EXPLOITING DIVERSITY IN WIRELESS CHANNELS WITH BIT-INTERLEAVED CODED MODULATION AND ITERATIVE DECODING (BICM-ID)

A Dissertation Submitted
to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in the Department of Electrical & Computer Engineering
University of Saskatchewan

by

Nghi H. Tran

Saskatoon, Saskatchewan, Canada

PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Post-graduate degree from the University of Saskatchewan, it is agreed that the Libraries of this University may make it freely available for inspection. Permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised this dissertation work or, in their absence, by the Head of the Department of Electrical & Computer Engineering or the Dean of the College of Graduate Studies and Research at the University of Saskatchewan. Any copying, publication, or use of this dissertation, or parts thereof, for financial gain without the written permission of the author is strictly prohibited. Proper recognition shall be given to the author and to the University of Saskatchewan in any scholarly use which may be made of any material in this dissertation.

Request for permission to copy or to make any other use of material in this dissertation in whole or in part should be addressed to:

Head of the Department of Electrical & Computer Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A9

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Ha H. Nguyen, who guided me to this exciting area of wireless communications. I can never forget the first wonderful lectures in digital communications and error control coding I took from him when I started my M.Sc. studies five years ago. Working closely with him thereafter, I have benefited tremendously from his great source of knowledge and his enthusiasm of doing research. Without his valuable advice, encouragement and unconditional support, this dissertation could not have been finished.

My deepest gratitude and appreciation also go to my co-advisor, Prof. Tho Le-Ngoc from McGill University, for his spiritual and financial support, as well as many helpful and useful research discussions. I must say a ton of thanks to him for his invaluable advice and suggestions on my Ph.D. studies and my future career. I admire his knowledge of science and life, both in depth and broadness.

I want to thank the committee members, Prof. Brian Daku, Prof. Rajesh Karki, and Prof. Richard Burton, for their constructive comments and suggestions on my research.

The financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Discovery and Strategic Grants, the University of Saskatchewan under the Graduate Scholarship, and Telecommunication Research Laboratories (TRLabs) Saskatoon under a Fellowship is gratefully acknowledged. These financial sources are crucial in order to help me keep working on my research and finish this dissertation.

My heartfelt thanks go to my wife, Minh Ha, for her love, encouragement, and support during this long-term of study. She has made my life more fun, fulfilled, and enjoyable. To her, I dedicate this dissertation.

Finally, my love and gratitude is devoted to my parents, my grandmother, and my sister. They are always a source of love, supporting and encouragement. Their importance to me is beyond any words can express.

UNIVERSITY OF SASKATCHEWAN

Electrical & Computer Engineering Abstract

EXPLOITING DIVERSITY IN WIRELESS CHANNELS WITH BIT-INTERLEAVED CODED MODULATION AND ITERATIVE DECODING (BICM-ID)

Student: Nghi H. Tran

Advisors: Prof. Ha H. Nguyen and Prof. Tho Le-Ngoc (McGill University)

Ph.D. Dissertation Submitted to the

College of Graduate Studies and Research

April, 2008

ABSTRACT

This dissertation studies a state-of-the-art bandwidth-efficient coded modulation technique, known as bit interleaved coded modulation with iterative decoding (BICM-ID), together with various diversity techniques to dramatically improve the performance of digital communication systems over wireless channels.

For BICM-ID over a single-antenna *frequency non-selective* fading channel, the problem of mapping over multiple symbols, i.e., multi-dimensional (multi-D) mapping, with 8-PSK constellation is investigated. An explicit algorithm to construct a good multi-D mapping of 8-PSK to improve the asymptotic performance of BICM-ID systems is introduced. By comparing the performance of the proposed mapping with an unachievable lower bound, it is conjectured that the proposed mapping is the global optimal mapping. The superiority of the proposed mapping over the best conventional (1-dimensional complex) mapping and the multi-D mapping found previously by computer search is thoroughly demonstrated.

In addition to the mapping issue in single-antenna BICM-ID systems, the use of signal space diversity (SSD), also known as linear constellation precoding (LCP), is considered in BICM-ID over *frequency non-selective* fading channels. The performance analysis of BICM-ID and complex N-dimensional signal space diversity is carried out to study its performance limitation, the choice of the rotation matrix and the design of a low-complexity receiver. Based on the design criterion obtained from a tight error bound, the optimality of the rotation matrix is established. It is shown that using the class of optimal rotation matrices, the performance of BICM-ID systems over a *frequency non-selective* Rayleigh fading channel approaches that of the BICM-ID systems over an additive white Gaussian noise (AWGN) channel when the dimension of the signal constellation increases. Furthermore, by exploiting the sigma mapping for any M-ary quadrature amplitude modulation (QAM) constellation, a very simple sub-optimal, yet effective iterative receiver structure suitable for signal constellations with large dimensions is proposed. Simulation results in various cases and conditions indicate that the proposed receiver can achieve the analytical performance bounds with low complexity.

The application of BICM-ID with SSD is then extended to the case of cascaded Rayleigh fading, which is more suitable to model mobile-to-mobile communication channels. By deriving the error bound on the asymptotic performance, it is first illustrated that for a small modulation constellation, a cascaded Rayleigh fading causes a much more severe performance degradation than a conventional Rayleigh fading. However, BICM-ID employing SSD with a sufficiently large constellation can close the performance gap between the Rayleigh and cascaded Rayleigh fading channels, and their performance can closely approach that over an AWGN channel.

In the next step, the use of SSD in BICM-ID over *frequency selective* Rayleigh fading channels employing a multi-carrier modulation technique known as orthogonal frequency division multiplexing (OFDM) is studied. Under the assumption of correlated fading over subcarriers, a tight bound on the asymptotic error performance for the general case of applying SSD over all N subcarriers is derived and used to

establish the best achievable asymptotic performance by SSD. It is then shown that precoding over subgroups of at least L subcarriers per group, where L is the number of channel taps, is sufficient to obtain this best asymptotic error performance, while significantly reducing the receiver complexity. The optimal joint subcarrier grouping and rotation matrix design is subsequently determined by solving the Vandermonde linear system. Illustrative examples show a good agreement between various analytical and simulation results.

Further, by combining the ideas of multi-D mapping and subcarrier grouping, a novel power and bandwidth-efficient bit-interleaved coded modulation with OFDM and iterative decoding (BI-COFDM-ID) in which multi-D mapping is performed over a group of subcarriers for broadband transmission in a *frequency selective* fading environment is proposed. A tight bound on the asymptotic error performance is developed, which shows that subcarrier mapping and grouping have independent impacts on the overall error performance, and hence they can be independently optimized. Specifically, it is demonstrated that the optimal subcarrier mapping is similar to the optimal multi-D mapping for BICM-ID in *frequency non-selective* Rayleigh fading environment, whereas the optimal subcarrier grouping is the same with that of OFDM with SSD. Furthermore, analytical and simulation results show that the proposed system with the combined optimal subcarrier mapping and grouping can achieve the full channel diversity without using SSD and provide significant coding gains as compared to the previously studied BI-COFDM-ID with the same power, bandwidth and receiver complexity.

Finally, the investigation is extended to the application of BICM-ID over a multiple-input multiple-output (MIMO) system equipped with multiple antennas at both the transmitter and the receiver to exploit both time and spatial diversities, where neither the transmitter nor the receiver knows the channel fading coefficients. The concentration is on the class of unitary constellation, due to its advantages in terms of both information-theoretic capacity and error probability. The tight error bound with respect to the asymptotic performance is also derived for any given unitary constellation

and mapping rule. Design criteria regarding the choice of unitary constellation and mapping are then established. Furthermore, by using the unitary constellation obtained from orthogonal design with quadrature phase-shift keying (QPSK or 4-PSK) and 8-PSK, two different mapping rules are proposed. The first mapping rule gives the most suitable mapping for systems that do not implement iterative processing, which is similar to a Gray mapping in coherent channels. The second mapping rule yields the best mapping for systems with iterative decoding. Analytical and simulation results show that with the proposed mappings of the unitary constellations obtained from orthogonal designs, the asymptotic error performance of the iterative systems can closely approach a lower bound which is applicable to any unitary constellation and mapping.

Table of Contents

List of Tables

List of Figures

ABBREVIATIONS

AF	Amplify-and-Forward
ASK	Amplitude Shift Keying
AUB	Asymptotic Union Bound
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BEP	Bit Error Probability
BICM	Bit-Interleaved Coded Modulation
BICM-EX	Bit-Interleaved Coded Modulation Expurgated
BICM-ID	Bit-Interleaved Coded Modulation with Iterative Decoding
BI-COFDM-ID	Bit-Interleaved Coded Modulation with OFDM and Iterative Decoding
BI-STCM-ID	Bit-Interleaved Space Time Coded-Modulation with Iterative Decoding
BIOS	Binary-Input Output-Symmetric
BPSK	Binary Phase Shift Keying
BSA	Binary Switching Algorithm
CC	Convolutional Code
$\mathcal{CN}(\cdot)$	Circularly Symmetric Complex Gaussian Random Variable
CDMA	Code-Division Multiple-Access
CSI	Channel State Information
dB	Decibel
DFT	Discrete Fourier Transform
DSL	Digital Subscriber Lines
EF	Error-free Feedback
EXIT	Extrinsic Information Transfer
FED	Free Euclidean Distance

FFT	Fast Fourier Transform
FLOP	Floating Point Operation
GA	Gaussian Approximation
ICI	Inter-Carrier Interference
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
i.i.d	Independent and Identically Distributed
ISI	Inter-Symbol Interference
ITU	International Telecommunication Union
LCP	Linear Constellation Precoding
LDPC	Low-Density Parity-Check
MAP	Maximum A Posteriori Probability
MIMO	Multi-Input Multi-Output
ML	Maximum Likelihood
MLC	Multilevel Coding
MMSE	Minimum Mean Square Error
N-D	N-dimensional
OFDM	Orthogonal Frequency Division Multiplexing
PAM	Pulse Amplitude Modulation
pdf	Probability Density Function
PEP	Pairwise Error Probability
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QAP	Quadratic Assignment Problem
QPSK	Quadrature Phase Shift Keying
SISO	Soft-Input Soft-Output
SNR	Signal-to-Noise Ratio
SSD	Signal Space Diversity
SSP	Semi Set Partitioning
TCM	Trellis-Coded Modulation

TTCM	Turbo Trellis-Coded Modulation
WiMAX	Worldwide Interoperability for Microwave Access

1. Introduction and Organization of The Dissertation

1.1 Introduction

Over the last decades, wireless communications has emerged and played an important role in the daily activities of human beings. This is due to the fact that wireless connectivity is making exciting inroad to everywhere, from home to office as well as from urban to rural areas. As far as the physical layer is concerned, the primary challenge in the design of wireless communication systems is to increase the data transmission rate over the bandwidth-limited wireless radio channel. Moreover, such a high data rate transmission has to be conducted with a high reliability (i.e., good quality) and, at the same time, at as low power consumption as possible.

A general digital communication system over a wireless channel includes three basic elements: a transmitter, a wireless radio channel and a receiver. The transmitter translates the information bits to signals that can be effectively transmitted over the channel. The transmitter includes a channel encoder and a modulator. The purpose of the channel encoder is to introduce, in a controlled manner, some redundancy in the binary information sequence. This redundant information can be used at the receiver to overcome the adverse effects of noise and interference encountered during the transmission. The modulator performs a process called modulation, in which a digital bit stream is converted into a waveform that passes through the channel. The wireless channel is the physical medium used to send the signal from the transmitter to the receiver. The receiver consists of a demodulator and a decoder, where the opposite operations are performed accordingly. In general, the receiver tries to recover

the transmitted information as correctly as possible.

Due to the presence of various sources of impairments such as multi-path propagation, and the relative mobility between the transmitter and the receiver (which constitute the notorious multi-path fading effect), digital communication systems operating over wireless channels often experience a poor error performance and low data rate. Such disadvantages make the design of a digital communication system over a wireless channel more challenging and interesting compared to the design for a traditional additive white Gaussian noise (AWGN) channel.

An effective way to mitigate the adverse effects of fading is to apply diversity techniques. The basic idea of all diversity techniques is to send separate copies of the same information over multiple independently faded paths in order to increase reliability and the probability of successful transmission. Common diversity techniques that have been studied intensively in the literature include:

- Time diversity (e.g., channel coding),
- Signal space diversity,
- Spatial diversity (e.g., multiple-input multiple-output (MIMO) systems),
- Frequency diversity (e.g., orthogonal frequency division multiplexing (OFDM)).

Typically, several types of diversity can be combined and incorporated in wireless systems to further improve the overall performance of the systems. This is also the main objective of this research.

One of the state-of-the-art techniques, known as bit-interleaved coded modulation (BICM) [1,2], was recently introduced and demonstrated to partially address the main challenges in wireless communications. The basic idea of a bandwidth efficient BICM system is to extract time diversity in a fading channel by combining a channel encoder, such as a linear block code or a convolutional code, and a random bit-interleaver. By doing so, it was shown that the diversity gain of the system can be increased to

the *minimum Hamming distance* of the channel encoder [1, 2]. The other advantage of BICM is that it offers more flexibility in the system design, because the channel encoder and the modulator can be designed independently. Therefore, this technique has been widely accepted in the current wireless standards [3].

Since the invention of turbo codes [4], interleaving and iterative processing have also been applied to many digital communication systems. It was shown in [5–11] that BICM with iterative decoding (BICM-ID) can provide a significant performance improvement compared to the non-iterative decoding scheme over a fading channel with a reasonable increase in complexity.

Motivated by the above discussions, this dissertation concentrates on the analysis and design of BICM-ID systems combining and incorporating various diversity techniques in order to combat fading effectively, and as a result, to dramatically improve the performance of digital communication systems over wireless channels. The motivations and contributions of the research will be given in detail in each individual chapter.

1.2 Organization of the Dissertation

The remainder of the dissertation is organized as follows.

Chapter 2 provides some relevant background and knowledge that are useful for the development of various techniques in the subsequent chapters.

Chapter 3 studies a problem of mapping from binary information bits to signals transmitted over channels. In particular, a BICM-ID system with multidimensional constellation and mapping using 8-PSK constellation is first described. By evaluating the error floor of the system, distance criteria that characterize the influence of the signal constellation and mapping to the error performance are provided for different channel models. Based on the distance profile of the 8-PSK constellation, unachievable lower bounds for these distance criteria are establihsed. An explicit algorithm to construct a good multi-D mapping of 8-PSK to improve the asymptotic perfor-

mance of BICM-ID systems is then introduced. By comparing the performance of the proposed mapping with the unachievable lower bound, it is conjectured that the proposed mapping is the global optimal mapping. Analytical and simulation results are provided to demonstrate the advantages of the proposed systems and the tightness of the error bound.

Chapter 4 investigates the application of BICM-ID with signal space diversity over a fully interleaved *frequency non-selective* Rayleigh fading channel. Analytical derivation of the system's bit error probability (BEP) is first presented. Then a distance criterion for given signal constellation, signal mapping and rotation matrix $\boldsymbol{G}$ that affects the error performance is obtained. The design problem of the rotation matrix $\boldsymbol{G}$ for given constellation and mapping is then discussed. It is demonstrated that using the class of optimal rotation matrices, the performance of BICM-ID systems over a Rayleigh fading channel approaches that of the BICM-ID systems over an AWGN channel when the dimension of the signal constellation increases. By exploiting the sigma mapping for any M-ary QAM constellation, a very simple sub-optimal, yet effective iterative receiver structure suitable for signal constellations with large dimensions is proposed. Simulation results in various cases and conditions indicate that the proposed receiver can achieve the analytical performance bounds with low complexity.

Chapter 5 extends the use of SSD in BICM-ID over a cascaded Rayleigh fading channel, which is more suitable to model the mobile-to-mobile communications channels. A tight bound on the asymptotic error performance is first derived. It is then illustrated that for small modulation constellations, a *cascaded* Rayleigh fading causes a much more severe performance degradation than a conventional Rayleigh fading. However, BICM-ID employing SSD with a sufficiently large constellation can close the performance gap between the Rayleigh and cascaded Rayleigh fading channels, and their performance can closely approach that over an AWGN channel.

Chapter 6 is concerned with the application of BICM-ID with signal space diversity over *frequency selective* fading channel using a multi-carrier modulation technique,

known as orthogonal frequency division multiplexing (OFDM). The general structure of OFDM systems using BICM-ID with SSD and their corresponding system model is first described. Based on this system model, the performance analysis for the general case when SSD is employed over all N subcarriers is presented. Distance criterion for given signal constellation, signal mapping and rotation matrix is subsequently established. Based on the design criterion, the best achievable asymptotic performance by using SSD is provided. It is then shown that precoding over subgroups of at least L subcarriers per group, where L is the number of channel taps, is sufficient to obtain this best asymptotic error performance, while significantly reducing the receiver complexity. Then the optimum joint subcarrier grouping and rotation matrix design is solved. Illustrative analytical and simulation results are finally presented and discussed.

Chapter 7 proposes a novel power and bandwidth-efficient bit-interleaved coded modulation with OFDM and iterative decoding (BI-COFDM-ID) using combined multi-dimensional (multi-D) mapping and subcarrier grouping. The structure of OFDM systems using BICM-ID, with or without SSD, and the implementation of multi-dimensional mapping over subcarriers are first described. Then the performance analysis for the system under consideration is presented. The design criterion for a given subcarrier mapping is established to address optimal choices of signal mapping and the subcarrier grouping. It is demonstrated through both analytical and simulation results that the technique of multi-D mapping over groups of subcarriers can successfully exploit the multi-path diversity and, at the same time, provide remarkable coding gains without the need of implementing SSD.

Chapter 8 analytically evaluates the asymptotic bit error probability (BEP) for unitary bit-interleaved space-time coded modulation with iterative decoding over non-coherent multiple-input multiple-output (MIMO) channels. The evaluation is based on the assumption of ideal feedback from the channel decoder to the demodulator and can be applied for any given unitary constellation. Simulation results of the bit-error-rate performance confirms that the derived asymptotic BEP is very tight at high

signal to noise ratio (SNR). Based on the asymptotic performance, design criteria for the choice of unitary constellation and mapping are also established. Using the class of unitary constellations obtained from orthogonal designs with 4-PSK and 8-PSK signal sets, the optimal mappings with respect to the asymptotic performances of systems with and without iterative processing are then introduced. It is shown that the use of the proposed mappings for iterative systems results in error performance that is very close to the lower bound of the asymptotic performance for any unitary constellation and mapping.

Finally, Chapter 9 draws the conclusions and gives suggestions for further studies.

2. Background

This chapter provides some background on wireless fading channels, diversity techniques over a wireless channels, and a general framework of BICM-ID and various tools to analyze the system.

2.1 Wireless Fading Channels

In wireless communications, the channel strength varies over both time and frequency. The variation is due to the following two reasons:

- Large-scale fading: The variation in the received signal power is caused by path loss of signal as a function of distance from the transmitter to the receiver and by shadowing due to obstacles such as buildings, trees, hills between the transmitter and receiver. The variation due to path loss and shadowing occurs over a large distance e.g., meters, and is typically frequency independent.

- Small-scale fading: The variation in the received signal power is due to the constructive and destructive addition of multipath signal components between the transmitter and receiver. Variation due to multipath occurs over very short distances, on the order of the signal wavelength, and is frequency dependent.

In this dissertation, it is of interest to emphasis only on the small-scale fading, since small-scale fading is one of the main challenges in the design of reliable and efficient wireless communications systems.

Because of the presence of multiple objects in the environment that produce reflected, diffracted, or scattered copies of the transmitted signal, there can be different versions of the transmitted signal at the receiver side. Mathematically, the continuous-time input-output model of a wireless channel with an arbitrary input $x(t)$ can be represented as [12]:

$$r(t) = \sum_i a_i(t)x(t - \tau_i(t)) + w(t), \tag{2.1}$$

where $a_i(t)$ and $\tau_i(t)$ are the overall attenuation and propagation delay at time t from the transmitter to the receiver on path i. The noise $w(t)$ is zero-mean additive white Gaussian noise with power spectral density $N_0/2$. In practice, the attenuations and the propagation delays might vary slowly over frequency. However, since the bandwidth used to transmit the signal is narrow relative to the carrier frequency, it is reasonable to assume that $a_i(t)$ and $\tau_i(t)$ do not depend on the carrier frequency f. The linear model in (2.1) therefore can be described by the impulse response $h(t, \tau)$ at time t to an impulse transmitted at time $t - \tau$ as follows:

$$h(t, \tau) = \sum_i a_i(t)\delta(\tau - \tau_i(t)), \tag{2.2}$$

where $\delta(\cdot)$ is the Dirac delta function.

An useful discrete-time baseband model of (2.1) can also be obtained by applying the sampling theorem. Assume the transmitted signal $x(t)$ is bandlimited to B. Then the discrete input-output model of (2.1) is given as [12]:

$$r[m] = \sum_l h_l[m]x[m - l + 1] + w[m], \tag{2.3}$$

where $h_l[m]$ is the lth channel filter tap at time m. This channel tap depends on the gains $a_i(t)$ of the paths whose delays $\tau_i(t)$ are close to l/B. The discrete-time noise $w[m]$ is a circularly symmetric complex Gaussian random variable, $\mathcal{CN}(0, N_0)$, i.e., the real and imaginary parts of $w[m]$ are zero-mean independent and identically distributed (i.i.d.) Gaussian random variables with variances $N_0/2$.

An important characteristic of a multipath fading channel is the time delay spread $T_d = \max_{i,j} |\tau_i(t) - \tau_j(t)|$ that it causes to the received signal, which is equal to the

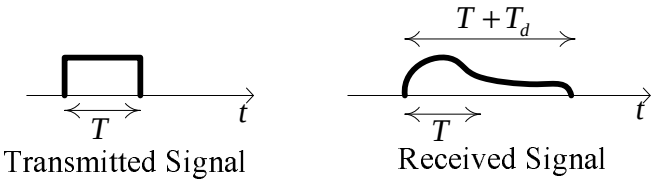


Figure 2.1 ISI phenomenon.

time delay between the arrival of the first received signal component and the last received signal component. Due to the time delay spread, the signal is dispersive in time, which results in the phenomenon of Inter-Symbol Interference (ISI) as shown in Fig. 2.1: distortion of a signal that causes the currently transmitted signal to have an effect on the next received signal. If the delay spread T_d is small compared to the signal duration T, or equivalently, inverse of the signal bandwidth B, the interference is negligible. In this case, most of the paths arrive during one symbol time and as a result, the channel model in (2.3) can be approximately represented by a single tap $r[m] = h[m]x[m] + w[m]$. This model is suitable for narrowband transmission and the wireless fading channel is called *frequency non-selective* fading.

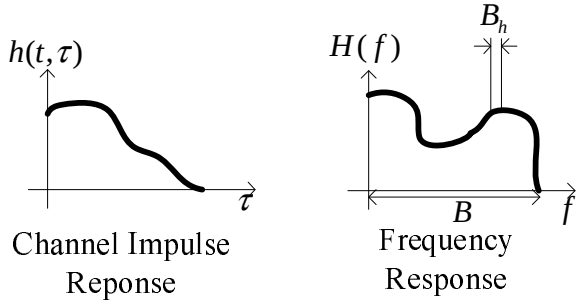


Figure 2.2 Channel impulse response and frequency response of a frequency selective fading channel.

In wideband high-data-rate systems, the symbol duration T is typically very small and $T_d \gg T$. Therefore, one needs to take into account the effect of ISI. Basically, dispersion in time domain means selectivity in frequency domain. Given a delay spread T_d, the frequency response does not change significantly in a band of width $B_h = 1/(2T_d)$, which is called the *coherence bandwidth* of the channel. Since $T_d \gg T$, one has $B \gg B_h$, and as a consequence, the frequency response varies over the entire

band B, as illustrated in Fig. 2.2. Equivalently, one can say the signal undergoes different attenuations over different frequency ranges. The fading channel is therefore called *frequency selective* fading and has to be represented by multiple taps as shown in (2.3). The number of channel taps L is basically the length of the ISI, which can be computed as $L = T_d/T$.

The other important property of fading in a wireless channel is whether it is fast or slow fading, which characterizes how rapidly the channel changes over time. This time variation arises because either the transmitter or the receiver is moving, and therefore the location of reflectors in the transmission path will change over time. Whether the channel can be categorized as fast or slow depends on both the channel and the application, which is discussed in details later for each specific system under consideration.

In this research, a simple but important fading model, called Rayleigh fading, shall be mainly adopted for consideration. Under Rayleigh fading, the channel gains $\{h_l[m]\}_{l=1}^{L}$ are modeled as:

$$h_l[m] \sim \mathcal{CN}(0, \delta_l^2). \tag{2.4}$$

Basically, this fading model implies that there is no line-of-sight path from the transmitter to the receiver. This model is a reasonable model when there are many objects in the environment that scatter the radio signal before it arrives at the receiver.

Also, throughout this research, the knowledge of the receiver about fading coefficients $\{h_l[m]\}$ is restricted only to the two extreme cases: perfect channel state information (CSI), where the receiver (not the transmitter) has perfect information of fading coefficients, and no CSI, where neither the transmitter nor the receiver knows the channel fading coefficients.

2.2 Diversity Techniques over a Wireless Channel

The wireless model in (2.3) might suffer a sudden decline in the receive power. This is due to the fact that there is a high probability that the channel gains $\{h_l[m]\}$

are dropped dramatically in magnitude. When it happens, the channel is said to be in a *deep fade*. An efficient way to combat fading, especially the deep fade, is to apply diversity techniques. The basic idea behind various diversity techniques (e.g., time, frequency, space) is to provide statistically independent copies of the same transmitted information at the receiver and appropriately process them to make the detection more reliable. In order to measure the effectiveness of each diversity technique, the relationship between the signal-to-noise ratio (SNR) and the overall error probability P_e is useful, which can be described by the *diversity gain* as follows:

$$G_d = -\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e}{\log \mathrm{SNR}}. \tag{2.5}$$

In essence, equation (2.5) implies that the error probability P_e decreases as the G_dth power of the SNR in a log-log scale.

2.2.1 Time Diversity

Time diversity can be achieved by dispersing the information signal over different time intervals in such a way that the signal experiences independent fades over time. Naturally, the simplest method is a *repetition code* in which the signal is repeated exactly the same over a number of time intervals. The use of this simple code often decreases the bandwidth efficiency drastically as one increases the number of times the signal is duplicated. In this work, a more sophisticated method, namely the error-control coding [13], shall be employed to extract the time diversity.

Basically, the purpose of error-control coding is to introduce, in a controlled manner, some redundancy in the binary information sequences. This redundancy can be used at the receiver to overcome the adverse effects of fading and noise encountered during the transmission of the signal over the channel. There are two types of classical error-control codes, namely algebraic block codes and convolutional codes [13]. More powerful and high-performance error correction codes can be constructed, at the price of increased decoding complexity. For example, parallel turbo codes [4] are obtained by concatenating several component convolutional codes. On the other hand, a low-density parity-check (LDPC) code [14,15] is a linear block code scattered

in a huge space but with the sparse parity check matrix. A major disadvantage of any error-control coding scheme is that it expands the transmission bandwidth to accommodate the redundancy in order to maintain the same transmission rate as with the uncoded systems. Such disadvantage can be partially overcome by combining error-control coding with high-order modulation schemes.

Concentrating on the framework of BICM-ID systems, this research only considers the class of standard convolutional codes with efficient "soft decoding" methods for these codes.

2.2.2 Signal Space Diversity

Signal space diversity is a means to exploit the time diversity without any power or bandwidth expansion. The general technique in this method is to apply a certain transformation to a conventional signal constellation to maximize the number of distinct components between two signal points [16,17]. More specifically, in signal space diversity, each group of N consecutive signal symbols is first mapped to an element of a complex N-dimensional (N-D) constellation, which is generally carved from a complex N-D lattice, and then a transformation matrix is applied to the lattice constellation in order to maximize both the diversity order and the minimum product distance of the complex N-D lattice [17]. It was shown in [17] that by increasing the dimensionality N of the signal set, a significant gain can be achieved compared to a non-transformed signal set over fading channels. Hereafter, with a slight abuse of definition, the common terminology "rotation" shall be used, instead of "transformation". A rotation is actually a distance-preserving transformation, i.e., the distance between any two points unchanged after the transformation. In fact, all transformation matrices studied later on are unitary. Therefore, they can be considered as rotation matrices.

As a simple example to illustrate the SSD technique, Fig. 2.3 shows two different constellations: a conventional quadrature phase shift keying (QPSK) constellation in Fig. 2.3-(a) and a rotated QPSK constellation in Fig. 2.3-(b). Any signal point in

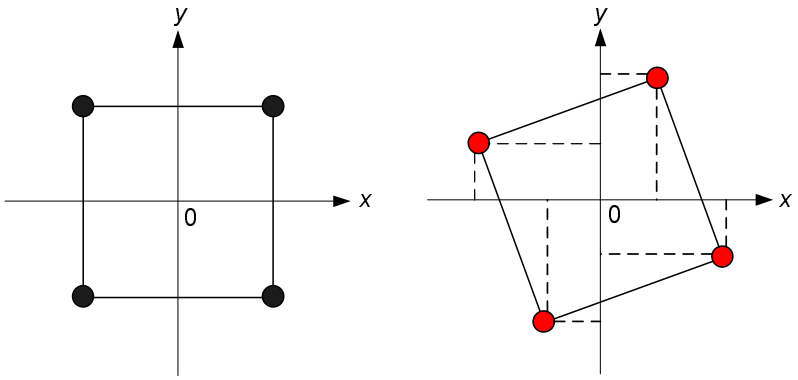


Figure 2.3 QPSK constellation and its rotated version.

these two constellations can be represented by two coordinates (x, y) in a complex one-dimensional signal space. Observe that for any signal point in a conventional QPSK constellation, given one coordinate x or y, it is not possible to unambiguously recognize it. On the other hand, with a suitable rotation angle, one can easily discriminate any signal point in a rotated QPSK constellation by just one coordinate x or y. In the context of digital communications over a fading channel, if a deep fade hits only one component of the transmitted signal, e.g., x or y, it is easier to reconstruct the transmitted signal by using the rotated constellation. In this case, it is said that the rotated QPSK constellation provides a signal space diversity of 2, while it is only 1 for the conventional QPSK constellation.

2.2.3 Spatial Diversity

Spatial diversity, also known as antenna diversity, can be obtained by using multiple antennas at the transmitter and/or the receiver. When the antennas are widely separated, the signals transmitted between different antenna pairs are faded independently and more independent paths are created. By using multiple antennas, two types of diversity can be achieved. The first one is called *receive diversity* when multiple antennas are employed at the receiver. On the other hand, multiple antennas at the transmitter provide *transmit diversity*.

With *receive diversity*, N_r antennas are employed at the receiver. There are two types of gain that can be achieved. The first one is called a *power gain*, which is due to the linearly increasing with N_r in the total received signal power when coherently combining N_r receive signals. The second gain is the *diversity gain*, since there are N_r independent signal paths from the transmitter to the receiver.

