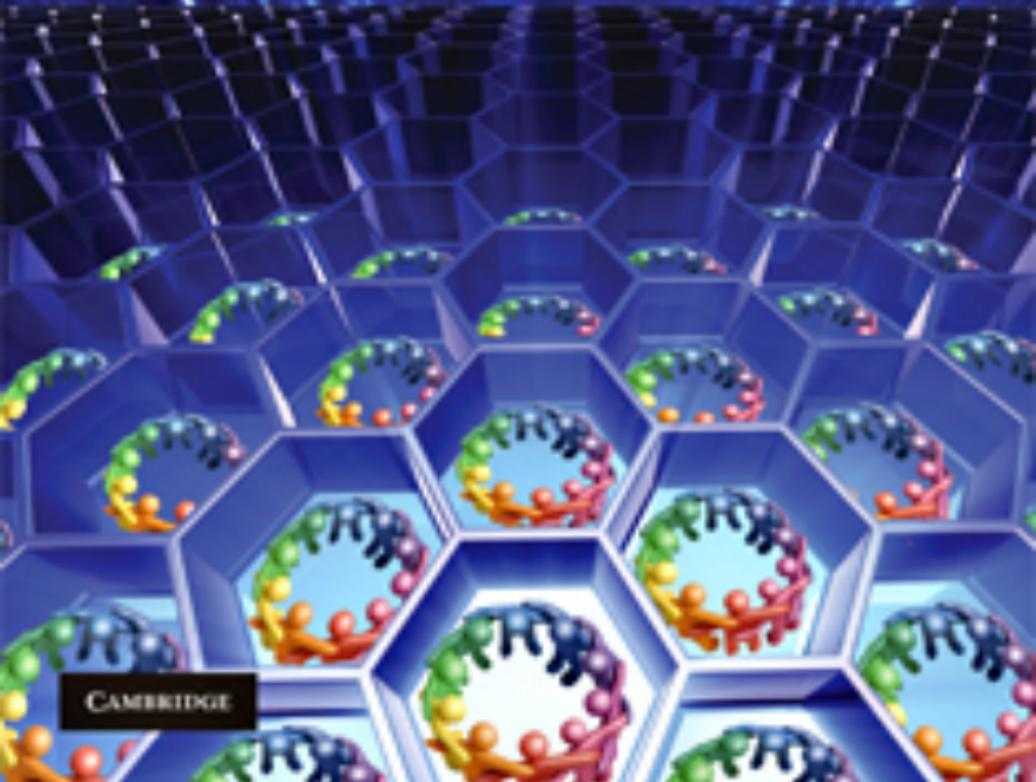


# Cooperative Communications and Networking

K. J. Ray Liu, Ahmed K. Sadek,  
Weifeng Su and Andres Kwasinski



CAMBRIDGE

CAMBRIDGE

[www.cambridge.org/9780521895132](http://www.cambridge.org/9780521895132)

This page intentionally left blank

## Cooperative Communications and Networking

Presenting the fundamental principles of cooperative communications and networking, this book treats the concepts of space, time, frequency diversity, and MIMO, with a holistic approach to principal topics where significant improvements can be obtained.

Beginning with background and MIMO systems, Part I includes a review of basic principles of wireless communications, space–time diversity and coding, and broadband space–time–frequency diversity and coding. Part II then goes on to present topics on physical layer cooperative communications, such as relay channels and protocols, performance bounds, optimum power control, multi-node cooperation, distributed space–time and space–frequency coding, relay selection, differential cooperative transmission, and energy efficiency. Finally, Part III focuses on cooperative networking including cooperative and content–aware multiple access, distributed routing, source–channel coding, source–channel diversity, coverage expansion, broadband cooperative communications, and network lifetime maximization.

With end-of-chapter review questions included, this text will appeal to graduate students of electrical engineering and is an ideal textbook for advanced courses on wireless communications. It will also be of great interest to practitioners in the wireless communications industry.

Presentation slides for each chapter and instructor-only solutions are available at [www.cambridge.org/9780521895132](http://www.cambridge.org/9780521895132)

**K. J. Ray Liu** is Professor in the Electrical and Computer Engineering Department, and Distinguished Scholar-Teacher, at the University of Maryland, College Park. Dr. Liu has received numerous honours and awards including best paper awards from IEEE Signal Processing Society, IEEE Vehicular Technology Society, and EURASIP, the IEEE Signal Processing Society Distinguished Lecturer, and National Science Foundation Young Investigator Award.

**Ahmed K. Sadek** is Senior Systems Engineer with Corporate Research and Development, Qualcomm Incorporated. He received his Ph.D. in Electrical Engineering from the University of Maryland, College Park, in 2007. His research interests include communication theory and networking, information theory and signal processing, with current focus on cognitive radios, spectrum sharing, cooperative communications, and interface management.

**Weifeng Su** is Assistant Professor at the Department of Electrical Engineering, State University of New York (SUNY) at Buffalo. He received his Ph.D. in Applied Mathematics from Nankai University, China in 1999, followed by his Ph.D. in Electrical Engineering from the University of Delaware, Newark in 2002. His research interests span a broad range of areas from signal processing to wireless communications and networking, and he won the Invention of the Year Award from the University of Maryland in 2005.

**Andres Kwasinski** is with Texas Instruments Inc., Communication Infrastructure Group. After receiving his Ph.D. in Electrical and Computer Engineering from the University of Maryland, College Park in 2004, he became Faculty Research Associate in the University's Department of Electrical and Computer Engineering. His research interests are in the areas of multimedia wireless communications, cross layer designs, digital signal processing, and speech and video processing.

# Cooperative Communications and Networking

K. J. RAY LIU

University of Maryland, College Park

AHMED K. SADEK

Qualcomm, San Diego, California

WEIFENG SU

State University of New York (SUNY) at Buffalo

ANDRES KWASINSKI

Texas Instruments, Germantown, Maryland



**CAMBRIDGE**  
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press  
The Edinburgh Building, Cambridge CB2 8RU, UK  
Published in the United States of America by Cambridge University Press, New York

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521895132](http://www.cambridge.org/9780521895132)

© Cambridge University Press 2009

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2008

ISBN-13 978-0-511-46079-1 eBook (Adobe Reader)

ISBN-13 978-0-521-89513-2 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

**To my parents Dr. Chau-Han Liu and Tama Liu – KJRL**

**To my parents Dr. Kamel and Faten and my wife Dina – AKS**

**To my wife Ming Yu and my son David – WS**

**To my wife Mariela and my daughters Victoria and Emma – AK**



# Contents

*Preface*

*page xi*

<b>Part I</b>	<b>Background and MIMO systems</b>	1
<b>1</b>	<b>Introduction</b>	3
	1.1 Wireless channels	4
	1.2 Characterizing performance through channel capacity	22
	1.3 Orthogonal frequency division multiplexing (OFDM)	25
	1.4 Diversity in wireless channels	29
	1.5 Cooperation diversity	40
	1.6 Bibliographical notes	42
<b>2</b>	<b>Space–time diversity and coding</b>	43
	2.1 System model and performance criteria	43
	2.2 Space–time coding	47
	2.3 Chapter summary and bibliographical notes	60
	Exercises	61
<b>3</b>	<b>Space–time–frequency diversity and coding</b>	64
	3.1 Space–frequency diversity and coding	64
	3.2 Space–time–frequency diversity and coding	98
	3.3 Chapter summary and bibliographical notes	113
	Exercises	114
<b>Part II</b>	<b>Cooperative communications</b>	117
<b>4</b>	<b>Relay channels and protocols</b>	119
	4.1 Cooperative communications	119
	4.2 Cooperation protocols	121

4.3	Hierarchical cooperation	138
4.4	Chapter summary and bibliographical notes	148
	Exercises	150
<b>5</b>	<b>Cooperative communications with single relay</b>	<b>152</b>
5.1	System model	152
5.2	SER analysis for DF protocol	155
5.3	SER analysis for AF protocol	170
5.4	Comparison of DF and AF cooperation gains	181
5.5	Trans-modulation in relay communications	186
5.6	Chapter summary and bibliographical notes	190
	Exercises	192
<b>6</b>	<b>Multi-node cooperative communications</b>	<b>194</b>
6.1	Multi-node decode-and-forward protocol	194
6.2	Multi-node amplify-and-forward protocol	217
6.3	Chapter summary and bibliographical notes	234
	Exercises	235
<b>7</b>	<b>Distributed space–time and space–frequency coding</b>	<b>238</b>
7.1	Distributed space–time coding (DSTC)	238
7.2	Distributed space–frequency coding (DSFC)	256
7.3	Chapter summary and bibliographical notes	273
	Appendix	274
	Exercises	275
<b>8</b>	<b>Relay selection: when to cooperate and with whom</b>	<b>278</b>
8.1	Motivation and relay-selection protocol	278
8.2	Performance analysis	282
8.3	Multi-node scenario	289
8.4	Optimum power allocation	295
8.5	Chapter summary and bibliographical notes	301
	Exercises	302
<b>9</b>	<b>Differential modulation for cooperative communications</b>	<b>306</b>
9.1	Differential modulation	306
9.2	Differential modulations for DF cooperative communications	308
9.3	Differential modulation for AF cooperative communications	347
9.4	Chapter summary and bibliographical notes	370
	Exercises	372

---

<b>10</b>	<b>Energy efficiency in cooperative sensor networks</b>	374
	10.1 System model	374
	10.2 Performance analysis and optimum power allocation	377
	10.3 Multi-relay scenario	381
	10.4 Experimental results	383
	10.5 Chapter summary and bibliographical notes	390
	Exercises	391
<b>Part III Cooperative networking</b>		393
<b>11</b>	<b>Cognitive multiple access via cooperation</b>	395
	11.1 System model	396
	11.2 Cooperative cognitive multiple access (CCMA) protocols	399
	11.3 Stability analysis	401
	11.4 Throughput region	423
	11.5 Delay analysis	424
	11.6 Chapter summary and bibliographical notes	429
	Exercises	429
<b>12</b>	<b>Content-aware cooperative multiple access</b>	432
	12.1 System model	433
	12.2 Content-aware cooperative multiple access protocol	437
	12.3 Dynamic state model	438
	12.4 Performance analysis	442
	12.5 Access contention–cooperation tradeoff	452
	12.6 Chapter summary and bibliographical notes	455
	Exercises	456
<b>13</b>	<b>Distributed cooperative routing</b>	457
	13.1 Network model and transmission modes	458
	13.2 Link analysis	461
	13.3 Cooperation-based routing algorithms	463
	13.4 Simulation examples	469
	13.5 Chapter summary and bibliographical notes	474
	Exercises	475
<b>14</b>	<b>Source–channel coding with cooperation</b>	478
	14.1 Joint source–channel coding bit rate allocation	478
	14.2 Joint source–channel coding with user cooperation	480
	14.3 The Source–channel–cooperation tradeoff problem	482
	14.4 Source codec	484
	14.5 Channel codec	488

---

14.6	Analysis of source–channel–cooperation performance	490
14.7	Validation of D-SNR characterization	504
14.8	Effects of source–channel–cooperation tradeoffs	505
14.9	Chapter summary and bibliographical notes	510
	Exercises	512
<b>15</b>	<b>Asymptotic performance of distortion exponents</b>	<b>514</b>
15.1	Systems setup for source–channel diversity	515
15.2	Multi-hop channels	519
15.3	Relay channels	532
15.4	Discussion	545
15.5	Chapter summary and bibliographical notes	547
	Exercises	548
<b>16</b>	<b>Coverage expansion with cooperation</b>	<b>550</b>
16.1	System model	550
16.2	Relay assignment: protocols and analysis	553
16.3	Relay assignment algorithms	557
16.4	Numerical results	563
16.5	Chapter summary and bibliographical notes	566
	Exercises	566
<b>17</b>	<b>Broadband cooperative communications</b>	<b>569</b>
17.1	System model	569
17.2	Cooperative protocol and relay-assignment scheme	571
17.3	Performance analysis	573
17.4	Performance lower bound	577
17.5	Optimum relay location	578
17.6	Chapter summary and bibliographical notes	580
	Exercises	581
<b>18</b>	<b>Network lifetime maximization via cooperation</b>	<b>583</b>
18.1	Introduction	583
18.2	System models	584
18.3	Lifetime maximization by employing a cooperative node	588
18.4	Deploying relays to improve device lifetime	597
18.5	Simulation examples	601
18.6	Chapter summary and bibliographical notes	605
	Exercises	607
	<i>References</i>	609
	<i>Index</i>	623

# Preface

Wireless communications technologies have seen a remarkably fast evolution in the past two decades. Each new generation of wireless devices has brought notable improvements in terms of communication reliability, data rates, device sizes, battery life, and network connectivity. In addition, the increase homogenization of traffic transports using Internet Protocols is translating into network topologies that are less and less centralized. In recent years, ad-hoc and sensor networks have emerged with many new applications, where a source has to rely on the assistance from other nodes to forward or relay information to a desired destination.

Such a need of cooperation among nodes or users has inspired new thinking and ideas for the design of communications and networking systems by asking whether cooperation can be used to improve system performance. Certainly it means we have to answer what and how performance can be improved by cooperative communications and networking. As a result, a new communication paradigm arose, which had an impact far beyond its original applications to ad-hoc and sensor networks.

First of all, why are cooperative communications in wireless networks possible? Note that the wireless channel is broadcast by nature. Even directional transmission is in fact a kind of broadcast with fewer recipients limited to a certain region. This implies that many nodes or users can “hear” and receive transmissions from a source and can help relay information if needed. The broadcast nature, long considered as a significant waste of energy causing interference to others, is now regarded as a potential resource for possible assistance. For instance, it is well known that the wireless channel is quite bursty, i.e., when a channel is in a severe fading state, it is likely to stay in the state for a while. Therefore, when a source cannot reach its destination due to severe fading, it will not be of much help to keep trying by leveraging repeating-transmission protocols such as ARQ. If a third party that receives the information from the source could help via a channel that is independent from the source–destination link, the chances for a successful transmission would be better, thus improving the overall performance.

Then how to develop cooperative schemes to improve performance? The key lies in the recent advances in MIMO (multiple-input multiple-output) communication technologies. In the soon-to-be-deployed fourth-generation (4G) wireless networks, very high data rates can only be expected for full-rank MIMO users. More specifically, full-rank MIMO users must be equipped multiple transceiver antennas. In practice, most

users either do not have multiple antennas installed on small-size devices, or the propagation environment cannot support MIMO requirements. To overcome the limitations of achieving MIMO gains in future wireless networks, one must think of new techniques beyond traditional point-to-point communications.

A wireless network system is traditionally viewed as a set of nodes trying to communicate with each other. However, from another point of view, because of the broadcast nature of wireless channels, one may think of those nodes as a set of antennas distributed in the wireless system. Adopting this point of view, nodes in the network may cooperate together for distributed transmission and processing of information. A cooperating node can act as a relay node for a source node. As such, cooperative communications can generate independent MIMO-like channel links between a source and a destination via the introduction of relay channels.

Indeed, cooperative communications can be thought of as a generalized MIMO concept with different reliabilities in antenna array elements. It is a new paradigm that draws from the ideas of using the broadcast nature of the wireless channels to make communicating nodes help each other, of implementing the communication process in a distribution fashion, and of gaining the same advantages as those found in MIMO systems. Such a new viewpoint has brought various new communication techniques that improve communication capacity, speed, and performance; reduce battery consumption and extend network lifetime; increase the throughput and stability region for multiple access schemes; expand the transmission coverage area; and provide cooperation tradeoff beyond source–channel coding for multimedia communications.

The main goals of this textbook are to introduce the concepts of space, time, frequency diversity, and MIMO techniques that form the foundation of cooperative communications, to present the basic principles of cooperative communications and networking, and to cover a broad range of fundamental topics where significant improvements can be obtained by use of cooperative communications. The book includes three main parts:

- **Part I: Background and MIMO systems** In this part, the focus is on building the foundation of MIMO concepts that will be used extensively in cooperative communications and networking. Chapter 1 reviews of fundamental material on wireless communications to be used in the rest of the book. Chapter 2 introduces the concept of space–time diversity and the development of space–time coding, including cyclic codes, orthogonal codes, unitary codes, and diagonal codes. The last chapter in this part, Chapter 3, concerns the maximum achievable space–time–frequency diversity available in broadband wireless communications and the design of broadband space–frequency and space–time–frequency codes.
- **Part II: Cooperative communications** This part considers mostly the physical layer issues of cooperative communications to illustrate the differences and improvements under the cooperative paradigm. Chapter 4 introduces the concepts of relay channels and various relay protocols and schemes. A hierarchical scheme that can achieve linear capacity scaling is also considered to give the fundamental reason

for the adoption of cooperation. Chapter 5 studies the basic issues of cooperation in the physical layer with a single relay, including symbol error rate analysis for decode-and-forward and amplify-and-forward protocols, performance upper bounds, and optimum power control. Chapter 6 analyses multi-node scenarios. Chapter 7 presents distributed space–time and space–frequency coding, a concept similar to the conventional space–time and space–frequency coding but different in that it is now in a distributed setting where assumptions and conditions vary significantly. Chapter 8 concerns the issue of minimizing the inherent bandwidth loss of cooperative communications by considering when to cooperate and whom to cooperate with. The main issue is on devising a scheme for relay selection and maximizing the code rate for cooperative communications while maintaining significant performance improvement. Chapter 9 develops differential schemes for cooperative communications to reduce transceiver complexity. Finally, Chapter 10 studies the issues of energy efficiency in cooperative communications by taking into account the practical transmission, processing, and receiving power consumption and illustrates the trade-off between the gains in the transmit power and the losses due to the receive and processing powers when applying cooperation.

- **Part III: Cooperative networking** This part presents impacts of cooperative communications beyond physical layer, including MAC, networking, and application layers. Chapter 11 considers the effect of cooperation on the capacity and stability region improvement for multiple access. Chapter 12 studies how special properties in speech content can be leveraged to efficiently assign resources for cooperation and further improve the network performance. Chapter 13 discusses cooperative routing with cooperation as an option. Chapter 14 develops the concept of source–channel–cooperation to consider the tradeoff of source coding, channel coding, and diversity for multimedia content. Chapter 15 focuses on studying how source coding diversity and channel coding diversity interact with cooperative diversity, and the system behavior is characterized and compared in terms of the asymptotic performance of the distortion exponent. Chapter 16 presents the coverage area expansion with the help of cooperation. Chapter 17 considers the various effects of cooperation on OFDM broadband wireless communications. Finally, Chapter 18 discusses network lifetime maximization via the leverage of cooperation.

This textbook primarily targets courses in the general field of cooperative communications and networking where readers have a basic background in digital communications and wireless networking. An instructor could select Chapters 1, 2, 4, 5, 6, 7.1, 8, 10, 11, 13, 14, and 16 to form the core of the material, making use of the other chapters depending on the focus of the course.

