf{n}$$

where, $\mathbf{W}_{ZF}^{H} = \mathbf{H}^{+} = \left(\mathbf{H}^{H}\mathbf{H}\right)^{-1}\mathbf{H}^{H}$

We have,

$$\hat{\mathbf{s}} = \mathbf{W}_{ZF}^{H}\mathbf{r} = \left(\begin{pmatrix}\mathbf{H}^{H} & \sqrt{\frac{N_0}{E_s}}\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{H} \\ \sqrt{\frac{N_0}{E_s}}\mathbf{I}\end{pmatrix}\right)^{-1}\begin{pmatrix}\mathbf{H}^{H} & \sqrt{\frac{N_0}{E_s}}\mathbf{I}\end{pmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{0}\end{bmatrix}$$

$$= \left(\mathbf{H}^{H}\mathbf{H} + \frac{N_0}{E_s}\mathbf{I}\right)^{-1}\left(\mathbf{H}^{H}\mathbf{y}\right)$$

This is exactly what we do in the MMSE detector (see section 9.3.3).

We can do the QR factorization of this new extended ZF detector and follow the same procedure of sequential detection which will behave like MMSE-SIC detector.

Review question 10.4 *Write down the equivalent ZF detection of MMSE detection.*

10.4.3 Conservation theorem

Linear detectors are discussed in section 9.3. The use of multiple antennas at the receiver mitigates interference as well as overcome multipaths. As usual, there is trade-off between this two for linear detectors. Receiver with multiple antennas: higher diversity gain implies lesser interference mitigation and vice versa according to conservation theorem (J. R. Barry et al., 2010).

Conservation theorem:

For ZF MIMO detector over a $N_R \times N_T$ Rayleigh flat fading channel, the diversity order is $N_R - N_T + 1$ since it makes the $N_T - 1$ interferers null. In other words, sum of diversity gain (d) plus number of interferers (N_{inter}) equals the number of receive antennas (N_R), i.e.

$$d + N_{inter} = N_R \tag{10.7}$$

Example 10.4

Explain that ZF-SIC has higher diversity order than ZF using conservation theorem.

Solution

Note that ZF-SIC involves nulling and cancelling operation simultaneously. First nulling vector must null the $N_T - 1$ interferers. Hence, from conservation theorem, the diversity order for first detection of the 1st symbol is $N_R - N_T + 1$. For the detection of the second symbol, we need to null the $N_T - 2$ interferers. Hence the diversity order for second detection of symbol is $N_R - N_T + 2$. Therefore, for detecting the k^{th} symbol, the diversity order is $N_R - N_T + k$. So the last symbol detected will have full diversity order. Note that if any symbols are not detected correctly, this diversity order decreases. To sum up, there is diversity order gain for ZF-SIC over ZF MIMO detectors.

Example 10.5

Explain that MMSE detector has higher diversity order than ZF using conservation theorem.

Solution

There is some difference in the way MMSE detector tackles the interferers. MMSE detectors ignore those interferers whose strength is below noise floor level, thereby, the diversity order for MMSE detector can be expressed as $N_R - N_{Teff} + 1$ where N_{Teff} is the number of significant interferers. Hence the effective diversity order of MMSE detectors could be higher than the ZF detector.

Review question 10.5 *What is conservation theorem?*

10.5 Ordered successive interference cancellation detector

In SIC, if the detected stream in one step is incorrect, its subtraction from the received vector will increase the interference and results in performance degradation. This is also known as error propagation. Hence the critical issue in ordering the detection of each stream so that error propagation is minimized. There are techniques which combine ordered successive interference cancellation (OSIC) and linear detection techniques like ZF-OSIC and MMSE-OSIC. To mitigate error propagation, multiple candidate symbols rather than a single detected symbol can be employed for SIC and this leads to list based detections. List based SIC MIMO detection outperforms the traditional SIC MIMO detection methods and they are discussed in depth in L. Bai et al., (2012). A. Zanella et al., (2005) investigated the performance of MMSE detectors in a flat Rayleigh-fading MIMO environment and generalize this methodology to derive the SER for MMSE-SIC with (or without) error propagation (EP). As mentioned before, the main problem with the SIC detector is that there may be error propagation. If we employ ordered successive interference cancellation (OSIC), then this error propagation may be minimized. The idea is to detect the signal with minimal error first so that the error propagation may be minimized. In the process, we may decide the order in which we detect the signals by various criteria listed below:

(a) Signal to noise interference ratio (SINR)
Signals with the higher SINR are detected earlier than the other signals. This is applicable for MMSE based OSIC detectors.

(b) Signal to noise ratio (SNR)
Signals with the higher SNR are detected earlier than the other signals. This is applicable for ZF based OSIC detectors.

(c) Log-likelihood ratio (LLR)
The ordering is based on the LLR. S. W. Kim (2003) has proposed LLR based detection ordering for V-BLAST and it has better performance. As we know that for V-BLAST detection (after nulling operation), we have,

$$y_i = x_i + \mathbf{w}_i\mathbf{n}$$

The LLR λ_i for x_i (assume equi-probable BPSK symbols) (S. W. Kim, 2003) is given by

$$\lambda_i = \ln \frac{P\left(x_i = +\sqrt{E_s}\,|y_i\right)}{P\left(x_i = -\sqrt{E_s}\,|y_i\right)} = \frac{4\sqrt{E_s}}{N_0}\frac{\mathrm{Re}\left(y_i\right)}{\|\mathbf{w}_i\|^2} \tag{10.8}$$

It can be shown that the bit error probability and LLR is related by

$$P_{e,i} = P\left(\hat{x}_i \neq x_i\right) = \frac{1}{1 + e^{|\lambda_i|}} \tag{10.9}$$

Since the bit error probability decreases with increasing LLR, the detection ordering is to detect the component of **x** that provides the largest $|\lambda_i|$ first.

Example 10.6

What is LLR?

Solution

For a given observation x, the likelihood function is defined as

$$f_i(x) = f(x/H_i),\ i = 0, 1$$

The ML decision is to choose the hypothesis (either H_0 or H_1) that maximizes the likelihood function.

$$f_0(x) \underset{H_1}{\overset{H_0}{\gtrless}} f_1(x) \Rightarrow \frac{f_0(x)}{f_1(x)} \underset{H_1}{\overset{H_0}{\gtrless}} 1 \Rightarrow LLR(x) = \ln\left(\frac{f_0(x)}{f_1(x)}\right) \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$\frac{f_0(x)}{f_1(x)}$ is called likelihood ratio (LR) and $\ln\left(\frac{f_0(x)}{f_1(x)}\right)$ is known as LLR.

Example 10.7

What are hard and soft decision?

Solution

For coded system, it is better to use soft decision than hard decision. Usually LLR is used for soft decision for channel decoders. For instance, let us consider a simple binary alphabet of $S = \{-1, +1\}$. So when we send 1 and -1, the received vectors are

$$\mathbf{r} = (+1)\mathbf{h} + \mathbf{n};\ \mathbf{r} = (-1)\mathbf{h} + \mathbf{n}$$

Then ML decision for signal s using the LR is given by

$$LR = \frac{f\left(\mathbf{r}|s = +1\right)}{f\left(\mathbf{r}|s = -1\right)} = \frac{\exp\left(-\left(\mathbf{r} - \mathbf{h}\right)^H \mathbf{R}_n^{-1}\left(\mathbf{r} - \mathbf{h}\right)\right)}{\exp\left(-\left(\mathbf{r} + \mathbf{h}\right)^H \mathbf{R}_n^{-1}\left(\mathbf{r} + \mathbf{h}\right)\right)}$$

$$= \exp\left\{\left(-(\mathbf{r}-\mathbf{h})^H \mathbf{R}_n^{-1}(\mathbf{r}-\mathbf{h})\right) + \left((\mathbf{r}+\mathbf{h})^H R_n^{-1}(\mathbf{r}+\mathbf{h})\right)\right\}$$

$$\because \qquad (x+a)^2 - (x-a)^2 = 4xa$$

We can define LLR (J. Choi, 2010) as

$$LLR = \ln(LR) = 4\mathrm{Re}\{\mathbf{h}^H\mathbf{R}_n^{-1}\mathbf{r}\}$$

The sign of LLR is like ML detection (hard decision) and the absolute value of LLR will give an idea of how reliable is the decision. This analysis could be extended for any M-ary symbol whose k^{th} bit is either 0 or 1 and finding the LLR for the k^{th} bit. Soft decision or LLR based linear MIMO detectors outperforms the hard decision based linear MIMO detectors (Y. S. Cho et al., 2010).

Review question 10.6 *What are three criteria for deciding the order of signal detection in OSIC?*

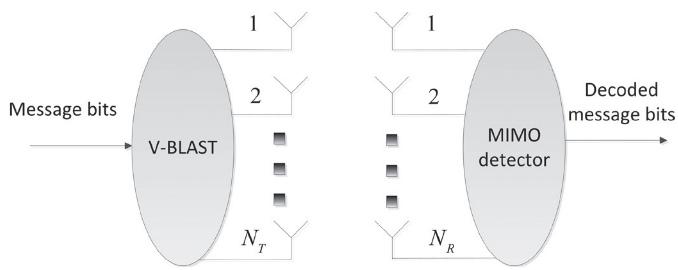


Fig. 10.3 V-BLAST MIMO system

10.5.1 Performance analysis

In the performance analysis, it is quite involved to find the exact performance analysis. Hence, we will derive the upper and lower bound of the ZF-SIC detector's performance. Consider the I–O model of a V-BLAST MIMO system depicted in Fig. 10.3. The received signal vector **r** for a narrowband MIMO channel **H** can be obtained as

$$\mathbf{r} = \mathbf{Hs} + \mathbf{n}$$

In the above equation, **s** is the signal vector and **n** is the AWGN noise vector whose elements are distributed as σ_n^2. The above equation can be expressed as

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{N_R} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_R} \end{bmatrix} \tag{10.10}$$

We can also express channel matrix **H** in terms of column vectors $\mathbf{H} = \begin{bmatrix} \mathbf{h_1} & \mathbf{h_2} & \cdots & \mathbf{h_{N_T}} \end{bmatrix}$

where, $\mathbf{h}_k = \begin{bmatrix} h_{1k} \\ h_{2k} \\ \vdots \\ h_{N_R k} \end{bmatrix}$ for $k = 1, 2, \ldots, N_T$.

Hence, $\mathbf{r} = \begin{bmatrix} \mathbf{h_1} & \mathbf{h_2} & \cdots & \mathbf{h_{N_T}} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_R} \end{bmatrix} = \sum_{k=1}^{N_T} \mathbf{h}_k s_k + \mathbf{n}$

We will assume SINR based ordering which is optimal, i.e., only the layer with highest SINR is detected in each recursion which gives the lowest SER in overall. We will also assume that perfect SIC (PSIC which means that the cancelled interference are accurate). Hence optimal ordered perfect ZF-SIC detector (ZF-OOPSIC) will give the lower bound of ZF-OSIC.

Let $\mathbf{r}^{(j)}$ denote the received signal for the j^{th} recursive step after interference cancellation. For PSIC, we can write

$$\mathbf{r}^{(j)} = \sum_{a=j}^{N_T} \mathbf{h}_{k_a} s_{k_a} + \sum_{a=1}^{j-1} \mathbf{h}_{k_a} \left(s_{k_a} - Q_y\left(\hat{s}_{k_a}\right)\right) + \mathbf{n} \tag{10.11}$$

where, $Q_y\left(\hat{s}_{k_a}\right)$ is the hard decision of the estimated value of $\hat{s}_{k_a}$.

In the above equation, the first summation term is for undetected symbols, and the second summation is the interference cancellation of the already detected symbols. For PSIC, the detected symbols are exactly what we have transmitted and hence the second summation term should be zero for ideal case (J. Han et al., 2010).

$$\mathbf{r}^{(j)} = \sum_{a=j}^{N_T} \mathbf{h}_{k_a} s_{k_a} + \mathbf{n} \tag{10.12}$$

If $\mathbf{w}_l^{(j)}$ is the nulling vector (this is the l^{th} row vector of Moore–Penrose pseudo-inverse of ($\mathbf{H}^{(j)} = \left[\mathbf{h}_1^{(j)}, \mathbf{h}_2^{(j)}, \ldots, \mathbf{h}_{N_R - j+1}^{(j)}\right]$) for the l^{th} undetected symbol in the j^{th} recursive step where $l = 1, 2, \ldots, N_T - j + 1$.

We have

$$\mathbf{w}_l^{(j)} \mathbf{h}_p^{(j)} = \begin{cases} 1 & l = p \\ 0 & i \neq p \end{cases},$$

the $N_T - j + 1$ estimated symbols derived in the j^{th} recursive step could be expressed as

$$\begin{aligned} \hat{s}_l^{(j)} &= \mathbf{w}_l^{(j)} \mathbf{r}^{(j)} = \mathbf{w}_l^{(j)} \left(\sum_{a=j}^{N_T} \mathbf{h}_{k_a} s_{k_a} + \mathbf{n} \right) \\ &= \mathbf{w}_l^{(j)} \mathbf{h}_l^{(j)} s_l + \mathbf{w}_l^{(j)} \sum_{a=j, k_a \neq l}^{N_T} \mathbf{h}_{k_a} s_{k_a} + \mathbf{w}_l^{(j)} \mathbf{n} = s_l + \mathbf{w}_l^{(j)} \mathbf{n} \end{aligned} \tag{10.13}$$

Therefore, the SINR of the l^{th} undetected layer in the j^{th} recursive step is given by

$$SINR_l^{(j)} = \frac{E\left(|s_l|^2\right)}{E\left(\left|\mathbf{w}_l^{(j)}\mathbf{n}\right|^2\right)} = \frac{P}{N_0\left|\mathbf{w}_l^{(j)}\right|^2} = \frac{\gamma_0}{\left|\mathbf{w}_l^{(j)}\right|^2} \tag{10.14}$$

Note that $\frac{1}{\left|\mathbf{w}_l^{(j)}\right|^2}$ follows Chi-square distribution with $2(N_R - N_T + j)$ degrees of freedom and variance $\frac{1}{2}$ and, therefore, $x_l^{(j)} = \sqrt{\frac{1}{\left|\mathbf{w}_l^{(j)}\right|^2}}; l = 1,2\ldots,N_T - j + 1$ for any l will have generalized Rayleigh distribution with $2(N_R - N_T + j)$ degrees of freedom and variance $\frac{1}{2}$. Its pdf and CDF (for even $2(N_R - N_T + j)$) are given by (J. G. Proakis et al., 2008)

$$P\left(x^{(j)}\right) = \frac{2}{\left(N_R - N_T + j - 1\right)!}\left(x^{(j)}\right)^{2(N_R - N_T + j) - 1} e^{-\left(x^{(j)}\right)^2} \tag{10.15}$$

$$P\left(x^{(j)}\right) = 1 - e^{-\left(x^{(j)}\right)^2} \sum_{k=0}^{N_R - N_T + j - 1} \frac{\left(x^{(j)}\right)^{2k}}{k!}$$

For optimal ordering using order statistics (G. Casella et al., 2002), the maximum SINR in each recursive step should be selected from continuous population of pdf $p(x^{(j)})$ and CDF $P(x^{(j)})$

Order statistics (H. A. David et al., 2003):

To select the l^{th} undetected layer with the maximum SINR in the j^{th} recursive step, $x_l^{(j)}$s are rearranged in ascending order of SINR say $\left(x_1^{(j)} \le x_2^{(j)} \le x_3^{(j)} \le \cdots \le x_{N_T - j}^{(j)} \le x_{N_T - j + 1}^{(j)}\right)$, the r^{th} highest value of SINR is selected. According to order statistics, the PDF of $x_r^{(j)}$ is given by

$$p_{x_r}^{j}\left(x^{(j)}\right) = \frac{1}{B\left(r, N_T - j + 2 - r\right)}\left\{P_{x_l}^{(j)}\left(x^{(j)}\right)\right\}^{r-1}\left\{1 - P_{x_l}^{(j)}\left(x^{(j)}\right)\right\}^{N_T - j + 1 - r} p_{x_l}^{(j)}\left(x^{(j)}\right) \tag{10.16}$$

where, $B(.,.)$ is the beta function and for positive arguments $B(m,n) = \frac{(m-1)!(n-1)!}{(m+n-1)!}$, hence,

$$B\left(r, N_T - j + 2 - r\right) = \frac{(r-1)!\left(N_T - j + 1 - r\right)!}{\left(N_T - j + 1\right)!} \tag{10.17}$$

Therefore the pdf of maximum SINR $x^{(j)}$ i.e., $x_{N_T - j + 1}^{(j)}$ (or minimal interference) for our case is given as

$$p_{N_T - j + 1}^{(j)}(x) = \frac{1}{B\left(N_T - j + 1, 1\right)}\left\{P_{N_T - j + 1}^{(j)}(x)\right\}^{N_T - j} p_{N_T - j + 1}^{(j)}(x)$$

$$= \frac{(N_T - j + 1)!}{(N_T - j)!}\left\{P_{N_T-j+1}^{(j)}(x)\right\}^{N_T-j} p_{N_T-j+1}^{(j)}(x) \tag{10.18}$$

The pdf and cdf of SINR of any undetected layer $x^{(j)}$ for our case are

$$p\left(x^{(j)}\right) = \frac{2}{(N_R - N_T + j - 1)!}\left(x^{(j)}\right)^{2(N_R-N_T+j)-1} e^{-\left(x^{(j)}\right)^2} \tag{10.19}$$

$$P\left(x^{(j)}\right) = 1 - e^{-\left(x^{(j)}\right)^2} \sum_{k=0}^{N_R-N_T+j-1} \frac{\left(x^{(j)}\right)^{2k}}{k!} \tag{10.20}$$

Hence, the pdf of maximum SINR $x^{(j)}$ substituting $N_R - N_T + j - 1 = q$ is

$$p_{N_T-j+1}^{j}(x) = \frac{2(N_T - j + 1)}{q!}\left[1 - e^{-x^2}\sum_{k=0}^{q}\frac{x^{2k}}{k!}\right]^{N_T-j} x^{2q+1}e^{-x^2} \tag{10.21}$$

The CEP of M-QAM could be expressed as

$$P_b(E|SINR) = 4\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{2g_{QAM}SINR}\right) - 4\left(1-\frac{1}{\sqrt{M}}\right)^2 Q^2\left(\sqrt{2g_{QAM}SINR}\right); g_{QAM} = \frac{3}{2(M-1)}$$

$$= 2\left(1 - \frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{3SINR}{2(M-1)}}\right) - \left(1 - \frac{1}{\sqrt{M}}\right)^2 erfc^2\left(\sqrt{\frac{3SINR}{2(M-1)}}\right)$$

where erfc is the complementary error function (G. L. Stuber, 2001), defined as

$$erfc(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}dy = 1 - erf(x) \tag{10.22}$$

Note that in the above equation, we have used the following relation between the Q function and erfc, i.e.,

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{y^2}{2}}dy = \frac{1}{2}erfc\left(\frac{x}{\sqrt{2}}\right)$$

There are N_T layers to be detected. Hence, we need to sum and average the probability of error for all layers. The SER for M-QAM could be obtained as

$$\begin{aligned}
P_b^{ZF-OOPSIC}(E) &= \frac{1}{N_T}\sum_{j=1}^{N_T} P_b^{ZF-OOPSIC(j)}(E) \\
&= \frac{1}{N_T}\sum_{j=1}^{N_T}\int_0^\infty P_b^{ZF-OOPSIC(j)}(E/SINR)p_{N_T-j+1}^{(j)}(x)dx \qquad (10.23) \\
&= \frac{1}{N_T}\sum_{j=1}^{N_T}\left\{\begin{array}{l}\frac{4A(N_T - j + 1)}{q!}\int_0^\infty erfc(Bx)\left[1 - e^{-x^2}\sum_{k=0}^{q}\frac{x^{2k}}{k!}\right]^{N_T-j} x^{2q+1}e^{-x^2}dx - \\ \frac{2A^2(N_T - j + 1)}{q!}\int_0^\infty erfc^2(Bx)\left[1 - e^{-x^2}\sum_{k=0}^{q}\frac{x^{2k}}{k!}\right]^{N_T-j} x^{2q+1}e^{-x^2}dx\end{array}\right\}
\end{aligned}$$

where, $A = \left(1 - \frac{1}{\sqrt{M}}\right); B = \sqrt{\frac{3\gamma_0}{2(M-1)}}$.

This is the lower bound of ZF-OSIC.

For upper bound we can calculate the SER of ZF without SIC since in ZF we do not do the interference cancellation, probability of error will be higher for this case than the ZF-OSIC. We can write SINR for ZF as

$$SINR^{(j)} = \frac{\gamma_0}{\left|\mathbf{w}^{(j)}\right|^2}$$

Note that $\frac{1}{\left|\mathbf{w}^{(j)}\right|^2}$ follows Chi-square distribution with $2(N_R - N_T + 1)$ degrees of freedom and variance $\frac{1}{2}$. Then $u^{(j)} = \sqrt{\frac{1}{\left|\mathbf{w}^{(j)}\right|^2}}$ will have generalized Rayleigh distribution with $2(N_R - N_T + 1)$ degrees of freedom and variance $\frac{1}{2}$. The PDF of $u^{(j)}$ for $j = 1$ is given by

$$P\left(u^{(j)}\right) = \frac{2}{(N_R - N_T)!}\left(u^{(j)}\right)^{2(N_R - N_T)+1} e^{-\left(u^{(j)}\right)^2} \tag{10.24}$$

In this case, the sum and average of probability of all undetected layers will be equal to the probability of error for a single undetected layer since there is no successive interference cancellation done for ZF. Therefore, for M-QAM, the SER for ZF is given by

$$P_b^{ZF}(E) = \frac{1}{N_T}\sum_{j=1}^{N_T} P_b^{ZF(j)}(E) = P_b^{ZF(j)}(E)$$

$$= \int_0^\infty P_b^{ZF}(E / SINR)\, p(u)\, du$$

$$= \frac{4A}{(N_R - N_T)!}\int_0^\infty erfc(Bu)(u)^{2(N_R - N_T)+1} e^{-(u)^2} du - \frac{2A^2}{(N_R - N_T)!}\int_0^\infty erfc^2(Bu)(u)^{2(N_R - N_T)+1} e^{-(u)^2} du \tag{10.25}$$

where, A and B were defined in Eq. (10.23).

Hence SER for ZF-OSIC is bounded as $P_b^{ZF-OOPSIC}(E) \le P_b^{ZF-OSIC}(E) \le P_b^{ZF}(E)$.

Review question 10.7 *What is OOPSIC?*

Review question 10.8 *What is the lower and upper bound on the SER for ZF-SIC?*

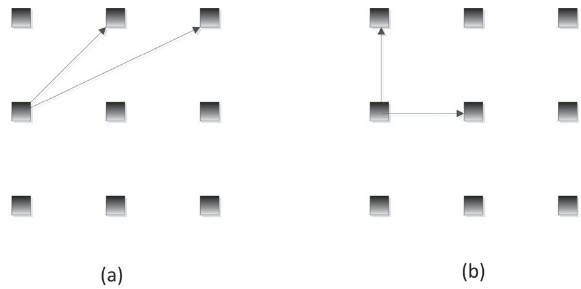


Fig. 10.4 (a) Long and non-orthogonal basis vectors (b) Short and orthogonal basis vectors

10.6 Lattice reduction based detector

One of the major issues with linear detectors we have considered so far was the noise enhancement. This effect becomes more pronounced when the channel matrix is not well behaved. Lattice reduction algorithm could be employed to reduce the condition number of channel matrix (bring it closer to 1). The condition number of a matrix is defined as the ratio of the largest and smallest singular value of the matrix. Unless and otherwise specified, all matrix or vector norm are assumed to be L2-norm. In MATLAB, one can calculate norm of a matrix **H** using command norm(**H**). The L2-norm of a matrix equals its largest singular value and the L2-norm of inverse of a matrix equals the reciprocal of the smallest singular value. It can be verified from the following example. Hence, the condition number of matrix (V. Kuhn, 2006) can be expressed as

$$\text{cond}\,(\mathbf{H}) = \frac{\sigma_{\max}}{\sigma_{\min}} = \|\mathbf{H}\|_2 \left\|\mathbf{H}^{-1}\right\|_2 \geq 1 \tag{10.26}$$

Hence for real orthogonal matrices with $\mathbf{H}^{-1} = \mathbf{H}^T$, the condition number is one and no noise amplification for linear detectors. Hence the matched filter and ZF are equivalent, since $(\mathbf{H}^H\mathbf{H})^{-1}\,\mathbf{H}^H = \mathbf{H}^H$. MMSE also has an equivalent ZF by defining extended matrices. Hence matched filter and MMSE will also be identical. Hence for well-conditioned orthogonal matrices, the performance of linear detectors is good. Therefore, it is desirable to have a roughly orthogonal matrix with condition number close to 1.

Example 10.8

Consider the matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

The eigenvalues of the matrix $\mathbf{H}^H\mathbf{H} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$ are $\lambda_1 = 6.8541, \lambda_2 = 0.1459$. Hence the singular values of **H** are $\sigma_1 = 2.618, \sigma_2 = 0.3820$.

$$\|\mathbf{H}\|_2 = \left\| \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \right\|_2 = \left(1^2 + 1^2 + 2^2 + 1^2\right)^{\frac{1}{2}} = 2.618 = \sigma_{\max}$$

$$\|\mathbf{H}^{-1}\|_2 = \left\| \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \right\|_2 = 2.618 = \frac{1}{0.3820} = \frac{1}{\sigma_{\min}}$$

Lattice is regular arrangements of points in Euclidean space (D. Micciancio et al., 2002). It is a set of points in n-D space with periodic structure. For instance, consider the two sets of lattice vectors shown in Fig. 10.4. In Fig. 10.4 (b), the basis vectors are short and orthogonal, it will have no noise enhancement. Using the basis vectors $\mathbf{b}_1(1, 0)$ and $\mathbf{b}_2(0, 1)$, we can generate the lattice points in 2-D space as depicted in Fig. 10.5 (a). They generate all the intersection points of the grid also known as lattice points. Similarly, using new basis vectors $\mathbf{b}_1' = \mathbf{b}_1 + \mathbf{b}_2$ (1, 1), $\mathbf{b}_2' = 2\mathbf{b}_1 + \mathbf{b}_2(2, 1)$, we can also generate the lattice points in 2-D space as shown in Fig. 10.6 (b). But we can't generate a lattice from the following two vectors $\mathbf{b}_1'' = \mathbf{b}_1 + \mathbf{b}_2$ (1, 1), $\mathbf{b}_2'' = 2\mathbf{b}_1$ (2, 0) and they are not basis vectors because the basic parallelepiped generated from these two vectors contains the lattice point (1,0) and (2,1). How will one represent the lattice point (1,0) with linear integer combination of (1,1) and (2,0)? It is not possible. The area of the parallelepiped shown in Fig. 10.5 (a) and (b) (shown in shaded region) are exactly same since these two generator matrices are equivalent. But, the shaded area depicted in Fig. 10. 5 (c) will be different.

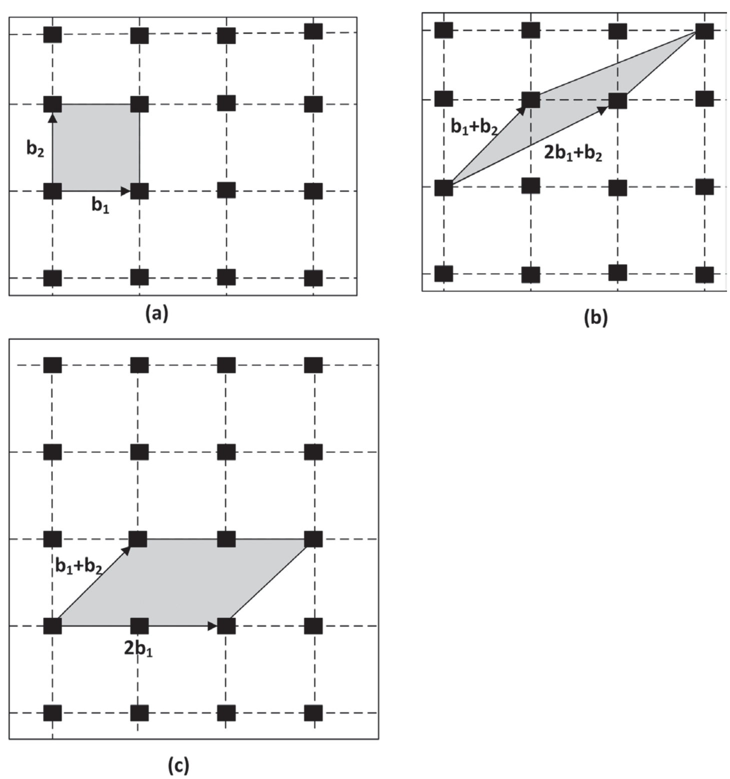


Fig. 10.5 (a) Basis vectors in R^2, (b) another equivalent basis vectors in R^2 and (c) not suitable basis vectors in R^2

As we are aware, the equivalent channel for Alamouti space-time code has two orthogonal column vectors. In Lattice reduction (LR) based MIMO detection, we need to find short and near orthogonal basis vectors of the channel matrix and perform MIMO detection over the equivalent MIMO system which is well conditioned.

Example 10.9

What is a lattice?

Solution

A lattice $\mathbf{L}$ (regularly arranged arrays of points) is a set of vectors that are obtained by the integer linear combination of a set of linearly independent vectors known as basis vectors $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots \mathbf{b}_k]$. Mathematically,

$$\mathbf{L} = \left\{ \mathbf{B} \middle| \mathbf{Bu} = \sum_{k=1}^{K} \mathbf{b}_k u_k, u_k \in Z \right\}$$

where, Z is the set of integers and $\mathbf{b}_k \in R^N, N \geq K$.

Example of lattice in communication theory is QAM constellation which is a finite subset of the complex integer lattice.

Example 10.10

What is generator matrix of a lattice?

Solution

A vector $\mathbf{v}$ in the lattice $\mathbf{L}$ could be expressed as

$$\mathbf{v} = \mathbf{Bu},\ \mathbf{u} \in Z^k$$

where $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots \mathbf{b}_k]$ is the generator matrix whose columns $[\mathbf{b}_k]$ are basis/generator vectors.

Example 10.11

What is a unimodular matrix?

Solution

A square matrix whose determinant is ±1 is called as unimodular. Then an integer matrix $\mathbf{U}$ (whose elements are integer) whose determinant is ±1 is called as integer unimodular.

It is quite possible that for the same lattice $\mathbf{L}$, there could be two different generator matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ as depicted in Fig. 10.6. Then it can be shown that $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$.

$$\det(\mathbf{U}\mathbf{U}^{-1}) = 1 = \det(\mathbf{U}^{-1})\det(\mathbf{U})$$

$$\Rightarrow \quad \det(\mathbf{U}^{-1}) = \det(\mathbf{U}) = \pm 1$$

Hence if $\mathbf{U}$ is unimodular, inverse of matrix $\mathbf{U}$ is also unimodular.

For example, the following matrices are unimodular.

$$\mathbf{U_1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}; \mathbf{U_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \mathbf{U_1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

Two generator matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ are equivalent iff $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$ for some unimodular matrix **U.**

Physical interpretation:

In MIMO detection, we want to have the short basis vectors (or approximately orthogonal). *Det* (**L**) is the n-dimensional volume of the parallelepiped spanned by the basis vectors. For example, see the shaded region of Fig. 10.5 (a) and (b). If **L** is full rank lattice, **B** is a square matrix, we have det (**L**) = |det (**B**)|. In other words, if we assume that ϕ_i are the acute angle between the $\mathbf{b}_i$ and the hyperplane spanned by $\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}$, then

$$\det(\mathbf{L}) = \|\mathbf{b}_1\| \cdots \|\mathbf{b}_n\| \sin\varphi_2 \cdots \sin\varphi_n$$

where the $\|\mathbf{x}\| = \sqrt{x_i^2}$ is the Euclidean length of a vector.

If $\mathbf{b}_i$ is short, then ϕ_i is large and $\mathbf{b}_i$ is nearly orthogonal to previous basis vectors. *Det* (**L**) is basically volume of the parallelepiped. From the Hadamard's inequality

$$\|\mathbf{b}_1\| \cdots \|\mathbf{b}_n\| \geq \det(L) \tag{10.27}$$

The equality holds when the basis vectors are mutually orthogonal. One can define orthogonality defect (od) as

$$od = \frac{\|\mathbf{b}_1\| \cdots \|\mathbf{b}_n\|}{\det(\mathbf{L})} \geq 1 \tag{10.28}$$

From the above inequality we can infer that shorter basis vectors will be closer to orthogonal.

Example 10.12

Explain in few words the lattice reduction based MIMO detection.

Solution

Lattice reduction (LR) basically assumes that the channel matrix (**H**) is a generator matrix for a lattice since **H** contains many column vectors. Using efficient algorithms like Lenstra, Lenstra and Lovasz (LLL), discussed in section 10.7, it will find an equivalent generator matrix of the lattice which is well behaved. Lattice reduction is concerned with selecting the short basis vectors. Here well-behaved means the new generator matrix for the lattice will have a smaller condition number. We will apply the MIMO detection techniques to the new well behaved equivalent generator matrix of the channel. Sub-optimal MIMO detectors like ZF, ZF-SIC are surprisingly efficient when they are employed on a reduced basis since the equivalent channel **G** will be better behaved and hence the noise enhancement will be minimized.

Note that designing an efficient decoder for MIMO is a challenging task. We have seen from example 9.1 that although ML decoder performance is optimal but the decoder complexity grows exponentially with the number of transmit antennas. It becomes almost impossible to realize a ML decoder for higher number of antennas and larger alphabet size. So we have resorted to low complexity sub-optimal decoders like ZF and MMSE decoders. We have observed that for such decoders, the diversity order was one for $N_T = N_R = N$ MIMO system. According to conservation theorem, MMSE has slightly higher diversity order but still there is a large gap in the performance from the ML decoder which has diversity order of N. SIC decoder also has slightly higher diversity order since the diversity order increases for detection of each information layer since the previously detected information layer interference has been cancelled. Lattice decoding on the other hand will narrow down the gap in the performance with the ML detection with reduced complexity in comparison to ML decoding. It has been shown in X. Ma et al., (2008) that the diversity order of LR based ZF/MMSE linear detector is the same as achieved by ML detector which is N for $N \times N$ MIMO systems.

From lattice view point, ML decoding is basically finding the closest vector problem (CVP) in the lattice. Finding CVP problem consists of two steps: (a) lattice reduction and (b) sphere decoding. Finding CVP exactly using sphere decoder is still computationally intensive and complexity is dependent on the SNR as well. There are two strategies for finding CVP using sphere decoding (First technique: find all lattice points inside a sphere and second technique: find the closest lattice point). First technique (U. Fincke et al., 1985) average complexity grows exponentially and the second technique (C. P. Schnorr et al., 1994) worst case complexity also grows exponentially. Instead of finding CVP exactly we may find it approximately using standard algorithms like LLL. LLL algorithm has polynomial time complexity and finds a reduced lattice which is approximately shortest. The computational complexity of LLL algorithm is $O(N^4)$ for integer bases (A. K. Lenstra et al., 1982) and $O(N^3 \log N)$ for real valued bases with i.i.d. normal distribution (C. Ling et al., 2007). This way of MIMO detection on a reduced lattice is also known as lattice reduction aided decoding which is illustrated in Fig. 10.6 (a). This method of finding the reduced lattice and applying linear detectors on the reduced lattice was suggested by L. Babai (1986) and H. Yao et al. (2002) also known as approximate lattice decoding. Basically we apply the low complexity MIMO decoders like ZF and MMSE on the reduced lattice. With LLL algorithm (refer to section 10.7), we find a reduced lattice (**G**) of **H**.

$$\mathbf{G} = \mathbf{HU} \tag{10.29}$$

where, **U** is a unimodular integer matrix.

We will apply the ZF and MMSE on this equivalent channel **G**.

LR–ML detection:

In ML detection we search for the closest point in the lattice. From lattice theory, since **H** and **G** generate the same lattice, they are related by a unimodular matrix **U** as follows.

$$\mathbf{G} = \mathbf{HU}$$

Therefore, the received signal vector can be rewritten as

$$\mathbf{r} = \mathbf{Hx} + \mathbf{n} = \mathbf{GU}^{-1}\mathbf{x} + \mathbf{n} = \mathbf{Gc} + \mathbf{n} \tag{10.30}$$

where, $\mathbf{c} = \mathbf{U}^{-1}\mathbf{x}$.

We can apply the ML detection and find an estimate for **c** as follows.

$$\hat{\mathbf{c}} = \underset{\mathbf{c} \in Z^K}{\arg\min} \left\| \mathbf{r} - \mathbf{G}\mathbf{c} \right\|^2$$

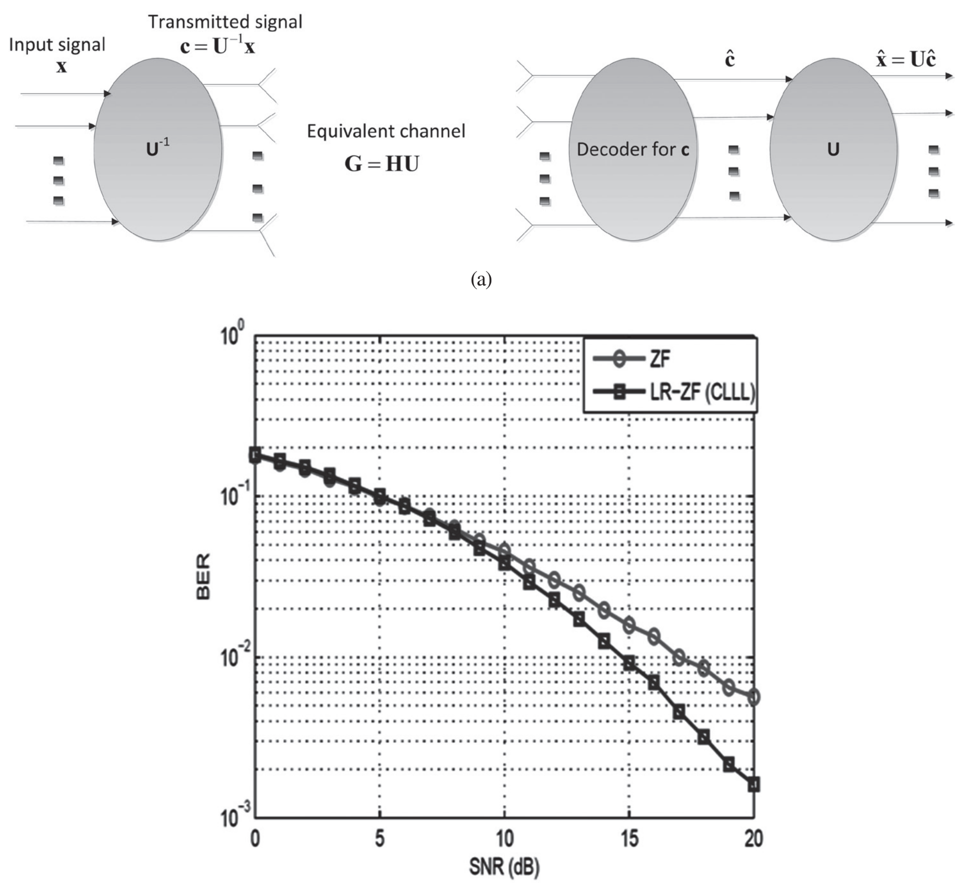
From estimate of **c**, we can find the estimate of **x** as follows.

$$\hat{\mathbf{x}} = \mathbf{U}\hat{\mathbf{c}}$$

LR–ZF detection:

Similarly, we can apply the ZF over the new matrix **G** instead of **H** which is more well-behaved and will minimize the noise enhancement. Note that $\mathbf{G}^+$ is the Moore–Penrose inverse of **G**.

$$\mathbf{G}^+\mathbf{r} = \mathbf{G}^+\left(\mathbf{H}\mathbf{x} + \mathbf{n}\right) = \mathbf{G}^+\left(\mathbf{G}\mathbf{U}^{-1}\mathbf{x} + \mathbf{n}\right) = \mathbf{G}^+\left(\mathbf{G}\mathbf{c} + \mathbf{n}\right) = \mathbf{c} + \mathbf{G}^+\mathbf{n} \tag{10.31}$$



(a)

(b)

Fig. 10.6 (a) Lattice reduction aided MIMO detection (b) BER performance comparison between ZF and LR-ZF for 2 × 2 MIMO system employing BPSK modulation over i.i.d. Rayeigh fading MIMO channel

Hence we can estimate **x** from **c** as follows.

$$\hat{\mathbf{x}} = \lceil \hat{\mathbf{c}} \rfloor$$

where, $\lceil \hat{\mathbf{c}} \rfloor$ denotes integer closest to $\hat{\mathbf{c}}$.

For example, $\lceil 0.7 \rfloor = 1, \lceil 0.3 \rfloor = 0$

BER performance comparison between ZF and LR-ZF for 2 × 2 MIMO system employing BPSK modulation scheme over i.i.d. Rayleigh fading MIMO channel is depicted in Fig. 10.6 (b). It can be observed that LR-ZF has superior BER performance than ZF.

LR-ZF-SIC detection:

For SIC detection, do QR decomposition of $\mathbf{G} = \mathbf{QR}$. Then multiply by $\mathbf{Q}^H$ to the received signal vector.

$$\mathbf{Q}^H\mathbf{r} = \mathbf{Q}^H\left(\mathbf{G}\mathbf{U}^{-1}\mathbf{x} + \mathbf{n}\right) = \mathbf{Q}^H\left(\mathbf{QRc} + \mathbf{n}\right) = \mathbf{Rc} + \mathbf{Q}^H\mathbf{n} \tag{10.32}$$

where, $\mathbf{c} = \mathbf{U}^{-1}\mathbf{x}$.

We can proceed with the same SIC detection techniques with this new equivalent system.

Review question 10.9 *What is LR-aided ML detection?*

Review question 10.10 *What is LR-aided ZF detection?*

Review question 10.11 *What is LR-aided ZF-SIC detection?*

10.7 Lattice reduction algorithms

A lattice is generated as the integer linear combination of some set of linearly independent vectors. A lattice in the n-D Euclidean space R^n is a set of the form

$$L = L(\mathbf{B}) = \left\{\sum_{i=1}^{n} u_i\mathbf{b}_i; u_i \in Z, i = 1,\cdots,n\right\}; \mathbf{B} = \left[\mathbf{b}_1,\cdots,\mathbf{b}_n\right]$$

A lattice L can be generated by different bases for $n \geq 2$ and hence there is no unique basis. We can obtain one basis from another basis by multiplying an integer unimodular matrix **U**.