In the case of *transmit diversity*, there are N_t transmit antennas at the transmitter. This is a common situation in the downlink of a mobile system, since it is more feasible to have multiple antennas at the base station rather than having multiple antennas at the mobile phone. To achieve the diversity gain, a similar approach to repetition code can be used. For example, one can transmit the same symbol over N_t different antennas during N_t time intervals with the assumption that only one antenna is on and the rest are silent at any one time. More sophisticated methods can be used to exploit the diversity gain for the transmit diversity systems, which are referred as *space-time codes* [18, 19]

The use of multiple antennas at both the transmitter and the receiver might offer a higher potential, which results in a so-called multiple-input multiple-output (MIMO) system. A general MIMO channel model with N_t transmit antennas and N_r receive antennas is shown in Fig. 2.4. At a discrete time t, let $x_{t,i}$ be the transmitted signal at transmit antenna i. The received signal $r_{t,j}$ measured at receive antenna j, $1 \leq j \leq N_r$, at discrete time t is given by:

$$r_{t,j} = \sum_{i=1}^{N_t} h_{i,j} x_{t,i} + w_{t,j} \tag{2.6}$$

where $w_{t,j}$ is the zero-mean complex Gaussian noise with variance $N_0/2$ per dimension. The complex coefficient $h_{i,j}$ is the channel gain from transmit antenna i to receive antenna j. Under the assumption of Rayleigh fading, $h_{i,j}$ is $\mathcal{CN}(0, 1)$

2.2.4 Frequency Diversity

Over a *frequency selective* fading channel, one can exploit frequency diversity. This is because the transmitted signal arrives over a number of time intervals. As

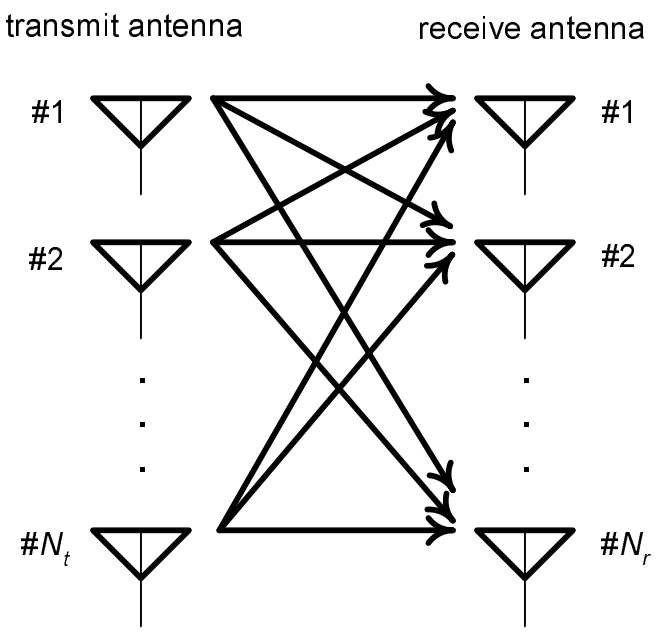


Figure 2.4 A MIMO channel model.

an example, consider a one-shot communications scenario, where only a signal $x[1]$ is transmitted. Assume that the channel is represented by L channel filter taps. The received signal is then given as:

$$r[l] = h_l[l]x[1] + w[l], \;\; 1 \leq l \leq L. \tag{2.7}$$

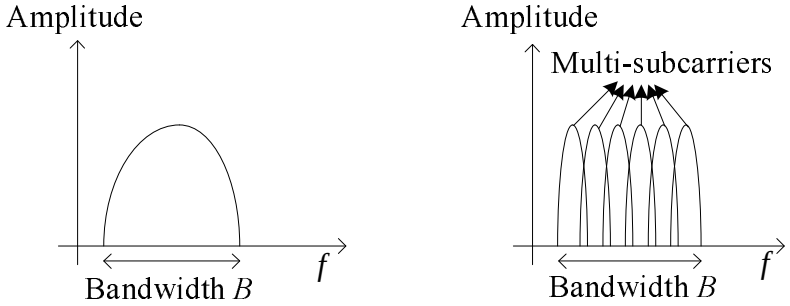
It can be seen that the information signal $x[1]$ is now transmitted over L independent paths, which provides a diversity order of L. In the case that signals are transmitted more frequently, the current signal is interfered with the other ones, due to the dispersion in time of the signals. This results in *inter-symbol interference* (ISI). Three common methods to obtain the frequency diversity and at the same time, deal with the ISI over *frequency selective* fading channel are as follows:

- *Single-carrier transmission with equalization*: This is the conventional method to deal with the ISI channel. As shown in Fig. 2.5-(a), the signals in single-carrier transmission are transmitted over only single carrier spanning the whole available spectrum B. By using an equalizer, the receiver manages to detect the transmitted signal, while depressing the interferences from the other symbols,

or equivalently, combating the ISI.

- *Multi-carrier transmission*: The basic idea of multi-carrier transmission is to divide the available spectrum into several subchannels (or subcarriers) as illustrated in Fig. 2.5-(b). The data is then shared among subcarriers and simultaneously transmitted. The important feature of multi-carrier system is that it transforms the broadband *frequency selective* fading channel into a set of correlated *frequency non-selective* (flat) fading subchannels, therefore significantly reducing the receiver complexity in both equalization and detection. Under a multi-carrier system, there is no ISI, but subchannels which are separated far apart fade independently and therefore, frequency diversity exists. To obtain a high spectral efficiency, the frequency responses of the subcarriers can be overlapped, as shown in Fig. 2.5-(b).

- *Direct sequence spread spectrum*: In this approach, the information signal is modulated by a pseudonoise sequence and then transmitted over a bandwidth B much larger than the data rate. The availability of such a larger bandwidth is made possible by multiplexing many distinct users simultaneously both in time and frequency. In this technique, users' signals are separated by means of pseudonoise sequences. If the data rate is low, ISI is negligible. The frequency diversity can be extracted by using the Rake receiver [12].



a) Frequency domain signal in single-carrier transmission.

b) Frequency domain signal in multi-carrier transmission.

Figure 2.5 Frequency domain signals in single-carrier and multi-carrier systems.

In this dissertation, the implementation of a multi-carrier modulation technique, named as orthogonal frequency division multiplexing (OFDM), is considered and it is briefly reviewed in the following.

OFDM

OFDM is a multi-carrier modulation technique that has been widely adopted in high data rate communication systems over *frequency selective* fading channels, including Digital Subscriber Lines (DSL), Wireless LANs (802.11a/g/n), Digital Video Broadcasting, and WiMAX-Worldwide Interoperability for Microwave Access. The basic idea behind OFDM is to divide the high-rate stream with short duration T, or equivalently, with a large bandwidth $B = 1/T$, into N lower-rate substreams with duration $T_s = N \times T$. A total of N different streams are transmitted simultaneously in subchannels (or subcarriers), with the narrow bandwidth of B/N each. Assume that the delay spread T_d of the channel is much greater than T, which leads to a severe ISI. By choosing a sufficiently large number of subcarriers N, the subcarrier bandwidth B/N would be much less than the coherence bandwidth $B_h = 1/(2T_d)$. Then each subchannel experiences *frequency non-selective* fading and the ISI is negligible. However, since all streams are transmitted at the same time, one needs to design orthogonal subchannels in order to to distinguish them at the receiver side, or equivalently, to avoid Inter-Carrier Interference (ICI).

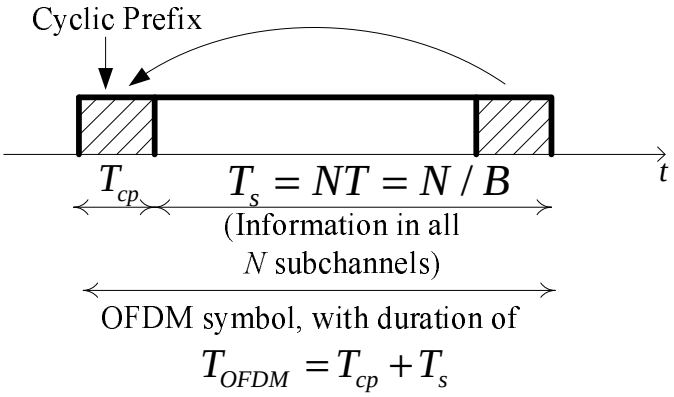


Figure 2.6 An OFDM symbol with cyclic prefix.

The key to make the parallel transmission over orthogonal subchannels possible is the use of a cyclic prefix: A copy of the last part with duration T_{cp} of an OFDM symbol is prepended to the transmitted symbol, as shown in Fig. 2.6. Since the cyclic prefix acts as a guard time, the transmitted OFDM signal would avoid ISI completely, as long as $T_{cp} > T_d$. Furthermore, the cyclic prefix makes the channel provide a circular convolution, which plays a crucial role to eliminate ICI.

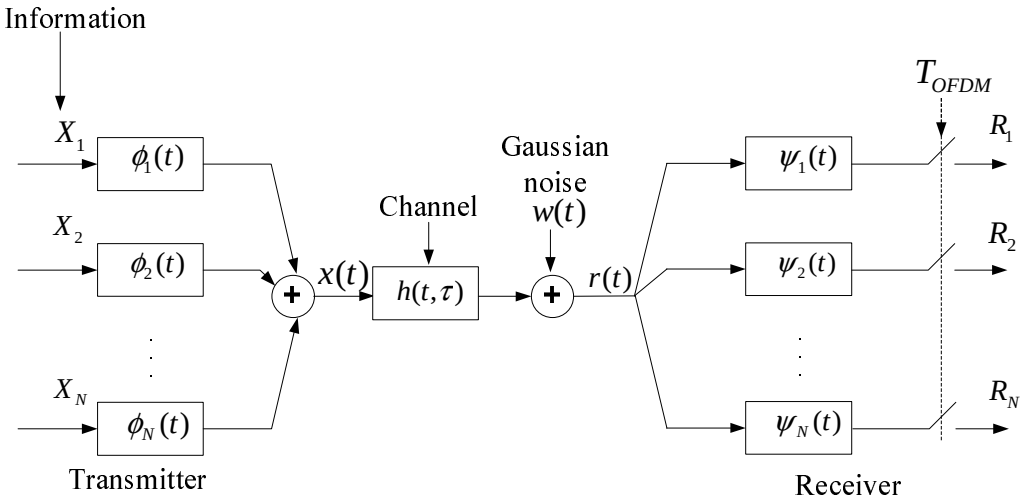


Figure 2.7 A continuous-time baseband-equivalent model of an OFDM system.

To give an insight how an OFDM system works, Fig. 2.7 shows a continuous-time baseband-equivalent model of an OFDM system. Let X_n, $1 \leq n \leq N$, be the nth information symbol at the transmitter. The transmitter uses the following rectangular pulse modulated on the subcarrier $f_n = (n-1)B/N$ to carry the information X_n:

$$\phi_n(t) = \begin{cases} \frac{1}{\sqrt{T_s}} \exp\left(j2\pi f_n(t - T_{cp})\right) & \text{if } t \in [0, T_{OFDM}] \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

Note that $\phi_n(t) = \phi_n(t + T_s)$ when t is within the cyclic prefix $[0, T_{cp}]$. Furthermore, the orthogonality among different subcarriers is maintained because of the following reason:

$$\int_{T_{cp}}^{T_{OFDM}} \phi_n(t)\phi_m^{\dagger}(t)\mathrm{d}t = 0 \;\text{ when }\; n \neq m \tag{2.9}$$

The transmitted baseband OFDM signal is then given as:

$$x(t) = \sum_{n=1}^{N} X_n \phi_n(t) \tag{2.10}$$

Over a *frequency selective* fading channel $h(t,\tau)$, the received signal is given as:

$$r(t) = (h \star x)(t) = \int_0^{T_{cp}} h(t,\tau)x(t-\tau)\mathrm{d}\tau + w(t) \tag{2.11}$$

where $w(t)$ is additive white Gaussian noise with the power spectral density $N_0/2$. The OFDM receiver consists of N matched filters $\{\psi_n(t)\}$, each matched to the last part $[T_{cp}, T_{OFDM}]$ of $\phi_n(t)$ as follows:

$$\psi_n(t) = \begin{cases} \phi_n^\dagger(T_{OFDM} - t) & \text{if } t \in [0, T_{OFDM} - T_{cp}] \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

It means that the cyclic prefix is removed at the receiver. Since the cyclic prefix contains all ISI from previous OFDM symbols, the outputs of the matched filters contain no ISI and one obtains the following N parallel channels:

$$R_n = H_n X_n + W_n \tag{2.13}$$

In (2.13), H_n is simply equal to the frequency response $H(f)$ at f_n and W_n is $\mathcal{CN}(0, N_0)$ representing the Gaussian noise.

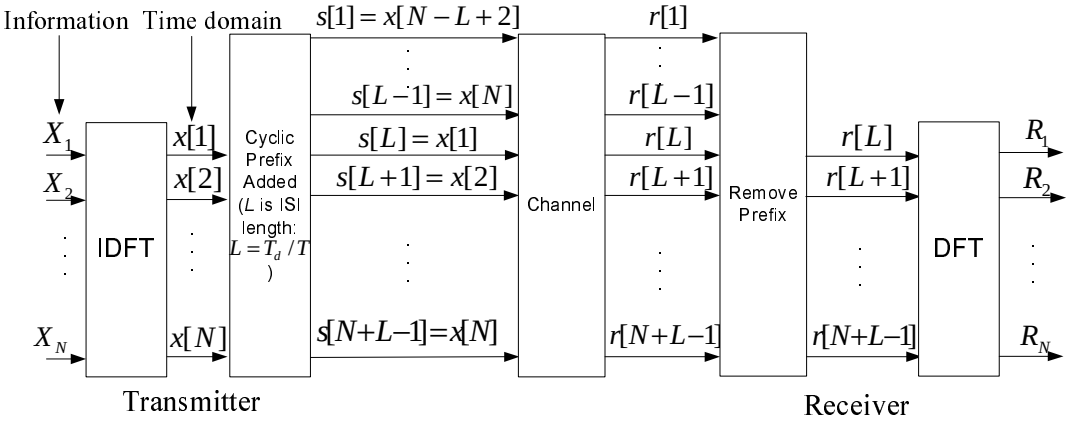


Figure 2.8 A discrete-time model of an OFDM system using IDFT and DFT.

The above model requires N independent radio frequency (RF) units and demodulation paths. In order to overcome this shortcoming, one can employ the efficient computational technique known as Discrete Fourier Transform (DFT) and inverse DFT (IDFT) at the baseband modulation and demodulation for an OFDM system. The discrete-time model for this method is shown in Fig. 2.8. It should be noted that

when N is a power of 2, IDFT and DFT can be very efficiently implemented using Fast Fourier Transform (FFT).

2.3 Bit-Interleaved Coded Modulation with Iterative Decoding (BICM-ID)

It is well known that over a fading channel, the diversity order of a communications system plays an important role in determining its error performance [20]. The potential of bit-interleaved coded modulation (BICM) systems over a fading channel was first recognized by Zehavi in [1] where he suggested a coded system built from a convolutional encoder followed by a random bit interleaver. Recently, an analytical framework for the evaluation of Zehavi's BICM systems was presented in [2]. The advantage of a BICM system is that it treats coding and modulation as separate components, hence, the code and the modulator can be flexibly selected. The diversity order can now be maximized by using the best convolutional code, which provides the largest Hamming distance d_H for given code rate and constraint length.

Recently, the success of turbo codes has demonstrated the advantages of iterative processing in the decoding of concatenated coding schemes. Reference [5] appears to be the first paper that studies iterations between the demodulator and the decoder to provide a significant performance gain over non-iterative BICM systems. In that paper, a simple iterative decoding processing with a hard-decision feedback was proposed. By using the soft-decision feedback in the iterative processing, it has been shown in [7] that the performance of BICM-ID can be further improved. Thereafter, BICM-ID has become a hot topic in wireless communications research.

In the following, a model of BICM-ID with soft-decision feedback is introduced, along with different tools to analyze the system. The model is quite general in the sense that it can apply to any specific system considered in subsequent chapters.

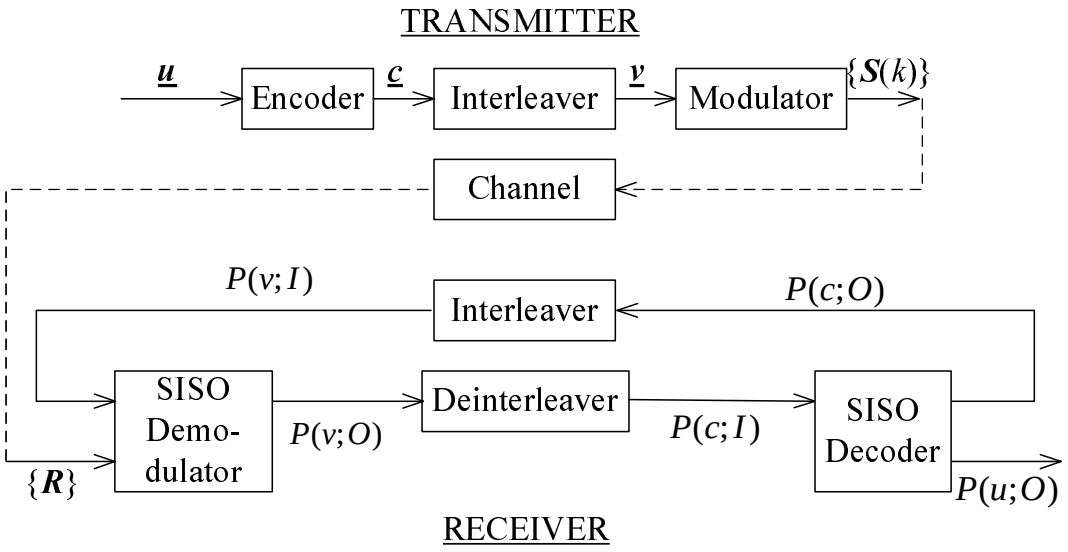


Figure 2.9 Block diagram of a BICM-ID system.

2.3.1 System Model

Figure 2.9 plots the general block diagram for a BICM-ID system. At the transmitter, the binary information sequence $\underline{\boldsymbol{u}}$ is first encoded into a coded sequence $\underline{\boldsymbol{c}}$ by a standard convolutional code. The coded sequence $\underline{\boldsymbol{c}}$ is then interleaved by a bit-wise interleaver to become the interleaved sequence $\underline{\boldsymbol{v}}$. The purpose of the interleaver is to break the fading correlation and enhance the independent among the coded bits. In general, the modulator maps a group of M_c binary coded bits to one of $K = 2^{M_c}$ signals, $\{\boldsymbol{S}(\kappa)\}_{\kappa=1}^{K}$, according to some mapping rule ξ. The set $\{\boldsymbol{S}(\kappa)\}_{\kappa=1}^{K}$ shall be referred to as the constellation $\boldsymbol{\Psi}$. The group of M_c coded bits carried by each signal $\boldsymbol{S}(\kappa)$ is called the *label* of $\boldsymbol{S}(\kappa)$. Furthermore, the Hamming weight of a label is defined as the number of bits "1" in the label, whereas the label Hamming distance between two labels is the number of bit positions that the two labels differ. As an example, Fig. 2.10 shows the 8-phase shift keying (8-PSK) constellation $\boldsymbol{\Psi}$ with semi-set partitioning (SSP) mapping ξ that can be employed at the modulator in a conventional BICM-ID system [7]. This constellation includes $K = 8$ signal points $\boldsymbol{S}(\kappa)_{\kappa=1}^{K}$, where each signal point carries $M_c = 3$ coded bits. In each symbol duration,

depending on a mapping rule ξ, a signal $\boldsymbol{S}(\kappa)$ shall be chosen accordingly to carry a group of M_c coded bits at the output of the interleaver. For instance, by using the SSP mapping shown in Fig. 2.10, if (0,1,0) are three coded bits at the output of the interleaver, the signal $\boldsymbol{S}(5)$ with the label 010 will be transmitted over the channel.

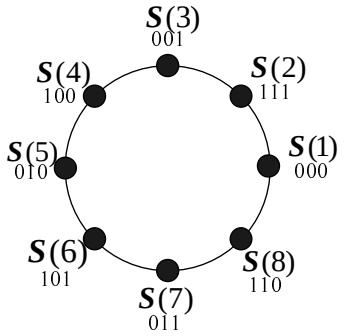


Figure 2.10 The 8-PSK constellation $\boldsymbol{\Psi}$ with semi-set partitioning (SSP) mapping ξ.

In general, depending on a specific channel and application, the constellation $\boldsymbol{\Psi}$ will be properly chosen to effectively combat fading, noise and interference from other users. Loosely speaking, the modulator is a method to adapt to the physical channel. Therefore, the signal symbols of the constellation $\boldsymbol{\Psi}$, and as a consequence, the mapping rule ξ, may span over time, space and frequency. For example, with space-time codes proposed for MIMO channels in [18,19], the signals $\{\boldsymbol{S}(\kappa)\}_{\kappa=1}^{K}$ are defined over both time and spatial domains. On the other hand, in multi-carrier systems employing orthogonal frequency division multiplexing (OFDM) [21], the signals are composed of many subcarriers (i.e., using different frequencies).

An ideal BICM receiver would perform a maximum likelihood decoding, which requires joint demodulation and convolutional decoding. Due to the presence of the random bit interleaver, the implementation of this receiver is intractable in practice. Instead, as shown in Fig. 2.9, the suboptimal receiver of the system includes two separate blocks: the soft-input soft-output (SISO)[1] demodulator and the SISO covo-

[1]Throughout this dissertation, the abbreviation SISO always stands for soft-input soft-output. In the literature, SISO can also be used for single-input single-output.

lutional decoder. The SISO channel decoder uses the *maximum a posteriori* (MAP) decoding algorithm in [22]. There is an iterative processing performed between the SISO decoder and the SISO demodulator. It has been widely acknowledged in the literature that error performance obtained from the iterative techniques can approach the full maximum likelihood performance [23].

Using the same notation as in [7, 22], the *a priori* information and the *extrinsic* information of a random variable z are denoted by $P(z;I)$ and $P(z;O)$, respectively. At the first iteration, since there is no *a priori* information at the input of the demodulator, the information at the output of the demodulator is simply the *a posteriori* probability. Given the received signal $\boldsymbol{R}$, this information for M_c coded bits v_k, $1 \leq k \leq M_c$, is computed using the MAP criterion as follows:

$$P(v_k = b|\boldsymbol{R}) = \sum_{\boldsymbol{S}(\kappa)\in\boldsymbol{\Psi}_b^k} P\left(\boldsymbol{S}(\kappa)|\boldsymbol{R}\right). \tag{2.14}$$

In (2.14), $P\left(\boldsymbol{S}(\kappa)|\boldsymbol{R}\right)$ is the *a posteriori* probability of the transmitted signal $\boldsymbol{S}(\kappa)$ given the received signal $\boldsymbol{R}$. The set $\boldsymbol{\Psi}_b^k$, $b \in \{0,1\}$, denotes the subset of $\boldsymbol{\Psi}$ that contains all signals whose labels have the value b at the kth position, $1 \leq k \leq M_c$. Using Bayes rule [24], the *a posterior* probability $P\left(\boldsymbol{S}(\kappa)|\boldsymbol{R}\right)$ can be computed as:

$$P\left(\boldsymbol{S}(\kappa)|\boldsymbol{R}\right) = \frac{p(\boldsymbol{R}|\boldsymbol{S}(\kappa))P(\boldsymbol{S}(\kappa))}{p(\boldsymbol{R})}, \tag{2.15}$$

where $P(\boldsymbol{S}(\kappa))$ is the *a priori* probability that the signal $\boldsymbol{S}(\kappa)$ is transmitted at the transmitter and $p(\boldsymbol{R}|\boldsymbol{S}(\kappa))$ is the probability density function of the received signal $\boldsymbol{R}$ given the signal $\boldsymbol{S}(\kappa)$ was transmitted. The denominator of (2.15) can be expressed as:

$$p(\boldsymbol{R}) = \sum_{\kappa=1}^{K} p(\boldsymbol{R}|\boldsymbol{S}(\kappa))P(\boldsymbol{S}(\kappa)). \tag{2.16}$$

From (2.15) and (2.16), it can be observed that the computation of the *a posterior* probability $P\left(\boldsymbol{S}(\kappa)|\boldsymbol{R}\right)$ requires knowledge of the *a priori* probability $P(\boldsymbol{S}(\kappa))$ and the conditional probability density function $p(\boldsymbol{R}|\boldsymbol{S}(\kappa))$. Hence, $P(v_k = b|\boldsymbol{R})$ can be determined as follows:

$$P(v_k = b|\boldsymbol{R}) \sim \sum_{\boldsymbol{S}(\kappa)\in\boldsymbol{\Psi}_b^k} p(\boldsymbol{R}|\boldsymbol{S}(\kappa))P(\boldsymbol{S}(\kappa)). \tag{2.17}$$

The notation $\sim$ in (2.17) indicates replacement by an equivalent statistic. For the first iteration, the transmitted signals can be assumed to be equally likely, i.e., $P(\boldsymbol{S}(\kappa))$ can be set to $1/K$.

From the second iteration, the *extrinsic* information $P(c_k; O)$ produced by SISO decoder of the coded bits is transferred back to the demodulator. After being interleaved, it becomes the *a priori* information $P(v_k; I)$ at the input of the demodulator. Since the (ideal) interleaver makes M_c bits in one symbol independent, the *a priori* information $P(\boldsymbol{S}(\kappa))$ of each signal $\boldsymbol{S}(\kappa) \in \boldsymbol{\Psi}$ can be obtained as:

$$P(\boldsymbol{S}(\kappa)) = P\left(v_1(\boldsymbol{S}(\kappa)), \ldots, v_{M_c}(\boldsymbol{S}(\kappa))\right) = \prod_{k=1}^{M_c} P(v_k = v_k(\boldsymbol{S}(\kappa)); I), \tag{2.18}$$

where $v_k(\boldsymbol{S}(\kappa)) \in \{0, 1\}$, $1 \leq k \leq M_c$, is the value of the kth bit in the label of $\boldsymbol{S}(\kappa)$. Using (2.18) and (2.17), the *extrinsic a posteriori* information from the second iteration can be computed as:

$$\begin{aligned} P(v_k = b; O) &= \frac{P(v_k = b|\boldsymbol{R})}{P(v_k = b; I)} \sim \frac{\sum_{\boldsymbol{S}(\kappa) \in \boldsymbol{\Psi}_b^k} p(\boldsymbol{R}|\boldsymbol{S}(\kappa)) P(\boldsymbol{S}(\kappa))}{P(v_k = b; I)} \\ &= \sum_{\boldsymbol{S}(\kappa) \in \boldsymbol{\Psi}_b^k} \left[p(\boldsymbol{R}|\boldsymbol{S}(\kappa)) \prod_{j \neq k} P(v_j = v_j(\boldsymbol{S}(\kappa)); I) \right]. \end{aligned} \tag{2.19}$$

Observe from (2.19) that the bit metric $P(v_k = b; O)$ is computed from the *a priori* probabilities $P(v_j; I)$ of the other labelling bits ($j \neq k$) of the same channel symbol. The regenerated bit metrics are then delivered to the SISO decoder. The SISO decoder treats that information as the updated *a priori* information for the coded bits and the iterative demodulation and decoding keep running. The hard-decision for the information bits can be made at any iteration based on the *extrinsic* information $P(u_k; O)$.

Finally, it is worth mentioning that the complexity of the MAP demodulator grows exponentially with the number of coded bits per symbol M_c, which becomes intractable for a large value of M_c (i.e., a large constellation $\boldsymbol{\Psi}$). Therefore, the suboptimal low-complexity methods of SISO demodulation might be more suitable. This issue will be further discussed for each specific system.

2.3.2 Performance Evaluation

Due to the large coding gain produced by the iterative processing in BICM-ID, one is most interested in the asymptotic error performance to which the system's performance converges. This usually happens at high enough SNRs and after a number of iterations. Such asymptotic performance can be obtained by assuming that the iterations between the SISO decoder and the SISO demodulator work perfectly, or equivalently, one has the perfect *a priori* information of the coded bits fed-back to the demodulator [7, 25]. By relating the behavior of BICM systems similar to a memoryless binary-input output-symmetric (BIOS) channel [2], the union bound on the bit error probability (BEP) for a rate-k_c/r_c convolutionally coded modulation can be written in a general form as [2]:

$$P_b \leq \frac{1}{k_c} \sum_{d=d_H}^{\infty} c_d f(d, \mathbf{\Psi}, \xi). \tag{2.20}$$

In (2.20), c_d is the total information weight of all the error events at Hamming distance d and d_H is the free Hamming distance of the code. The function $f(d, \mathbf{\Psi}, \xi)$ is the average pairwise error probability (PEP), which depends on the Hamming distance d, the constellation $\mathbf{\Psi}$ and the mapping rule ξ. In the following, the computation of the function $f(d, \mathbf{\Psi}, \xi)$ from the pairwise error probability (PEP) of two codewords is briefly discussed.

Let $\underline{\boldsymbol{c}}$ and $\breve{\underline{\boldsymbol{c}}}$ denote the input and the estimated sequences of coded bits, respectively, with Hamming distance d between them. Without loss of generality, assume that $\underline{\boldsymbol{c}}$ and $\breve{\underline{\boldsymbol{c}}}$ differ in the first d consecutive bits. Furthermore, with the use of a sufficiently long interleaver, it can be assumed that each of these different bits appears in a different block of M_c bits before mapping to the constellation symbols in $\mathbf{\Psi}$. Therefore, these binary sequences correspond to codewords $\underline{\boldsymbol{S}}$ and $\breve{\underline{\boldsymbol{S}}}$, defined as sequences of d signal points as $\underline{\boldsymbol{S}} = [\boldsymbol{S}_1, \ldots, \boldsymbol{S}_d]$ and $\breve{\underline{\boldsymbol{S}}} = \left[\breve{\boldsymbol{S}}_1, \ldots, \breve{\boldsymbol{S}}_d\right]$. Here, $\boldsymbol{S}_e$ and $\breve{\boldsymbol{S}}_e$, $1 \leq e \leq d$, belong to the constellation $\mathbf{\Psi}$. In general, the pair-wise error probability (PEP) $P(\underline{\boldsymbol{S}} \rightarrow \breve{\underline{\boldsymbol{S}}})$ needs to be computed. The function $f(d, \mathbf{\Psi}, \xi)$ is then obtained by averaging over all the codewords $\underline{\boldsymbol{S}}$ and $\breve{\underline{\boldsymbol{S}}}$ taking into account all the

corresponding bit positions [2]. The problem is that this computation quickly becomes infeasible with large values of d and the constellation size K. By assuming the ideal feedback from the decoder to the demodulator, it follows that the labels of two signal points $\boldsymbol{S}_e$ and $\breve{\boldsymbol{S}}_e$ differ in only 1 bit. Furthermore, with memoryless channels (i.e., the channel imperfectness is independent from symbol to symbol), the function $f(d, \boldsymbol{\Psi}, \xi)$ can be simply calculated by averaging over the constellation $\boldsymbol{\Psi}$.

The above evaluation of the system's error performance will be applied throughout this dissertation for various BICM-ID systems. The advantage of this method is that it can provide an insight on how to optimize the system error performance, as shown later for each specific system. It should be noted that there exist some research works that analyze the error performance of BICM systems, either with the assumption of an ideal infinite-length interleaver [26] or a uniform finite-length interleaver [27]. However, these studies do not provide a practical design guideline, which might not be useful in optimizing the system error performance.