It can also be used for courses on wireless communications that partially cover the basic concepts of MIMO and/or cooperative communications which can be considered as generalized MIMO scenarios. A possible syllabus may include selective chapters from Parts I and II. If it is a course on wireless networking, then material can be drawn from Chapter 4 and the chapters in Part III.

This book comes with presentation slides for each chapter to aid instructors with the preparation of classes. A solution manual is also available to instructors upon request. Both can be obtained from the publisher via the proper channels.

This book could not have been made possible without the contributions of the following people: Amr El-Sherif, T. Kee Himsoon, Ahmed Ibrahim, Zoltan Safar, Karim Seddik, and W. Pam Siriwongpairat. We also would like to thank them for their technical assistance during the preparation of this book.

# Part I

---

## Background and MIMO systems



# 1 Introduction

---

Wireless communications have seen a remarkably fast technological evolution. Although separated by only a few years, each new generation of wireless devices has brought significant improvements in terms of link communication speed, device size, battery life, applications, etc. In recent years the technological evolution has reached a point where researchers have begun to develop wireless network architectures that depart from the traditional idea of communicating on an individual point-to-point basis with a central controlling base station. Such is the case with ad-hoc and wireless sensor networks, where the traditional hierarchy of a network has been relaxed to allow any node to help forward information from other nodes, thus establishing communication paths that involve multiple wireless hops. One of the most appealing ideas within these new research paths is the implicit recognition that, contrary to being a point-to-point link, the wireless channel is broadcast by nature. This implies that any wireless transmission from an end-user, rather than being considered as interference, can be received and processed at other nodes for a performance gain. This recognition facilitates the development of new concepts on distributed communications and networking via cooperation.

The technological progress seen with wireless communications follows that of many underlying technologies such as integrated circuits, energy storage, antennas, etc. Digital signal processing is one of these underlying technologies contributing to the progress of wireless communications. Perhaps one of the most important contributions to the progress in recent years has been the advent of MIMO (multiple-input multiple-output) technologies. In a very general way, MIMO technologies improve the received signal quality and increase the data communication speed by using digital signal processing techniques to shape and combine the transmitted signals from multiple wireless paths created by the use of multiple receive and transmit antennas.

Cooperative communications is a new paradigm that draws from the ideas of using the broadcast nature of the wireless channel to make communicating nodes help each other, of implementing the communication process in a distribution fashion and of gaining the same advantages as those found in MIMO systems. The end result is a set of new tools that improve communication capacity, speed, and performance; reduce battery consumption and extend network lifetime; increase the throughput and stability region for multiple access schemes; expand the transmission coverage area; and provide cooperation tradeoff beyond source–channel coding for multimedia communications.

In this chapter we begin with the study of basic communication systems and concepts that are highly related to user cooperation, by reviewing a number of concepts that will be useful throughout this book. The chapter starts with a brief description of the relevant characteristics of wireless channels. It then follows by discussing orthogonal frequency division multiplexing followed by the different concepts of channel capacity. After this, we describe the basic ideas and concepts of MIMO systems. The chapter concludes by describing the new paradigm of user cooperative communications.

## 1.1 Wireless channels

Communication through a wireless channel is a challenging task because the medium introduces much impairment to the signal. Wireless transmitted signals are affected by effects such as noise, attenuation, distortion and interference. It is then useful to briefly summarize the main impairments that affect the signals.

### 1.1.1 Additive white Gaussian noise

Some impairments are additive in nature, meaning that they affect the transmitted signal by adding noise. Additive white Gaussian noise (AWGN) and interference of different nature and origin are good examples of additive impairments. The additive white Gaussian channel is perhaps the simplest of all channels to model. The relation between the output  $y(t)$  and the input  $x(t)$  signal is given by

$$y(t) = x(t)/\sqrt{\Gamma} + n(t), \quad (1.1)$$

where  $\Gamma$  is the loss in power of the transmitted signal  $x(t)$  and  $n(t)$  is noise. The additive noise  $n(t)$  is a random process with each realization modeled as a random variable with a Gaussian distribution. This noise term is generally used to model background noise in the channel as well as noise introduced at the receiver front end. Also, the additive Gaussian term is frequently used to model some types of inter-user interference although, in general, these processes do not strictly follow a Gaussian distribution.

### 1.1.2 Large-scale propagation effects

The *path loss* is an important effect that contributes to signal impairment by reducing its power. The path loss is the attenuation suffered by a signal as it propagates from the transmitter to the receiver. The path loss is measured as the value in decibels (dB) of the ratio between the transmitted and received signal power. The value of the path loss is highly dependent on many factors related to the entire transmission setup. In general, the path loss is characterized by a function of the form

$$\Gamma_{\text{dB}} = 10\nu \log(d/d_0) + c, \quad (1.2)$$

where  $\Gamma_{\text{dB}}$  is the path loss  $\Gamma$  measured in dB,  $d$  is the distance between transmitter and receiver,  $\nu$  is the path exponent,  $c$  is a constant, and  $d_0$  is the distance to a power

measurement reference point (sometimes embedded within the constant  $c$ ). In many practical scenarios this expression is not an exact characterization of the path loss, but is still used as a sufficiently good and simple approximation. The path loss exponent  $\nu$  characterizes the rate of decay of the signal power with the distance, taking values in the range of 2 (corresponding to signal propagation in free space) to 6. Typical values for the path loss exponent are 4 for an urban macro cell environment and 3 for urban micro cell. The constant  $c$  includes parameter related to the physical setup of the transmission such as signal wavelength, antennas height, etc.

Equation (1.2) shows the relation between the path loss and the distance between the transmit and the receive antenna. In practice, the path losses of two receive antennas situated at the same distance from the transmit antenna are not the same. This is, in part, because the transmitted signal is obstructed by different objects as it travels to the receive antennas. Consequently, this type of impairment has been named *shadow loss* or *shadow fading*. Since the nature and location of the obstructions causing shadow loss cannot be known in advance, the path loss introduced by this effect is a random variable. Denoting by  $S$  the value of the shadow loss, this effect can be added to (1.2) by writing

$$\Gamma_{\text{dB}} = 10\nu \log(d/d_0) + S + c. \quad (1.3)$$

It has been found through experimental measurements that  $S$  when measured in dB can be characterized as a zero-mean Gaussian distributed random variable with standard deviation  $\sigma$  (also measured in dB). Because of this, the shadow loss value is a random value that follows a log-normal distribution and its effect is frequently referred as *log-normal fading*.

### 1.1.3 Small-scale propagation effects

From the explanation of path loss and shadow fading it should be clear that the reason why they are classified as large-scale propagation effects is because their effects are noticeable over relatively long distances. There are other effects that are noticeable at distances in the order of the signal wavelength; thus being classified as small-scale propagation effects. We now review the main concepts associated with these propagation effects.

In wireless communications, a single transmitted signal encounters random reflectors, scatterers, and attenuators during propagation, resulting in multiple copies of the signal arriving at the receiver after each has traveled through different paths. Such a channel where a transmitted signal arrives at the receiver with multiple copies is known as a multipath channel. Several factors influence the behavior of a multipath channel. One is the already mentioned random presence of reflectors, scatterers and attenuators. In addition, the speed of the mobile terminal, the speed of surrounding objects and the transmission bandwidth of the signal are other factors determining the behavior of the channel. Furthermore, due to the presence of motion at the transmitter, receiver, or surrounding objects, the multipath channel changes over time. The multiple copies of the transmitted signal, each having a different amplitude, phase, and delay, are added at the receiver creating either constructive or destructive interference with each other. This

results in a received signal whose shape changes over time. Therefore, if we denote the transmitted signal by  $x(t)$  and the received signal by  $y(t)$ , we can write their relation as

$$y(t) = \sum_{i=1}^L h_i(t)x(t - \tau_i(t)), \quad (1.4)$$

where  $h_i(t)$  is the attenuation of the  $i$ -th path at time  $t$ ,  $\tau_i(t)$  is the corresponding path delay, and  $L$  is the number of resolvable paths at the receiver. This relation implicitly assumes that the channel is linear, for which  $y(t)$  is equal to the convolution of  $x(t)$  and the channel response at time  $t$  to an impulse sent at time  $\tau$ ,  $h(t, \tau)$ . From (1.4), this impulse response can be written as

$$h(t, \tau) = \sum_{i=1}^L h_i(t)\delta(t - \tau_i(t)), \quad (1.5)$$

Furthermore, if it is safe to assume that the channel does not change over time, the received signal can be simplified as

$$y(t) = \sum_{i=1}^L h_i x(t - \tau_i),$$

and the channel impulse response as

$$h(t) = \sum_{i=1}^L h_i \delta(t - \tau_i). \quad (1.6)$$

In many situations it is convenient to consider the discrete-time baseband-equivalent model of the channel, for which the input–output relation derived from (1.4) for sample  $m$  can be written as

$$y[m] = \sum_{k=l}^L h_k[m]x[m - k], \quad (1.7)$$

where  $h_k[m]$  represents the channel coefficients. In this relation it is implicit that there is a sampling operation at the receiver and that all signals are considered as in the baseband equivalent model. The conversion to a discrete-time model combines all the paths with arrival time within one sampling period into a single channel response coefficient  $h_l[m]$ . Also, note that the model in (1.7) is nothing more than a time-varying FIR digital filter. In fact, it is quite common to call the channel model based on the impulse response as the tapped-delay model. Since the nature of each path, its length, and the presence of reflectors, scatterers, and attenuators are all random, the channel coefficients  $h_k$  of a time-invariant channel are random variables (and note that the redundant time index needs not be specified). If, in addition, the channel changes randomly over time, then the channel coefficients  $h_k[m]$  are random processes. Such an effect needs to be taken into consideration with functions that depend on the coefficients, since now they become random functions.

### 1.1.4 Power delay profile

The function determined by the average power associated with each path is called the *power delay profile* of the multipath channel. Figure 1.1 shows the power delay profile for a typical wireless channel slightly modified from the ITU reference channel model called “Vehicular B” [87]. Several parameters are derived from the power delay profile or its spectral response (Fourier transform of the power delay profile), which are used to both characterize and classify different multipath channels:

- The *channel delay spread* is the time difference between the arrival of the first measured path and the last. If the duration of the symbols used for signaling over the channel exceeds the delay spread, then the symbols will suffer from inter-symbol interference. Note that, in principle, there may be several signals arriving through very attenuated paths, which may not be measured due to sensitivity of the receiver. This makes the concept of delay spread tied to the sensitivity of the receiver.
- The *coherence bandwidth* is the range of frequencies over which the amplitude of two spectral components of the channel response are correlated. The coherence bandwidth provides a measurement of the range of frequencies over which the channel shows a flat frequency response, in the sense that all the spectral components have approximately the same amplitude and a linear change of phase. This means that if the transmitted signal bandwidth is less than the channel coherence bandwidth, then all the spectral components of the signal will be affected by the same attenuation and by a linear change of phase. In this case, the channel is said to be a *flat fading channel*. In another way, since the signal sees a channel with flat frequency response, the channel is often called a *narrowband channel*. If on the contrary, the transmitted signal

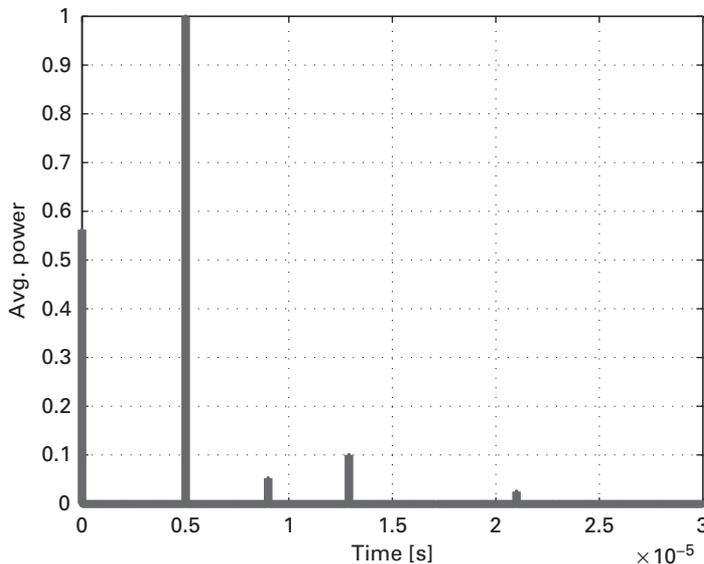


Fig. 1.1 The power delay profile of a typical wireless channel.

bandwidth is more than the channel coherence bandwidth, then the spectral components of the signal will be affected by different attenuations. In this case, the channel is said to be a *frequency selective channel* or a *broadband channel*.

**Example 1.1** There are a large number of different channel models that have been used over time for evaluation of communications systems. The large number is due to the different settings found in the plethora of communication systems already in the market or under development. In Tables 1.1 through 1.4 we summarize the parameters of the power delay profile for some of the channels defined in the ITU recommendation M.1225, which is intended for a system operating at a carrier frequency of 2 GHz. In the ITU recommendation, several channel models are discussed so as to account for typically large variability of wireless channels. In this example, Tables 1.1 and 1.2 show the parameters for channels corresponding to a pedestrian setting. As its names indicates, this environment is designed to model pedestrian users, either outside on a street or inside a residence, with small cells, low transmit power and outside base stations with low antenna heights. Tables 1.3 and 1.4 show the parameters for channels corresponding to a vehicular setting. In contrast with the pedestrian environment, the vehicular case models larger cell sizes and transmit power. Also to account for the large variability of wireless channels, two types of channel models are specified for both the pedestrian and vehicular cases. The two types of channels are called “type A” and “type B”, where the channel type A is defined as that of a low delay spread case that occurs frequently and channel type B is defined as that of the median delay spread case.

**Table 1.1** ITU-R M.1225 Pedestrian A channel parameters.

Tap	Relative delay [ns]	Average power [dB]
1	0	0
2	110	−9.7
3	190	−19.2
4	410	−22.8

**Table 1.2** ITU-R M.1225 Pedestrian B channel parameters.

Tap	Relative delay [ns]	Average power [dB]
1	0	0
2	200	−0.9
3	800	−4.9
4	1200	−8.0
4	2300	−7.8
4	3700	−23.9

**Table 1.3** ITU-R M.1225 Vehicular A channel parameters.

Tap	Relative delay [ns]	Average power [dB]
1	0	0
2	310	-1.0
3	710	-9.0
4	1090	-10.0
4	1730	-15.0
4	2510	-20.0

**Table 1.4** ITU-R M.1225 Vehicular B channel parameters.

Tap	Relative delay [ns]	Average power [dB]
1	0	-2.5
2	300	0
3	8900	-12.8
4	12900	-10.0
4	17100	-25.2
4	20000	-16.0

Figures 1.2 and 1.3 show in the time and frequency domain, respectively, the impulse response in Tables 1.1 through 1.4. The figures illustrate the typical variability of channel models, both in terms of delay spread and coherence bandwidth. Also note how, within the same type A or type B channels, the vehicular channels exhibit a larger delay spread. ▲

Whether a particular channel will appear as flat fading or frequency selective depends, of course, on the channel delay spread, but it also depends on the characteristics of the signal being sent through the channel. Figure 1.4 shows a section of the spectral response of the channel with power delay profile shown in Figure 1.1. We can see that if the transmitted signal has a bandwidth larger than a few tens of kilohertz, then the channel will affect differently those spectral components of the transmitted signal that are sufficiently apart.

This can be seen in Figure 1.5, which shows the time and frequency domain input and output signals to the channel in Figures 1.1 and 1.2. In Figure 1.5, the input signal is a raised cosine pulse with roll off factor 0.25 and symbol period 0.05  $\mu$ s. For this pulse, the bandwidth is approximately 2 MHz. This makes the channel behave like a frequency selective channel. As can be seen in the frequency domain representation of the output pulse in Figure 1.5, the typical result of the frequency selectivity is that there are large differences in how each spatial component is affected. In the time domain, it can be

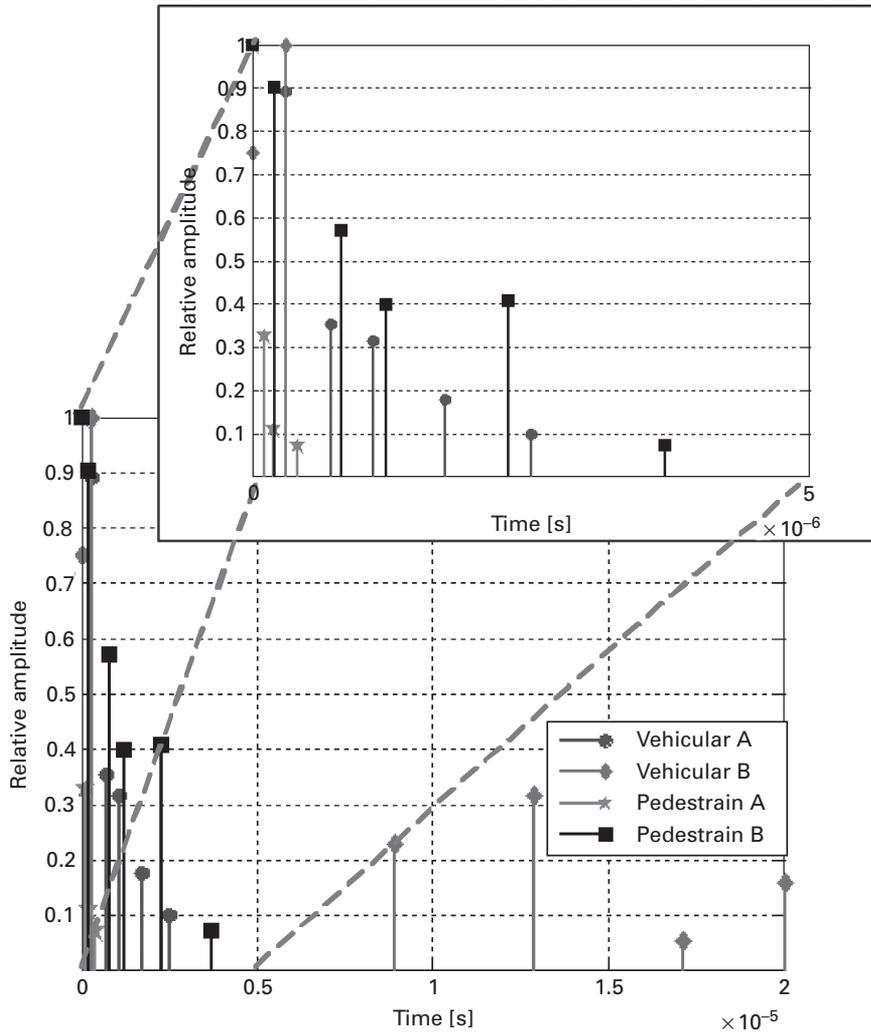
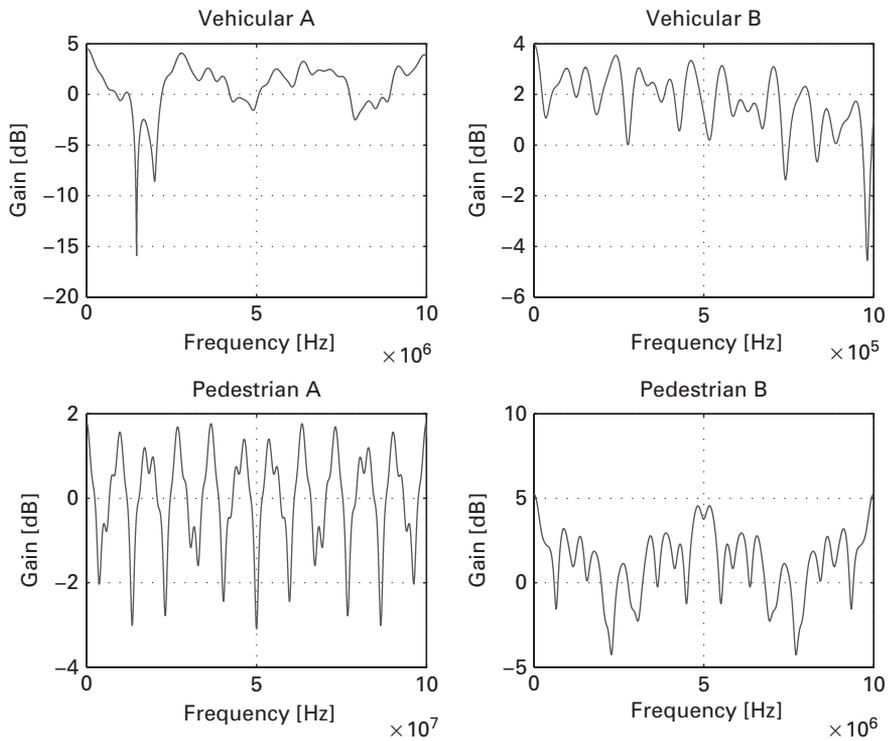