Basically this is the result of three operations (illustrated in Fig. 10.7):

(a) exchanging two columns
(b) multiplying any column by -1 and
(c) adding an integer multiple of one column to another.

V-BLAST OSIC is an example of such operations.

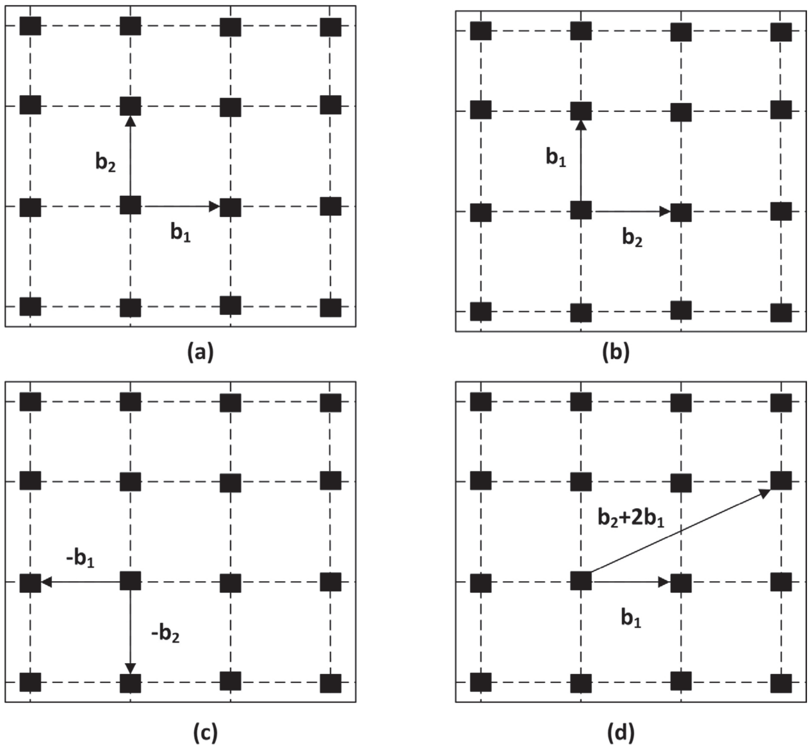


Fig. 10.7 (a) $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ basis vectors for the integer lattice Z^2 (b) exchanging two columns $\mathbf{b}_1$ and $\mathbf{b}_2$ (c) multiplying both columns by -1 (d) adding twice of column $\mathbf{b}_1$ to column $\mathbf{b}_2$

Example 10.13

What are Lattice reduction techniques?

Solution

There are many Lattice reduction techniques. Some of them are Lenstra, Lenstra and Lovasz (LLL), Korkin–Zolotarev (KZ) and Minkowski. Only LLL algorithm has polynomial time complexity to find the reduced basis vectors (C. P. Schnorr et al., 1994), hence, we will explore it in this section.

Most algorithms are developed for real valued lattices. For complex MIMO system,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Use equivalent real channel model as follows.

$$\begin{bmatrix} \mathrm{Re}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathbf{H}) & -\mathrm{Im}(\mathbf{H}) \\ \mathrm{Im}(\mathbf{H}) & \mathrm{Re}(\mathbf{H}) \end{bmatrix} \begin{bmatrix} \mathrm{Re}(\mathbf{x}) \\ \mathrm{Im}(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \mathrm{Re}(\mathbf{n}) \\ \mathrm{Im}(\mathbf{n}) \end{bmatrix} \quad (10.33)$$

Now the dimension will be doubled (say $m = 2N_T$ and $n = 2N_R$).

LLL reduced lattice:

A basis $\mathbf{H}_{red}$ with QL decomposition $\mathbf{H}_{red} = \mathbf{Q}_{red}\mathbf{L}_{red}$ is LLL reduced with parameter δ (usually taken $\frac{3}{4}$), $\frac{1}{4} \le \delta \le 1$, if the following two conditions hold true.

(a) *Size reduction*

$$\left|\mathbf{L}_{l,k}\right| \le \frac{1}{2}\left|\mathbf{L}_{l,l}\right|; 1 \le k < l \le m \tag{10.34}$$

Note that column l is to the right of column k. The condition (a) says the diagonal components of the $\mathbf{L}_{red}$ are at least double as the off-diagonal components of the same row. This is called the size-reduction condition and this ensures that there is no significant projection of one column on another. If it is not satisfied for (l, k) pair, we deduct an integer multiple of the l^{th} column from the k^{th} column so that this condition is satisfied. The size reduction is carried out by subtracting integer multiples of the right column with index l from the left column with index k. This condition does not guarantee a minimal basis, there is another condition called as Lovasz condition which will ensure the correct sorting of the columns.

(b) *Sorting*

$$\delta\left|\mathbf{L}_{red}\left[k+1,k+1\right]\right|^2 \le \left|\mathbf{L}_{red}\left[k,k\right]\right|^2 + \left|\mathbf{L}_{red}\left[k+1,k\right]\right|^2; 1 \le k \le m-1 \tag{10.35}$$

The condition (b) ensures proper sorting since the lengths of the columns are only compared on the basis of a little 2 × 2 submatrix. If the above condition is not satisfied, we will interchange the columns. Why are we considering 2 × 2 submatrices? This will reduce the computational complexity at the price of lower performance especially for big channel matrices. The columns have to be ordered according to their lengths, shortest columns right and largest on the left. It may be noted that QL decomposition gives such a lower triangular matrix, that's why the sorting has been much simplified. So, one starts LLL algorithm with initial inputs as Q and L.

Example 10.14

Explain the above condition (b) with the help of an example.

Solution

For instance for 5 × 5 lower triangular matrix $\mathbf{L} = \begin{bmatrix} L_{11} & 0 & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 & 0 \\ L_{41} & L_{42} & L_{43} & L_{44} & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{bmatrix}$; we will consider the 2×2 sub-matrices $\begin{bmatrix} L_{1,1} & \\ L_{2,1} & L_{2,2} \end{bmatrix}$, $\begin{bmatrix} L_{2,2} & \\ L_{3,2} & L_{3,3} \end{bmatrix}$, $\begin{bmatrix} L_{3,3} & \\ L_{4,3} & L_{4,4} \end{bmatrix}$ and $\begin{bmatrix} L_{4,4} & \\ L_{5,4} & L_{5,5} \end{bmatrix}$. If this condition is not satisfied, columns will be exchanged. This condition is also known as Lovasz condition. Small value of δ leads to fast convergence, whereas large value δ leads to better basis. Usual choice of $\delta = \frac{3}{4}$. Lovasz condition for 2 × 2 submatrix $\begin{bmatrix} L_{1,1} & \\ L_{2,1} & L_{2,2} \end{bmatrix}$ is $\frac{3}{4}L_{2,2}^2 \le L_{1,1}^2 + L_{2,1}^2$.

Review question 10.12 *What is size reduction?*

Review question 10.13 *What is Lovasz condition?*

10.7.1 LLL algorithm for lattice reduction

The LLL algorithm for reduction of lattice basis does not give optimal solution or minimal basis, but has polynomial time complexity. All the vectors and matrices are converted from complex to real hence the matrix sizes have been doubled.

The inputs to the LLL algorithm are **Q**, **L** and **U**.

$$\mathbf{L} = \begin{bmatrix} L_{11} & 0 & 0 & \cdots & 0 \\ L_{21} & L_{22} & 0 & \cdots & 0 \\ L_{31} & L_{32} & L_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{m1} & L_{m2} & L_{m3} & \cdots & L_{mm} \end{bmatrix}; \mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

The outputs are reduced matrices viz., $\mathbf{Q}_{red}$, $\mathbf{L}_{red}$ and $\mathbf{U}_{red}$.

LLL LR algorithm (A. K. Lenstra et al., 1982)

~~~~~~

Initialization:

$\mathbf{Q}_{red} = \mathbf{Q}, \mathbf{L}_{red} = \mathbf{L}, \mathbf{U} = \mathbf{I}_m$ (%initial inputs, U is a unimodular matrix)

$k = m - 1$ (% k is the column under consideration and start from the last but second column, note that m=2N_T)

while $k \geq 1$ (% for all columns of the matrix from the 1st to the last)

for $l = k + 1, \ldots, m$ (% l is larger than k, l^{th} is in the right side of k^{th} column)

$\mu[l, k] = \left\lceil \frac{L_{red}[l,k]}{L_{red}[l,l]} \right\rfloor$ (% ratio of the off-diagonal and diagonal element in the same row)

if $\mu[l, k] \neq 0$ (% off-diagonal element is larger than diagonal element)

$L_{red}[l : m,k] = L_{red}[l : m,k] - \mu L_{red}[l : m,l]$ (% subtract integer multiple of l^{th} column from k^{th} column of $\mathbf{L}_{red}$ which has only l:m elements for l^{th} column)

$U[:,k] = U[:,k] - \mu U[:,l]$

(% subtract integer multiple of l^{th} column from k^{th} column of **U**)

end

if $\delta \left|\mathbf{L}_{red}[k+1,k+1]\right|^2 \geq \left|\mathbf{L}_{red}[k,k]\right|^2 + \left|\mathbf{L}_{red}[k+1,k]\right|^2$

Exchange columns k and $k + 1$ in $\mathbf{L}_{red}$ and **U** (% when we
~~~~~~

interchange k and k+1 column, now the new matrix will no longer be lower triangular, we need to force zero the **L**(k,k+1) element zero)

$$\boldsymbol{\Theta} = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}; \alpha = \frac{\mathbf{L}_{red}\left[k+1, k+1\right]}{\left\|\mathbf{L}_{red}\left[k : k+1, k+1\right]\right\|}; \beta = \frac{\mathbf{L}_{red}\left[k, k+1\right]}{\left\|\mathbf{L}_{red}\left[k : k+1, k+1\right]\right\|}$$

$\mathbf{L}_{red}[k : k+1,\ 1 : k+1] = \boldsymbol{\Theta}\mathbf{L}_{red}\ [k : k+1,\ 1 : k+1]\boldsymbol{\Theta}^T$

(%Calculate Givens rotation matrix $\boldsymbol{\Theta}$ such that element **L**(k,k+1) become zero)

$\mathbf{Q}_{red}[;\ k : k+1] = \mathbf{Q}_{red}\ [;\ k : k+1]\boldsymbol{\Theta}^T$

(% consider all columns, but only the k and k+1 rows only, Givens

rotation operate on columns only)

$k = \min\ \{k+1,\ m-1\}$

else

$k = k - 1$

end

end

~~~~~~

Example 10.15

What is Givens rotation?

Solution

Consider a matrix $\boldsymbol{\Theta}(i,\ k,\ \theta)$ which is an $N \times N$ identity matrix except for the elements $\boldsymbol{\Theta}^*_{i,i} = \boldsymbol{\Theta}_{k,k} = \cos\theta = \alpha$ and $-\boldsymbol{\Theta}^*_{i,k} = \boldsymbol{\Theta}_{k,i} = \sin\theta = \beta$ (V. Kuhn, 2006). It gives a rotation of θ in the N-D vector space.

$$\boldsymbol{\Theta}(i,\ k,\ \theta) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \alpha & \cdots & -\beta & & \\ & & \vdots & & \vdots & & \\ & & \beta^* & \cdots & \alpha^* & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

Define

$$\boldsymbol{\Theta}(i,\ k,\ \theta) = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 0 & \mathbf{I} & 0 & 0 \\ 0 & \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{bmatrix}$$
~~~~~~

where, i is the row that contains $\alpha = \cos\theta$ and $-\beta = -\sin\theta$; k is the row that contains $\beta = \sin\theta$ and $\alpha = \cos\theta$.

If θ is chosen properly it can force the i^{th} element of a column vector equal to zero as shown below. It also force the k^{th} element as $\sqrt{x_i^2 + x_k^2}$. Assume x_i and x_k are the i^{th} and k^{th} element of the column vector, then

$$\alpha = \cos\theta = \frac{x_k}{\sqrt{|x_i|^2 + |x_k|^2}} \text{ and } \beta = \sin\theta = \frac{x_i}{\sqrt{|x_i|^2 + |x_k|^2}}$$

$$\mathbf{y} = \Theta(i, k, \theta)\mathbf{x}$$

$$\Rightarrow \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha = \frac{x_k}{\sqrt{|x_i|^2 + |x_k|^2}} & \cdots & -\beta = -\frac{x_i}{\sqrt{|x_i|^2 + |x_k|^2}} & \\ & & \vdots & & \\ & & \beta = \frac{x_i^*}{\sqrt{|x_i|^2 + |x_k|^2}} & \cdots & \alpha^* = \frac{x_k^*}{\sqrt{|x_i|^2 + |x_k|^2}} & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_k \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ 0 \\ \vdots \\ \sqrt{|x_i|^2 + |x_k|^2} \\ \vdots \\ x_N \end{bmatrix}$$

Example 10.16

Explain LLL algorithm to find $\mathbf{Q}_{red}$, $\mathbf{R}_{red}$, $\mathbf{U}_{red}$ for a simple 2 × 2 matrix $\mathbf{H} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Solution

We can find the condition number of **H** by using MATLAB command *cond(**H**)=14.9330.*

Condition number of **H** should be closer to 1.

Apply LLL algorithm, the inputs are:

The input to the LLL algorithm above is **Q**, **L** and **U=I.**

In MATLAB, one can find

[Q R]=qr(fliplr(H));

L=fliplr(flipud(R));

Q=fliplr(Q);

Now $\mathbf{Q} = \begin{bmatrix} -0.8944 & -0.4472 \\ 0.4472 & -0.8944 \end{bmatrix}; \mathbf{L} = \begin{bmatrix} 0.4472 & 0 \\ -3.1305 & -4.4721 \end{bmatrix}; \mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Size reduction:

$$\mu_{21} = \frac{|L_{21}|}{|L_{22}|} = \frac{3.1305}{4.4472} = 0.7039$$

Since the nearest integer of μ_{21} is 1, hence,
Subtract column 1 from column 2 of $\mathbf{Q}_{red}$ and $\mathbf{L}_{red}$.

$$\mathbf{Q}_{red} = \begin{bmatrix} -0.4472 & -0.4472 \\ 0.4472 & -0.8944 \end{bmatrix}; \mathbf{L}_{red} = \begin{bmatrix} 0.4472 & 0 \\ 1.3167 & -4.4721 \end{bmatrix}; \mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Sorting:

$$\frac{3}{4}\left|\mathbf{L}_{red}[2,2]\right|^2 = 14.9998 \geq \left|\mathbf{L}_{red}[1,1]\right|^2 + \left|\mathbf{L}_{red}[2,1]\right|^2 = 1.9337$$

Column 1 and 2 should be interchanged for $\mathbf{Q}_{red}$ and $\mathbf{L}_{red}$.

$$\mathbf{Q}_{red} = \begin{bmatrix} -0.4472 & -0.4472 \\ -0.8944 & 0.4472 \end{bmatrix}; \mathbf{L}_{red} = \begin{bmatrix} 0 & 0.4472 \\ -4.4721 & 1.3167 \end{bmatrix}; \mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now $\mathbf{L}_{red}$ is no more lower triangular. We need to force zero the (1,2) element of $\mathbf{L}_{red}$, apply Given rotation.

$$\mathbf{G}_{1,2,\theta} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \frac{1.3167}{\sqrt{1.3167^2 + 0.4472^2}} = 0.9469 & -\frac{0.4472}{\sqrt{1.3167^2 + 0.4472^2}} = -0.3216 \\ 0.3216 & 0.9469 \end{bmatrix}$$

Hence,

$$\mathbf{Q}_{red} = \begin{bmatrix} -0.4472 & -0.4472 \\ -0.8944 & 0.4472 \end{bmatrix}\left[\mathbf{G}_{1,2,\theta}\right]^T = \begin{bmatrix} -0.2796 & -0.5673 \\ -0.9907 & 0.1358 \end{bmatrix};$$

$$\mathbf{L}_{red} = \mathbf{G}_{1,2,\theta}\begin{bmatrix} 0 & 0.4472 \\ -4.4721 & 1.3167 \end{bmatrix} = \begin{bmatrix} 1.4382 & 0 \\ -4.2346 & 1.3906 \end{bmatrix}; \mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, $\mathbf{H}_{red} = \mathbf{Q}_{red}\mathbf{L}_{red} = \begin{bmatrix} -0.4022 & -0.8159 \\ -0.1936 & 2.5911 \end{bmatrix}$.

Now the *cond*($\boldsymbol{H}_{red}$) = *6.1529.*
It is closer to 1.

10.8 Summary

Figure 10.8 gives the summary of this chapter. In this chapter on Advanced MIMO detection, we have discussed about the two different transmission schemes for spatially multiplexed MIMO systems, viz., V-BLAST and D-BLAST. At the receiver of such spatially multiplexed MIMO systems, we have discussed about the advanced MIMO detection schemes like SIC, OSIC and LR based detectors.

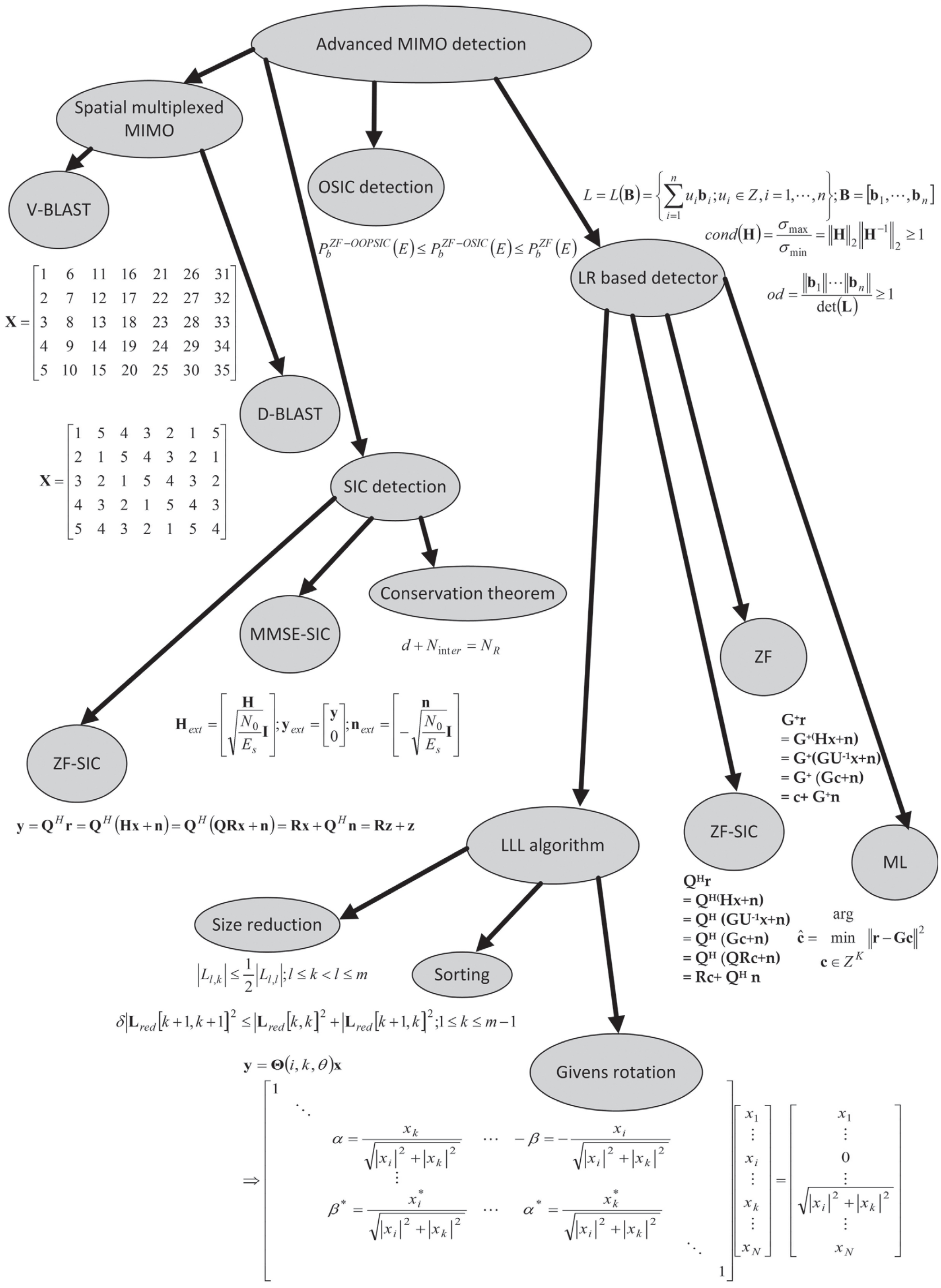


Fig. 10.8 Chapter 10 in a nutshell

Exercises

Exercise 10.1

Is it possible to design an extended ZF detector which will behave like MMSE detector? Explain in few words.

Exercise 10.2

What are hard and soft decision? Explain this with respect to ML and LLR based detection.

Exercise 10.3

Show that approximate BER for MMSE MIMO detector is given by

$$\text{BER} \cong \left(\frac{1}{1+\bar{\gamma}}\right)^{N_R-N_T+1}\left(\frac{N_R+\frac{1}{\bar{\gamma}}}{N_R+1+\frac{1}{\bar{\gamma}}}\right)^{N_T-1}$$

Exercise 10.4

Show that the bit error probability ($P_{e,i}$) and LLR (λ_i) are related by

$$P_{e,i} = P\left(\hat{x}_i \neq x_i\right) = \frac{1}{1+e^{|\lambda_i|}}$$

Exercise 10.5

Given two basis of a lattice as $\mathbf{h}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}; \mathbf{h}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$. Find new basis which are nearly orthogonal that generate the same lattice.

Exercise 10.6

Prove that $\left(1+\frac{1}{4}\frac{\beta^j-\beta}{\beta-1}\right) \leq \beta^{j-1}$ using method of induction.

Exercise 10.7

Apply standard LLL algorithm to find the reduced $\mathbf{Q}_{red}$, $\mathbf{R}_{red}$, $\mathbf{U}_{red}$ for a simple 2 × 2 matrix $\mathbf{H} = \begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$.

References

1. Babai, L. 1986. 'On Lovasz lattice reduction and the nearest lattice point problem'. *Combinatorica*. 6(1). 1–13.
2. Bai, L. and J. Choi. 2012. *Low Complexity MIMO Detection*. New York: Springer.
3. Barry, J. R., E. A. Lee, and D. G. Messerschmitt. 2010. *Digital Communications*. New Delhi: Springer.

4. Bohnke, R., D. Wubben, V. Kuhn, and K. D. Kammeyer. 2003. ‘Reduced complexity MMSE detection for BLAST architectures’. In *Proc. IEEE Proceedings of Global Conference on Telecommunications*. San Francisco, CA.
5. Casella, G. and R. L. Berger. 2002. *Statistical Interference*. Pacific Grove: Thomson Learning.
6. Cho, Y. S., J. Kim, W. Y. Yang, and C.-G. Kang. 2010. *MIMO–OFDM wireless communications using MATLAB*. Singapore: Wiley.
7. Choi, J. 2010. *Optimal Combining and Detection*. Cambridge: Cambridge University Press.
8. David, H. A. and H. N. Nagaraja. 2003. *Order Statistics*, 3rd ed. New York: Wiley Interscience.
9. Duman, T. M. and A. Ghrayeb. 2007. *Coding for MIMO Communication Systems*. Chichester: John Wiley & Sons.
10. Fincke, U. and M. Pohst. Apr. 1985. ‘Improved methods for calculating vectors of short length in a lattice, including complexity analysis’. *Math. Comput.* 44. 463–471.
11. Foschini, G. J. 1996. ‘Layered space-time architecture for wireless communications in a fading environment when using multiple-element antenna’. *Bell Lab. Tech. Journ.* 1. 41–59.
12. Foschini, G. J., G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky. Nov. 1999. ‘Simplified processing for high spectral efficiency wireless communication employing multielement arrays’. *IEEE Trans. Commun.* 17(11). 1841–1852.
13. Han, J., Q.-M. Cui, X.-F. Tao, and P. Zhang. Feb. 2010. ‘SER bound for ordered ZF-SIC receiver in M-QAM MIMO system’. *Journal of China Universities of Posts and Telecommunications*. 51–55.
14. Kim, S. W. Dec. 2003. ‘Log-likelihood ratio based detection ordering for the V-BLAST’. In *Proc. GLOBECOM*. 1. 292–296.
15. Kuhn, V. 2006. *Wireless Communications over MIMO Channels*. Chichester: John Wiley & Sons.
16. Lenstra, A. K., J. H. W. Lenstra, and L. Lovasz. 1982. ‘Factorizing polynomials with rational coefficients’. *Math. Ann.* 216(4). 515–534.
17. Ling, C. and N. Howgrave-Graham. June 2007. ‘Effective LLL reduction for lattice decoding’. In *Proc. Int. Symp. Inform. Theory*. Nice, France.
18. Ma, X. and W. Zhang. Feb. 2008. ‘Performance analysis of MIMO systems with lattice-reduction aided linear equalization’. *IEEE Trans. Comm.* 56(2). 309–318.
19. Micciancio, D. and S. Goldwasser. 2002. *Complexity of Lattice Problems A Cryptographic Perspective*. Boston: Kluwer Academic Publishers.
20. Proakis, J. G. and M. Salehi. 2008. *Digital Communications*. New York: McGraw-Hill.
21. Schnorr, C. P. and M. Euchner. 1994. ‘Lattice basis reduction: improved practical algorithms and solving subset sum problems’. *Math. Program.* 66. 181–191.
22. Shiu, D.-S. and M. Kahn. June 1999. ‘Layered space-time codes for wireless communications using multiple transmit antennas’. *IEEE International Conference on Communications*. 1. 436–440.
23. Stuber, G. L. 2001. *Principles of Mobile Communication*. Dordrecht: Kluwer Academic Publishers.
24. Yao, H. and G. W. Wornell. Nov. 2002. ‘Lattice-reduction-aided detectors for MIMO communication systems’. In *Proc. Globecom*. Taipei, China.
25. Zanella, A., M. Chiani, and M. Z. Win. May 2005. ‘MMSE reception and successive interference cancellation for MIMO systems with high spectral efficiency’. *IEEE Trans. Comm.* 4(3). 1244–1253.

Antenna Selection and Spatial Modulation

CHAPTER 11

11.1 Introduction

Multiple-input multiple-output (MIMO) systems' capacity increases and the bit error rate minimizes with the number of antennas as compared to single-input single-output (SISO) system. But, they have higher fabrication cost and energy consumption due to multiple radio frequency (RF) chains. An RF chain usually has low noise amplifier, frequency converters, analog–digital and digital–analog converters, filters, etc. A dedicated RF chain is needed for every antenna which makes implementation cost and hardware complexity higher. Antenna selection minimizes this by using lesser number of RF chains and by connecting selected antennas with RF chains with the help of switches. Select the best set of antennas at the transmitter or receiver so as to maximize the channel capacity or received signal-to-noise ratio (SNR). In the next section, we will discuss about the transmit antenna selection (also called as hard antenna selection) over η–μ fading channels. Then we will discuss about the soft antenna selection for closely spaced antennas. Then we will briefly discuss about spatial modulation and combine spatial modulation with antenna selection. In spatial modulation, there are two information bearing units: (a) Transmit antenna index and (b) Symbol from signal constellation, which is transmitted from antenna corresponding to the selected transmit antenna index. It is a relatively new MIMO technique, which was proposed R. Mesleh et al. (2008). Some of the advantages are (a) higher capacity (b) reduced hardware complexity and (c) avoidance of transmit antenna synchronization. One of the problems with spatial modulation is that the link for the selected antenna may be down, and then the performance of the spatial modulation is worst. In order to overcome this, one can do antenna selection before applying spatial modulation. Select a subset of antenna with the best links and apply spatial modulation on those selected antennas. As reported by B. Kumbhani et al., (2014), there is significant performance improvement in MIMO systems which combine antenna selection with spatial modulation.

11.2 Transmit antenna selection (TAS) over η–μ fading channels

We have considered a $N_T \times N_R$ MIMO system with only one RF chain at the transmitter as depicted in Fig. 11.1. This kind of antenna selection is also called as hard antenna selection. Soft antenna selection for closely spaced antennas at the transmitter and receiver will be discussed in the next section. A single transmitting antenna that maximizes the SNR at the receiver is selected to transmit the message bits as shown by dashed lines in Fig. 11.1. We have shown antenna 2 as the selected antenna in Fig. 11.1 for illustration purpose only, but it could be any one of the transmit antennas which maximizes the received SNR. Basically, there will be a RF switch which will connect to the selected transmitting antenna and sends signals from that particular transmitting antenna only. All other antennas at the transmitter are sitting idle. Hence there are no issue for inter-antenna interference. In this MIMO technique, after single antenna selection at the transmitter, the whole system looks like a SIMO system and one can apply suitable diversity combining scheme like maximal ratio combining (MRC) or selection combining (SC) at the receiver. Based on the type of the diversity scheme employed at the receiver, there are two types of TAS as follows.

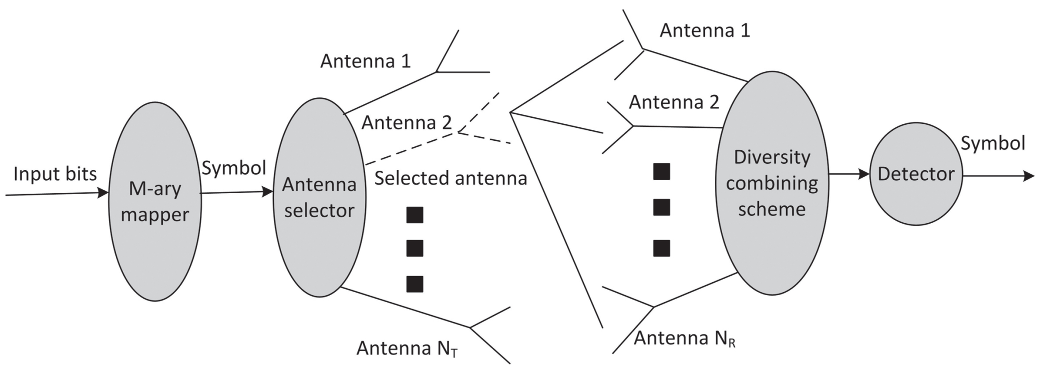


Fig. 11.1 Transmit antenna selection (hard antenna selection) in MIMO systems

11.2.1 Types of TAS

Selection of the best antenna at the transmitter (TAS) can be used in conjunction with either selection combining (SC) or maximal ratio combining (MRC) at the receiver. Hence there are two types of TAS: (a) TAS/MRC and (b) TAS/SC.

Let the fading coefficients from i^{th} transmitting antenna to j^{th} receiving antenna is denoted by h_{ij}, $i \in [1, N_T]$, $j \in [1, N_R]$. Two types of TAS are described below.

(a) *TAS/SC:* In this case, TAS is applied at the transmitter and SC is applied at the receiver. We select only one antenna at the receiver that receives the highest SNR. We refer to such systems as TAS/SC ($N_T \rightarrow 1; N_R \rightarrow 1$) systems. The link which gives the highest received SNR is determined by

$$I_{ij} = \underset{\substack{1 \le i \le N_T \\ 1 \le j \le N_R}}{\arg\max} \left\{ h_{t,ij} = \left| h_{i,j} \right|^2 \right\} \tag{11.1}$$

where, i,j represent the antennas corresponding to the best link at the transmitter and receiver, respectively, and I_{ij} denotes the best link. Only the antennas that correspond to the best link are active at a time in this case. So we have single RF chain at the transmitter as well as at the receiver.

(b) *TAS/MRC:* In this case, TAS is applied at the transmitter and MRC is applied at the receiver. It is assumed that the number of RF chains at receiver is same as the number of receiver antennas. The resulting received SNR for MRC combining scheme is given by $\gamma_t = \sum_{n=1}^{N_r} \gamma_n$, where γ_n is the instantaneous SNR of the n^{th} branch. The transmitting antenna that maximizes SNR at the receiver can be determined by

$$I_i = \underset{1 \le i \le N_T}{\arg\max} \left\{ h_{t,i} = \sum_{j=1}^{N_R} \left| h_{i,j} \right|^2 \right\} \tag{11.2}$$

where, I_i is the transmitting antenna that maximizes the received SNR. Such systems are referred to as TAS/MRC ($N_T \rightarrow 1; N_R$) systems.

11.2.2 η–μ channel model

We will assume the fading envelope ($|h_{ij}|$) to be η–μ distributed. The pdf of SNR at any single antenna output is given as (M. Yacoub, 2007)

$$p_{\gamma_{\eta-\mu}}(x) = \frac{2\sqrt{\pi}\mu^{\mu+\frac{1}{2}} h^{\mu} x^{\mu-\frac{1}{2}} e^{-\frac{2\mu x h}{\overline{\gamma}}}}{\Gamma(\mu) H^{\mu-\frac{1}{2}} \overline{\gamma}^{\mu+\frac{1}{2}}} I_{\mu-\frac{1}{2}}\left(\frac{2\mu H x}{\overline{\gamma}}\right); \overline{\gamma} = E(\gamma) \tag{11.3}$$

where all parameters were defined in chapter 2, and $\overline{\gamma}$ is the average SNR.

Note that h and H parameters are dependent on η (refer to section 2.4.3). The cdf of received SNR at any single antenna output is given as

$$P_{\gamma_{\eta-\mu}}(x) = \frac{2\sqrt{\pi}(\mu)^{\mu+\frac{1}{2}} h^{\mu}}{\Gamma(\mu) H^{\mu-\frac{1}{2}} (\overline{\gamma})^{\mu+\frac{1}{2}}} \int_0^x \gamma^{\mu-\frac{1}{2}} e^{\left(-\frac{2\mu\gamma h}{\overline{\gamma}}\right)} I_{\mu-\frac{1}{2}}\left(\frac{2\mu\gamma H}{\overline{\gamma}}\right) d\gamma \tag{11.4}$$

The modified Bessel's function can be represented in infinite series form as

$$I_r(\omega) = \sum_{i=0}^{\infty} \frac{1}{i!\Gamma(i+r+1)} \left(\frac{\omega}{2}\right)^{r+2i},$$

hence, $$I_{\mu-\frac{1}{2}}\left(\frac{2\mu H\gamma}{\overline{\gamma}}\right) = \sum_{i=0}^{\infty} \frac{1}{i!\Gamma(i+\mu+0.5)} \left(\frac{\mu H\gamma}{\overline{\gamma}}\right)^{\mu-0.5+2i}$$

The integral in the above expression can be solved by using (I. S. Gradshteyn et al., 2008) (8.445, 3.351.1). After further mathematical simplifications, the cdf can be given as

$$P_{\gamma}^{\eta-\mu} = \frac{2^{1-2\mu}\sqrt{\pi}}{h^{\mu}\Gamma(\mu)} \sum_{i=0}^{\infty} \frac{\gamma_{inc}^{low}\left(2\mu+2i, \frac{2\mu h x}{\overline{\gamma}}\right)}{i!\Gamma(\mu+i+0.5)} \left(\frac{H}{2h}\right)^{2i} \tag{11.5}$$

where, $\gamma_{inc}^{low}(\bullet,\bullet)$ is lower incomplete gamma function.

Gamma functions

1. Complete Gamma function:

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$$

For integer a, $\Gamma(a) = (a-1)!$

MATLAB command is "gamma(a)".

2. Upper incomplete Gamma function:

$$\Gamma_{inc}^{up}(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$$

Hence, $\Gamma_{inc}^{up}(a,0) = \Gamma(a)$

For integer a, $\Gamma_{inc}^{up}(a,x) = (a-1)!e^{-x}\sum_{k=0}^{a-1}\frac{x^k}{k!}$

MATLAB command is "gammainc(a,x, 'upper')".

3. Lower incomplete Gamma function

$$\gamma_{inc}^{low}(a,x) = \int_0^x t^{a-1}e^{-t}dt = a^{-1}x^a\,{}_1F_1(a;1+a;-x)$$

MATLAB command is "gammainc(a, x, 'lower')".

For integer a, $\gamma_{inc}^{low}(a,x) = (a-1)!\left(1-e^{-x}\sum_{k=0}^{a-1}\frac{x^k}{k!}\right)$

Hence, $\Gamma_{inc}^{up}(a,x) + \gamma_{inc}^{low}(a,x) = \Gamma(a)$

Review question 11.1 *What is the Gamma function?*

Review question 11.2 *What is the lower incomplete Gamma function?*

Review question 11.3 *What is the upper incomplete Gamma function?*

11.2.3 Probability density function (PDF) and moment generating function (MGF) of received SNR after antenna selection

Each $h_{t,ij}$ or $h_{t,i}$ for $i \in [1, N_T]$ & $j \in [1, N_R]$ obtained from Eq. (11.1) or Eq. (11.2) are arranged in ascending order such that $h_{t,(1)} \le h_{t,(2)} \le \cdots \le h_{t,(N)}$, where, $h_{t,(\bullet)}$ is the random variable (RV) obtained after the arrangement in ascending order and $N = N_R N_T$ for TAS/SC and $N = N_T$ TAS/MRC systems, respectively. In TAS/MRC system, we select the transmitting antenna corresponding to the highest channel gain, h_{t,N_T} when MRC diversity technique is used at the receiver. In TAS/SC, the transmit and receive antenna pair that maximizes the received SNR $h_{t,N_T N_R}$ is selected. The pdf of maximum received SNR in such a system can be given by (H. A. David et al., 2003 and B. Kumbhani et al., 2015).

$$p_{h_t^{\max}}(x) = N\left[P_\gamma^{\eta-\mu}(x)\right]^{N-1} p_{\gamma_{\eta-\mu}}(x)$$

$$= N\left[\frac{2^{1-2\mu}\sqrt{\pi}}{h^\mu\Gamma(\mu)}\sum_{i=0}^{\infty}\frac{\gamma_{inc}^{low}\left(2\mu+2i,\frac{2\mu hx}{\overline{\gamma}}\right)}{i!\,\Gamma(\mu+i+0.5)}\left(\frac{H}{2h}\right)^{2i}\right]^{N-1}\frac{2\sqrt{\pi}\mu^{\mu+\frac{1}{2}}h^\mu x^{\mu-\frac{1}{2}}e^{-\frac{2\mu xh}{\overline{\gamma}}}}{\Gamma(\mu)H^{\mu-\frac{1}{2}}\overline{\gamma}^{\mu+\frac{1}{2}}}I_{\mu-\frac{1}{2}}\left(\frac{2\mu Hx}{\overline{\gamma}}\right) \quad (11.6)$$

In the above expression, modified Bessel function and incomplete gamma function can be represented in the series form (I. S. Gradshteyn et al., 2008) as follows

$$I_r(\omega) = \sum_{i=0}^{\infty}\frac{1}{i!\,\Gamma(i+r+1)}\left(\frac{\omega}{2}\right)^{r+2i} \quad (11.7)$$

$$\gamma_{inc}^{low}(a,x) = \frac{e^{-a}}{a}\sum_{k=0}^{\infty}\frac{x^{a+k}}{(a+1)_k} \quad (11.8)$$

So, the PDF and MGF of the received instantaneous SNR (B. Kumbhani et al., 2015) can be given as Eq. (11.9) and Eq. (11.10).

$$p_\gamma(x) = 2Nl_\mu\sum_{i_{1,2,\cdots,N}=0}^{\infty}\sum_{j_{1,2,\cdots,N-1}=0}^{\infty}\frac{2^{j_\Sigma}H^{2i_\Sigma}q_\mu e^{-\overline{\lambda}_\mu x}x^{r-1}}{h^{N\mu+2i_\Sigma}}\frac{1}{\prod_{p=1}^{N}\Gamma(\mu')i_p!}\frac{1}{\prod_{p=1}^{N-1}(2\mu')_{j_p}(\mu+i_p)} \quad (11.9)$$

$$M_\gamma(s) = 2Nl_\mu\sum_{i_{1,2,\cdots,N}=0}^{\infty}\sum_{j_{1,2,\cdots,N-1}=0}^{\infty}\frac{2^{j_\Sigma}H^{2i_\Sigma}q_\mu\Gamma(r)\left(\overline{\lambda}_\mu+s\right)^{-r}}{h^{N\mu+2i_\Sigma}}\frac{1}{\prod_{p=1}^{N}\Gamma(\mu')i_p!}\frac{1}{\prod_{p=1}^{N-1}(2\mu')_{j_p}(\mu+i_p)} \quad (11.10)$$

where the compact notations used are

(a) $\sum_{i_{1,2,\cdots,N}=0}^{\infty}\sum_{j_{1,2,\cdots,N-1}=0}^{\infty} = \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_N=0}^{\infty}\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty}\cdots\sum_{j_{N-1}=0}^{\infty}$,

(b) $r = 2N\mu + 2i_\Sigma + j_\Sigma = 2N\mu + 2\sum_{p=1}^{N} i_p + \sum_{p=1}^{N-1} j_p$,

(c) $\mu' = \mu + i_p + 0.5$,

(d) $l_\mu = \left(\frac{\sqrt{\pi}}{\Gamma(\mu)}\right)^N$,

(e) $\overline{\lambda}_\mu = \frac{2N\mu h}{\overline{\gamma}}$,

(f) $q_\mu = \left(\frac{\mu h}{\bar{\gamma}}\right)^r$

It should be noted that Eq. (11.9) applies to TAS/MRC as well as TAS/SC systems. For TAS/MRC γ_t is $\eta - N_R\mu$ distributed as it is sum of N_R independent and identically distributed (i.i.d.) η–μ square variates (M. Yacoub, 2007). So, in (11.9), (11.10) and all subsequent expressions, μ shall be interpreted as $N_R\mu$ for TAS/MRC systems. N_R may be interpreted as the number of antennas selected at the receiver, i.e.,

(a) $N_R = 1$ for TAS/SC ($N_T \to 1$; $N_R \to 1$) system while
(b) $N_R = N_R$ for TAS/MRC ($N_T \to 1$; N_R) system.

Review question 11.4 *What is the distribution of sum of N_R i.i.d. η–μ square variates?*

Review question 11.5 *What is the infinite series form for lower incomplete Gamma function?*

Review question 11.6 *What is the infinite series form for modified Bessel's functions?*

11.2.4 Exact probability of error

In this section, we derive expression for exact symbol error rate (SER) of TAS/MRC systems over η–μ fading channels for various modulation techniques. SER for a wireless communication system can be calculated by averaging the conditional error probability (CEP) over the pdf of received SNR. CEP for various modulation schemes can be given by

$$P_e(\gamma) = aQ\left(\sqrt{b\gamma}\right) - cQ^2\left(\sqrt{b\gamma}\right) \tag{11.11}$$

where, $Q(\cdot)$ is the Gaussian Q-function, γ is SNR and a, b and c are modulation dependent parameters. The values of a, b and c are listed in Table 11.1 for different digital modulation techniques.