2.3.3 EXIT Chart Analysis

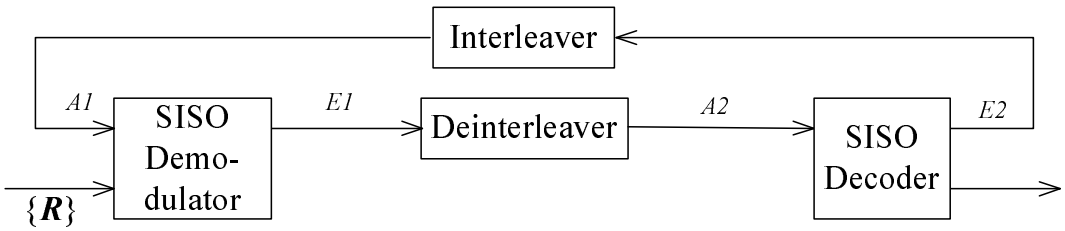


Figure 2.11 Input and output extrinsic information.

The convergence of a BICM-ID system can also be analyzed by the histogram method using extrinsic information transfer (EXIT) chart [11]. Following the notations in [11], let ($A1$ and $E1$) and ($A2$ and $E2$) denote the input and output extrinsic information in log-likelihood ratio domain associated with the SISO demodulator and SISO decoder, respectively (see Fig. 2.11).

From simulations of iterative decoder, it is observed in [11] that the priori input $A1$ fed back from SISO decoder is Gaussian distributed. In particular, with the known

unmapped bit $v \in \{0, 1\}$, $A1$ can be expressed as follows [11]:

$$A1 = \sigma_{A1}^2/2 \cdot (2v - 1) + n_{A1}, \tag{2.21}$$

where n_{A1} is a Gaussian random variable with mean zero and variance σ_{A1}^2. Hence, the mutual information $I_{A1} = I(A1; V)$ between the priori knowledge $A1$ and the coded bit V at the input of the demodulator is calculated as:

$$I_{A1}(\sigma_{A1}) = \frac{1}{2} \sum_{v=0,1} \int_{-\infty}^{+\infty} p_{A1}(\beta|V = v) \cdot \log_2 \frac{2p_{A1}(\beta|V = v)}{p_{A1}(\beta|V = 1) + p_{A1}(\beta|V = 0)} \mathrm{d}\beta \tag{2.22}$$

where

$$p_{A1}(\beta|V = v) = \frac{1}{\sqrt{2\pi}\sigma_{A1}} \exp\left(-\frac{(\beta - \frac{\sigma_{A1}^2}{2} \cdot (2v - 1))}{2\sigma_{A1}^2}\right). \tag{2.23}$$

It can be verified that $I_{A1}(\sigma_{A1})$ is a monotonically increasing function of σ_{A1} with $I_{A1}(0) = 0$ and $I_{A1}(\infty) = 1$. Mutual information can also be used to quantify the extrinsic information $E1$ and the coded bit V at the output of the demodulator, which is denoted as $I_{E1} = I(E1; V)$. Given I_{A1} calculated from a certain value of σ_{A1}, one can obtain I_{E1} as a function of I_{A1} as:

$$I_{E1} = T_1(I_{A1}, SNR). \tag{2.24}$$

The function $T_1(\cdot)$ in (2.24) is called the extrinsic information transfer characteristic of the demodulator and can be determined by means of Monte Carlo simulation of the demodulator (histogram measurements).

Similarly, let I_{E2} and I_{A2} be the mutual information representing the a priori knowledge and the extrinsic information of the coded bits at the input and output of the SISO decoder, respectively. The extrinsic transfer characteristic of the SISO decoder $I_{E2} = T_2(I_{A2})$ can also obtained by means of Monte Carlo simulation. Due to random bit interleaving, this function does not depend on the channel property, and is only determined by the convolutional code itself.

To visualize the exchange of extrinsic information between the demodulator and decoder in iterative processing, one needs to draw the demodulator and decoder

transfer characteristics into a single diagram, which is referred to as Extrinsic Information Transfer Chart. As an example, Fig. 2.12 provides EXIT charts for a BICM-ID system over a frequency non-selective Rayleigh fading channel employing 8-PSK constellation $\mathbf{\Psi}$ with the semi-set partitioning (SSP) mapping rule ξ shown in Fig. 2.10, along with a rate-2/3, 4-state convolutional code whose generator matrices are $\mathbf{g}_1 = (6, 2, 6)$ and $\mathbf{g}_2 = (2, 4, 4)$ [13].

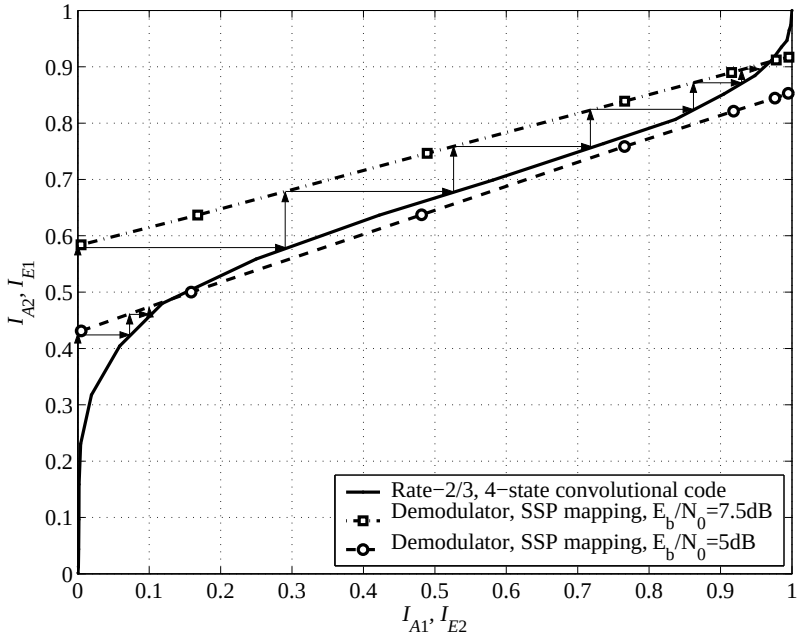


Figure 2.12 EXIT charts of the BICM-ID system employing 8-PSK constellation and SSP mapping at different E_b/N_0.

It can be seen in Fig. 2.12 that on the ordinate, the inner extrinsic output I_{E_1} becomes the outer a priori input I_{A2} (interleaving does not change mutual information). On the abscissa, the outer extrinsic output I_{E2} becomes the inner a priori input I_{A1}. For comparison, the demodulator characteristic is provided at two values of signal-to-noise ratio (SNR), $E_b/N_0 = 7.5$dB and $E_b/N_0 = 5$dB, where E_b is the energy per information bit.

Observe from Fig. 2.12 the mutual information I_{A1} at the first iteration is zero for both SNRs. It is due to the reason that one has no priori knowledge about the coded bits at the first iteration. At $E_b/N_0 = 7.5$dB, the demodulator produces an

extrinsic information $E1$ with the mutual information $I_{E1} = 0.58$, which is much higher that that at $E_b/N_0 = 5$dB. After being deinterleaved, the extrinsic output of the demodulator becomes the a priori input to the decoder, i.e., $I_{A2} = I_{E1}$. The extrinsic information at the output of the decoder then becomes the a priori information for the demodulator, i.e., $I_{A1} = I_{E2}$, after being interleaved and the information exchange process continues in an iterative manner. However, it can be observed that at $E_b/N_0 = 5$dB, the demodulator and the decoder transfer characteristics do intersect after a few iterations and the decoding trajectory gets stuck at quite low mutual information, which results in high error performance. On the other hand, at higher SNR, i.e., $E_b/N_0 = 7.5$dB, the demodulator transfer characteristic is raised and there is an open tunnel between the demodulator and the decoder transfer characteristics, which allows for convergence of iterative demodulation and decoding towards higher mutual information, or equivalently, lower error performance.

From the above example, it is clear that EXIT charts provide a very effective tool to predict the convergence behavior of a BICM-ID system. Further more, as pointed out in [28], by taking into account the mutual information on the information bits, EXIT charts can also be used to predict the error performance of a BICM-ID system.

3. A Novel Multi-Dimensional Mapping of 8-PSK for BICM-ID[1]

In a conventional BICM-ID system, the deployment of a mapping rule ξ is only considered in a real one or two-dimensional constellation $\mathbf{\Psi}$, such as pulse amplitude modulation (PAM), phase shift keying (PSK), or quadrature amplitude modulation (QAM) constellations. Under this circumstance, a substantial amount of work has been carried out to address the mapping problem for real one or two-dimensional constellations in single-antenna systems [7–9, 11, 31–34] as well as in multiple-antenna systems [35, 36]. It has been shown that the signal constellation $\mathbf{\Psi}$ and the mapping rule ξ play an important role in determining the bit error rate (BER) performance for a BICM-ID system.

As mentioned in Chapter 2, the binary mapping rule ξ and the constellation $\mathbf{\Psi}$ might span over time, space or frequency in order to provide a more reliable transmission. Recently, considerable attention has been paid to the use of multi-dimensional mapping where ξ and $\mathbf{\Psi}$ span over time, to further improve the error performance of BICM-ID. In particular, reference [25] studied hypercube mappings of QPSK constellation for BICM-ID in single-antenna (i.e., single-input single-output) systems. It was shown in [25] that a significant coding gain can be achieved without any bandwidth or power expansion and with a reasonable increase in the system complexity, which is required to address the soft-output demodulation of the multi-dimensional constellation. A parallel research work was also carried out in [37] but it only concentrates on an AWGN channel. The technique of multi-dimensional mapping

[1] A part of this chapter was presented in [29, 30]

has also been extended to BICM-ID with multiple-input multiple-output (MIMO) systems in [23, 38, 39]. It was observed in [23] that performance of the system with multi-dimensional mapping can achieve near turbo-code performance with only a simple convolutional code.

In general, symbol mapping is simply the mapping from a binary sequence to the sequence of complex constellation symbols in $\mathbf{\Psi}$. As mentioned earlier, this procedure is crucial for the error performance of BICM-ID. However, only a few efforts have been devoted to address the problem of *optimal* mapping design. Except references [25] and [37], all research work related to the mapping design problem is only based on some computer searching techniques. For example, the brute-force computer searching can be carried out for the single-antenna systems that use a low-order signal constellation and the conventional two-dimensional mapping [7, 32, 33]. In the case of multi-dimensional mapping with a high-order constellation in both single-antenna and multiple-antenna systems, the exhaustive computer search to find the optimal mapping is impossible due to the huge complexity of the search. As an example, with a very simple single-antenna system employing 8-PSK constellation, there are $64! = 1.27 \times 10^{89}$ possible real 4-dimensional (or complex 2-dimensional) mappings. The binary switching algorithm (BSA), originally adopted in [40] for mapping optimization of the conventional BICM-ID, can then be used to find a good mapping. However, it should be emphasized that the BSA only gives locally optimal mappings. Furthermore, for a large constellation (such as multi-dimensional constellations), the BSA quickly becomes intractable and a more complicated searching inside a selected list must be applied [23]. Another searching technique, related to the quadratic assignment problem (QAP), was also presented in [36]. This technique, however, has the same disadvantages with that of the BSA for larger constellations.

It is observed that all the available computer searching techniques do not take into account the symmetry of the signal constellations. They can be applied to the constellation with any shape and with an arbitrary number of signal points. On the other hand, the algorithm proposed in [25] to construct the global optimal mapping

for multi-dimensional BICM-ID is only valid for QPSK constellation and cannot be used for any higher-order signal constellations.

Motivated by the above observations and discussions, the contribution of this chapter is to introduce a multi-dimensional mapping design for BICM-ID employing 8-PSK constellation over a *frequency non-selective* Rayleigh fading channel, which is important for practical applications. The proposed algorithm exploits the symmetry of the 8-PSK constellation and the simple fact that the 8-PSK constellation can be decomposed into two QPSK constellations. In a real 2-dimensional signal space, it can be easily verified that the proposed mapping is indeed the global optimal mapping. For a higher dimension, unfortunately, whether the proposed mapping is a global optimal mapping still remains to be answered. Nevertheless, by comparing the performance of the proposed mapping with an unachievable lower bound, it is shown that there is only a very small gap between the performance of the proposed mapping and the bound. It is therefore conjectured that the proposed mapping is actually the global optimal mapping.

The chapter starts with a brief introduction of a BICM-ID system using multi-dimensional constellation and mapping. The union bound on the bit error probability with the assumption of error-free feedback is described next. From the derivation of the error bound, distance criteria that characterize the influence of the signal constellation and mapping to the error performance are presented. Based on the distance profile of the 8-PSK constellation, unachievable lower bounds for these distance criteria are established. The proposed mapping design algorithm is then presented. Finally, numerical and simulation results are provided to demonstrate the advantages of the proposed mapping.

It should be noted that in this chapter, a conventional one or two-dimensional mapping refers to a mapping in a real one or two-dimensional space, respectively. The later is equivalent to a mapping in a complex one-dimensional space.

3.1 BICM-ID with Multi-Dimensional Mapping

In a conventional BICM-ID system, the transmitter and receiver are similar to those described in Fig. 2.9 in Chapter 2. A signal mapping ξ is the mapping from a group of binary coded bits to a constellation symbol in a real one or two-dimensional constellation, as similar to the 8-PSK constellation employing semi-set partitioning (SSP) mapping shown in Fig. 2.10 in Chapter 2. In general, for conventional BICM-ID systems, every group of $M_c = m = \log_2 M$ coded bits is mapped to one signal point in a real one or two-dimensional M-ary constellation, denoted as Ω.

In contrast to these conventional systems, by considering a multiple symbol interval, one could simultaneously map $M_c = Nm$ coded bits, $N > 1$, to N consecutive M-ary signal points. As a result, a larger constellation $\mathbf{\Psi}$ in real $2N$-dimensional signal space, or equivalently, complex N-dimensional (N-D) signal space, is created, having in total $K = M^N$ signal points. Each N-D signal point can be represented by the following vector[2]:

$$\boldsymbol{s}(\kappa) = [s_{\kappa,1}, s_{\kappa,2}, \ldots, s_{\kappa,2n-1}, s_{\kappa,2N}], 1 \leq \kappa \leq K. \tag{3.1}$$

Equivalently,

$$\boldsymbol{s}(\kappa) = [q_{\kappa,1}, \ldots, q_{\kappa,N}], \tag{3.2}$$

where $q_{\kappa,p} = s_{\kappa,2p-1} + j s_{\kappa,2p}, 1 \leq p \leq N$, represents the pth complex M-ary symbol. Each signal $\boldsymbol{s}(\kappa)$ is labelled by $M_c = Nm$ bits as follows:

$$\boldsymbol{s}(\kappa) \sim \boldsymbol{a}(\kappa) = (a_{\kappa,1}, a_{\kappa,2}, \ldots, a_{\kappa,M_c}) = \xi^{-1}(\boldsymbol{s}(\kappa)), \tag{3.3}$$

where $a_{\kappa,k}, 1 \leq k \leq M_c$, is either 0 or 1 and ξ denotes the mapping from M_c bits to one N-D signal, i.e., $\boldsymbol{s}(\kappa) = \xi(\boldsymbol{a}(\kappa))$. Clearly, any two-dimensional mapping used in a conventional system with M-ary constellation is just a special case of the above general mapping. Moreover, there is no change in bandwidth efficiency due to the use of this general mapping. Let $\boldsymbol{r} = [r_1, \ldots, r_N]$ represent the received signal in

[2]Since a N-D signal point is represented by a row vector, bold lower letter $\boldsymbol{s}$ has been substituted for bold capital letter $\boldsymbol{S}$ to refer to a signal point in $\mathbf{\Psi}$

a N-D signal space. For a *frequency non-selective* slowly Rayleigh fading channel and coherent detection (i.e., the channel state information is perfectly known at the receiver), it suffices to write the received signal r_p as follows [7]:

$$r_p = h_p q_{\kappa,p} + w_p \tag{3.4}$$

where $1 \leq p \leq N$ and κ is the index of the transmitted signal. In (3.4), w_p is $\mathcal{CN}(0, N_0)$. Since CSI is perfectly known at the receive, one can consider the fading gain as the scalar h_p, which is the amplitude of a $\mathcal{CN}(0, 1)$ random variable[3]. This scalar is a Rayleigh random variable with the probability density function given by:

$$\lambda_{h_p}(h_p) = \begin{cases} 2h_p \exp(-h_p^2), & h_p \geq 0 \\ 0, & h_p < 0 \end{cases} \tag{3.5}$$

It is then straightforward to see that the random variable $z = h_p^2$ follows the exponential distribution with the probability density function is given as:

$$f_Z(z) = \exp(-z). \tag{3.6}$$

Two cases of fading amplitude are considered as follows: (i) The coefficients h_p are different for different M-ary symbols in one N-D signal (fast fading). Furthermore, they can be modeled as independent and identical distributed (i.i.d) Rayleigh random variables, i.e., fully interleaved *frequency non-selective* fading channel; (ii) The coefficients h_p are constant over N consecutive M-ary symbols, i.e., over the duration of the N-D signal, and change independently from one duration of the N-D signal to the next. The channel model in the second case is usually referred to as a quasistatic fading channel [18]. Note that in the case of AWGN channel, h_p equals 1 for all p.

The receiver in multi-dimensional BICM-ID systems includes the SISO demodulator and SISO decoder as described in Chapter 2. The SISO demodulator works in a multi-dimensional constellation $\boldsymbol{\Psi}$, with the conditional density function $p(\boldsymbol{r}|\boldsymbol{s}(\kappa))$

[3]In general, the channel gain representing the fading coefficient is $\mathcal{CN}(0, 1)$ for Rayleigh fading channels

given the channel gains $\{h_p\}$ and signal $\boldsymbol{s}(\kappa)$ was transmitted calculated as:

$$p(\boldsymbol{r}|\boldsymbol{s}(\kappa)) = \frac{1}{(\pi N_0)^N} \exp\left(-\frac{\sum_{p=1}^{N} ||r_p - h_p q_{\kappa,p}||^2}{N_0}\right). \tag{3.7}$$

3.2 Union Bound on The Bit Error Probability (BEP)

The union bound on the bit error probability (BEP) presented in Section 2.3.2, Chapter 2 for a general BICM-ID system can be applied straightforwardly for the BICM-ID system that employs a rate-k_c/n_c convolutional code, a multi-dimensional constellation $\boldsymbol{\Psi}$ and a mapping ξ considered in this chapter. In particular, in order to compute the upper bound in (2.20), one needs to consider the average pairwise error probability $f(d, \boldsymbol{\Psi}, \xi)$, which depends on the Hamming distance d, the multi-dimensional constellation $\boldsymbol{\Psi}$ and the mapping ξ.

Similar to the assumption in Chapter 2, consider $\underline{\boldsymbol{c}}$ and $\breve{\underline{\boldsymbol{c}}}$ as the input and the estimated coded sequences, respectively, with Hamming distance d between them and they differ in the first d consecutive bits. These binary sequences correspond to codewords $\underline{\boldsymbol{s}}$ and $\breve{\underline{\boldsymbol{s}}}$, defined as sequences of d signal points as $\underline{\boldsymbol{s}} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d]$ and $\breve{\underline{\boldsymbol{s}}} = [\breve{\boldsymbol{s}}_1, \ldots, \breve{\boldsymbol{s}}_d]$. Here, $\boldsymbol{s}_e$ and $\breve{\boldsymbol{s}}_e$, $1 \leq e \leq d$, belong to the multi-dimensional constellation $\boldsymbol{\Psi}$, i.e., $\boldsymbol{s}_e$ and $\breve{\boldsymbol{s}}_e$ can be one of K signals $\{\boldsymbol{s}(\kappa)\}_{\kappa=1}^{K}$ in $\boldsymbol{\Psi}$. These two signals can be expressed as follows:

$$\boldsymbol{s}_e = [y_{e,1}, \ldots, y_{e,N}] \tag{3.8}$$

$$\breve{\boldsymbol{s}}_e = [\breve{y}_{e,1}, \ldots, \breve{y}_{e,N}] \tag{3.9}$$

Also let

$$\underline{\boldsymbol{h}} = [h_{1,1}, h_{1,2}, \ldots h_{1,N}, \ldots, h_{d,1}, h_{d,2}, \ldots, h_{d,N}] \tag{3.10}$$

where $h_{e,p}$, $1 \leq e \leq d$, $1 \leq p \leq N$, represents the Rayleigh fading coefficient corresponding to the pth M-ary symbol $y_{e,p}$ in the eth signal $\boldsymbol{s}_e$. Then the pairwise error

probability (PEP) conditioned on $\underline{\boldsymbol{h}}$ can be computed as[4]

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}|\underline{\boldsymbol{h}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d}\sum_{p=1}^{N} h_{e,p}^2 \|y_{e,p} - \breve{y}_{e,p}\|^2}\right). \tag{3.11}$$

First, consider the case of fast fading channel. Let $f_A(d, \boldsymbol{\Psi}, \xi)$ denote the function $f(d, \Psi, \xi)$ in this case. By using the Gaussian probability integral $Q\left(\sqrt{2\gamma}\right) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{\gamma}{\sin^2\theta}\right) \mathrm{d}\theta$ and averaging (3.11) over the sequence $\underline{\boldsymbol{h}}$ of $N \cdot d$ independent and identically distributed (i.i.d.) Rayleigh random variables, it can be shown that

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}) = E_{\underline{\boldsymbol{h}}}\left\{P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}|\underline{\boldsymbol{h}})\right\} = \frac{1}{\pi}\int_0^{\pi/2}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}\theta \tag{3.12}$$

where

$$\Delta_e = \prod_{p=1}^{N}\left(1 + \frac{1}{4N_0}\frac{\|y_{e,p} - \breve{y}_{e,p}\|^2}{\sin^2\theta}\right)^{-1}. \tag{3.13}$$

Note that the term inside the brackets of (3.13) is obtained by invoking the equality $E_h\{\exp(-\gamma h^2)\} = 1/(1+\gamma)$, where h is a Rayleigh random variable.

Owing to the success of decoding steps, one needs to consider the two signals $\boldsymbol{s}_e = [y_{e,1}, \ldots, y_{e,N}]$ and $\breve{\boldsymbol{s}}_e = [\breve{y}_{e,1}, \ldots, \breve{y}_{e,N}]$ whose labels differ in only one bit. Also observe that Δ_e is an i.i.d. random variable. Hence, the function $f_A(d, \boldsymbol{\Psi}, \xi)$ can be obtained by averaging over all signal points $\boldsymbol{s}(\kappa)_{\kappa=1}^{K}$ of $\boldsymbol{\Psi}$ [2, 41]. Using the same notations as in [25], let $\boldsymbol{s}(\kappa) = [q_{\kappa,1}, \ldots, q_{\kappa,N}]$ and $\boldsymbol{s}(j(\kappa,k)) = [q_{j(\kappa,k),1}, \ldots, q_{j(\kappa,k),N}]$ be the two signal points whose labels differ at position k. Thus $j(\kappa, k)$ is understood as an integer index of a signal in $\boldsymbol{\Psi}$ and it depends on κ, k and the specific mapping ξ. Now taking the average over all the signals $\boldsymbol{s}(\kappa) = [q_{\kappa,1}, \ldots, q_{\kappa,N}]$ and the corresponding signal $\boldsymbol{s}(j(\kappa,k)) = [q_{j(\kappa,k),1}, \ldots, q_{j(\kappa,k),N}]$ in $\boldsymbol{\Psi}$ gives the following upper bound on $f_A(d, \boldsymbol{\Psi}, \xi)$:

$$f_A(d, \boldsymbol{\Psi}, \xi) \le \frac{1}{\pi}\int_0^{\pi/2} E\left\{\prod_{p=1}^{N}\left(1 + \frac{\|q_{\kappa,p} - q_{j(\kappa,k),p}\|^2}{4N_0\sin^2\theta}\right)^{-1}\right\}^{d}\mathrm{d}\theta \tag{3.14}$$

[4]The Q-function is defined as $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \mathrm{e}^{-t^2/2}\mathrm{d}t$.

where $E\{\cdot\}$ denotes the average over all the signals in $\mathbf{\Psi}$ and is computed as

$$\begin{aligned} & E\left\{\prod_{p=1}^{N}\left(1+\frac{\|q_{\kappa,p}-q_{j(\kappa,k),p}\|^2}{4N_0\sin^2\theta}\right)^{-1}\right\} \\ = & \frac{1}{M_c 2^{M_c}}\sum_{\boldsymbol{s}(\kappa)\in\mathbf{\Psi}}\sum_{k=1}^{M_c}\left[\prod_{p=1}^{N}\left(1+\frac{\|q_{\kappa,p}-q_{j(\kappa,k),p}\|^2}{4N_0\sin^2\theta}\right)^{-1}\right] \end{aligned} \tag{3.15}$$

and $M_c = Nm$ is the number of coded bits carried by one N-D signal point.

In the case of quasistatic Rayleigh fading channel, let $f_B(d, \mathbf{\Psi}, \xi)$ denote the function $f(d, \mathbf{\Psi}, \xi)$. Then using a similar analysis as above, one has the following upper bound on $f_B(d, \mathbf{\Psi}, \xi)$ [25]:

$$f_B(d,\mathbf{\Psi},\xi) \le \frac{1}{\pi}\int_0^{\pi/2} E\left\{\left(1+\frac{\|\boldsymbol{s}(\kappa)-\boldsymbol{s}(j(\kappa,k))\|^2}{4N_0\sin^2\theta}\right)^{-1}\right\}^d \mathrm{d}\theta \tag{3.16}$$

where

$$E\left\{\left(1+\frac{\|\boldsymbol{s}(\kappa)-\boldsymbol{s}(j(\kappa,k))\|^2}{4N_0\sin^2\theta}\right)^{-1}\right\} = \frac{1}{M_c 2^{M_c}}\sum_{\boldsymbol{s}(\kappa)\in\mathbf{\Psi}}\sum_{k=1}^{M_c}\left(1+\frac{\|\boldsymbol{s}(\kappa)-\boldsymbol{s}(j(\kappa,k))\|^2}{4N_0\sin^2\theta}\right)^{-1} \tag{3.17}$$

It is clear from (3.14) and (3.16) that $f_A(d, \mathbf{\Psi}, \xi)$ and $f_B(d, \mathbf{\Psi}, \xi)$ can be efficiently computed with a single integral. As shown later, by using (3.14) and (3.16) in (2.20), one obtains an accurate approximation for the asymptotic error performance of a BICM-ID system employing multi-dimensional constellation $\mathbf{\Psi}$ and mapping ξ.

3.3 The Distance Criteria and The Lower Bound for The Optimal Mapping

The disadvantage of (3.14) and (3.16) is that they do not give an insight on how to design good mappings for a given constellation $\mathbf{\Psi}$. Using the inequality $Q\left(\sqrt{2\gamma}\right) < \frac{1}{2}\exp(-\gamma)$, the two functions $f_A(d, \mathbf{\Psi}, \xi)$ and $f_B(d, \mathbf{\Psi}, \xi)$ can be approximated at high signal to noise ratio (SNR) as [25]:

$$f_A(d,\mathbf{\Psi},\xi) \sim \frac{1}{2}\left[\delta_A(\mathbf{\Psi},\xi)\right]^d \tag{3.18}$$

and

$$f_B(d,\mathbf{\Psi},\xi) \sim \frac{1}{2}\left[\delta_B(\mathbf{\Psi},\xi)\right]^d \tag{3.19}$$

where $\delta_A(\mathbf{\Psi}, \xi)$ and $\delta_B(\mathbf{\Psi}, \xi)$ are two distance parameters that characterize the influences of the multi-dimensional signal constellation and mapping to the error performance of BICM-ID over the two channel models. The distance criterion for the case of fast Rayleigh fading channel is written as [25]:

$$\delta_A(\mathbf{\Psi}, \xi) = \frac{1}{M_c 2^{M_c}} \sum_{\boldsymbol{s}(\kappa) \in \mathbf{\Psi}} \sum_{k=1}^{M_c} \prod_{p=1}^{N} \left(1 + \frac{\|q_{\kappa,p} - q_{j(\kappa,k),p}\|^2}{4N_0}\right)^{-1}. \tag{3.20}$$

Over the quasistatic Rayleigh fading channel, the distance criterion is given as follows [25]:

$$\delta_B(\mathbf{\Psi}, \xi) = \frac{1}{M_c 2^{M_c}} \sum_{\boldsymbol{s}(\kappa) \in \mathbf{\Psi}} \sum_{k=1}^{M_c} \left(1 + \frac{\|\boldsymbol{s}(\kappa) - \boldsymbol{s}(j(\kappa,k))\|^2}{4N_0}\right)^{-1}. \tag{3.21}$$

At high SNR, one can use the approximation $(1+x)^{-1} \approx x^{-1}$ when $x \to \infty$ to simplify $\delta_B(\mathbf{\Psi}, \xi)$ as

$$\delta_B(\mathbf{\Psi}, \xi) \approx (4N_0)\hat{\delta}_B(\mathbf{\Psi}, \xi), \tag{3.22}$$

where

$$\hat{\delta}_B(\mathbf{\Psi}, \xi) = \frac{1}{M_c 2^{M_c}} \sum_{\boldsymbol{s}(\kappa) \in \mathbf{\Psi}} \sum_{k=1}^{M_c} \frac{1}{\|\boldsymbol{s}(\kappa) - \boldsymbol{s}(j(\kappa,k))\|^2}. \tag{3.23}$$

Note that $\hat{\delta}_B(\mathbf{\Psi}, \xi)^{-1}$ is exact the Harmonic mean Euclidean distance with ideal feedback defined in [7]. Furthermore, observe that for the set of mappings whose components $||q_{\kappa,p} - q_{j(\kappa,k),p}\|^2 > 0$, one can also approximate $\delta_A(\mathbf{\Psi}, \xi)$ as:

$$\delta_A(\mathbf{\Psi}, \xi) \approx (4N_0)^N \hat{\delta}_A(\mathbf{\Psi}, \xi) \tag{3.24}$$

where

$$\hat{\delta}_A(\mathbf{\Psi}, \xi) = \frac{1}{M_c 2^{M_c}} \sum_{\boldsymbol{s}(\kappa) \in \mathbf{\Psi}} \sum_{k=1}^{M_c} \prod_{p=1}^{N} \frac{1}{\|q_{\kappa,p} - q_{j(\kappa,k),p}\|^2}. \tag{3.25}$$

Similar to $\hat{\delta}_B(\mathbf{\Psi}, \xi)$, the parameter $\hat{\delta}_A(\mathbf{\Psi}, \xi)$ above does not depend on N_0 and it is therefore more convenient to use it rather than $\delta_A(\mathbf{\Psi}, \xi)$ for comparing different mappings as long as $||q_{\kappa,p} - q_{j(\kappa,k),p}\|^2 > 0$. Also, when this happens, it is observed from (3.18) and (3.24) that a maximum diversity gain of Nd can be achieved over fast fading channels, as far as the asymptotic performance is concerned. On the other hand, over a quasistatic Rayleigh fading channel, the diversity gain is only d, as shown in (3.19) and (3.22).

It can be seen that the two parameters $\delta_A(\mathbf{\Psi}, \xi)$ and $\delta_B(\mathbf{\Psi}, \xi)$ should be made as small as possible to lower the asymptotic bit error rate performance of BICM-ID systems. Loosely speaking, this objective can be fulfilled by labelling the signals in the constellation such that a signal pair at a larger Euclidean distance corresponds to a smaller label Hamming distance.

As discussed earlier, an exhaustive search to find the mappings that yield the smallest values of these parameters is impossible for a high-dimensional constellation. Therefore, different searching algorithms (e.g., the BSA) have been widely used to address the above mapping problem. In contrast to computer searching, proposed here is an explicit and simple algorithm to construct a good multi-dimensional mapping. The limitation, however, is that the proposed algorithm is only applicable for 8-PSK constellation.