Fig. 1.2

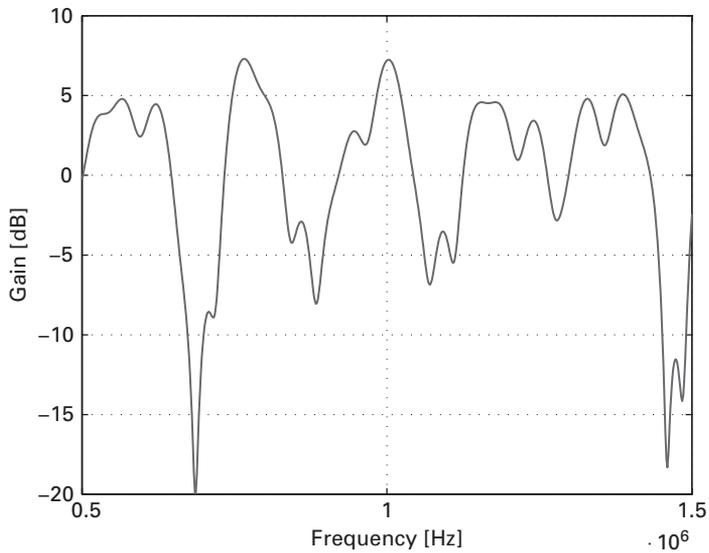
The amplitude of the different paths for the channels in Tables 1.1 through 1.4. The amplitudes of each path are shown relative to the value of the path with larger gain.

seen that the single pulse at the input of the channel appears repeated at the output with different delays corresponding to each path.

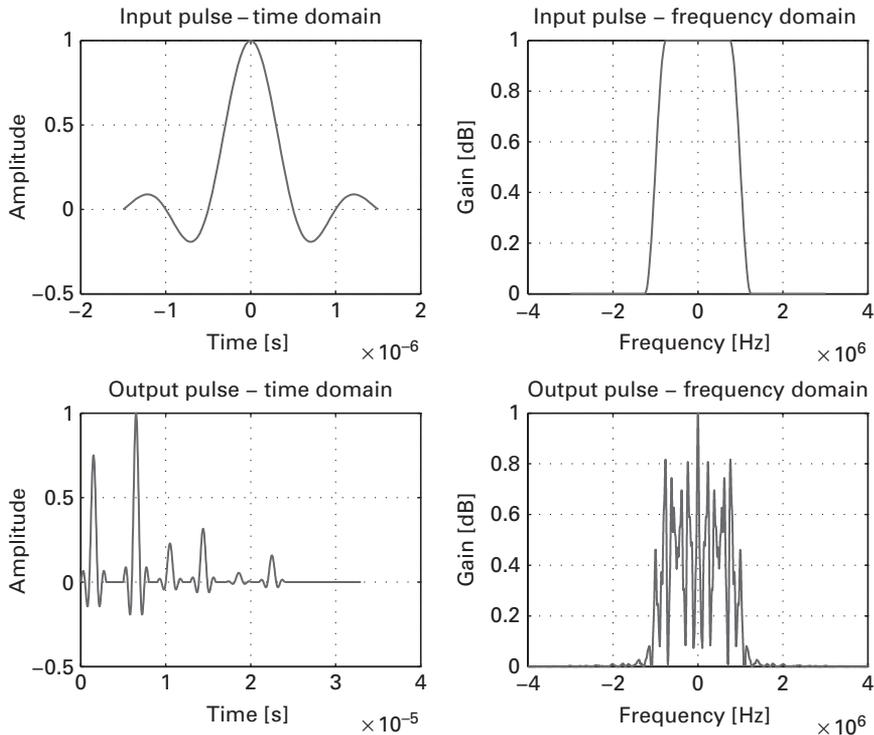
Such a phenomenon can also be seen in detail in Figure 1.6, which shows the output pulse and each of the pulses arriving through a different path, with their corresponding delay. Since the delay associated with some path is larger than the symbol period, the multipath, frequency selective channel is suffering from intersymbol interference (ISI). The fact that a time domain phenomenon such as instances of a signal arriving with different delays, translate into a frequency domain effect, such as frequency selectivity, can be understood in the following way. When the signals with different delays from the multipath get superimposed at the receive antenna, the different delay translates



**Fig. 1.3** Spectral response of the channels with power delay profile in Tables 1.1 through 1.4.



**Fig. 1.4** Section of the spectral response of the channel with power delay profile shown in Figure 1.1.

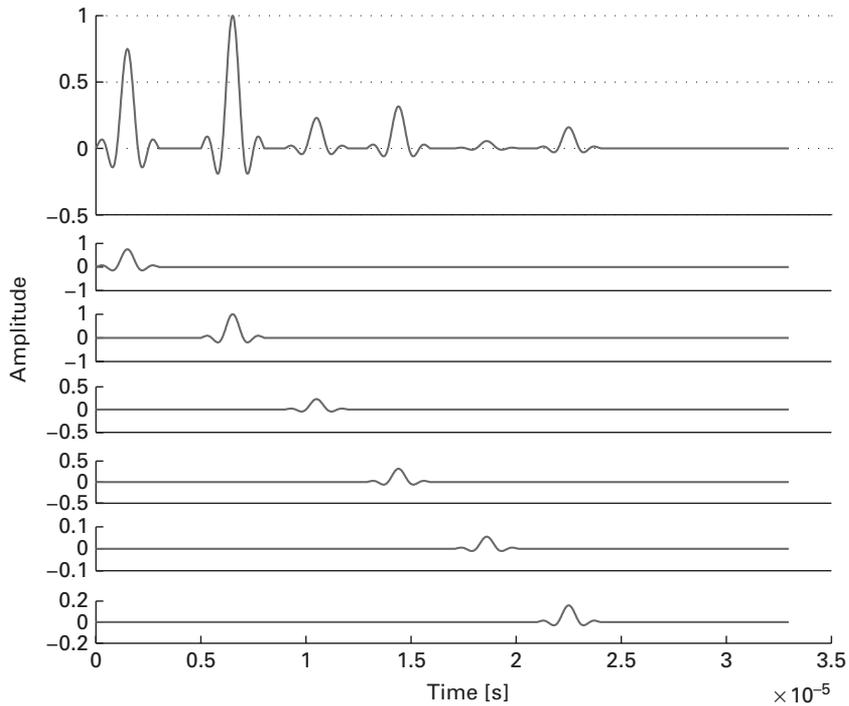


**Fig. 1.5** The input and output pulses to a frequency selective channel.

into different phases. Depending on the phase difference between the spectral components, their superposition may result into destructive or constructive interference. Even more, because the relation between phase and path delay for each spectral component of the arriving signal varies with the frequency of the spectral component, the signal will undergo destructive or constructive interference of different magnitude for each spectral component, resulting in the frequency response of the channel not appearing of constant amplitude.

Figures 1.7 and 1.8 show the time and frequency domains input and output signals to the channel in Figures 1.1 and 1.4 when the input pulse have a transmission period long enough that the channel behaves as non frequency selective. In this case, the input pulse has a bandwidth of approximately 2 KHz, for which the frequency response of the channel appears roughly flat. Consequently, the transmitted pulse suffers little alterations in both time and frequency domains. Also, note that now with the longer duration of the pulse, the delays associated with different channel paths can be practically neglected and there is no ISI.

In addition to power delay profile and channel delay spread, there are other parameters related to time-varying characteristics of the wireless channel. As we have said, the motions of the transmitter, the receiver or the reflectors along the signal propagation path creates a change of the channel transfer characteristics over time. Such motions also introduce frequency shifts due to the Doppler shift effect. To characterize the channel in



**Fig. 1.6** The pulse at the output of a frequency selective channel and each of the component pulses due to multipath.

terms of Doppler shift it is necessary to look at the variation of the channel power profile over time. In other words, instead of considering the statistics of the channel between two frequencies at a fixed time instant, we now look at the same frequency component as it changes over time. The parameters usually considered in these cases are:

- For a time-invariant channel, the coefficient of the channel impulse response corresponding to one path is a random variable (generally a complex-valued Gaussian random variable). When the channel changes over time, the coefficient becomes a random process, with each realization at different time instants being a random variable. These random variables may or may not be correlated. The *channel coherence time* is the time difference that makes the correlation between two realizations of the channel impulse response be approximately zero. The Fourier transform of the correlation function between the realizations of a channel coefficient is known as the *channel Doppler power spectrum*, or simply the Doppler spectrum. The Doppler spectrum characterizes in a statistical sense how the channel response widens an input signal spectrum due to Doppler shift, i.e., if a single tone of frequency  $f_c$  is sent through a channel with a Doppler shift  $f_d$ , the Doppler spectrum will have components in the range from  $f_c - f_d$  to  $f_c + f_d$ .

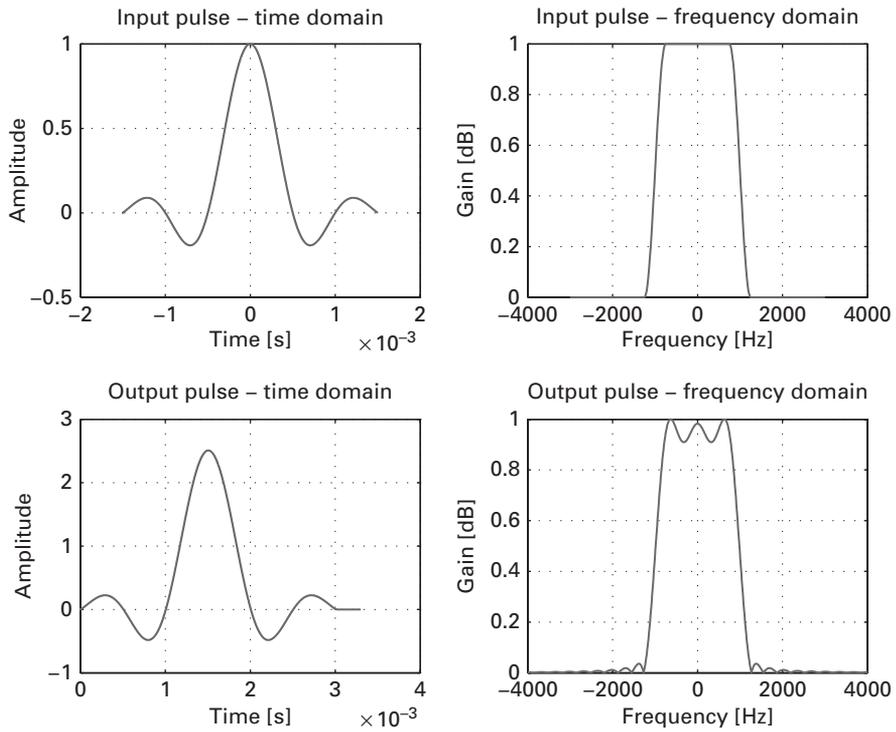
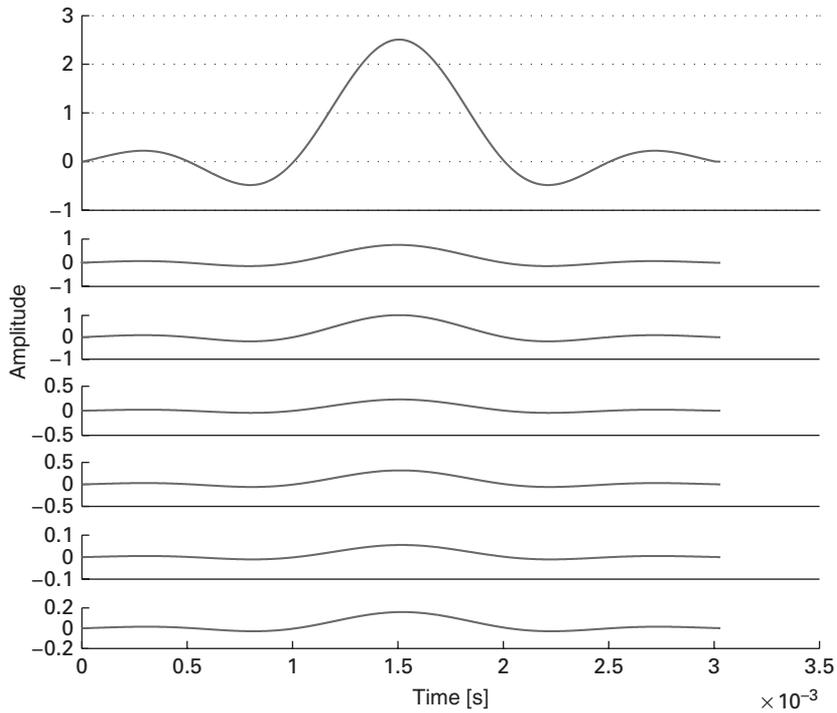


Fig. 1.7 The input and output pulses to a frequency non-selective channel.

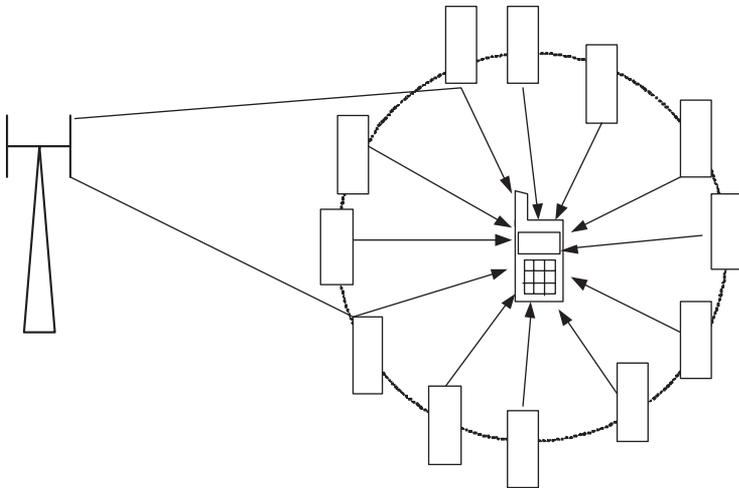
- The *Doppler spread* is defined as the range of frequencies over which the Doppler power spectrum is nonzero. The Doppler spread is the inverse of the channel coherence time and, as such, provides information on how fast the channel changes over time. Here, again, the notion on how fast the channel is changing depends also on the input signal. If the channel coherence time is larger than the transmitted signal symbol period; or equivalently, if the Doppler spread is smaller than the signal bandwidth, the channel will be changing over a period of time longer than the input symbol duration. In this case, the channel is said to have *slow fading*. If the converse applies, the channel is said to have *fast fading*.

### 1.1.5 Uniform scattering environment models

As previously mentioned, the channel coefficients are complex-valued random variables or processes. This raises the important question of what are the statistical properties of the coefficients and what kind of mathematical model can characterize this behavior. One of the most common models for the random channel coefficients is based on an environment known as the “uniform scattering environment.” Since this model was introduced by R. H. Clarke and later developed by W. C. Jakes, the model is also known as Clarke’s model or Jakes’ model. In the model, it is assumed that a waveform arrives at a receiver after being scattered on a very large number of scatterers. These scatterers



**Fig. 1.8** The pulse at the output of a frequency non-selective channel and each of the component pulses due to multipath.



**Fig. 1.9** The uniform scattering environment.

are assumed to be randomly located on a circle centered on the receiver (see Figure 1.9). In the environment it is assumed that there is no line-of-sight (LOS) signal with a power notably larger than the rest. Consequently, the received waveform is made of the superposition of many waveforms arriving from the scatterers at an angle that is uniformly

distributed between 0 and  $2\pi$ . For the purpose of this study, let us first introduce the complex baseband representation of the transmitted bandpass signals  $s(t)$ ,

$$s(t) = \Re\left\{x(t)e^{j2\pi f_c t}\right\},$$

where  $f_c$  is the carrier frequency. Expanding this expression, we get

$$\begin{aligned} s(t) &= \Re\{x(t)\} \cos(2\pi f_c t) + \Im\{x(t)\} \sin(2\pi f_c t) \\ &= s_I(t) \cos(2\pi f_c t) + s_Q(t) \sin(2\pi f_c t), \end{aligned}$$

where  $s_I(t)$  and  $s_Q(t)$  are the in-phase and quadrature components of  $s(t)$ , respectively. When this signal is transmitted through a channel with baseband impulse response  $h(t)$ , the resulting received signal is

$$y(t) = \Re\left\{(x(t) * h(t))e^{j2\pi f_c t}\right\}.$$

If the channel has  $L$  paths, with path  $n$  having amplitude  $h_n(t)$ , an associated delay  $\tau_n(t)$ , and a Doppler phase shift  $\varphi_n$  (which accounts for the Doppler shift due to the motion of the receiver of each received wave), the received signal can be written as

$$y(t) = \Re\left\{\sum_{n=1}^L h_n(t)x(t - \tau_n(t))e^{j[2\pi f_c(t - \tau_n(t)) + \varphi_n]}\right\}.$$

If, for the purpose of this characterization, we assume that the transmitted signal is a single tone with the same frequency as the carrier frequency, the received signal becomes

$$\begin{aligned} y(t) &= \Re\left\{\sum_{n=1}^L \left[h_n(t)e^{-j(2\pi f_c \tau_n(t) - \varphi_n)}\right]e^{j2\pi f_c t}\right\} \\ &= y_I(t) \cos(2\pi f_c t) + y_Q(t) \sin(2\pi f_c t), \end{aligned} \quad (1.8)$$

where

$$y_I(t) = \sum_{n=1}^L h_n(t) \cos(2\pi f_c \tau_n(t) - \varphi_n), \quad (1.9)$$

$$y_Q(t) = \sum_{n=1}^L h_n(t) \sin(2\pi f_c \tau_n(t) - \varphi_n). \quad (1.10)$$

This result shows that both the in-phase and the quadrature components of the received signal are actually composed of the superposition of multiple copies of the signal arriving with a change of amplitude and phase as determined by the characteristics of each of the channel paths.