Table 11.1

Modulation parameters for various modulation schemes (Y. Chen et al., 2004 and A. Goldsmith, 2005)

Modulation scheme	*a*	*b*	*c*
Binary phase shift keying (BPSK)	1	2	0
M-ary phase shift keying (MPSK)	2	$2\sin\left(\frac{\pi}{M}\right)$	0
Quadrature phase shift keying (QPSK)	1	2	1
M-ary quadrature amplitude modulation (MQAM)	$4\left(\frac{\sqrt{M}-1}{\sqrt{M}}\right)$	$\frac{3}{M-1}$	$4\left(\frac{\sqrt{M}-1}{\sqrt{M}}\right)^2$
M-ary pulse amplitude modulation (MPAM)	$2\left(\frac{M-1}{M}\right)$	$\frac{6}{M^2-1}$	0

Review question 11.7 *Write down the CEP for MQAM.*

Review question 11.8 *Write down the CEP for BPSK.*

Review question 11.9 *Write down the CEP for MPAM.*

Review question 11.10 *Write down the CEP for QPSK.*

Review question 11.11 *Write down the CEP for MPSK.*

The SER can be given by

$$\overline{P_e} = \int_0^\infty P_e(\gamma)\, p_{\gamma_{t(N)}}(\gamma)\, d\gamma \tag{11.12}$$

The exact SER can be evaluated using Craig's alternate form of Q function

$\left(Q(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta, Q^2(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{4}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta\right)$ as

$$\overline{P_e} = \int_0^\infty \left[aQ\left(\sqrt{bx}\right) - cQ^2\left(\sqrt{bx}\right)\right] p_{\gamma_{t(N)}}(x)\, dx$$

$$= \underbrace{\frac{a}{\pi}\int_0^{\frac{\pi}{2}} M_\gamma\left(\frac{b}{2\sin^2\theta}\right) d\theta}_{I_1} - \underbrace{\frac{c}{\pi}\int_0^{\frac{\pi}{4}} M_\gamma\left(\frac{b}{2\sin^2\theta}\right) d\theta}_{I_2}$$

The integrals I_1 and I_2 of Eq. (11.12) are evaluated using Mathematica to get Eq. (11.13) and Eq. (11.14), respectively where ${}_pF_q(\{a_1, a_2, \ldots, a_p\}; \{b_1, b_2, \ldots, b_q\}; z)$ is Hypergeometric function, $F_A^{(1)}(a; b_1, b_2; c; x, y)$ is Appell Hypergeometric function of two variables.

$$I_1 = \frac{a}{\pi}\int_0^{\frac{\pi}{2}} M_\gamma\left(\frac{b}{2\sin^2\theta}\right) d\theta$$

$$= \frac{2Nal_\mu}{\pi} \sum_{i_{1,2,\cdots,N}=0}^{\infty} \sum_{j_{1,2,\cdots,N-1}=0}^{\infty} \frac{2^{j_\Sigma} H^{2i_\Sigma} q_\mu \Gamma(r)}{h^{N\mu+2i_\Sigma}} \frac{1}{\prod_{p=1}^{N} \Gamma(\mu') i_p!} \frac{1}{\prod_{p=1}^{N-1} (2\mu')_{j_p} (\mu + i_p)} \int_0^{\frac{\pi}{2}} \left(\overline{\lambda}_\mu + \frac{b}{2\sin^2\theta}\right)^{-r} d\theta$$

$$\therefore I_1 = \frac{2Nal_\mu}{\sqrt{\pi}} \sum_{i_{1,2,\cdots,N}=0}^{\infty} \sum_{j_{1,2,\cdots,N-1}=0}^{\infty} \frac{2^{j_\Sigma} H^{2i_\Sigma} q_\mu \overline{\lambda}^{-r} \left(2r\left(\overline{\lambda}_\mu + b\right)\right)^{-1}}{h^{N\mu+2i_\Sigma} \prod_{p=1}^{N} \Gamma(\mu') i_p! \prod_{p=1}^{N-1} (2\mu')_{j_p} (\mu + i_p)} \left(f_1 + f_2 + f_3\right) \tag{11.13}$$

where the expressions for three functions are

$$f_1 = r\sqrt{\pi}\left(\overline{\lambda}_\mu + b\right)\Gamma(r)$$

$$f_2 = {}_2F_1\left(1, r+0.5, -0.5, \frac{b}{\overline{\lambda}_\mu + b}\right)\left(\frac{\overline{\lambda}_\mu}{\overline{\lambda}_\mu + b}\right)^r \left(\frac{b}{\overline{\lambda}_\mu + b}\right)^{-0.5} \Gamma(r+0.5)\,\overline{\lambda}_\mu$$

$$f_3 = {}_2F_1\left(1, r+0.5, 0.5, \frac{b}{\overline{\lambda}_\mu + b}\right)\left\{2b(r+1) - \overline{\lambda}_\mu\right\}$$

$$\therefore \qquad I_2 = \frac{c}{\pi}\int_0^{\frac{\pi}{4}} M_\gamma\left(\frac{b}{2\sin^2\theta}\right)d\theta$$

$$= \frac{2cNl_\mu}{\sqrt{\pi}} \sum_{i_{1,2,\cdots,N}=0}^{\infty} \sum_{j_{1,2,\cdots,N-1}=0}^{\infty} \frac{2^{j_\Sigma} H^{2i_\Sigma} q_\mu F_A^{(1)}\left(r+0.5, r, 1, r+1.5, -\frac{\overline{\lambda}_\mu}{\overline{\lambda}_\mu + b}, -1\right)}{h^{N\mu+2i_\Sigma} \prod_{p=1}^{N} \Gamma(\mu') i_p! \prod_{p=1}^{N-1}\left((2\mu')_{j_p}\left(\mu + i_p\right)\right)(1+2r)b^r} \tag{11.14}$$

Hypergeometric and Appell Hypergeometric functions:

(i) $(a)_n$ is Pochhammer symbol defined as $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} = a(a+1)\cdots(a+n-1)$. For example, $(a)_0 = 1$, $(a)_1 = a$, $(a)_2 = a(a+1) = a^2 + a$, …. Note that if a is a positive integer, $\Gamma(a) = (a-1)!$.

(ii) The hypergeometric function is defined for two complex vectors $\mathbf{a} = \{a_1, a_2, \ldots, a_p\}$ and $\mathbf{b} = \{b_1, b_2, \ldots, b_q\}$ as arguments and single variable z. Note that $\mathbf{a}$ is a vector of p elements and $\mathbf{b}$ is a vector of q elements, that's why the Hypergeometric function is denoted as ${}_pF_q$. It is defined as

$${}_pF_q(\mathbf{a}; \mathbf{b}; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k \cdots (a_{p-1})_k (a_p)_k}{(b_1)_k (b_2)_k \cdots (b_q)_k} \frac{z^k}{k!}$$

For $\mathbf{a} = \mathbf{b}$, we have,

$${}_pF_q(\mathbf{a}; \mathbf{a}; z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$$

Assume p = 2 and q = 1. Then vector **a** have 2 elements and **b** has one element only. Assume **a** = $\{a, b\}$ and **b** = $\{c\}$. Then

$${}_2F_1(a; b; a; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots$$

(iii) An extension of Hypergeometric function to two variables resulted in four new kinds of special functions. We are interested in the Appell's hypergeometric functions. It is denoted as $F_A^{(1)}(a; b_1, b_2; c; x, y)$, one in the superscript means it is the first of the four functions and subscript A means Appell. It has two variables (x, y) and three arguments $(a; b_1, b_2; c)$. It can be obtained as

$$F_A^{(1)}(a; b_1, b_2; c; x, y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(a)_{m+n} (b_1)_m (b_2)_n}{m!\,n!\,(c)_{m+n}} x^m y^n$$

All arguments are scalar. There are two variables and two summations. Note that as of now, these above two special functions are not available in MATLAB, but Mathematica has in-built functions for Hypergeometric function and Appell Hypergeometric function. Some commands are Hypergeometric1F1[a,b,z] and AppelF1[a,b1,b2,c,x,y].

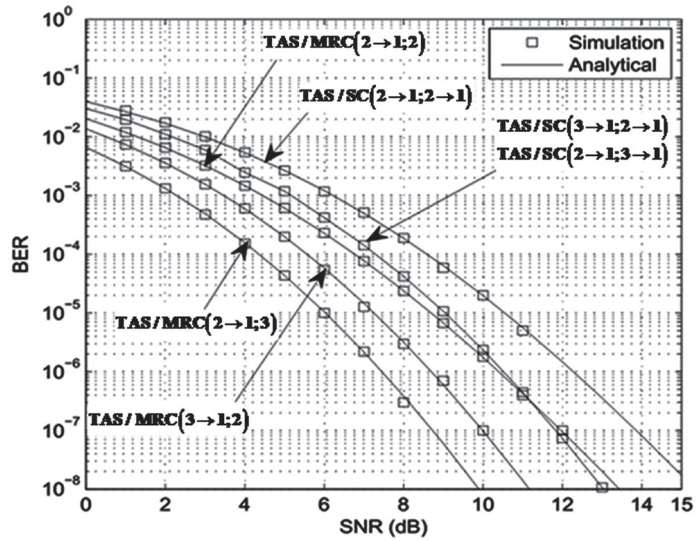


Fig. 11.2 Performance comparison of TAS/MRC and TAS/SC systems for $\eta = 1$, $\mu = 1$fading MIMO channel

In Fig. 11.2, the results are shown for a $TAS/MRC\ (N_T \to 1; N_R)$ system and $TAS/SC\ (N_T \to 1; N_R \to 1)$ system for $\eta = 1$ and $\mu = 1$ (one cluster). It compares the performance of TAS/MRC system and TAS/SC systems. The analytical and simulation results of error performance are plotted for BPSK modulation scheme. From Fig. 11.2, it can be observed that to achieve same BER of 10^{-4}, TAS/MRC $(2 \to 1;\ 1)$ require 2dB less SNR than TAS/SC $(2 \to 1;\ 2 \to 1)$ system. Similarly, to achieve the BER of 10^{-5}, TAS/MRC $(3 \to 1;\ 2)$ require 3dB less SNR when compared with the performance of TAS/SC $(2 \to 1;\ 3 \to 1)$ or TAS/SC $(3 \to 1;\ 2 \to 1)$ system. It is important to note that the BER performance of TAS/SC $(N_T \to 1;\ N_R \to 1)$ system is same as that of TAS/SC $(N_R \to 1;\ N_T \to 1)$ system. Similarly, it is intuitive that the BER performance of TAS/SC $(N_T \to 1;\ N_R \to 1)$ system is same as that of TAS/SC $(N_T N_R \to 1;\ 1)$ system or TAS/SC $(1;\ N_T N_R \to 1)$ system.

Review question 11.12 *What is Pochhammer symbol?*

Review question 11.13 *What is hypergeometric function?*

Review question 11.14 *What is Appell's hypergeometric function?*

Example 11.1

Find the probability of error for TAS/MRC over i.i.d. Rayleigh fading MIMO channel.

Solution

We may select the antenna which maximize the receive SNR

$$I_i = \underset{1 \le i \le N_T}{\arg\max} \left\{ h_{t,i} = \sum_{j=1}^{N_R} \left| h_{i,j} \right|^2 \right\}$$

For Rayleigh fading channel, I_i are i.i.d. Chi square distributed with the probability density function (PDF) and cumulative distribution function (CDF) as

$$p(x) = \frac{x^{N_r - 1} e^{-x}}{\Gamma(N_r)}, x \ge 0$$

$$P(x) = 1 - e^{-x} \sum_{i=0}^{N_r - 1} \frac{x^i}{i!}, x \ge 0$$

Using order statistics, PDF of $I_{(N_T)}$ such that $I_1 \le I_2 \le \cdots \le I_{(N_T)}$ can be given as

$$p_{(N_t)}(x) = N_t \left[P(x) \right]^{N_t - 1} p(x) = N_t \left[1 - e^{-x} \sum_{i=0}^{N_r - 1} \frac{x^i}{i!} \right]^{N_t - 1} \frac{x^{N_r - 1} e^{-x}}{\Gamma(N_r)}$$

Assume binary phase shift keying (BPSK) for TAS/MRC MIMO system, then instantaneous SNR at the MRC receiver

$$\gamma_b = \gamma \sum_{j=1}^{N_r} \left| h_{N_t, j} \right|^2 = \gamma I_{(N_t)}$$

The average bit error rate (BER) can be obtained from conditional error probability (CEP) for BPSK

$$P_e = \int_0^{\infty} Q\left(\sqrt{2\gamma_b} \right) p_{\gamma_b}(\gamma_b) \, d\gamma_b$$

Thus, closed form expression for BER (Z. Chen et al., 2005) is

$$P_e = \frac{N_t}{\Gamma(N_r)} \sum_{k=0}^{N_t - 1} \left[\frac{(-1)^k}{\left[2(k+1) \right]^{N_r}} \binom{N_t - 1}{k} \sum_{t=0}^{k(N_r - 1)} f_4 \sum_{j=0}^{N_r + t - 1} f_5 \right]$$

where the expression for functions are

$$f_4 = a_t(N_r, k)(N_r + t - 1)! \left(1 - \sqrt{\frac{\gamma}{\gamma + k + 1}} \right)^{N_r + t}$$

where $a_t(N_r, k)$ is the coefficient of z^{2t} in the expansion of $\left(\sum_{i=0}^{N_r - 1} \frac{\left[\frac{z^2}{2(k+1)} \right]^i}{i!} \right)^k$

$$f_5 = 2^{(-j)} \binom{N_r + t - 1 + j}{j} \left(1 + \sqrt{\frac{\gamma}{\gamma + k + 1}} \right)^j$$

11.2.5 Channel capacity

In this section, we derive the expression for ergodic channel capacity for TAS/MRC systems over η–μ fading channels. Ergodic capacity is the maximum rate at which information can be transmitted to have reliable reception. The average ergodic capacity per unit bandwidth can be given as (M. K. Simon et al., 2005)

$$\frac{\bar{C}_{erg}}{B} = \int_{0}^{\infty} \log_2(1+\gamma)\, p_{N_{(t)}}(\gamma)\, d\gamma \tag{11.15}$$

Using Eq. (11.9) in the above expression and evaluating it by using the method discussed in (M.-S. Alouini et al., 1999) for integer values of 2μ, we get the average ergodic capacity per unit bandwidth as

$$\frac{\bar{C}_{erg}}{B} = 2Nl_{\mu} \sum_{i_{1,2,\cdots,N}=0}^{\infty} \sum_{j_{1,2,\cdots,N-1}=0}^{\infty} \frac{2^{j_\Sigma} H^{2i_\Sigma} q_\mu \Gamma(r) e^{\bar{\lambda}_\mu} \sum_{t=1}^{r} \frac{\Gamma\left(t-r,\bar{\lambda}_\mu\right)}{\left(\bar{\lambda}_\mu\right)^{t}}}{h^{N\mu+2i_\Sigma} \prod_{p=1}^{N} \Gamma(\mu') i_p! \prod_{p=1}^{N-1} \left((2\mu')_{j_p} \left(\mu+i_p\right)\right)} \tag{11.16}$$

Simulations and numerical evaluations were done for the ergodic capacity of TAS/MRC systems over η–μ fading channels. The numerical and simulation results are plotted in Fig. 11.3 for various values of fading parameters.

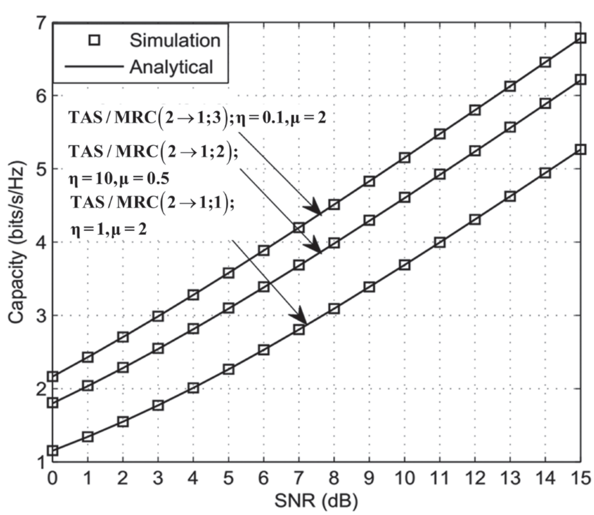


Fig. 11.3 Capacity of TAS/MRC (2 → 1; N_R) for different values of η and μ

11.3 Soft antenna selection for closely spaced antennas

The conventional antenna selection like that of previous section is called hard antenna selection (HAS). In HAS, only a set of antennas is active and selection is implemented in the RF domain by means of a set of switches. HAS does not have good performance for practical situation like closely spaced antennas which results in high antenna correlation (Z. Xu et al., 2010). This problem can be overcome by using soft antenna selection (SAS). In this, all the antennas are active and a transformation is performed in RF domain upon the received signals across all the antennas and select antennas after the transformation. Some SAS schemes are:

(a) Fast Fourier Transform (FFT) based selection
(b) Phase shift based selection

Suppose we have $N_T \times N_R$ MIMO system. Let L be the number of antennas to be selected at the receiver. At the receiver, the best L antenna elements are selected. In HAS, we choose the best subset of antenna elements for which the capacity of the system is highest among all capacity values achieved by any other possible antenna subset. The instantaneous capacity expression for a MIMO system with uniform power allocation (see chapter 6) is given by

$$C = \log_2 \det\left(\mathbf{I}_{N_T} + \frac{\rho}{N_T}\mathbf{H}^H\mathbf{H}\right). \tag{11.17}$$

With antenna selection under capacity maximization, the effective capacity now becomes

$$C = \log_2 \det\left(\mathbf{I}_{N_T} + \frac{\rho}{N_T}\breve{\mathbf{H}}^H\breve{\mathbf{H}}\right) \tag{11.18}$$

To perform receive antenna selection a matrix $\breve{\mathbf{H}}$ of $L \times N_T$ is selected from the full channel matrix $\mathbf{H}$, such that the chosen subset created by striking $N_R - L$ rows from $\mathbf{H}$ results in maximum capacity. For spatial multiplexing systems, capacity due to antenna selection at the receiver has been shown to be comparable to the full complexity system as long as the number of selected chains at the receiver is greater than or equal to the number of transmit antenna elements i.e., $L \geq N_T$. For correlated MIMO channels, HAS schemes perform considerably worse than a full complexity system. SAS schemes have been proved to be better for correlated channel. Let us assume correlated Rayleigh fading MIMO channel (Kronecker model, see chapter 3). The $(i, j)^{th}$ entry of $\mathbf{R}_{\mathbf{T_X}}$ and $\mathbf{R}_{\mathbf{R_X}}$ is given by $J_0\left(\frac{2\pi d_{ij}}{\lambda}\right)$, where, J_0 is the zeroth order Bessel function of the first kind, and d_{ij} denotes the distance between the $(i, j)^{th}$ antenna elements.

Hybrid antenna selection

n_R antennas are selected out of N_R receiving antennas like in conventional hard antenna selection. Note that hybrid antenna selection maximizes the capacity but hard antenna selection maximizes the SNR usually. The capacity for this scheme is

$$C_{Hybrid} = \max_{\mathbf{S} \in \mathbf{S}_{col}} \log_2 \det\left(\mathbf{I}_{N_T} + \frac{\rho}{N_T}(\mathbf{SH})^H \mathbf{SH}\right) \tag{11.19}$$

where, $\mathbf{S}$ is an a $n_R \times N_R$ matrix, defined as selection matrix that extracts n_R rows from $\mathbf{H}$ that are associated with the selected subset of antennas, whose cardinality is given by $|\mathbf{S}_{col}| = \binom{N_R}{n_R}$ where $\mathbf{S}_{col}$ is the collection of all possible selection matrices.

FFT-based Selection

For this SAS scheme, the antenna selection is performed on the virtual channel **FH** where **F** is $N \times N$ FFT matrix**.** All the received observation streams are sent through a Fourier transform before selection.

The (k,l)th entry of **F** is given by:

$$F(k,l) = \frac{1}{\sqrt{N}} \exp\left\{\frac{-j2\pi(k-1)(l-1)}{N}\right\}, \forall k, l \in [1, N] \tag{11.20}$$

This system capacity can be expressed as

$$C_{FFT} = \max_{\mathbf{S} \in \mathbf{S}_{col}} \log_2 \det\left(\mathbf{I}_{N_T} + \frac{\rho}{N_T} (\mathbf{SFH})^H \mathbf{SFH}\right) \tag{11.21}$$

Phase-Shift Based Selection

This is another type of soft selection scheme in which all the received observation streams are passed through a phase shift matrix, $\mathbf{\Theta}$. The matrix $\mathbf{\Theta}$ is an $n_R \times N_R$ matrix which performs phase shift implementation in the RF domain. It serves as a $N_R - \text{to} - n_R$ switch with n_R output streams. Let the largest singular value of **H** be denoted as $\lambda_{\mathbf{H}}^1$ and $\mathbf{u}_{\mathbf{H}}^1$ as the left singular vector of **H** associated with $\lambda_{\mathbf{H}}^1$, second largest singular value of **H** be denoted as $\lambda_{\mathbf{H}}^2$ and $\mathbf{u}_{\mathbf{H}}^2$ as the left singular vector of **H** associated with $\lambda_{\mathbf{H}}^2$ and so on. The phase shift matrix $\mathbf{\Theta}$ can be expressed as (Y. Yang et al., 2009)

$$\mathbf{\Theta} = \exp\left\{j \times angle\left\{\mathbf{u}_{\mathbf{H}}^1, \mathbf{u}_{\mathbf{H}}^2, \cdots, \mathbf{u}_{\mathbf{H}}^{n_R}\right\}\right\} \tag{11.22}$$

where, angle{•} gives the phase angles, in radians, of a matrix with complex elements, exp{•} denotes the element-by-element exponential of a matrix.

The capacity expression for this system can be written as

$$C_{PS} = \log_2 \det\left(\mathbf{I}_{N_T} + \frac{\rho}{N_T} (\mathbf{\Theta H})^H \left(\mathbf{\Theta\Theta}^H\right)^{-1} \mathbf{\Theta H}\right) \tag{11.23}$$

The performance analysis of the above soft antenna selection schemes are reported in Y. Yang et al. (2009). Interested readers are requested to explore the above mentioned reference.

11.4 What is spatial modulation?

In Spatial modulation, as the name suggests, we do modulation over space. Assume *k* bit information blocks are to be sent from the spatial modulation (SM) based transmitter. First it makes sub-blocks of *n* bits and *m* bits where *n* bits are spatially modulated (n bits decides which antenna will be active and transmitting the m bits) and *m* bits are modulated using digital modulation schemes like M-ary modulation. In other words, *m* bits are transmitted physically, but effectively similar to transmitting $k = m+n$ bits. There is a restriction on the number of transmit antennas and it should be a integer exponent of 2. The SM system model is shown in Fig. 11.4. It can be explained with the help of an example.

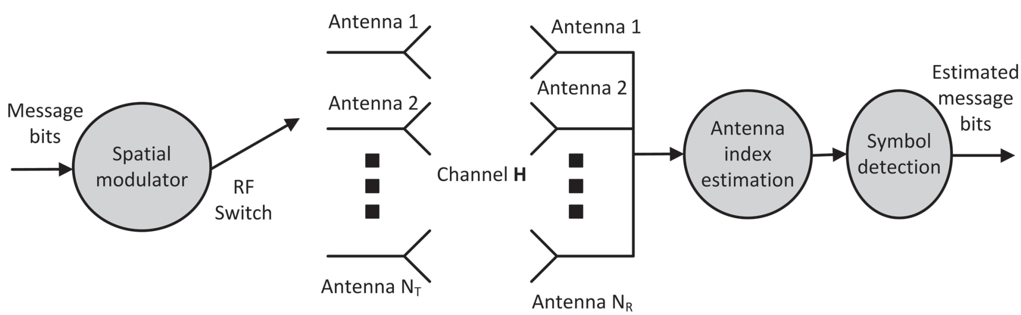


Fig. 11.4 SM MIMO system model

Example 11.2

Explain the transmission of 011 message bit using 2×2 and 4×4 SM MIMO systems.

Solution

The transmission of spatial modulation can be explained with the help of Table 11.2 and Table 11.3. Assume that one wants to transmit 3 bits at a time. There are two possible ways of doing spatial modulation for this as depicted in Table 11.2 and Table 11.3.

Table 11.2
SM mapping table for $N_T = 2$ and M = 4 (QAM)

Input bits	*Antenna index*	*Transmit symbol*
000	1	1 + j
001	1	1 – j
010	1	–1 – j
011	1	–1 + j
100	2	1 + j
101	2	1 – j
110	2	–1 – j
111	2	–1 + j

Table 11.3
SM mapping table for $N_T = 4$ and M = 2 (BPSK)

Input bits	*Antenna index*	*Transmit symbol*
000	1	–1
001	1	1
010	2	–1
011	2	1
100	3	–1
101	3	1
110	4	–1
111	4	1

First case, consider the input bit is 011, 0 bit signifies that for a 2×2 MIMO system, we will send symbol from the first antenna only, second antenna sit idle. The 11 in the message bit will be sent from the first antenna and the corresponding symbol for QAM is -1+j as illustrated in Fig. 11.5.

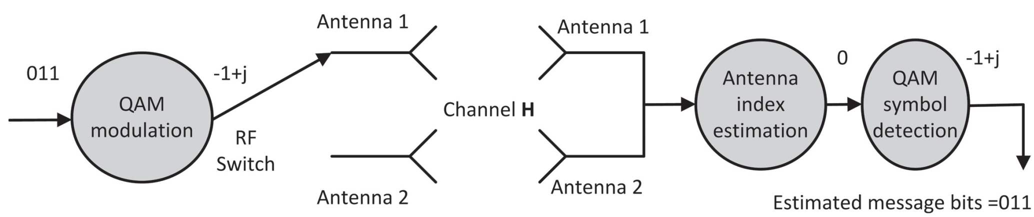


Fig. 11.5 2 × 2 SM MIMO system's transmission of input message bit 011

Second case, for a 4×4 MIMO system (see Table 12.2), for the same input bit, 01 signifies that we will send the symbol from the second antenna, 1st, 3rd and 4th antenna sit idle. The 1 message bit will be sent from antenna 2 and the corresponding symbol is 1 for BPSK as illustrated in Fig. 11.6.

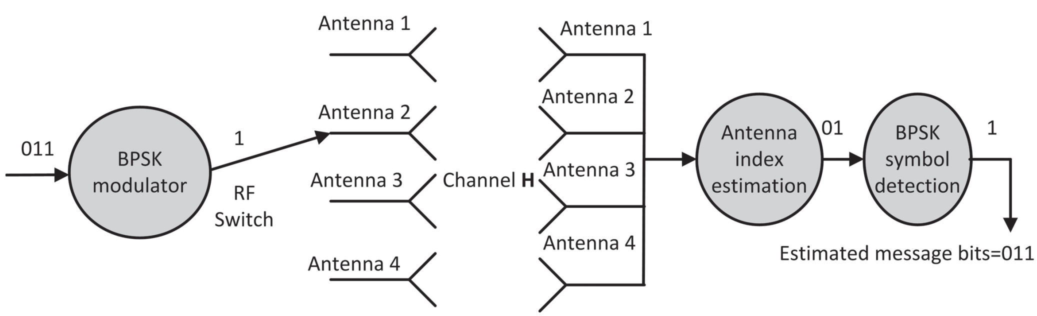


Fig. 11.6 4 × 4 SM MIMO system's transmission of input message bit 011

Example 11.3

Explain the reception of 011 message bit using 2×2 and 4×4 SM MIMO systems.

Solution

In simple sense, the first effort of the receiver is to estimate from which antenna the symbol has been sent. This will give some part of the message bit. Then it estimates the symbol. This will recover all the message bits. For example, assume that receiver of a 2×2 MIMO system (Table 11.2) estimates that the message has been sent from antenna 1, then the first part of the message bit is 0. Then it estimates that the symbol sent is -1+j, then the remaining part of the message bit is 11. Hence the decoded message bit is 011 as shown in Fig. 11.5. Whereas, consider the receiver of a 4×4 MIMO system (Table 11.3) estimates that the message has been sent from antenna 2, the first part of the message bit is 01. Then it estimates that the symbol sent is 1, then the remaining part of the message bit is 1. Therefore, the decoded message bit is 011 as depicted in Fig. 11.6.

Review question 11.15 *What is spatial modulation?*

11.5 Performance analysis of spatial modulation

Let us have a brief discussion on the SM receiver in order to find the performance analysis of SM MIMO systems. At the receiver there are two steps for detection of the transmitted message bits. Assume that the transmit antenna j is active at a particular instant of time and the corresponding channel vector is $\mathbf{h}_j$. Hence the received signal $\mathbf{y}$ can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{jq} + \mathbf{n} \tag{11.24}$$

which can be further simplified as

$$\mathbf{y} = \mathbf{h}_j x_q + \mathbf{n}$$

$$\text{where, } \mathbf{x}_{jq} = \begin{bmatrix} & 0 \\ & 0 \\ & \vdots \\ j^{th}\ position \rightarrow & x_q \\ & \vdots \\ & 0 \\ & 0 \end{bmatrix}.$$

First part of the receiver detection is the transmit antenna index (j) estimation.

$$\hat{j} = \underset{j}{\arg\max} \frac{\left|\mathbf{h}_j^H \mathbf{y}\right|}{\left\|\mathbf{h}_j\right\|^2} \tag{11.25a}$$

Second part of the detection process detects the symbol which has been transmitted from the j^{th} transmit antenna (J. Jeganathan et al., 2008) as follows.

$$\mathbf{x}_{\hat{q}} = \underset{q}{\arg\min} \left\|\mathbf{h}_{\hat{j}} x_q\right\|^2 - 2\,\mathrm{Re}\left\{\mathbf{h}_{\hat{j}}^H \mathbf{y} x_q^*\right\} \tag{11.25b}$$

Assume two detection processes are independent i.e., transmit antenna index estimate and estimation of the transmit symbol. Let us denote P_a is probability that the antenna index estimation is incorrect and P_s is probability is that the transmitted symbol estimation is incorrect. Then the probability of correct estimation can be represented as

$$P_c = \left(1 - P_a\right)\left(1 - P_s\right) \tag{11.26a}$$

Hence the probability of error is given as

$$P_e = 1 - P_c = 1 - \left(1 - P_a\right)\left(1 - P_s\right) = P_a + P_s - P_a P_s \tag{11.26b}$$

If we assume that the M-ary modulation employed is QAM, then the conditional error of probability (CEP) is represented as

$$P_s\,(\mathrm{E}/\gamma) = aQ\left(\sqrt{b\gamma}\right) - cQ^2\left(\sqrt{d\gamma}\right) \tag{11.27}$$

where, $a = 2$, $b = 1$, $c = 1$ and $d = 1$ for 4-QAM

Q-function can be approximated as a sum of two exponentials (M. Chiani et al., 2003) as follows.

$$Q(x) \cong \frac{1}{12}e^{-\frac{x^2}{2}} + \frac{1}{4}e^{-\frac{2x^2}{3}} \tag{11.28}$$

Then the above CEP can be approximated as

$$P_s\,(\mathrm{E}/\gamma) \cong \frac{a}{12}e^{-\frac{b\gamma}{2}} + \frac{a}{4}e^{-\frac{2b\gamma}{3}} - \frac{c}{144}e^{-b\lambda} - \frac{c}{16}e^{-\frac{4b\gamma}{3}} - \frac{c}{24}e^{-\frac{7b\gamma}{6}} \tag{11.29}$$

In order to find the average probability of error, we need to integrate the CEP over the pdf of the received SNR, hence,

$$P_s(E) = \int_0^\infty P_e\left(E\,/\,\gamma\right)p_\gamma\left(\gamma\right)d\gamma = \int_0^\infty \left\{aQ\left(\sqrt{b\gamma}\right) - cQ^2\left(\sqrt{d\gamma}\right)\right\}p_\gamma\left(\gamma\right)d\gamma$$

$$\Rightarrow \qquad P_s(E) \cong \int_0^\infty\left(\frac{a}{12}e^{-\frac{b\gamma}{2}} + \frac{a}{4}e^{-\frac{2b\gamma}{3}} - \frac{c}{144}e^{-b\gamma} - \frac{c}{16}e^{-\frac{4b\gamma}{3}} - \frac{c}{24}e^{-\frac{7b\gamma}{6}}\right)p_\gamma\left(\gamma\right)d\gamma$$

For 4-QAM, putting the values a, b, c and d, we have,

$$\Rightarrow \qquad P_s(E) \cong \int_0^\infty\left(\frac{1}{6}e^{-\frac{\gamma}{2}} + \frac{1}{2}e^{-\frac{2\gamma}{3}} - \frac{1}{144}e^{-\gamma} - \frac{1}{16}e^{-\frac{4\gamma}{3}} - \frac{1}{24}e^{-\frac{7\gamma}{6}}\right)p_\gamma\left(\gamma\right)d\gamma$$

The above integrals fit the definition of MGF of the received SNR and hence can be expressed in the form.

$$P_s(E) \cong \sum_i \zeta_i \int_0^\infty e^{-\vartheta_i\gamma}p_\gamma\left(\gamma\right)d\gamma = \sum_i \zeta_i MGF_\gamma\left(\vartheta_i\right)$$

Therefore,

$$P_s(E) \cong \frac{1}{6}MGF_\gamma\left(\frac{1}{2}\right) + \frac{1}{2}MGF_\gamma\left(\frac{2}{3}\right) - \frac{1}{144}MGF_\gamma\left(1\right) - \frac{1}{16}MGF_\gamma\left(\frac{4}{3}\right) - \frac{1}{24}MGF_\gamma\left(\frac{7}{6}\right) \tag{11.30}$$

The probability of error in transmit antenna index estimation can be obtained as

$$P_a = Q\left(\sqrt{\gamma_{eff}}\right) \cong \frac{1}{12}MGF_{\gamma_{eff}}\left(\frac{1}{4}\right) + \frac{1}{4}MGF_{\gamma_{eff}}\left(\frac{1}{3}\right) \tag{11.31}$$

where, $\gamma_{eff} = \frac{\gamma}{2}\left\|\mathbf{h}_j - \mathbf{h}_{\hat{j}}\right\|$.

The PDF of η-μ fading distribution is given in (N. Ermolova, 2008).

$$p_{\eta-\mu}(x) = \frac{2\sqrt{\pi}h^\mu}{\Gamma\left(\mu\right)}\left(\frac{\mu}{\Omega_h}\right)^\mu\left(\frac{x}{H}\right)^{\mu-\frac{1}{2}}e^{-x'}I_{\mu-\frac{1}{2}}\left(x'\right);\, x' = \frac{2\mu hx}{\Omega_h} \tag{11.32}$$

The MGF of η-μ fading distribution (refer to chapter 2) can be obtained as

$$MGF_{\eta-\mu}(s) = \left(\frac{4\mu^2 h}{\left(2\mu(h-H)+s\right)\left(2\mu(h+H)+s\right)} \right)^{\mu} \tag{11.33}$$

Hence, we can calculate the P_s and P_a using Eqs. (11.30) and (11.31) and find the error probability of SM MIMO system over η-μ fading channel using the MGF of Eq. (11.33). Figure 11.7 depicts the SER vs SNR of 2×2 SM MIMO system over Nakagami-q fading channel which is a particular case of η-μ fading channel (refer to chapter 2). It has been observed that for fixed μ = 0.5, as we increase ν from 0.1, 0.2, 0.3, 0.5 and 0.9, the SER improves consistently. The improvement in the SER is more visible for higher SNR.

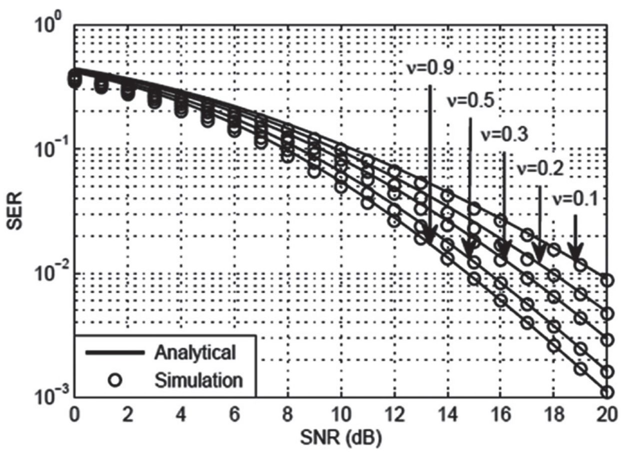


Fig. 11.7 SER vs. SNR (dB) for 2 × 2 SM MIMO system considering Nakagami-q fading as a special case of η-μ fading channel for $\eta = \nu^2$ and $\mu = 0.5$

Review question 11.16 *What is the relation between Nakagami-q or Hoyt fading distribution with η-μ fading distribution?*

Review question 11.17 *Write the Chiani's approximation of the Q-function.*

Review question 11.18 *Write down the expression of probability of error of SM in terms of probability of error in antenna index estimation and symbol detection assuming that the two processes are independent.*

11.6 Performance analysis of SM with antenna selection

The main problem with SM is that some of the antenna from which we are transmitting symbols may be experiencing the worst performance or dead link. Then we will have a very bad performance. In order to overcome this issue, we may do antenna selection at the beginning and apply SM on those antennas having the best links (B. Kumbhani et al., 2014). This gives a way to avoid applying SM on the bad links and improves the performance of SM. Hence, SM systems combined with antenna selection can be divided into two phases: (a) transmit antenna selection (b) SM applied over the selected antennas. In order to do this, we need CSI available at transmitter. A_j is the received SNR due to transmission from antenna j at the transmitter.

$$A_j = \sum_{i=1}^{N_r} \left|h_{i,j}\right|^2 \tag{11.34}$$

Assuming i.i.d. MIMO Rayleigh fading channel, then A_j has Chi-square distribution with the PDF $f_{A_j}(x) = \frac{x^{N_R-1}e^{-x}}{\Gamma(N_R)}, x \geq 0$ where Γ(.) is the Gamma function and CDF $F_{A_j}(x) = 1 - e^{-x}\sum_{i=0}^{N_R-1}\frac{x^i}{i!}$, $x \geq 0$. From order statistics, we may select the best subset (S out of N_T) of antennas at the transmitter. It may be summarized as: (a) received SNR A_js when only one transmit antenna (j^{th} antenna) is active at the transmitter are arranged in ascending order (b) S antennas out of N_T corresponding to highest A_js (received SNR) are selected. PDF of $A_{(r)}$ such that $A_{(1)} \leq A_{(2)} \leq \cdots \leq A_{(r)} \leq \cdots \leq A_{(N_T-1)} \leq A_{(N_T)}$ can be given as (H. A. David, 2003).

$$f_{A_{(r)}}(x) = \frac{1}{B(r, N_T - r + 1)}\left\{F_{A_j}(x)\right\}^{r-1}\left\{1 - F_{A_j}(x)\right\}^{N_T - r} f_{A_j}(x) \tag{11.35}$$

where, $B(\bullet, \bullet)$ is the Beta function and $r = N_T - S + 1$.

The PDF of received SNR $A_{(r)}$ can be given as

$$p_{A_{(r)}}(x) = \frac{1}{(N_T - r + 1)\Gamma(N_R)}\sum_{i=r}^{N_T}\sum_{j=0}^{i-1}\sum_{k=0}^{K}\frac{1}{B(i, N_T - i + 1)}\binom{i-1}{j}(-1)^j C_k(j, N_R)\, x^{N_R+k-1}e^{-x(N_T-i+j+1)} \tag{11.36}$$

where, $C_k(j, N_R)$ is the coefficient of x^k in the expansion of $\left(\sum_{l=0}^{N_R-1}\frac{x^l}{l!}\right)^{N_T-i+j}$ and $K = (N_R - 1)(N_T - i + j)$.

Outage probability is the probability of outage. The transmission over a channel is in outage whenever the data rate for the transmission exceeds the capacity of the channel.

$$P_{out}(A_{(R)}, R) = P_r\left(A_{(r)} < \frac{2^R - 1}{\bar{A}_{(r)}} = A_{th}\right)$$

Hence, we can find the outage probability for a given data rate R from the CDF of $A_{(r)}$ and it is given as

$$P_{out}(A_{(R)}, R) = \frac{1}{(N_T - r + 1)\Gamma(N_R)}\sum_{i=r}^{N_T}\frac{1}{B(i, N_T - i + 1)}\sum_{j=0}^{i-1}\binom{i-1}{j}(-1)^j\sum_{k=0}^{K}C_k(j, N_R)\frac{\gamma_{inc}\left(N_R + k, A_{th}(N_T - i + j + 1)\right)}{(N_T - i + j + 1)^{N_R+k}} \tag{11.37}$$

where, γ_{inc} (. , .) is the incomplete Gamma function, $C_k(j, N_R)$ is the coefficient of x^k in the expansion of $\left(\sum_{l=0}^{N_R-1} \frac{x^l}{l!}\right)^{N_T-i+j}$ and $K = (N_R - 1)(N_T - i + j)$.

In the Fig. 11.8, 2×2 SM MIMO is the traditional SM system without any antenna selection. For the 2×2 SM MIMO system, we want to send two bits at a time. We use 1 bit for antenna selection, this bit will decide whether we have to send the symbol from antenna 1 or 2 and the second bit will decide which BPSK symbol will be sent from the selected antenna. 4/2×2 SM MIMO means it is a 4×2 MIMO system in which we select the 2 transmit antennas with the best links out of the 4 transmit antennas and apply SM MIMO system on the corresponding 2×2 SM MIMO system. 6/2×2 SM MIMO means it is a 6×2 MIMO system in which we select the 2 transmit antennas with the best links out of the 6 transmit antennas and apply SM MIMO system on the corresponding 2×2 SM MIMO system. It can be observed that 6/2×2 SM MIMO outperforms 4/2×2 SM MIMO whereas 4/2×2 SM MIMO has better performance than the traditional 2×2 SM MIMO system.