To assess the performance of the proposed mapping, lower bounds on the distance parameters $\delta_A(\mathbf{\Psi}, \xi)$ and $\delta_B(\mathbf{\Psi}, \xi)$ shall be first stated. It will also be demonstrated that there does not exist any mapping that achieves these lower bounds. For simplicity of notation, hereafter, a signal $\boldsymbol{s}(\kappa) \in \mathbf{\Psi}$ shall be simply denoted as $\boldsymbol{s}$.

Consider the 8-PSK constellation with four possible inter-signal Euclidean distances d_1, d_2, d_3 and d_4 as shown in Fig. 3.1. By studying the Euclidean distance

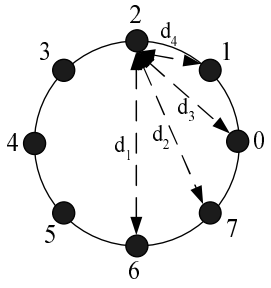


Figure 3.1 Distance spectrum of 8-PSK constellation.

profile of a multi-dimensional constellation constructed from 8-PSK, it is straightforward to see that the best mapping possible in terms of minimizing $\delta_A(\mathbf{\Psi}, \xi)$ and $\delta_B(\mathbf{\Psi}, \xi)$ should have the following property.

Property 3.1: For any signal point $\boldsymbol{s} \in \boldsymbol{\Psi}$, the $3N$ signal points whose labels differ in only 1 bit compared to the label of $\boldsymbol{s}$ have the following properties regarding the squared Euclidean distances: (i) There is only 1 signal point at distance Nd_1^2; (ii) There are $2N$ signal points at distance $(N-1)d_1^2 + d_2^2$; and (iii) There are $(N-1)$ signal points at distances $(N-2)d_1^2 + d_2^2 + d_2^2$. Therefore, one has the following lower bounds:

$$\begin{aligned} \delta_A(\boldsymbol{\Psi}, \xi) &\geq \frac{1}{3N}\left(1+\frac{d_1^2}{4N_0}\right)^{-(N-2)}\left[\left(1+\frac{d_1^2}{4N_0}\right)^{-2}\right. \\ &+ 2N\left(1+\frac{d_1^2}{4N_0}\right)^{-1}\left(1+\frac{d_2^2}{4N_0}\right)^{-1} \\ &+ \left.(N-1)\left(1+\frac{d_2^2}{4N_0}\right)^{-2}\right] \end{aligned} \tag{3.26}$$

and

$$\delta_B(\boldsymbol{\Psi}, \xi) \geq \frac{4N_0}{3N}\left[\frac{1}{Nd_1^2} + \frac{2N}{(N-1)d_1^2 + d_2^2} + \frac{(N-1)}{(N-2)d_1^2 + d_2^2 + d_2^2}\right]. \tag{3.27}$$

Unfortunately, as proved in Appendix A, there does not exist any mappings that satisfy *Property 3.1*. This implies that the lower bounds in (3.26) and (3.27) are not achievable. Alternatively, the next section proposes an algorithm to design the mapping with the following property.

Property 3.2: For any signal point $\boldsymbol{s}$, the $3N$ signal points whose labels differ in only 1 bit compared to that of $\boldsymbol{s}$ have the following properties: (i) There is only 1 signal point at distance Nd_1^2; (ii) There are N signal points at distance $(N-1)d_1^2 + d_2^2$; and (iii) There are $(2N-1)$ signal points at distance $(N-1)d_1^2 + d_3^2$.

Comparing the proposed mapping with the non-existing mapping, it is clear that the significant contribution to the asymptotic error performance is the squared distance d_1^2. It can also be seen that as N increases, the gap between the distance parameters corresponding to these two mappings decreases and becomes very small. Moreover, it is observed later that for $N = 1$, the proposed mapping is the same as the global optimal mapping.

3.4 The Proposed Mapping Algorithm

First, start with the 8-PSK constellation in Fig. 3.1, where each signal point is assigned an integer index. Divide the signal points into two subsets as follows:

- The odd subset, denoted by **O**, includes the signal points with odd indices $1, 3, 5$ and 7. That is $\mathbf{O} = \{1, 3, 5, 7\}$.
- The even subset, denoted by **E**, includes the remaining signal points with even indices $0, 2, 4$ and 6. Thus, $\mathbf{E} = \{0, 2, 4, 6\}$.

Clearly, each subset is simply a QPSK or a rotated (by $\pi/4$) QPSK constellation with two different inter-signal Euclidean distances d_1 and d_3.

For convenience, represent the signal point $\boldsymbol{s} \in \boldsymbol{\Psi}$ as a string of N integer indices as:

$$\boldsymbol{s} = [s_1, \ldots, s_N] \tag{3.28}$$

where s_n takes a value in **E** or **O**. The signal point $\boldsymbol{s}$ is labelled with $3N$ binary bits as follows:

$$\boldsymbol{s} \sim \boldsymbol{a} = (a_1, \ldots, a_{3N}). \tag{3.29}$$

For each signal point $\boldsymbol{s}$, associate with it the binary N-tuple $\boldsymbol{x} = (x_1, \ldots, x_N)$ computed as:

$$x_n = s_n \bmod 2. \tag{3.30}$$

Also denote the corresponding decimal number of $\boldsymbol{x}$ by X. Note that there are $K = 8^N$ distinct signal points $\boldsymbol{s}$, while there are only 2^N distinct N-tuple $\boldsymbol{x}$. According to (3.30), associated to a given N-tuple $\boldsymbol{x}$ are 4^N signal points $\boldsymbol{s}$. These 4^N signal points are the vertices of the twisted hypercube[5] H_X. In other words, the multi-dimensional constellation $\boldsymbol{\Psi}$ can be decomposed into 2^N twisted hypercubes, each containing 4^N signal points. As an example, any signal point $\boldsymbol{s}$ whose components $\{s_n\}$ are all in the even subset **E** will be found in the hypercube H_0 (since $\boldsymbol{x} = (0, \ldots, 0)$ and $X = 0$).

[5] H_X is called a twisted hypercube if each of its faces is a QPSK or a rotated QPSK.

The objective of the proposed algorithm is to obtain the label $\boldsymbol{a}$ for each signal point $\boldsymbol{s}$ that satisfies *Property 3.2* mentioned earlier. The proposed algorithm is given and discussed in the following three main steps.

Step 1: For each signal point $\boldsymbol{s}$ with the associated vector $\boldsymbol{x}$, the first N bits of its label vector $\boldsymbol{a}$ are simply assigned as:

$$a_n = x_n = s_n \bmod 2,\ 1 \leq n \leq N. \tag{3.31}$$

As a direct consequence of Step 1, all the signal points in the same twisted hypercube H_X have the same first N bits in their labels.

Step 2: This step considers the labelling strategy for the special hypercube H_0. More specifically, the remaining $2N$ binary labelling bits $(a_{N+1}, \ldots, a_{3N})$ for any $\boldsymbol{s}$ in H_0 are determined by applying the the optimal mapping algorithm of a hypercube proposed in [25]. As a result, the signal points in H_0 and their labels possess the following properties [25]:

Distance Properties of H_0: For any signal point $\boldsymbol{s} \in H_0$, the $2N$ signal points whose labels differ in only 1 bit at positions from $(N+1)$ to $3N$ compared to the label of $\boldsymbol{s}$ are also in H_0. Furthermore, among these $2N$ signal points, there are 1 signal point at squared distance Nd_1^2 and $(2N-1)$ signal points at squared distance $(N-1)d_1^2 + d_3^2$.

After Step 2, 4^N signal points in H_0 have been labelled.

Step 3: This final step constructs the labels of all the remaining signal points in $\boldsymbol{\Psi}$ based on the labels of H_0. In particular, for each decimal number X, $1 \leq X \leq 2^N - 1$, with the corresponding N-tuple $\boldsymbol{x} = (x_1, \ldots, x_N)$, first associate each signal point $\boldsymbol{s} = [s_1, \ldots, s_N] \in H_0$ with the signal point $\boldsymbol{z} = [z_1, \ldots, z_N] \in H_X$ as follows:

$$z_n = [s_n + 4(x_1 + \ldots + x_N) - x_n] \bmod 8,\ 1 \leq n \leq N. \tag{3.32}$$

Then label the signal point $\boldsymbol{z} \in H_X$ by:

$$\boldsymbol{z} \sim (x_1, \ldots, x_N, a_{N+1}, \ldots, a_{3N})\,. \tag{3.33}$$

With the above mapping construction, one has the following theorems.

Theorem 3.1: For any $1 \leq X \leq 2^N - 1$, the mapping defined in (3.32) and (3.33) preserves the distance properties of H_0.

Proof: Since $\boldsymbol{s} \in H_0$, all components s_n are even and one has:

$$(z_n - x_n) \bmod 2 = 0. \tag{3.34}$$

This ensures that the signal point $\boldsymbol{z}$ is in the hypercube H_X. Clearly, with a given $\boldsymbol{s}$, the operation in (3.32) with $(2^N - 1)$ N-tuple $\boldsymbol{x}$ ($\boldsymbol{x} \neq 0$) gives $(2^N - 1)$ signal points $\boldsymbol{z}$ in the hypercube H_X, $1 \leq X \leq 2^N - 1$. Furthermore, there is a one-to-one correspondence between $\boldsymbol{s}$ and $\boldsymbol{z}$ (i.e., given $\boldsymbol{x}$, there is only one pair of $\boldsymbol{s}$ and $\boldsymbol{z}$ that satisfy (3.32)). Consider two signal points $\boldsymbol{s}$ and $\boldsymbol{t}$ in H_0 and the corresponding signal points $\boldsymbol{z}$ and $\boldsymbol{y}$ in H_X. It follows from (3.32) that $||s_n - t_n||^2 = ||z_n - y_n||^2$, $1 \leq n \leq N$, and hence

$$||\boldsymbol{s} - \boldsymbol{t}||^2 = ||\boldsymbol{z} - \boldsymbol{y}||^2. \tag{3.35}$$

From the labelling rule in (3.33), it is clear that the labels of the two signal points $\boldsymbol{s}$ and $\boldsymbol{z}$ have the same last $2N$ bits. It then can be concluded that the *distance properties of* H_0 stated before are preserved for any hypercube H_X. This also implies that such properties are preserved for any signal points in $\boldsymbol{\Psi}$.

Theorem 3.2: By the association in (3.32) and the labelling rule in (3.33), the squared Euclidean distance between any two signal points in $\boldsymbol{\Psi}$ whose labels differ in only 1 bit at the position n, $1 \leq n \leq N$, is $(N-1)d_1^2 + d_2^2$.

Proof: Consider any two signal points $\boldsymbol{z}$ and $\boldsymbol{w}$ whose labels differ in only 1 bit at position n. Without loss of generality, assume that:

$$\boldsymbol{z} = [z_1, \ldots, z_N] \sim (x_1, \ldots, x_{n-1}, 1, x_{n+1}, \ldots, x_N, \ldots, x_{3N}) \tag{3.36}$$

and

$$\boldsymbol{w} = [w_1, \ldots, w_N] \sim (x_1, \ldots, x_{n-1}, 0, x_{n+1}, \ldots, x_N, \ldots, x_{3N}). \tag{3.37}$$

Furthermore, consider the following signal point $\boldsymbol{s} \in H_0$:

$$\boldsymbol{s} = [s_1, \ldots, s_N] \sim (\underbrace{0, \ldots, 0, 0, 0, \ldots, 0}_{N\,0's}, x_{N+1}, \ldots, x_{3N}). \tag{3.38}$$

It then follows from (3.32) that:

$$z_p = [s_p + 4(x_1 + \ldots + x_{n-1} + 1 + x_{n+1} + \ldots + x_N) - x_p] \bmod 8 \tag{3.39}$$

and

$$w_p = [s_p + 4(x_1 + \ldots + x_{n-1} + 0 + x_{n+1} + \ldots + x_N) - x_p] \bmod 8. \tag{3.40}$$

Therefore the absolute difference between the two integers z_p and w_p is

$$|z_p - w_p| \;=\; \begin{cases} 4 \bmod 8 = 4, & p \neq n \\ 3 \bmod 8 = 3, & p = n \end{cases} \tag{3.41}$$

Hence the squared Euclidean distance between two 8-PSK signal points corresponding to the integer indices z_p and w_p can be determined as

$$||z_p - w_p||^2 \;=\; \begin{cases} d_1^2, & p \neq n \\ d_2^2, & p = n \end{cases} \tag{3.42}$$

It then can be concluded that:

$$||\boldsymbol{z} - \boldsymbol{w}||^2 = (N-1)d_1^2 + d_2^2. \tag{3.43}$$

From Theorems 3.1 and 3.2, it is simple to see that the proposed mapping satisfies *Property 3.2*.

As aforementioned, the common technique to find good multi-dimensional mappings for BICM-ID relies on the BSA. It is therefore of interest to evaluate the proposed mapping against the BSA. Appendix B shows that the proposed mapping is immune (i.e., unaffected) by the BSA. This means that if the BSA starts with the proposed mapping, it can never find a better one. This result also implies that the proposed mapping is at least one of the locally optimum mappings produced by the BSA.

Example 3.1: As an example, consider the proposed mapping for the simplest case of $N = 1$, i.e., the case of 8-PSK constellation in a complex 1-D signal space. There are only two hypercubes H_0 and H_1.

Step 1: Label the first bit for all signal points as follows: 0 for signal points in H_0 and 1 for signal points in H_1. The result of this step is shown in Fig. 3.2-(a).

Step 2: For signal points H_0, label the last two bits with the optimal mapping algorithm in [25], which is simply the anti-Gray mapping of QPSK constellation. The results after this step are shown in Fig. 3.2-(b).

Step 3: Using the association in (3.32) and the labelling rule in (3.33), label the last two binary bits of the signal points in H_1. For example, since $1 = (6 + 3) \bmod 8$, the last binary two bits of the signal point with index 1 are the same with that of the signal point with index 6 and they are $(1, 1)$. The complete mapping is shown in Fig. 3.2-(c). This mapping is exactly one version of the semi-set partitioning (SSP) mapping shown in Fig. 2.10, which is the best mapping of 8-PSK constellation found by a brute-force search [7, 34].

Example 3.2: For the case of $N = 2$, the proposed mapping is provided in Table 3.1. Observe that any signal points in the same hypercube H_X have the same first 2 labelling bits, while the labelling of the last 4 bits follow from the optimal mapping rule in [25]. Table 3.2 compares the parameters $\hat{\delta}_A(\mathbf{\Psi}, \xi)$ and $\hat{\delta}_B(\mathbf{\Psi}, \xi)$ of the proposed mapping against the unachievable lower bound when $N = 2, 4$ and 8. Note that the parameter $\hat{\delta}_A(\mathbf{\Psi}, \xi)$ can be used since all components $||q_{\kappa,p} - q_{j(\kappa,k),p}||^2$ of the proposed and non-exiting mappings are strictly positive. Observe from Table 3.2 that the differences in the two distance parameters of the two mappings are very small, especially when N increases. This suggests that there is almost no difference between the two systems in terms of the asymptotic performance.

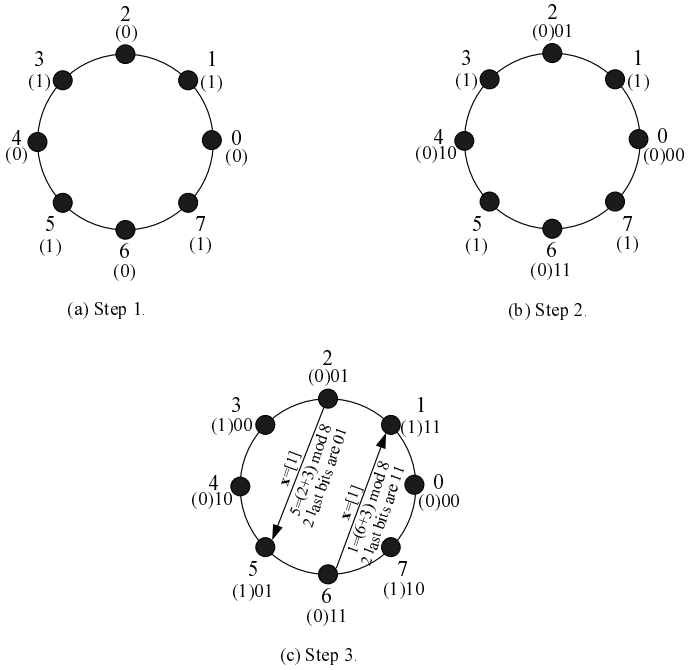


Figure 3.2 Three steps to construct proposed mapping for 8-PSK constellation.

3.5 Illustrative Results

This section provides the analytical and simulation results to confirm the advantage of the proposed multi-dimensional mapping of 8-PSK for BICM-ID. The analysis based on EXIT chart discussed in Section 2.3.3, Chapter 2 is also briefly presented.

3.5.1 The Error Performances

In all simulation results, unless otherwise specified, a simple convolutional code of rate-2/3, 4-state and generator sequences $\mathbf{g}_1 = (6, 2, 6)$ and $\mathbf{g}_2 = (2, 4, 4)$ [42] is used. This code together with an 8-PSK constellation yields a spectral efficiency of 2 bits/s/Hz. The bit-wise interleaver with a length of 12,000 coded bits is designed according to the rules outlined in [7]. Each point in the BER curves is simulated with 10^7 to 3×10^8 coded bits. The error floors calculated according to (2.20), (3.14) and (3.16) with the first 20 Hamming distances are also drawn to show how the iterations converge to the error floor. The lower bound corresponding to the non-

Table 3.1 The proposed complex 2-D mapping. It includes 4 twisted hypercube H_0, H_1, H_2 and H_3. In each hypercube, the signal point $\boldsymbol{s}$ is represented by a string of two integers. The corresponding label $\boldsymbol{a}$ is a binary 6-tuple.

H_0		H_1		H_2		H_3	
$\boldsymbol{s}$	$\boldsymbol{a}$	$\boldsymbol{s}$	$\boldsymbol{a}$	$\boldsymbol{s}$	$\boldsymbol{a}$	$\boldsymbol{s}$	$\boldsymbol{a}$
$[0,0]$	(000000)	$[0,1]$	(010010)	$[1,0]$	(101111)	$[1,1]$	(111101)
$[0,2]$	(001010)	$[0,3]$	(011011)	$[1,2]$	(100101)	$[1,3]$	(110100)
$[0,4]$	(000011)	$[0,5]$	(010001)	$[1,4]$	(101100)	$[1,5]$	(111110)
$[0,6]$	(001001)	$[0,7]$	(011000)	$[1,6]$	(100110)	$[1,7]$	(110111)
$[2,0]$	(000111)	$[2,1]$	(010101)	$[3,0]$	(100011)	$[3,1]$	(110001)
$[2,2]$	(001101)	$[2,3]$	(011100)	$[3,2]$	(101001)	$[3,3]$	(111000)
$[2,4]$	(000100)	$[2,5]$	(010110)	$[3,4]$	(100000)	$[3,5]$	(110010)
$[2,6]$	(001110)	$[2,7]$	(011111)	$[3,6]$	(101010)	$[3,7]$	(111011)
$[4,0]$	(001011)	$[4,1]$	(011001)	$[5,0]$	(100100)	$[5,1]$	(110110)
$[4,2]$	(000001)	$[4,3]$	(010000)	$[5,2]$	(101110)	$[5,3]$	(111111)
$[4,4]$	(001000)	$[4,5]$	(011010)	$[5,4]$	(100111)	$[5,5]$	(110101)
$[4,6]$	(000010)	$[4,7]$	(010011)	$[5,6]$	(101101)	$[5,7]$	(111100)
$[6,0]$	(001100)	$[6,1]$	(011110)	$[7,0]$	(101000)	$[7,1]$	(111010)
$[6,2]$	(000110)	$[6,3]$	(010111)	$[7,2]$	(100010)	$[7,3]$	(110011)
$[6,4]$	(001111)	$[6,5]$	(011101)	$[7,4]$	(101011)	$[7,5]$	(111001)
$[6,6]$	(000101)	$[6,7]$	(010100)	$[7,6]$	(100001)	$[7,7]$	(110000)

existing mapping mentioned in previous sections[6] is also plotted from (2.20), (3.14) and (3.16) to show how close it is in terms of the BER performance between the proposed and the non-existing mappings. For comparison, the performance of the conventional system with SSP 8-PSK mapping is also provided.

First, for the quasistatic Rayleigh fading channel, Fig. 3.3 presents the perfor-

[6]Recall that this non-existing mapping satisfies *Property 3.1*.

Table 3.2 The distance parameters $\hat{\delta}_A(\mathbf{\Psi}, \xi)$ and $\hat{\delta}_B(\mathbf{\Psi}, \xi)$ for the non-existing and the proposed mappings. The energy of each 8-PSK symbol is normalized to 3.

	$\hat{\delta}_A(\Psi, \xi)$			$\hat{\delta}_B(\Psi, \xi)$		
n	2	4	8	2	4	8
The non-existing mapping	0.0082	$5.8e-5$	$2.8e-9$	0.0451	0.0218	0.0107
The proposed mapping	0.0108	$7.9e-5$	$3.9e-9$	0.0497	0.0228	0.0109

mance of a BICM-ID system employing the proposed complex 2-D mapping after $1, 4, 8$ and 12 iterations. The performance after 12 iterations of the conventional system using SSP 8-PSK is also shown. It can be seen from Fig. 3.3 that, the simulation

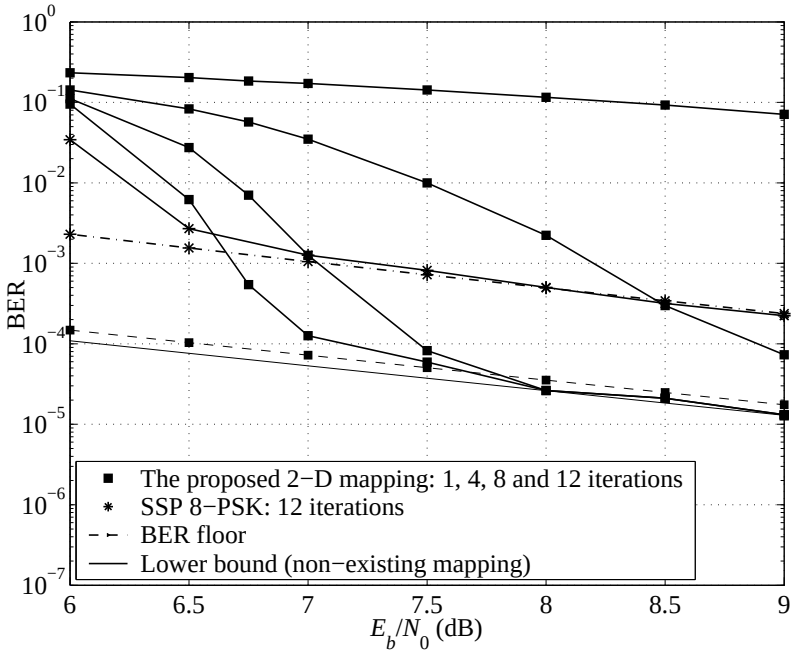


Figure 3.3 BER performance of BICM-ID systems over quasistatic fading channel: 4-state convolutional code and different mappings.

results converge to the error floor bounds for both systems. It is also obvious from Fig. 3.3 that a significant coding gain is achieved by the proposed mapping over the SSP 8-PSK mapping. Moreover, it can be clearly observed that there is almost no difference between the asymptotic performance of the proposed mapping and that of

the non-existing mapping.

Since the asymptotic performances and the real BER are very tight at medium and high SNR, the asymptotic performances can be used to accurately predict the BER. To see what is the real coding gain provided by the 2-D mapping over the conventional SSP 8-PSK mapping, Fig. 3.4 plots the error floor bounds for the systems employing the two mappings over the SNR range from 6 to 18 dB. Observe that

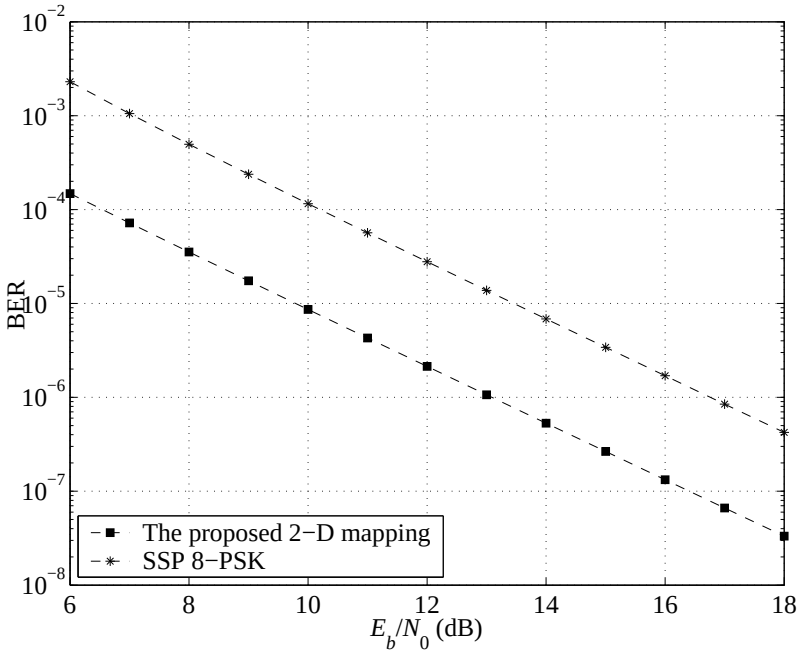


Figure 3.4 Error floor bounds over the quasistatic fading channel.

with the selected 4-state convolutional code, the coding gain offered by the proposed mapping at the BER level of 10^{-6} is about 3.8 dB. Such a coding gain is achieved with a reasonable increase in the receiver complexity which is required to address the soft-output demodulation of a higher-order multi-dimensional constellation.

For the case of fast Rayleigh fading channel, Fig. 3.5 shows the performance of the BICM-ID system using the proposed complex 2-D mapping with $1, 4, 8$ and 12 iterations and the conventional system with SSP 8-PSK mapping after 12 iterations. Comparing Figs. 3.3 and 3.5, it is observed that there is no difference in BER performance between the two models of Rayleigh fading channels if the SSP mapping is used. It can be seen from Fig. 3.5 that the actual BER performance of two systems

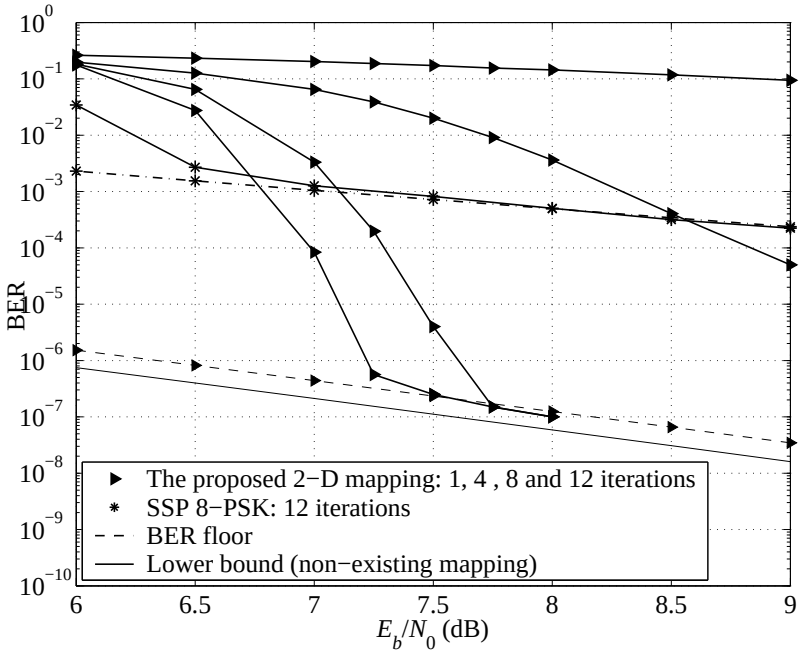


Figure 3.5 BER performance of BICM-ID systems over fast fading channel: 4-state convolutional code and different mappings.

also converges to the BER floors. Though the convergence happens at a quite low BER for the proposed mapping with $12,000$ bit interleaver, the analytical asymptotic performance is still very useful to predict the actual BER. A faster convergence can be achieved by using a longer interleaving length. Similar to the previous channel model, it can be seen that the gap in terms of the asymptotic BER perforamnce between the proposed mapping and the non-existing mapping is very slight.

The error floor bounds for the two systems are also plotted in Fig. 3.6 for comparison. At the BER level of 10^{-6}, the coding gain is about 10.5dB in terms of the asymptotic performance. However, as can be observed from Fig. 3.5, the proposed 2-D mapping with $12,000$-bit interleaver length and 12 iterations achieves BER level of 10^{-6} at around 7.2dB. Hence, the real coding gain is about 9.6dB. Obviously, this is still a very significant gain. When compared to the case of quasistatic fading channel, the coding gain is different in about 5.8 dB. This observation is similar to the performance of a BICM-ID system with signal space diversity studied in [8].

Finally, Fig. 3.7 compares the asymptotic performances of the systems that use

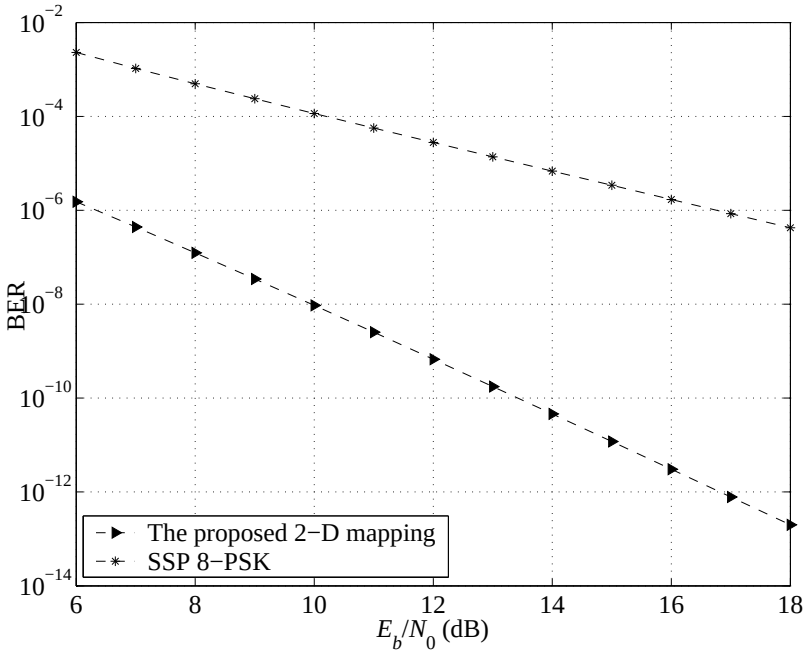


Figure 3.6 Error floor bounds over the fast fading channel.

the proposed 2-D and the 2-D mapping obtained by the BSA in [43] over both channel models. It can be observed that the proposed mapping is superior than the mapping

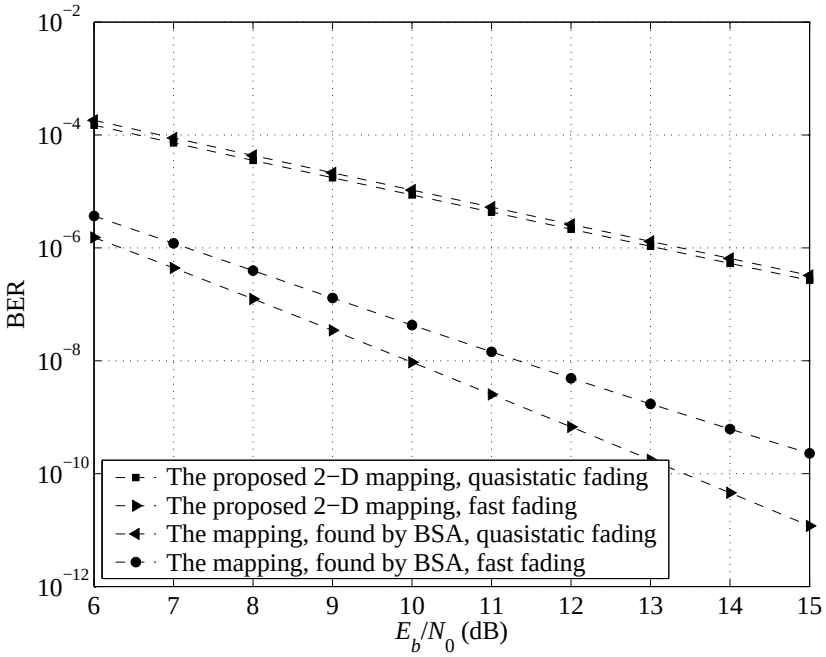


Figure 3.7 The asymptotic performances for systems using the proposed mapping and the mapping obtained by the BSA.

found by the BSA, especially for the case of fast fading channel. Such a superior mapping is easily obtained by the proposed algorithm instead of the time-consuming computer search based on the BSA.