Next, as part of the settings associated with the uniform scattering environment, we assume that all the received signals arrive with the same amplitude. This is a reasonable assumption because, given the geometry of the uniform scattering environment, in the absence of a direct LOS path, each signal arriving at the receiver would experience similar attenuations. Furthermore, in the uniform scattering environment, it is reasonable

to assume that the number of paths  $L$  is very large. Therefore, resorting to the Central Limit Theorem, it follows that each coefficient can be modeled as a circularly symmetric complex Gaussian random variable with zero mean (i.e., as a random variable made of two quadrature components, with each component being a zero mean Gaussian random variable with the same variance as the other component  $\sigma^2$ ). We denote this observation as  $y_I \sim \mathcal{N}(0, \sigma^2)$ ,  $y_Q \sim \mathcal{N}(0, \sigma^2)$ .

To better understand the channel behavior, it is important to find the statistics (in terms of probability density function (pdf)) of the envelope and phase of the channel coefficients. To get this, it is necessary to consider the transformation of those random variables representing a channel coefficient in Cartesian coordinates into those representing the coefficient in polar coordinates. This means that, if we write the coefficient as  $h = h_I + jh_Q$  ( $h_I$  and  $h_Q$  represent the in-phase and the quadrature phase components, respectively), we want to find the pdf of the random variables  $r$  and  $\theta$ , obtained through the transformations

$$\begin{aligned} r &= \sqrt{h_I^2 + h_Q^2}, \\ \theta &= \arctan(h_Q/h_I), \end{aligned} \quad (1.11)$$

which represent a channel coefficient as  $h = r e^{j\theta}$ . Equivalently, we may consider the inverse transform

$$\begin{aligned} h_I &= r \cos \theta, \\ h_Q &= r \sin \theta. \end{aligned} \quad (1.12)$$

For the general case of transforming random variables  $V = t_1(X, Y)$  and  $W = t_2(X, Y)$ , the transformation of the joint pdf  $f_{X,Y}$  of the random variables  $X$  and  $Y$  into the joint pdf  $f_{V,W}$  of the random variables  $V$  and  $W$ , is given by the expression [112]

$$\begin{aligned} f_{V,W}(v, w) &= \frac{f_{X,Y}(s_1(v, w), s_2(v, w))}{|\mathcal{J}(x, y)|}, \\ \mathcal{J}(x, y) &= \det \begin{bmatrix} \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{bmatrix}, \end{aligned}$$

where  $\mathcal{J}(x, y)$  is the Jacobian of the transformation and where  $x = s_1(v, w)$  and  $y = s_2(v, w)$  are the inverse transformations of  $t_1$  and  $t_2$ , respectively. Applying this relation to the transformation (1.12) results in the Jacobian  $\mathcal{J}(r, \theta) = r/\sigma^2$ , which leads to the joint pdf

$$f(r, \theta) = \frac{r}{2\pi\sigma^2} e^{-r^2/(2\sigma^2)}, \quad r \geq 0, 0 \leq \theta \leq 2\pi,$$

where we have used for easier readability a slightly modified notation for the pdf. From the joint pdf it is possible to find the marginal pdfs

$$f(r) = \int_0^{2\pi} f(r, \theta) d\theta = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}, \quad r \geq 0,$$

$$f(\theta) = \int_0^\infty f(r, \theta) dr = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi.$$

This result shows that the magnitude of the channel coefficients is a random variable with a Rayleigh distribution and the phase is also a random variable with a uniform distribution in the range  $[0, 2\pi]$ . Because the magnitude of the channel coefficients follow a Rayleigh distribution, this model is frequently called a *Rayleigh fading* model.

Also, for the case of two nonnegative random variables related by the transformation  $Y = X^2$ , using similar random variables transformation techniques yields the relation between pdfs

$$f_Y(y) = \frac{f_X(x)}{dy/dx}.$$

With this relation it can be shown that a random variable  $X$  that is defined as the squared magnitude of a Rayleigh-distributed channel coefficient ( $X = |h|^2$ ) follows an exponential distribution, with pdf

$$f_X(x) = \frac{1}{\sigma^2} e^{-x/\sigma^2}, \quad x \geq 0. \quad (1.13)$$

In addition, the sum of the squared magnitude of channel coefficient,  $\sum_i |h_i|^2$ , where each is the sum of two real i.i.d. Gaussian random variables representing the in-phase and quadrature components (i.e.,  $h_i = h_{I_i} + jh_{Q_i}$ ), results in a Chi-square random variable with  $2L$  ( $L$  being the number of channel coefficients in the sum) degrees of freedom. The pdf of this distribution is

$$f(x) = \frac{x^{L-1}}{(L-1)!} e^{-x}, \quad x \geq 0. \quad (1.14)$$

To consider the statistics of the received signal and the channel as they change over time, the two most important results are the time correlation and its Fourier transform, the power spectral density (PSD). Using as a starting point (1.8), (1.9), and (1.10), the time correlation of the received signal is

$$\begin{aligned} C_y(\tau) &= E[y(t)y(t+\tau)] \\ &= C_{y_I}(\tau) \cos(2\pi f_c \tau) + C_{y_I, y_Q}(\tau) \sin(2\pi f_c \tau), \end{aligned} \quad (1.15)$$

where  $E[\cdot]$  is the expectation operator. The autocorrelation of  $y_I$ ,  $C_{y_I}(\tau)$  equals

$$C_{y_I}(\tau) = E[y_I(t)y_I(t+\tau)].$$

Considering the expression for  $y_I(t)$  in (1.9), the magnitude of the Doppler phase shift,  $\varphi_n$ , depends on the velocity of the receiver relative to the scatterer from where the wave comes from. This relative velocity is equal to  $v \cos \alpha_k$ , where  $v$  is the absolute velocity

of the receiver. If we denote the angle associated with the  $k$ -th wave as  $\alpha_k$ , the Doppler shift for waveform  $k$  equals

$$\varphi_k = 2\pi \frac{v}{\lambda} t \cos \alpha_k = 2\pi f_D t \cos \alpha_k,$$

where  $\lambda$  is the wavelength and  $f_D = v/\lambda$  is the Doppler frequency. Now we can write (1.9) as

$$y_I(t) = \sum_{n=1}^L h_n(t) \cos(2\pi f_c \tau_n(t) - 2\pi \frac{v}{\lambda} t \cos \alpha_n). \quad (1.16)$$

In the uniform scattering environment the phase  $2\pi f_c \tau_n(t)$  changes more rapidly than the Doppler phase shift. In addition, since the distance from the scatterers to the mobile is much larger than the signal wavelength, it is possible to assume that the angle associated with the  $k$ -th wave,  $\alpha_k$ , is a random variable uniformly distributed in  $[0, 2\pi]$  and independent of the angle associated with other paths. Under these conditions, the autocorrelation of  $y_I$ ,  $C_{y_I}(\tau)$  equals

$$\begin{aligned} C_{y_I}(\tau) &= \frac{P_r}{2\pi} \int_0^{2\pi} \cos(\pi v \tau \cos \alpha / \lambda) d\alpha \\ &= P_r J_0(2\pi f_c v \tau / c), \end{aligned}$$

where  $c$  is the speed of light,  $P_r$  is the received power, and  $J_0(\cdot)$  is the Bessel function of the first kind and zeroth order, defined as

$$J_0(x) = \frac{1}{\pi} \int_0^\pi e^{-jx \cos \theta} d\theta.$$

Using similar reasoning, the cross-correlation between the in-phase and quadrature components of the received signal,  $C_{y_I, y_Q}(\tau)$ , can be found to equal zero. Therefore, using (1.15), the time correlation of the received signal is

$$C_y(\tau) = P_r J_0(2\pi f_c v \tau / c) \cos(2\pi f_c \tau). \quad (1.17)$$

In this result, the  $\cos(2\pi f_c \tau)$  component indicates the correlation of the received signal to a complete period shift due to being a single tone of frequency  $f_c$ . Taking the Fourier transform of (1.17) yields the power spectral density of the received signal

$$S_y(f) = \begin{cases} \frac{P_r}{4\pi f_D} \frac{1}{\sqrt{1 - \left(\frac{|f - f_c|}{f_D}\right)^2}} & \text{if } |f - f_c| \leq f_D \\ 0 & \text{else.} \end{cases} \quad (1.18)$$

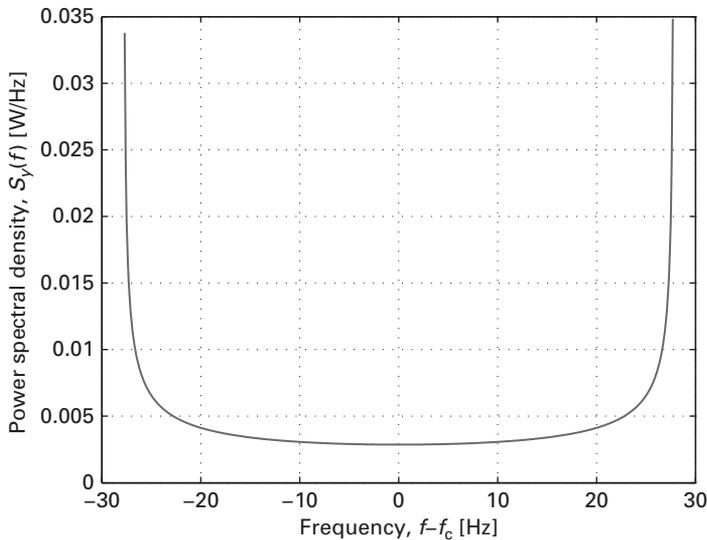
Note here that the frequency shift  $f - f_c$  is a consequence of the  $\cos(2\pi f_c \tau)$  component in (1.17) and the frequency shift property of Fourier transforms. Since the input signal is a single tone, ignoring the frequency shift in the power spectral density of the channel effects, which consequently has a time correlation equal to

$$C_h(\tau) = P_r J_0(2\pi f_c v \tau / c). \quad (1.19)$$

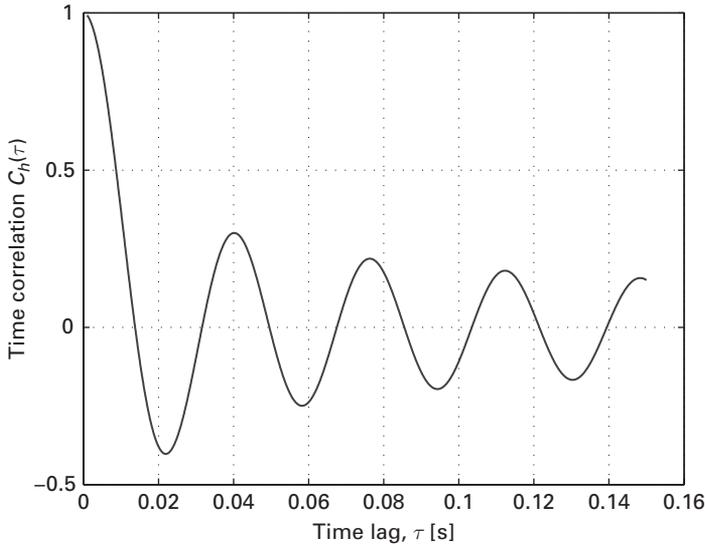
**Example 1.2** In this example we show typical cases of the cross-correlation and the power spectral density functions we just have discussed. We assume a system setup where  $f_c = 1$  GHz,  $v = 30$  km/h, and  $P_r = 1$ . Figure 1.10 shows the power spectral density as obtained from (1.18). For the settings in this example,  $f_D \approx 27$  Hz, so in the figure the rapid increase of PSD at frequencies near  $f = f_c \pm f_D$  can be observed. Since the uniform scattering environment is nothing more than a model to represent physical channel behaviors, the PSD in (1.18) could become infinite at frequencies  $f = f_c \pm f_D$ . This cannot happen in practice, but (1.18) tells us that the PSD will be maximum at these frequencies. Next, Figure 1.11 shows the time correlation as given by (1.19). Note that there are values of  $\tau$  for which the correlation is 0, which means that the channel will decorrelate signals arriving with these delays. ▲

### 1.1.6 Other channel coefficients models

The Rayleigh fading model is not the only model for the channel coefficients. In fact, when we derived the Rayleigh fading for the uniform scattering environment from (1.8)–(1.10), we assumed that all the received signals arrive with the same amplitude due to the absence of a direct LOS path and the symmetric geometry of the environment. When there is one line-of-sight path, then it can no longer be assumed that both the in-phase and quadrature components can be approximated as a zero mean Gaussian random variable. Now, the two components are Gaussian random variables but one has



**Fig. 1.10** Power spectral density of a received tone of frequency  $f_c = 1$  GHz for a mobile in a uniform scattering environment moving at  $v = 30$  km/h.



**Fig. 1.11** Time correlation for a signal in a uniform scattering environment over a channel with central frequency  $f_c = 1$  GHz and for a mobile speed of  $v = 30$  km/h.

mean  $A$ , which is the peak amplitude of the signal from the line-of-sight path, and the other still has zero mean. In this case, we can still use the useful transformations (1.11), for which the cumulative distribution function (CDF) of  $r$  is

$$\begin{aligned}
 F_r(z) &= \Pr[r \leq z] \\
 &= \frac{1}{2\pi\sigma^2} \iint_{\sqrt{h_1^2+h_2^2} \leq z} \exp\left[-\frac{1}{2}\left(\left[\frac{h_1-A}{\sigma}\right]^2 + \left[\frac{h_2}{\sigma}\right]^2\right)\right] dx dy \\
 &= \frac{e^{-\frac{1}{2}\left(\frac{A}{\sigma}\right)^2}}{2\pi\sigma^2} \int_0^z e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2} \left(\int_0^{2\pi} e^{rA \cos \frac{\theta}{\sigma^2}} d\theta\right) u(z)r dr, \quad (1.20)
 \end{aligned}$$

where  $u(z)$  is the unit step function so that (1.20) is valid only for  $z \geq 0$ . Next, the integral in  $\theta$  in (1.20) can be written in terms of the modified Bessel function of the first kind and zeroth order, defined as

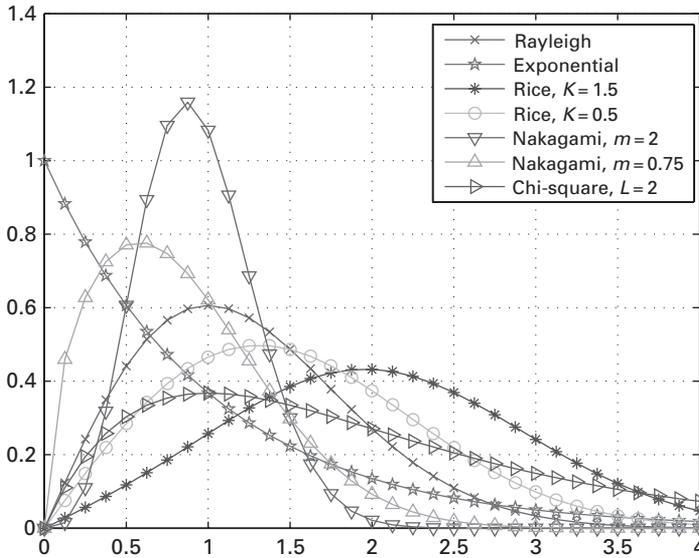
$$I_0(x) = J_0(jx) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} d\theta, \quad (1.21)$$

to get

$$F_r(z) = \frac{e^{-\frac{1}{2}\left(\frac{A}{\sigma}\right)^2}}{\sigma^2} \int_0^z r I_0\left(\frac{rA}{\sigma^2}\right) e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2} u(z) dr. \quad (1.22)$$

From this result, the pdf is obtained by differentiating with respect to  $z$ , resulting in

$$f_r(z) = \frac{z}{\sigma^2} e^{-\left(\frac{z^2}{2\sigma^2} + K\right)} I_0\left(\frac{2Kz}{A}\right), \quad z \geq 0. \quad (1.23)$$



**Fig. 1.12** Different probability density functions used to model random fading.

This is the pdf of a Rician distribution. In (1.23),  $K$  is a parameter of the Rician distribution defined as  $K = A^2/(2\sigma^2)$ . Note that, when  $K = 0$ , the Rician pdf becomes equal to the Rayleigh pdf, which is consistent with the fact that  $K = 0$  means there is no LOS path.

In some cases, it is convenient to model the channel by taking samples of channel realizations and then matching them to a mathematical model. For this it is useful to have a probability density function that can be easily matched to the data samples. This function is provided by the Nakagami fading distribution, which is given by

$$f(x) = \frac{2m^m x^{2m-1}}{\Gamma(m)\sigma^{2m}} e^{-mx^2/\sigma^2}, \quad m \geq 1/2, \quad (1.24)$$

where  $\Gamma(\cdot)$  is the Gamma function and  $m$  is a parameter used to adjust the pdf of the Nakagami distribution to the data samples. For example, if  $m = 1$ , then the Nakagami distribution becomes equal to the Rayleigh distribution. One advantage of the Nakagami distribution is that it matches empirical data better than other distributions. In fact, the Nakagami distribution was originally proposed due to this reason.

Figure 1.12 shows the different probability density functions that were discussed in this section.