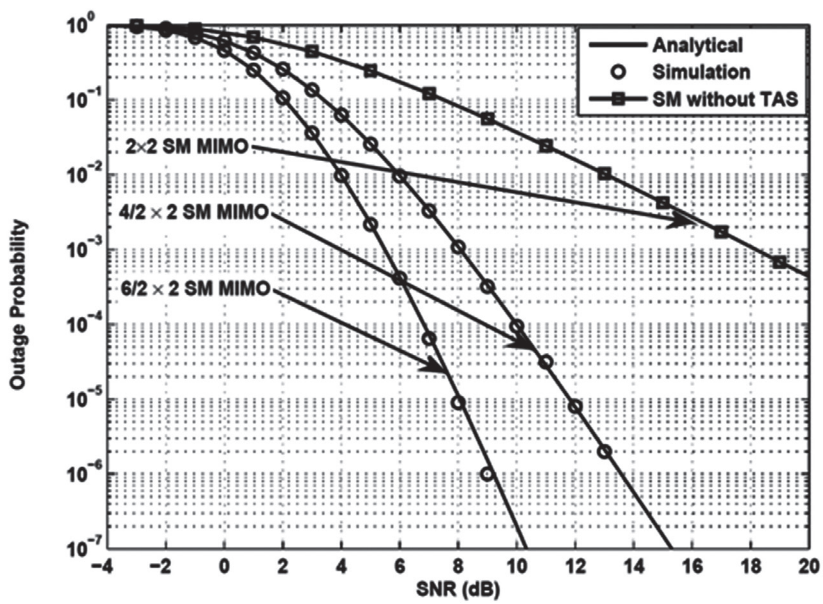


Fig. 11.8 Outage probability vs. SNR curve for TAS SM MIMO systems with antenna selection (R = 2 bits/s/Hz)

Review question 11.19 *What is beta function?*

Review question 11.20 *What is the need for SM with antenna selection?*

Review question 11.21 *What is the advantage of SM with antenna selection over SM?*

Review question 11.22 *What is order statistics?*

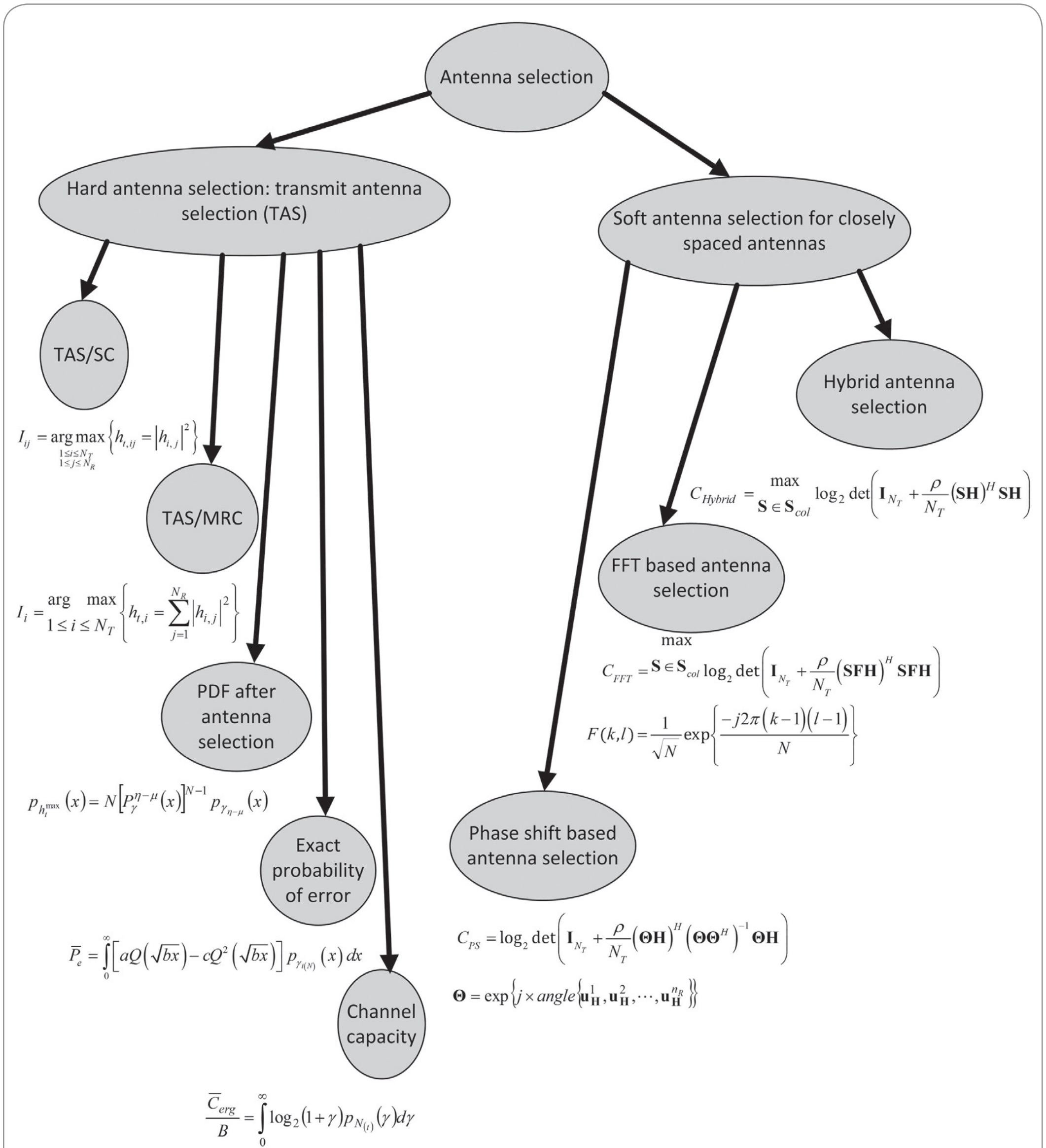
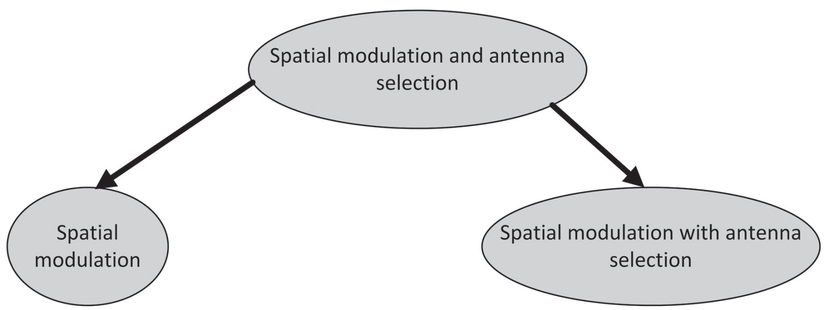


Fig. 11.9 Antenna selection in a nutshell

11.7 Summary

In this chapter, we have discussed about the antenna selection and spatial modulation. In antenna selection, we have learnt about the hard and soft antenna selection. In conjunction with transmit antenna selection (TAS), we can employ selection combining (SC) at the receiver (TAS/SC) or Maximal ratio combining (MRC) at the receiver (TAS/MRC). Using order statistics, we have derived the PDF after antenna selection and calculated the exact probability of error for antenna selection and its capacity. Soft antenna selection is suitable for closely spaced antennas. In this, we have studied FFT based, phase shift based and hybrid antenna selection which are used for maximizing the channel capacity. Spatial modulation is a relatively new MIMO technique in which one does modulation over space. The antenna index of the transmitting antenna also conveys some amount of information. In practical scenarios, some of the links for transmitting antennas may be completely down. If we send the signal from such antennas, then there may be complete outage. In order to avoid such an unwanted situation, we may do antenna selection before applying spatial modulation. Among all the transmitting antennas, select a subset of antennas with the best links and apply spatial modulation on those selected subset of antennas. Then what we have is a spatial modulation with antenna selection. Nutshell of what we have learnt in spatial modulation and antenna selection is shown in Fig. 11.10 and Fig. 11.9, respectively.



Spatial modulation:

$$\mathbf{y} = \sqrt{\rho}\mathbf{h}_j x_q + \mathbf{n}$$

$$\hat{j} = \arg\max_j \frac{|\mathbf{h}_j^H \mathbf{y}|}{\|\mathbf{h}_j\|^2} \qquad \mathbf{x}_{\hat{q}} = \arg\min_q \|\mathbf{h}_{\hat{j}} x_q\|^2 - 2\mathrm{Re}\{\mathbf{h}_{\hat{j}}^H \mathbf{y} x_q^*\}$$

$$P_e = 1 - P_c = 1 - (1 - P_a)(1 - P_s) = P_a + P_s - P_a P_s$$

$$P_a = Q\left(\sqrt{\gamma_{eff}}\right) \cong \frac{1}{12} MGF_{\gamma_{eff}}\left(\frac{1}{4}\right) + \frac{1}{4} MGF_{\gamma_{eff}}\left(\frac{1}{3}\right) \qquad \gamma_{eff} = \frac{\gamma}{2}\|\mathbf{h}_j - \mathbf{h}_{\hat{j}}\|$$

$$P_s(E) \cong \frac{1}{6} MGF_\gamma\left(\frac{1}{2}\right) + \frac{1}{2} MGF_\gamma\left(\frac{2}{3}\right) - \frac{1}{144} MGF_\gamma(1) - \frac{1}{16} MGF_\gamma\left(\frac{4}{3}\right) - \frac{1}{24} MGF_\gamma\left(\frac{7}{6}\right)$$

Spatial modulation with antenna selection:

$$A_{(1)} \le A_{(2)} \le \cdots \le A_{(r)} \le \cdots \le A_{(N_T-1)} \le A_{(N_T)}$$

$$f_{A_{(r)}}(x) = \frac{1}{B(r, N_T - r + 1)} \left\{F_{A_j}(x)\right\}^{r-1} \left\{1 - F_{A_j}(x)\right\}^{N_T - r} f_{A_j}(x)$$

$$P_{out}(A_{(r)}, R) = P_r\left(A_{(r)} < \frac{2^R - 1}{\bar{A}_{(r)}} = A_{th}\right)$$

Fig. 11.10 Spatial modulation in a nutshell

Exercise

Exercise 11.1

Find the expression of approximate SER for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. η–μ fading channels.

Exercise 11.2

Find the expression of asymptotic SER for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. η–μ fading channels.

Exercise 11.3

Find the expression of exact SER for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. k–μ fading channels.

Exercise 11.4

Find the expression of approximate SER for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. k–μ fading channels.

Exercise 11.5

Find the expression of asymptotic SER for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. k–μ fading channels.

Exercise 11.6

Find the expression of ergodic capacity for transmit antenna selection with maximal ratio combining (TAS/MRC) MIMO systems for several modulation schemes over i.i.d. k–μ fading channels.

Exercise 11.7

Repeat the exercises 1–6 for transmit antenna selection with selection combining (TAS/SC) MIMO systems.

Exercise 11.8

Draw a block diagram of SM MIMO system. Explain each block in few words.

Exercise 11.9

For i.i.d. MIMO fading channel, find the error in antenna index estimation of SM MIMO system.

Exercise 11.10

For i.i.d. MIMO fading channel, find the error in symbol detection of SM MIMO system in terms of MGF of the received SNR of the fading channel.

Exercise 11.11

If we assume that the M-ary modulation employed is QAM, then the conditional error of probability (CEP) is represented as

$$P_s(E/\gamma) = aQ\left(\sqrt{b\gamma}\right) - cQ^2\left(\sqrt{d\gamma}\right)$$

Find the simplified expression the above CEP using Chiani's approximation.

Exercise 11.12

From order statistics for i.i.d. Rayleigh fading MIMO channel, find the pdf and cdf of any $A_{(r)}$ such that

$$A_{(1)} \le A_{(2)} \le \cdots \le A_{(r)} \le \cdots \le A_{(N_t-1)} \le A_{(N_t)}$$

A_j is the received SNR due to transmission from antenna j (j = 1, 2, … N_T) at the transmitter.

$$A_j = \sum_{i=1}^{N_r} \left|h_{i,j}\right|^2$$

Exercise 11.13

Find the outage probability of SM with antenna selection over i.i.d. Rayleigh fading MIMO channel.

Exercise 11.14

Find the outage probability of SM with antenna selection over i.i.d. η-μ/k-μ/α-μ fading MIMO channel.

References

1. Alouini, M.-S. and A. Goldsmith. Jul 1999. 'Capacity of Rayleigh fading channels under different adaptive transmission and diversity combining techniques'. *IEEE Trans. Veh. Technol*. 48(4). 1165–1181.
2. Chen, Y. and C. Tellambura. 2004. 'Distribution functions of selection combiner output in equally correlated Rayleigh, Rician, and Nakagami-m fading channels'. *IEEE Transactions on Communications*. 52(11). 1948–1956.
3. Chen, Z., J. Yuan, and B. Vucetic. July 2005. 'Analysis of transmit antenna selection/maximal-ratio combining in Rayleigh fading channels'. *IEEE Trans. on Vehicular Technology*. 54(4). 1312–1321.
4. Chiani, M., D. Dardari, and M. K. Simon. 2003. 'New exponential bounds and approximations for the computation of error probability in fading channels'. *IEEE Trans. Wireless Commun*. 2(4). 840–845.
5. David, H. A. and H. N. Nagaraja. 2003. *Order Statistics*. 3rd ed. New York: Wiley Interscience.
6. Ermolova, N. 2008. 'Moment generating functions of the generalized η-μ and k-μ distributions and their applications to performance evaluations of communication systems'. *IEEE Communications Letters*. 12(7). 502–504.
7. Goldsmith, A. 2005. *Wireless Communications*. Cambridge: Cambridge University Press.
8. Gradshteyn, I. S. and I. M. Ryzhik. 2007. *Table of Integrals, Series, and Products*. 7th ed. New York: Academic Press.
9. Jeganathan, J., A. Ghrayeb, and L. Szczecinski. Aug. 2008. 'Spatial modulation: optimal detection and performance analysis'. *IEEE Communications Letters*. 12. 545–547.
10. Kumbhani, B. and R. S. Kshetrimayum. 2015. 'Analysis of TAS/MRC based MIMO systems over η-μ fading channels'. *IETE Technical Review*. 32(4). 252–259.
11. Kumbhani, B. and R. S. Kshetrimayum. 2015. 'MGF based approximate SER calculation of SM MIMO systems over generalized η-μ and k-μ fading channels'. *Wireless Personal Communications*. 83(3). 1903–1913.

12. Kumbhani, B. and R. S. Kshetrimayum. Jan 2014. 'Outage probability analysis of spatial modulation systems with antenna selection'. *Electronics Letters*. 50(2). 125–126.
13. Mesleh, R., S. Engelken, S. Sinanovic, and H. Haas. Aug. 2008. 'Analytical SER calculation of spatial modulation'. In *Spread Spectrum Techniques and Applications, 2008. ISSSTA '08. IEEE 10th International Symposium on*. 272 –276.
14. Simon, M. K. and M. S. Alouini. 2005. *Digital Communication over Fading Channels*. 2nd ed. New York: Wiley.
15. Xu, Z., S. Sfar, and R. S. Blum. Feb. 2010. 'Receive antenna selection for closely-spaced antennas with mutual coupling'. *IEEE Trans. Wireless Communications*. 9(2). 652–661.
16. Yacoub, M. 2007. 'The κ–μ distribution and the η–μ distribution'. *IEEE Antennas and Propagation Magazine*. 49(1). 68–81.
17. Yang, Y., R. S. Blum, and S. Sfar. 2009. 'Antenna selection for MIMO systems with closely spaced antennas'. *EURASIP Journal on Wireless Communications and Networking*. 11.

Advanced Topics in MIMO Wireless Communications

CHAPTER 12

12.1 Introduction

In this chapter on advance MIMO communications, we will study about Space-Time Block Coded Spatial Modulation (STBC–SM), MIMO based cooperative communications and Large-scale (LS) MIMO systems. In STBC–SM, we will combine STBC and SM techniques to improve the rate and diversity of MIMO systems. In particular, we will find an upper bound on the bit error probability (BEP) of STBC–SM analysis over correlated Rayleigh and Rician fading MIMO channels. In MIMO based cooperative communications, we will apply cooperative communication employing MIMO based source, relay and destination nodes. We will find an approximate close-form formula for BER of MIMO based cooperative communication over i.i.d. $\alpha - \mu$ fading MIMO channel. Next we will discuss about the much-awaited topic on large-scale (LS) MIMO systems which is one of the proponents for fifth generation (5G) mobile wireless communications. In LS MIMO, we will consider three scenarios viz., single user (SU) LS MIMO, multiuser (MU) LS MIMO and multi-cell LS MIMO. The section concludes with a discussion on Coordinated Multipoint transmission and Heterogeneous networks. The last section will discuss about the MIMO Cognitive radios.

12.2 Space-time block coded spatial modulation

It is a well-known fact that MIMO systems give higher spectral efficiencies than single-input single-output (SISO) systems without increasing signal bandwidth and power. Spatial modulation is a recent MIMO technique which provides higher spectral efficiencies through antenna indexing. In this section, we will provide an upper bound on the bit error probability (BEP) of space-time block coded spatial modulation (STBC–SM) over correlated Rayleigh and Rician fading channels which is well validated by Monte Carlo simulations. S. Alamouti (1998) proposed a simple two branch transmit diversity scheme which has been extended for any number of transmit antennas by V. Tarokh et al., (1999) also commonly referred to as space-time block codes (STBC). R. Mesleh et al., (2006) presented a new transmission scheme called as spatial modulation (SM) in which symbols are transmitted through a particular antenna and that antenna index itself carries useful information. E.

Basar et al., (2011) combined SM and STBC taking advantages of both techniques while overcoming their disadvantages. They have reported that STBC–SM have performance advantages over SM, Vertical Bell Laboratories Layered Space Time (V-BLAST), etc. However their analysis was limited to independent and identically distributed (i.i.d.) Rayleigh channel only, albeit, they have provided simulation results for exponentially correlated Rayleigh channels. In this section, we will obtain an upper bound for STBC–SM BEP for correlated Rayleigh and Rician fading channels which is verified by Monte Carlo simulations.

In Space-time Block Coded Spatial Modulation *(*STBC–SM) technique (E. Basar et al., 2011), the input data is divided into three streams. Two streams carry the STBC symbols and the other streams carry the transmit antenna indices. We will use Alamouti's STBC, where two complex symbols taken from an M-PSK or M-QAM constellations are transmitted from two transmit antennas in two symbol intervals in an orthogonal manner (S. M. Alamouti, 1998). The input–output relation for the STBC–SM $N_T \times N_R$ MIMO system for frequency flat fading case is given below.

$$\mathbf{Y} = \sqrt{\frac{\rho}{\mu}}\mathbf{HX} + \mathbf{N} \tag{12.1}$$

where, μ is a normalization factor to ensure that ρ is the average signal-to-noise ratio (SNR) at each receive antenna. Here **Y** is the $N_R \times 2$ received signal matrix, **N** is the $N_R \times 2$ zero mean circularly symmetric complex Gaussian (ZMCSCG) noise, **X** is the $N_T \times 2$ transmit codeword matrix and **H** is the $N_R \times N_T$ channel matrix which is assumed to be quasi-static correlated Rayleigh or Rician fading. The maximum likelihood (ML) decoding makes an exhaustive search over all transmission matrices and decodes in favour of the matrix that minimizes the metric given below

$$\hat{\mathbf{X}} = \underset{\mathbf{X} \in \chi}{\arg\min} \left\| \mathbf{Y} - \sqrt{\frac{\rho}{\mu}}\mathbf{HX} \right\|^2 \tag{12.2}$$

where, χ is the signal matrix alphabets.

The conditional pairwise error probability (PEP) of decoding STBC–SM codeword matrix $\mathbf{X}_l$ when STBC–SM codeword matrix $\mathbf{X}_k$ was transmitted is given by

$$P\left(\mathbf{X}_k \rightarrow \mathbf{X}_l \mid \mathbf{H}\right) = Q\left(\sqrt{\frac{\rho}{\mu}\left\|\mathbf{H\Delta}\right\|^2}\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} e^{-\frac{\rho\left\|\mathbf{H\Delta}\right\|^2}{2\mu\sin^2\theta}} d\theta \tag{12.3}$$

where, $\mathbf{\Delta} = \mathbf{X}_k - \mathbf{X}_l$ is the codeword difference matrix.

The unconditional PEP for μ = 1 and $E\left\{trace\left(\mathbf{X}^H\mathbf{X}\right)\right\} = 2$ is given by

$$P\left(\mathbf{X}_k \rightarrow \mathbf{X}_l\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \Phi_{\left\|\mathbf{H\Delta}\right\|^2}\left(-\frac{\rho}{4\sin^2\theta}\right) d\theta \tag{12.4}$$

where, $\Phi_{\left\|\mathbf{H\Delta}\right\|^2}$ denotes the moment generating function (MGF) of $\left\|\mathbf{H\Delta}\right\|^2$.

We will consider Kronecker MIMO channel model which is expressed as

$$\mathbf{H} = \mathbf{R}_{R_X}^{1/2}\tilde{\mathbf{H}}\mathbf{R}_{T_X}^{T/2} \tag{12.5}$$

where, $\tilde{\mathbf{H}}$ is the i.i.d. channel matrix, $\mathbf{R}_{R_X}$ and $\mathbf{R}_{T_X}$ are the correlation matrices at the transmitter and receiver, respectively.

The MGF for correlated Rayleigh fading channel (A. Hedayat et al., 2005) is

$$\Phi(s) = \left|\mathbf{I}_{n_R n_T} - s\boldsymbol{\psi}\right| = \prod_{i=1}^{r}\prod_{j=1}^{\hat{r}}\left(1 - s\sigma_i\lambda_j\right)^{-1} \tag{12.6}$$

where, $\boldsymbol{\psi} = \mathbf{R}^{H/2}\left(\mathbf{I}_{n_R} \otimes \Delta\Delta^H\right)\mathbf{R}^{1/2}$, $s = -\dfrac{\rho}{4\sin^2\theta}$, $\mathbf{R} = \mathbf{R}_{R_X} \otimes \mathbf{R}_{T_X}$, σ_i are the eigenvalues of $\Delta\Delta^H\mathbf{R}_{T_X}$, λ_j are the eigenvalues of $\mathbf{R}_{R_X}$, $r = rank\left(\Delta\Delta^H\mathbf{R}_{T_X}\right)$ and $\hat{r} = rank\left(\mathbf{R}_{R_X}\right)$.

The union bound on BEP can be calculated as

$$P_b \le \frac{1}{2^{2m}}\sum_{k=1}^{2^{2m}}\sum_{l=1}^{2^{2m}} n_{k,l}\frac{P\left(\mathbf{X}_k \to \mathbf{X}_l\right)}{2m} \tag{12.7}$$

where, $n_{k,l}$ is the number of bits in error between the codeword matrices.

$\mathbf{X}_k$ and $\mathbf{X}_l$ assuming 2m bits are transmitted during two consecutive symbol intervals using one of the 2^{2m} different STBC–SM codeword matrices.

Hence applying Chernoff bound on PEP (put sin $\theta = 1$ in the integrand of Eq. 12.4), we have the union bound as

$$P_b \le \frac{1}{2^{2m}}\sum_{k=1}^{2^{2m}}\sum_{l=1}^{2^{2m}} \frac{n_{k,l}}{4m}\prod_{i=1}^{r}\prod_{j=1}^{\hat{r}}\left(1 + \frac{\rho\sigma_i\lambda_j}{4}\right)^{-1} \tag{12.8}$$

The MGF for Rician fading channel (A. Hedayat et al., 2005) is given as:

$$\Phi(s) = \frac{\exp\left[s\bar{\mathbf{h}}^H\boldsymbol{\psi}\left\{\mathbf{I}_{N_R N_T} - \dfrac{s\boldsymbol{\psi}}{(K+1)}\right\}^{-1}\bar{\mathbf{h}}\right]}{\left|\mathbf{I}_{N_R N_T} - \dfrac{s\boldsymbol{\psi}}{(K+1)}\right|} \tag{12.9}$$

where, $\bar{\mathbf{h}} = vect\left(\bar{\mathbf{H}}^H\right)$ and $\bar{\mathbf{H}} = \sqrt{\dfrac{K}{K+1}}\breve{\mathbf{H}}\left(N_R, N_T\right)$ is the mean channel matrix of the Rician channel with Rice parameter K.

Therefore, PEP is given as

$$P\left(\mathbf{X}_k \to \mathbf{X}_l\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{e^{\left[-\frac{\rho}{4\sin^2\theta}\bar{\mathbf{h}}^H\boldsymbol{\psi}\left\{\mathbf{I}_{N_R N_T} + \frac{\rho\boldsymbol{\psi}}{4\sin^2\theta(K+1)}\right\}^{-1}\bar{\mathbf{h}}\right]}}{\left|\mathbf{I}_{N_R N_T} + \dfrac{\rho\boldsymbol{\psi}}{4\sin^2\theta(K+1)}\right|}d\theta \tag{12.10}$$

Applying Chernoff bound on the PEP, we can obtain the union bound as

$$P_b \le \frac{1}{2^{2m}}\sum_{k=1}^{2^{2m}}\sum_{l=1}^{2^{2m}} \frac{n_{k,l}}{4m}e^{\left[-\frac{\rho}{4}\bar{\mathbf{h}}^H\boldsymbol{\psi}\left\{\mathbf{I}_{N_R N_T} + \frac{\rho\boldsymbol{\psi}}{4(K+1)}\right\}^{-1}\bar{\mathbf{h}}\right]}\prod_{i=1}^{r}\prod_{j=1}^{\hat{r}}\left(1 + \frac{\rho\sigma_i\lambda_j}{4(K+1)}\right)^{-1} \tag{12.11}$$

The total number of transmit and receive antennas are assumed to be four each. Figure 12.1 shows the bit error rate (BER) performance of STBC–SM for equi-correlated and uncorrelated Rayleigh fading at 3 bits/s/Hz. It depicts that the analytical results give an upper error bound on Monte Carlo simulations for different values of correlation coefficients at the transmitter (pt) and receiver (pr). The BER increase with higher correlation coefficients as expected. Note that for Rician fading channel we have taken $\breve{\mathbf{H}}(N_R, N_T)$ as ones (N_R, N_T). Figure 12.2 shows that the Monte Carlo simulation BER is bounded by the analytical results for Rice parameter K = 2 at 2 bits/s/Hz as well. It may be also observed that the analytical bound is tighter for high SNR regions.

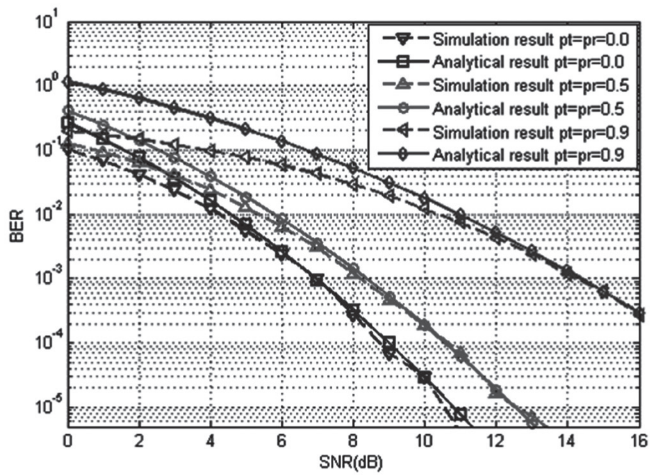


Fig. 12.1 Monte Carlo simulation result vs. BER analytical bound for equi-correlated Rayleigh fading at 3 bits/s/Hz

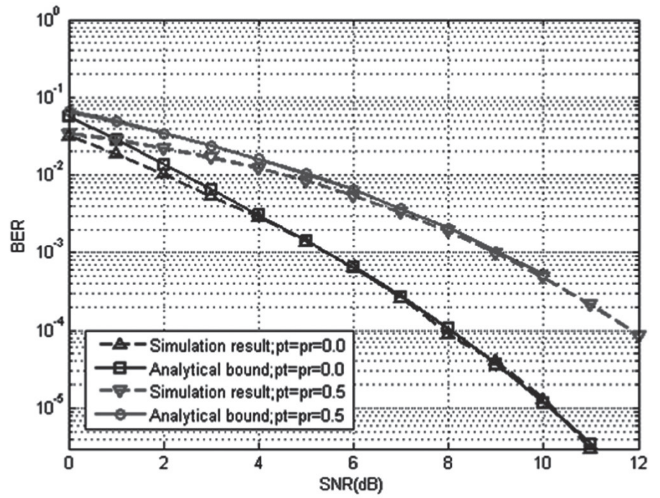


Fig. 12.2 Monte Carlo simulation result vs. BER analytical bound for equi-correlated Rician fading (K=2) at 2 bits/s/Hz

The error bound on BEP has been found out for STBC–SM over correlated Rayleigh and Rician channel. Monte Carlo simulation has been carried out to verify the error bounds (B. Kumbhani et al., 2015). This work could be continued further to analyze STBC–SM in other correlated channels viz., Nakagami, Hoyt, α-μ, k-μ and η-μ fading channels.

Review question 12.1 *Write down the union bound on BEP for STBC–SM over correlated Rician and Rayleigh fading MIMO channel.*

12.3 MIMO based cooperative communication

12.3.1 SISO based cooperative communication

Cooperative communication is a particular case of relay based communication. In this communication, we assume that there are three nodes viz. source (S), destination (D) and relay (R). The R not only receives signal from the S but also forwards it to D (if it decides to forward the message). It also sends its own information. Similarly S may also become R for other user. Generally two protocols are used at the R node (G. Menghwar et al., 2009) and they are:

(a) **Amplify and Forward (AF) Protocol:** In AF protocol, every R node, amplifies and re-transmits to the D the signal it received from the S.
(b) **Decode and Forward (DF) Protocol:** In DF protocol, every R decodes transmitted symbol from the S and if the decoding is successful, the R sends the re-encoded symbol to D otherwise R sits idle.

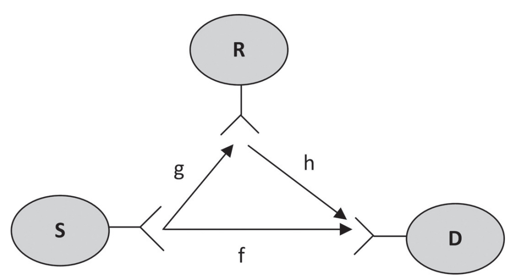


Fig. 12.3 Single relay based cooperative communication system

In cooperative communication, signal reaches from S to D in two phases. In first phase signal is transmitted from the S to R. In the second phase, the signal is transmitted from R to D, if R decides to re-transmit the received signal. Let us consider a single relay based cooperative communication system with one S communicating with one D and one R as depicted in Fig. 12.3 (K. J. R. Liu, 2009). In the first phase, the S broadcasts message data to R and D. The signals received at D and R ($y_{s,d}$ and $y_{s,r}$) are expressed as

$$y_{s,d} = \sqrt{p_1} fx + n_1$$
$$y_{s,r} = \sqrt{p_1} gx + n_2 \tag{12.12}$$

p_1 is the transmitted signal power from S to R and S to D. x symbol is transmitted from S. n_1 and n_2 are ZMCSCG RVs with variance N_0. Channel gain coefficients f and g are assumed to be k-μ or η-μ distributed fading coefficients between S and D and/or between S and R, respectively.

In second phase, for DF protocol, if the R is able to decode the transmitted symbol correctly, then it forwards the decoded symbol with transmission power p_2 to the D. Otherwise R sits idle. The signal received at the D($y_{r,d}$) can be expressed as

$$y_{r,d} = \sqrt{\tilde{p}_2} hx + n_3 \tag{12.13}$$

The transmission power $\tilde{p}_2 = p_2$ if the R decodes the transmitted symbol correctly, otherwise $\tilde{p}_2 = 0$. n_3 is ZMCSCG with variance N_0. The channel gain coefficient h is assumed to be k-μ or η-μ distributed fading coefficients between R and D. The fading gain coefficients are f (S to D), g (S to R) and h (R to D). Assume CSIR is available, no CSIT. The D jointly combines the phase 1 and 2 received signals from the S. The detector used at the D is MRC (D. G. Brennan, 2003). The total transmitted power has to satisfy, $p_1 + p_2 = p$. The combined signal at the D can be expressed as

$$y = \frac{\sqrt{p_1} f^*}{N_0} y_{s,d} + \frac{\sqrt{\tilde{p}_2} h^*}{N_0} y_{r,d} \tag{12.14}$$

where, f^* and h^* are the complex conjugates of f and h, respectively.

The received SNR at D is expressed as

$$\gamma = \frac{p_1 |f|^2 + \tilde{p}_2 |h|^2}{N_0} \tag{12.15}$$

Note that Q function is defined and approximated (Chiani et al., 2003) as

$$Q(\mathrm{x}) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy \cong \frac{1}{12}\exp\left(-\frac{x^2}{2}\right) + \frac{1}{4}\exp\left(-\frac{2x^2}{3}\right)$$

Hence, for BPSK

$$\psi_{BPSK}(\gamma) = Q\left(\sqrt{2\gamma}\right) \cong \frac{1}{12}\exp(-\gamma) + \frac{1}{4}\exp\left(-\frac{4\gamma}{3}\right)$$

where, γ is SNR, $\psi_{BPSK}(\gamma) = Q\left(\sqrt{2\gamma}\right)$ is conditional probability of error for BPSK.

Hence, the conditional BER of cooperative communication over the fading channel employing BPSK can be calculated as sum of the conditional BER for the two phases. In first case, assume that there may be no successful decoding at the R ($\tilde{p}_2 = 0$), hence MRC receiver combines two signals one from the R and the other from S. In second case, assume that there is successful decoding at the R ($\tilde{p}_2 = p_2$) and MRC receiver combines two signals one from the R and the other from S. In case

1, $\psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)$ is the chance of incorrect decoding at the R and for such case $\tilde{p}_2 = 0$. In case 2 for BPSK, the chance of correct decoding at the R is $\left(1-\psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)\right)$ and for this case $\tilde{p}_2 = p$.

Hence,

$$P_{BPSK}^{cond} = P_{BPSK}^{cond,1st} + P_{BPSK}^{cond,2nd} \tag{12.16}$$

where, $P_{BPSK}^{cond,1st}$ and $P_{BPSK}^{cond,2nd}$ are respectively, for first and second cases mentioned above.

Therefore,

$$P_{BPSK}^{cond} = \psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)\times\psi_{BPSK}(\gamma)\Big|_{\tilde{p}_2=0} + \left(1-\psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)\right)\times\psi_{BPSK}(\gamma)\Big|_{\tilde{p}_2=p_2}$$

$$\because \qquad \gamma = \frac{p_1|f|^2+\tilde{p}_2|h|^2}{N_0};\ \tilde{p}_2 = \begin{cases}0, & case1\\ p_2, & case2\end{cases}$$

$$\therefore P_{BPSK}^{cond} = \psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)\times\psi_{BPSK}\left(\frac{p_1|f|^2}{N_0}\right)+\left(1-\psi_{BPSK}\left(\frac{p_1|g|^2}{N_0}\right)\right)\times\psi_{BPSK}\left(\frac{p_1|f|^2+p_2|h|^2}{N_0}\right)$$

Assume that the fading channels, *f*, *g* and *h* are independent of each other and are identically distributed. We can calculate the approximate average BER by averaging the conditional BER over the channel gain coefficients, *f*, *g* and *h* as

$$P_{BPSK}(e) = \zeta_{1st}^{s,r}(p_1)\,\zeta_{1st}^{s,d}(p_1) + \left(1-\zeta_{1st}^{s,r}(p_1)\right)\zeta_{2nd}^{s,r,d}(p_1+p_2) \tag{12.17}$$

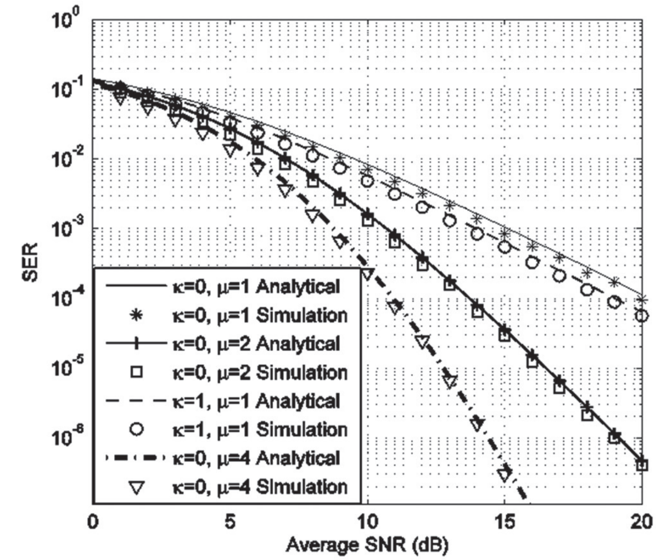
where,

$$\zeta_{1st}^{s,r}(p_1) = \frac{1}{12}MGF_{p_1}(1)+\frac{1}{4}MGF_{p_1}\left(\frac{4}{3}\right);\ \zeta_{1st}^{s,d}(p_1) = \frac{1}{12}MGF_{p_1}(1)+\frac{1}{4}MGF_{p_1}\left(\frac{4}{3}\right);$$

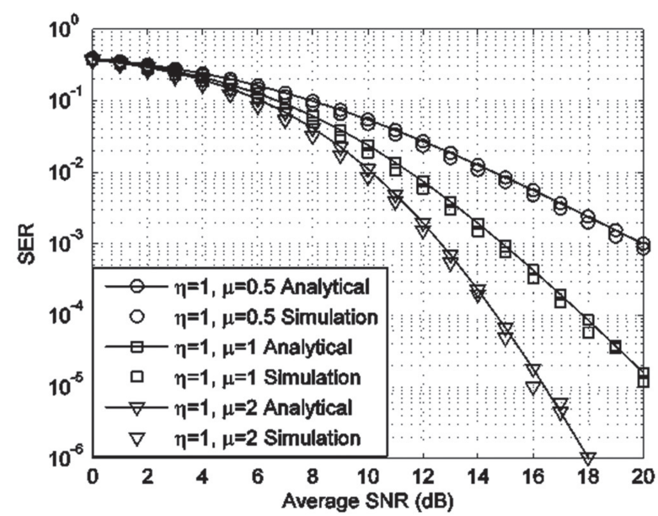
$$\zeta_{2nd}^{s,r,d}(p_1+p_2) = \frac{1}{12}MGF_{p_1}(1)\times MGF_{p_2}(1)+\frac{1}{4}MGF_{p_1}\left(\frac{4}{3}\right)\times MGF_{p_2}\left(\frac{4}{3}\right).$$

The MGF of k-μ or η-μ distributed fading are given in section 2.4.

The derived approximate expressions of SER has been evaluated numerically and plotted with respect to average SNR, and are compared with the simulation results for different values of k-μ and η-μ fading channels. They are found to be in good agreement (B. Kumbhani et al., 2014) for BPSK and QPSK modulation schemes.



(a)



(b)

Fig. 12.4 SER vs. SNR for SISO based cooperative communication system employing (a) BPSK modulation over k-μ fading channel (b) QSPK modulation scheme over η-μ fading channel

Example 12.1

Write a short MATLAB program to generate the BER vs. SNR for BPSK modulation scheme for Rayleigh ($k = 0; \mu = 1$) as a special case of k-μ distribution.

Solution

```
%### ANALYTICAL BIT ERROR RATE FOR COOPERATIVE COMMUNICATION #######
clc;
clear all;
close all;
kapa = 0;%input('enter kappa value');
%ratio of dominant component to total power
mu = 1;%input('enter no of clusters');
P11 = 0:20; %SNR in dB
P = 10.^(P11/10);

P1 = P/2; % equal power allocation for both
P2 = P/2; % relay and transmitter

for i = 1:length(P11)

    z11 = mu*(1+kapa)/(mu*(1+kapa)+P1(i));
    z12 = exp(mu*kapa*(z11-1));
    z13 = mu*(1+kapa)/(mu*(1+kapa)+(4/3)*P1(i));
    z14 = exp(mu*kapa*(z13-1));

    z21 = mu*(1+kapa)/(mu*(1+kapa)+P2(i));
    z22 = exp(mu*kapa*(z21-1));
    z23 = mu*(1+kapa)/(mu*(1+kapa)+(4/3)*P2(i));
    z24 = exp(mu*kapa*(z23-1));

     mgf1p1 = z11^mu * z12;
     mgf13p1 = z13^mu * z14;

     mgf1p2 = z21^mu * z22;
     mgf13p2 = z23^mu * z24;

     RICBER(i) = (1/12)* mgf1p1 +(1/4)* mgf13p1;
     DICBER1(i) = (1/12)* mgf1p2 +(1/4)* mgf13p2;
     DICBER(i) = (1/12)*mgf1p1* mgf1p2 + (1/4)* mgf13p1* mgf13p2;
     RCBER(i) = 1-RICBER(i);

    ber(i) = RICBER(i)*DICBER1(i)+RCBER(i)*DICBER(i);

end

figure
semilogy(P11,ber)
%grid on

clc;
clear all;
N=10^6; % number of bits or symbols
ip = rand(1,N)>0.5; % generating 0,1 with equal probability

s = 2*ip-1; % BPSK modulation 0 -> -1; 1 -> 0
SNR1 = 0:20; % multiple Eb/N0 values
    total = 10.^(SNR1/10);
        SNR = 10*log10(total/2);
```

```
        clear i;

%%%%%%%% COOPERATIVE COMMUNICATION %%%%%%%%%%%%%

    for ii = 1:length(SNR1)

        n1 = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N));
        % white gaussian noise, 0dB variance
        n2 = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N));
        % white gaussian noise, 0dB variance
        n3 = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N));
        % white gaussian noise, 0dB variance
        f = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N)); % Rayleigh channel
        g = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N)); % Rayleigh channel
        h = 1/sqrt(2)*(randn(1,N) + 1j*randn(1,N)); % Rayleigh channel

        y1 = (f.*s)+ n1.*10^(-SNR(ii)/20); % Channel and Noise addition

        y2 = (g.*s) + n2.*10^(-SNR(ii)/20); % Channel and Noise addition

        y2Hat = (conj(g).*y2); % equalization maximal ratio combining

        ip1Hat = real(y2Hat)>0; % receiver - hard decision decoding

        rem=2*ip1Hat-1;%relay message

        y3 =(h.*rem) + n3.*10^(-SNR(ii)/20);
        % Channel and noise Noise addition


    clear kk;
    for kk = 1:10^6

        if  ip(kk)==ip1Hat(kk)

            y3Hat(kk) = conj(h(kk))*y3(kk)+conj(f(kk))*y1(kk);

        else
            y3Hat(kk) = conj(f(kk))*y1(kk);

        end
    end
         ip3Hat = real(y3Hat)>0;% receiver - hard decision decoding
        nErr1(ii) = size(find(ip- ip3Hat),2);% counting the errors

  end

%figure
hold on
simBer1 = nErr1/N; % simulated ber
semilogy(SNR1,simBer1,'--');
grid on
xlabel('SNR-->');
ylabel('BER-->');
```

Review question 12.2 *What are the two commonly used protocols at the relay nodes?*

Review question 12.3 *For single relay based cooperative communication as shown in Fig. 12.3, write down the approximate BER for BPSK over i.i.d. channels.*

12.3.2 MIMO based cooperative communication over α-μ fading channels

Recently, dual hop transmission has attracted the attention of many researchers because of their significant advantage over direct transmission, such as extending the coverage area, increasing the connectivity and improving the SER performance, etc. Let us consider a dual hop relaying system (see Fig. 12.5 a) over i.i.d. α–μ fading channel, where an S node communicate to the D with the help of $\{R_1 \quad R_2 \quad \ldots \quad R_K\}$ using DF protocol. Assume the S node is equipped with N_T transmit antennas, k[th] relay node equipped with N_K receive antennas and N_T transmit antennas and D node equipped with N_D receive antennas. The OSTBC technique is applied at every transmit and receive nodes. We assume S_D data symbols are transmitted in T time slots which are encoded with OSTBC codeword. We also assume all Rs and D have S-R and R-D channel gain information respectively, while S and all Rs do not have S-Rs and Rs-D channel gain information respectively. We also assume that relays are half duplex, which means Rs either receive or transmit the information at any time. The instantaneous SNR output at the k[th] R can be expressed as in (Q. Yang et al., 2010).

$$\gamma = \frac{TE_S}{S_D \sigma_n^2} \sum_{i=1}^{N_K} \sum_{j=1}^{N_T} \left| h_{i,j}^k \right|^2 \tag{12.18}$$

with E_S is power of transmitted signal per antenna and σ_n^2 is noise power, $h_{i,j}^k$ is the channel gain between the i[th] receive antenna and j[th] transmit antenna of the S-k[th] R. We consider orthogonal transmission approach to transmit information from R nodes to D nodes. We can transmit signal through time-division technique or through frequency-division technique. In time-division technique, total $(\bar{K}+1)T$ time slots will be required to transmit S_D data symbols information from S to D, where $\bar{K}$ is number of Rs participating in R nodes to D nodes transmission. In first T time slots, S nodes transmit the information to all R nodes. Only those Rs will participate for decoding sets that have good enough S–R link to provide successful decoding of all S_D data symbols. In time-division technique, we also assume that channel conditions do not change for $(\bar{K}+1)T$ time slots. At the D nodes, for MRC receiver, the instantaneous SNR output can be expressed as

$$\gamma = \frac{TE_S}{S_D \sigma_n^2} \sum_{k=0}^{\bar{K}} \sum_{i=1}^{N_K} \sum_{j=1}^{N_T} \left| \bar{h}_{i,j}^k \right|^2 \tag{12.19}$$

where, $\bar{h}_{i,j}^k \left(k \in [1, \bar{K}]\right)$ represents channel gain between the i[th] receive antenna and j[th] transmit antenna of k[th] R and D node link and $\bar{h}_{i,j}^0$ represents channel gain between the i[th] receive antenna and j[th] transmit antenna of S–D link.