3.5.2 EXIT Chart Analysis

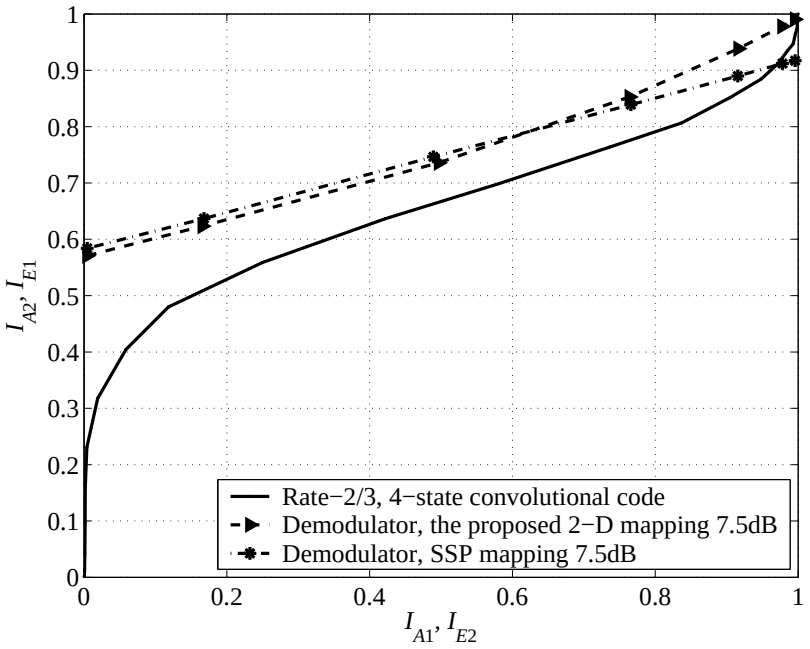


Figure 3.8 EXIT charts of the BICM-ID system employing the proposed 2-D mapping and the conventional SSP mapping at $E_b/N_0 = 7.5$dB over a fast fading channel.

The advantage of the proposed 2-D mapping compared to the SSP mapping can also be analyzed by using EXIT charts. Figure 3.8 shows the EXIT charts for BICM-ID systems over a fast fading channel using the proposed 2-D mapping and the conventional SSP 8-PSK mapping at $E_b/N_0 = 7.5$dB. As before, the rate-2/3, 4-state convolutional code with generator sequences $\mathbf{g}_1 = (6, 2, 6)$ and $\mathbf{g}_2 = (2, 4, 4)$ [42] is used. It can be seen from Fig. 3.8 that at 7.5dB, there is an open tunnel between the demodulator and the decoder transfer characteristics for both mappings, which leads the convergence of iterative demodulation and decoding towards high mutual information. Observe that when I_{A1} approaches 1, i.e., one has the perfect *a priori* information of the coded bits fed-back to the demodulator, the mutual information

provided by the proposed 2-D mapping is much higher over that of the SSP mapping. Equivalently, lower error performance can be achieved by using the proposed mapping. This clearly agrees with the results in Fig. 3.5.

4. Performance of BICM-ID with SSD over a Rayleigh Fading Channel[1]

In Chapter 3, it was demonstrated that by simply implementing the mapping ξ over multiple symbol intervals, a significant performance gain can be achieved over a conventional BICM-ID system where only a real two-dimensional mapping rule is considered. The use of multi-dimensional mapping in Chapter 3 is actually a processing technique at bit level to provide further protection to the transmitted information. Although the constellation $\mathbf{\Psi}$ studied in Chapter 3 spans over time in a multi-dimensional signal space, component symbols transmitted over the channel are still the same with that in a conventional BICM-ID system, e.g., symbols in the 8-PSK constellation. This chapter deals with a different approach using signal space diversity (SSD) technique, which is somewhat more general. In particular, considered in this chapter is a processing method at symbol level in which a signal symbol is spread over N different time intervals before being transmitted over the channel. The technique of multi-dimensional mapping can be also applied straightforwardly. However, for simplicity, this chapter only considers the conventional mapping, i.e., the mapping that is implemented independently and identically for each complex component, as similar to conventional BICM-ID. As presented in detail later, this chapter shall partially address a fundamental question for a conventional BICM-ID system over a fully interleaved *frequency non-selective* Rayleigh fading channel: what is its performance limitation?

As mentioned earlier, signal space diversity is an attractive method to exploit

[1] A part of this chapter was presented in [44, 45]

time diversity without any power or bandwidth expansion. It was shown in [17] that the error performance of an uncoded system (without error control coding) using SSD becomes insensitive to fading for a lattice constellation with a sufficiently large dimension. This implies that a fading channel can be converted to an AWGN channel to achieve the same error probability. Furthermore, the problem of constructing the rotation matrix in order to optimize the error performance of an uncoded system was also addressed in [17]. Some best known constellations for uncoded systems over fading channels obtained from algebraic number theory are provided by Viterbo in [46].

Applications of SSD have been considered in various uncoded and coded systems. For example, lattice-based space-time block codes were studied in [47] to exploit the advantages of signal space diversity for further performance improvement. In [48, 49], bit-interleaved coded modulation (BICM) combined with coordinate interleaving and rotated constellations to double the diversity order was demonstrated. Further work on conventional BICM-ID in [8] examined SSD with symbols in the real two-dimensional signal space. With coordinate interleaving, the in-phase and quadrature components of the same transmitted signal in a real two-dimensional signal space experience independent fades. Therefore, the diversity order can be increased by choosing the proper constellations and mappings.

SSD offers performance improvement at the cost of high computational complexity of the receiver, especially for constellations with large dimensions. Recently, several techniques have been proposed to reduce the receiver complexity, including sphere decoding [50] and soft interference cancelation [51]. In particular, simulation results on performance of BICM-ID using a complex 32-dimensional lattice with QPSK constellation and Gray mapping under Rayleigh fading showed that the iterative scalar Gaussian approximation (GA) offers error probability performance close to that in an AWGN channel with relatively low complexity [51].

This chapter presents a performance analysis of BICM-ID with complex N-dimensional (N-D) signal space diversity in a Rayleigh fading channel to investigate its perfor-

mance limitation, the choice of the rotation matrix and the design of a low-complexity receiver. At first, a tight error bound on the bit error probability (BEP) is derived. The expression of this error bound is useful in order to understand the individual contributions of the signal constellation, signal mapping and rotation matrix $\boldsymbol{G}$ to the overall performance of the system. Optimal rotation matrices $\boldsymbol{G}$ for a given constellation and mapping are then developed. It is shown that, for the class of optimal rotation matrices $\boldsymbol{G}$, the system performance over a fading channel approaches that in an AWGN channel with a reasonably large dimension N of the lattice constellations. With a proper mapping for an M-QAM constellation, a simple sub-optimal but yet effective iterative receiver structure suitable for large dimension N based on the minimum mean square error (MMSE) estimation [52] is developed. The simple sub-optimal receiver helps to obtain simulation results to compare with the analytical bounds for constellations with very large dimensions.

4.1 System Model

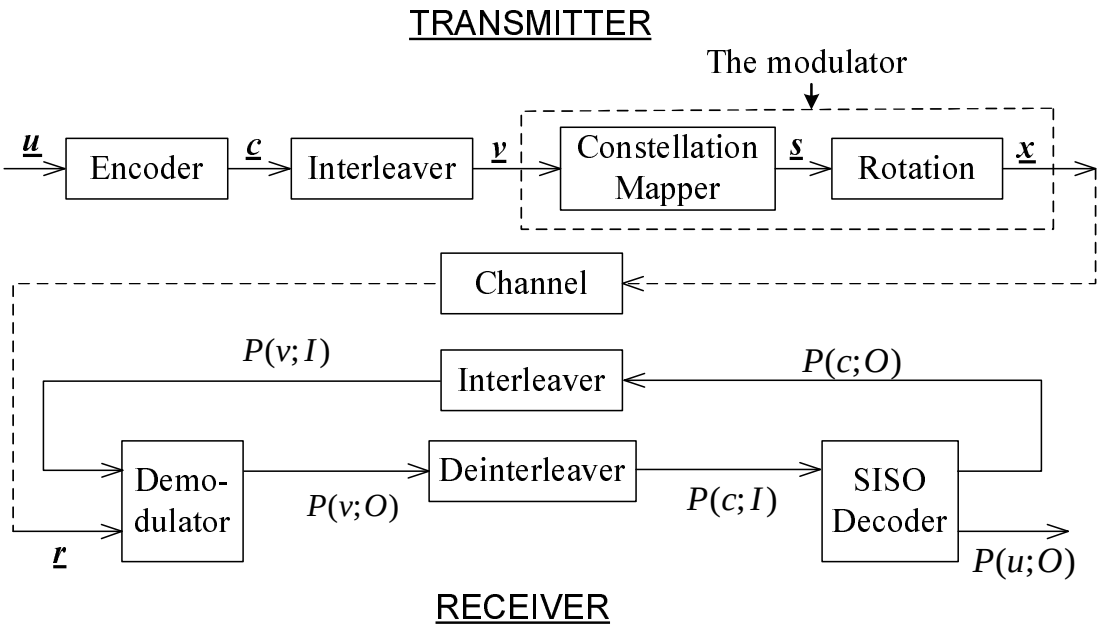


Figure 4.1 Block diagram of a BICM-ID system with signal space diversity.

The block diagram of a BICM-ID system with signal space diversity (BICM-ID-SSD) is shown in Fig. 4.1. For this system, a sequence of $M_c = Nm$ coded bits at the

output of the interleaver is mapped to one complex N-D constellation symbol at the modulator to produce the symbol sequence $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]$, where $s_i \in \Omega$, $\forall i$, with Ω a real two-dimensional constellation of size $M = 2^m$, such as QAM constellations. Similar to the previous chapter, it can be seen that a constellation $\boldsymbol{\Psi}$ in a N-D signal space is constructed, with in total of $K = M^N$ signal points $\boldsymbol{s}$. As mentioned earlier, this chapter only considers the conventional mapping ξ, i.e., the mapping that is implemented independently and identically for each complex component s_i, which is similar to conventional BICM-ID. The technique of SSD is applied by rotating a signal $\boldsymbol{s}$ in N-D constellation $\boldsymbol{\Psi}$ with an $N \times N$ complex rotation matrix $\boldsymbol{G}$. The rotated symbol $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$ corresponding to a new rotated constellation $\boldsymbol{\Psi}_{\mathrm{r}}$ is given by:

$$\boldsymbol{x}^\top = \boldsymbol{G}\boldsymbol{s}^\top, \tag{4.1}$$

where the superscript $\top$ denotes matrix transpose operation. Given $E\{\boldsymbol{s}^\dagger \boldsymbol{s}\} = \boldsymbol{I}_N$, i.e., the covariance of $\boldsymbol{s}$ is an identity matrix, where $\dagger$ denotes conjugate transpose, the entries $g_{i,u}$, $1 \leq i, u \leq N$, of $\boldsymbol{G}$ satisfy the following condition to ensure the total power constraint $E\{||\boldsymbol{x}||^2\} = E\{||\boldsymbol{s}||^2\}$:

$$\sum_{i=1}^{N} \sum_{u=1}^{N} ||g_{i,u}||^2 = N. \tag{4.2}$$

Let $\boldsymbol{r} = [r_1, r_2, \ldots, r_N]$ denote the N-D received signal. Assume that the channel is *frequency non-selective* Rayleigh fading and it changes independently in each component duration of the N-dim signal symbol, i.e., it is the fast fading case considered in Chapter 3. Furthermore, assume coherent detection at the receiver. It then suffices to represent the received signal as follows:

$$\boldsymbol{r}^\top = \boldsymbol{H}\boldsymbol{G}\boldsymbol{s}^\top + \boldsymbol{w}^\top. \tag{4.3}$$

Here, the entries of $\boldsymbol{w} = [w_1, \ldots, w_N]$ are i.i.d. circularly symmetric Gaussian random variables with variance N_0, i.e., $\mathcal{CN}(0, N_0)$. The matrix $\boldsymbol{H} = \mathrm{diag}(h_1, \ldots, h_N)$ contains the fading coefficients in its diagonal, where each fading coefficient h_i can be simply modeled as independent scalar Rayleigh random variables with unit mean squared value.

The receiver in BICM-ID with SSD systems includes the SISO demodulator and SISO decoder as described in Chapter 2. For the soft-input soft-output demodulator, the optimal one, i.e., the MAP demodulator, is similar to the one in multi-dimensional mapping BICM-ID systems in Chapter 3. The conditional density function $p(\boldsymbol{r}|\boldsymbol{s})$ given the channel gains $\{h_i\}$ and signal $\boldsymbol{s}$ was transmitted is calculated as:

$$p(\boldsymbol{r}|\boldsymbol{s}) = \frac{1}{(\pi N_0)^N} \exp\left(-\frac{\sum_{i=1}^{N} ||r_i - h_i x_i||^2}{N_0}\right). \tag{4.4}$$

However, since the complexity of the MAP demodulator grows exponentially with the number of coded bits per signal point Nm, the implementation of the MAP demodulator becomes intractable for medium to large values of Nm, which is of interest for SSD method. For such cases, the sub-optimal low-complexity methods using Gaussian approximations proposed in [51] can be attractive alternatives. Moreover, with a careful mapping design proposed for M-QAM constellations, Section 4.4 presents another sub-optimal soft-output demodulator based on the minimum mean square error (MMSE) estimation principle.

4.2 Performance Evaluation

Similar to any BICM-ID system, the key point to derive the union bound on the BEP for a BICM-ID employing signal space diversity with the assumption of error-free feedback is the evaluation of the average pairwise error probability $f(d, \boldsymbol{\Psi}, \xi)$, which depends on the Hamming distance d, the constellation $\boldsymbol{\Psi}$, the mapping rule ξ. Since BICM-ID with SSD takes into account the rotation $\boldsymbol{G}$, it can be seen that $f(d, \boldsymbol{\Psi}, \xi)$ also depends on $\boldsymbol{G}$. Therefore, the notation $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ is used in this chapter to denote the function $f(d, \boldsymbol{\Psi}, \xi)$.

Let $\underline{\boldsymbol{s}}$ and $\underline{\breve{\boldsymbol{s}}}$ be two sequences of d N-D symbols as $\underline{\boldsymbol{s}} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d]$ and $\underline{\breve{\boldsymbol{s}}} = [\breve{\boldsymbol{s}}_1, \ldots, \breve{\boldsymbol{s}}_d]$ representing the binary input and estimate sequences $\underline{\boldsymbol{c}}$ and $\underline{\breve{\boldsymbol{c}}}$, respectively, with the Hamming distance d between them. Also let $\underline{\boldsymbol{H}} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_d]$, where $\boldsymbol{H}_e = \text{diag}(h_{e,1}, \ldots, h_{e,N})$, $1 \leq e \leq d$, represents the channel gains that affect the transmitted symbol $\boldsymbol{s}_e$. The two complex N-D signal symbols $\boldsymbol{s}_e$ and $\breve{\boldsymbol{s}}_e$ correspond

to the two rotated symbols $\boldsymbol{x}_e$ and $\breve{\boldsymbol{x}}_e$, i.e., $\boldsymbol{x}_e^\top = \boldsymbol{G}\boldsymbol{s}_e^\top$ and $\breve{\boldsymbol{x}}_e^\top = \boldsymbol{G}\breve{\boldsymbol{s}}_e^\top$. Then the PEP conditioned on $\underline{\boldsymbol{H}}$ can be computed as follows:

$$P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}}|\underline{\boldsymbol{H}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d} d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e)}\right), \tag{4.5}$$

where $d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e)$ is the squared Euclidean distance between the two received signals corresponding to $\boldsymbol{x}_e$ and $\breve{\boldsymbol{x}}_e$ conditioned on $\boldsymbol{H}_e$ and in the absence of the AWGN, given by

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \left|\left|\boldsymbol{H}_e\boldsymbol{G}(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right|\right|^2 = \sum_{i=1}^{N} h_{e,i}^2 \left|\left|\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right|\right|^2, \tag{4.6}$$

where $\boldsymbol{g}_i$ is the ith row of $\boldsymbol{G}$. Using the Gaussian probability integral and averaging (4.5) over $\underline{\boldsymbol{H}}$ gives

$$P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}}) = \frac{1}{\pi}\int_0^{\pi/2}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}\theta, \tag{4.7}$$

where

$$\Delta_e = \prod_{i=1}^{N}\left(1 + \frac{||\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top||^2}{4N_0\sin^2\theta}\right)^{-1}. \tag{4.8}$$

Similar to the analysis in previous chapters, by assuming that the iterative processing works perfectly, the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ for BICM-ID with SSD can be written as follows:

$$f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G}) \le \frac{1}{\pi}\int_0^{\pi/2}\left[\underbrace{\mathcal{E}_{\boldsymbol{s},\boldsymbol{p}}\left\{\prod_{i=1}^{N}\left(1 + \frac{||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2}{4N_0\sin^2\theta}\right)^{-1}\right\}}_{\gamma(\boldsymbol{\Psi},\xi,\boldsymbol{G})}\right]^d \mathrm{d}\theta, \tag{4.9}$$

where the expectation is over all pairs of signal symbols $\boldsymbol{s}$ and $\boldsymbol{p}$ of $\boldsymbol{\Psi}$ whose labels differ in only 1 bit. The result of the expectation operation in (4.9) can be shown to be

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{M_c 2^{M_c}}\sum_{\boldsymbol{s}\in\boldsymbol{\Psi}}\sum_{k=1}^{M_c}\left[\prod_{i=1}^{N}\left(1 + \frac{||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2}{4N_0\sin^2\theta}\right)^{-1}\right] \tag{4.10}$$

where $\boldsymbol{p}$ is the symbol in $\boldsymbol{\Psi}$ whose label differs in only one bit at the position k compared to the label of $\boldsymbol{s}$ and $M_c = Nm$.

For a large value of $M_c = Nm$, the computation of (4.10) becomes intractable due to the huge number of signal symbols in $\boldsymbol{\Psi}$. For simplicity, consider the mapping ξ that is implemented independently and identically for each signal component in the real two-dimensional constellation Ω. First, by interchanging the summations in (4.10) one can write $\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G})$ as a sum of N terms as:

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{N} \sum_{u=1}^{N} \gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G}), \tag{4.11}$$

where

$$\gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m2^{Nm}} \sum_{k=(u-1)m+1}^{um} \sum_{\boldsymbol{s} \in \boldsymbol{\Psi}} \left[\prod_{i=1}^{N} \left(1 + \frac{||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2}{4N_0 \sin^2 \theta} \right)^{-1} \right] \tag{4.12}$$

is essentially obtained by averaging over all pairs of $\boldsymbol{s}$ and $\boldsymbol{p}$ whose labels differ in 1 bit at positions from $k = (u-1)m+1$ to $k = um$. Note that the inner sum in (4.12) is taken over $m2^{Nm}$ possible symbols $\boldsymbol{s}$, which is still computationally impossible when N and m are large. Fortunately, by independently mapping for each component, it is easy to verify that there is only one distinct component between $\boldsymbol{s}$ and $\boldsymbol{p}$ at the uth position. Hence, for a given u, one has:

$$||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2 = ||g_{i,u}(s_u - p_u)||^2, \tag{4.13}$$

where s_u and p_u are the uth components of $\boldsymbol{s}$ and $\boldsymbol{p}$, respectively. It then follows that:

$$\gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m2^{Nm}} \sum_{k=(u-1)m+1}^{um} \sum_{\boldsymbol{s} \in \boldsymbol{\Psi}} \left[\prod_{i=1}^{N} \left(1 + \frac{||g_{i,u}(s_u - p_u)||^2}{4N_0 \sin^2 \theta} \right)^{-1} \right]. \tag{4.14}$$

Observe that s_u can be any signal point in the constellation Ω and p_u is also a signal point in Ω whose label differs in only 1 bit at position $j = (k - (u-1)m)$ compared to that of s_u. Therefore, instead of averaging over $m2^{Nm}$ cases of $\boldsymbol{s} \in \boldsymbol{\Psi}$ as in (4.14), $\gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be computed more efficiently by averaging over $m2^m$ cases of $s_u \in \Omega$. Furthermore, $\gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be written by dropping the subscript u for s_u and p_u as follows:

$$\gamma_u(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m2^m} \sum_{s \in \Omega} \sum_{j=1}^{m} \prod_{i=1}^{N} \left(1 + \frac{||g_{i,u}(s - p)||^2}{4N_0 \sin^2 \theta} \right)^{-1}, \tag{4.15}$$

where s and p are two signal points in the conventional constellation Ω whose labels differ in the position j. Therefore, the average in (4.10) can be computed much easier as follows:

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s \in \Omega} \sum_{j=1}^{m} \prod_{i=1}^{N} \left(1 + \frac{||g_{i,u}(s-p)||^2}{4N_0 \sin^2 \theta} \right)^{-1} \right]. \tag{4.16}$$

Applying (4.16) in (4.9), the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be efficiently and accurately computed via a single integral. Also observe that $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be computed by essentially averaging over the constellation Ω. It suggests that $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be denoted as $f(d, \Omega, \xi, \boldsymbol{G})$

To give an insight on how to design the matrix $\boldsymbol{G}$ and good mappings for a given constellation, use the inequality $Q\left(\sqrt{2\gamma}\right) < \frac{1}{2}\exp(-\gamma)$ to approximate the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G}) = f(d, \Omega, \xi, \boldsymbol{G})$ as:

$$f(d, \Omega, \xi, \boldsymbol{G}) \approx \frac{1}{2} \delta^d(\Omega, \xi, \boldsymbol{G}), \tag{4.17}$$

where

$$\delta(\Omega, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s \in \Omega} \sum_{j=1}^{m} \prod_{i=1}^{N} \left(1 + \frac{||g_{i,u}(s-p)||^2}{4N_0} \right)^{-1} \right]. \tag{4.18}$$

The parameter $\delta(\Omega, \xi, \boldsymbol{G})$ can be used to characterize the influence of the rotation matrix $\boldsymbol{G}$, the constellation Ω and the mapping rule ξ to the BER performance of BICM-ID with signal space diversity. In particular, for a given constellation Ω and the mapping ξ, one would prefer the rotation matrix $\boldsymbol{G}$ that minimizes $\delta(\Omega, \xi, \boldsymbol{G})$. In the next section, the optimal choice of $\boldsymbol{G}$ is discussed in more details.

Before closing this section, it is worth mentioning that as far as the asymptotic performance is concerned, the system can achieve a diversity order of Nd, which can be observed from (4.17) and (4.18).

4.3 Optimal Rotation Matrix G and Convergence to an AWGN Channel

This section addresses the design problem to obtain the optimal matrix $\boldsymbol{G}$ for a given constellation and mapping. It is then shown that by using the optimal rotation

matrix $\boldsymbol{G}$, the asymptotic performance of a BICM-ID system over a Rayleigh fading channel can approach that over an AWGN channel.

4.3.1 Optimal Rotation Matrix $\boldsymbol{G}$

By interchanging the summations in (4.18), $\delta(\Omega, \xi, \boldsymbol{G})$ is rewritten as:

$$\delta(\Omega, \xi, \boldsymbol{G}) = \frac{1}{m2^m} \sum_{s \in \Omega} \sum_{j=1}^{m} \kappa(s, p, \boldsymbol{G}). \tag{4.19}$$

where s and p are two signal points in Ω whose labels differ in only 1 bit at position j, $1 \leq j \leq m$ and the parameter $\kappa(s, p, \boldsymbol{G})$ is computed as follows:

$$\kappa(s, p, \boldsymbol{G}) = \frac{1}{N} \sum_{u=1}^{N} \prod_{i=1}^{N} \left(1 + \frac{||g_{i,u}(s-p)||^2}{4N_0}\right)^{-1} \tag{4.20}$$

Under the power constraint in (4.2), it is clear that for a fixed N_0, one has:

$$\sum_{i=1}^{N} \sum_{u=1}^{N} \left(4N_0 + ||g_{i,u}(s-p)||^2\right) = N(4NN_0 + ||s-p||^2). \tag{4.21}$$

Using Cauchy inequality, it is straightforward to show that $\kappa(s, p, \boldsymbol{G})$ in (4.20) achieves the minimum value when $||g_{i,u}|| = \frac{1}{\sqrt{N}}$ for all i and u. This means that the best asymptotic error performance of BICM-ID with SSD is achieved by using the rotation matrix $\boldsymbol{G}$ whose all entries $\{g_{i,u}\}$ are equal in magnitude.

The above class of rotation matrices $\boldsymbol{G}$ is only optimal with respect to the asymptotic performance of the system. However, it is also of interest to study the effect of $\boldsymbol{G}$ on the performance of the system after the first iteration (i.e., the performance of a BICM system). This is because such performance influences the convergence behavior of BICM-ID.

In general, for BICM systems, without the assumption of error-free feedback, the closed-form expression for the function $f(d, \boldsymbol{\Psi}, \xi) = f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ in (2.20) is not simple. Thus, a reasonable approach is to consider the worst case of PEP $P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}})$ in (4.7) for a given Hamming distance d. As pointed out in [53], it is straightforward to see that the worst case of $P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}})$ is equivalent to the worst case of the PEP for

the rotated constellation, $\mathbf{\Psi}_{\mathrm{r}}$, of $\mathbf{\Psi}$. It turns out that for a given outer channel code, the design guideline for BICM is related to uncoded transmission of the signal points in $\mathbf{\Psi}_{\mathrm{r}}$. Specifically, a good choice of $\boldsymbol{G}$ based on the error performance of uncoded system with the constellation $\mathbf{\Psi}_{\mathrm{r}}$ also guarantees a good error performance of BICM. This observation agrees with the study of BICM in [2] where it was shown that the channel code and the modulator can be treated separately. Therefore, the two design criteria in [17, 54] related to the diversity order and the minimum product distance for an uncoded system with SSD can be readily applied for BICM.

From the above analysis, design criteria of the rotation matrix $\boldsymbol{G}$ to achieve good error performance of BICM-ID over a Rayleigh fading channel are summarized as follows:

1. The first priority is to obtain $\boldsymbol{G}$ with all entries equal in magnitude. This guarantees the best asymptotic error performance.

2. $\boldsymbol{G}$ should be chosen to provide a full diversity rotation, i.e., for any $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{\Psi}_{\mathrm{r}}$ and $\boldsymbol{x} \neq \boldsymbol{y}$, $||x_n - y_n|| \neq 0$ for all $1 \leq n \leq N$.

3. When the full diversity is achieved, one would prefer $\boldsymbol{G}$ with a large minimum product distance $\min_{\boldsymbol{x}\neq\boldsymbol{y}\in\mathbf{\Psi}_{\mathrm{r}}} \prod_{n=1}^{N} ||x_n - y_n||$.

At this point, it is natural to ask whether there exists a class of optimal rotation matrices $\boldsymbol{G}$ that satisfies the above design criteria for any value of N. Fortunately, thanks to the algebraic number theory, this class of $\boldsymbol{G}$ is well studied in the literature. For example, when N is a Euler number, $\phi(P)$, with $P \neq 0 \bmod 4$ or N is a power of 2, one might use the following algebraic construction of the rotation matrix (called LCP-A class) proposed in [55] as follows:

$$\boldsymbol{G}_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \alpha_1 & \dots & \alpha_1^{N-1} \\ \vdots & \vdots & & \vdots \\ 1 & \alpha_N & \dots & \alpha_N^{N-1} \end{bmatrix} \tag{4.22}$$

where $\{\alpha_i\}$, $1 \leq i \leq N$, are specified as follows:

- If N is an Euler number, i.e., $N \in \mathcal{N}_1 := \{\phi(P)\}$ with $P \neq 0 \bmod 4$, where $\phi(P)$ is the number of positive integers less than P and relatively prime to P, then $\{\alpha_i\}$ are the roots of $\psi_P(x) = 0$, where $\psi_P(x) = \prod_{p\in\mathcal{P}}(x - \mathrm{e}^{j2\pi p/P})$ and $\mathcal{P} := \{1 \leq p < P : \gcd(p, P) = 1\}$.

- If N is a power of 2, i.e., $N \in \mathcal{N}_2 := \{2^p : p \text{ is an integer}\}$, then $\{\alpha_i\}$ are the roots of $x^N - \sqrt{-1} = 0$.

As proved in [55], the use of $\boldsymbol{G}_N$ along with any BPSK, PAM, or QAM constellations provides full diversity order and the optimal minimum product distance. It is not hard to verify that $||g_{i,u}|| = 1/\sqrt{N}$ for all i, u. Finally, it can be seen that there does not exist any real rotation $\boldsymbol{G}$ to satisfy the first two criteria.

4.3.2 Converting a Fading Channel to an AWGN Channel with BICM-ID

First, consider BICM-ID over an AWGN channel that uses a real two-dimensional constellation Ω and the mapping rule ξ. The tight bound on the function $f(d, \Omega, \xi)$ for an AWGN channel, denoted by $f_{\mathrm{AWGN}}(d, \Omega, \xi)$, is given as [25]:

$$f_{\mathrm{AWGN}}(d, \Omega, \xi) \leq \frac{1}{\pi}\int_0^{\pi/2}\left\{\frac{1}{m2^m}\left[\sum_{s\in\Omega}\sum_{j=1}^{m}\exp\left(-\frac{||s-p||^2}{4N_0\sin^2\theta}\right)\right]\right\}^d \mathrm{d}\theta, \tag{4.23}$$

where, as before, p is the signal point in Ω whose label differs in 1 bit at position j compared to the label of s.