## 1.2 Characterizing performance through channel capacity

In this book we will be studying different communication schemes. One important way of characterizing their achievable performance is through the use of information theory concepts, most notably through the use of concepts such as mutual information and the

characterization of performance limits through system capacity. At its core, information theory deals with the information provided by the outcome of a random variable. The information provided by the outcome  $x$  of a discrete random variable  $X$  is defined as

$$I_X(x) = \log \frac{1}{\Pr[X = x]} = -\log \Pr[X = x], \quad (1.25)$$

where  $\Pr[X = x]$  is the probability of the outcome  $X = x$  and the logarithm can be, in principle, of any base but is most frequently taken with base 2, followed by base  $e$  in some fewer cases. Intuitively, the rarer an event is, the more information it provides. Since the communication process is inherently a process relating more than one random variable (e.g. the input and output of a channel, an uncompressed and a compressed representation of a signal, etc.), it is also important to define a magnitude relating the information shared by two random variables. This magnitude is the *mutual information*, which for two discrete random variables  $X$  and  $Y$  is defined as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} \Pr[X = x, Y = y] \log \frac{\Pr[X = x, Y = y]}{\Pr[X = x]\Pr[Y = y]},$$

where  $\Pr[X = x, Y = y]$  is the joint probability mass function and  $\Pr[X = x]$  and  $\Pr[Y = y]$  are marginal probability mass functions. Following Bayes theorem ( $\Pr[X = x, Y = y] = \Pr[X = x|Y = y]\Pr[Y = y]$ , with  $\Pr[X = x|Y = y]$  being the conditional probability mass function of  $X$  given that  $Y = y$ ), the mutual information can also be written as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} \Pr[X = x, Y = y] \log \frac{\Pr[X = x|Y = y]}{\Pr[X = x]}.$$

Furthermore, we can write

$$\begin{aligned} I(X; Y) &= - \sum_{x \in X} \log \Pr[X = x] \sum_{y \in Y} \Pr[X = x, Y = y] \\ &\quad + \sum_{x \in X} \sum_{y \in Y} \Pr[X = x, Y = y] \log \Pr[X = x|Y = y] \\ &= - \sum_{x \in X} \Pr[X = x] \log \Pr[X = x] \\ &\quad + \sum_{x \in X} \sum_{y \in Y} \Pr[X = x, Y = y] \log \Pr[X = x|Y = y]. \end{aligned} \quad (1.26)$$

The first term in this result is called the *entropy* of the random variable  $X$ ,

$$H(X) = - \sum_{x \in X} \Pr[X = x] \log \Pr[X = x],$$

and the second term can be written in terms of the *conditional entropy* of  $X$ ,

$$H(X|Y) = - \sum_{x \in X} \Pr[X = x, Y = y] \log \Pr[X = x|Y = y].$$

Considering (1.25), the entropy of the random variable can also be read as the mean value of the information provided by all its outcomes. Likewise, the conditional entropy

can be regarded as the mean value of the information provided by all the outcomes of a random variable ( $X$ ) given than the outcome of a second random variable ( $Y$ ) is known, or how much uncertainty about a random variable ( $X$ ) remains after knowing the outcome of a second random variable ( $Y$ ). Therefore, the mutual information as in (1.26) can now be written as

$$I(X; Y) = H(X) - H(X|Y),$$

and intuitively be interpreted as the mean amount of uncertainty about one random variable ( $X$ ) that is resolved after learning about the outcome of another random variable ( $Y$ ), or the average amount of information shared by the two random variables. Finally, we note here that although the concepts we have been introducing were tailored to discrete random variables, the same concepts apply to continuous random variables with the only differences that the sums are replaced by integrals and the probability mass functions by probability density functions.

In information theory, one of the main measures of performance is system capacity. Nevertheless, due to the fact that the calculation of capacity always involves a number of assumptions and simplifications, the measurement of capacity does not come in a “one size fits all” solution. In particular, the notion of capacity is influenced by how much the channel changes over the duration of a coding interval and the properties of the random process associated with the channel fluctuations.

When the random variations of the channel are a stationary and ergodic process it is possible to consider the traditional notion of capacity as introduced by Claude Shannon [181]. In this case, coding is assumed to be done using arbitrary long blocks. Also, the random process driving the channel changes needs to be stationary and ergodic. Because of this, this notion of capacity is known as ergodic capacity or Shannon capacity. The capacity of an AWGN channel with fast flat fading, when only the receiver has knowledge of the channel state,

$$C = \mathbb{E} \left[ \log \left( 1 + \frac{|h|^2 P}{N_0} \right) \right], \quad (1.27)$$

where  $\mathbb{E}[\cdot]$  is the expectation operator (operating on the random channel attenuation),  $P$  is the power of the transmitted signal (assumed i.i.d. Gaussian, with zero mean so as to achieve capacity),  $N_0$  is the variance of the background noise, and  $|h|^2$  is the envelope of the channel attenuation.

Although the notion of Shannon capacity is quite useful, there are also many design settings where the assumptions of using arbitrary long codes or that the channel is a stationary and ergodic random process do not hold. In these cases, Shannon capacity may not yield useful results. For example, in the case of a non-ergodic slowly fading channel following a Rayleigh distribution, the Shannon capacity is arbitrary small or zero. This is because the result is affected by those realizations of the channel corresponding to deep fades. Nevertheless, an arbitrary small capacity is not the true depiction of many realizations of the fading process (which is confirmed by the many communications taking place every day under these conditions!). Therefore, for these cases it is more

appropriate to consider the notion of outage capacity. This notion is tied to the concept of an outage event.

There are many ways of defining an outage event but, from an information theory point of view, an outage event is defined as the set of channel realizations that cannot support reliable transmission at a rate  $R$ . In other words, the outage event is the set of channel realizations with an associated capacity less than a transmit rate  $R$ . Considering now that the setup that led to (1.27) corresponds to a non-ergodic channel, the outage condition for a realization of the fading can be written as

$$\log \left( 1 + \frac{|h|^2 P}{N_0} \right) < R. \quad (1.28)$$

From here, the outage probability is calculated as the one associated with the outage event,

$$P_{\text{out}} = \Pr \left[ \log \left( 1 + \frac{|h|^2 P}{N_0} \right) < R \right], \quad (1.29)$$

where  $\Pr[\cdot]$  is the probability operator (once again on the random channel attenuation).

Once we have introduced the concepts of outage event and outage probability, the  $P_{\text{out}}$  outage capacity,  $C_{\text{out}}$ , is defined as the information rate that can be reliably communicated with a probability  $1 - P_{\text{out}}$ , that is

$$\Pr [C \leq C_{\text{out}}] = P_{\text{out}}, \quad (1.30)$$

where  $C$  is the Shannon capacity associated with the channel.

We finally note here that, as mentioned above, the outage probability may be defined differently from (1.29). Another way of defining the outage probability is by considering the event that the received signal to noise ratio is below a threshold. This definition can be related to (1.29) by simple algebraic operations that expose the received signal-to-noise ratio as the random variable, i.e., for the signal-to-noise ratio (SNR)  $\gamma$  and SNR threshold  $\gamma_T$ ,

$$P_{\text{out}} = \Pr [\gamma < 2^R - 1] = \Pr [\gamma < \gamma_T].$$

## 1.3 Orthogonal frequency division multiplexing (OFDM)

In Section 1.1.3 we discussed that when the signal bandwidth is much larger than the channel coherence bandwidth, the channel is frequency selective. We also explained that these channels present such impairments as intersymbol interference, which deform the shape of the transmitted pulse, risking the introduction of detection errors at the receiver. This impairment can be addressed with different techniques. One of these techniques is multicarrier modulation. In multicarrier modulation, the high bandwidth signal to be transmitted is divided over multiple mutually orthogonal signals of a bandwidth small enough such that the channel appears to be non-frequency selective. Different multicarrier modulation techniques may differ based on the choice of orthogonal signals. Among

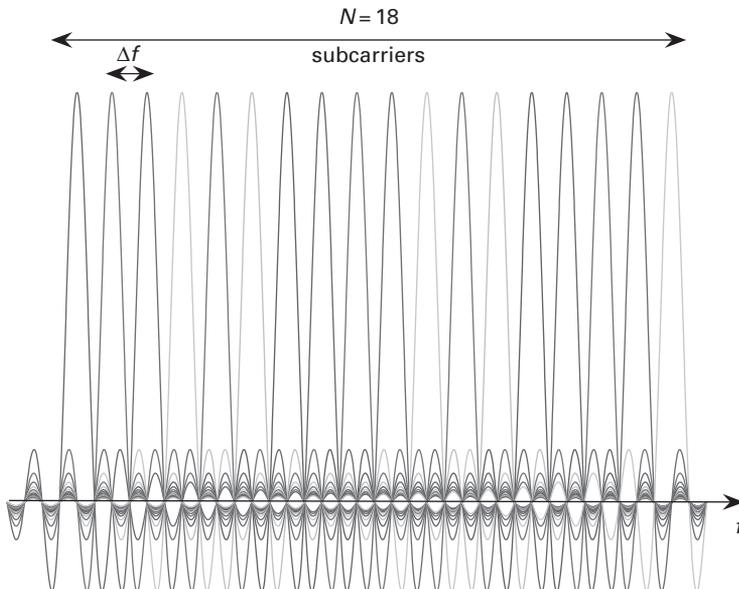
the many possible multicarrier modulation techniques, orthogonal frequency division multiplexing (OFDM) is the one that has gained more acceptance as the modulation technique for high-speed wireless networks and 4G mobile broadband standards. In OFDM, the orthogonal signals used for multicarrier modulation are truncated complex exponentials. Assume an OFDM transmitter where the high rate serial input stream is split into  $N$  parallel substreams. Assume also that, at some instant of time, the sequence  $\{d_k\}_{k=0}^{N-1}$  represents the  $N$  complex symbols that are input to the OFDM modulator for transmission as a single OFDM symbol of duration  $T_s$ . This translates in practice into an operation where the input stream to the OFDM modulator is divided and organized into blocks of  $N$  symbols, which are modulated into a single OFDM symbol. The resulting OFDM modulated symbol is, for  $0 \leq t \leq T_s$ ,

$$s(t) = \sum_{k=0}^{N-1} d_k \phi_k(t) = \sum_{k=0}^{N-1} d_k e^{j2\pi f_k t}, \quad (1.31)$$

where  $f_k = f_0 + k\Delta f$  and  $\Delta f = 1/T_s$ . In (1.31) the signals  $\phi_k(t)$ , which are defined as

$$\phi_k(t) = \begin{cases} e^{j2\pi f_k t} & \text{if } 0 \leq t \leq T_s \\ 0 & \text{else,} \end{cases} \quad (1.32)$$

form an orthonormal set that are used as the carrier signal of each subcarrier in this multicarrier modulation technique. Because these signals are truncated complex exponential, in frequency domain they are of the form  $\sin(x)/x$ . These signals are shown in Figure 1.13, which also illustrates how OFDM splits a carrier with large bandwidth into multiple orthogonal subcarriers of much smaller bandwidth.



**Fig. 1.13** The OFDM orthonormal set of modulation signals in the frequency domain.

Assume, next, that the OFDM symbol in (1.31) is sampled with a period  $T_{sa} = T_s/N$ . Then, we can write the resulting sampled signal  $s[n]$  as

$$s[n] = s(nT_{sa}) = \sum_{k=0}^{N-1} d_k e^{j2\pi f_k n T_s/N}, \quad 0 \leq n \leq N-1.$$

If we assume, without loss of generality, that  $f_0 = 0$  we get  $f_k = k\Delta f = k/T_s$ , leading to

$$s[n] = \sum_{k=0}^{N-1} d_k e^{j2\pi nk/N}. \quad (1.33)$$

This result can be read as  $s[n]$  being the inverse Fourier transform of  $d_k$ , which is a simple way of generating an OFDM symbol and one of its main advantages.

In practice, the OFDM symbol as defined in (1.33) is extended with the addition of a cyclic prefix. To understand the construction of the prefix, assume a multipath channel with  $L$  taps defined through the coefficients  $h[0], h[1], \dots, h[L-1]$ . With this in mind, the original channel input sequence  $s[0], s[1], \dots, s[N-L], s[N-L+1], \dots, s[N-1]$ , becomes  $s[N-L+1], \dots, s[N-1], s[0], s[1], \dots, s[N-L], s[N-L+1], \dots, s[N-1]$  after adding the cyclic prefix. Note that the prefix is built by just copying the last  $L-1$  elements of the original channel input sequence at the beginning of it. This operation does not affect the OFDM signal or its properties, such as the orthogonality between the multicarrier modulated signals, because it is simply a reaffirmation of the periodicity of the OFDM symbol (period equal to  $N$ ), as follows from (1.33). Also note that, following the assumption of a multipath channel with delay spread  $L$ , the samples corresponding to the cyclic prefix will be affected by intersymbol interference from the previous OFDM symbol. To combat this interference, the prefix can be eliminated at the receiver without any loss of information in the original sequence and without intersymbol interference affecting the original sequence. Next, let us illustrate the effect of adding the cyclic prefix on transforming a frequency selective fading channel into a set of parallel flat fading channels.

Let us call the channel input sequence, after adding the cyclic prefix, as  $\mathbf{x}$  where

$$\mathbf{x} = [s[N-L+1], \dots, s[N-1], s[0], s[1], \dots, s[N-L], s[N-L+1], \dots, s[N-1]]. \quad (1.34)$$

The output of the channel can be written as

$$y[n] = \sum_{l=0}^{L-1} h[l]x[n-l] + v[n], \quad n = 1, 2, \dots, N+L-1, \quad (1.35)$$

where  $v[n]$  is additive white Gaussian noise.

The multipath channel affects the first  $L-1$  symbols and therefore the receiver ignores these symbol. The received sequence is then given by

$$\mathbf{y} = [y[L], y[L+1], \dots, y[N+L-1]]. \quad (1.36)$$

Equivalently, one can write the received signal in terms of the original channel input as

$$y[n] = \sum_{l=0}^{L-1} h[l]s[(n-l-L)\bmod N] + v[n]. \quad (1.37)$$

This can be also written in terms of cyclic convolution as follows:

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{s} + \mathbf{v}, \quad (1.38)$$

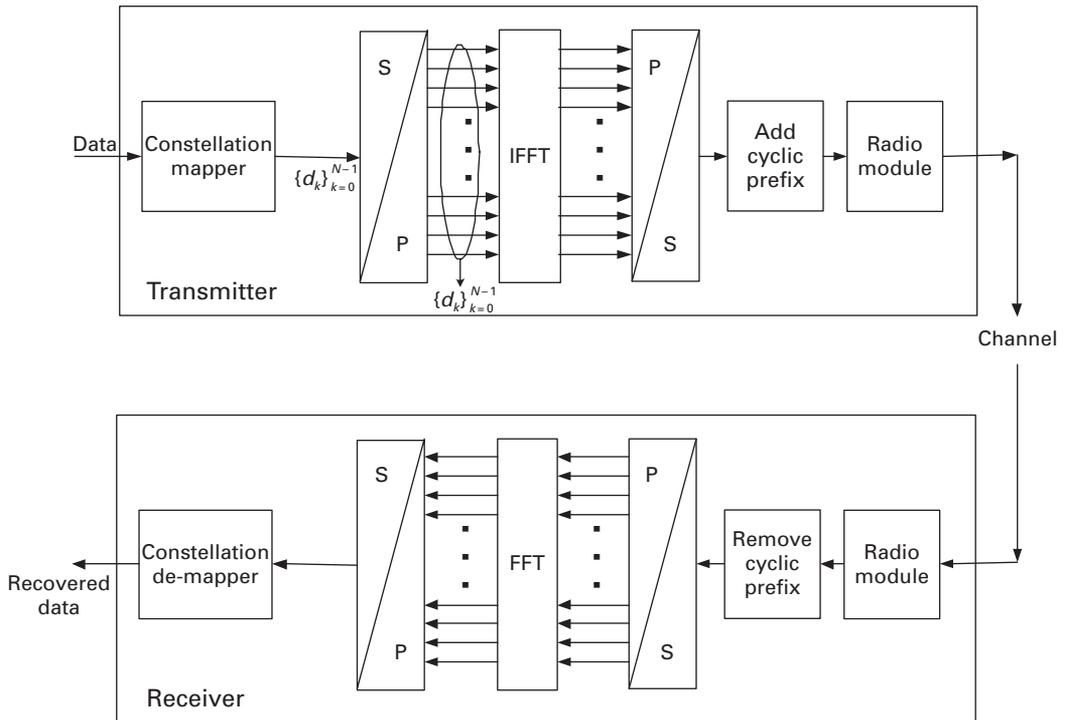
where  $\otimes$  denotes cyclic convolution. At the receiver, after taking the discrete Fourier transform (DFT) of the received signal and after removing the cyclic prefix, we get

$$Y_n = H_n S_n + V_n, \quad (1.39)$$

where  $Y_n$ ,  $H_n$ ,  $S_n$ , and  $V_n$  are the  $n$ -th point of the  $N$ -point DFT of the received signal, channel response, channel input, and noise vector, respectively.

From (1.39), at the receiver side the frequency selective fading channel has been transformed to a set of parallel flat fading channels. Therefore, one can see the benefit of OFDM and how it reduces the complexities associated with time equalization.

Finally, Figure 1.14 summarizes the operations involved in the OFDM communication link by showing a block diagram of a transmitter and a receiver. It is worth highlighting here that an OFDM symbol is made of a block of  $N$  input symbols. At the OFDM transmitter the  $N$  input symbols are converted into an OFDM symbol with



**Fig. 1.14** Block diagram of an OFDM transmitter and a receiver.

$N$  subcarriers. If we now consider the successive transmission of several OFDM symbols, the data organization on the channel can be conceptually pictured as a grid in a *frequency*  $\times$  *time* plane with a width of  $N$  subcarriers (in the frequency dimension) and a depth equal to the number of transmitted OFDM symbols (in the time dimension).

## 1.4 Diversity in wireless channels

As we have explained, fading wireless channels present the challenge of being changing over time. In communication systems designed around a single signal path between source and destination, a crippling fade on this path is a likely event that needs to be addressed with such techniques as increasing the error correcting capability of the channel coding block, reducing the transmission rate, using more elaborate detectors, etc. Nevertheless, these solutions may still fall short for many practical channel realizations.

Viewing the problem of communication through a fading channel with a different perspective, the overall reliability of the link can be significantly improved by providing more than one signal path between source and destination, each exhibiting a fading process as much independent from the others as possible. In this way, the chance that there is at least one sufficiently strong path is improved. Those techniques that aim at providing multiple, ideally independent, signal paths are collectively known as *diversity techniques*. In its simplest form, akin to repetition coding (where signal redundancy is achieved by simply repeating the signal symbols multiple times), the multiple paths may carry multiple distorted copies of the original message. Nevertheless, better performance may be achieved by applying some kind of coding across the signals sent over the multiple paths and by combining in a constructive way the signals received through the multiple paths.