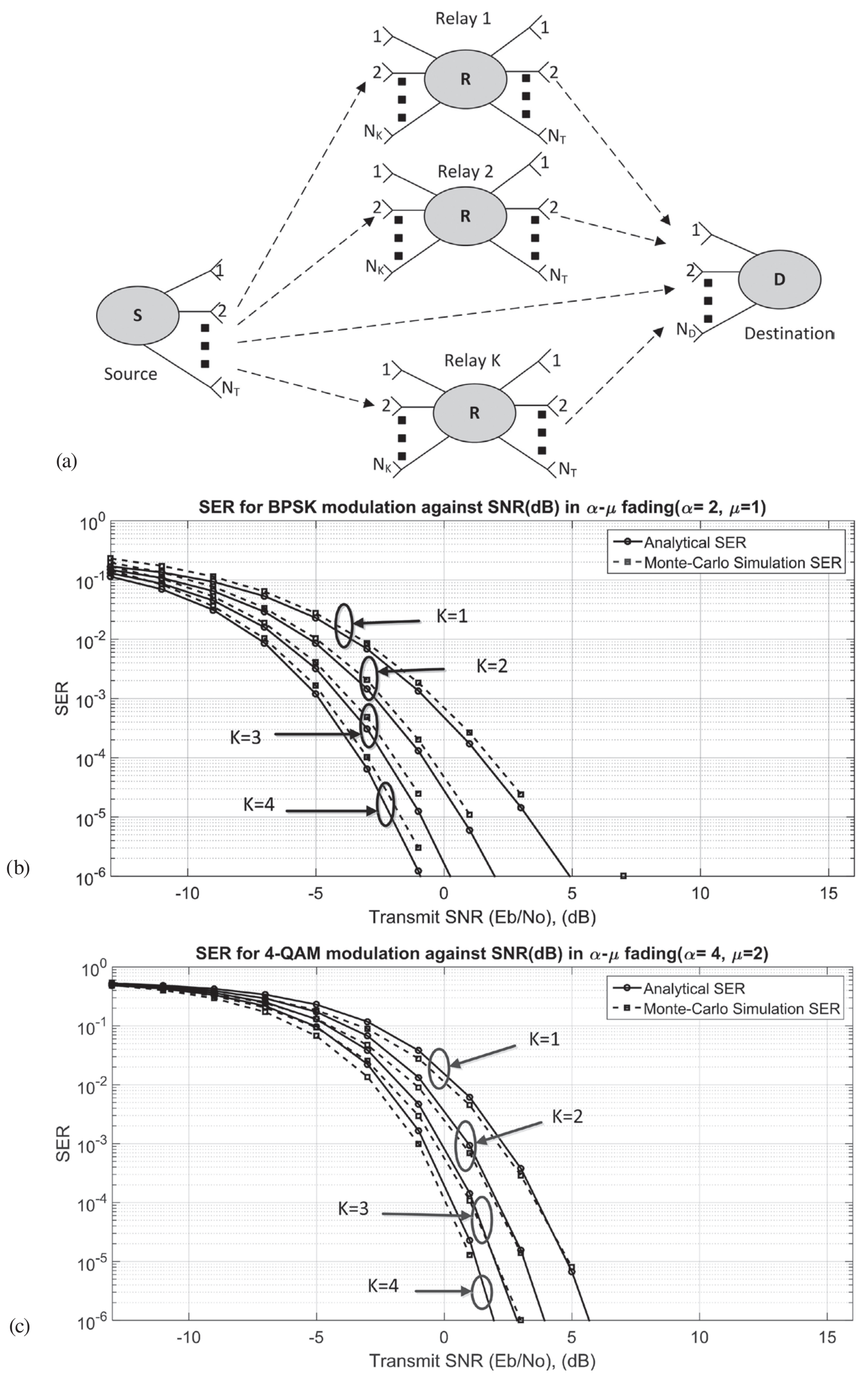


Fig. 12.5 (a) Two hop multiple relays model; approximate and Monte Carlo simulations of SER performance of MIMO based cooperative communication employing (b) BPSK modulation over the $\alpha = 2$ and $\mu = 1$ (Rayleigh) fading channel and (c) 4-QAM modulation over the $\alpha = 4$ and $\mu = 2$ fading channel

We assume DF relays, so Rs will forward the information to D iff, S–R link is sufficient to decode all S_D data symbols correctly. It means, all K relays may not participate for Rs–D transmission. By using the theorem on total probability, we have,

$$P_{SER} = \sum_{\bar{K}=1}^{K} \Pr\left(SER / \bar{K}\right)\Pr\left(\bar{K}\right) \tag{12.20}$$

where, P_{SER} denotes total symbol-error probability, K is total number of relays that participate for source-relays transmission, $\bar{K}$ is total number of relays that participate for relays-destination transmission, $\Pr\left(\bar{K}\right)$ denotes probability that only $\bar{K} \in \left[1, K\right]$ relays participate for relays-destination transmission and $\Pr\left(SER / \bar{K}\right)$ is the conditional SER for given $\bar{K}$.

Assuming identical fading conditions for all the relays, the probability of any relay being able to decode the symbol correctly becomes A. Then, it can be shown that

$$P\left(\bar{K}\right) = {}^{\bar{K}}C_K A^{\bar{K}} \left(1 - A\right)^{K-\bar{K}} \tag{12.21}$$

Since $\bar{K}$ relays are used, the mgf of the i.i.d. α-μ wireless channels between the source to destination and relay to the destination can be written as (A. Magableh et al., 2009).

$$M_\gamma\left(s\right) = \left(\frac{\alpha\mu^\mu\sqrt{\frac{k}{l}}}{2\Gamma\left(\mu\right)\left(\frac{\bar{\gamma}s}{l}\right)^{\frac{\alpha\mu}{2}}\left(2\pi\right)^{\frac{l+k-2}{2}}}\right)^{N_T N_D\left(\bar{K}+1\right)} \left(G_{l,k}^{k,l}\left(\left(\frac{\mu}{k\bar{\gamma}^{\frac{\alpha}{2}}}\right)^k \left(\frac{l}{s}\right)^l \middle| \begin{matrix} I\left(l, 1-\frac{\alpha\mu}{2}\right) \\ I\left(k,0\right) \end{matrix}\right)\right)^{N_T N_D\left(\bar{K}+1\right)} \tag{12.22}$$

For the k^th relay, the mgf for source to relay channel can be obtained from the above mgf by putting $\bar{K} = 0$ and substituting $N_k = N_D$. The probability of error when $\bar{K}$ relays are used for transmission can be calculated as follows:

$$P\left(SER\middle|\bar{K}\right) = \int_0^\infty P\left(SER\middle|\bar{K},\gamma\right)p_\gamma\left(\gamma\right)d\gamma$$

where, $P\left(SER\middle|\bar{K},\gamma\right)$ is the conditional probability of error (CPE) for different modulation schemes used.

For BPSK,

$$P\left(SER\middle|\bar{K}\right) = \int_0^\infty Q\left(\sqrt{2\gamma}\right)p_\gamma\left(\gamma\right)d\gamma \approx \int_0^\infty \left(\frac{1}{12}e^{-\gamma} + \frac{1}{4}e^{-\frac{4\gamma}{3}}\right)p_\gamma\left(\gamma\right)d\gamma = \frac{1}{12}M_\gamma\left(1\right) + \frac{1}{4}M_\gamma\left(\frac{4}{3}\right) \tag{12.23}$$

For 4-QAM,

$$P\left(SER\middle|\bar{K}\right) = \int_0^\infty \left(2Q\left(\sqrt{\gamma}\right) - Q^2\left(\sqrt{\gamma}\right)\right)p_\gamma\left(\gamma\right)d\gamma$$

$$\approx \int_0^\infty \left(\frac{1}{6} e^{-\frac{\gamma}{2}} + \frac{1}{2} e^{-\frac{2\gamma}{3}} - \frac{1}{144} e^{-\gamma} - \frac{1}{16} e^{-\frac{4\gamma}{3}} - \frac{1}{24} e^{-\frac{7\gamma}{6}} \right) p_\gamma(\gamma) d\gamma$$

$$= \frac{1}{6} M_\gamma\left(\frac{1}{2}\right) + \frac{1}{2} M_\gamma\left(\frac{2}{3}\right) - \frac{1}{144} M_\gamma(1) - \frac{1}{16} M_\gamma\left(\frac{4}{3}\right) - \frac{1}{24} M_\gamma\left(\frac{7}{6}\right) \quad (12.24)$$

Using Eq. (12.20), approximate SER of MIMO based cooperative communication over $\alpha-\mu$ fading channel is plotted vs. SNR for BPSK and 4-QAM modulation schemes in Fig. 12.5 (b) and (c) (A. K. Saxena et al., 2016). It can be observed that the SER for MIMO based cooperative communication improves when more number of relays (Rs) participates in the cooperative communication. Some of the parameters chosen are $N_D = N_T = N_K = 2$ and $S_D = 2$. Equal power allocation is applied for all transmit nodes and Alamouti space-time code is employed. The simulation and analytical results are in close agreement verifying the accuracy of the analytical results.

Review question 12.4 *What is the approximate SER of 4-QAM MIMO based cooperative communication over i.i.d. α-μ fading channel?*

12.4 Large-scale MIMO systems

12.4.1 Introduction

Large-scale (LS) MIMO systems are aiming to employ hundreds or thousands of antennas at the transmitter and receiver. What are the advantages? We can have large diversity and rate gains. For example, for a 100 × 100 MIMO system, the achievable diversity gain is 10,000 and rate gain is 100. Till today, such large antenna arrays are not used in wireless communications like in Long term evolution (LTE) based fourth generation (4G) mobile wireless communications have allowed up to a maximum of 8 antennas at the base station (BS). But in the future generation wireless systems like 5G, such large antenna arrays employing hundreds or thousands of antennas are envisaged to be used.

Let us make some typical calculations for large antenna arrays. Consider a mobile phone whose maximum dimension is 9 cm which is a typical size. How many antennas can be placed on such mobile phones? From our knowledge of antenna arrays, antenna spacing for antenna arrays should be at least $\frac{\lambda}{2}$. For GSM, the operating frequency is 900 MHz. In order to simplify our calculation, let us approximate this frequency to 1 GHz which is quite close. $\frac{\lambda}{2} = 15cm$ for 1GHz. So we cannot place more than single antenna on such mobile phone with the present available technology.

How about millimetre wave communication at 60 GHz? $\frac{\lambda}{2} = 0.25cm$ for 60 GHz. Now we have the scope for placing at least 36 antennas on the same mobile phone for 60 GHz wireless communications. We have neglected the space occupied by antenna in our analysis. This is one of the reasons why researchers are proposing millimetre wave communications for 5G mobile communications. What if we want to place more than 36 antennas on the same mobile phone? We can do so by placing antennas in three dimensions, which is also known as full dimension MIMO (Y.-H. Nam et al., 2013). For example, we can place several antennas in a MIMO cube (B. N. Getu et al., 2005 and C.-Y. Chiu et al., 2008).

An important question to be raised is "Can we use the same techniques we have used for conventional MIMO systems at the transmitter and receiver for LS MIMO systems?". At the transmitter one may employ full rate and full diversity codes like Threaded Algebraic Space Time Codes, Perfect STBC and Rateless STBC (A. H. Alqahtani et al., 2014).

What happen to the MIMO channel when $N_T, N_R \geq 100$? For large antenna arrays, channel hardening effect becomes more prominent. In other words, the channel becomes more and more deterministic for N_R, $N_T \to \infty$. According to Marcenko–Pastur law from random matrix theory, for a $N_R \times N_T$ channel matrix $\mathbf{H}$ ($N_T, N_R \to \infty$) whose elements are zero mean i.i.d. with variance $\frac{1}{N_R}$ (irrespective of the actual distribution of each elements), the empirical distribution of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ converge almost surely to the density (A. Tulino et al., 2004)

$$f_\beta(x) = \left(\frac{\beta - 1}{\beta}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+ (b-x)^+}}{2\pi\beta x} \tag{12.25}$$

where, $\beta = \frac{N_T}{N_R}$, $a = \left(1 - \sqrt{\beta}\right)^2$, and $(z)^+ = \max(z, 0)$.

Because of channel hardening, we can employ low complexity MIMO detection techniques. So far we have seen that ML detection is the optimal detection techniques for conventional MIMO systems. But we cannot use ML detection for large MIMO systems since it has prohibitively large computational complexity for large MIMO systems (see example 9.1). The decoding complexity of an ML decoder is the number of metric computations required to reach the ML decision. For instance, for a 10 × 10 MIMO system employing 16-QAM modulation scheme such metric calculations will be $10^{16} = 1.0995 \times 10^{12}$ which is prohibitively large for LS MIMO that will employ hundreds to thousands of antennas at the transmitter and receiver. But certain algorithms from machine learning and artificial intelligence have shown to achieve near-optimal performance with low complexities (A. Chockalingam et al., 2014) for large MIMO systems viz. Reactive tabu search (N. Srinidhi et al., 2009), Gibbs sampling (M. Hansen et al., 2009). Alamouti like STBC for transmission and its detection using heuristic based search algorithms like Tabu and Hill-climbing search (F. Glover, 1989 and 1990, S. Gupta et al., 2015) are also reported in the literature. It is better to initialize the initial solution with ZF and MMSE solution to improve the performance of such heuristic based detection methods. A good survey on LS-MIMO detection is provided by S. Yang et al., (2015).

Review question 12.5 *What is Marcenko–Pastur law?*

Review question 12.6 *Name some possible detection schemes for LS-MIMO.*

12.4.2 Single user LS-MIMO: capacity and hardware impacts

The received signal vector $\mathbf{y} \in C^{N_R \times 1}$ for a point-to-point or single user (SU) MIMO, we have considered till now, can be expressed as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}$$

where, $\mathbf{x} \in C^{N_T \times 1}$ is the transmit signal vector and $\mathbf{n} \in C^{N_R \times 1}$ represents the noise vector with zero mean and covariance matrix $\sigma_n^2 \mathbf{I}_{N_R}$.

The instantaneous capacity for MIMO channel (Eq. 4.4 b) is given by

$$C = \log_2 \left| \mathbf{I}_{R_H} + \frac{P}{\sigma_n^2 N_T} \mathbf{Q} \right| bits / s / Hz$$

Assume full rank MIMO channel matrix, then $R_H = \min\{N_R, N_T\} = m$. Hence the above instantaneous capacity (Eq. 4.4 a) can be expressed as

$$C = \sum_{i=1}^{m} \log_2 \left(1 + \frac{\lambda_i P}{N_T \sigma_n^2} \right) bits / s / Hz$$

Since the trace of a square matrix **Q** is equal to sum of its eigenvalues (see Appendix A), i.e.,

$$\sum_{i=1}^{m} \lambda_i = \sum_{i=1}^{m} \sigma_i^2 = \text{trace}(\mathbf{Q}).$$

(a) The worst case for capacity is when only one singular value of the channel matrix H is not zero, i.e., $\lambda_1 = \sigma_1^2 = \text{trace}\ (\mathbf{Q})$. Such cases are appropriate for line-of-sight (LOS) propagation. Hence, the instantaneous capacity for MIMO channel is lower bounded by

$$C \geq \log_2 \left(1 + \frac{P\left(trace\left(\mathbf{Q}\right)\right)}{N_T \sigma_n^2} \right) \quad bits / s / Hz$$

(b) The best case is when all the singular values are equal, i.e., $m\lambda_e = \text{trace}\ (\mathbf{Q})$. This is suitable for i.i.d. channel matrix. Hence, the instantaneous capacity for MIMO channel is upper bounded by

$$C \leq m \log_2 \left(1 + \frac{P\left(trace\left(\mathbf{Q}\right)\right)}{m N_T \sigma_n^2} \right) \quad bits / s / Hz$$

If we normalize the magnitude of the channel gain coefficients equal to one, then, trace (Q) ≈ $N_T N_R$. Hence, the instantaneous capacity for MIMO channel is bounded as follows:

$$\log_2 \left(1 + \frac{P N_R}{\sigma_n^2} \right) bits / s / Hz \leq C \leq m \log_2 \left(1 + \frac{nP}{N_T \sigma_n^2} \right) bits / s / Hz; n = \max\left\{N_R, N_T\right\} \quad (12.26)$$

Review question 12.7 *What is the capacity lower and upper bound for SU-MIMO?*

Large-scale asymptotic analysis

Case 1: Let $N_T \to \infty$, keeping N_R fixed. Assume the favourable condition of channel orthogonalization where the rows of the channel matrix are asymptotically orthogonal, i.e., $\left(\frac{\mathbf{HH}^H}{N_T}\right)_{N_T >> N_R} \approx \mathbf{I}_{N_R}$.

Note that $m = N_R$ for this case. Hence,

$$C_{N_T >> N_R} \approx \log_2 \left| \mathbf{I}_{N_R} + \frac{P\mathbf{I}_{N_R}}{\sigma_n^2} \right| = N_R \log_2 \left(1 + \frac{P}{\sigma_n^2} \right) \tag{12.27}$$

Case 2: Let $N_R \to \infty$, keeping N_T fixed. Assume the favourable condition of channel orthogonalization where the columns of the channel matrix are asymptotically orthogonal, i.e., $\left(\frac{\mathbf{H}^H\mathbf{H}}{N_R}\right)_{N_R >> N_T} \approx \mathbf{I}_{N_T}$.

Note that $m = N_T$ for this case. Hence,

$$C_{N_R >> N_T} \approx \log_2 \left| \mathbf{I}_{N_T} + \frac{PN_R\mathbf{I}_{N_T}}{N_T\sigma_n^2} \right| = N_T \log_2 \left(1 + \frac{PN_R}{N_T\sigma_n^2} \right) \tag{12.28}$$

Note that $C_{N_T >> N_R}$ and $C_{N_R >> N_T}$ are both highly favourable scenarios since they achieve the capacity upper bound mentioned above (F. Rusek et al., 2013).

Review question 12.8 *What is $C_{N_T >> N_R}$ and $C_{N_R >> N_T}$ for SU LS-MIMO?*

Impact of hardware impairments in SU LS-MIMO

The I–O model of MIMO system taking into account the hardware impairments can be expressed as

$$\mathbf{y} = \mathbf{H}\,(\mathbf{x} + \boldsymbol{\eta}_t) + \mathbf{n} + \boldsymbol{\eta}_r$$

where the additive distortion noise terms $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_r$ for the hardware impairments at the transmitter and the receiver can be modelled (Central limit theorem) as $\boldsymbol{\eta}_t \sim N_C\left(0, \delta_t^2\left(diag\left(r_1, r_2, \cdots, r_{N_T}\right)\right)\right)$ and $\boldsymbol{\eta}_r \sim N_C\left(0, \delta_r^2\left(trace\left(\mathbf{R}_x\right)\right)\mathbf{I}_{N_R}\right)$. Note that $\mathbf{R}_x = E[\mathbf{xx}^H]$ and $r_1, r_2, \ldots r_{N_T}$ are the diagonal elements of the transmitted signal covariance matrix ($\mathbf{R}_x$). Error vector magnitude (EVM) quantifies the mismatch between the expected signal and the actual signal in RF transreceivers (H. Holma et al., 2011). For $\delta_t = \delta_r = 0$, there are no hardware impairments or ideal case. A higher value of δ_t and δ_r signifies higher hardware impairments. A typical range of values for δ_t, for instance for LTE, is $0.08 \leq \delta_t \leq 0.175$. Let us define the average SNR per receive antenna as,

$$\rho = \frac{E\left[trace\left(\mathbf{R}_x\right)\right]}{\sigma_n^2} = \frac{E\left[trace\left(\frac{P}{N_T}\mathbf{I}_{N_T}\right)\right]}{\sigma_n^2}$$

We have defined complex Wishart matrix in chapter 3 as, $\mathbf{Q} = \begin{cases} \mathbf{H}^H\mathbf{H} & if \quad N_T < N_R \\ \mathbf{H}\mathbf{H}^H & if \quad N_T \geq N_R \end{cases}$. In the following analysis, **H** represents the Rician fading MIMO channel. Let us define a new matrix which takes into account the hardware impairments as $\mathbf{\Phi} = \frac{\rho\delta_t^2}{N_T}\mathbf{Q} + \left(\rho\delta_r^2 + 1\right)\mathbf{I}_{R_H}$ where the rank of the full rank channel matrix is $R_H = \begin{cases} N_T & for \quad N_T < N_R \\ N_R & for \quad N_T \geq N_R \end{cases}$. The ergodic achievable rate (J. Zhang et al.) can be expressed as

$$C^{hardware} = E\left\{\log_2\left|\mathbf{I}_{R_H} + \frac{\rho\mathbf{Q}\mathbf{\Phi}^{-1}}{N_T}\right|\right\} \tag{12.29}$$

Review question 12.9 *Write down the I–O model for SU MIMO taking into account hardware impairments.*

Example 12.2

Carry out the asymptotic LS-MIMO analysis taking into account the hardware impairments.

Solution

Let us do the asymptotic LS-MIMO analysis.

Case 1: Let $N_T \to \infty$, keeping N_R fixed. Assume the favourable condition of channel orthogonalization where the rows of the channel matrix are asymptotically orthogonal, i.e., $\left(\frac{\mathbf{H}\mathbf{H}^H}{N_T}\right)_{N_T >> N_R} \approx \mathbf{I}_{N_R}$.

Note that $m = N_R$ for this case. Hence, using the dominated convergence theorem (R. Couillet et al., 2011), we have,

$$C_{N_T>>N_R}^{hardware} \approx N_R \log_2\left(1 + \frac{\rho}{\rho\delta_t^2 + \rho\delta_r^2 + 1}\right) \tag{12.30}$$

Therefore, the capacity $C_{N_T>>N_R}^{hardware}$ taking into account the hardware impairments depends on N_R, ρ, δ_t and δ_r.

Case 2: Let $N_R \to \infty$, keeping N_T fixed. Assume the favourable condition of channel orthogonalization where the columns of the channel matrix are asymptotically orthogonal, i.e., $\left(\frac{\mathbf{H}^H\mathbf{H}}{N_R}\right)_{N_R >> N_T} \approx \mathbf{I}_{N_T}$.

Note that $m = N_T$ for this case. Hence, using the dominated convergence theorem (R. Couillet et al., 2011), we have,

$$C_{N_R>>N_T}^{hardware} \approx N_T \log_2\left(1 + \frac{1}{\delta_t^2}\right) \tag{12.31}$$

Therefore, the capacity $C_{N_R >> N_T}^{hardware}$ taking into account the hardware impairments depends only N_T and δ_t and independent of ρ and δ_r. It shows that it is better to employ low-cost hardware at the receiver.

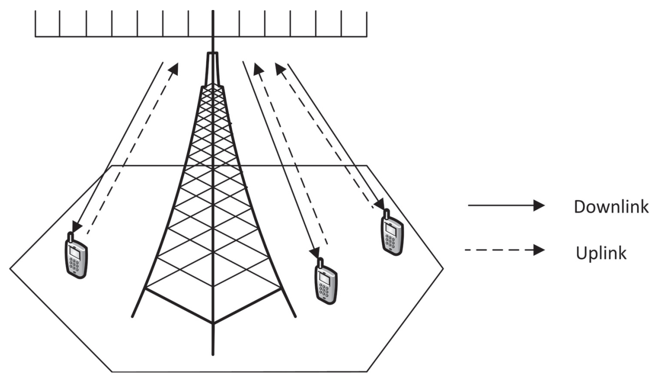


Fig. 12.6 Illustration of uplink (multiple access channel) and downlink (broadcast channel) for MU-MIMO system (N_{BS} = 14, N_{MS} = 1, N_U = 3)

Review question 12.10 *What are achievable rates for SU LS-MIMO taking into account hardware effect?*

Review question 12.11 *Is it true that it is better to employ low-cost hardware at the receiver for SU LS-MIMO?*

12.4.3 Multiuser LS-MIMO: capacity and matched filter processing

Multi-user MIMO scenario (H. Huang et al., 2012 and T. L. Marzetta, 2010) is the most applicable scenario for cellular communications where multiple users transmit and receive signals from a base station. In multiuser (MU) LS-MIMO as depicted in Fig. 12.6, there are two types of MU-MIMO channel models (X. Ma et al., 2014 and Y. S. Cho et al., 2010):

- Multiple-access channel (MAC)
- Broadcast channel (BC)

Multiple-access channel

In MAC, we assume that a single base station (BS) is receiving signals from multiple mobile station (MS) or users. We will denote the number of users N_U and each MS or user is equipped with N_{MS} antennas. Signals from multiple users are up linked to a BS equipped with N_{BS} antennas. Assume each user is sending transmit signal vector $\mathbf{x}_i^{UL} \in C^{N_{MS} \times 1}, i = 1, 2, \cdots, N_U$ to the BS. The corresponding channel for each user with the BS is $\mathbf{H}_i^{MAC} \in C^{N_{BS} \times N_{MS}}, i = 1, 2, \cdots, N_U$. The received signal vector at BS $\mathbf{y}^{UL} \in C^{N_{BS} \times 1}$ can be expressed as

$$\mathbf{y}^{UL} = \mathbf{H}_1^{MAC}\mathbf{x}_1^{UL} + \mathbf{H}_2^{MAC}\mathbf{x}_2^{UL} + \cdots + \mathbf{H}_{N_U}^{MAC}\mathbf{x}_{N_U}^{UL} + \mathbf{n}^{UL}$$
$$= \sum_{i=1}^{N_U}\mathbf{H}_i^{MAC}\mathbf{x}_i^{UL} + \mathbf{n}^{UL}$$

where, $\mathbf{n}^{UL} \in C^{N_{BS}\times 1}$ is the additive white Gaussian noise vector with zero mean and covariance matrix of $\sigma_n^2\mathbf{I}_{N_{BS}}$.

The above equation may be expressed as

$$\mathbf{y}^{UL} = \mathbf{H}^{MAC}\mathbf{x}^{UL} + \mathbf{n}^{UL} \tag{12.32}$$

where, $\mathbf{H}^{MAC} = \begin{bmatrix}\mathbf{H}_1^{MAC} & \mathbf{H}_2^{MAC} & \cdots & \mathbf{H}_{N_U}^{MAC}\end{bmatrix}, \mathbf{x}^{UL} = \begin{bmatrix}\mathbf{x}_1^{UL} \\ \mathbf{x}_2^{UL} \\ \vdots \\ \mathbf{x}_{N_U}^{UL}\end{bmatrix}$

Broadcast channel

The BC is a single BS sending signals to multiple MSs or users in the downlink. Note that each MS will receive independent signal vector, $\mathbf{x}^{DL} \in C^{N_{BS}\times 1}$. The corresponding channel vector for BS to each user will be a matrix, $\mathbf{H}_i^{BC} \in C^{N_{MS}\times N_{BS}}, i = 1,2,\cdots,N_U$. The received signal vector at the i[th] MS $\mathbf{y}_i^{DL} \in C^{N_{MS}\times 1}, i = 1,2,\cdots,N_U$ can be expressed as

$$\mathbf{y}_i^{DL} = \mathbf{H}_i^{BC}\mathbf{x}^{DL} + \mathbf{n}_i^{DL}$$

where, $\mathbf{n}_i^{DL} \in C^{N_{MS}\times 1}$ is the additive white noise vector with zero mean and covariance matrix $\sigma_n^2\mathbf{I}_{N_{MS}}$.

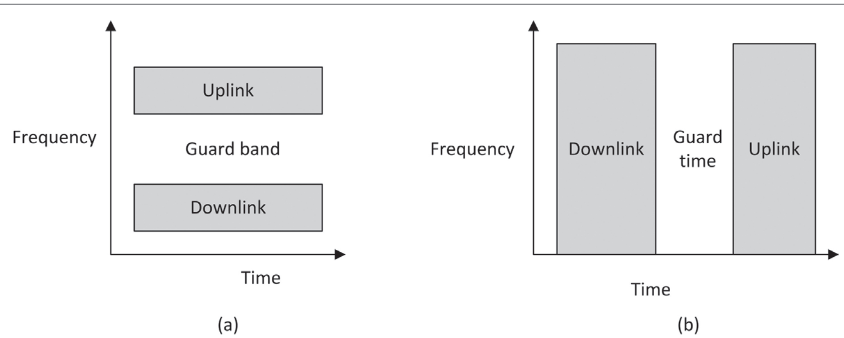


Fig. 12.7 (a) Frequency-division duplex (FDD) and (b) Time-division duplex (TDD)

If we assume time-division duplex (TDD) transmission (R. Brandt, 2014) as depicted in Fig. 12.7 then $\mathbf{H}_i^{BC} = \left(\mathbf{H}_i^{MAC}\right)^T$ then

$$\mathbf{y}_i^{DL} = \left(\mathbf{H}_i^{MAC}\right)^T \mathbf{x}^{DL} + \mathbf{n}_i^{DL} \tag{12.33}$$

For all users,
$$\mathbf{y}^{DL} = \left(\mathbf{H}^{MAC}\right)^T \mathbf{x}^{DL} + \mathbf{n}^{DL}$$

where, $\mathbf{y}^{DL}$, $\mathbf{H}^{MAC} = \begin{bmatrix} \mathbf{H}_1^{MAC} & \mathbf{H}_2^{MAC} & \cdots & \mathbf{H}_{N_U}^{MAC} \end{bmatrix}$ and $\mathbf{y}^{DL} = \begin{bmatrix} \mathbf{y}_1^{DL} \\ \mathbf{y}_2^{DL} \\ \vdots \\ \mathbf{y}_{N_U}^{DL} \end{bmatrix}$

Review question 12.12 *What are the two types of MU-MIMO channels?*

Review question 12.13 *Write down the I–O model for MU-MIMO MAC and BC.*

Capacity and Matched Filter Precoding

Consider BS equipped with N_{BS} antennas serving N_U single antenna users. Let us denote the i^th^ user to the j^th^ antenna of the BS (L. Lu et al., 2014) as

$$h_{ji} = g_{ji}\sqrt{d_i}$$

where g_{ji} and d_i represent the complex small–scale fading and large-scale fading coefficients respectively.

Therefore,

$$\mathbf{H}^{MAC} = \mathbf{G}\mathbf{D}^{\frac{1}{2}} \tag{12.34}$$

where, $\mathbf{D} = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_{N_U} \end{pmatrix}$ and $\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1N_U} \\ g_{21} & g_{22} & \cdots & g_{2N_U} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N_{BS}1} & g_{N_{BS}2} & & g_{N_{BS}N_U} \end{pmatrix}$

Assume favourable condition where all column vectors of the channel are orthogonal, i.e., $\left(\frac{\mathbf{G}^H\mathbf{G}}{N_{BS}}\right)_{N_{BS}>>N_U} \approx \mathbf{I}_{N_U}$ therefore, $\left(\left(\mathbf{H}^{MAC}\right)^H \mathbf{H}^{MAC}\right)_{N_{BS}>>N_U} \approx \mathbf{D}^{\frac{1}{2}}\mathbf{G}^H\mathbf{G}\mathbf{D}^{\frac{1}{2}} = N_{BS}\mathbf{D}$. Let us assume uplink transmit power for each user is $\frac{P}{N_U}$. Hence the capacity of the multiuser MIMO (BS equipped with N_{BS} antennas acting as receiver and N_U single antenna users as transmitter) for uplink (MAC) is given as

$$C_{N_{BS}>>N_U}^{MAC} \approx \log_2\left|\mathbf{I}_{N_U} + \frac{PN_{BS}\mathbf{D}}{N_U\sigma_n^2}\right| = \sum_{i=1}^{N_U}\log_2\left(1 + \frac{PN_{BS}d_i}{N_U\sigma_n^2}\right) \tag{12.35}$$

Let us consider a simple linear processing matched-filter (MF) at the BS.

$$\left(\mathbf{H}^{MAC}\right)^H \mathbf{y}^{UL} = \left(\mathbf{H}^{MAC}\right)^H \left(\mathbf{H}^{MAC}\mathbf{x}^{UL} + \mathbf{n}^{UL}\right) \approx N_{BS}\mathbf{D}\mathbf{x}^{UL} + \left(\mathbf{H}^{MAC}\right)^H \mathbf{n}^{UL} \tag{12.36}$$

Note that **D** is a diagonal matrix; therefore, MF processing at the BS decouples the signals from each user. The SNR for each user can be evaluated as, $N_{BS}\frac{P}{\sigma_n^2 N_U}d_i$. Since we have parallel independent Gaussian channels, the capacity of this will be the same with that of $C^{MAC}_{N_{BS}>>N_U}$. Hence MF is an optimal processing at the BS when the number of antennas at the BS grows large.

BS has full CSI, so adaptive power allocation could be carried out. Hence, the capacity for downlink (BC) may be expressed as

$$C^{BC} = \max_{\mathbf{D_P}} \log_2\left(\mathbf{I}_{N_{BS}} + \frac{P}{\sigma_n^2}\mathbf{H}^{BC}\mathbf{D_P}\left(\mathbf{H}^{BC}\right)^H\right)$$

where, $\mathbf{D_P}$ is a positive diagonal matrix with power allocation for N_U users as

$$\mathbf{D_P} = \begin{pmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_{N_U} \end{pmatrix}, \sum_{i=1}^{N_U} p_i = P.$$

Assume favourable condition where all row vectors of the channel are orthogonal, we have the capacity for downlink (BC) as

$$C^{BC}_{N_{BS}>>N_U} \approx \max_{\mathbf{D_P}} \log_2\left(\mathbf{I}_{N_U} + \frac{PN_{BS}}{\sigma_n^2}\mathbf{D_P}\mathbf{D}\right) \tag{12.37}$$

Like MAC case, we may use MF precoder and send the transmitted signal vector as

$$\mathbf{x}^{DL}_{pre} = \left(\mathbf{H}^{MAC}\right)^* \mathbf{D}^{-\frac{1}{2}}\mathbf{D}_P^{\frac{1}{2}}\mathbf{x}^{DL}$$

Hence,

$$\begin{aligned} \mathbf{y}^{DL} &= \left(\mathbf{H}^{MAC}\right)^T \mathbf{x}^{DL}_{pre} + \mathbf{n}^{DL} \\ &= \left(\mathbf{H}^{MAC}\right)^T \left(\mathbf{H}^{MAC}\right)^* \mathbf{D}^{-\frac{1}{2}}\mathbf{D}_P^{\frac{1}{2}}\mathbf{x}^{DL} + \mathbf{n}^{DL} \\ &= N_{BS}\mathbf{D}^{\frac{1}{2}}\mathbf{D}_P^{\frac{1}{2}}\mathbf{x}^{DL} + \mathbf{n}^{DL} \end{aligned} \tag{12.38}$$

Since both **D** and $\mathbf{D}_P$ are diagonal matrices, hence, all the signals transmitted from BS to each user are decoupled totally. The above analysis is also referred to as Massive MIMO in the literature.

Review question 12.14 *Write down the capacity expression for UL and DL of MU LS-MIMO.*

12.4.4 Multi-cell LS-MIMO: precoders

It is well-known that there are many cells in a cellular network and users in each cell are served by a BS. Figure 12.8 illustrates a multi-cell MIMO based cellular network. Only seven cells are shown for illustration purpose. The UL, DL and interference signals are also shown for two neighbouring cells for four mobile users. In TDD based LS-MIMO system, pilot sequences are transmitted from the users to the BS to help BS in estimating the channels. Usually pilot sequences are orthogonal and they are limited in number for a given period and bandwidth. Generally same set of pilot sequences are assigned in all cells. So users in the neighbouring cells with the same set of pilot sequences will have high level of interference also popularly known as pilot contamination in the literature. As the number of cells or BSs increases this interference will increase.

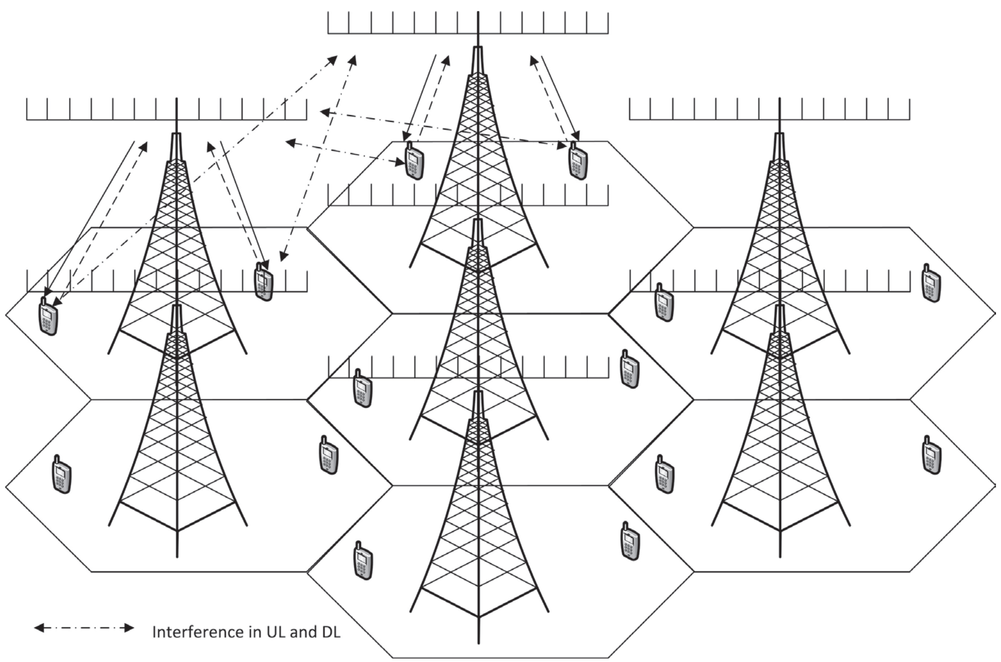


Fig. 12.8 Multi-cell MIMO based cellular network (BS equipped with $N_{BS} = 14$ antennas and single antenna MS or user, each cell has $N_U = 2$ users for illustration purpose)

As we have seen for MU-MIMO, in which we have considered a single BS, MF processing decouples the signals from each user completely. But for multi-cell MIMO, the estimated channel vector in each cell is a linear combination of channel vectors of users in other cells that use the same pilot. Hence, MF processing will not work well. Other than MF processing, other precoders are reported in the literature.

(a) **ZF precoding**
In ZF beamforming we pre-multiply the transmit signal vector by

$$\mathbf{W}_{ZF} = \frac{1}{\sqrt{\gamma_l}}\left(\hat{\mathbf{H}}_l\right)^H\left(\hat{\mathbf{H}}_l\left(\hat{\mathbf{H}}_l\right)^H\right)^{-1} \quad (12.39)$$

where, $\hat{\mathbf{H}}_l = \mathbf{d}_{ll}^{-1}\hat{\mathbf{G}}_l, \mathbf{d}_{ll} = \begin{pmatrix} \sqrt{d_{l1l}} & & & \\ & \sqrt{d_{l2l}} & & \\ & & \ddots & \\ & & & \sqrt{d_{lN_Ul}} \end{pmatrix}$, $\gamma_l = \frac{trace\left(\left(\hat{\mathbf{H}}_l\right)^H \hat{\mathbf{H}}_l\right)}{N_U}$ and $\hat{\mathbf{G}}_l$ is the estimated CSI matrix of users in the l^{th} BS.

(b) **Regularized ZF precoding**
In RZF beamforming we pre-multiply the transmit signal vector by

$$\mathbf{W}_{RZF} = \frac{1}{\sqrt{\gamma_l}}\left(\hat{\mathbf{H}}_l\right)^H\left(\hat{\mathbf{H}}_l\left(\hat{\mathbf{H}}_l\right)^H + \delta\mathbf{I}_{N_U}\right)^{-1} \quad (12.40)$$

where, δ is a parameter that balances the interference suppression and SNR decrease. Usually δ is chosen as $\frac{N_{BS}}{20}$ (F. Rusek et al., 2013). $\delta = 0$ and $\delta = \infty$ reduces to MF and ZF, respectively.

(c) **MMSE precoding**
It is a special case of RZF (M. Joham et al., 2005) where

$$\delta = \frac{N_U \sigma_n^2}{2SNR^{DL}\log_2 M} \quad (12.41)$$

It has been observed that RZF and MMSE based precoders outperforms ZF and MF based precoders for multi-cell MIMO systems (J. Jose et al., 2011).