Now consider the asymptotic performance of BICM-ID over the Rayleigh fading channel where an optimal rotation matrix $\boldsymbol{G}$ is used. Recall that for the optimal rotation matrix $\boldsymbol{G}$, $||g_{i,u}|| = \frac{1}{\sqrt{N}}$ for all $1 \leq i, u \leq N$. After some manipulations, (4.16) can be written in a more compact form as:

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m2^m}\left[\sum_{s\in\Omega}\sum_{j=1}^{m}\left(1 + \frac{1}{4N_0}\frac{||s-p||^2}{N\sin^2\theta}\right)^{-N}\right]. \tag{4.24}$$

For each pair of (s, p), define

$$\beta(s, p) = \frac{||s-p||^2}{4N_0\sin^2\theta}. \tag{4.25}$$

Let $P = 1/N$. When N goes to infinity (i.e., P goes to zero), the following identities can be established:

$$\lim_{N\to\infty} \ln\left[\left(1+\frac{\beta(s,p)}{N}\right)^{-N}\right] = -\lim_{P\to 0}\left(\frac{\ln\left[1+\beta(s,p)P\right]}{P}\right)$$

$$= -\lim_{P\to 0}\frac{\left(\frac{1}{1+\beta(s,p)P}\right)\cdot\beta(s,p)}{1} = -\beta(s,p). \tag{4.26}$$

It then follows that:

$$\lim_{N\to\infty}\left[\left(1+\frac{\beta(s,p)}{N}\right)^{-N}\right] = \exp\left(-\beta(s,p)\right). \tag{4.27}$$

Combining (4.27), (4.25), (4.24) and (4.9) shows that the right-hand side of (4.9) approaches the right-hand side of (4.23) when the optimal rotation matrix $\boldsymbol{G}$ is used and N goes to infinity. This means that the use of signal space diversity with a very large dimension N makes the asymptotic performance of a BICM-ID system over a Rayleigh fading channel approach that over an AWGN channel. This result is in parallel with the previous result established in [17] for uncoded systems.

4.4 Proposed M-QAM Mapping and Sub-Optimal Receiver

The SSD technique offers a significant performance improvement, at the cost of high complexity receiver, especially for a high dimensional signal space. For example, the complexity of the optimal MAP demodulator grows exponentially with the number of coded bits per signal point Nm, which makes the implementation become infeasible. Based on a so-called sigma mapping, this section proposes a sub-optimal yet effective demodulator to deal with the complexity issue of SSD in very large dimensions.

4.4.1 Sigma Mapping

The basic idea behind the proposed mapping of M-QAM is to relate the vectors of coded bits to the transmitted QAM symbols in a linear manner. In essence, such a mapping allows one to view a single-user M-ary system as a code-division

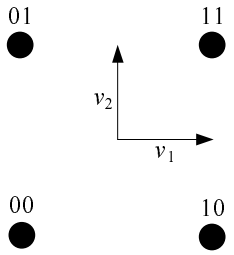
multiple-access (CDMA), or multi-user system, where "each user" employs antipodal binary signaling (e.g., the BPSK). Then a sub-optimal soft-output MMSE demodulator, which is similar to the soft-output MMSE multi-user detector, can be applied. It should be mentioned here that this type of mapping was originally proposed in [56] as a capacity-approaching mapping method for a multilevel coding scheme. A more detailed investigation of this mapping (called sigma mapping) for M-PAM constellations was also recently carried out in [57].

In what follows, the sigma mapping for a complex constellation will be generalized and presented for M-QAM. Consider the set of m basis complex vectors $\{v_j\}$, $1 \leq j \leq m$, which satisfy $\sum_{j=1}^{m} ||v_j||^2 = m$. For each vector of m (coded) bits $\boldsymbol{b} = (b_1, \ldots, b_m)$, define the following sigma mapping from $\boldsymbol{b}$ to a complex signal point s in a two-dimensional signal space:

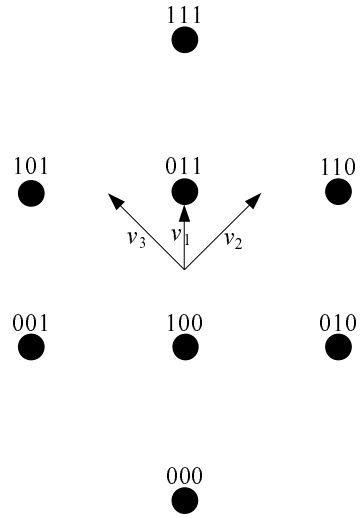
$$s = \boldsymbol{v}(2\boldsymbol{b} - 1)^{\top}, \tag{4.28}$$

where $\boldsymbol{v} = [v_1, \ldots, v_m]$. Thus, for a fixed set of basis vectors, the $M = 2^m$ binary tuples $\{\boldsymbol{b}\}$ create an M-ary sigma constellation, denoted by $\Omega_{\sum}$, with an average symbol energy of m. Observe that (4.28) can be interpreted as a received signal (in the absence of background noise) in a binary CDMA system with m users whose signature waveforms (or signature sequences) are $v_1, \ldots, v_m$.

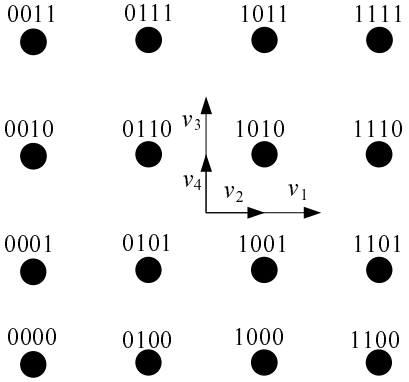
Obviously, the shape of $\Omega_{\sum}$ depends on the set of basis vectors $\{v_j\}_{j=1}^{m}$. Furthermore, it can be shown that any M-QAM constellation can be considered as an $\Omega_{\sum}$ constellation by using a proper set of the basis vectors. As examples, Fig. 4.2 presents the sigma mappings for QPSK, 8-QAM and 16-QAM constellations, where the corresponding basis vectors are also indicated. Note that the sigma mapping is exactly the Gray mapping for QPSK constellation. For larger constellations, e.g., M=64 or 256, there might exist multiple and nonequivalent basis vectors $\boldsymbol{v}$ to apply the sigma mapping and the selection of a particular one might affect the performance of the iterative demodulation/decoding. In such a case, care should be taken to choose the basis vector that yields the best performance. Furthermore, the technique of sigma mapping can also be extended to non-traditional constellations (i.e., with arbitrary



Sigma mapping for QPSK.



(b) Sigma mapping for 8-QAM.



(c) Sigma mapping for 16-QAM.

Figure 4.2 Sigma mappings for various M-QAM constellations: (a) Sigma mapping for QPSK with $v_1 = \exp(j0)$ and $v_2 = \exp(j\frac{\pi}{2})$; (b) Sigma mapping for 8-QAM with $v_1 = \sqrt{0.6}\exp(j\frac{\pi}{2})$, $v_2 = \sqrt{1.2}\exp(j\frac{\pi}{4})$ and $v_3 = \sqrt{1.2}\exp(j\frac{3\pi}{4})$; (c) Sigma mapping for 16-QAM with $v_1 = 2v_2 = \sqrt{1.6}\exp(j0)$ and $v_3 = 2v_4 = \sqrt{1.6}\exp(j\frac{\pi}{2})$.

shapes) for more flexibility in the design of the basis vector $\boldsymbol{v}$.

4.4.2 Suboptimal Soft-Output Demodulator

Consider a signal space diversity system in a complex N-D signal space. Let each complex component be implemented using a sigma mapping with the basis vectors $\{v_j\}$, $1 \leq j \leq m$. With $\boldsymbol{v} = [v_1, \ldots, v_m]$, create an $N \times Nm$ block diagonal matrix $\boldsymbol{V}$ as follows:

$$\boldsymbol{V} = \text{diag}\underbrace{(\boldsymbol{v}, \ldots, \boldsymbol{v})}_{N\,\boldsymbol{v}'s}. \tag{4.29}$$

Let $\boldsymbol{C}$ be the $1 \times Nm$ coded bit vector whose entries are either 1 or 0 and $\boldsymbol{B} = 2\boldsymbol{C} - 1$. From (4.28) and (4.29), the symbol sequence $\boldsymbol{s}$ in $\boldsymbol{\Psi}$ corresponding to $\boldsymbol{C}$ is given as:

$$\boldsymbol{s}^\top = \boldsymbol{V}(2\boldsymbol{C} - 1)^\top = \boldsymbol{V}\boldsymbol{B}^\top. \tag{4.30}$$

Hence, the received signal $\boldsymbol{r}$ in (4.3) can be rewritten as follows:

$$\boldsymbol{r}^\top = \boldsymbol{H}\boldsymbol{G}\boldsymbol{V}\boldsymbol{B}^\top + \boldsymbol{w}^\top. \tag{4.31}$$

Define the real matrices $\boldsymbol{R} = [\text{real}(\boldsymbol{r}^\top); \text{imag}(\boldsymbol{r}^\top)]$, $\boldsymbol{X} = [\text{real}(\boldsymbol{H}\boldsymbol{G}\boldsymbol{V}); \text{imag}(\boldsymbol{H}\boldsymbol{G}\boldsymbol{V})]$ and $\boldsymbol{n} = [\text{real}(\boldsymbol{w}^\top); \text{imag}(\boldsymbol{w}^\top)]$. Then one obtains:

$$\boldsymbol{R} = \boldsymbol{X}\boldsymbol{B}^\top + \boldsymbol{n}. \tag{4.32}$$

With the above linear real matrix system, the efficient MMSE estimation proposed in [52] can be readily applied to compute the extrinsic information for the coded bits as the output of the demodulator.

It is of interest to compare the complexity of the proposed sub-optimal MMSE demodulator and the scalar and vector GA algorithms introduced in [51]. To examine and quantitatively compare the complexity of different sub-optimal demodulating methods, the complexity of arithmetic computations can be measured in terms of the floating point operation (FLOP) [58]. Specifically, one FLOP is counted for a real subtraction or addition and two FLOPS are counted for the complex ones. Furthermore, multiplications and divisions are counted as one FLOP each if the result is real. Otherwise they are counted as six FLOPS. In the following, the complexity estimation of each sub-optimal receiver will be presented. The counting is largely based on [59],

although reference [59] counts one FLOP for a single addition, substraction, multiplication or division between two complex numbers. Furthermore, an exponential, logarithm or a trivial operation with a constant are counted as one FLOP. Obviously, it is infeasible to count exactly all FLOPS. Nevertheless, the most important ones shall be counted. For a fair comparison, the number of FLOPS per each iteration corresponds to 1 signal point in N-D signal space, or equivalently Nm coded bits.

The Vector GA Algorithm

First, to compute the *a priori* probabilities for all signal points in 1-D space requires $N(4m+2)2^m$ FLOPS. Second, computing the expectations and variances for signal points in 1-D space needs $16N2^m - 3N$ FLOPS. Third, the soft estimation cancellation uses $8N^3 - 2N^2$ FLOPS. Fourth, evaluating the density function involves the expectation, the covariance matrix, the inverse of the covariance matrix and it requires $14N^2 2^m + 9N^4 + 20N^4/3 + 4N^3 - 5N^2 + (8N^3 + 8N^2 + N)2^m$ FLOPS. Finally, computing the extrinsic information for coded bits needs $9Nm2^m$ FLOPS. Putting the above numbers together, the total number of FLOPS for vector GA algorithm is $2^m(8N^3 + 22N^2 + 13Nm + 19N) + 46N^4/3 + 12N^3 - 7N^2 - 3N$ FLOPS per Nm coded bits[2].

The Scalar GA Algorithm

The computation of the *a priori* probabilities for signal points in 1-dim space uses $N^2(4m+2)2^m$ FLOPS. Computing the expectations and variances for signal points in 1-dim space needs $16N^2 2^m - 3N^2$ FLOPS. The soft estimation cancellation requires $8N^3 - 2N^2$ FLOPS. Evaluating the density function needs $22N^2 2^m + 8N^3$ FLOPS. And finally it requires $9N^2 m2^m$ FLOPS to compute the extrinsic information

[2]It should be mentioned here that there is a mistake in computing the number of real multiplications for vector GA method in [51]. Specifically, the parameter N^3 should not be counted for 2^{m-1} signal points. This is because the inversion operation of the covariance matrix does not depend on each signal point. This fact has been verified and confirmed with an author of [51] via email communication.

for coded bits. Thus the total computation cost for the scalar GA algorithm is $2^m(13N^2m + 40N^2) + 16N^3 - 5N^2$ FLOPS per Nm coded bits.

The MMSE Demodulator

The complexity of the soft-output MMSE demodulator is dominated by the matrix inversion operation. For an $N \times N$ real matrix, the number of FLOPS required to implement the inverse operation is $5N^3/3$. For a special case with diagonal matrix, it requires only N FLOPS. The total cost for MMSE demodulator is determined to be $N^3(11/3m^3 + 4m^2 + 16m) + N^2(5m^2 + 4m) + N(7m + 2)$ FLOPS per Nm coded bits.

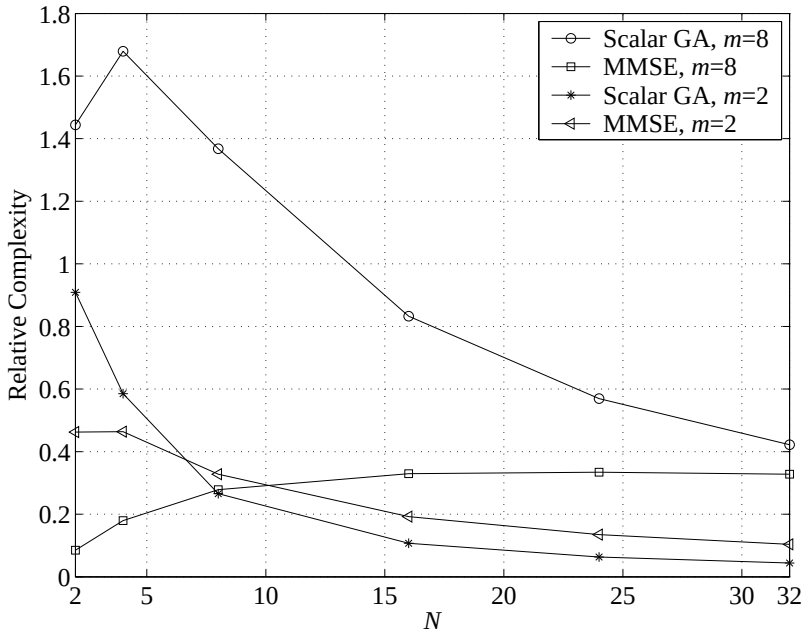


Figure 4.3 Relative complexity of the scalar GA and MMSE methods compared to the vector GA algorithm.

Figure 4.3 shows the relative complexity of the scalar GA and MMSE methods over the complexity of the vector GA (i.e., the complexity of vector GA is normalized to be 1 for any values of N and m). Observe from Fig. 4.3 that with small values of m (e.g., $m = 2$), the demodulator using the scalar GA algorithm is simpler than the proposed MMSE demodulator for all high values of N shown in this figure. However,

as m increases to 8 and with practical values of N (such as $N = 16,\ 32$), the MMSE demodulator is a better choice in terms of computational complexity. Figure 4.3 also clearly demonstrates that the vector GA algorithm gives the most complicated sub-optimal demodulator for almost all cases. The only exceptions are for small values of N and large values of m, where the scalar GA algorithm is more complicated than the vector GA algorithm. This is due to the fact that the scalar GA algorithm needs to compute N^2m extrinsic information values, instead of Nm values as in the case of the vector GA method [51].

4.5 Illustrative Results

In this section, analytical and simulation results are provided to confirm the performance analysis carried out in the previous sections. In all the computations of the bound on the bit error probability, the first 20 Hamming distances of the convolutional codes are retained to make sure the accuracy of the bound. The convergence analysis of the sub-optimal iterative receiver based on the soft-output MMSE demodulator by means of the extrinsic information transfer (EXIT) charts is also briefly presented.

4.5.1 Signal Space Diversity with Sigma Mapping of QPSK

Consider a BICM-ID-SSD system that employs a QPSK constellation with sigma mapping (i.e., Gray mapping), a rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ and $N = 2$. Furthermore, a random interleaver of length 1920 bits is used. Figure 4.4 presents the BER performance of the system after 1 and 4 iterations using different soft-output demodulators, namely the proposed MMSE, the scalar GA, the vector GA, and the MAP demodulators. The rotation matrix is chosen as in (4.22) with $\alpha_1 = \alpha = \exp\left(j\frac{2\pi}{4N}\right)$ and $\alpha_i = \alpha \exp\left(j\frac{2\pi(i-1)}{N}\right)$. It can be observed that the performance of the sub-optimal demodulators is significantly improved with the number of iterations. After 4 iterations, the simulation results on the performance of all the demodulation methods under consideration are almost the same and close to the analytical BER bound as shown in Fig. 4.4.

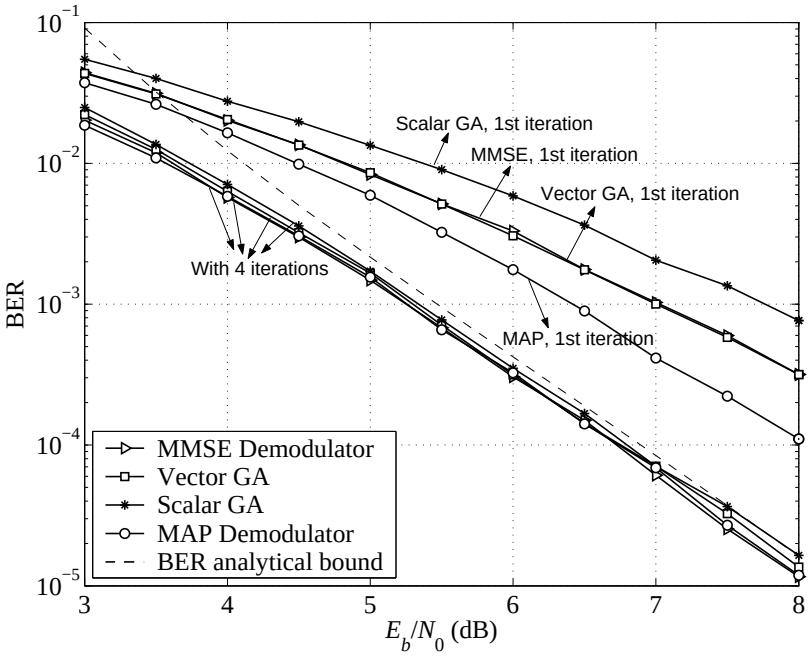


Figure 4.4 BER performance of BICM-ID-SSD with various demodulation methods and $N = 2$.

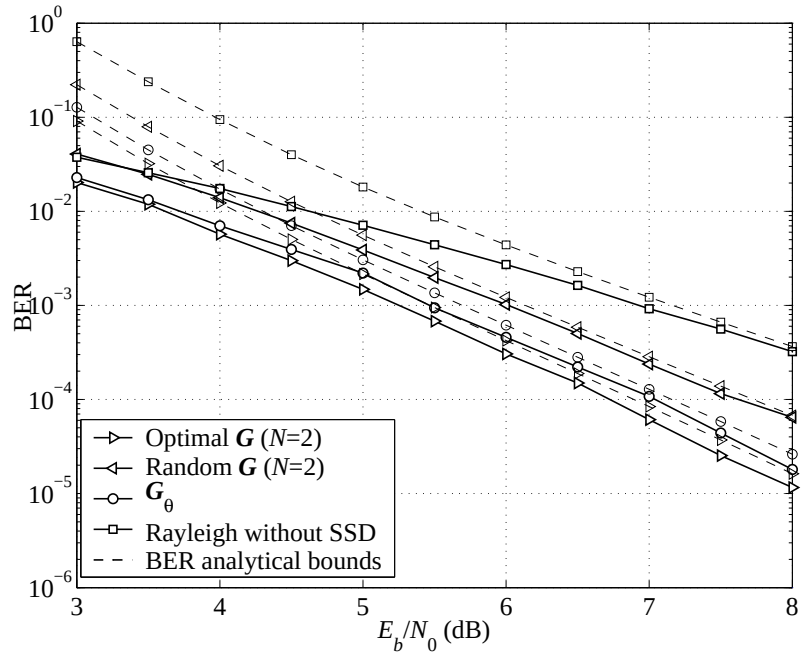


Figure 4.5 BER performance for BICM-ID with signal space diversity: Comparison of different rotation matrices, $N = 2$ and proposed MMSE demodulator.

Figure 4.5 demonstrates the role of the rotation matrix $\boldsymbol{G}$, where the BER performance after 4 iterations of BICM-ID-SSD with 2-D constellation ($N = 2$) and the proposed MMSE demodulator are compared for difference choices of $\boldsymbol{G}$. These choices include (i) an optimal matrix $\boldsymbol{G}$ in (4.22); (ii) a randomly-generated $\boldsymbol{G}$ as considered in [51]; and (iii) the following optimal rotation matrix $\boldsymbol{G}_\theta$ that maximizes the minimum product distance for real unitary transformation with a full modulation diversity [17, 46, 60, 61]:

$$\boldsymbol{G}_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{4.33}$$

with $\tan(2\theta) = 2$. Note that $\boldsymbol{G}_\theta$ is not optimal in terms of the asymptotic performance for BICM-ID-SSD, since their elements have different magnitudes. Also plotted in this figure are the error bounds for all the systems under consideration (the broken lines) and the performance of BICM-ID without signal space diversity. Observe that, compared to the random $\boldsymbol{G}$ and $\boldsymbol{G}_\theta$, the use of the optimal rotation matrix results in coding gains as high as 0.25dB and 0.8dB at the BER level of 10^{-4}, respectively. Note also that the error bounds are tight for all the systems.

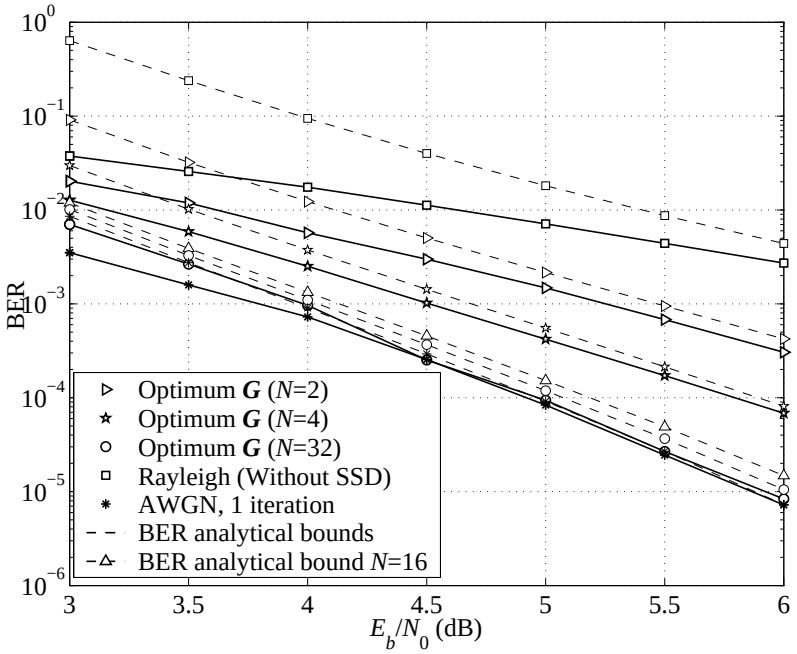


Figure 4.6 BER performance for BICM-ID with signal space diversity employing MMSE demodulator: Optimal rotation matrices for various N.

To understand the influence of the SSD and the impact of the dimensionality of the constellation (i.e., the parameter N) to the error performance, Fig. 4.6 shows the BER performance of BICM-ID-SSD over a Rayleigh fading channel after 4 iterations with $N = 2$, $N = 4$ and $N = 32$. The proposed MMSE soft-output demodulator is used with rotation matrix $\boldsymbol{G}$ selected from (4.22). For comparison, the performance of BICM-ID without SSD and using the MAP demodulator is also shown. Note that, because Gray mapping is used for QPSK, iterative processing is useless for the systems without SSD. The error bounds derived in the previous section for $N = 2$, $N = 4$, $N = 16$ and $N = 32$ are also provided in this figure in order to verify the tightness of the bound[3]. Furthermore, both the analytical bound and the simulation result of system performance with MAP demodulator over an AWGN channel are provided to serve as performance benchmarks for the Rayleigh fading channel. Since Gray mapping is used, the error performance over an AWGN channel is obtained with only 1 iteration.

It can be observed that by using sub-optimal MMSE demodulator, BICM-ID systems can fully exploit the advantage of SSD technique for various values of N. The error bounds of all the systems are very tight, except at the region of low SNR. Thus, the error bound derived in this chapter can be very useful to predict the BER performance of BICM-ID with SSD. Analytical BER bounds show significant performance gains when increasing N from 2 to 16. The gains become saturated when N is larger than 16. The error bounds for N =16 and 32 have a negligible difference of less than 0.05dB. Figure 4.6 also shows that the simulation results after 4 iterations for $N = 32$ are very close to the analytical BER bound. Furthermore, they approach the performance over an AWGN channel at $E_b/N_0 > 4.0$dB. These results clearly support the analytical evaluation made in the previous sections.

Finally, Fig. 4.7 compares the error performances of the systems using two different sub-optimal demodulators, namely the MMSE and the scalar GA methods with $N = 32$. It can be seen that both methods exploit efficiently the benefit of SSD

[3]For readability, only the BER analytical bound is shown for $N = 16$.

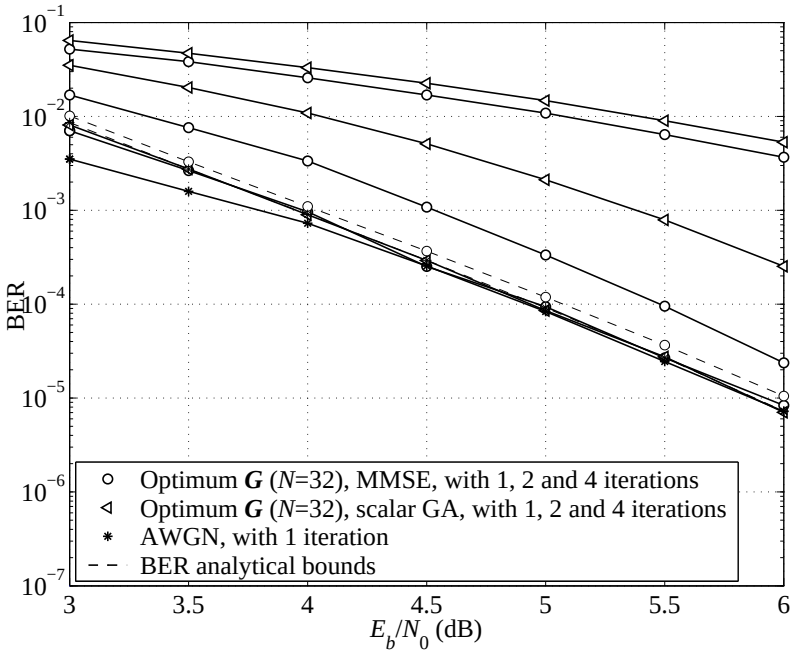


Figure 4.7 BER performances for BICM-ID with signal space diversity and $N = 32$ employing MMSE and scalar GA demodulators. The sigma (Gray) mapping of QPSK is used.

technique. The error performances of both systems converge to the error bound and are close to the error performance over an AWGN channel. However, the figure shows that the performance of the MMSE demodulator converges to the error bound faster than that of the scalar GA algorithm. It is worth mentioning that the vector GA algorithm becomes too complicated for implementation with $N = 32$ [51].

4.5.2 Further Results With Higher-Order Modulation

This subsection provides more analytical and simulation results for the systems employing 16-QAM and 8-QAM constellations and the respective sigma mappings and with large values of N (e.g., $N = 32$). The two sub-optimal MMSE and scalar GA demodulators are used with a random interleaver of 9600 bits. Since large values of N are used, the vector GA demodulator is too complicated to implement.

Figure 4.8 presents the BER performance after 1, 5, 9 and 12 iterations for the systems employing 16-QAM and the sigma mapping with $N = 32$. Again the optimal

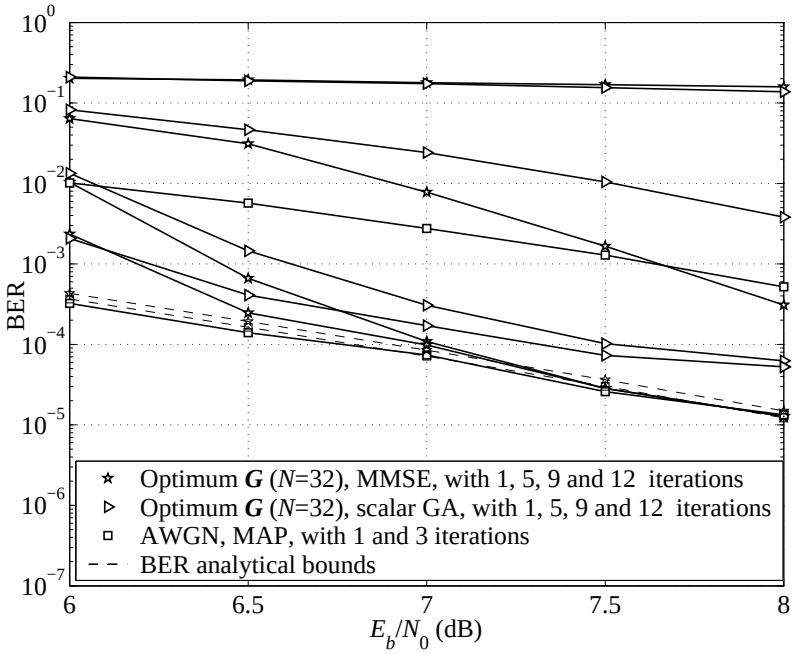


Figure 4.8 BER performances for BICM-ID-SSD using sigma mapping of 16-QAM with optimal rotation matrix $\boldsymbol{G}$ for $N = 32$. MMSE and scalar GA methods are employed.

rotation matrix is the matrix constructed as in (4.22). The same rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ is also used in these systems. This code together with a 16-QAM constellation yields a spectral efficiency of 2 bits/s/Hz. For comparison, the BER of the system employing MAP receiver with 1 and 3 iterations over an AWGN channel is also provided.

Similar to the case of QPSK, Fig. 4.8 shows that the performance of the BICM-ID-SSD system with $N = 32$ employing MMSE demodulator over a Rayleigh fading channel approaches that in an AWGN channel and the simulation results and the analytical BER bound are very close. However, it is noticed that, compared to the case of QPSK, the required number of iterations for 16-QAM to approach the bound is larger, i.e., 12 or more. In the case of scalar GA demodulator, it fails to fully exploit the advantage of SSD technique, where its performance does not converge to the error bound. From the extensive simulation results with different QAM constellations and values of N, though not shown here in the interest of space, it is observed that the error

performance of BICM-ID-SSD with scalar GA method cannot closely approach the performance in an AWGN channel except for the case of QPSK and Gray mapping.