Also important is the processing performed at the receiver, where the signals arriving through the multiple paths are constructively combined. The goal in combining is to process the multiple received signals so as to obtain a resulting signal of better quality or with better probability of successful reception than each of the received ones. The nature of the processing that is applied to each signal during combining is a function of the particular design goals. If the goal is to linearly combine the signals so that the signal-to-noise ratio (SNR) is maximized at the resulting signal, then the resulting mechanism is called a *maximal ratio combiner* (MRC). Suppose that at the input of the MRC there are  $L$  signal samples,  $y_0, y_1, \dots, y_{L-1}$ , that are to be combined into a signal sample  $y_M$ . Each of the received signals correspond to a unit-energy transmitted signal that have been received through the corresponding  $L$  different paths characterized as  $h_0 e^{j\phi_0}, h_1 e^{j\phi_1}, \dots, h_{L-1} e^{j\phi_{L-1}}$ . Since the MRC is a linear combiner, the input and output are related through the relation

$$y_M = \sum_{k=0}^{L-1} c_k e^{-j\phi_k} y_k, \quad (1.40)$$

where  $c_k$  are the coefficients of the MRC combiner and the complex exponential is used for equalizing the phases of each term (cophasing). Assuming that the signals to be combined are equally affected by noise at the receiver with power density  $N_0$ , the SNR at the output of the MRC,  $\gamma_M$  is

$$\gamma_M = \frac{\left( \sum_{k=0}^{L-1} c_k h_k \right)^2}{N_0 \sum_{k=0}^{L-1} c_k^2}, \quad (1.41)$$

because the noise is also processed as part of the received signal samples. The MRC coefficients that maximize (1.41) also maximize its numerator. Then, the maximizing coefficients can be found by using the Cauchy–Schwarz inequality,

$$\left( \sum_{k=0}^{L-1} c_k h_k \right)^2 \leq \left( \sum_{k=0}^{L-1} c_k^2 \right) \left( \sum_{k=0}^{L-1} h_k^2 \right). \quad (1.42)$$

The SNR in (1.41) is maximized when (1.42) is an equality. This is achieved by letting

$$c_k = \frac{h_k}{\sqrt{N_0}}.$$

The resulting maximized SNR at the output of the MRC is

$$\gamma_M = \frac{\left( \sum_{k=0}^{L-1} h_k^2 \right)}{N_0}. \quad (1.43)$$

Intuitively, the MRC combines multiple signals by first cophasing them, followed by weighting each sample proportionally to the corresponding path SNR and finally adding them. The resulting signal at the output of the MRC will have an SNR equal to the sum of the SNRs corresponding to each path.

As mentioned earlier, the MRC is not the only known combiner. Other cases are the *selection combiner*, where the output is the input with best SNR, and the *threshold combiner*, which sequentially scans the received signals and outputs the first one with SNR exceeding a threshold.

For any diversity technique, the performance improvement is manifested by the communication error probability decreasing at a much larger rate at a high channel signal-to-noise ratio (SNR) than systems with less or no diversity. When using log–log scales, this rate of decrease in the communication error probability becomes the slope of the line representing the communication error probability at high SNR and is known as the *diversity gain*. Strictly speaking, the diversity gain is defined as [239]

$$m = - \lim_{\gamma \rightarrow \infty} \frac{\log P_{\text{SER}}}{\log \gamma}, \quad (1.44)$$

where  $\gamma$  is the SNR and  $P_{\text{SER}}$  is the probability of symbol error (a function of the SNR). This definition establishes an implicit behavior at high SNR for the probability

of symbol error as being a linear function of the SNR when seen in a plot with log–log scales. Then it can be seen that, as previously stated, in these conditions the diversity gain is the slope of the linear relation. It is better to have as large a diversity gain as possible, since it means that the probability of symbol error is reduced at a faster rate.

For MIMO systems, as will be shown in Chapters 2 and 3, it is possible to achieve a diversity gain equal to the product between the number of transmit and receive antennas. Also, it is important to note that, depending on the particular diversity scheme and the system setup, other measures of probability of error can be used. For example, the outage probability is used in some cases, instead of the probability of symbol error.

There are many different forms of diversity in addition to the spatial diversity mentioned above. For example, in time diversity, multiple (possibly coded) copies of a symbol are sent at different time instants, and in frequency diversity, multiple (possibly coded) copies of a symbol are sent through channels of different carrier frequency. Furthermore, multiple diversity techniques can be combined to provide even greater performance improvement. Next, a succinct introduction to time, frequency, and antenna diversity systems is provided. Some of the following chapters will consider techniques that are derived from or that combine these forms of diversity.

### 1.4.1 Time diversity

It is quite common to find communication scenarios where the channel coherence time equals or exceeds several symbol transmission periods. This implies that two symbols transmitted with a separation in time longer than the coherence time will experience channel realizations that are highly uncorrelated and can be used to obtain diversity. The simplest way to achieve this is to form the two symbols by using a repetition coding scheme. Also, to guarantee that the repeated symbols will be transmitted over uncorrelated channel realization, an appropriate interleaver is applied to the stream of symbols to be transmitted.

At the receiver, the copies of the symbol will have to be combined together. As explained above, if the transmission of each symbol can be represented by an input–output expression of the form

$$y_i = h_i x + n_i, \quad (1.45)$$

where  $x$  is the unit-energy transmitted symbol,  $y_i$  is the symbol received over path  $i$ ,  $n_i$  is the background noise modeled as a circularly symmetric Gaussian random variable with zero mean and variance  $N_0$ , and  $h_i$  is the channel realization over path  $i$ , assumed to be following a Rayleigh fading, optimal combining that maximizes the received SNR is achieved with a maximal ratio combiner. Recall that the SNR at the output of an MRC equals the sum of the SNRs of the branches at the input of the MRC. From the explanation above on Rayleigh fading, the MRC output SNR will have a Chi-squared distribution as in (1.14).

If, for example, we assume BPSK modulation, and  $M$  copies of the transmitted symbol being combined with path channel gains,  $h_1, h_2, \dots, h_M$ , the error probability that can be obtained using MRC combiner is

$$\mathcal{Q}\left(\sqrt{\frac{2}{N_0} \sum_{n=1}^M |h_n|^2}\right).$$

From here, the average probability of symbol error is [45]

$$P_{\text{SER}} = \int_0^\infty \mathcal{Q}(\sqrt{2\gamma}) f(\gamma) d\gamma \quad (1.46)$$

$$= \left(\frac{1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}}{2}\right)^M \sum_{m=0}^{M-1} \binom{M+m-1}{m} \left(\frac{1 + \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}}}{2}\right)^m, \quad (1.47)$$

where  $\gamma = (1/N_0) \sum_{n=1}^M |h_n|^2$  is the SNR at the output of the MRC,  $f(\gamma)$  is the probability density function of the Chi-squared distribution, as in (1.14), and  $\bar{\gamma}$  is the mean SNR at the output of the MRC.

**Example 1.3** Although it is possible to derive useful results from (1.47), in the study of systems with some form of diversity (and, in fact, in many other communication systems also), it is sometimes better to work with upper bounds derived from the Chernoff bound. The Chernoff bound is a useful inequality in applications of probability and random processes to signal processing problems that establishes an upper bound on the tail probability of a random variable  $X$ ,

$$\Pr[X \geq a] \leq \min_t \{e^{-at} E[e^{tX}]\}, \quad (1.48)$$

where  $a$  is a constant. When  $X$  is a standard Gaussian random variable  $\mathcal{N}(0, 1)$  (zero mean, unit variance), the Chernoff bound becomes

$$\Pr[X \geq a] \leq \min_t \{e^{(-at+t^2/2)}\} = e^{-a^2/2}, \quad (1.49)$$

where the minimization is done by simply equating the derivative of  $e^{(-at+t^2/2)}$  to zero. When applying the Chernoff bound to the  $\mathcal{Q}$  function in (1.46), we obtain the upper bound on average probability [45]

$$P_{\text{SER}} \leq \prod_{m=1}^M \frac{1}{1 + \frac{\bar{\gamma}}{2}}. \quad (1.50)$$

At large SNR, this upper bound becomes tighter, leading to the approximation

$$P_{\text{SER}} \approx \left(\frac{\bar{\gamma}}{2}\right)^{-M}. \quad (1.51)$$

▲

The result (1.51) is revealing in presenting the physical meaning of diversity gain. If now, we look at this expression using log–log scales we get

$$\log(P_{\text{SER}}) \approx -M \log(\bar{\gamma}), \quad (1.52)$$

which is the equation of a linear function. Furthermore, notice that by applying the definition of diversity gain (1.44), it can be seen not only that the diversity gain equals  $-M$  (the number of repetitions of the symbol), but also that the diversity gain represents the slope of the approximately linear relation (1.52). Since the diversity gain equals the number of repetitions of the symbol, we may say that the time-diversity system with repetition coding achieves full diversity gain. Nevertheless, the use of repetition coding sacrifices the total bit rate. This drawback can be addressed through other coding schemes as will be seen in later chapters.

### 1.4.2 Frequency diversity

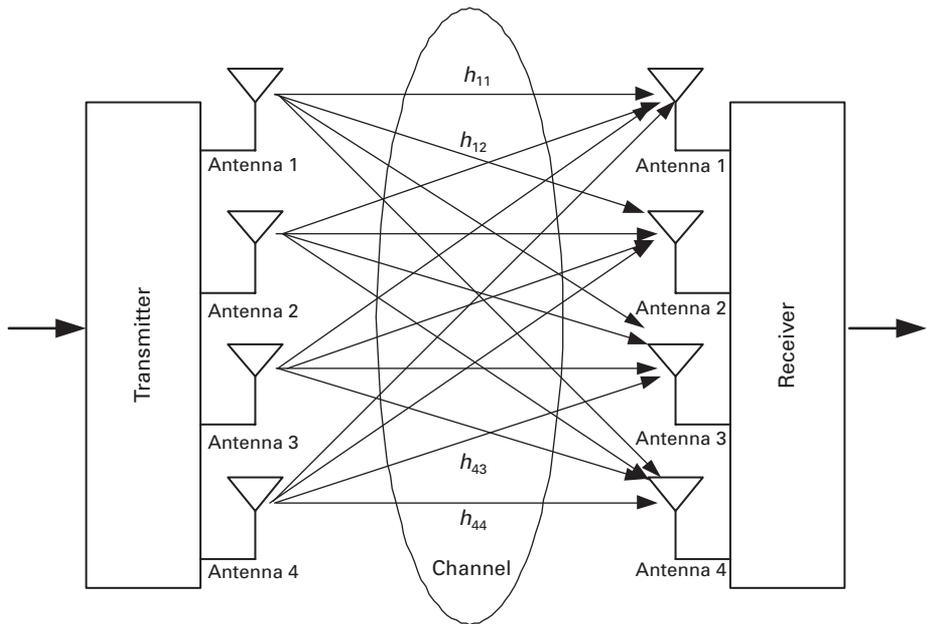
Analogous to time diversity, in those wideband systems where the available bandwidth exceeds the channel coherence bandwidth, it is possible to realize diversity by using channels that are a partition of the available bandwidth and that are separated by more than the channel coherence bandwidth.

Realizing frequency diversity as a partition of the whole system bandwidth into channels with smaller bandwidth and independent frequency response is perhaps the most intuitively natural approach. This approach is applicable in multicarrier systems, where transmission is implemented by dividing the wideband channel into non-overlapping narrowband subchannels. The symbol used for transmission in each subchannel has a transmission period long enough for the subchannel to appear as a flat fading channel. Different subchannels are used together to achieve frequency diversity by ensuring that each is separated in the frequency domain from the rest of the subchannels in the transmission by more than the coherence bandwidth. In this way, the fading processes among the subchannels will show a small cross-correlation. As will be seen in later chapters, an example of these systems are those using orthogonal frequency division multiplexing (OFDM).

Although not as intuitively natural, frequency diversity can also be achieved through processing based on a time-domain phenomenon. Recall that the frequency response of multipath channels is not of constant amplitude and linear phase because each spectral component of the signal undergoes destructive or constructive interference of different magnitude depending on the delay of each path and the frequency of the spectral component. These multipath channels provide diversity through each of the copies of the signal arriving through each path. Because of this, the overall channel appears as frequency selective (see Figures 1.5 and 1.6). It is then possible to achieve diversity of an order equal to the number of independent paths.

### 1.4.3 MIMO systems

By using multiple antennas at the transmitter side and/or the receiver side as shown in Figure 1.15, we may exploit diversity in the spatial domain which is called spatial diversity (also called antenna diversity). This configuration of deploying multiple antennas is



**Fig. 1.15** A four-transmit, four-receive MIMO system.

often referred as multiple-input-single-output (MISO) systems if only a single antenna is deployed at the receiver side, single-input-multiple-output (SIMO) systems if a single transmit antenna is used, or, in general, multiple-input-multiple-output (MIMO) systems with multiple transmit antennas and multiple receive antennas. With more than one transmit/receive antenna, different channels are established between each pair of transmit and receive antennas. With such a configuration, the transmitted information can go through different channels to arrive at the receiver side. As long as one of the channels is strong enough, the receiver should be able to recover the transmitted information. If we assume that different channels are independent or correlated with a low correlation, then the chance that all channel links fail is low. The greater the number of antenna pairs, the greater the redundancy (diversity) of the received signals, i.e., the higher the reliability of the transceiver detection. The assumption of low-correlated or independent channel links can be achieved by appropriate separation of the antennas at both the transceiver sides. The necessary antenna separation at each side depends on the scattering in the neighborhood of the antenna and on the signal carrier frequency. For a mobile, the typical separation is between half to one carrier wavelength, for base stations the necessary separation is in the order of tens of wavelengths, which can be easily satisfied.

#### 1.4.3.1 Two motivated examples

Note that, in a MISO system, the signal present at the receive antenna is the combination of signals from all transmit antennas after having traveled through the different fading channels established from each transmit antenna to the receive antenna. The redundancy is also termed as transmit diversity, which depends on the number of transmit

antennas. While in a SIMO system, a transmitted signal goes through different channels and is received at each receive antenna. Signals from each receive antenna are combined and jointly detected at the receiver side. The corresponding redundancy is often called receive diversity, which is related to the number of receive antennas. In general, in a MIMO system, both transmit and receive diversities are realized. As we will discuss in Chapter 2, the overall signal redundancy, or diversity order, is in this case the product of the numbers of transmit and receive antennas. The MIMO configuration can be exploited through different designs that differ, among other factors, in the antenna configuration at the receiver and the transmitter, as well as the particular form of performance improvement that it is intended to obtain. For better understanding the spatial diversity, we consider the following two motivated examples.

**Example 1.4** Consider a system that implements receive diversity by using one antenna at the transmitter and  $M_r$  antennas at the receiver (i.e., a SIMO system). There are  $M_r$  paths between transmitter and receiver. The signal arriving from all paths need to be combined at the receiver. If the additive background noise is complex-valued, zero-mean circularly symmetric Gaussian, and independent from each path, then the optimal combiner is the maximal ratio combiner (MRC) (a fact that can be accepted after realizing the similarity of the combining problem here and the one in diversity from repetition coding over time, with the diversity branches now over space instead of over time). Assuming BPSK modulation, and conditioning on the  $M_r$  channel gains of the paths,  $h_1, h_2, \dots, h_{M_r}$ , the error probability that can be obtained using MRC combiner is  $\mathcal{Q}\left(\sqrt{2\gamma} \sum_{j=1}^{M_r} |h_j|^2\right)$ . This error probability can be written in the following form, which provides more insight into the achievable gains,

$$P_e = \mathcal{Q}\left(\sqrt{2(M_r\gamma) \left(\frac{\sum_{j=1}^{M_r} |h_j|^2}{M_r}\right)}\right).$$

In this expression, the factor  $M_r\gamma$  shows that the use of the MRC results in a linear increase of SNR with the number of paths. This gain is called the array gain. Also, the factor  $(1/M_r) \sum_{j=1}^{M_r} |h_j|^2$  shows an averaging effect on the paths gain where the paths in deeper fade are compensated by those in a good condition, thus resulting in a lower probability that the overall link attenuation will be too large. This effect is the materialization of the diversity gain in this scheme. To see this more clearly, consider that each  $h_j$  is the sum of two i.i.d. real Gaussian random variables, and recall from Section 1.1.3 that  $\sum_{j=1}^{M_r} |h_j|^2$  follows a Chi-square distribution with  $2M_r$  degrees of freedom (see (1.14)). Integrating over this distribution to obtain the average probability of symbol error results in

$$P_{\text{SER}} = \left(\frac{1 - \sqrt{\frac{\gamma}{1+\gamma}}}{2}\right)^{M_r} \sum_{m=0}^{M_r-1} \binom{M_r+m-1}{m} \left(\frac{1 + \sqrt{\frac{\gamma}{1+\gamma}}}{2}\right)^m. \quad (1.53)$$

Unsurprisingly, this expression is very similar to (1.47). This is due to the analogy between achieving diversity by transmitting at different time instants and by transmitting through different paths. The essence of this similarity resides in the common use of an MRC combiner, which results in a similar expression for the probability of error. Furthermore, we can now draw on this similarity and apply the Chernoff bound to derive expressions similar to (1.50) and (1.51) to get

$$P_{\text{SER}} \approx \alpha_{M_r} \left( \frac{\beta_{M_r} \gamma}{2} \right)^{-M_r},$$

which shows the diversity order  $M_r$ , since the SNR decays as  $\gamma^{-M_r}$ . ▲

**Example 1.5** (Alamouti scheme) Consider now a system that implements transmit diversity, with two antennas at the transmitter and one antenna at the receiver. Transmission is done by sending two symbols,  $s_1$  and  $s_2$ , over the duration of two symbol periods. We assume a flat fading channel, i.e., the channel does not change during the two symbol periods. Instead of sending the symbol  $s_1$  during the first symbol period and the symbol  $s_2$  during the second period, the Alamouti scheme treats the two symbols as a block and applies a form of precoding to the symbols. Specifically, during the first transmission period,  $s_1$  is sent from antenna 1 and  $s_2$  is sent from antenna 2; and during the second transmission period,  $-s_2^*$  is sent from antenna 1 and  $s_1^*$  from antenna 2 (here  $s_1^*$  denotes the complex conjugate of  $s_1$ ).

At the receiver, we denote by  $y_1$  and  $y_2$  the signals received during the first and second symbol periods, respectively. The two received signals are processed as a vector given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

The above transceiver signals can be rearranged as

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2^* \end{bmatrix}.$$

Denote  $\mathbf{y} = [y_1 \ y_2^*]^T$  and

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}.$$

Then the receiver processes the received vector  $\mathbf{y}$  by taking  $\tilde{\mathbf{y}} = \mathbf{H}^H \mathbf{y}$ , resulting in

$$\tilde{\mathbf{y}} = \begin{bmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_1^2| + |h_2^2| \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \mathbf{H}^H \begin{bmatrix} z_1 \\ z_2^* \end{bmatrix}.$$

Since the processed noise vector is still complex Gaussian with zero mean, but now with a covariance matrix equal to the diagonal matrix  $\text{diag}(|h_1^2| + |h_2^2|, |h_1^2| + |h_2^2|)$  times the received noise power, the received SNR for a symbol detected with  $\mathbf{z}$  is

$$\gamma = \frac{|h_1^2| + |h_2^2|}{2} \gamma_S,$$

where  $\gamma_S$  is the SNR of a symbol transmitted without the Alamouti scheme and the factor 2 is due to the fact that each  $s_i$  is transmitted at half  $\gamma_S$ . We finally note here that this result shows in a fashion similar to Example 1.4 that the Alamouti scheme can achieve a diversity order 2.  $\blacktriangle$

### 1.4.3.2 MIMO capacity

In this subsection, we address advantages of MIMO systems from an information theory aspect by reviewing a fundamental capacity result indicating that the capacity of a MIMO system increases at least linearly with the minimum number of transmit or receive antennas.