Review question 12.15 *What is pilot contamination in multi-cell LS-MIMO?*

Review question 12.16 *Name some precoding schemes for multi-cell LS-MIMO.*

12.4.5 Interference suppression: Coordinated multipoint transmission (CoMP) and Heterogeneous networks

The instantaneous capacity of i^{th} user at time slot t in a multi-cell MIMO network can be calculated as

$$C_i(t) = \log_2\left(1 + SINR_i(t)\right) \quad (12.42)$$

where signal-to-interference-plus-noise ratio (SINR) is defined for i^{th} user as

$$SINR_i = \frac{P_{ds}(t)}{P_{noise}(t) + P_{IUI}(t) + P_{ICI}(t)} \quad (12.43)$$

Note that

- $P_{ds}(t)$ is the power of the desired signal at time slot t
- $P_{IUI}(t)$ is the power of the inter-user interference (intracell) at time slot t
- $P_{ICI}(t)$ is the power of the intercell interference at time slot t
- $P_{noise}(t)$ is the power of the noise at time slot t

As illustrated in Fig. 12.9, users at the cell edge will have very high level of $P_{ICI}(t)$ (quite near to the BS of neighbouring cells) and very low level of $P_{ds}(t)$ (far away from the serving BS). Consequently they have a lower value of SINR and capacity.

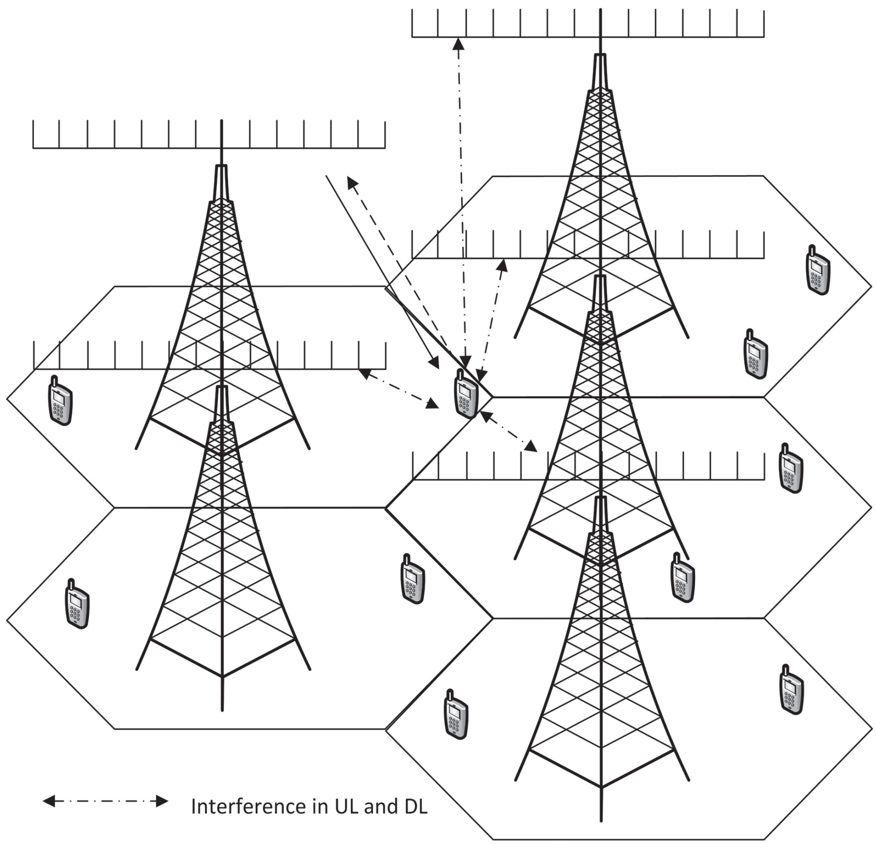


Fig. 12.9 Interference issues for users at cell edge

Review question 12.17 *Define instantaneous capacity of i^{th} user at time slot t in a multi-cell MIMO network.*

Coordinated multipoint transmission

ICI can be reduced by coordination among BSs also popularly known as coordinated multipoint transmission (CoMP). Two main objectives of CoMP are

- Mitigate ICI at the cell edge
- Improve data rate at the cell-edge

There are two kinds of CoMP employing either full or partial cooperation among BSs. In full cooperation of BSs, also popularly called as Network MIMO, all the coordinating BSs share data and CSI of all users (see Fig. 12.10). It behaves as a single distributed MIMO system to serve the users. When the concept of serving BS vanishes, all the coordinating BSs act as single distributed MIMO network. In dynamic cell selection, data to each user is transmitted from the coordinating cell with the best channel condition keeping the other cells muted. Hence there are no ICI and best possible data rate is given to the users. It has been observed that fully cooperative CoMP has enormous gain in terms of spectral efficiency (G. J. Foschini et al., 2006).

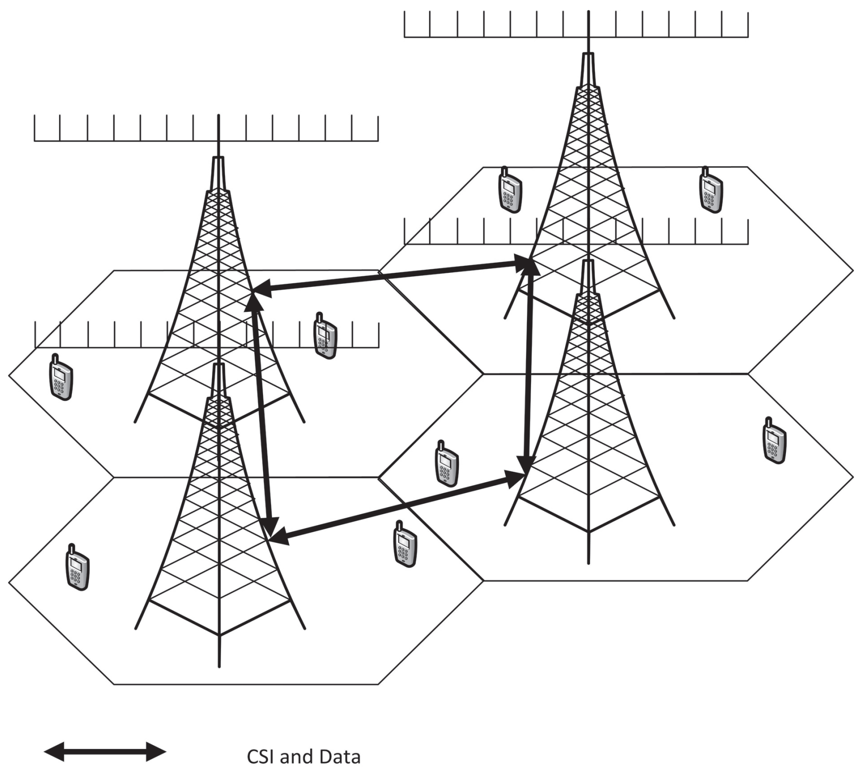


Fig. 12.10 A cluster of fully cooperating four BSs

In partial cooperation, each user is served by only one BS but the scheduling/ beamforming is shared among coordinating BSs. In coordinated scheduling/ beamforming, user data is available only at the serving BS and are not shared over the backhaul links. But CSI of users are shared via the backhaul links in order to have coordination in beamforming/ scheduling. Note that coordinated scheduling/ beamforming with no data sharing among coordinated BSs is more practical than Network MIMO for limited backhaul capacity. Interested readers may refer to N. Seifi et al., (2016) for 3D coordinated beamforming for reducing ICI in multi-cell MIMO systems.

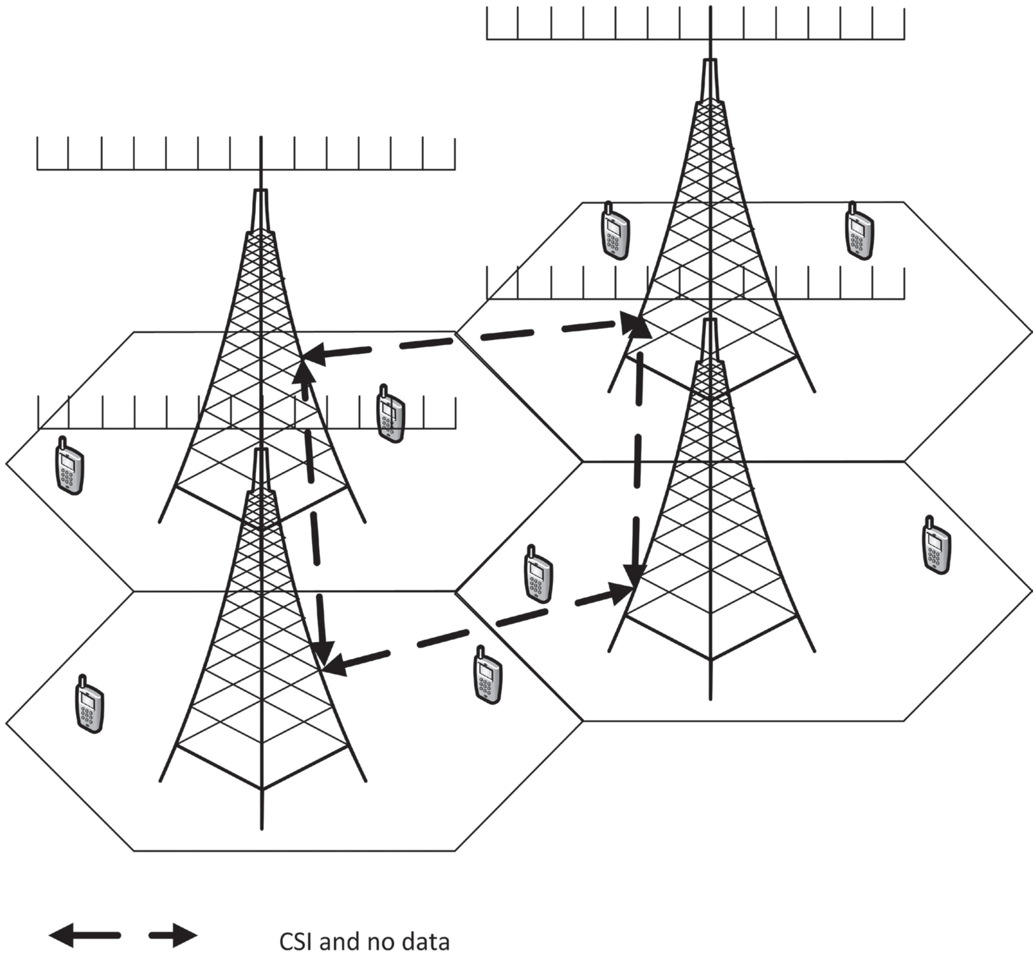


Fig. 12.11 Coordinated beamforming/ scheduling

Heterogeneous networks:

Another solution to the problem of cell edge users is to have a small cell networks (T. Q. S. Quek et al., 2013) like pico cells within the macro cell networks. In such heterogeneous networks (networks of macro cells and pico cells), there is a high level of co-channel interference as shown in Fig. 12.12. Interference suppression in such networks could be carried out in three steps (Y. Li et al., 2016):

- Triangular decomposition of joint channel matrix **H** and extraction of the equivalent interference channel model which reduces the inter-cell interference to half
- From equivalent interference channel model, use signal-to-leakage-plus-noise-ratio to compute the equivalent pre-coding matrices to suppress rest of the inter-cell interference
- Compute the intra-cell interference suppressing precoding matrices for each user

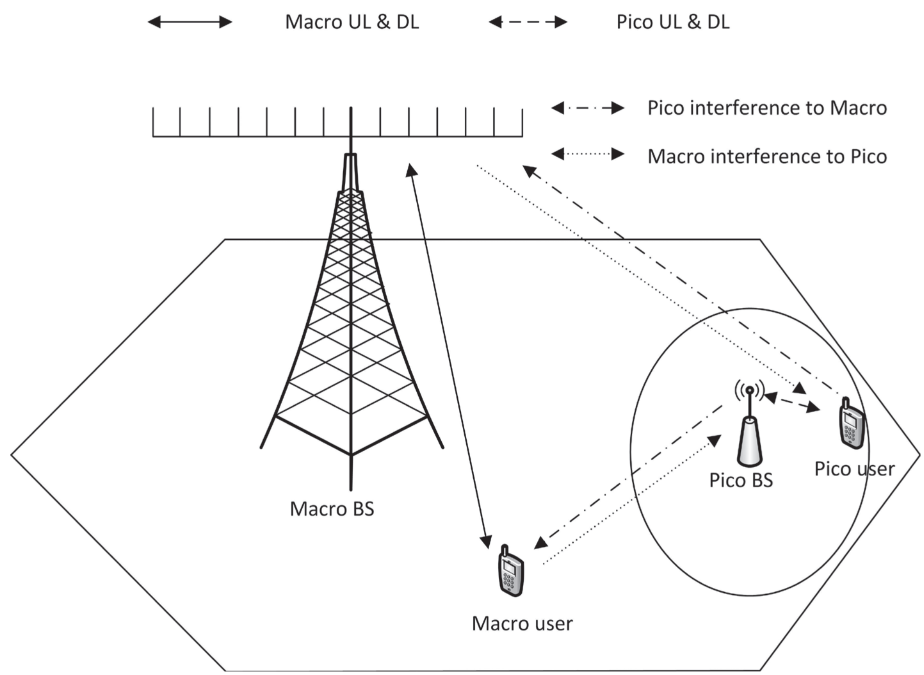


Fig. 12.12 DL and UL interference model in multi-cell MIMO heterogeneous networks

Review question 12.18 *What are the problems faced by users at the cell edge?*

Review question 12.19 *Name two possible solutions to the cell edge user problem.*

12.5 MIMO cognitive radios

12.5.1 What is cognitive radio?

Cognitive radio (CR) is an intelligent wireless communication system that is aware of its surrounding environment (S. Haykin, 2005). It is a smart and flexible radio (Secondary User (unlicensed) device) that monitors and senses its radio environment for potential spectrum opportunities.

The three main tasks of cognitive radio system are:

- Radio scene analysis (sensing) at the receiver i.e., estimation of the interference temperature and detection of spectrum holes.

  What is spectrum hole?

  It is a band of frequencies assigned to a Primary User (PU), but at a particular time and location, that band is not used by that PU. Spectrum utilization can be made efficient by allowing Secondary User (SU), who is not being serviced to gain access to the spectrum hole.

- Channel identification at the receiver i.e., CSI and prediction of channel capacity for use by the transmitter.
- Transmitting power and dynamic spectrum management at the transmitter.

In MIMO cognitive radios, SU and PU employ multiple antennas for transmission of signals which can reap the benefits of MIMO wireless communications like higher spectral efficiency and link reliability. A very detailed literature survey on MIMO cognitive radio is provided in book by R. C. Qiu et al., (2012). Conventionally, regulated radio frequency bands use spectrum allocations (using licensing procedures). Measurements on the licensed band show severe temporal and/or spatial underutilization of the assigned spectral resources. Hence, there is imbalance between the spectrum shortage and spectrum underutilization. An innovative spectrum access strategy called spectrum pooling can overcome this.

12.5.2 What is spectrum pooling?

If we allow opportunistic Secondary User (unlicensed) access to spectral resources unused by Primary User (licensed), then there will be significant improvement in spectrum utilization. But Secondary User (SU) transmission must avoid any harmful interference to Primary User (PU) systems. PU generally broadcasts these interference constraints to SU. Usually specified by two parameters I_{th} and $\Pr_{I_{th}}$, I_{th} is the maximum allowable interference power at the PU and $\Pr_{I_{th}}$ is the probability that the interference power level at PU exceeds I_{th}. SU has to transmit under those interference constraints. This can be mathematically expressed as, $\Pr(I_n > I_{th}) < \Pr_{I_{th}}$ where, I_n is the interference at the PU. C. Sun et al., (2010) reported an algorithm for calculating adaptive power control at the SU transmitter (Tx) of CR system which allows the SU receiver (Rx) to maintain a constant output SNR simultaneously limiting the interference to the PU. M. F. Hanif et al., (2011) have shown that antenna selection may be exploited to jointly satisfy interference constraint at the PU Rxs while improving the rates of SU devices. SUs regularly perform efficient radio scene analysis to detect the presence of PU signals. For a detailed understanding of spectrum sensing techniques viz. energy detection, cyclostationary detection, pilot-based coherent detection, covariance based detection and wavelet-based detection, one may refer to J. Ma et al., (2009). Error in sensing causes missed detection and false alarm.

What is missed detection and false alarm?

The SU senses the PU to be idle (sleeping) even though it is active (transmitting). Hence SU transmit with full power causing more interference to the PU. It is an unwanted situation called missed detection. We will denote the probability of missed detection by P_D. In another scenario, the SU senses that PU is active even if it is sitting idle. It forces the SU to stop transmission or transmit under interference constraint reducing the throughput of SU unnecessarily. We will denote probability of false alarm P_{FA}. The following CR models are widely used in the literature.

12.5.3 Interweave model

In CR model where SU transmits only when the PU is sleeping, is called interweave model. Let us denote two switching functions one at the SU transmitter and SU receiver (K.-C. Chen et al., 2009) which are defined as

- S_t equals 1 i.e., switch on SU transmission when SU transmitter detects no active PU nearby. S_t equals 0 i.e., switch off SU transmission when SU transmitter detects active PU nearby.
- S_r equals 1 i.e., switch on SU transmission when SU receiver detects no active PU nearby. S_r equals 0 i.e., switch off SU receiver when SU transmitter detects active PU nearby.

We may write the I-O model for secondary user MIMO link as

$$\mathbf{y} = S_r\left[\mathbf{H}\left(S_t\mathbf{x}\right) + \mathbf{n}\right]$$

12.5.4 Underlay and overlay model

In underlay model, simultaneous transmission of PU and SU is allowed if the interference to PU from SU transmission is below some acceptable threshold. In overlay model, SU allocates a part of its power to help the PU in its transmission and remaining power is used for its own transmission. Overlay model has many implementation problems since it allows the SU to decode PU transmission which raises many privacy and security issues for PU (E. Biglieri et al., 2013).

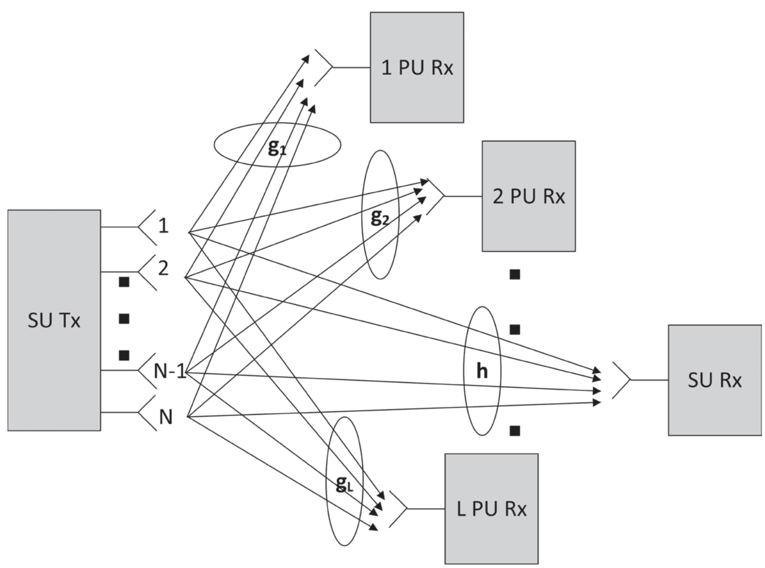


Fig. 12.13 CR network with L single-antenna PUs and 1 SU with Tx employing N antennas and Rx employing single antenna

12.5.5 Robust beamforming

Consider a CR network (depicted in Fig. 12.13) with a SU and L PUs. SU transmitter (Tx) uses N antennas and SU receiver (Rx) employs single antenna. All PUs use single antenna. The interfering channels between the SU Tx and PUs are denoted by $\mathbf{g}_l$, $l = 1, 2, \ldots, L$. The channel between the SU Tx and SU Rx is denoted by $\boldsymbol{h}$. The objective of robust beamforming is to maximize SU received power under constraint P and control the PUs interference to acceptable limits denoted by I_l, $l = 1, 2, \ldots, L$. Mathematically,

$$\max_{\|\mathbf{w}\|^2 \le P} \left|\mathbf{h}^H \mathbf{w}\right|^2 \quad s.t. \quad \left|\mathbf{g}_l^H \mathbf{w}\right|^2 \le I_l \forall l \tag{12.44}$$

where, **w** is the beamforming vector at the SU and s.t. is the abbreviation for such that.

Assuming the CSI errors as additive complex Gaussian noise we have

$$\mathbf{h} = \hat{\mathbf{h}} + \Delta\mathbf{h}; \mathbf{g}_l = \hat{\mathbf{g}}_l + \Delta\mathbf{g}_l \forall l$$

where, $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}_l$ are the channel estimates available at the SU Tx and assume

$$\Delta\mathbf{h} \sim N_C\left(0, \sigma_h^2 \mathbf{I}\right); \Delta\mathbf{g}_l \sim N_C\left(0, \sigma_{g_l}^2 \mathbf{I}\right) \forall l.$$

The Eq. (12.44) becomes

$$\max_{\|\mathbf{w}\|^2 \le P} \Pr\left(\left|\mathbf{h}^H \mathbf{w}\right|^2 \ge P_{th}\right) \quad s.t. \quad \Pr\left(\left|\mathbf{g}_l^H \mathbf{w}\right|^2 < I_l\right) \ge \varepsilon_l \forall l \tag{12.45}$$

where, ε_l is some predetermined level, P_{th} is the threshold of SU transmission power and Pr is short form of Probability. Note that $y_h = \left|\mathbf{h}^H \mathbf{w}\right|^2 = \left|\hat{\mathbf{h}}^H \mathbf{w} + \Delta\mathbf{h}\mathbf{w}\right|^2$ is non-central Chi square distributed with two degrees of freedom with variance, $\sigma_{y_h} = \frac{\|\mathbf{w}\|^2 \sigma_h^2}{2}$, non-centrality parameter, $nc_{y_h}^2 = \left|\hat{\mathbf{h}}^H \mathbf{w}\right|^2$. Therefore, $\Pr\left(\left|\mathbf{h}^H \mathbf{w}\right|^2 \ge P_{th}\right) = Q_M\left(\frac{nc_{y_h}}{\sigma_{y_h}}, \frac{\sqrt{P_{th}}}{\sigma_{y_h}}\right)$, where, Q_M is the generalized Marcum Q function. Similarly, $y_{g_l} = \left|\mathbf{g}_l^H \mathbf{w}\right|^2 = \left|\hat{\mathbf{g}}_l^H \mathbf{w} + \Delta\mathbf{g}_l \mathbf{w}\right|^2$ is also non-central Chi square distributed with two degrees of freedom with variance, $\sigma_{y_{g_l}} = \frac{\|\mathbf{w}\|^2 \sigma_{g_l}^2}{2}$, non-centrality parameter, $nc_{y_{g_l}}^2 = \left|\hat{\mathbf{g}}_l^H \mathbf{w}\right|^2$. So, $\Pr\left(\left|\mathbf{g}_l^H \mathbf{w}\right|^2 < I_l\right) = Q_M\left(\frac{nc_{y_{g_l}}}{\sigma_{y_{g_l}}}, \frac{\sqrt{I_l}}{\sigma_{y_{g_l}}}\right)$. Hence we can rewrite the Eq. (12.45) as follows.

$$\max_{\|\mathbf{w}\|^2 \le P} Q_M\left(\frac{nc_{y_h}}{\sigma_{y_h}}, \frac{\sqrt{P_{th}}}{\sigma_{y_h}}\right) \quad s.t. \quad Q_M\left(\frac{nc_{y_{g_l}}}{\sigma_{y_{g_l}}}, \frac{\sqrt{I_l}}{\sigma_{y_{g_l}}}\right) \le 1 - \varepsilon_l \forall l \tag{12.46}$$

G. Zheng et al., (2010) showed that the Eq. (12.46) can be solved optimally using second-order cone programming (SOCP) in tandem with simple 1-D search on the transmit power. They have also reported that for single PU and single SU case, the close form solution of optimal **w** exists.

12.5.6 Precoding

The basic idea of precoding is to minimize the interference in CR network. Assume that there is a single PU Tx, single SU Tx, single SU Rx and L PU Rx all of them employing multiple antennas. The received signal at L PUs Rx and 1 SU Rx (depicted in Fig. 12.14) are

$$\mathbf{y}_{P_i} = \frac{\sqrt{\rho_{PP_i}}\sqrt{P_P}}{\sqrt{N_T}}\mathbf{H}_{PP_i}\mathbf{x}_P + \frac{\sqrt{\rho_{SP_i}}\sqrt{P_S}}{\sqrt{N_T}}\mathbf{H}_{SP_i}\mathbf{x}_S + \mathbf{n}_{P_i}, i = 1,2,\cdots,L$$

$$y_S = \frac{\sqrt{\rho_{SS}}\sqrt{P_S}}{\sqrt{N_T}}\mathbf{H}_{SS}\mathbf{x}_S + \mathbf{n}_S$$

where,

- N_T is the number of transmit antennas at the Tx
- N_R is the number of receive antennas at the Rx
- $\mathbf{H}_{PP_i}$ is the $N_R \times N_T$ channel matrix between the PU Tx and ith PU Rx
- $\mathbf{H}_{SP_i}$ is the $N_R \times N_T$ interference channel matrix between the SU Tx and ith PU Rx
- $\mathbf{H}_{SS}$ is the $N_R \times N_T$ channel matrix between the SU Tx and SU Rx
- ρ_{PP_i} are the channel gains for $\mathbf{H}_{PP_i}$
- ρ_{SP_i} are the channel gains for $\mathbf{H}_{SP_i}$
- ρ_{SS} are the channel gains for $\mathbf{H}_{SS}$
- P_P is the transmit power of PU Tx
- P_S is the transmit power of SU Tx
- $\mathbf{x}_P$ is the $N_T \times 1$ signal vector for PU Tx
- $\mathbf{x}_S$ is the $N_T \times 1$ signal vector for SU Tx
- $\mathbf{n}_{P_i}$ is the $N_R \times 1$ AWGN vector at ith PU Rx
- $\mathbf{n}_S$ is the $N_R \times 1$ AWGN vector at SU Rx

Note that all the elements of the channel matrix $\mathbf{H}_{PP_i}$, $\mathbf{H}_{SP_i}$ and $\mathbf{H}_{SS}$ are distributed as $N_C(0, 1)$. All the elements of the noise vector $\mathbf{n}_{P_i}$ and $\mathbf{n}_S$ are distributed as $N_C(0, N_0)$. $E[\mathbf{x}_P\mathbf{x}_P^H] = \mathbf{I}_{N_T} = E[\mathbf{x}_S\mathbf{x}_S^H]$. Assume that PU Tx has the knowledge of $\mathbf{H}_{PP_i}$. Besides SU Tx has the knowledge of both $\mathbf{H}_{SS}$ and $\mathbf{H}_{SP_i}$.

(a) **Pre-whitening precoding**:
Precoding matrix for L=1 (M. S. Kang et al., 2008) can be simply expressed as

$$\mathbf{w} = \mathbf{H}_{SP_1}^{-1} \tag{12.47}$$

(b) **Water-filling based precoding**:
This precoder at the SU Tx (A. J. Paulraj et al., 2003) is given by

$$\mathbf{w} = \mathbf{V}_{SS}\mathbf{\Sigma}_w \tag{12.48}$$

where, $\mathbf{H}_{SS} = \mathbf{U}_{SS}\mathbf{\Sigma}_{SS}\mathbf{V}_{SS}^H$ is the SVD of $\mathbf{H}_{SS}$ matrix and $\mathbf{\Sigma}_w$ is the optimal covariance matrix.

(c) **Interference minimized precoding**:
This precoder at the SU Tx (M. Jung et al., 2011) is given by

$$\mathbf{w} = \mathbf{V}_{I_{N_T}} \tag{12.49}$$

where the interference channel is given by $\mathbf{H}_I = \begin{bmatrix} \sqrt{\rho_{SP_1}} H_{SP_1}^T \\ \sqrt{\rho_{SP_2}} H_{SP_2}^T \\ \vdots \\ \sqrt{\rho_{SP_L}} H_{SP_L}^T \end{bmatrix}$ and its SVD

$$\mathbf{H}_t = \mathbf{U}_I \mathbf{\Sigma}_I \mathbf{V}_I^H = \begin{bmatrix} \mathbf{U}_{I_1} & \mathbf{U}_{I_2} & \cdots & \mathbf{U}_{I_{N_T}} \end{bmatrix} \begin{bmatrix} \sigma_{I_1} & 0 & \cdots & 0 \\ 0 & \sigma_{I_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{I_{N_T}} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{I_1}^H \\ \mathbf{V}_{I_2}^H \\ \vdots \\ \mathbf{V}_{I_{N_T}}^H \end{bmatrix}$$

The transmitter select a channel with minimum singular value of the interference channel and send the signals in that channel.

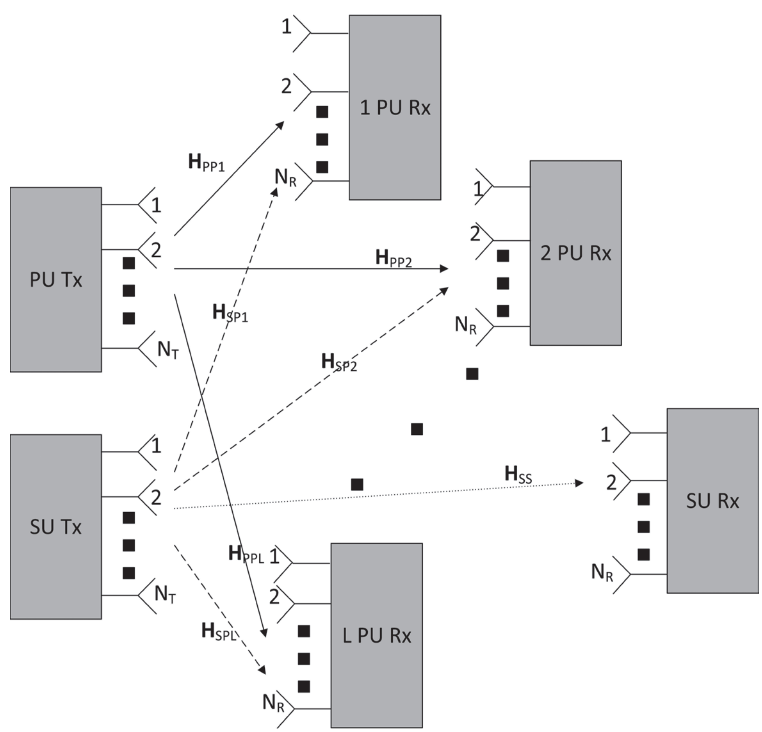


Fig. 12.14 CR network with 1 PU Tx and 1 SU Tx and L PU Rx and 1 SU Rx all employing multiple antennas

12.5.7 Game theory

A brief introduction to Game theory is given in Appendix G. With two or more players being an integral part of a game, it is natural for the study of cognitive radio to be motivated by certain ideas in game theory. In CR network, the players are the cognitive radios, their actions are their choice of transmission parameters (e.g., transmission powers, access probability, etc.) and their payoff or utility function are their defined performance measures such as QoS which can be a combination of throughput, energy, interference related parameter (e.g., SIR or SINR). A game theoretical model to obtain the optimal payoff (pricing) for dynamic spectrum sharing in CR network is proposed in (D. Niyato et al., 2008). In this multiple PU compete to give spectrum usage to the SU and firm gives prices as payoff to them. Service provider wants to maximize its revenue and user wants to have maximum QoS and lowest price. (F. Wang et al., 2008) proposed a price-based iterative water-filling algorithm which reaches to a Nash equilibrium. G. Scutani et al., (2010) proposed a game theoretical approach for MIMO cognitive radio. They have tried to solve resource allocation problem in CR network. How to allow simultaneous communication over MIMO channels among SUs under interference power constraints to PUs? Their reported algorithms have overcome the main drawback of MIMO iterative Water-filling algorithm (IWFA) i.e., the violation of the temperature-interference limits.

Review question 12.20 *What is MIMO based cognitive radios?*

Review question 12.21 *What is spectrum hole?*

Review question 12.22 *What is spectrum sensing?*

Review question 12.23 *What is probability of missed detection and false alarm?*

Review question 12.24 *Explain in few words about the interweave, underlay and overlay models of cognitive radio networks.*

Review question 12.25 *Why do we need robust beamforming in cognitive radio networks?*

Review question 12.26 *Mention three precoding techniques in cognitive radio networks.*

12.6 Summary

Figure 12.15 shows the chapter 12 in a nutshell. In this chapter, we have discussed about four recent MIMO techniques viz., STBC–SM, MIMO based cooperative communications, LS-MIMO systems and MIMO Cognitive radios. STBC–SM is the hybrid of STBC and SM; it will have advantages of both the techniques. In STBC–SM, we have derived union bound on BEP of STBC–SM over correlated Rayleigh and Rician fading MIMO channel. Then we discuss briefly about the single antenna based cooperative communication. In MIMO based cooperative communication, we have derived the approximate SER of 4-QAM over i.i.d. α-μ fading MIMO channel. In LS-MIMO systems, we have considered single user, multi-user and multi-cell MIMO systems. We have derived the capacity of LS-MIMO systems for the above three scenarios. Finally, we discuss about the interference issue for users at the cell-edge and how to overcome this by applying coordinated multipoint transmission and heterogeneous networks. In MIMO CR, we have studied about the robust beamforming, precoding and game theory briefly.

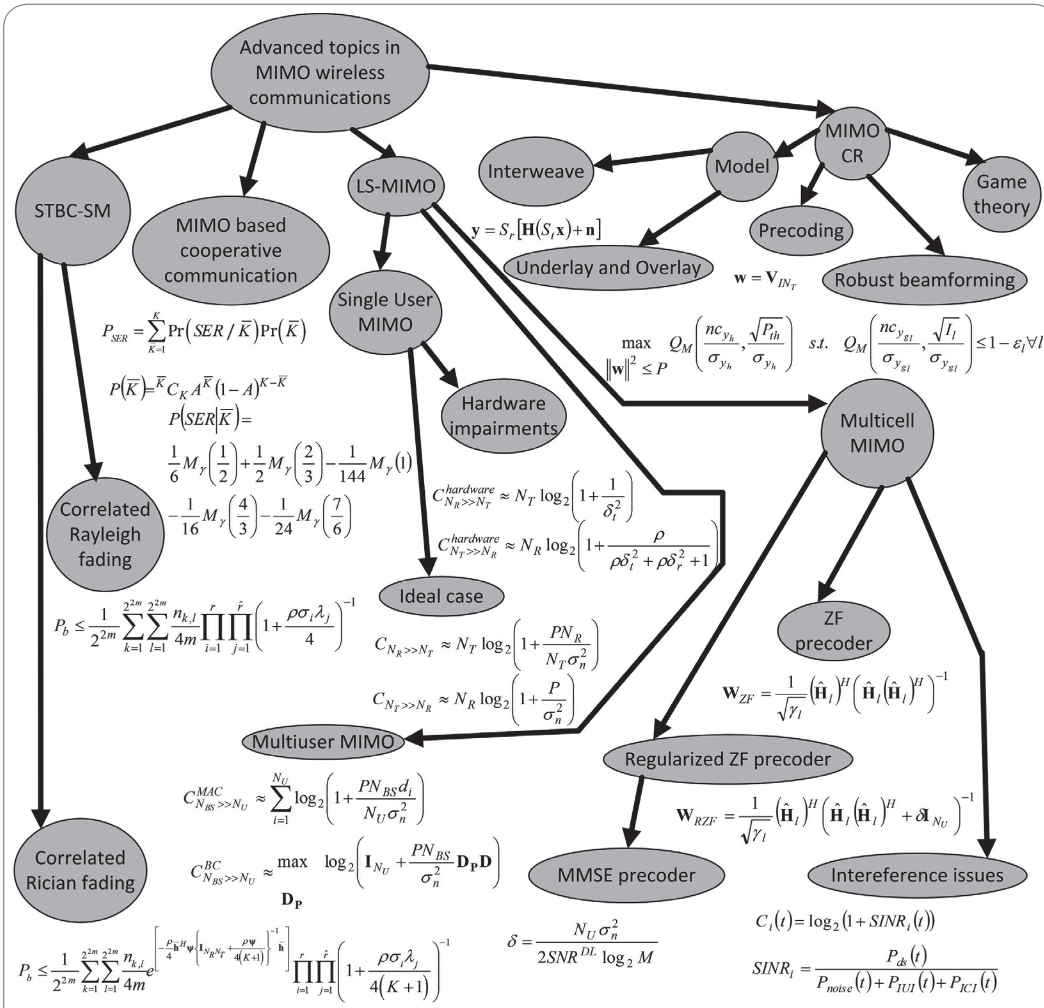


Fig. 12.15 Chapter 12 in a nutshell

Exercises

Exercise 12.1

Derive the union bound on BEP for STBC–SM over correlated Rician and Rayleigh fading MIMO channel.

Exercise 12.2

Find SER of 4-QAM modulated SISO based cooperative communication over i.i.d. k-μ fading channels.

Exercise 12.3

Find SER of 4-QAM modulated SISO based cooperative communication over i.i.d. η-μ fading channels.

Exercise 12.4

Find SER of 4-QAM modulated MIMO based cooperative communication over i.i.d. k-μ fading channels.

Exercise 12.5

Find SER of 4-QAM modulated MIMO based cooperative communication over i.i.d. η-μ fading channels.

Exercise 12.6

How to take into account the path loss in Eq. (12.12)?

Exercise 12. 7

Find $C_{N_T >> N_R}$ and $C_{N_R >> N_T}$ for SU LS-MIMO.

Exercise 12.8

Find the $C_{N_T \to \infty}$ and $C_{N_R \to \infty}$ for SU LS-MIMO taking into account hardware effects.

Exercise 12.9

Find the capacity expression for UL and DL of MU LS-MIMO.

Exercise 12.10

Explain MF processing for MU LS-MIMO.

Exercise 12.11

Explain some precoding schemes for multi-cell LS-MIMO.

Exercise 12.12

Explain Coordinated multipoint transmission in brief.

Exercise 12.13

Write down the three steps in interference suppression algorithm in a multi-cell MIMO heterogeneous network.

Exercise 12.14

What is interference minimized precoding in Cognitive radio networks?

Exercise 12.15

Explain the applications of game theory in Cognitive radio networks.

References

1. Alamouti, S. 1998. ‘A simple transmit diversity technique for wireless communications’. *IEEE Journal on Selected Areas in Communications*. 16(8). 1451–1458.
2. Alqahtani, A. H., A. I. Sulyman, and A. Alsanie. 2014. ‘Rateless space-time block code for massive MIMO systems’. *International Journal of Antennas and Propagation*.
3. Basar, E., H. Aygölü, E. Panayirci, and H. V. Poor. 2011. ‘Space-time block coded spatial modulation’. *IEEE Transactions on Communications*. 59 (3). 823-832.
4. Biglieri, E., A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor. 2013. *Principles of Cognitive Radio*. Cambridge: Cambridge University Press.
5. Brandt, R. 2014. ‘Coordinated precoding for multi-cell MIMO networks’. *PhD thesis*, KTH Sweden.
6. Brennan, D. G. 2003. ‘Linear diversity combining techniques’. *IEEE Proceedings*. 91(2). 331–356.
7. Chen, K.-C. and R. Prasad. 2009. *Cognitive Radio Networks*. Chichester: John Wiley & Sons.
8. Chiani, M., D. Dardari, and M. K. Simon. 2003. ‘New exponential bounds and approximations for the computation of error probability in fading channels’. *IEEE Transactions on Wireless Communications*. 2(4). 840–845.
9. Chiu, C.-Y., J.-B. Yan, and R. D. Murch. Apr. 2008. ‘24-port and 36-port antenna cubes suitable for MIMO wireless communications’. *IEEE Trans. Ant. And Propagat*. 56(4). 1170–1176.
10. Cho, Y. S., J. Kim, W. Y. Yang, and C.-G. Kang. 2010. *MIMO−OFDM wireless communications with MATLAB*. Singapore: John Wiley & Sons.
11. Chockalingam, A. and B. S. Rajan. 2014. *Large MIMO Systems*. Cambridge: Cambridge University Press.
12. Couillet, R. and M. Debbah. 2011. *Random Matrix Methods for Wireless Communications*. Cambridge: Cambridge University Press.
13. Foschini, G. J., K. Karakayali, and R. A. Valenzuela. Aug. 2006. ‘Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency’. *IEE Proc. Communications*. 153(4). 548–555.
14. Getu, B. N. and J. B. Andersen. May 2005. ‘The MIMO cube—a compact MIMO antenna’. *IEEE Trans. Wireless Commun*. 4(3). 1136–1141.
15. Glover, F. Summer 1989. ‘Tabu search – part I’. *ORSA Journal of Computing*. 1(3). 190–206.
16. Glover, F. Winter 1990. ‘Tabu search – part II’. *ORSA Journal of Computing*. 2(1). 4–32.
17. Gupta, S., J. Yadav, and R. S. Kshetrimayum. July 2015. ‘Low complexity detection algorithms for Alamouti like STBC for large MIMO systems’. In *Proc. IEEE Connect*. Bangalore.
18. Hanif, M. F., P. J. Smith, D. P. Taylor, and P. A. Martin. 2011. ‘MIMO cognitive radios with antenna selection’. *IEEE Transactions on Wireless Communications*. 10(11). 3688–3699.