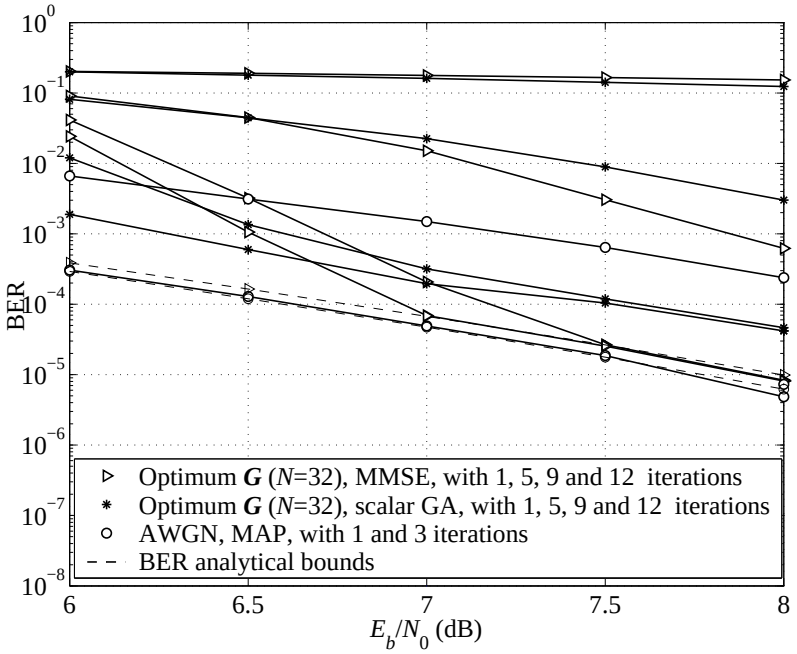


Figure 4.9 BER performance for BICM-ID-SSD using the sigma mapping of 8-QAM with optimal rotation matrix $\boldsymbol{G}$ for $N = 32$. MMSE and scalar GA methods are employed.

Figure 4.9 shows similar analytical and simulation results for a BICM-ID-SSD system employing the sigma mapping of 8-QAM constellation. Furthermore, a rate-2/3, 4-state convolutional code with generator matrix $\mathbf{g}_1 = (6, 2, 6)$ and $\mathbf{g}_2 = (2, 4, 4)$ is used, which results in a spectral efficiency of 2 bits/s/Hz. Clearly, similar observations can be made with respect to the tightness of the analytical error bounds and the performance convergence between the fading and the AWGN channels for the MMSE and scalar GA methods.

4.5.3 Convergence Analysis with EXIT Charts

The performance convergence of the BICM-ID-SSD system over a Rayleigh fading channel employing the low-complexity MMSE demodulator can also be analyzed by the histogram method using extrinsic information transfer (EXIT) chart as introduced

in subsection 2.3.3, Chapter 2.

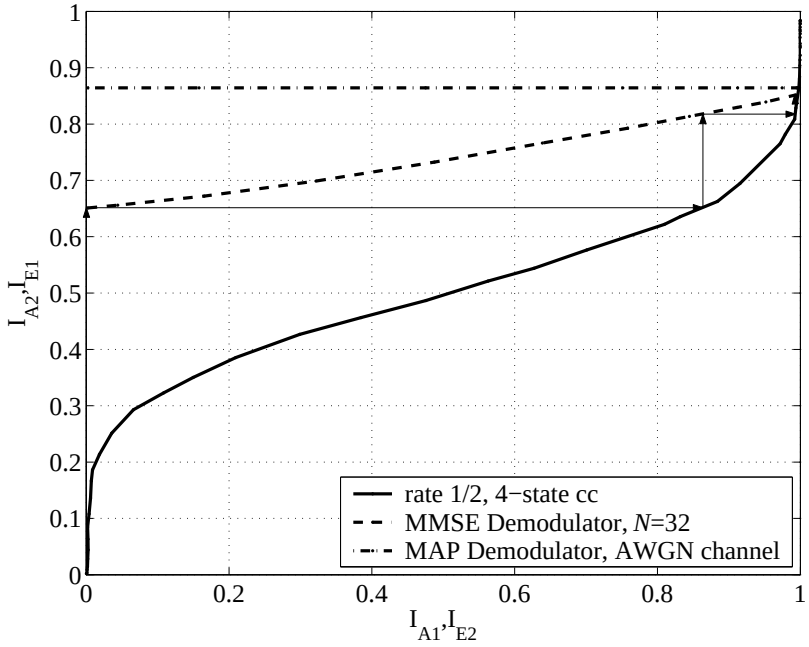


Figure 4.10 EXIT charts with iterative decoding trajectories of BICM-ID-SSD employing sigma mapping of QPSK, $N = 32$ and rate-1/2, 4-state convolutional code.

Figure 4.10 plots the EXIT charts for the BICM-ID-SSD system employing QPSK with sigma mapping and $N = 32$ over a Rayleigh fading channel. Here the MMSE demodulator and the outer rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ are employed. For this figure, the signal to noise ratio E_b/N_0 is set at 5dB. Also for comparison, the EXIT charts for BICM-ID employing the MAP demodulator over an AWGN channel are also provided.

As can be seen from Fig. 4.10, the iterative demodulation and decoding is useless for sigma (Gray) mapping of QPSK over an AWGN channel. In particular, the bit-wise mutual information for sigma mapping remains constant and does not depend on the *a priori* information I_{A1}. More interestingly, for BICM-ID-SSD, the mutual information I_{E1} approaches that of BICM-ID-SSD over an AWGN channel when the perfect knowledge of I_{A1} is achieved. Based on the decoding trajectory, it can be seen that it takes just only 3 iterations for the convergence of the iterative decoding. This

prediction by the EXIT charts closely matches with the simulation results presented in Fig. 4.7.

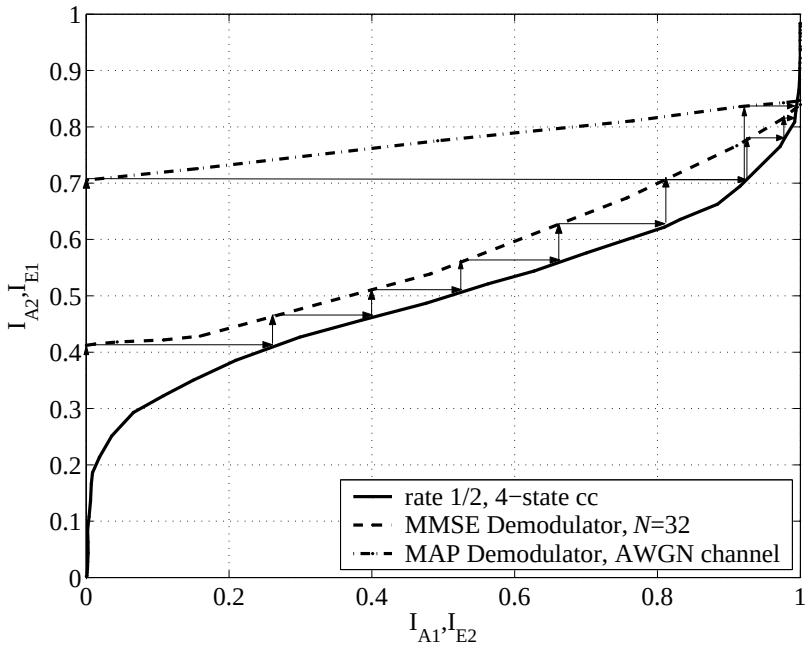


Figure 4.11 EXIT charts with iterative decoding trajectories of BICM-ID-SSD employing sigma mapping of 16-QAM, $N = 32$ and a rate-1/2, 4-state convolutional code.

Similar EXIT charts can be obtained for higher modulation schemes. In Fig. 4.11, the EXIT charts for systems using sigma mapping of 16-QAM and with the same convolutional code are provided at $E_b/N_0 = 7.0$dB. Obviously, when a perfect knowledge of I_{A1} is fed back to the demodulator, the same information I_{E1} can be achieved for both systems (over the Rayleigh and AWGN channels). It again confirms the convergence between the performance over an AWGN channel and the performance over a Rayleigh fading channel with SSD. As can be seen from the decoding trajectory, the iterative demodulation and decoding work well at $E_b/N_0 = 7.0$dB. However, it needs more iterations to converge to the asymptotic performance. This prediction also agrees with the simulation results presented in Fig. 4.8.

5. BICM-ID with Signal Space Diversity over Cascaded Rayleigh Fading Channels[1]

It was demonstrated in Chapter 4 that BICM-ID employing a standard convolutional code together with signal space diversity (SSD) can provide a low-complexity, but yet effective method in combining channel coding with high-level modulation schemes. In particular, Chapter 4 showed that using SSD with a sufficiently large dimension of the lattice constellation makes the performance of BICM-ID over a Rayleigh fading channel approach that over an AWGN channel. In essence, Chapter 4 has partially addressed the current and important research problem in wireless communications over a Rayleigh fading channel: Practically eliminate the adverse effects of fading with a low complexity receiver structure.

Even though the Rayleigh fading model has been considered in many research works in the literature, this fading model is only suitable to characterize the statistical properties of the conventional cellular radio systems where the mobile station is moving and the base-station is stationary. On the other hand, in several new applications such as mobile ad-hoc networks and dedicated short-range communication systems for intelligent highway, both the transmitter and receiver are in motion, which results in different statistical properties of the wireless channel. Under this circumstance, one needs to take into account the double scattering effect in the channel, where scatters are both around the transmitter and the receiver. Both theoretical analysis and experimental results [64–67] recently indicated that cascaded Rayleigh fading, induced by double scattering, is more suitable to model the mobile-to-mobile

[1] A part of this chapter was presented in [62, 63]

communications channels. This fact thus brings new challenges in performance analysis, and as a consequence, the design criterion for mobile-to-mobile communications. Surprisingly, to date, only a few efforts have been devoted to performance analysis of communications systems operating over cascaded Rayleigh fading channels [66–69]. For example, by studying the pairwise error probability (PEP) of uncoded systems over cascaded Rayleigh fading channels, it was demonstrated in [68, 69] that only a partial diversity order can be achieved, which implies a performance degradation when compared to a conventional Rayleigh fading channel.

This chapter is therefore concerned with the application of SSD in BICM-ID to improve the error performance of communications systems operating over cascaded Rayleigh fading channels. This can be understood as an extension of the study in Chapter 4, with a similar derivation framework, but for cascaded Rayleigh fading.

The chapter starts with a brief introduction of cascaded Rayleigh fading. Then a tight bound on the asymptotic error performance of BICM-ID with SSD is derived. Based on the error bound, it is established that, similar to the case of a Rayleigh fading channel, the entries of the optimal rotation matrix (with respect to the asymptotic performance) for a cascaded Rayleigh fading channel also have equal magnitudes and the optimal rotation matrix presented in Chapter 4 can be readily applied for the case of cascaded Rayleigh fading. By comparing the error performance of BICM-ID with SSD over the conventional and cascaded Rayleigh fading channels, it is shown that employing the technique of SSD in a sufficiently large dimension can close the performance gap between the two fading channels. Moreover, it is observed that by employing the optimal rotation matrix, the error performance of BICM-ID with SSD over a cascaded Rayleigh fading channel can closely approach the performance of BICM-ID over an AWGN channel. This is similar to the observation made in Chapter 4 for the case of BICM-ID with SSD over a conventional Rayleigh fading channel.

5.1 Cascaded Rayleigh Fading and Its Probability Density Function

For BICM-ID with SSD over a cascaded Rayleigh fading channel, the system model is similar to the model presented in Chapter 4, except that the fading coefficients $\{h_i\}$ follow a different distribution.

Due to double scattering, information-theoretic and error probability studies [66–72] have indicated that there is always a severe performance degradation in a cascaded Rayleigh fading channel compared to that over conventional Rayleigh fading channel, where only single scattering is assumed. In the case of cascaded Rayleigh fading, each fading coefficient h_i is represented by the product of two independent circularly symmetric complex Gaussian random variables $h_i = a_i b_i$, where both a_i and b_i have unit variance, i.e., $\mathcal{CN}(0, 1)$. It is not hard to verify that the random variable $z = ||h_i||^2$ has the following density function:

$$f_Z(z) = 2\mathcal{K}_0\left(2\sqrt{z}\right) \tag{5.1}$$

where $\mathcal{K}_\mu(\cdot)$ is the modified Bessel function of the second kind [73].

Figure 5.1 plots the probability density functions (pdf) of a random variable z for cascaded Rayleigh fading given in (5.1) and Rayleigh fading model given in (3.6). It can be seen that for cascaded Rayleigh fading, there is a much higher probability that the channel is in deep fade, i.e., z is small. That simply explains for the reason why the error performance over cascaded Rayleigh fading is degraded.

5.2 Performance Evaluation for BICM-ID with SSD over Cascaded Rayleigh Fading Channels

This section derives a similar error bound for BICM-ID with SSD as in Chapter 4 but for a cascaded Rayleigh fading channel. As mentioned earlier, the derivation framework is the same with Chapter 4. However, the details are quite different and much more involved for the case of cascaded Rayleigh fading.

Again, in order to derive the union bound on the asymptotic BEP for a BICM-ID

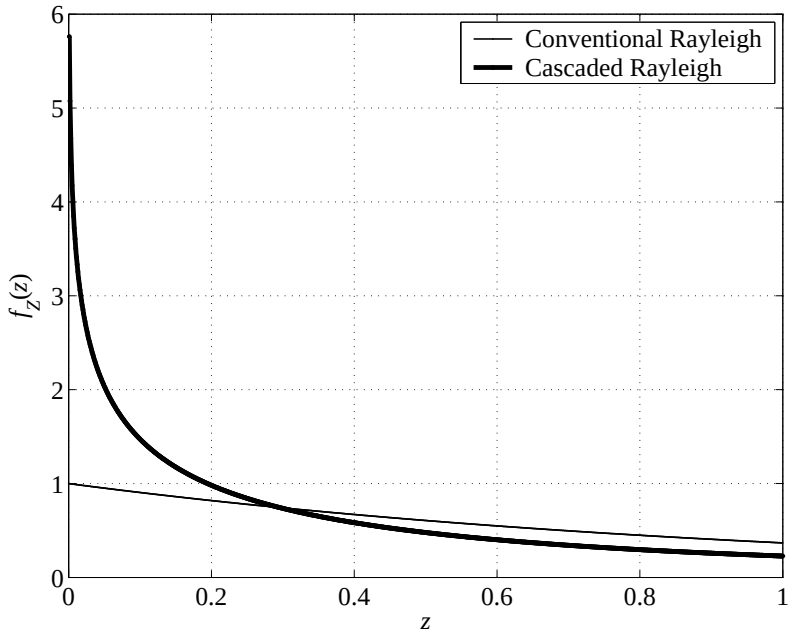


Figure 5.1 The probability density functions $f_Z(z)$ of a random variable z for cascaded Rayleigh and Rayleigh fading.

with signal space diversity that employs a rate-k_c/n_c convolutional code, a complex N-dim constellation $\boldsymbol{\Psi}$ and a mapping rule ξ, one needs to compute the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$, which is the average pair-wise error probability (PEP). Using the same notations as in Chapter 4 and let $\underline{\boldsymbol{H}} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_d]$, where $\boldsymbol{H}_e = \text{diag}(h_{e,1}, \ldots, h_{e,N})$, $1 \leq e \leq d$, represents the path gains that affect the transmitted symbol $\boldsymbol{s}_e$. As mentioned before, over a cascaded Rayleigh fading channel, the path gain $h_{e,i}$, $1 \leq i \leq N$, is expressed as $h_{e,i} = a_{e,i}b_{e,i}$. Then the PEP conditioned on the fading coefficients $\underline{\boldsymbol{H}}$ can be computed as follows:

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}|\underline{\boldsymbol{H}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d} d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e)}\right), \tag{5.2}$$

where

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \left|\left|\boldsymbol{H}_e\boldsymbol{G}(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right|\right|^2 = \sum_{i=1}^{N} ||a_{e,i}||^2 \cdot ||b_{e,i}||^2 \cdot \left|\left|\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right|\right|^2. \tag{5.3}$$

with $\boldsymbol{g}_i$ is the ith row of $\boldsymbol{G}$. Using the Gaussian probability integral and averaging over

the Rayleigh random variables $\{||a_{e,i}||\}$ and $\{||b_{e,i}||\}$ as similar to [69], one obtains

$$P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\prod_{e=1}^{d} \Delta_e \right) \mathrm{d}\theta, \tag{5.4}$$

where

$$\Delta_e = \prod_{i=1}^{N} \frac{E_1\left(\frac{4N_0 \sin^2\theta}{||\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top||^2}\right)}{E_0\left(\frac{4N_0 \sin^2\theta}{||\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top||^2}\right)}, \tag{5.5}$$

and the functions $E_1(x) = \int_x^\infty \frac{\exp(-t)}{t}\mathrm{d}t$ and $E_0(x) = \frac{\exp(-x)}{x}$ [73].

Thanks to the success of iterative decoding steps, the union bound on $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ for cascaded Rayleigh fading can be computed by averaging over the constellation $\boldsymbol{\Psi}$ as

$$f(d, \Psi, \xi, \boldsymbol{G}) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[\underbrace{\mathcal{E}_{\boldsymbol{s},\boldsymbol{p}}\left\{\Delta(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{G})\right\}}_{\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G})} \right]^d \mathrm{d}\theta, \tag{5.6}$$

where

$$\Delta(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{G}) = \prod_{i=1}^{N} \frac{E_1\left(\frac{4N_0 \sin^2\theta}{||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2}\right)}{E_0\left(\frac{4N_0 \sin^2\theta}{||\boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^\top||^2}\right)}. \tag{5.7}$$

The expectation in (5.6) is over all the pairs of $\boldsymbol{s}$ and $\boldsymbol{p}$ in $\boldsymbol{\Psi}$ whose labels differ in only 1 bit at position k, which is computed as:

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{M_c 2^{M_c}} \sum_{\boldsymbol{s} \in \boldsymbol{\Psi}} \sum_{k=1}^{M_c} \Delta(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{G}). \tag{5.8}$$

where $M_c = Nm$ is a number of coded bits per signal point in $\boldsymbol{\Psi}$.

Similar to the analysis in Chapter 4, it can be verified that if ξ is implemented independently for each component of $\boldsymbol{s}$, then there is only one distinct component between $\boldsymbol{s}$ and $\boldsymbol{p}$ as long as their labels differ in only 1 bit. After some manipulations, the average in (5.8) can be computed by simply averaging over all pairs of signal points (s, p) in the conventional constellation Ω whose labels differ in only 1 bit at position j, $1 \leq j \leq m$. That is,

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m 2^m} \sum_{s \in \Omega} \sum_{j=1}^{m} \alpha(s, p, \boldsymbol{G}), \tag{5.9}$$

where

$$\alpha(s,p,\boldsymbol{G}) = \frac{1}{N}\sum_{u=1}^{N}\prod_{i=1}^{N}\frac{E_1\left(\frac{4N_0\sin^2\theta}{||g_{i,u}(s-p)||^2}\right)}{E_0\left(\frac{4N_0\sin^2\theta}{||g_{i,u}(s-p)||^2}\right)}. \tag{5.10}$$

Applying (5.9) to compute $f(d,\boldsymbol{\Psi},\xi,\boldsymbol{G})$ in (5.6), the tight union bound on the asymptotic BEP in (2.20) can be efficiently computed via a single integral. Since the function $f(d,\boldsymbol{\Psi},\xi,\boldsymbol{G})$ now depends on the constellation Ω instead of the constellation $\boldsymbol{\Psi}$, $f(d,\boldsymbol{\Psi},\xi,\boldsymbol{G})$ can be simply denoted as $f(d,\Omega,\xi,\boldsymbol{G})$ as in Chapter 4.

For a fixed constellation Ω and mapping rule ξ, the problem of choosing good rotation matrix $\boldsymbol{G}$ in order to minimize the asymptotic performance can be answered by applying the Chernoff bound $Q\left(\sqrt{2x}\right) < \frac{1}{2}\exp(-x)$. Using the same analysis as before, one obtains the approximated version of $f(d,\Omega,\xi,\boldsymbol{G})$ which does not involve integral:

$$f(d,\Omega,\xi,\boldsymbol{G}) \sim \frac{1}{2}\delta^d(\Omega,\xi,\boldsymbol{G}), \tag{5.11}$$

where

$$\delta(\Omega,\xi,\boldsymbol{G}) = \frac{1}{m2^m}\sum_{s\in\Omega}\sum_{j=1}^{m}\kappa(s,p,\boldsymbol{G}), \tag{5.12}$$

with

$$\kappa(s,p,\boldsymbol{G}) = \frac{1}{N}\sum_{u=1}^{N}\prod_{i=1}^{N}\frac{E_1\left(\frac{4N_0}{||g_{i,u}(s-p)||^2}\right)}{E_0\left(\frac{4N_0}{||g_{i,u}(s-p)||^2}\right)}. \tag{5.13}$$

It follows from the above analysis that, as far as minimizing the asymptotic BEP is concerned, the optimal rotation matrix $\boldsymbol{G}$ is the one that minimizes $\kappa(s,p,\boldsymbol{G})$. In the next section, the optimal choice of $\boldsymbol{G}$ is addressed.

5.3 Optimal Rotation Matrix $\boldsymbol{G}$

Applying the Cauchy inequality gives the following lower bound on $\kappa(s,p,\boldsymbol{G})$ in (5.13):

$$[\kappa(s,p,\boldsymbol{G})]^N \geq \prod_{u=1}^{N}\prod_{i=1}^{N}\frac{E_1(x_{i,u})}{E_0(x_{i,u})}, \tag{5.14}$$

where $x_{i,u} = \frac{4N_0}{||g_{i,u}(s-p)||^2}$. It is easy to see that the equality holds when $\{x_{i,u}\}$, or equivalently, $\{||g_{i,u}||\}$ are equal for all $1 \leq i,u \leq N$. This is because $\frac{d}{dx}\left[\frac{E_1(x)}{E_0(x)}\right] > 0$ [73,

5.1.21], which implies that $\frac{E_1(x)}{E_0(x)}$ is an increasing function. Furthermore, from the power constraint $\sum_{i=1}^{N}\sum_{u=1}^{N}||g_{i,u}||^2 = N$ (see (4.2), Chapter 4), one has:

$$\begin{aligned}\sum_{u=1}^{N}\sum_{i=1}^{N}\frac{1}{x_{i,u}} &= \frac{||(s-p)||^2}{4N_0}\sum_{u=1}^{N}\sum_{i=1}^{N}||g_{i,u}||^2 \\ &= \frac{N\cdot||(s-p)||^2}{4N_0} = A = \text{constant}. \end{aligned} \tag{5.15}$$

Denote the right-hand side of (5.14) as $q(\{x_{i,u}\})$. Then the minimization of $\kappa(s,p,\boldsymbol{G})$ now becomes the minimization over $q(\{x_{i,u}\})$ subject to the constraint in (5.15). This optimization problem can be solved using the theorem of majorization [74], which is similar to the method using Jensen's inequality [75]. In particular, one has the following Lemma.

Lemma 5.1: Let $h(y)$ be a log-convex function, i.e., $\log h(y)$ is a convex function. If $y_1,\ldots,y_n$ are positive numbers with $\sum_{i=1}^{n} y_i = n\cdot y$ then:

$$\prod_{i=1}^{n} h(y_i) \geq h(y)^n. \tag{5.16}$$

Proof: Similar to the proof in [75, 3.6.48], due to the fact that $\log h(y)$ is a convex function in the interval $(0,\infty)$, one can apply the Jensen's inequality to obtain:

$$\sum_{i=1}^{n}\log h(y_i) \geq n\log h(y), \tag{5.17}$$

which is equivalent to (5.16).

In order to apply Lemma 5.1 to find the minimum value of $q(\{x_{i,u}\})$ subject to the constraint in (5.15), one needs to show that the function $\frac{E_1(1/y)}{E_0(1/y)}$ is a log-convex of the positive variable y. To this end, the following Lemma is useful.

Lemma 5.2: Let $z(y,t)$ be a log-convex function of y with parameter $t \geq 0$. Also let $v(t)$ be a non-negative function of t for $t \geq 0$. Then the function $h(y) = \int_0^\infty z(y,t)v(t)\mathrm{d}t$ is a log-convex function of y on $(0,\infty)$.

Proof: See Appendix C.

Now, start with

$$\frac{E_1(x)}{E_0(x)} = x\exp(x)\int_x^\infty \frac{\exp(-t)}{t}\mathrm{d}t = x\int_x^\infty \frac{\exp(-(t-x))}{t}\mathrm{d}t. \tag{5.18}$$

Changing the variable $u = t - x$, one obtains:

$$\frac{E_1(x)}{E_0(x)} = \int_0^\infty \frac{x}{x+u}\exp(-u)\mathrm{d}u = \int_0^\infty \frac{1}{1+1/x\cdot u}\exp(-u)\mathrm{d}u. \tag{5.19}$$

Substituting $y = 1/x$ yields:

$$\frac{E_1(1/y)}{E_0(1/y)} = \int_0^\infty \frac{1}{1+y\cdot u}\exp(-u)\mathrm{d}u. \tag{5.20}$$

Therefore, it can be seen from Lemma 5.2 that the function $\frac{E_1(1/y)}{E_0(1/y)}$ of y is a log-convex function, since $1/(1+y\cdot u)$ is a log-convex function of y.

Consider the function $q(\{x_{i,u}\})$ with variables $y_{i,u} = 1/x_{i,u}$. From (5.15), it can be seen that $\sum_{i=1}^N \sum_{u=1}^N y_{i,u} = A$. Then one has:

$$\begin{aligned} q(\{x_{i,u}\}) &= \prod_{u=1}^N \prod_{i=1}^N \frac{E_1(x_{i,u})}{E_0(x_{i,u})} = \prod_{u=1}^N \prod_{i=1}^N \frac{E_1(1/y_{i,u})}{E_0(1/y_{i,u})} \\ &\geq \left[\frac{E_1(\frac{N^2}{A})}{E_0(\frac{N^2}{A})}\right]^{N^2} = \text{constant}, \end{aligned} \tag{5.21}$$

where the inequality follows from Lemma 5.1.

From (5.21), it can be observed that the function $q(\{x_{i,u}\})$ and the parameter $\kappa(s,p,\boldsymbol{G})$ achieve the minimum values if and only if $x_{i,u}$ are equal for all i and u. Equivalently, it can be stated that the optimal rotation matrix $\boldsymbol{G}$ is the matrix with all entries equal in magnitude. This turns out to be the same class of optimal rotation matrix with respect to the asymptotic performance obtained for a conventional Rayleigh fading channel in Chapter 4.

The above analysis is only concerned with the asymptotic performance. Typically, one is also interested in the convergence behavior of the performance with iterations. To this end, one can follow the method presented in Chapter 4 by considering the worst case of the PEP. In [53], it was shown that the three design criteria for the rotation matrix $\boldsymbol{G}$ over a Rayleigh fading channel can be readily applied for the case

of a cascaded Rayleigh fading channel. As a consequence, one can employ the class of rotation matrices in (4.22) over a cascaded Rayleigh fading channel when N is a Euler number, $\phi(P)$, with $P \neq 0 \bmod 4$ or N is a power of 2.

5.4 Performance Comparison Between The Conventional and Cascaded Rayleigh Fading Models

By using the optimal rotation matrix $\boldsymbol{G}$ with all elements equal in magnitude, i.e., $||g_{i,u}||^2 = \frac{1}{N}$ for all i and u, the parameter $\alpha(s, p, \boldsymbol{G})$ in (5.10) for given signal pair s and p whose label differs in only 1 bit at the jth position can be computed as:

$$\alpha(s, p, \boldsymbol{G}) = \left(\frac{E_1 \left(N \cdot z \right)}{E_0 \left(N \cdot z \right)} \right)^N, \tag{5.22}$$

where

$$z = \frac{4N_0 \sin^2 \theta}{||(s-p)||^2}. \tag{5.23}$$

In the case of BICM-ID with SSD over a conventional Rayleigh fading channel, by using an optimal $\boldsymbol{G}$, it can be verified from (4.9) and (4.16) that one can compute the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$, which is now denoted as $f_{\text{Ray}}(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ to avoid any confusion, as:

$$f_{\text{Ray}}(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G}) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[\gamma_{\text{Ray}}(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) \right]^d \mathrm{d}\theta, \tag{5.24}$$

where

$$\gamma_{\text{Ray}}(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{m 2^m} \sum_{s \in \Omega} \sum_{j=1}^{m} \alpha_{\text{Ray}}(s, p, \boldsymbol{G}) \tag{5.25}$$

$$\alpha_{\text{Ray}}(s, p, \boldsymbol{G}) = \left(\frac{N \cdot z}{N \cdot z + 1} \right)^N \tag{5.26}$$

and z is also given as in (5.23). It should be noted that the parameter z in (5.23) is an inverse of $\beta(s, p)$ defined in (4.25).

For a given value of N, the comparison between the asymptotic performances of the systems operating over two fading channel models can be made by using the parameters $\alpha(s, p, \boldsymbol{G})$ and $\alpha_{\text{Ray}}(s, p, \boldsymbol{G})$. From (5.22) and the fact that $E_0(x) =$

$\exp(-x)/x$, one has:

$$\begin{aligned} \alpha(s,p,\boldsymbol{G}) &= \left(Nz\cdot\exp(Nz)\cdot E_1\left(N\cdot z\right)\right)^N \\ &> \left(Nz/(Nz+1)\right)^N = \alpha_{\text{Ray}}(s,p,\boldsymbol{G}). \end{aligned} \tag{5.27}$$

where the inequality follows from [73, 5.1.19] with $1/(x+1) < \exp(x)E_1(x)$. The inequality in (5.27) shows that for a given value of N, the error performance of the system over a cascaded Rayleigh fading channel is always poorer than the performance over a conventional Rayleigh fading channel.

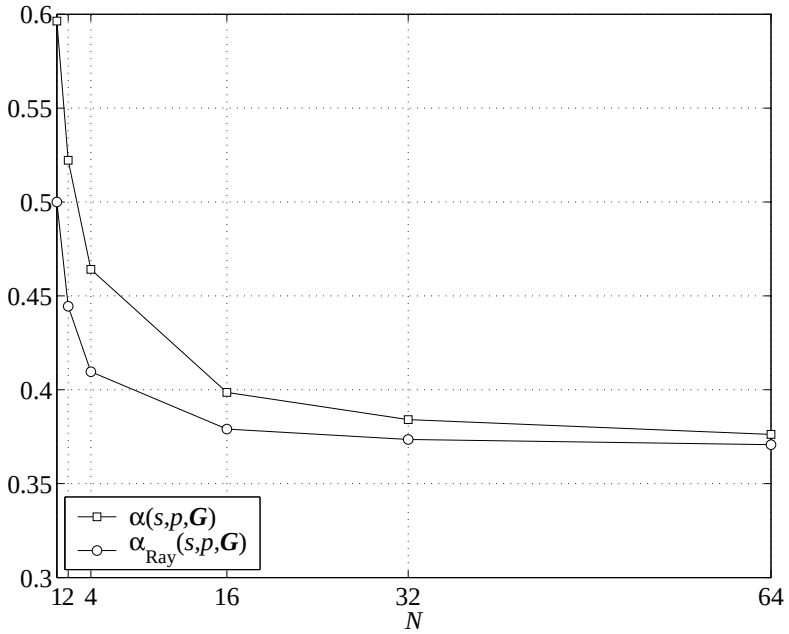


Figure 5.2 Plots of $\alpha(s,p,\boldsymbol{G})$ and $\alpha_{\text{Ray}}(s,p,\boldsymbol{G})$ with the optimal rotation $\boldsymbol{G}$ for different values of N.