We consider a MIMO system with  $M_t$  transmit antennas and  $M_r$  receive antennas. The channel coefficient between transmit antenna  $i$ ,  $1 \leq i \leq M_t$ , and receive antenna  $j$ ,  $1 \leq j \leq M_r$ , is denoted by  $h_{i,j}$ . These coefficients are modeled as independent circularly symmetric complex Gaussian random variables with zero mean and variance one. The MIMO transceiver can be modeled as

$$Y = \sqrt{\frac{\rho}{M_t}} XH + Z, \quad (1.54)$$

where  $X = [x_1 \ x_2 \ \dots \ x_{M_t}]$  is a signal vector transmitted by the  $M_t$  antennas,  $Y = [y_1 \ y_2 \ \dots \ y_{M_r}]$  is a signal vector received by the  $M_r$  receive antennas, and  $Z = [z_1 \ z_2 \ \dots \ z_{M_r}]$  is a noise vector whose elements are modeled as independent circularly symmetric complex Gaussian random variables with zero mean and variance one. The channel coefficient matrix  $H = \{h_{i,j} : 1 \leq i \leq M_t, 1 \leq j \leq M_r\}$  is assumed to be known at the receiver side, but unknown at the transmitter side. The signal vector is assumed to satisfy the energy constraint  $E\|X\|_F^2 = M_t$ , where  $\|X\|_F$  is the Frobenius norm of  $X$ , defined as

$$\|X\|_F^2 = \sum_{i=1}^{M_t} |x_i|^2.$$

In (1.54), the factor  $\sqrt{\rho/M_t}$  ensures that  $\rho$  is the average signal to noise ratio (SNR) at each receive antenna, and it is independent of the number of transmit antennas.

If the input signal  $X$  is a circularly symmetric complex Gaussian random vector with zero mean and variance  $E\{X^H X\} = Q$ , then the output signal is also a circularly symmetric complex Gaussian random vector with zero mean and variance

$$E\{Y^H Y\} = \frac{\rho}{M_t} H^H Q H + I_{M_r},$$

in which  $I_{M_r}$  is an identity matrix of size  $M_r$  by  $M_r$ . So, for any given channel  $H$ , the mutual information between the input  $X$  and the output  $Y$  is

$$I(X; Y|H) = \log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} H^H Q H \right). \quad (1.55)$$

Since  $E\|X\|_{\mathbb{F}}^2 = M_t$ ,  $\text{trace}(Q) = \text{trace}(E\{X^H X\}) = M_t$ . Next, we would like to maximize the mutual information  $I(X; Y|H)$  as in (1.55) over the choice of nonnegative  $Q$  with the constraint  $\text{trace}(Q) = M_t$ . Since we can write  $Q$  as  $Q = U^H Q_0 U$ , where  $U$  is unitary and  $Q_0$  is diagonal, so

$$\log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} H^H Q H \right) = \log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} \tilde{H}^H Q_0 \tilde{H} \right), \quad (1.56)$$

in which  $\tilde{H} = UH$ . Note that  $\tilde{H}$  has the same distribution as that of  $H$  since  $U$  is unitary. So we just need to maximize

$$\log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} \tilde{H}^H Q_0 \tilde{H} \right)$$

over the nonnegative diagonal matrix  $Q_0$  with the constraint  $\text{trace}(Q_0) = M_t$ . It has been shown in [215] that

$$\log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} \tilde{H}^H Q_0 \tilde{H} \right)$$

is maximized when  $Q_0$  has equal diagonal elements, i.e.,  $Q_0 = I_{M_t}$ . Intuitively, it is easy to understand that if the transmitter has no prior information about the channel, each transmit antenna should be treated equally and allocated the same weight, i.e., the variance of the input signal vector over  $M_t$  transmit antennas should be an identity matrix. Therefore, the maximal mutual information is

$$\log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} \tilde{H}^H \tilde{H} \right).$$

Finally, we assume that the channel is memoryless, i.e., the channel  $H$  changes independently from each use of the channel to another. Thus the average capacity of the MIMO system is given by

$$C = E_H \left\{ \log_2 \det \left( I_{M_r} + \frac{\rho}{M_t} H^H H \right) \right\}, \quad (1.57)$$

in which the expectation is taken over the fading channel  $H$ . We can see that when  $M_t = M_r = 1$ , the above result reduces to the capacity in (1.27) for a conventional single-input-single-output (SISO) system.

In the following, we further interpret the capacity result in (1.57) for SIMO, MISO, and MIMO systems, respectively. For a SIMO system, i.e.,  $M_t = 1$  and  $M_r > 1$ , the channel  $H$  is a vector as  $H = [h_{1,1} \ h_{1,2} \ \dots \ h_{1,M_r}]$ . Since

$$\begin{aligned} \det \left( I_{M_r} + \rho H^H H \right) &= \det \left( I_{M_t} + \rho H H^H \right) \\ &= 1 + \rho \sum_{j=1}^{M_r} |h_{1,j}|^2, \end{aligned}$$

in which the first equality follows from the determinant identity  $\det(I_m + AB) = \det(I_n + BA)$  for any matrices  $A$  and  $B$  of sizes  $m \times n$  and  $n \times m$  respectively. Therefore, the capacity of the SIMO system is

$$C = E_H \left\{ \log_2 \left( 1 + \rho \sum_{j=1}^{M_r} |h_{1,j}|^2 \right) \right\}, \quad (1.58)$$

in which  $\sum_{j=1}^{M_r} |h_{1,j}|^2$  is a Chi-square random variable with  $2M_r$  degrees of freedom, compared with the SISO case where  $C = E_h \{ \log_2 (1 + \rho |h|^2) \}$  and  $|h|^2$  is a Chi-square random variable with two degrees of freedom.

For a MISO system, i.e.,  $M_r = 1$  and  $M_t > 1$ , the channel  $H$  is a column vector as  $H = [h_{1,1} \ h_{1,2} \ \dots \ h_{M_t,1}]^T$ . In this case,

$$\det \left( I_{M_r} + \frac{\rho}{M_t} H^H H \right) = 1 + \frac{\rho}{M_t} \sum_{i=1}^{M_t} |h_{i,1}|^2.$$

So the corresponding capacity can be specified as

$$C = E_H \left\{ \log_2 \left( 1 + \rho \frac{\sum_{i=1}^{M_t} |h_{i,1}|^2}{M_t} \right) \right\}. \quad (1.59)$$

We can see that when  $M_t$  is large,  $(\sum_{i=1}^{M_t} |h_{i,1}|^2) / M_t \approx E\{|h|^2\}$ , in which  $|h|^2$  is a Chi-square random variable with two degrees of freedom. Thus, the capacity in (1.59) is almost the same as that of the SISO system.

For a MIMO system, without loss of generality, we assume  $M_t = M_r > 1$ . Note that

$$H^H H = \begin{bmatrix} \sum_{i=1}^{M_t} |h_{i,1}|^2 & \sum_{i=1}^{M_t} h_{i,1}^* h_{i,2} & \cdots & \sum_{i=1}^{M_t} h_{i,1}^* h_{i,M_r} \\ h_{i,2}^* h_{i,1} & \sum_{i=1}^{M_t} |h_{i,2}|^2 & \cdots & \sum_{i=1}^{M_t} h_{i,2}^* h_{i,M_r} \\ \vdots & \vdots & \ddots & \vdots \\ h_{i,M_r}^* h_{i,1} & \sum_{i=1}^{M_t} h_{i,M_r}^* h_{i,2} & \cdots & \sum_{i=1}^{M_t} |h_{i,M_r}|^2 \end{bmatrix}.$$

Since  $h_{i,j}$  are i.i.d. complex Gaussian random variables with zero mean and variance one, so for large  $M_t$ ,  $(\sum_{i=1}^{M_t} |h_{i,j}|^2) / M_t \rightarrow 1$  for each  $1 \leq j \leq M_r$  and  $(\sum_{i=1}^{M_t} h_{i,j_1}^* h_{i,j_2}) / M_t \rightarrow 0$  for any  $1 \leq j_1 \neq j_2 \leq M_r$ . Thus, for large  $M_t$ ,  $(1/M_t) H^H H \rightarrow I_{M_r}$ . Therefore, for large  $M_t$ , the capacity in (1.57) is

$$\begin{aligned} C &\rightarrow \log_2 \det (I_{M_r} + \rho I_{M_r}) \\ &= M_r \log_2 (1 + \rho), \end{aligned} \quad (1.60)$$

which increases linearly with the number of receive antennas. Note that the above discussion is also true for any  $M_t \geq M_r$ . Based on the above discussion for SIMO, MISO, and MIMO systems, we can see that the capacity of a multiple-antenna system increases at least linearly with the minimum number of transmit or receive antennas.

### 1.4.3.3 Diversity–multiplexing tradeoff

In MIMO systems, the multiple paths created between any pair of transmit–receive antennas can be used to obtain diversity gain. On the other hand, these paths can also be used to transmit independent messages from each transmit antenna, in which case it is possible to achieve an increase in transmit bit rate given by a *multiplexing gain*.

At the receiver, the MIMO configuration allows for separation of each data stream. It is readily apparent that there should be a tradeoff between diversity and multiplexing gains because the later is achieved at the expense of signal paths that otherwise could be used to increase the former.

The diversity–multiplexing tradeoff is specified through the choice of an achievable combination of diversity and multiplexing gains, or, in other words, by specifying the achievable diversity gain as a function of the multiplexing gain. Let  $d^*(r)$  be the diversity–multiplexing tradeoff curve for the slow fading channel. A point on this curve, a diversity gain  $d^*(r)$ , is achieved at multiplexing gain  $r$  if

$$d^*(r) = - \lim_{\gamma \rightarrow \infty} \frac{\log P_{\text{out}}(r \log \gamma)}{\log \gamma}. \quad (1.61)$$

This definition means that when communicating at a rate  $R = r \log \gamma$ , the achievable diversity gain is that for which the outage probability decays as  $P_{\text{out}}(r \log \gamma) \approx \gamma^{-d^*(r)}$  at arbitrary large SNR. Also, this formulation can be extended to any type of fading channel, beyond the slow fading channel, by replacing the outage probability with the probability of error, i.e.,

$$d^*(r) = - \lim_{\gamma \rightarrow \infty} \frac{\log P_e(r \log \gamma)}{\log \gamma}. \quad (1.62)$$

---

**Example 1.6** Consider a system that transmits a single symbol using QAM modulation at SNR  $\gamma$  over a channel with unit-average power Rayleigh fading and complex-valued, circularly symmetric additive Gaussian background noise with zero mean and unit power. At the receiver, the detection performance is driven by the minimum distance between constellation points  $d_{\text{min}}$ , which at high SNR is approximately given by [219]

$$d_{\text{min}} \approx \frac{\sqrt{\gamma}}{2^{R/2}}.$$

This leads to an error probability at high SNR approximately equal to

$$P_e \approx \frac{1}{d_{\text{min}}^2} = \frac{2^R}{\gamma},$$

which results in the diversity–multiplexing tradeoff

$$d(r) = 1 - r,$$

for  $r$  between 0 and 1. ▲

---

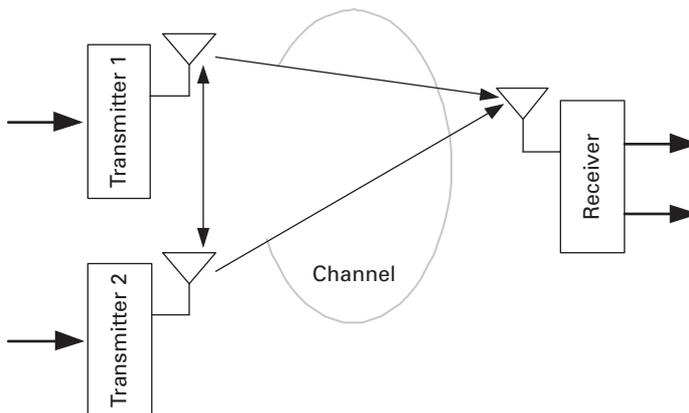
## 1.5 Cooperation diversity

The proliferation of wireless communication applications in the last few years is unprecedented. Voice communication is no longer the only application people need.

High data rate applications, wireless broadband Internet, gaming, and many other applications have emerged recently. Most future wireless systems such as ultra mobile broadband (UMB), Long Term Evolution (LTE), and IEEE 802.16e (WiMAX) promise very high data rates per user over high bandwidth channels (5, 10, and 20 MHz). For example, in the fourth generation wireless networks to be deployed in the next couple of years, namely, mobile broadband wireless access (MBWA) or IEEE 802.20, peak data rates of 260 Mbps can be achieved on the downlink, and 60 Mbps on the uplink [80]. These data rates can, however, only be achieved for full-rank MIMO users. More specifically, full-rank MIMO users must have multiple antennas at the mobile terminal, and these antennas must see independent channel fades to the multiple antennas located at the base station. In practice, not all users can guarantee such high rates because they either do not have multiple antennas installed on their small-size devices, or the propagation environment cannot support MIMO because, for example, there is not enough scattering. In the later case, even if the user has multiple antennas installed, full-rank MIMO is not achieved because the paths between several antenna elements are highly correlated.

To overcome the above limitations of achieving MIMO gains in future wireless networks, we must think of new techniques beyond traditional point-to-point communications. The traditional view of a wireless system is that it is a set of nodes trying to communicate with each other. From another point of view, however, because of the broadcast nature of the wireless channel, we can think of those nodes as a set of antennas distributed in the wireless system. Adopting this point of view, nodes in the network can cooperate together for distributed transmission and processing of information. The cooperating node acts as a relay node for the source node.

Cooperative communications is a new communication paradigm which generates independent paths between the user and the base station by introducing a relay channel. The relay channel can be thought of as an auxiliary channel to the direct channel



**Fig. 1.16** Two transmitters associated in a user-cooperative configuration.

between the source and destination. Since the relay node is usually several wavelengths distant from the source, the relay channel is guaranteed to fade independently from the direct channel, which introduces a full-rank MIMO channel between the source and the destination. In the cooperative communications setup, there are a-priori few constraints to different nodes receiving useful energy that has been emitted by another transmitting node. The new paradigm in user cooperation is that, by implementing the appropriate signal processing algorithms at the nodes, multiple terminals can process the transmissions overheard from other nodes and be made to collaborate by relaying information for each other (Figure 1.16). The relayed information is subsequently combined at a destination node so as to create spatial diversity. This creates a network that can be regarded as a system implementing a distributed multiple antenna where collaborating nodes create diverse signal paths for each other.

Hence, cooperative communications is a new paradigm shift for the fourth generation wireless system that will guarantee high data rates to all users in the network, and we anticipate that it will be the key technology aspect in fifth generation wireless networks.

In terms of research “ascendance,” cooperative communications can be seen as related to research in relay channel and MIMO systems. The concept of user cooperation itself was introduced in two-part series of papers [179, 180]. In these works, Sendonaris *et al.* proposed a two user cooperation system, in which pairs of terminals in the wireless network are coupled to help each other forming a distributed two-antenna system. Parts II and III of this book will study in detail the design and analysis of cooperative communications, so we defer further explanation until then.

## 1.6 Bibliographical notes

Because this is an introductory chapter, we have not covered in detail any of the topics studied here. Nevertheless, there are plenty of excellent textbooks and research papers that complement our presentation. On the topic of wireless channels, the reader can find more information in the books by Proakis [146], Rappaport [150], Tse and Viswanath [219], and Goldsmith [45]. For a practical presentation, the description of channel models used in different standards such as [87] are always a good reference. The books by Tse and Viswanath [219] and by Goldsmith [45] were the main sources for most of the topics covered in this introduction and are where further explanations can be found. The book by Cover and Thomas [26] presents very good in-depth study of many topics in information theory. Also, in-depth study of the concept of outage capacity can be found in [141]. Extra coverage on the topic of diversity can be found in books such as [91], tutorial papers such as [144], or books covering communications over fading channels [189]. Finally, the topic of diversity–multiplexing tradeoff has been studied in [239] and related papers.

# 2 Space–time diversity and coding

---

The idea of using multiple transmit and receive antennas in wireless communication systems has attracted considerable attention with the aim of increasing data transmission rate and system capacity. A key issue is how to develop proper transmission techniques to exploit all of the diversities available in the space, time, and frequency domains. In the case of narrow-band wireless communications, the channel fading is frequency non-selective (flat) and diversities are available only in the space and time domains. The modulation and coding approach that is developed for this scenario is termed space–time (ST) coding, exploiting available spatial and temporal diversity.

In this chapter, we first describe the MIMO communication system architecture with frequency-non-selective fading channels, which are often termed as narrow-band wireless channels, and discuss design criteria in achieving the full space–time diversity. Then, we introduce several well-known ST coding techniques that can be guaranteed to achieve full space–time diversity.

## 2.1 System model and performance criteria

Assume that the MIMO systems have  $M_t$  transmit and  $M_r$  receive antennas. Channel state information (CSI) is assumed to be known at the receiver, but not at the transmitter. In narrowband transmission scenario, the fading channel is frequency-non-selective or flat, and is assumed to be quasi-static, i.e., the channel stays constant during one codeword transmission and it may change independently from one codeword transmission to another. In this case, diversity is available only in the space and time domains.