19. Hansen, M., B. Hassibi, A. G. Dimakis, and W. Xu. Dec 2009. 'Near-optimal detection in MIMO systems using Gibbs sampling'. In *Proc. ICC*.
20. Haykin, S. Feb 2005. 'Cognitive radio: brain-empowered wireless communications'. *IEEE Journal on Selected Areas in Communications*. 23(2). 201–220.
21. Hedayat, A., H. Shah, and A. Nosratinia. 2005. 'Analysis of space-time coding in correlated fading channels'. *IEEE Transactions on Wireless Communications*. 4(6). 2882–2889.
22. Holma, H. and A. Toskala. 2011. *LTE for UMTS: Evolution to LTE-Advanced*. Chichester: John Wiley & Sons.
23. Huang, H., C. B. Papadias, and S. Venkatesan. 2012. *MIMO Communication for Cellular Networks*. New York: Springer.
24. Joham, M., W. Utschick, and J. Noseek. Aug 2005. 'Linear transmit processing in MIMO communication systems'. *IEEE Trans. on Signal Processing*. 53(8). 2700–2712.
25. Jose, J., A. Ashikhmin, T. L. Marzetta, and S. Vishwanath. Aug. 2011. 'Pilot contamination and precoding in multi-cell TDD systems'. *IEEE Trans. Wireless Communications*. 10(8). 2640–2651.
26. Jung, M., K. Hwang, and S. Choi. 2011. 'Interference minimization approach to precoding scheme in MIMO-based cognitive radio networks'. *IEEE Communications Letters*. 99. 789–791.
27. Kang, M. S., B. C. Jung, D. K. Sung, and W. Choi. Nov 2008. 'A pre-whitening scheme in a MIMO-based spectrum-sharing environment'. *IEEE Commun. Lett.* 12(11). 831–833.
28. Kumbhani, B., L. N. B. Reddy, and R. S. Kshetrimayum. Jan. 2014. 'Approximate SER of cooperative communications over generalized $\eta-\mu$ and $k-\mu$ fading channels'. *IET Journal of Engineering*. 1–4.
29. Kumbhani, B., V. K. Mohandas, R. P. Singh, S. Kabra, and R. S. Kshetrimayum. July 2015. 'Analysis of space-time block coded spatial modulation in correlated Rayleigh and Rician fading channels'. In *Proc. IEEE International Conference on Digital Signal Processing (DSP)*. Singapore. 516–520.
30. Liu, K. J. R., A. Kwasinski, W. Su, and A. K. Sadek. 2009. *Cooperative Communications and Networking*. Cambridge: Cambridge University Press.
31. Lu, L., G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang. Oct 2014. 'An overview of massive MIMO: benefits and challenges'. *IEEE Journal on Selected Topics in Signal Processing*. 8(5). 742–758.
32. Ma, J., G. Y. Li, and B. H. Juang. May 2009. 'Signal processing in cognitive radio'. *Proc. of IEEE*. 805–823.
33. Ma, X. and Q. Zhou. 2014. 'Massive MIMO and its detection'. In *MIMO Processing for 4G and Beyond*, M. M. da Silva and F. A. Monteiro, Eds. Boca Raton: CRC Press. 449–471.
34. Magableh, A. and M. M. Matalgah. June 2009. 'Moment generating function of the generalized α-μ distribution with applications'. *IEEE Communications Letters*. 13. 411–413.
35. Marzetta, T. L. Nov 2010. 'Noncooperative cellular wireless with unlimited number of base station antennas'. *IEEE Trans. Wireless Communications*. 9(11). 3590–3600.
36. Menghwar, G. and C. Mecklenbrauker. Feb 2009. 'Cooperative versus non-cooperative communications'. In *Proc. 2nd International Conference on Computer, Control and Communication*. 1–3.
37. Mesleh, R., H. Haas, C. W. Ahn, and S. Yun. 2006. 'Spatial modulation a new low complexity spectral efficiency enhancing technique'. *Proc. Int. Conf. on Communications and Networking in China*.
38. Nam, Y.-H., B. L. Ng, K. Sayana, Y. Li, and J. Zhang. June 2013. 'Full-dimension MIMO (FD-MIMO) for next generation cellular technology'. June 2013. *IEEE Communications Magazine*. 172–179.
39. Niyato, D. and E. Hossain. 2008. 'Competitive pricing for spectrum sharing in cognitive radio networks: dynamic game, inefficiency of Nash equilibrium and collusion'. *IEEE J. Sel. Areas Commun.* 26(1). 192–202.
40. Paulraj, A. J., R. Nabar, and D. Gore. 2003. *Introduction to Space-time Wireless Communications*. Cambridge: Cambridge University Press.
41. Qiu, R. C., Z. Hu, H. Li, and M. C. Wicks. 2012. *Cognitive radio communications and networking principle and practice*. Chichester: Wiley.

42. Quek, T. Q. S., G. de la Roche, I. Guvenc, and M. Kountouris. 2013. *Small Cell Networks: Deployment, PHY Techniques and Resource Management*. Cambridge: Cambridge University Press.
43. Rusek, F., D. Persson, B. K. Lau, E. G. Larsson, T. M. Marzetta, O. Edfors, and F. Tufvesson. Jan. 2013. 'Scaling up MIMO: opportunities and challenges with very large arrays'. *IEEE Signal Processing Magazine*. 40–60.
44. Saxena, A. K., S. Goyal, and R. S. Kshetrimayum. 'Approximate analysis of MIMO-based cooperative communication in α–μ channel distribution'. In *Proc. IEEE COMSNETS 2016*. Bangalore, India.
45. Scutari, G. and D. P. Palomar. 2010. 'MIMO cognitive radio: A game theoretical approach'. *IEEE Transactions on Signal Processing*. 58(2). 761–780.
46. Seifi, N., J. Zhang, R. W. Heath, Jr., T. Svensson, and M. Coldrey. October 2014. 'Coordinated 3D beamforming for interference management in cellular networks'. *IEEE Trans. on Wireless Communications*. 13(10). 5396–5410.
47. Srinidhi, N., S. K. Mohammed, A. Chockalingam, and B. S. Rajan. July 2009. 'Low complexity near-ML decoding of large non-orthogonal STBCs using reactive tabu search'. In *Proc. IEEE ISIT*.
48. Sun, C., Y. D. Alemseged, H. N. Tran, and H. Harada. May 2010. 'Transmit power control for cognitive radio over a Rayleigh fading channel'. *IEEE Transactions on Vehicular Technology*. 59(4). 1847–1857.
49. Tarokh, V., H. Jafarkhani, and A. R. Calderbank. 1999. 'Space-time block codes from orthogonal designs'. *IEEE Transactions on Information Theory* 45(5). 1456–1467.
50. Tulino, A. and S. Verdu. 2004. *Random Matrix Theory and Wireless Communications*. Delft: Now Publishers.
51. Wang, F., M. Krunz, and S. Cui. 2008. 'Price-based spectrum management in cognitive radio networks'. *IEEE J. Sel. Topics Signal Process.* 2(1). 74–87.
52. Yang, Q., K. S. Kwak, and F. Fu. 2010. 'Closed-form symbol error rate expression of decode-and-forward relaying using orthogonal space-time block coding'. *IET Communications*. 4(3). 368–375.
53. Yang, S. and L. Hanzo. Fourth quarter, 2015. 'Fifty years of MIMO detection: the road to large-scale MIMOs'. *IEEE Communication Surveys and Tutorials*. 17(4). 1941–1988.
54. Zhang, J., L. Dai, X. Zhang, E. Bjornson, and Z. Wang. 'Achievable rate of Rician large-scale MIMO channels with transreceiver hardware impairments'. To appear in *IEEE Trans. Vehicular Technology*.
55. Zheng, G., S. Ma, K. K. Wong, and T. S. Ng. 2010. 'Robust beamforming in cognitive radio'. *IEEE Transactions on Wireless Communications*. 9(2). 570–576.

Matrices

APPENDIX A

Trace of a square matrix

Trace of a square matrix **A,** denoted by trace(**A**)**,** is the summation of its diagonal components.

$$Trace\,(\mathbf{A}) = \sum_{i=j} A_{ij}$$

Frobenius norm of a matrix

Frobenius norm of an N × M matrix is square root of summation of square of its components.

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} A_{ij}^2}$$

Note that $$trace\,(\mathbf{A}\mathbf{A}^H) = trace\left(\mathbf{A}^H\mathbf{A}\right) = \|\mathbf{A}\|^2.$$

$$\text{Determinant}\,(\mathbf{I} + \mathbf{AB}) = \text{Determinant}\,(\mathbf{I} + \mathbf{BA}).$$

Square root of a matrix

Square root of a matrix $\mathbf{A} \geq 0$ whose SVD is

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H$$

is defined as

$$\mathbf{A}^{1/2} = \mathbf{U}\,\mathbf{\Sigma}^{1/2}\,\mathbf{V}^H$$

We can also obtain the square root of matrix from Cholesky decomposition.

Vectorization of a matrix

Let **A** be an N×M matrix and vectorization of a matrix **A** is defined as a vector constructed from stacking all the columns of **A** into a column vector as shown below.

$$\mathbf{A} = \begin{bmatrix} A_{11} & \cdots & A_{1M} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NM} \end{bmatrix} = \begin{bmatrix} \mathbf{c_1} & \cdots & \mathbf{c_M} \end{bmatrix}$$

$$\Rightarrow \qquad \mathbf{a} = vect(\mathbf{A}) = \begin{bmatrix} \mathbf{c}_1^T & \cdots & \mathbf{c}_M^T \end{bmatrix}^T$$

Inner product of two vectors

Inner product of two vectors **x** and **y** is defined as

$$\mathbf{x}^H\mathbf{y} = \sum_{i=1}^{N} x_i^* y_i; \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Rank of a matrix

Rank of a matrix equals the minimum number of linearly independent rows which is always identical to the number of linearly independent columns.

Kronecker product

Let **A** be an N×M matrix and **B** be an L×K matrix. The Kronecker product of **A** and **B** represented by **A** ⊗ **B** is an NL×MK matrix. It can be obtained as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \cdots & A_{1M}\mathbf{B} \\ \vdots & \ddots & \vdots \\ A_{N1}\mathbf{B} & \cdots & A_{NM}\mathbf{B} \end{bmatrix}$$

Some useful identities:

(a) $vect(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vect(\mathbf{B})$

(b) $(\mathbf{A} \otimes \mathbf{B})vect(\mathbf{C}) = vect(\mathbf{BCA}^T)$

(c) $vect(\mathbf{AB}) = vect(\mathbf{ABI}) = (\mathbf{I}^T \otimes \mathbf{A})vect(\mathbf{B})$

(d) $vect(\mathbf{AB}) = vect(\mathbf{IAB}) = (\mathbf{B}^T \otimes \mathbf{I})vect(\mathbf{A})$

(e) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$

(f) $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{B}^H \otimes \mathbf{A}^H$

(g) $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$

(h) $rank(\mathbf{A} \otimes \mathbf{B}) = rank(\mathbf{A})rank(\mathbf{B})$

(i) If $\lambda_\mathbf{A}$ is an eigenvalue of **A** and $\lambda_\mathbf{B}$ is an eigenvalue of **B**, then $\lambda_\mathbf{A}\lambda_\mathbf{B}$ is an eigenvalue of **A** ⊗ **B**.

Block matrix inverse formula

If we partition an arbitrary matrix **A** as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

Then, we can find its inverse as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$$

where,
$$\mathbf{A}^{11} = \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}$$

Relation between eigenvalues and its matrix ***A***

A square matrix **A** has an eigenvector $\mathbf{e}_i$ (usually $\mathbf{e}_i$ is normalized $\left\|\mathbf{e}_i\right\| = 1$) and eigenvalue λ_i if

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i.$$

If rank of matrix **A** is R_A, then there are R_A eigenvalues. Then from spectral theorem, we can write

$$\mathbf{A} = \mathbf{E}\boldsymbol{\lambda}\mathbf{E}$$

where, **E** is unitary matrix ($\mathbf{E}\mathbf{E}^H = \mathbf{E}^H\mathbf{E} = \mathbf{I}$) defined as $\mathbf{E} = [\mathbf{e}_1\mathbf{e}_2 \ldots \mathbf{e}_{R_\mathbf{A}}]$ and $\boldsymbol{\lambda}$ is a diagonal matrix of R_A eigenvalues given as $\boldsymbol{\lambda} = diag\,(\lambda_1, \lambda_2, \ldots \lambda_{R_\mathbf{A}})$.

Assuming $R_A \times R_A$ matrix **A** is full rank, eigenvalues could be obtained from the roots of the characteristic polynomial, $p_{R_\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}_{R_\mathbf{A}})$. For every eigenvalue λ_i the solution of the equation $(\mathbf{A} - \lambda_i\mathbf{I}_{R_\mathbf{A}})\mathbf{x}_i = 0$ gives the eigenvectors.

An orthogonal real-valued matrix has mutually orthogonal columns.

$$\mathbf{c}_i^T\mathbf{c}_j = 0.$$

If entire columns of an orthogonal matrix have unity length, then

$$\mathbf{c}_i^T\mathbf{c}_j = \delta(i, j)$$

Some other properties of eigenvalues are:

- Trace of a square matrix ***A*** is identical to the summation of all its eigenvalues.
- For a full-rank square matrix ***A***, the multiplication of all its eigenvalues is identical to the determinant of the square matrix ***A***.
- If an eigenvalue equals zero, then the matrix is singular {det(***A***)=0}.
- For a rank deficient matrix of rank r, the multiplication of non-zero eigenvalues is identical to the determinant of r×r submatrix of **A**.
- Eigenvectors corresponding to distinct eigenvalues are linearly independent of one another.

QR decomposition of a matrix

QR decomposition (also called the QR factorization) of a matrix is a decomposition of the matrix into an orthogonal matrix **Q** and a triangular matrix **R**. One can find QR decomposition from Gram Schmidt orthogonalization (GSO). Let us quickly review about GSO (J. G. Proakis, 2008). It is basically used for finding a set of orthonormal vectors from a given set of N-dimensional vectors. The procedure is as follows:

Example A.1

Explain Gram Schmidt orthogonalization (GSO).

Solution

Consider the Gram Schmidt procedure for column vectors of a matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_N \end{bmatrix}$$

Then

1. Select the first vector from the set; name it as $\mathbf{v}_1$ and normalize it

$$\mathbf{v}_1 = \mathbf{a}_1, \mathbf{u}_1 = \frac{\mathbf{a}_1}{|\mathbf{a}_1|}$$

2. Select second vector from the set and subtract its projection onto $\mathbf{u}_1$ and normalize

$$\mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{u}_1 \rangle \mathbf{u}_1, \mathbf{u}_2 = \frac{\mathbf{v}_2}{|\mathbf{v}_2|}$$

where the inner product is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{m} x_i y_i$$

3. Select the third vector from the set and subtract its projection onto $\mathbf{u}_1, \mathbf{u}_2$ and normalize it

$$\mathbf{v}_3 = \mathbf{a}_3 - \langle \mathbf{a}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{a}_3, \mathbf{u}_2 \rangle \mathbf{u}_2, \mathbf{u}_3 = \frac{\mathbf{v}_3}{|\mathbf{v}_3|}$$

One can continue this process to construct N orthonormal vectors.

Example A.2

How does one find the QR factorization?

Solution

Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$$

Hence
$$\mathbf{v}_1 = \mathbf{a}_1, \mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{u}_1, \mathbf{a}_2 \rangle \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \frac{\sqrt{2}}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

and
$$\mathbf{u}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Once we have $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, we can find the QR factorization as follows.

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_N \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \end{bmatrix} \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{u}_1 \rangle & \langle \mathbf{a}_2, \mathbf{u}_1 \rangle & \cdots & \langle \mathbf{a}_N, \mathbf{u}_1 \rangle \\ 0 & \langle \mathbf{a}_2, \mathbf{u}_2 \rangle & & \langle \mathbf{a}_N, \mathbf{u}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{a}_N, \mathbf{u}_N \rangle \end{bmatrix} = \mathbf{QR}$$

Note that **Q** is a N×N unitary matrix whose column vectors are orthonormal.

References

1. Choi, J. 2010. *Optimal Combining and Detection*. Cambridge: Cambridge University Press.
2. Kuhn, V. 2006. *Wireless Communications over MIMO channels*. Chichester: Wiley.
3. Strang, G. 2006. *Linear Algebra and its Applications*. New Delhi: Cengage Learning India.
4. Yuen, C., Y. L. Guan, and T. Tjhung. 2007. *Quasi-orthogonal Space-time Block Codes*. London: Imperial College Press.

APPENDIX B

MGF of Hermitian Quadratic Form in Complex Gaussian Variate

Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$ be a complex Gaussian vector $\mathbf{v} \sim N_c^N(\boldsymbol{\mu}_\mathbf{v}, \mathbf{R}_\mathbf{v})$ whose mean, $\boldsymbol{\mu}_\mathbf{v} = E\{\mathbf{v}\}$ and covariance matrix, $(\mathbf{R}_\mathbf{v})$ is defined as $\mathbf{R}_\mathbf{v} = E\{(\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})(\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})^\mathbf{H}\}$.

What is a complex Gaussian vector, **v**?

It is basically a vector whose components $v_1, v_2, \ldots v_N$ are complex Gaussian random variables.

Note $\mathbf{R}_\mathbf{v}$ is Hermitian (since $\mathbf{R}_\mathbf{v}^T = \mathbf{R}_\mathbf{v}^*$) and positive definite.

The pdf of **v** is

$$p(\mathbf{v}) = \frac{1}{\pi^N |\mathbf{R}_\mathbf{v}|} \exp\left[-(\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})^H \mathbf{R}_\mathbf{v}^{-1} (\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})\right]$$

Since the covariance matrix $\mathbf{R}_\mathbf{v}$ is not an identity matrix, complex random vector **v** is not i.i.d. How do we make the covariance matrix identity matrix?

From spectral theorem,

$$\mathbf{R}_\mathbf{v} = \mathbf{U}_R \boldsymbol{\Lambda}^\mathbf{R} \mathbf{U}_\mathbf{R}^\mathbf{H}$$

where, $\boldsymbol{\Lambda}^\mathbf{R}$ is a diagonal matrix with N eigenvalues $\Lambda_i^R, i = 1, \ldots, N$ and $\mathbf{U}_\mathbf{R}$ is a unitary matrix.

Since $\mathbf{R}_\mathbf{v}$ is positive definite, all eigenvalues are positive real, hence,

$$|\mathbf{R}_\mathbf{v}| = |\mathbf{U}_\mathbf{R}||\boldsymbol{\Lambda}^\mathbf{R}||\mathbf{U}_\mathbf{R}^H| = \prod_{i=1}^{N} \Lambda_i^R > 0 .$$

Therefore, we can find the inverse of $\mathbf{R_v}$ as

$$\mathbf{R_v}^{-1} = \mathbf{U_R}\left(\mathbf{\Lambda^R}\right)^{-1}\mathbf{U_R}^{\mathrm{H}}$$

Hence, $$p(\mathbf{v}) = \frac{1}{\pi^{\mathrm{N}}\left|\mathbf{R_v}\right|}\exp[-\{\mathbf{U_R}^{\mathrm{H}}\left(\mathbf{v}-\boldsymbol{\mu}_\mathbf{v}\right)\}^{\mathrm{H}}\left(\mathbf{\Lambda^R}\right)^{-1}\left\{\mathbf{U_R}^{\mathrm{H}}\left(\mathbf{v}-\boldsymbol{\mu}_\mathbf{v}\right)\right\}]$$

Since the eigenvalues are positive real, we can find $\mathbf{\Sigma^R}$ (Hermitian square root of $\mathbf{\Lambda^R}$) such that,

$$\mathbf{\Lambda^R} = \mathbf{\Sigma^R}(\mathbf{\Sigma^R})^H$$

Let us define a new transformation whose covariance matrix is $\mathbf{I_N}$.

$$\mathbf{w} = \left(\mathbf{\Sigma^R}\right)^{-1}\mathbf{U_R}^{\mathrm{H}}\mathbf{v}$$

Hence, $\mathbf{w} \sim \mathrm{N}_C^N\left(\boldsymbol{\mu}_\mathbf{w},\mathbf{I_N}\right)$, $\mathbf{w}$ becomes a i.i.d. random vector and we can find $\mathbf{v}$ from $\mathbf{w}$ as

$$\mathbf{v} = \mathbf{U_R}\mathbf{\Sigma^R}\mathbf{w}$$

Theorem:
For a Gaussian random vector, $\mathbf{v} \sim \mathrm{N_C}^{\mathrm{N}}\left(\boldsymbol{\mu}_\mathbf{v},\mathbf{R_v}\right)$, if $\mathbf{\Sigma^R}$ is an N x N matrix with property $\left(\mathbf{\Sigma^R}\right)\left(\mathbf{\Sigma^R}\right)^{\mathrm{H}} = \mathbf{\Lambda^R}$, then the random vector, $(\mathbf{\Sigma^R})^{-1}(\mathbf{v}-\boldsymbol{\mu}_\mathbf{v})$ is a standard normal random vector. Note that $(\mathbf{v}-\boldsymbol{\mu}_\mathbf{v})$ will have zero mean.

Hence pdf of $\mathbf{w}$ is

$$\mathrm{p}(\mathbf{w}) = \frac{1}{\pi^{\mathrm{N}}}\exp\left[-\left(\mathbf{w}-\boldsymbol{\mu}_\mathbf{w}\right)^{\mathrm{H}}\left(\mathbf{w}-\boldsymbol{\mu}_\mathbf{w}\right)\right] = \prod_{\mathrm{i}=1}^{\mathrm{N}}\frac{1}{\pi}\exp\left[\left(\mathrm{w_i}-\mu_{\mathrm{wi}}\right)^2\right]$$

In other words, $\mathbf{w}$ is i.i.d; therefore, the pdf of $\mathbf{w}$ is simply the multiplication of pdfs of $w_1, w_2, \ldots, w_N$.

Consider the random quadratic form of $\mathbf{A}$ in complex Gaussian multivariate $\mathbf{v}$ as

$$\mathrm{y} = \mathrm{Quad}_\mathbf{A}\left(\mathbf{v}\right) = \mathbf{v^H A v}$$

If $\mathbf{A}$ is Hermitian, $y = y^*$, then Hermitian quadratic form is real.

Hence, $$y = \mathrm{Quad}_\mathbf{A}\left(\mathbf{w}\right) = \mathbf{w}^{\mathrm{H}}\left(\mathbf{\Sigma^R}\right)^{\mathrm{H}}\mathbf{U_R}^{\mathrm{H}}\mathbf{A}\mathbf{U_R}\mathbf{\Sigma^R}\mathbf{w} = \mathbf{w}^{\mathrm{H}}\mathbf{B}\mathbf{w}$$

where, $\mathbf{B} = (\mathbf{\Sigma^R})^{\mathrm{H}}\mathbf{U_R}^{\mathrm{H}}\mathbf{A}\mathbf{U_R}\mathbf{\Sigma^R}$ is obviously Hermitian.

It is Hermitian quadratic form in independent complex Gaussian multivariate $\mathbf{w}$.

We can further simplify this quadratic form by introducing another transformation from $\mathbf{w}$ to $\mathbf{x}$. Note that $\mathbf{B}$ is such that $\mathrm{y} > 0$ for any arbitrary $\mathbf{w}$**,** both $\mathbf{B}$ and its quadratic form is said to be Hermitian positive definite. Since $\mathbf{B}$ is Hermitian, from spectral theorem,

$$\mathbf{B} = \mathbf{U_B}\mathbf{\Lambda^B}\mathbf{U_B}^{\mathrm{H}}$$

where, $\mathbf{U^B}$ is a unitary matrix and $\mathbf{\Lambda^B}$ is a diagonal matrix of eigenvalues $\Lambda_k^{\mathbf{B}}$, $k = 1, \ldots N$. Introducing a new transformation,

$$\mathbf{x} = \mathbf{U_B}^{\mathrm{H}}\mathbf{w} \quad \Rightarrow \mathbf{w} = \mathbf{U_B}\mathbf{x}$$

Since $\mathbf{U_B}$ is unitary (multiplying an i.i.d. complex Gaussian multivariate by a unitary matrix does not change its distribution) and $\mathbf{w} \sim \mathrm{N_C}^{\mathrm{N}}\left(\boldsymbol{\mu}_{\mathbf{w}}, \mathbf{I_N}\right)$; therefore, $\mathbf{x} \sim \mathrm{N_C}^{\mathrm{N}}\left(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{I_N}\right)$. Hence the Hermitian quadratic form in complex Gaussian multivariate $\mathbf{x}$ is

$$y = \mathrm{Quad_A}\left(\mathbf{x}\right) = \mathbf{w}^{\mathrm{H}}\mathbf{U_B}\mathbf{\Lambda^B}\mathbf{U_B}^{\mathbf{H}}\mathbf{w} = \mathbf{x}^{\mathrm{H}}\mathbf{\Lambda^B}\mathbf{x}$$

Note that random vector $\mathbf{x}$ is a (N x 1) column vector. $\mathbf{x}^{\mathrm{H}}$ is a (1 x N) row vector. Hence $\mathbf{x}^{\mathrm{H}}\mathbf{x}$ is a real valued random variable, whereas $\mathbf{x}\mathbf{x}^{\mathrm{H}}$ is an N x N matrix. $\mathbf{\Lambda^B}$ is a diagonal matrix. Therefore,

$$y = \mathrm{Quad}_A\left(\mathbf{x}\right) = \sum_{k=1}^{N}\Lambda_k{}^B\left|x_k\right|^2$$

Hence the name is given as quadratic form.

Note that $\sum_{k=1}^{N}\left|x_k\right|^2$ is distributed as complex non-central Chi-square distributed whose MGF is given as

$$\left(\frac{1}{1-s}\right)^N \exp\left(\frac{s\sum\left|\mu_{xk}\right|^2}{1-s}\right) = \prod_{i=1}^{N}\frac{1}{1-s}\exp\left(\frac{s\left|\mu_{xk}\right|^2}{1-s}\right)$$

Hence the MGF of $\sum_{k=1}^{N}\Lambda_k{}^B\left|x_k\right|^2$ could be written as (replace s by $\Lambda_k^B s$).

Let us try to find an alternate expression of the denominator of the above MGF.

$$\because \qquad \prod_{i=1}^{N}\left(1-\Lambda_k{}^B s\right) = \left|\boldsymbol{I} - s\boldsymbol{B}\right|$$

Also,
$$\mathbf{B} = \left(\mathbf{\Sigma^R}\right)^{\mathrm{H}}\mathbf{U_R}^{\mathrm{H}}\mathbf{A}\mathbf{U_R}\mathbf{\Sigma^R}$$

$$\Rightarrow \qquad \mathbf{A} = \left(\left(\mathbf{\Sigma^R}\right)^{\mathrm{H}}\mathbf{U_R}^{\mathrm{H}}\right)^{-1}\mathbf{B}\left(\mathbf{U_R}\mathbf{\Sigma^R}\right)^{-1}$$

and therefore
$$\mathbf{R_v A} = \mathbf{U_R}\mathbf{\Sigma^R}\left(\mathbf{\Sigma^R}\right)^H\mathbf{U_R}^H\mathbf{U_R}\left(\left(\mathbf{\Sigma^R}\right)^H\right)^{-1}\mathbf{B}\left(\mathbf{\Sigma^R}\right)^{-1}\mathbf{U_R}^H$$

$$= \mathbf{U_R}\mathbf{\Sigma^R}\mathbf{B}\left(\mathbf{\Sigma^R}\right)^{-1}\mathbf{U_R}^{\mathbf{H}}$$

$$\because \qquad \left|\mathbf{U_R}\right| = 1; \left|\mathbf{\Sigma^R}\right|^{-1} = \left|\left(\mathbf{\Sigma^R}\right)^{-1}\right|; \left|\mathbf{R_v}\boldsymbol{A}\right| = \left|\mathbf{B}\right|$$

$$\therefore \qquad \prod_{i=1}^{N}\left(1-\Lambda_k{}^{B}s\right) = \left|\mathbf{I}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right|$$

Also, let us find an alternate form of the numerator of the above MGF

$$\sum_{k=1}^{N}\frac{s\Lambda_k{}^{B}}{1-s\Lambda_k{}^{B}}\left|\mu_{xk}\right|^2 = \sum_{k=1}^{N}\left|\mu_{xk}\right|^2\frac{s\Lambda_k{}^{B}}{1-s\Lambda_k{}^{B}}$$

$$= -\sum_{k=1}^{N}\left|\mu_{xk}\right|^2\left(1-\frac{1}{1-s\Lambda_k{}^{B}}\right)$$

$$= -\left(\boldsymbol{\mu}_x\right)^H\left[\mathbf{I}_N-\left(\mathbf{I}_N-s\boldsymbol{\Lambda}^B\right)^{-1}\right]\left(\boldsymbol{\mu}_x\right)$$

$$= -\left(\boldsymbol{\mu}_{\mathbf{x}}\right)^H\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\left(\boldsymbol{\mu}_{\mathbf{x}}\right)$$

Since, $\mathbf{x} = \mathbf{U_B}^H\mathbf{w} = \mathbf{U_B}^H\left(\boldsymbol{\Sigma}^{\mathbf{R}}\right)^{-1}\mathbf{U_R}^H\mathbf{v}$.

$$\therefore \quad \left(\boldsymbol{\mu}_{\mathbf{x}}\right)^H\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}\left(\boldsymbol{\mu}_{\mathbf{x}}\right)$$

$$= \left[E\left\{\mathbf{U_B}^H\left(\boldsymbol{\Sigma}^{\mathbf{R}}\right)^{-1}\mathbf{U_R}^H\mathbf{v}\right\}\right]^H\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}E\left\{\mathbf{U_B}^H\left(\boldsymbol{\Sigma}^{\mathbf{R}}\right)^{-1}\mathbf{U_R}^H\mathbf{v}\right\}$$

Taking the expectation of the complex multivariate **v** only, we have,

$$\left(\boldsymbol{\mu}_{\mathbf{x}}\right)^H\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}\left(\boldsymbol{\mu}_{\mathbf{x}}\right)$$

$$= \left[E\{\mathbf{v}\}\right]^H\mathbf{U_R}\left[\left(\boldsymbol{\Sigma}^{\mathbf{R}}\right)^{-1}\right]^H\mathbf{U_B}\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}\mathbf{U_B}^H\left(\boldsymbol{\Sigma}^{\mathbf{R}}\right)^{-1}\mathbf{U_R}^H E\{\mathbf{v}\}$$

$$= \left[E\{\mathbf{v}\}\right]^H\mathbf{U_R}\left(\boldsymbol{\Lambda}^{\mathbf{R}}\right)^{-1}\mathbf{U_R}^H\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}\mathbf{E}\{\mathbf{v}\}$$

$$= \left[E\{\mathbf{v}\}\right]^H\left(\mathbf{R}_{\mathbf{v}}\right)^{-1}\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}E\{\mathbf{v}\}$$

$$= \left(\boldsymbol{\mu}_{\mathbf{v}}\right)^H\left(\mathbf{R}_{\mathbf{v}}\right)^{-1}\left\{\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\right\}\left(\boldsymbol{\mu}_{\mathbf{v}}\right)$$

Hence,

$$M_y^{(s)} = \frac{\exp\left[-\left(\boldsymbol{\mu}_{\mathbf{v}}\right)^H\left(\mathbf{R}_{\mathbf{v}}\right)^{-1}\left[\mathbf{I}_{\mathbf{N}}-\left(\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right)^{-1}\right]\left(\boldsymbol{\mu}_{\mathbf{v}}\right)\right]}{\left|\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right|}$$

For zero mean case, ($\mu_{\mathbf{v}} = 0$),

$$M_y^{(s)} = \frac{1}{\left|\mathbf{I}_{\mathbf{N}}-s\mathbf{R}_{\mathbf{v}}\mathbf{A}\right|}$$

We can further simplify $M_y^{(s)} = \dfrac{\exp\left[-s(\boldsymbol{\mu}_\mathbf{v})^H\left[\left(\mathbf{I_N} - s\mathbf{A}\mathbf{R}_\mathbf{v}\right)^{-1}\right]\mathbf{A}(\boldsymbol{\mu}_\mathbf{v})\right]}{\left|\mathbf{I_N} - s\mathbf{R}_\mathbf{v}\mathbf{A}\right|}$ using matrix inversion lemma, $\left(\mathbf{A} + \mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{D}\mathbf{A}^{-1}$. Here $\mathbf{A} = \mathbf{I_N}$, $\mathbf{B} = -s\mathbf{I_N}$, $\mathbf{C} = \mathbf{R_v}$ and $\mathbf{D} = \mathbf{A}$.

References

1. Janaswamy, R. 2001. *Radiowave Propagation and Smart Antennas for Wireless Communications*. New York: Kluwer Academic Publishers.
2. Schwartz, M., W. R. Bennett, and S. Stein. 1996. *Communication Systems and Techniques*. New York: IEEE Press.
3. Turin, G. L. June 1960. 'The characteristic function of Hermitian quadratic forms in complex normal variates'. *Biometrika*. 47. 199–201.

Basics of Information Theory

APPENDIX C

Shannon Information Content:

For an outcome with probability p, the Shannon information content (SIC) is defined as

$$-\log_2(p) = -\frac{\ln(p)}{\ln(2)}.$$

Example C.1

(a) The outcomes of tossing a coin are either head or tail with equal probability p = $\frac{1}{2}$ and their SIC equals 1 bit.

(b) Is it your birthday? There are two possible answers yes or no with probabilities $\frac{1}{365}$ and $\frac{364}{365}$ and their SIC are 8.512 and 0.004 bits, respectively.

An important observation is that unlikely outcomes give more information.

Entropy:

H(X) is defined as the average SIC of a RV *X*.

$$H(X) = E(-\log_2(p_X(x)))$$

where, *E* denotes the expectation operator.

Note that entropy depends only on $p_X(x)$ not on *X*. Also the $\log_2$ in the entropy expression transforms the multiplications in probabilities to additions in entropies.

Example C.2

(a) If you throw a dice, the possible outcomes are $X = \{1,2,3,4,5,6\}$ with probabilities $p_X(x) = \left\{\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right\}$. $p_X(x)$ is also known as probability mass function of *X*.

If *g(x)* is a function defined on a discrete RV *X*, we have,

$$E(g(x)) = \sum_{x \in X} p_X(x)\, g(x)$$

Therefore, $E(x) = \mu = 3.5; E(x^2) = \sigma^2 + \mu^2 = 15.17; E(-\log_2(p_X(x))) = 2.58 = H(X)$.

(b) Bernoulli RV

For a Bernoulli RV, the possible outcomes are $X = \{0,1\}$ with corresponding probabilities $p_X(x) = \{1-p, \quad p\}$.

Hence the entropy, $H(X) = E(-\log_2(p_X(x))) = -p\log_2 p - (1-p)\log_2(1-p) = H(p)$.

H(p) is purely a concave function. It is maximum when p = $\frac{1}{2}$ (supreme uncertainty) and it is zero for p = 1 or 0 (uncertainty is minimum). Entropy has a key role in information theory.

Mutual Information

It is defined as the decrement in the uncertainty (entropy) of *X* because of knowledge of *Y*. Mathematically,

$$\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= E(-\log_2(p(X))) - E(-\log_2(p(X/Y))) \\
&= E(-\log_2(p(X))) - E\left(-\log_2\left(\frac{p(X,Y)}{P(Y)}\right)\right) \\
&= E\left(-\log_2\left(\frac{P(X)P(Y)}{p(X,Y)}\right)\right) \\
&= H(X) + H(Y) - H(X,Y) \\
&= I(Y,X) \\
&= H(Y) - H(Y/X)
\end{aligned}$$

Note that conditioning cuts down entropy.

$$0 \le I(X;Y) = H(Y) - H(Y/X)$$

$$\Rightarrow \qquad H(Y/X) \le H(Y)$$

The equality is possible for independent *Y* and *X*.

Convex and Concave functions

Definition: *f(x)* is strictly convex over *(a,b)* if

$$f(\lambda u + (1-\lambda)v) < \lambda f(u) + (1-\lambda)f(v) \forall u \neq v \in (a,b), 0 < \lambda < 1$$

In other words, each chord in *f(x)* lies above *f(x)*. For convex function *f(x)*, *-f(x)* is a concave function.

Example C.3

x^2, x^4, e^x and $x\log(x)$ ($x \geq 0$) are strictly convex function.

$\log(x)$, $\sqrt{x}$ are strictly concave function.

x is a concave and convex function.

Jensen's inequality

(a) For convex function $f(x)$, $E(f(X)) \geq f(E(X))$.
(b) For strictly convex function $f(x)$, $E(f(X)) > f(E(X))$.

We can prove this by applying the method of induction.

Proof:

For a two-mass point distribution (from the definition of convex function),

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)$$

For a k-mass point distribution, expectation of *f(X)* is expressed as

$$E(f(X)) = \sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k) \sum_{i=1}^{k-1} \frac{p_i}{1-p_k} f(x_i)$$

$$\geq p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} \frac{p_i}{1-p_k}(x_i)\right)$$

In the above, we have assumed that the Jensen's inequality is satisfied for k-1 mass point distribution. Now, from the convexity of a two mass point distribution, we have,

$$E(f(X)) \geq f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} \frac{p_i}{1-p_k}(x_i)\right) = f(E(X))$$

Hence proved.

Example C.4

$f(x) = x^2$ is a strictly convex function.

Let us say the possible outcomes are $X=\{-1,+1\}$ with equal probabilities of p = {1/2,1/2}.

Then $E(X) = 0, f(E(X)) = 0$,

but, $E(f(X)) = 1$.

Hence, $E(f(X)) > f(E(X))$.

Differential entropy

Differential entropy h(X) is expressed as

$$h(X) = -\int_{-\infty}^{\infty} f_X(x)\log_2\left(f_X(x)\right)dx = E\left(-\log_2\left(f_X(x)\right)\right)$$

Example C.5

Complex Multidimensional Gaussian R.V.

Assume an N-dimensional complex Gaussian distributed vector, **x**. Every element of **x** is composed of an in-phase component x_I^k and a quadrature component x_Q^k so

$$x^k = x_I^k + jx_Q^k, \text{ where k=1,2,\ldots,N}$$

Or vectorially, $\mathbf{x} = \mathbf{x}_I + j\mathbf{x}_Q$

It is presumed that **x** has zero mean. We need to find the differential entropy of **x**. For orthogonal $\mathbf{x}_I$ and $\mathbf{x}_Q$, $E\left[\mathbf{x}_I\mathbf{x}_Q^T\right] = 0$. Besides they are statistically independent if both in-phase and quadrature components are Gaussian distributed. Mathematically,

$$f_{\mathbf{x}_I,\mathbf{x}_q}\left(x_\mathbf{I}, x_\mathbf{Q}\right) = f_{\mathbf{x}_I}\left(x_\mathbf{I}\right)f_{\mathbf{x}_q}\left(x_\mathbf{Q}\right)$$

Both $\mathbf{x}_I$ and $\mathbf{x}_Q$ has identical formula for their joint pdfs. Hence, we have two significant findings:

1. The components $\mathbf{x}_I$ and $\mathbf{x}_Q$ have identical entropy ($h(\mathbf{x}_I) = h(\mathbf{x}_Q)$).
2. Differential entropies of $\mathbf{x}_I$ and $\mathbf{x}_I$ being logarithmic, will add up ($h(\mathbf{x}) = h(\mathbf{x}_I) + h(\mathbf{x}_Q) = 2h(\mathbf{x}_I)$).

The joint pdf of the complex Gaussian vector **x** with zero mean and $\mathbf{R}_\mathbf{x}$ correlation matrix is expressed as,

$$f_\mathbf{x}(\mathbf{x}) = \frac{1}{(2\pi)^N\left|\mathbf{R}_\mathbf{x}\right|}\exp\left\{-\frac{1}{2}\mathbf{x}^T\left(\mathbf{R}_\mathbf{x}\right)^{-1}\mathbf{x}\right\}.$$

Therefore (see information inequality which will be discussed later for proving the following equation),

$$h(\mathbf{x}) = \log_2\left|2\pi e\mathbf{R}_\mathbf{x}\right| = N + N\log_2(2\pi) + \log_2\left|\mathbf{R}_\mathbf{x}\right| \text{ bits.}$$

For scalar complex Gaussian r.v. x, N=1 this simplifies to $h(x) = 1 + \log_2\left(2\pi\sigma_x^2\right)$ bits. If x is real, $h(x) = \dfrac{1 + \log_2\left(2\pi\sigma_x^2\right)}{2}$ bits.

Kullback–Leibler distance

Relative differential entropy of two pdf f and g is expressed as

$$D(f \parallel g) = \int f(x)\log_2\left(\frac{f(x)}{g(x)}\right)dx = -h_f(X) - E_f\left(\log_2\left(g(X)\right)\right)$$

Information inequality

$$D(f \parallel g) \geq 0$$

Proof:

Define $S = \{x : f(x) > 0\}$.

$$-D(f \parallel g) = -\int f(x)\log_2\left(\frac{g(x)}{f(x)}\right)dx = E_f\left(\log_2\left(\frac{g(X)}{f(X)}\right)\right) \leq \log_2\left(E_f\left(\frac{g(X)}{f(X)}\right)\right)$$

In the above, we have used Jensen's inequality and log being a concave function.

$$-D(f \parallel g) \leq \log_2 \int_S \left(f(x)\frac{g(x)}{f(x)}\right)dx = \log_2 \int_S g(x)\,dx = \log_2(1) = 0$$

Hence proved.

Example C.6

(a) Find the entropy maximizing distribution over the interval *(a,b)*.

Answer

Assume $f(x)$ is a distribution over the interval $x \in (a, b)$.

We have, $0 \leq D(f \parallel u) = -h_f(X) - E_f\left(\log_2(u(X))\right) = -h_f(X) + \log_2(b-a)$

$\Rightarrow$ $h_f(X) \leq \log_2(b-a)$

A uniform distribution ($u(x) = \dfrac{1}{b-a}$) maximizes entropy over the interval $x \in (a, b)$.

(b) For a given covariance matrix **K,** find the zero mean entropy maximizing distribution over the infinite interval $(-\infty, \infty)^n$.

Answer

A multivariate Gaussian distribution with the pdf $\varphi(\mathbf{x}) = |2\pi\mathbf{K}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right)$.

Proof:

$$0 \leq D(f \parallel u) = -h_f(\mathbf{X}) - E_f\left(\log_2(\varphi(\mathbf{X}))\right)$$

$$\Rightarrow \quad h_f(\mathbf{X}) \leq E_f\left(\log_2(\varphi(\mathbf{X}))\right)$$

$$= -(\log_2 e)E_f\left(-\frac{1}{2}\ln|2\pi\mathbf{K}| - \frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right)$$

$$= \frac{1}{2}(\log_2 e)E_f\left(\ln|2\pi\mathbf{K}| + \mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right)$$

$$= \frac{1}{2}(\log_2 e)\left[E_f\left(\ln|2\pi\mathbf{K}|\right) + E_f\left(\sum_{i,j} x_i\left(\mathbf{K}^{-1}\right)_{ij} x_j\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + E_f\left(\sum_{i,j} x_i x_j\left(\mathbf{K}^{-1}\right)_{ij}\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + \left(\sum_{i,j} E_f\left(x_j x_i\right)\left(\mathbf{K}^{-1}\right)_{ij}\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + \left(\sum_{i,j}(\mathbf{K})_{ji}\left(\mathbf{K}^{-1}\right)_{ij}\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + \left(\sum_{i,j}\left(\mathbf{K}\mathbf{K}^{-1}\right)_{jj}\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + \left(\sum_{i,j}(\mathbf{I})_{jj}\right)\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi\mathbf{K}|\right) + n\right]$$

$$= \frac{1}{2}(\log_2 e)\left[\left(\ln|2\pi e\mathbf{K}|\right)\right]$$

$$= \frac{1}{2}\left[\left(\log_2|2\pi e\mathbf{K}|\right)\right]$$

$$= h_\varphi(\mathbf{X})$$

This proof (RHS) also gives the differential entropy of a multivariate normal distribution.

Capacity of a parallel Gaussian channel

Let us consider n-independent Gaussian channel with I–O relation for the ith channel as

$$Y_i = X_i + N_i$$

where, N_i are zero mean Gaussian i.i.d. $Z_i \sim N\left(0, \sigma^2\right)$ with the power constraint $n^{-1}\sum_{i=1}^{n} E\left(X_i^2\right) \le P$.

Let us analyze and find the capacity of the i^{th} Gaussian channel first.

Assumption: X_i and N_i are independent RV with zero mean.

$$E\left(Y_i^2\right) = E\left(X_i^2 + N_i^2 + 2X_iN_i\right) = P + \sigma^2$$

We may express information capacity as

$$C = \max_{E\left(X_i^2\right) \le P} I\left(X_i; Y_i\right)$$

$$\because \quad \begin{aligned} I(X_i;Y_i) &= h(Y_i) - h(Y_i \mid X_i) \\ &= h(Y_i) - h(X_i + N_i \mid X_i) \\ &= h(Y_i) - h(N_i \mid X_i) \\ &= h(Y_i) - h(N_i) \end{aligned}$$

We know that optimal input is for Gaussian distributed X_i and the noise N_i is also Gaussian distributed.

Hence, $I(X_i;Y_i) \leq \frac{1}{2}\log_2\left(2\pi e\left(P + \sigma^2\right)\right) - \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right)$

The I–O relation can be represented in the vector form as

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}$$

Let us find the capacity for this case.

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= h(\mathbf{Y}) - h(\mathbf{Y} \mid \mathbf{X}) \\ &= h(\mathbf{Y}) - h(\mathbf{X} + \mathbf{N} \mid \mathbf{X}) \\ &= h(\mathbf{Y}) - h(\mathbf{N} \mid \mathbf{X}) \\ &= h(\mathbf{Y}) - h(\mathbf{N}) \end{aligned}$$

Note that mutual information is maximum when N_i and Y_i are i.i.d. Gaussian with zero mean.

Hence, $I(\mathbf{X};\mathbf{Y}) \leq C = \sum_{i=1}^{n} h(Y_i) - h(N_i) = \sum_{i=1}^{n} \frac{1}{2}\log_2\left(1 + \frac{P_i}{\sigma_i^2}\right)$

How to allocate power to each channel so that capacity is maximized?

This can be done with the help of Waterfilling algorithm. The idea is to put more power to least noisy channels such that power plus noise in each channel are equal.

References

1. Choi, J. 2010. *Optimal Combining and Detection*. Cambridge: Cambridge University Press.
2. Cover, T. M. and J. A. Thomas. 2006. *Elements of Information Theory*. New Delhi: Wiley.
3. Haykin, S. and M. Moher. 2005. *Modern Wireless Communications*. New Delhi: Pearson.

Basics of Convolutional Codes

APPENDIX D

Convolutional encoder

In convolutional code (B. P. Lathi, 2009; S. G. Wilson, 1996; E. Biglieri, 2005; T. Oberg, 2001), the block of n code bits produced by the encoder in a particular moment is dependent on

- the block of k message bits in that particular moment and
- the block of data bits for N-1 moments (N > 1) before.

A convolutional code with constraint length N consists of an N-stage shift register (SR) and ν modulo-2 adders.

Figure D.1 shows such a coder for the case N = 3 and $\nu = 2$. The output samples the ν modulo-2 adders in a sequence, once during each input-bit interval.

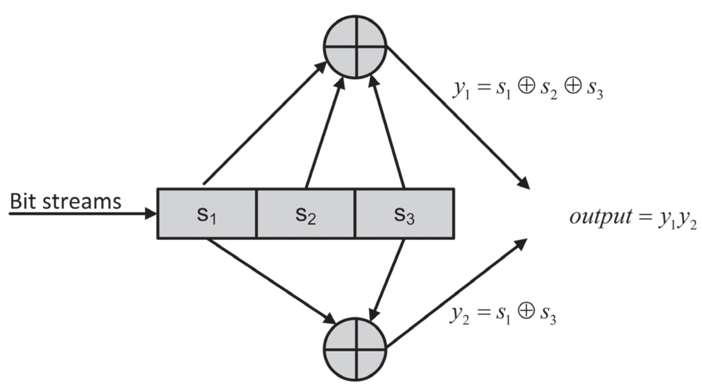


Fig. D.1 Convolutional coder

Example D.1

Assume that the input digits are 0101. Find the coded sequence output.

Solution

Initially, the SRs $s_1 = s_2 = s_3 = 0$.

Note that SR just shifts the input data to the next SR in the next time instant.

When the first message bit 1 enters the SR, $s_1 = 1$, $s_2 = s_3 = 0$.

Then $y_1 = y_2 = 1$.

The coder output is 11.

When the second message bit 0 enters the SR, $s_1 = 0$, $s_2 = 1$, $s_3 = 0$.

Then $y_1 = 1$, $y_2 = 0$.

The coder output is 10.

When the third message bit 1 enters the SR, $s_1 = 1$, $s_2 = 0$, $s_3 = 1$.

Then $y_1 = y_2 = 0$.

The coder output is 00.

When the fourth message bit 0 enters the SR, $s_1 = 0$, $s_2 = 1$, $s_3 = 0$.

Then $y_1 = 1$, $y_2 = 0$.

The coder output is 10.

In order to stop, one input N-1 0s and ensures that 0 passes all the way through the SR.

Therefore, for the 0101 input digits, one eventually input 000101 to the SR.

The coder output is 001101000111.

There are n = (N+k-1) ν digits in the coded output for each k input bits.

State diagram

When a message bit enters the SR (s_1) the coder outputs are dependent on both the message bit in s_1 and the two past bits already in s_3 and s_2. There are four feasible combinations of the two past bits in s_3 and s_2: 00, 01,10,11. We will name these four states as a, b, c, d respectively as depicted in Fig. D.2 (a). The count of states is 2^{N-1}. Based on the encoder state, a message bit 0 or 1 produces four different outputs. The overall behavior can be briefly shown using the state diagram of Fig. D.2 (b). This is a fourfold directed graph employed to show the input–output relationship of encoder. Convention: we will adopt solid lines for 0 input bit, and dashed lines for 1 input bit.

Interpretations from State diagram D.2 (b):

1. State a goes to State a for 0 input and 00 output
2. State a goes to State b for 1 input and 11 output
3. State b goes to State c for 0 input and 10 output
4. State b goes to State d for 1 input and 01 output
5. State c goes to State a for 0 input and 11 output
6. State c goes to State b for 1 input and 00 output
7. State d goes to State c for 0 input and 01 output
8. State d goes to State d for 1 input and 10 output

Note that the encoder cannot go directly from state a to state c or d. *From any given state, the encoder can go to only two states directly by inputting a single message bit.*

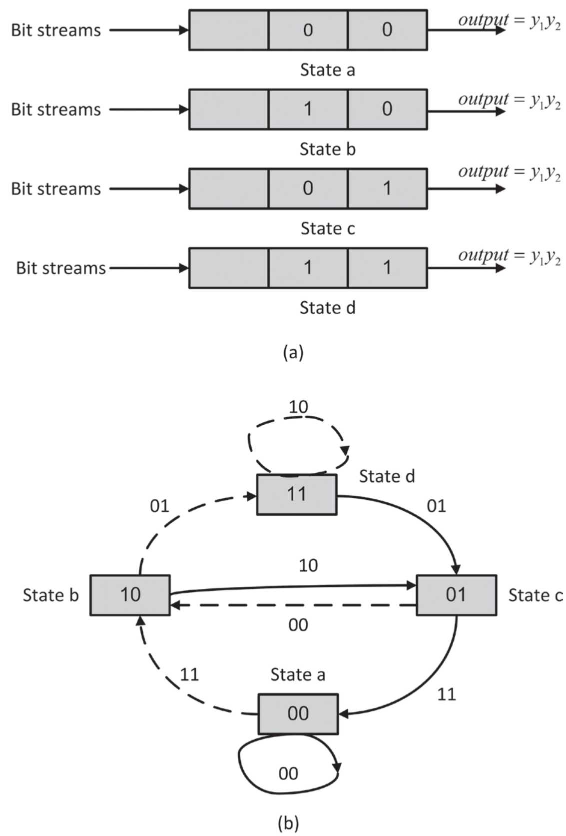


Fig. D.2 (a) 3-shift registers showing states *a*, *b*, *c* and *d* (b) State diagram for the coder

Trellis diagram

Trellis diagram may be easily drawn from the above state diagram. It commences from entire 0s in the SR, i.e., state a and makes transitions depending on every input data digit. Such changes are represented by a solid line (ensuing data digit 0) and by a dashed line (ensuing data digit 1). Hence for first input digit 0, the encoded output is 00 (solid line) and for input digit 1, the encoded output is 11 (dashed line). We continue in such a way for the second input digit and so on as depicted in Fig. D.3.

Decoding

We shall consider maximum-likelihood (ML) decoding based Viterbi's algorithm. Out of all decoders for convolutional codes, Viterbi's ML algorithm is an optimal technique. As usual, ML receiver selects a codeword nearest to the received codeword. Since there are 2^k codewords (k input data digits), the ML decoder needs to store 2^k codewords and compares with the received codeword. For large k, there is exponential increase in complexity of the decoder. Viterbi simplify such ML calculation by observing that every four nodes (a, b, c and d) has only two antecessors. Note that every node can be checked in from two nodes only. Besides, only the path that matches most with the received sequence (the minimum distance path) requires to be kept for each node. For a received bit stream, it is necessary to determine a path in the Trellis diagram with the output bit stream which matches most with the received stream.

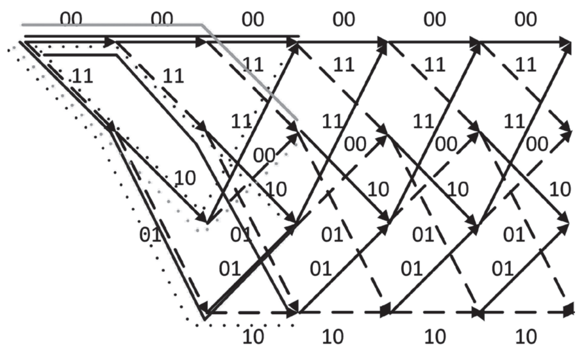


Fig. D.3 Survivor paths after the 3rd branch of the Trellis diagram for received sequence 01 00 01

Example D.2

Assume that the initial six received digits are 01 00 01. Find the survivor paths (minimum-distance path with the received sequence).

Solution

Table D.1

Survivor paths after the 3rd branch of the Trellis diagram for received sequence 01 00 01

After 3rd branches	*Paths*	*Distance with received sequence*	*Survivor?*
Node a	00 00 00	2	Yes
	11 10 11	3	
Node b	00 00 11	2	Yes
	11 10 00	3	
Node c	00 11 10	5	
	11 01 01	2	Yes
Node d	00 11 01	3	Yes
	11 01 10	4	

With four paths eliminated as illustrated in Table D.1, the four survivor paths are the only contestants. It is necessary for us to memorize the four survivor paths and their distances from the received sequences. Usually, the count of survivor paths is identical to the count of states, that is 2^{N-1}. As soon as we have survivors at all the third-level nodes, we observe at the ensuing two received digits.

To truncate the Viterbi algorithm and ultimately we need to resolve on single path rather than four. This is made possible for 00 given to the last two input bits. When the first 0 is input to the SR, we look for the survivors at nodes a and c only. The survivors at nodes b and d are discarded because these nodes can be checked in only when input bit is 1, as inferred from the trellis diagram. When the second 0 enters the SR, we scrutinize survivor at node a. We discard the survivor at node c since the last two bits 00 leads to the encoder state a. It implies that the count of states is diminished from four to two (a and c) by the first zero input and to one state (a) by the second zero input to the SR.

For Viterbi algorithm, storage and computational complexity reduces considerably (proportional to 2N) and are widely used for constraint length N<10.

References

1. Biglieri, E. 2005. *Coding for Wireless Channels*. London: Springer.
2. Lathi, B. P. 2009. *Modern Digital and Analog Communication Systems*. New York: Oxford University Press.
3. Oberg, T. 2001. *Modulation, Detection and Coding*. Chichester: John Wiley & Sons.
4. Wilson, S. G. 1996. *Digital Modulation and Coding*. New Delhi: Pearson.

Basics of Turbo Codes

APPENDIX E

Encoder

Turbo codes are basically parallel concatenation of 2 systematic recursive convolutional codes (SRCC). Length k-message **u** encoded by the 1st encoder produces parity bits $\mathbf{p}^{(1)}$. Interleaved **u**, i.e., $\Pi(\mathbf{u})$ encoded by the 2nd encoder produces parity bits $\mathbf{p}^{(2)}$. Systematic codes transmit the bits in **u,** $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ over the channel as depicted in Fig. E.1.

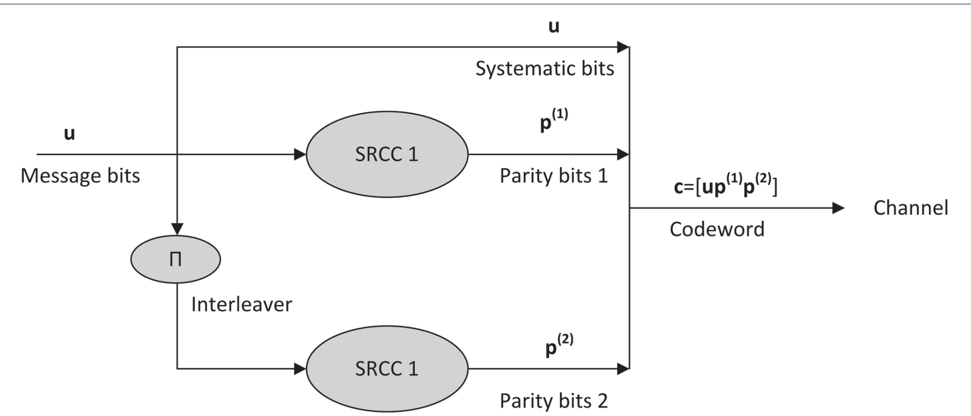


Fig. E.1 Turbo code encoder

Interleaver:

It is represented by a permutation sequence $\Pi = [\Pi_1, \Pi_2, \ldots, \Pi_n]$ where the sequences $\Pi = [\Pi_1, \Pi_2, \ldots, \Pi_n]$ are a permutation of the integers 1–n. Interleaver ensures that parity bits of encoder 2 is completely different than encoder 1. If low-weight parity sequence for encoder 1 then high-weight parity sequence for encoder 2, avoids low-weight turbo codewords.

Example E.1

(a) $\prod$ = [4 2 5 3 1] acting on the input vector $\mathbf{u} = [u_1 \quad u_2 \quad u_3 \quad u_4 \quad u_5]$ will produce $\Pi(\mathbf{u}) = [u_4 \quad u_2 \quad u_5 \quad u_3 \quad u_1]$.

(b) $\mathbf{u}$ = [1 0 1 1 0]→ $\prod(\mathbf{u})$=[1 0 0 1 1]

Code rate

Code rate is $\dfrac{k}{k + n_1 - k + n_2 - k}$ where code rates for encoder 1 and encoder 2 are chosen as $\dfrac{k}{n_1}$ and $\dfrac{k}{n_2}$ respectively.

Puncturing

To increase code rate, we may puncture the output of one or both convolutional codes.

(a) For example, encoding message bits [0 1 0 1 0 0] with an encoder produces the two codeword bits, $C_1 = [0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0]$; $C_2 = [0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0]$. Present code rate = $\dfrac{6}{12} = \dfrac{1}{2}$. Puncturing every third bit in codeword 2 will produce $C_2 = [0\ \ 1\ \ \times\ \ 0\ \ 1\ \ \times]$ where x indicates that the corresponding bit is not transmitted. That means, for every 6 message bits, there are only 10 codeword bits, which means the code rate = $\dfrac{6}{10} = \dfrac{3}{5}$.

The puncturing pattern is specified by puncturing matrix P. For encoder with n output bits the matrix P will have n rows one for each output stream. The zero entry in the third column of the second row indicates that every third bit in the output C_2 is to be punctured.

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

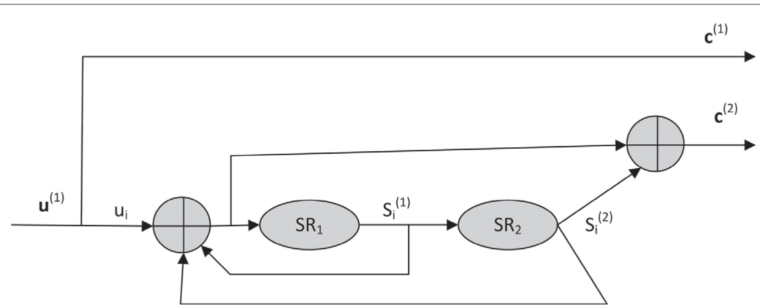


Fig. E.2 A SRCC encoder

Systematic recursive convolutional code (SRCC)

Note there is a feedback loop in encoder diagram unlike convolutional codes we have discussed in Appendix D; hence, the name recursive. So we can modify the generator polynomial for SRCC accordingly. Besides, the first part of the code $\mathbf{c}^{(1)}$ is the message bit itself; so the name systematic.

Example E.2

(a) Consider SRCC shown in Fig. E.2. Find the generator matrix of this SRCC.

Solution

Looking at the SRCC encoder diagram, the generator matrix will be

$$\mathbf{G} = \left[1 \quad \frac{1 + D^2}{1 + D + D^2} \right]$$

1 in the generator matrix means we are sending the first part of the codeword same as the message bit. In the feedforward loop only the outputs of SR_2 and input message bit to SR_1 mod-2 summation is done that's why the generator polynomial numerator is $1 + D^2$. In the feedback loop all the outputs of the SR_1 and SR_2, and input to the SR_1 are all connected; hence, the generator polynomial denominator is $1 + D + D^2$.

(b) Consider an input message bit of 100, find the output of the code.

Solution

At time t = 1,

First part of the code is equal to the input message bit at that time.

$$c_1^{(1)} = u_1 = 1$$

Second part of the code can be calculated as

$$c_1^{(2)} = \left(u_1 \oplus s_1^{(1)} \oplus s_1^{(2)}\right) \oplus s_1^{(2)} = 1$$

Note that subscript denote the time.

At time t = 2,

First part of the code is equal to the input message bit at that time.

$$c_2^{(1)} = u_2 = 0$$

Second part of the code can be calculated as

$$c_2^{(2)} = \left(u_2 \oplus s_2^{(1)} \oplus s_2^{(2)}\right) \oplus s_2^{(2)}$$

We need to find the states of the SR_1 and SR_2 first.

$$s_2^{(1)} = \left(u_1 \oplus s_1^{(1)} \oplus s_1^{(2)}\right) = 1$$

$$s_2^{(2)} = s_1^{(1)} = 0$$

Hence, $c_2^{(2)} = \left(u_2 \oplus s_2^{(1)} \oplus s_2^{(2)}\right) \oplus s_2^{(2)} = 1$

At time t = 3,

First part of the code is equal to the input message bit at that time.

$$c_3^{(1)} = u_3 = 0$$

Second part of the code can be calculated as

$$c_3^{(2)} = \left(u_3 \oplus s_3^{(1)} \oplus s_3^{(2)}\right) \oplus s_3^{(2)}$$

We need to find the states of the SR_1 and SR_2 first.

$$s_3^{(1)} = \left(u_2 \oplus s_2^{(1)} \oplus s_2^{(2)}\right) = 1$$

$$s_3^{(2)} = s_2^{(1)} = 1$$

Hence, $c_3^{(2)} = \left(u_3 \oplus s_3^{(1)} \oplus s_3^{(2)}\right) \oplus s_3^{(2)} = 1$

The final code word for the input message bit 100 is

$\mathbf{C} = \left[c_1^{(1)} \quad c_1^{(2)}; \quad c_2^{(1)} \quad c_2^{(2)}; \quad c_3^{(1)} \quad c_3^{(2)};\right] = [1 \quad 1; \quad 0 \quad 1; \quad 0 \quad 1;]$

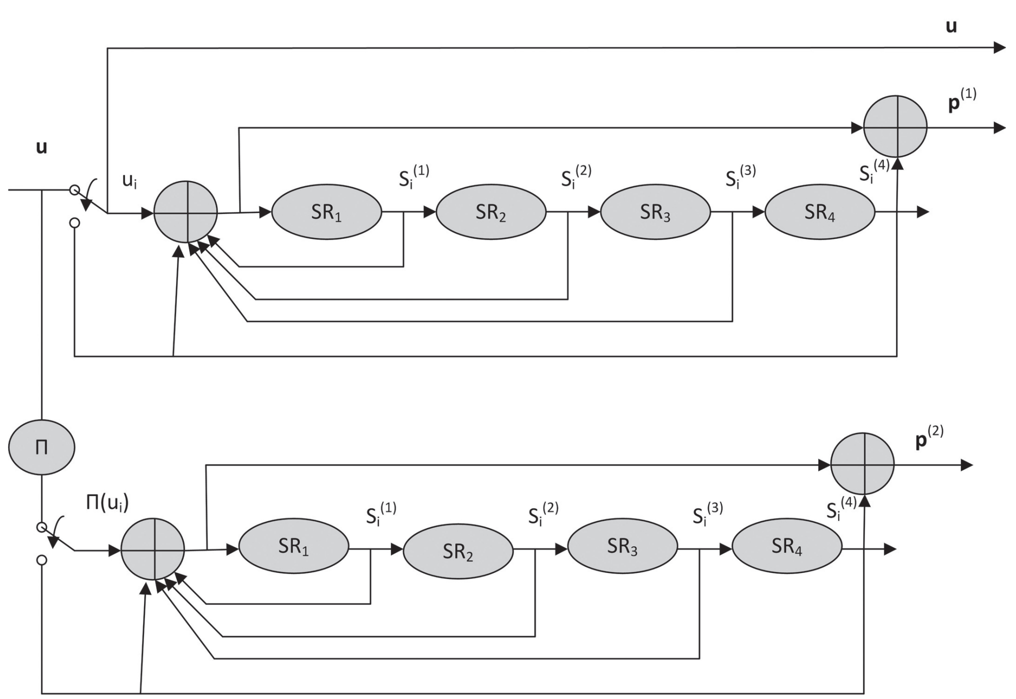


Fig. E.3 Berrou, Glavieux and Thitimajshima turbo code encoder

$\frac{1}{2}$ rate turbo code by Berrou, Glavieux and Thitimajshima (C. Berrou et al., 1993)

It uses the same encoder 1 and 2: rate $\frac{1}{2}$ SRCC shown in Fig. E.3. Hence it will produce rate $\frac{1}{3}$ turbo code. But we can increase the code rate to $\frac{1}{2}$ by puncturing both the SRCC encoders. Puncturing matrix for the encoder 1 and 2 are as follows.

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}; \mathbf{P}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

The generator matrix for the SRCC depicted in Fig. E.3.

$$\mathbf{G} = \left[1 \quad \frac{1 + D^4}{1 + D + D^2 + D^3 + D^4} \right]$$

Interleaver used is $\Pi = [3, 7, 6, 2, 5, 10, 1, 8, 9, 4]$

Assume a message bit $\mathbf{u}$ = [1 0 1 0 1 0 1 0 1 0] is entering the turbo code encoder. Let us find out what is the output codeword.

The input bit to SRCC 1 is $\mathbf{u}$.

The bit interleaved input for the SRCC 2 is given by

$$\mathbf{V} = \prod(\mathbf{u}) = [1\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 0]$$

At time t = 1,

First part of the code is equal to the input message bit at that time.

$$c_1^{(1)} = u_1 = 1$$

Second part of the code (for SRCC 1) can be calculated as

$$c_1^{(2)} = p_1^{(1)} = \left(u_1 \oplus s_1^{(1)} \oplus s_1^{(2)} \oplus s_1^{(3)} \oplus s_1^{(4)}\right) \oplus s_1^{(4)} = 1$$

Note that subscript denote the time.

Also note that $v_1 = 1$

Third part of the code (SRCC 2) can be calculated as

$$c_1^{(3)} = p_1^{(2)} = \left(v_1 \oplus s_1^{(1)} \oplus s_1^{(2)} \oplus s_1^{(3)} \oplus s_1^{(4)}\right) \oplus s_1^{(4)} = 1$$

At time t = 2,

First part of the code is equal to the input message bit at that time.

$$c_2^{(1)} = u_2 = 0$$

Second part of the code (for SRCC 1) can be calculated as

$$c_2^{(2)} = p_2^{(1)} = \left(u_2 \oplus s_2^{(1)} \oplus s_2^{(2)} + s_2^{(3)} + s_2^{(4)}\right) \oplus s_2^{(4)}$$

We need to find the SRCC 1 states of the SR_1, SR_2, SR_3 and SR_4 first.

$$s_2^{(1)} = \left(u_1 \oplus s_1^{(1)} \oplus s_1^{(2)} \oplus s_1^{(3)} \oplus s_1^{(4)}\right) = 1$$

$$s_2^{(2)} = s_1^{(1)} = 1$$

$$s_2^{(3)} = s_2^{(4)} = 0$$

Hence, $c_2^{(2)} = p_2^{(1)} = 0$.

Note that $v_2 = 1.$

Third part of the code (SRCC 2) can be calculated as

$$c_2^{(3)} = p_2^{(2)} = \left(v_2 \oplus s_2^{(1)} \oplus s_2^{(2)} + s_2^{(3)} + s_2^{(4)}\right) \oplus s_2^{(4)}$$

We need to find the SRCC 2 states of the SR_1, SR_2, SR_3 and SR_4 first.

$$s_2^{(1)} = \left(v_1 \oplus s_1^{(1)} \oplus s_1^{(2)} \oplus s_1^{(3)} \oplus s_1^{(4)}\right) = 1$$

$$s_2^{(2)} = s_1^{(1)} = 1$$

$$s_2^{(3)} = s_2^{(4)} = 0$$

Hence, $c_2^{(3)} = p_2^{(1)} = 1$.

Hence the output code from SRCC 1.

$$\mathbf{u} = [10\cdots]$$

$$\mathbf{p}^{(1)} = [10\cdots]$$

For puncturing matrix, $\mathbf{P}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, we have,

We will send **u** as it is, therefore, $\mathbf{u} = [10\cdots]$.

We will send 1st, 3rd, 5th, 7th and 9th bits from parity bit matrix 1, therefore, $\mathbf{p}^{(1)} = [1\cdots]$.

Hence, the output code from SRCC 2.

$$\mathbf{v} = [11\cdots]$$

$$\mathbf{p}^{(2)} = [11\cdots]$$

For puncturing matrix, $\mathbf{P}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, we have,

We will not send any bit from **v**.

We will send 2nd, 4th, 6th, 8th and 10th bits from parity bit matrix 1, therefore, $\mathbf{p}^{(1)} = [1\cdots]$.

We can observe that for 2 input bits, we are sending the same 2 input bits and 2 parity bits. The code rate is $\frac{1}{2}$.

References

1. Berrou, C., A. Glavieux, and P. Thitimajshima. 1993. 'Near Shannon limit error-correcting coding and decoding: turbo-codes'. In *Proc. International Conference on Communications*. Geneva, Switzerland. 1064–1070.
2. Haykin, S. and M. Moher. 2010. *Communication Systems*. New Delhi: John Wiley & Sons.
3. Johnson, S. J. 2010. *Iterative Error Corrections*. Cambridge: Cambridge University Press.

APPENDIX F

Algebraic Structures

Groups, rings and fields are common algebraic structures. Let Z be the set of integers; $Z = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}$. Q is the set of rational numbers; $Q = \left\{ \frac{a}{b} \mid a, b \in Z, b \neq 0 \right\}$. R is the set of real numbers and C is the set of complex numbers.

Groups

A Group (G) is a set of elements with a binary operation "•" that obeys the succeeding four properties (or axioms) and an optional fifth property (B. A. Forouzan, 2007).

- Closure: $\forall a, b \in G, a \bullet b \in G$
- Associativity: $\forall a, b, c \in G, (a \bullet b) \bullet c = a \bullet (b \bullet c)$
- Existence of identity: $\forall a \in G, \exists e, a \bullet e = e \bullet a = a$
- Existence of an inverse:
  $\forall a \in G, \exists -a, a - a = -a + a = 0$ or $\forall a \in G, \exists a^{-1}, a \times a^{-1} = a^{-1} \times a = 1$
- Commutativity (optional):
  $\forall a, b \in G, a \bullet b = b \bullet a$, which is true only for Abelian (commutative) group.

The order of a group $|G|$ is the count of elements in a group. If it is finite, it is a finite group.

Subgroups

A subgroup H of a group G is a group constructed from a subset of elements in a group with the same operation "•". If a subgroup of a group can be produced using the power of an element σ (generator), the subgroup is known as a cyclic subgroup, denoted by $\langle \sigma \rangle$. Elements in a finite cyclic subgroup can be written as $\{e, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$, where $\sigma^n = e$. The order of an element σ in a group is the smallest integer for which $\sigma^n = e$.

Lagrange's theorem

Assume G is a finite group and H is a subgroup of G. In that case, the order of H divides the order of G.

Rings

A ring is a set R along with two binary operations "+" and "×" obeying the axioms (R. B. J. T. Allenby, 1991):

- R is an Abelian group for the operation "+" (satisfies all five axioms for "+" operation) ,
- The operation "×" is associative and certainly closed also (satisfies first two axioms only for the "×" operation),
- The operations satisfy the Distributive Laws (the second operation is distributed over the first operation). $\forall a,b,c \in R, a \times (b + c) = a \times b + a \times c, (a + b) \times c = a \times c + b \times c$

If the second operation "×" is commutative, we call R a commutative ring. Sometimes the ring has a multiplicative identity and we say it a Ring with identity and the multiplicative identity is denoted by 1. The set of elements of a ring R that are invertible for the operation "•" is known as the set of units of R denoted by R*.

Example F.1

Assume R be a commutative ring with an identity. Consider a polynomial $p(\alpha)$ of degree n whose coefficients in R with an indeterminate α may be represented as, $p(\alpha) = \sum_{i=0}^{n} a_i \alpha^i, a_n \neq 0, a_i \in R$.

Addition and multiplication of polynomials are carried out as usual. The ring of such polynomials is denoted by $R[\alpha]$.

Subring

A subring S of a ring R is a subset of R that is a ring under the same operations as R. In other words, a non-empty subset S of R is a subring if $a, b \in S \Rightarrow a - b, ab \in S$. Hence S is closed under subtraction and multiplication. For example, the set $\{a + bi \in C \mid a, b \in Z\}$ makes a subring of C and is known as the ring of Gaussian integers (represented as $Z[i]$).

Ideal

An ideal is a special type of subring. A subring I of R is a left ideal if $a \in I, r \in R \Rightarrow ra \in I$. Hence I is closed under subtraction and multiplication on the left by elements of the ring. A right ideal may be defined alike. A two-sided ideal (usually referred to as an ideal) is both a left and right ideal. Hence $a, b \in I, r \in R \Rightarrow a - b, ar, ra \in I$. For instance, for a ring R with the subsets {0} and R are both two-sided ideals.

Principal ideal

A very great way of making ideals is stated below. Consider R to be a commutative ring with identity. Assume S is a subset of R. The ideal produced by S is the subset, $\langle S \rangle = r_1 s_1 + r_2 s_2 + \ldots + r_k s_k$; $r_1, r_2, \cdots \in R; s_1, s_2, \cdots \in S; k \in N$. More specifically, if S has a lone element s this is known as the principal ideal produced by s, expressed as $\langle S\rangle = rs \mid r \in R$. For non-commmutative ring, it is a left ideal. It is simple to change the above definition to make a right ideal or a two-sided ideal. If the ring is void of an identity then in most cases S will not be a subset of $\langle S\rangle$. Cyclic codes form principal ideals in a ring of polynomials. For example, the ideal 2Z of Z is the principal ideal $\langle 2\rangle$.

Fields: A field is a commutative ring with identity ($1 \neq 0$) in which each non-zero element has a multiplicative inverse.

The rings Q, R, C are fields. Consider an irreducible polynomial in α, $\alpha^2 - 2 = 0 \Rightarrow \alpha = \sqrt{2}$ is not an element of Q. One may extend Q by adding $\sqrt{2}$ to Q, denoted by $Q\left(\sqrt{2}\right)$ which contains both Q and $\sqrt{2}$ (E. Viterbo et. al., 2004). An element in $Q\left(\sqrt{2}\right)$ can be expressed in polynomial form, $a + b\sqrt{2}, a, b \in Q$. The basis vector for $Q\left(\sqrt{2}\right)$ is $\left\{1, \sqrt{2}\right\}$ and the dimension of $Q\left(\sqrt{2}\right)$ is 2 over Q.

Field extension from R → C (T. K. Moon, 2005)

A polynomial $p(\alpha)$ that cannot be factored into polynomials of lower degree is referred to as irreducible. For instance, $p(\alpha) = \alpha^2 + 1$ has no real roots (irreducible polynomial). Consider a polynomial notation of a complex number, $a + b\alpha$, $a, b \in R$, where R is a field of real numbers. We may assume $a + b\alpha$ is a polynomial of degree ≤ 1 in the indeterminate α.

- Addition: $a + b\alpha + c + d\alpha = (a + c) + (b + d)\alpha, a, b, c, d \in R$
- Multiplication: $(a + b\alpha)(c + d\alpha) = ac + (ad + bc)\alpha + (bd)\alpha^2$. It results in a quadratic polynomial but complex numbers are expressed as polynomial of degree ≤ 1 in the indeterminate α. How to convert it to this form? It can be done in the following ways. Divide $ac + (ad + bc)\alpha + (bd)\alpha^2$ by $p(\alpha) = \alpha^2 + 1$ and find the remainder which is $(ad + bc)\alpha + (ac - bd)$. In other words, any time $\alpha^2 + 1$ appears, we can replace it to 0 which implies that $\alpha^2 = -1$.

The new field C with the new element α in it is referred to as an extension field of the base field R.

Subfield

A subfield of a field is a subset of the field which is also a field.

Example F.2

Galois Field (GF) construction

A 4-tuple (a, b, c, d) expressing a number in $GF(2^4)$ can be expressed as, $a + b\alpha + c\alpha^2 + d\alpha^3$. Consider the irreducible polynomial over $GF(2)$ $p(\alpha) = 1 + \alpha + \alpha^4 \Rightarrow \alpha^4 = 1 + \alpha$. Hence we may simplify $\alpha^5 = \alpha\alpha^4 = \alpha + \alpha^2, \alpha^6 = \alpha\alpha^5 = \alpha^2 + \alpha^3, \alpha^7 = \alpha\alpha^6 = \alpha^3 + \alpha^4 = \alpha^3 + \alpha + 1$ and so on. The order of every finite GF must be power of a prime.

References

1. Allenby, R. B. J. T. 1991. *Rings, Fields and Groups*. London: Edward Arnold.
2. Forouzan, B. A. 2007. *Cryptography and Network Security*. New Delhi: Tata McGraw–Hill.
3. Moon, T. K. 2005. *Error Correction Coding*. Indianapolis: John Wiley & Sons.
4. Viterbo, E. and F. Oggier. 2004. *Algebraic Number Theory and Code Design for Rayleigh Fading Channels*. Delft: Now Publishers.

An Introduction to Game Theory

APPENDIX G

Introduction

Game theory presents a formal analysis of interaction among a group of players in a game. In game theory, the action of player directly affects the other players. There are two types of games:

(a) Cooperative game: Player form alliance to bring the result of the game in his/her favour.
(b) Non-cooperative game: Optimal strategy for such game is that leads to Nash equilibrium, a term coined after the Nobel Laureate John Nash. Any player cannot gain from varying un-alterably his/her strategy if the strategy of all other players is fixed. An action profile is a vector of player's actions. A Nash equilibrium is an action profile in which each action is the best response to the actions of all the other players. Nash equilibrium is a stable operating or equilibrium point. It means that there is no payoff for any player in a finite game to vary strategy when all the other players pursue the equilibrium policy. The learning process can be cast as a repeated stochastic game (i.e., recast version of a one-shot game). Every player gets or knows the past behavior of the other players, which may influence the present decision to be made. The job of a player is to choose the best mixed strategy, having the knowledge of mixed strategies of all other players in the game. The mixed strategy of a player is a RV whose values are the pure strategies of that player.

An example from wireless sensor network:
For instance, in wireless sensor network, a game needs three components (S. K. Das et al., 2004):

(a) Players (sensor nodes in the network)
(b) Strategy (criteria for selection of actions of players)
(c) Utility (payoff) function is the performance metric.

Let us assume a packet is sent from a source node (SN) to a destination node (DN) through an intermediate node (IN). IN may forward the packet to the DN by using the following criteria:

(a) Reputation: Have IN and DN made enough reputation to trust and co-operate each other to forward the packet?
(b) Distance: The farther are the two nodes, the more they don't trust each other.
(c) Traffic: Have IN and DN a joint operation history? Can they trust each other?

If IN and DN have sufficient reputation, closeness and joint operation history, then IN will forward the packet to DN. This is the strategy for players in the game. The payoff in this case is the number of packets each node receives and forwards at each time slot. The payoff function is defined based on cooperation, reputation and quality of security. In each cluster (subset of players or nodes in this case), a cluster head is chosen for a node with the highest cooperation, best reputation and high quality of security.

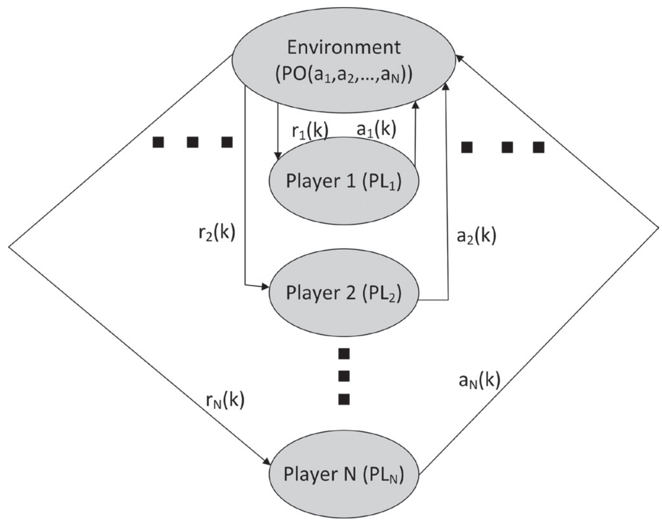


Fig. G.1 Automata game

Basic game theory

Let us formalize the basic game theory now. A learning automaton (player in our case) is an object that can select from a finite number of actions according to its environment (K. Narendra et al., 1989). In a multiple automata game depicted in Fig. G.1, N players viz., PL_1, PL_2, …, PL_N participate in a game. Each player PL_i is represented by 4-tuple $\{S_i, r_i, \mathbf{p}_i, LA_i\}$ where

- S_i are pure strategies or finite set of actions
- r_i is the response from the environments
- $\mathbf{p}_i(k) = \begin{bmatrix} p_{i1}(k) \\ p_{i2}(k) \\ \vdots \\ p_{im}(k) \end{bmatrix}$ is the action choice probability distribution vector of player i at time k and

  $p_{il}(k)$ is the probability that player i chooses the l^{th} pure strategy
- LA_i is the stochastic learning algorithm for updating $\mathbf{p}_i(k)$ that is, $\mathbf{p}_i(k + 1) = LA_i(\mathbf{p}_i(k), a_i(k), r_i(k))$ where $a_i \in S_i$ is the action selected by player i. The aim of each player is to augment the expected payoff. The payoff of each player is defined as $PO^i(a_1, a_2, \dots a_n) = E[r_i \mid \text{player j choose action } a_j; a_j \in S_j, j = 1,2,\cdots,N]$.

The pseudo code for learning algorithm is:

1. Initialize: choose $a_i \in S_i, i = 1,2,\cdots,N$ randomly with random $\mathbf{p}_i(0)$.
2. Do while

    $k = k + 1$

    for $i = 1$ to N (for all players)

    Apply a_i and observe response r_i from the environment

    Update $\mathbf{p}_i(k+1) = LA_i(\mathbf{p}_i(k), a_i(k), r_i(k))$

    End

    Until a stopping a criterion is reached or $\mathbf{p}_i(k)$ converges.

Persona models of players

There are four persona models in game theory (K. L. Du et al., 2010).

(a) Rational actor model (Homo economicus model): The players choose their action to maximize their own utility or payoff. In a non-cooperative game, players are selfish and take their action to maximize their own utility or payoff. In cooperative game, players cooperate by forming coalition and each coalition will have a single payoff. Assume a group of N players. Each player i is expected to invest to control the probability pd_i of damaging the system. The more players invest the lower is the pd_i expected. For selfish players who do not consider the overall condition of the system, the payoff for the ith player is

$$PO_i(pd_i) = pd_i\left(1 - \frac{1}{N}\sum_{i=1}^{N} pd_i\right) - c\left(pd_i^2\right)$$

where, c is the cost for not investing.

If no player invests, then $PO_i \rightarrow 0$. It is also called the tragedy of commons. If most players invest, then $PO_i \rightarrow pd_i$. The Nash equilibrium for this game is for

$$pd_i^{Nash} = \frac{1}{1 + 2c + \frac{1}{N}}, PO_i^{Nash} = \frac{c + \frac{1}{N}}{\left(1 + 2c + \frac{1}{N}\right)^2}.$$

(b) Homo equalis has desired to deal with fairness. They may be inclined to cut down their own payoff to increase the equality in the group when on top, but is displeased and crave to cut down inequality in the group when on the bottom. The utility function for ith player in a N-player game is given by

$$PO_i = PO_i - \frac{\alpha_i}{N-1}\sum_{PO_j > PO_i}\left(PO_j - PO_i\right) - \frac{\beta_i}{N-1}\sum_{PO_j < PO_i}\left(PO_i - PO_j\right), 0 \le \beta_i < \alpha_i \le 1$$

(c) Homo reciprocands tend to cooperate and share with others who are similarly disposed and punish those who violate cooperation even at present cost and even with no future rewards from doing so.

(d) Homo parochins segregate the world into insiders and outsiders and values insiders more profoundly than outsiders and partly suppress personal aims in favour of the aims of the group of insiders.

References

1. Das, S. K., A. Agah and K. Basu. 2004. 'Security in wireless mobile and sensor network'. In *Wireless Communication System and Network*. M. Guizani, Ed. NY: Springer. 550–552.
2. Du, K. L. and M. N. S. Swamy. 2010. *Wireless Communication Systems From RF Subsystems to 4G Enabling Technologies*. Cambridge: Cambridge University Press.
3. Narendra, K. and M. A. L. Thathachar. 1989. *Learning Automata: An Introduction*. New Jersey: Prentice Hall.

Index