An ST-coded MIMO system is shown in Figure 2.1. The ST encoder divides input data stream into  $b$  bit long blocks and, for each block, selects one ST codeword from the codeword set of size  $L = 2^b$ . The selected codeword is then transmitted through the channel over the  $M_t$  transmit antennas and  $T$  time slots. Each codeword can be represented as a  $T \times M_t$  matrix

$$C = \begin{bmatrix} c_1^1 & c_1^2 & \cdots & c_1^{M_t} \\ c_2^1 & c_2^2 & \cdots & c_2^{M_t} \\ \vdots & \vdots & \ddots & \vdots \\ c_T^1 & c_T^2 & \cdots & c_T^{M_t} \end{bmatrix} \triangleq \{c_t^i : i = 1, 2, \dots, M_t\}, \quad (2.1)$$

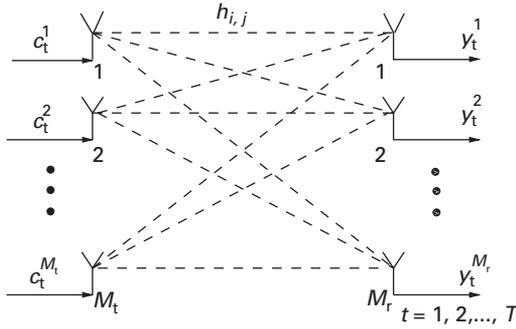


Fig. 2.1 MIMO communication system with  $M_t$  transmit and  $M_r$  receive antennas

where  $c_t^i$  denotes the channel symbol transmitted by transmit antenna  $i$ ,  $i = 1, 2, \dots, M_t$ , at discrete time  $t$ ,  $t = 1, 2, \dots, T$ . The codewords are assumed to satisfy the energy constraint  $E\|C\|_F^2 = M_t T$ , where  $E$  stands for the expectation and  $\|C\|_F$  is the Frobenius norm of  $C$ , defined as

$$\|C\|_F^2 = \text{tr}(C^H C) = \text{tr}(C C^H) = \sum_{t=1}^T \sum_{i=1}^{M_t} |c_t^i|^2.$$

The channel coefficient between transmit antenna  $i$  and receive antenna  $j$  is denoted by  $h_{i,j}$ . These coefficients are modeled as zero-mean, complex Gaussian random variables with unit variance. The received signal  $y_t^j$  at receive antenna  $j$  at time  $t$  can be expressed as

$$y_t^j = \sqrt{\frac{\rho}{M_t}} \sum_{i=1}^{M_t} c_t^i h_{i,j} + z_t^j, \quad t = 1, 2, \dots, T, \quad (2.2)$$

where  $z_t^j$  is the complex Gaussian noise component at receive antenna  $j$  at time  $t$  with zero mean and unit variance. The factor  $\sqrt{\rho/M_t}$  in (2.2) ensures that  $\rho$  is the average signal-to-noise ratio (SNR) at each receive antenna, and it is independent of the number of transmit antennas. The received signal (2.2) can be rewritten in a more compact form as

$$Y = \sqrt{\frac{\rho}{M_t}} C H + Z, \quad (2.3)$$

where  $Y = \{y_t^j : 1 \leq t \leq T, 1 \leq j \leq M_r\}$  is the received signal matrix of size  $T \times M_r$ ,  $H = \{h_{i,j} : 1 \leq i \leq M_t, 1 \leq j \leq M_r\}$  is the channel coefficient matrix of size  $M_t \times M_r$ ,  $Z = \{z_t^j : 1 \leq t \leq T, 1 \leq j \leq M_r\}$  is the noise matrix of size  $T \times M_r$ , and  $C$  is the space–time codeword, as defined in (2.1).

Assume that the perfect channel information is available at the receiver, then the maximum-likelihood (ML) decoding of the transmitted matrix is

$$\hat{C} = \arg \min_C \|Y - \sqrt{\frac{\rho}{M_t}} C H\|_F^2.$$

Suppose that codeword  $C$  is transmitted and the receiver erroneously in favor of codeword  $\tilde{C}$ . Since the noise term is a zero-mean Gaussian random variable, the pairwise error probability (PEP) for a fixed channel realization can be determined as (leave the proof as an exercise)

$$\begin{aligned} P(C \rightarrow \tilde{C}|H) &= Q\left(\sqrt{\frac{\rho}{2M_t}} \|(C - \tilde{C})H\|_F\right) \\ &= \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\rho}{4M_t \sin^2 \theta} \|(C - \tilde{C})H\|_F^2\right) d\theta, \end{aligned} \quad (2.4)$$

where  $Q(x) = 1/\sqrt{2\pi} \int_x^\infty \exp(-t^2/2) dt$  is the Gaussian error function, and the second equality comes from the Craig expression

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2 \sin^2 \theta}\right) d\theta.$$

Averaging over the Rayleigh fading channel  $H$ , the PEP can be determined as follows (leave the proof as an exercise):

$$P(C \rightarrow \tilde{C}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{\gamma} \left(1 + \frac{\rho \lambda_i}{4M_t \sin^2 \theta}\right)^{-M_r} d\theta, \quad (2.5)$$

where  $\gamma = \text{rank}(C - \tilde{C})$ , and  $\lambda_1, \lambda_2, \dots, \lambda_\gamma$  are the non-zero eigenvalues of  $(C - \tilde{C})(C - \tilde{C})^H$ . The superscript  $H$  stands for the complex conjugate and transpose of a matrix. By taking  $\theta = \pi/2$  in (2.5), we have the well known upper bound

$$P(C \rightarrow \tilde{C}) \leq \frac{1}{2} \prod_{i=1}^{\gamma} \left(1 + \frac{\rho \lambda_i}{4M_t}\right)^{-M_r} \quad (2.6)$$

$$\leq \frac{1}{2} \left(\frac{\rho}{4M_t}\right)^{-\gamma M_r} \left(\prod_{i=1}^{\gamma} \lambda_i\right)^{-M_r}. \quad (2.7)$$

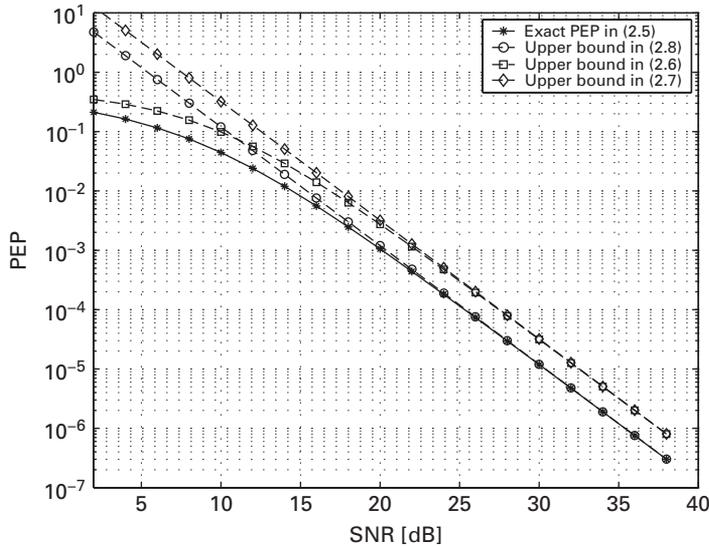
On the other hand, for high enough SNR, it is easy to see that the exact PEP in (2.5) can be upper bounded as

$$\begin{aligned} P(C \rightarrow \tilde{C}) &\leq \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{\gamma} \left(\frac{\rho \lambda_i}{4M_t \sin^2 \theta}\right)^{-M_r} d\theta \\ &= \left(\frac{2\gamma M_r - 1}{\gamma M_r - 1}\right) \left(\frac{\rho}{M_t}\right)^{-\gamma M_r} \left(\prod_{i=1}^{\gamma} \lambda_i\right)^{-M_r}, \end{aligned} \quad (2.8)$$

where the equality comes from

$$\frac{1}{\pi} \int_0^{\pi/2} (\sin \theta)^{2\gamma M_r} d\theta = 2^{2\gamma M_r} \left(\frac{2\gamma M_r - 1}{\gamma M_r - 1}\right).$$

Let us compare the three upper bounds in (2.6–2.8) with the exact PEP (2.5) in Figure 2.2. We can see that, at high SNR, the upper bound (2.8) is much tighter than



**Fig. 2.2** Comparison between the exact PEP and the three upper bounds. Assume that there are  $M_t = 2$  transmit and  $M_r = 1$  receive antennas,  $\gamma = 2$  and  $\lambda_1 = \lambda_2 = 1$ .

that in (2.7). Note that they share the same term on the product of the nonzero eigenvalues and the order of SNR, and the difference between them is a constant. We observe that the term  $(\rho/M_t)^{-\gamma M_r}$  in the upper bounds is a dominant term when the SNR  $\rho$  is high, thus for a given SNR, the rank  $\gamma$  should be maximized in order to minimize the PEP error rate. Two ST code design criteria can be developed based on the upper bound (2.7):

- *Rank criterion or diversity criterion:* The minimum rank of the code difference matrix  $C - \tilde{C}$  overall distinct codewords  $C$  and  $\tilde{C}$  should be as large as possible. If the matrix  $C - \tilde{C}$  is always of full rank for a specific ST code, we say that this ST code achieves full diversity.
- *Product criterion:* The minimum value of the product  $\prod_{i=1}^{\gamma} \lambda_i$  over all distinct codewords  $C$  and  $\tilde{C}$ , which is often termed as coding gain, should be as large as possible. This quantity is referred to as the coding advantage achieved by the ST code.

The diversity criterion is the more important of the two since it determines the slope of the performance curve. In order to achieve the maximum diversity, the difference matrix  $C - \tilde{C}$  has to be full rank for any pair of distinct codewords  $C$  and  $\tilde{C}$ . The product criterion is of secondary importance and should be optimized if the full diversity is achieved. If  $(C - \tilde{C})(C - \tilde{C})^H$  is of full rank, then the product  $\lambda_1 \lambda_2 \dots \lambda_n$  is equal to the determinant of  $(C - \tilde{C})(C - \tilde{C})^H$ . In this case, which implies  $M_t \geq T$ , a helpful quantity  $\zeta$  called *diversity product* is given by

$$\zeta = \frac{1}{2\sqrt{M_t}} \min_{C \neq \tilde{C}} \left| \det \left[ (C - \tilde{C})(C - \tilde{C})^H \right] \right|^{1/(2T)}, \quad (2.9)$$

which is a normalized coding gain. The factor  $1/(2\sqrt{M_t})$  guarantees that  $0 \leq \zeta \leq 1$ . When all codewords are square matrices, i.e.,  $T = M_t$ , the diversity product can be simplified as

$$\zeta = \frac{1}{2\sqrt{M_t}} \min_{C \neq \tilde{C}} \left| \det(C - \tilde{C}) \right|^{1/M_t}, \quad (2.10)$$

which is often used as a benchmark in designing ST codes.

## 2.2 Space–time coding

In the previous section, we analyzed the performance of the ST-coded MIMO systems and obtained ST code design criteria in achieving the spatial and temporal diversity. In this section, we introduce several well-known ST coding techniques that can guarantee to achieve the full space–time diversity.

### 2.2.1 Cyclic and unitary ST codes

A simple and effective coding scheme achieving full space–time diversity is the cyclic ST coding approach which follows the following code structure:

$$C_l = \sqrt{M_t} \text{diag}\{e^{ju_1\theta_l}, e^{ju_2\theta_l}, \dots, e^{ju_{M_t}\theta_l}\}, \quad l = 0, 1, \dots, L-1, \quad (2.11)$$

where  $\text{diag}\{e^{ju_1\theta_l}, e^{ju_2\theta_l}, \dots, e^{ju_{M_t}\theta_l}\}$  is a diagonal matrix with diagonal entries  $e^{ju_1\theta_l}, e^{ju_2\theta_l}, \dots$ , and  $e^{ju_{M_t}\theta_l}$ , in which  $\theta_l = (l/L)2\pi$  and  $u_1, u_2, \dots, u_{M_t} \in \{0, 1, \dots, L-1\}$  are some integers to be optimized. We observe that if we denote

$$V_l = \text{diag}\{e^{ju_1 \frac{2\pi}{L}}, e^{ju_2 \frac{2\pi}{L}}, \dots, e^{ju_{M_t} \frac{2\pi}{L}}\}, \quad (2.12)$$

then  $C_l = \sqrt{M_t} V_l^l$  for  $l = 0, 1, \dots, L-1$ , and  $V_l^L = V_l^0$ , which has a cyclic structure, hence the term *cyclic ST codes*.

In the following, we discuss how the cyclic structure can guarantee the full diversity and we will also optimize the parameters  $u_1, u_2, \dots, u_{M_t}$ . For any two distinct codewords  $C_l$  and  $C_{l'}$ ,  $l \neq l'$ , we have

$$\begin{aligned} C_l - C_{l'} &= \sqrt{M_t} \text{diag}\{e^{ju_1\theta_l} - e^{ju_1\theta_{l'}}, e^{ju_2\theta_l} - e^{ju_2\theta_{l'}}, \dots, e^{ju_{M_t}\theta_l} - e^{ju_{M_t}\theta_{l'}}\} \\ &= \sqrt{M_t} \text{diag}\{e^{ju_1\theta_{l'}}, e^{ju_2\theta_{l'}}, \dots, e^{ju_{M_t}\theta_{l'}}\} \\ &\quad \times \text{diag}\{e^{ju_1\theta_{l-l'}} - 1, e^{ju_2\theta_{l-l'}} - 1, \dots, e^{ju_{M_t}\theta_{l-l'}} - 1\} \\ &= C_{l'} \text{diag}\{e^{ju_1\theta_{l-l'}} - 1, e^{ju_2\theta_{l-l'}} - 1, \dots, e^{ju_{M_t}\theta_{l-l'}} - 1\}, \end{aligned}$$

in which  $\theta_{l-l'} = \frac{l-l'}{L}2\pi$ . So the determinant of the difference matrix is

$$\begin{aligned} |\det(C_l - C_{l'})| &= |\det(C_{l'})| \cdot \left| \det(\text{diag}\{e^{ju_1\theta_{l-l'}} - 1, \right. \\ &\quad \left. e^{ju_2\theta_{l-l'}} - 1, \dots, e^{ju_{M_t}\theta_{l-l'}} - 1\}) \right| \\ &= M_t^{\frac{M_t}{2}} \prod_{i=1}^{M_t} \left| e^{ju_i\theta_{l-l'}} - 1 \right| \\ &= M_t^{\frac{M_t}{2}} \prod_{i=1}^{M_t} \left| 2 \sin \frac{u_i(l-l')\pi}{L} \right|. \end{aligned} \quad (2.13)$$

From (2.13), we can see that if  $\sin \frac{u_i(l-l')}{L} \neq 0$  for  $u_1, u_2, \dots, u_{M_t}$  and for any  $l \neq l'$ , then the cyclic code achieves the full diversity according to the rank criterion discussed in the previous section.

From (2.13), we can calculate the diversity product of the cyclic code as follows:

$$\begin{aligned} \zeta &= \frac{1}{2\sqrt{M_t}} \min_{l \neq l'} |\det(C_l - C_{l'})|^{1/M_t} \\ &= \min_{l \neq l'} \left| \prod_{i=1}^{M_t} \sin \frac{u_i(l-l')\pi}{L} \right|^{1/M_t} \\ &= \min_{1 \leq l \leq L-1} \left| \prod_{i=1}^{M_t} \sin \frac{u_i l \pi}{L} \right|^{1/M_t}. \end{aligned} \quad (2.14)$$

Thus, the parameters  $u_i \in \{0, 1, \dots, L-1\}$  should be chosen such that the diversity product  $\zeta$  is maximized. For small  $L$  and  $M_t$ , exhaustive computer search can be performed to find the optimum parameters  $u_1, u_2, \dots, u_{M_t} \in \{0, 1, \dots, L-1\}$ .

---

**Example 2.1** For  $M_t = 2$  and  $L = 4$  (i.e.,  $R = 1$  bit/s/Hz), by exhaustive computer search, the optimum parameters in this case are  $[u_1 \ u_2] = [1 \ 1]$ . The corresponding cyclic ST codes are given by:

$$\begin{aligned} C_0 &= \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & C_1 &= \sqrt{2} \begin{bmatrix} j & 0 \\ 0 & j \end{bmatrix}, \\ C_2 &= \sqrt{2} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, & C_3 &= \sqrt{2} \begin{bmatrix} -j & 0 \\ 0 & -j \end{bmatrix}. \end{aligned}$$

The diversity product  $\zeta = \min_{1 \leq l \leq 3} \left| \prod_{i=1}^2 \sin \frac{u_i l \pi}{4} \right|^{1/2} = \frac{\sqrt{2}}{2}$ . ▲

---

**Example 2.2** For  $M_t = 2$  and  $L = 16$  (i.e.,  $R = 2$  bits/s/Hz), by exhaustive computer search, the optimum parameters in this case are  $[u_1 \ u_2] = [1 \ 7]$ . The corresponding cyclic ST codes are given by:

$$C_l = \sqrt{2} \begin{bmatrix} e^{j\theta_l} & 0 \\ 0 & e^{j\theta_l} \end{bmatrix}, \quad \theta_l = \frac{l\pi}{8}, \quad l = 0, 1, \dots, 15.$$

▲

Cyclic ST codes are special unitary codes in which all off-diagonal entries are zero. In general, unitary matrices with nonzero diagonal entries can also be used to achieve the full space–time diversity and provide a larger diversity product. Unitary ST codes with codewords  $C_0, C_1, \dots, C_{L-1}$  satisfy

$$\left( \frac{1}{\sqrt{M_t}} C_l \right)^H \left( \frac{1}{\sqrt{M_t}} C_l \right) = I_{M_t \times M_t}, \quad \text{or} \quad C_l^H C_l = M_t I_{M_t \times M_t}, \quad (2.15)$$

for  $l = 0, 1, \dots, L - 1$ . Unitary ST codes can be designed through Fourier transforms or by using unitary matrices with some special structures. For example, the following unitary ST code has a larger diversity product than the cyclic code shown in Example 2.1.

**Example 2.3** For  $M_t = 2$  and  $L = 4$  (i.e.,  $R = 1$  bit/s/Hz), there are four unitary matrices with size  $2 \times 2$  that can guarantee the full space–time diversity:

$$C_0 = \sqrt{\frac{2}{3}} \begin{bmatrix} j & 1-j \\ -1-j & -j \end{bmatrix}, \quad C_1 = \sqrt{\frac{2}{3}} \begin{bmatrix} -j & -1-j \\ 1-j & j \end{bmatrix},$$

$$C_2 = \sqrt{\frac{2}{3}} \begin{bmatrix} -j & 1+j \\ -1+j & j \end{bmatrix}, \quad C_3 = \sqrt{\frac{2}{3}} \begin{bmatrix} j & -1+j \\ 1+j & -j \end{bmatrix}.$$

The diversity product of this unitary ST code is  $\zeta = \sqrt{\frac{2}{3}} \approx 0.8165$ .

▲

## 2.2.2 ST codes from orthogonal designs

ST codes from orthogonal designs have received considerable attention in MIMO wireless communications. Such codes can guarantee to achieve the full space–time diversity and also provide simple fast ML decoding algorithms. Some codes have been adopted in the WLAN standard IEEE 802.11n.

### *A motivated example: Alamouti scheme*

Alamouti proposed in 1998 [5] a simple scheme for MIMO systems with two transmit antennas as follows:

$$G_2(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}, \quad (2.16)$$

in which  $x_1$  and  $x_2$  are arbitrary complex symbols and an energy constraint is  $E\|G_2\|_F^2 = 4$ . The proposed scheme in (2.16) has an interesting property that for arbitrary complex  $x_1$  and  $x_2$ , the columns of  $G_2$  are orthogonal to each other, i.e.,