Fig. 5.2 plots the two parameters $\alpha(s,p,\boldsymbol{G})$ and $\alpha_{\text{Ray}}(s,p,\boldsymbol{G})$ versus N when z is normalized to be 1. It indicates that when N increases, both $\alpha_{\text{Ray}}(s,p,\boldsymbol{G})$ and $\alpha(s,p,\boldsymbol{G})$ become smaller. This implies that the performance gains in both cases can be increased by implementing SSD in a larger dimension. Furthermore, when N is small, there is a big gap between $\alpha(s,p,\boldsymbol{G})$ and $\alpha_{\text{Ray}}(s,p,\boldsymbol{G})$. Therefore, it is expected that for small N, *cascaded* Rayleigh fading induces much greater performance degradation than Rayleigh fading. This analytical result agrees with those

in [68, 69] for uncoded systems. This performance gap becomes negligible when N is sufficiently large. It means that by employing SSD with sufficiently large dimension, the performance of the system over a cascaded Rayleigh fading channel can closely approach that over a Rayleigh fading channel.

The ultimate performance limit of the system under consideration using SSD is now examined. In Chapter 4, by letting N goes to infinity, it was shown that the parameter $\alpha_{\text{Ray}}(s, p, \boldsymbol{G})$ approaches $\exp(-1/z) = \exp\left(-\frac{||s-p||^2}{4N_0 \sin^2\theta}\right)$. Equivalently, one has:

$$\frac{1}{\pi}\int_0^{\pi/2} \left[\gamma_{\text{Ray}}(\boldsymbol{\Psi}, \xi, \boldsymbol{G})\right]^d \mathrm{d}\theta = \frac{1}{\pi}\int_0^{\pi/2} \left\{\frac{1}{m2^m}\left[\sum_{s\in\Omega}\sum_{j=1}^{m} \exp\left(-\frac{||s-p||^2}{4N_0 \sin^2\theta}\right)\right]\right\}^d \mathrm{d}\theta, \tag{5.28}$$

It can be seen that the right-hand side of (5.28) serves as the tight bound of $f_{\text{AWGN}}(d, \Omega, \xi)$ in (4.23), which is exactly the function $f(\cdot)$ in (2.20) for BICM-ID over an AWGN channel [25]. A similar result can be obtained for the case of BICM-ID over a cascaded Rayleigh fading channel, which is stated in the following Lemma.

Lemma 5.3: The asymptotic error performance of BICM-ID with SSD over a cascaded Rayleigh fading channel approaches that over an AWGN channel.

Proof: Proving this lemma is equivalent to show that $\alpha(s, p, \boldsymbol{G}) = \exp(-1/z)$ when N goes to infinity. This is also equivalent to prove that

$$\log \alpha(s, p, \boldsymbol{G}) = -1/z \tag{5.29}$$

when $N \to \infty$. Multiplying both sizes of (5.29) with z and considering the new variable $X = Nz$, one needs to show that:

$$\lim_{X\to\infty} X \cdot \log\left(X \cdot \exp(X) \cdot E_1(X)\right) = -1. \tag{5.30}$$

When X is sufficiently large, the function $X \cdot \exp(X) \cdot E_1(X)$ can be represented with an asymptotic expression as follows [73, 5.1.51]:

$$X \cdot \exp(X) \cdot E_1(X) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n n!}{X^n}. \tag{5.31}$$

Using the L'Hospital rule, it then follows that:

$$\begin{aligned}
& \lim_{X\to\infty} X \cdot \log\left(X \cdot \exp(X) \cdot E_1(X)\right) \\
= & \lim_{X\to\infty} \frac{\log\left(1+\sum_{n=1}^{\infty} \frac{(-1)^n n!}{X^n}\right)}{1/X} \\
= & \lim_{X\to\infty} \frac{(-X^2)}{1+\sum_{n=1}^{\infty} \frac{(-1)^n n!}{X^n}} \cdot \left(\sum_{n=1}^{\infty} \frac{(-1)^{n+1} \cdot n \cdot n!}{X^{n+1}}\right) \\
= & \lim_{X\to\infty} -\left(1+\sum_{n=2}^{\infty} \frac{(-1)^n \cdot n \cdot n!}{X^{n-1}}\right) = -1
\end{aligned} \tag{5.32}$$

which completes the proof.

5.5 Illustrative Results

For simplicity, the section only concentrates on the case of QPSK modulation with two-dimensional Gray mapping ξ. A rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ is applied as the outer channel code, along with a random interleaver of length 1920 coded bits. To compute the union bound from (2.20), as common, the first 20 Hamming distances of the convolutional code are retained. Unless otherwise stated, the optimal matrix $\boldsymbol{G}$ is chosen from (4.22) with different values of $N = 2^r$.

Fig. 5.3 shows the error performance of the systems without SSD (and 1 iteration) and with SSD (and 1 and 4 iterations) over a cascaded Rayleigh channel, when $N = 2$ and the optimal MAP demodulator is used at the receiver. It is observed that no performance improvement with iterative processing when Gray mapping is used for the system without SSD. The analytical error bounds are also provided for comparison. It can be seen that in case of the system with SSD, the derived error bound is very tight. For the system without SSD, the bound under-estimates the simulation result. This is because the asymptotic union bound is tight only at low levels of BER. Though not shown here, it was observed that the bound for the system without SSD is tight at a higher SNR, corresponding to the BER level of around 10^{-3}. As can be seen from Fig. 5.3, a significant performance gain can be achieved by using the SSD

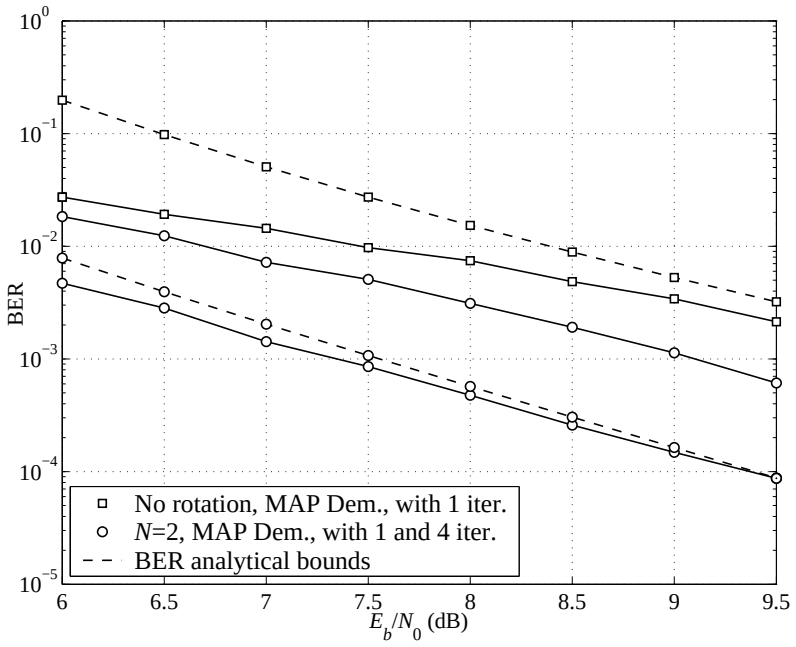


Figure 5.3 Performance of BICM-ID systems without SSD and with SSD when $N = 2$ over a cascaded Rayleigh fading channel.

technique with $N = 2$ compared to the system without SSD. By comparing the tight error bounds over a wider range of SNR, it is observed that the coding gain is about 4dB at the BER of 10^{-4}.

Fig. 5.4 demonstrates the role of rotation matrix $\boldsymbol{G}$ with 2-D constellation ($N = 2$), where the BER performances after 1 and 4 iterations are compared for various choices of $\boldsymbol{G}$. Besides $\boldsymbol{G}_2$ in (4.22), the class of real rotation $\boldsymbol{G}_\theta$ in (4.33) is also considered. More specifically, two cases of $\boldsymbol{G}_\theta$ with $\theta = \pi/4$ and $\theta = \arctan(2)/2$ are examined. The two rotation matrices are denoted by $\boldsymbol{G}_{\pi/4}$ and $\boldsymbol{G}_{\theta,2}$, respectively. Recall that $\boldsymbol{G}_{\theta,2}$ is the optimal rotation matrix for uncoded systems in the sense that it maximizes the modulation diversity and the minimum product distance [17,46,60,61]. However, its entries are not equal in magnitude. On the other hand, it can be verified that the rotation $\boldsymbol{G}_{\pi/4}$ does not achieve the full modulation diversity, but this rotation gives the best asymptotic error performance discussed earlier. In Fig. 5.4, the error bounds for all systems are also provided. The performance with 1 iteration of the system without SSD is plotted as a reference.

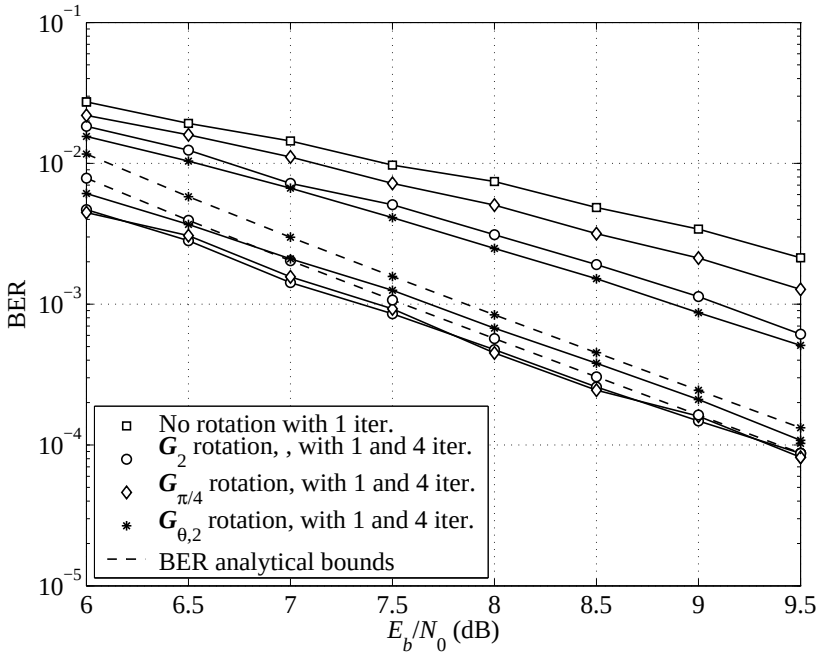


Figure 5.4 Performance of BICM-ID systems with SSD when $N = 2$ over a cascaded Rayleigh fading channel employing various rotation matrices.

As can be seen from Fig. 5.4, the error performances after 4 iterations with $\boldsymbol{G}_2$ and $\boldsymbol{G}_{\pi/4}$ are almost the same and converge to the same error bound, which confirms the analytical results. Furthermore, these performances are better than the error performance after 4 iterations[2] and the error bound of the system employing $\boldsymbol{G}_{\theta,2}$. By plotting the error bounds over a wider range of SNR, it is observed that a gain of 0.3dB is achieved at the BER level of 10^{-4}. For demonstration purpose, thereafter, the rotation matrix $\boldsymbol{G}_N$ in (4.22) is always chosen.

As discused in Chapter 3, the complexity of the MAP demodulator increases exponentially with N, which makes the receiver implementation impractical for large N. Therefore, one can use the suboptimal low-complexity MMSE demodulator as an attractive alternative.

Fig. 5.5 shows the BER performance with 1 and 4 iterations for $N = 2$ over conventional and cascaded Rayleigh fading channels. The MMSE demodulator is

[2]It was observed that additional iterations do not improve the error performance.

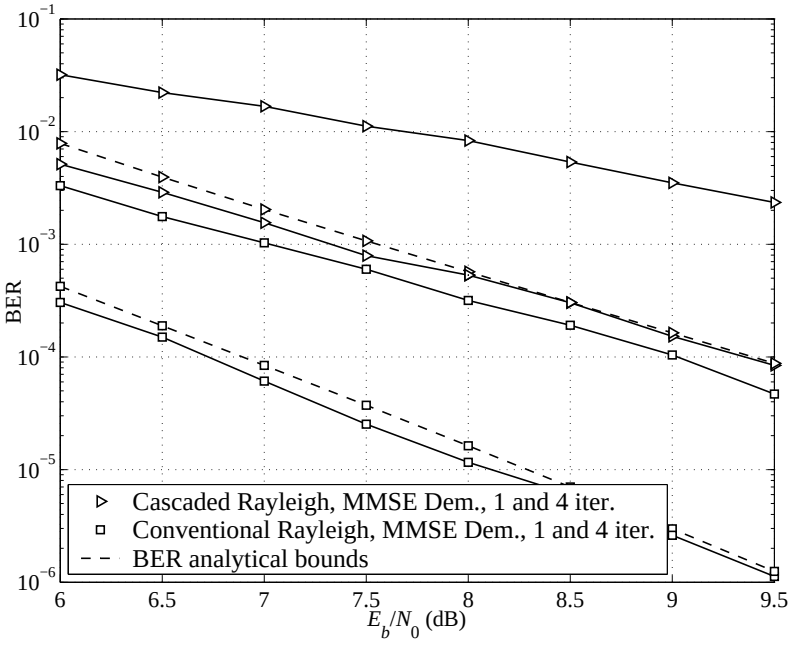


Figure 5.5 Performance of BICM-ID systems with SSD and $N = 2$ over different fading channels.

employed for both systems. It can be observed that the BER performances converge to the error bounds in both cases. Fig. 5.5 also confirms that, for small values of N, the cascaded Rayleigh fading induces a severe performance degradation as compared to Rayleigh fading at any number of iterations.

Finally, to see the benefit of SSD at a sufficiently large value of N, Fig. 5.6 presents the performances after 1 and 8 iterations of the systems over the two fading channel models for $N = 32$ and with the MMSE demodulator. The analytical error bounds are also presented to show the tightness of the bounds. Furthermore, both the analytical bound and the simulation result of system performance with MAP demodulator over an AWGN channel [25] are provided to serve as performance benchmarks for the fading channels. Since Gray mapping is used, the error performance over an AWGN channel is obtained with only 1 iteration. It can be observed that with $N = 32$, the error rates after 8 iterations for both fading models tightly converges to the error bounds. In the case of cascaded Rayleigh fading, the convergence is at higher SNR. This is because the error bounds are derived with the assumption of error-

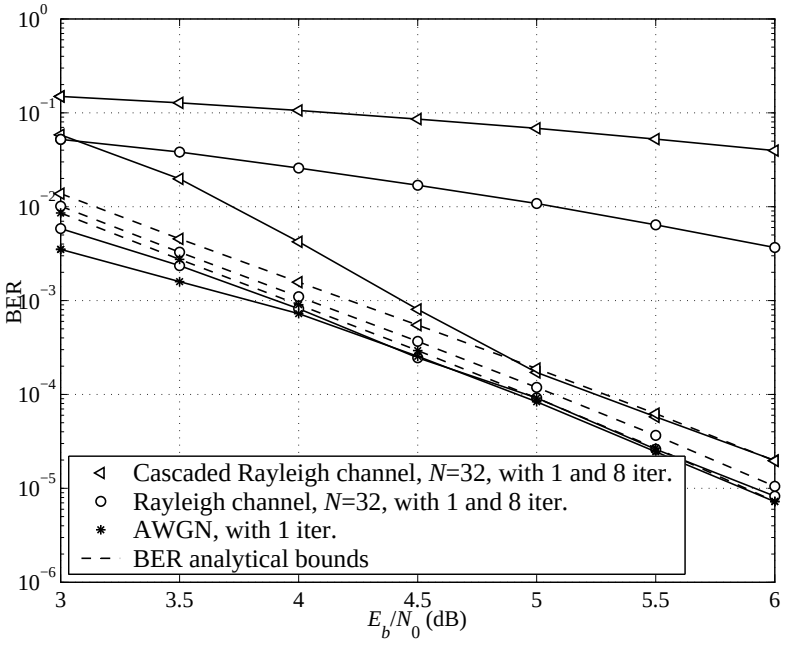


Figure 5.6 Performance of BICM-ID systems with SSD and $N = 32$ over different fading channels. For comparison, the performance of BICM-ID over an AWGN channel is also plotted.

free feedback. At low SNR, the iterative processing cannot work perfectly and the performance is also degraded, which makes the BER worse than the error bound. A faster convergence behavior can be observed at lower SNR by using a longer interleaver and/or running more iterations. When convergence starts to happen, as assumed by the analysis, there is only a small gap between the error performances of the two systems after 8 iterations and they both closely approach the performance of BICM-ID over an AWGN channel. For example, at the BER level of 10^{-4}, the gaps between the error performance over an AWGN channel and those over conventional and cascaded Rayleigh fading channels are only 0.12dB and 0.3dB, respectively.

6. Bit-Interleaved Coded OFDM with Signal Space Diversity[1]

Up to this point, various BICM-ID frameworks have been studied, but only for a *frequency non-selective* fading channel. As discussed in Chapter 2, over a broadband wireless channel, the channel delay spread T_d becomes an increasingly large multiple of the symbol duration T. This results in a *frequency selective* fading channel with severe inter-symbol interference (ISI). Single carrier modulation with equalization is an effective solution to deal with the ISI and to exploit the frequency diversity over *frequency selective* channels. However, the drawback of the single carrier architecture is in the receiver equalizer, which is sometime too complex to implement. On the other hand, as introduced in Chapter 2, by using orthogonal frequency division multiplexing (OFDM), the broadband *frequency selective* fading channel is transformed into a set of correlated *frequency non-selective* (flat) fading subchannels, one for each subcarriers, therefore significantly reducing the receiver complexity in both equalization and detection.

In OFDM, since the bandwidth occupied by each subcarrier is selected to be smaller than the coherence bandwidth B_h of the channel (in order to yield the flat fading effect over each subcarrier), or equivalently, the number of subcarriers is larger than the number of channel taps, the channel coefficients $\{H_n\}$ of subcarriers are generally correlated [12]. In particular, given a bandwidth B and the length of ISI L, there are approximately $N/2L$ neighboring subcarriers whose channel coefficients are heavily correlated, where N is the number of subcarriers. At a given subcarrier

[1] A part of this chapter was presented in [76–79]

n, if the subchannel is in deep fade, it is likely that the information transmitted in subchannel nth and its neighboring subchannels cannot be reconstructed properly at the receiver. In order to provide more reliable transmission and fully exploit the frequency diversity order of L, coding across subcarriers is necessary.

As studied in the previous chapters, signal space diversity (SSD) can provide performance improvement over *frequency non-selective* fading channels by spreading information over time. This idea can be readily applied to OFDM systems, where SSD is used to spread information over subcarriers in frequency domain. For instance, applications of signal space diversity have been considered in a number of uncoded orthogonal frequency-division multiplexing (OFDM) systems [80–85]. The work in [83] studied uncoded OFDM with linear block precoding based on the mean cutoff rate. The investigation of linear constellation precoding applied to uncoded OFDM in [82] indicates that signal space diversity with a specifically selected group of subcarriers can achieve the same maximum frequency diversity and coding gains as with spreading across all the available subcarriers. This technique is then applied to various uncoded OFDM systems [80, 81, 84, 85] and surprisingly, without the explicit solution for subcarrier grouping. For example, it was not clear if the specific subcarrier grouping suggested in [82] is the unique and general solution. Recently, in a connection with an investigation presented in this chapter, this problem has been partially addressed for an uncoded OFDM system in [77].

The application of SSD has also been carried out for coded OFDM systems in [86–88]. In particular, the work in [87] briefly studied the effect of signal space diversity in coded OFDM systems with a rather unrealistic assumption of statistically independent fading over subchannels. A more general OFDM system that employs joint channel coding and signal space diversity (via linear precoding) has also been presented in [89].

This chapter considers the application of signal space diversity in OFDM systems using BICM-ID over *frequency selective* Rayleigh fading channels, where fading across subcarrier is assumed to be correlated. Different from the work in [89], here the

problem of subcarrier grouping and rotation matrix design are jointly addressed to optimize the error performance while keeping the complexity at minimum.

The chapter starts with a general structure of OFDM systems using BICM-ID with signal space diversity and their corresponding system model. A general tight bound on the asymptotic error performance for precoding over all N subcarriers is derived next. Based on this bound, design criterion to achieve the best asymptotic performance with signal space diversity is established for a given constellation and mapping rule. It is then shown that signal space diversity can just be implemented over a set of subcarrier groups of at least L elements for the same optimum performance as in the case of precoding over all N subcarriers, where L is the number of channel taps. Subsequently, the joint optimum subcarrier grouping and rotation matrix design for signal space diversity is derived. Finally, illustrative analytical and simulation results are presented and discussed.

Some notations used in this chapter are as follows: All the vectors are row-wise. The bold lower letters indicate vectors, while the bold capital letters are used for matrices. The superscripts $\top$ and $\dagger$ denote transpose and conjugate transpose, respectively; $\mathrm{diag}(H_1, \ldots, H_N)$ is an $N \times N$ diagonal matrix with diagonal elements H_i, $i = 1, \ldots, N$; $\det(\cdot)$ denotes the determinant of a square matrix; and $\mathbf{0}$ stands for the all-zero vector.

6.1 System Model

Figure 6.1 illustrates the application of BICM-ID in an OFDM system that also employs signal space diversity, which is similar to the system model studied in [89]. As similar to a general BICM-ID system, the information sequence $\underline{\boldsymbol{u}}$ is first encoded into a coded sequence $\underline{\boldsymbol{c}}$. The coded sequence $\underline{\boldsymbol{c}}$ is then interleaved by a bit-wise interleaver to become the interleaved sequence $\underline{\boldsymbol{v}}$. Consider OFDM system with N subcarriers. Each group of Nm interleaved coded bits are mapped to an original OFDM symbol $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]$ with the mapping rule ξ, where $s_i \in \Omega$, $\forall i$, with Ω is a real two-dimensional constellation of size 2^m. Clearly, the symbol $\boldsymbol{s}$ can be

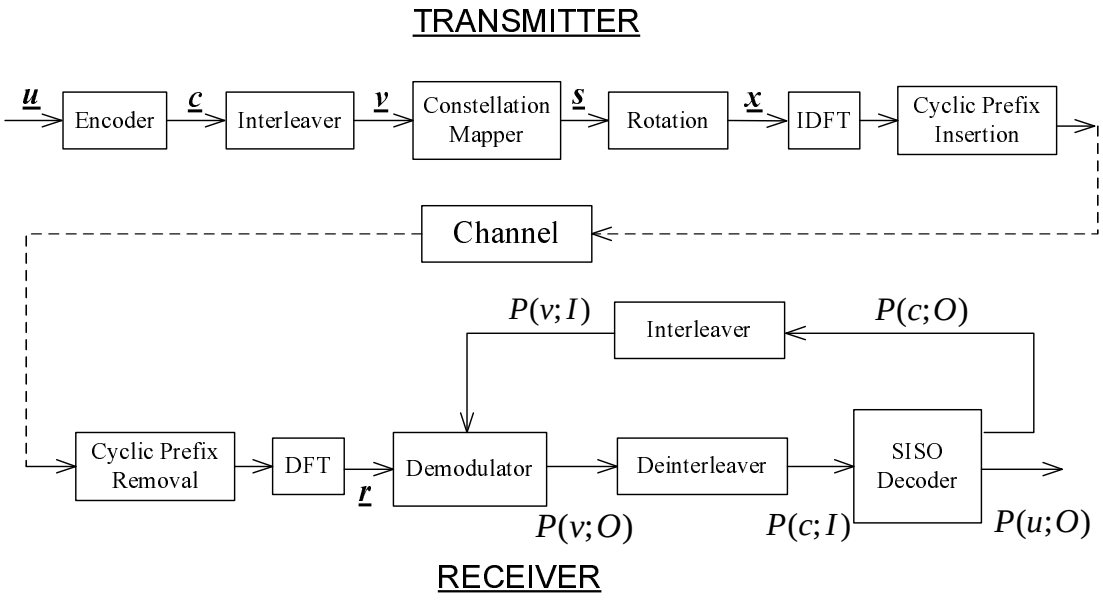


Figure 6.1 Block diagram of a BICM-ID for OFDM with signal space diversity.

considered as a signal point in a complex N-dimensional (N-D) constellation $\mathbf{\Psi}$, which includes in total of 2^{Nm} signal points. It can be observed that the above constellation $\mathbf{\Psi}$ is similar to the one considered in Chapter 4, except the fact that here, $\mathbf{\Psi}$ spans over subcarriers.

For simplicity, assume the mapping rule ξ is employed independently in each component s_i, which is the same with the assumption made in Chapter 4. The use of multi-dimensional mapping technique shall be further evaluated in Chapter 7. The rotated sequence $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$ is obtained as:

$$\boldsymbol{x}^\top = \boldsymbol{G}\boldsymbol{s}^\top, \tag{6.1}$$

where $\boldsymbol{G}$ can be any $N \times N$ matrix with complex elements $g_{i,u}$ which satisfies the power constraint $\sum_{i=1}^{N}\sum_{u=1}^{N} \|g_{i,u}\|^2 = N$.

In general, the number of carrier N can be very large, which leads to very high decoding complexity. The very attractive solution known as subchannel grouping, was proposed in [82] for uncoded OFDM systems to reduce the complexity at the receiver. The basic idea of this approach is to divide all the N carriers into nonintersecting

subsets and then apply the rotation matrix with a much smaller size to each subset. Following the notations in [82], denote the set of N indices of all the subcarriers as $\mathcal{I} := \{1, \ldots, N\}$. Assume that $N = VF$. Then N carriers can be divided into V nonintersecting subsets $\{I_1, \ldots, I_V\}$. Each subset I_v, $1 \leq v \leq V$, includes F carrier indices $I_v = (p_{v,1}, \ldots, p_{v,F})$. A rotation matrix $\boldsymbol{\Theta}$ with size $F \times F$ is then applied to each group I_v to obtain a sequence $\boldsymbol{x}_v$ of size F. V sequences $\{\boldsymbol{x}_v\}$ form the rotated sequence $\boldsymbol{x}$. Clearly, the subcarrier grouping approach is a special case of the general model in (6.1) with the following equivalent rotation matrix $\boldsymbol{G}$:

$$\boldsymbol{G} = \sum_{v=1}^{V} \boldsymbol{\pi}_v^T \boldsymbol{\Theta} \boldsymbol{\pi}_v, \tag{6.2}$$

where $\boldsymbol{\pi}_v$ is a $F \times N$ permutation matrix constructed from rows $\{p_{v,f}\}$, $1 \leq f \leq F$, of $\boldsymbol{I}_N$.

By employing a specific subcarrier grouping with F equals the number of channel taps L, it was shown in [82] that both diversity and coding gains (but not BER) of uncoded OFDM can be preserved as in the case of $F = N$. This result is simply due to the fact that for OFDM, there are totally L independent paths, giving at most a frequency diversity order of L.

The rotated sequence $\boldsymbol{x}$ is passed to the inverse discrete Fourier transform (IDFT) block and then inserted with a cyclic prefix of length L_p. Assume that DFT processing[2] and cyclic prefix removal are properly carried out at the receiver with coherent detection. Since OFDM converts the broadband *frequency selective* fading channel into N *frequency non-selective* subchannels, the received signal in the frequency domain can be presented as follows:

$$\boldsymbol{r}^\top = \boldsymbol{H}\boldsymbol{G}\boldsymbol{s}^\top + \boldsymbol{n}^\top. \tag{6.3}$$

Here, each entry n_i of $\boldsymbol{n} = [n_1, \ldots, n_N]$ is a complex white Gaussian noise $\mathcal{CN}(0, N_0)$. The matrix $\boldsymbol{H} = \text{diag}(H_1, \ldots, H_N)$ contains the correlated fading coefficients in its

[2]Recall that when N is a power of 2, IFFT and FFT can be efficiently implemented.

diagonal, where:

$$H_i = \sum_{l=1}^{L} h_l \exp(-j2\pi i(l-1)/N), \quad 1 \le i \le N. \tag{6.4}$$

In (6.4), the channel vector $\boldsymbol{h} = [h_1, \ldots, h_L]$ contains the channel gains of all the L taps, where each h_l, $l = 1, \ldots, L$, is modeled as a circularly symmetric complex Gaussian random variable. It is assumed that the channel gains remain constant within one OFDM symbol and change independently from one OFDM symbol to the next. This is a reasonable assumption when a sufficiently long interleaver is used to break the correlation of the channel in time. Furthermore, within each OFDM symbol, the general case of correlated channel taps is considered by assuming that the channel vector $\boldsymbol{h}$ has a full-rank correlation matrix $\boldsymbol{R_h} = E(\boldsymbol{h}^\dagger \boldsymbol{h})$ [80–82]. Note that the model of independent and identically distributed (i.i.d.) channel taps considered in [12, 86] with $\boldsymbol{R_h} = \frac{1}{L}\boldsymbol{I}_L$ and the model with an exponential power profile of the channel taps in [90] are just special cases of the above general model. For convenience, the coefficient H_i is rewritten as follows:

$$H_i = \boldsymbol{h}\boldsymbol{w}_i^\top, \tag{6.5}$$

where

$$\boldsymbol{w}_i = \left[\exp\left(-j\frac{2\pi i 0}{N}\right), \ldots, \exp\left(-j\frac{2\pi i(L-1)}{N}\right)\right] \tag{6.6}$$

for $1 \le i \le N$.

As shown in Fig. 6.1, the receiver of the system includes the soft-input soft-output (SISO) demodulator and the soft-input soft-output channel decoder for the convolutional code, which is the same with the pervious BICM-ID systems. As similar to Chapter 4, in the general case of the rotation matrix $\boldsymbol{G}$, the complexity of the optimal MAP demodulator grows exponentially with the number of coded bits per OFDM symbols Nm, which becomes quickly intractable for medium to large values of Nm. When subchannel grouping approach is used, the MAP demodulator can be applied for each group of subcarriers I_v, which significantly reduces the complexity, especially when $F \ll N$. The suboptimal low-complexity yet effective method pro-

posed in Chapter 4 using the MMSE receiver and the sigma mapping, or the Gaussian approximations proposed in [51] can be also attractive alternatives.

6.2 Performance Evaluation

In this section, given a constellation $\mathbf{\Psi}$ and the mapping rule ξ, the *asymptotic* bit error probability (BEP) of OFDM systems with BICM-ID and signal space diversity is investigated for the general $N \times N$ rotation matrix $\boldsymbol{G}$. The evaluation of such asymptotic BEP performance shall be carried out based on the assumption of error-free feedback from the decoder to the demodulator as normally done in the pervious chapters with the perfect interleaver (i.e., an interleaver of infinite length). For convenience, the union bound of the BEP performance for a rate-k_c/n_c convolutional code, a complex N-dim constellation $\mathbf{\Psi}$ and a mapping rule ξ shown in (2.20) is re-written here as:

$$P_b \leq \frac{1}{k_c} \sum_{d=d_H}^{\infty} c_d f(d, \mathbf{\Psi}, \xi, \boldsymbol{G}). \tag{6.7}$$

Again, it is clear that the key point to evaluate P_b is to compute the function $f(d, \mathbf{\Psi}, \xi, \boldsymbol{G})$.

Using the same notations as in Chapter 4, let $\underline{\boldsymbol{c}}$ and $\underline{\breve{\boldsymbol{c}}}$ denote the input and estimate sequences, respectively, with the Hamming distance d between them. These binary sequences correspond to the sequences $\underline{\boldsymbol{s}}$ and $\underline{\breve{\boldsymbol{s}}}$, whose elements are OFDM symbols in $\mathbf{\Psi}$. Without loss of generality, assume that $\underline{\boldsymbol{c}}$ and $\underline{\breve{\boldsymbol{c}}}$ differ in the first d consecutive bits. Hence, $\underline{\boldsymbol{s}}$ and $\underline{\breve{\boldsymbol{s}}}$ can be redefined as sequences of d OFDM symbols as $\underline{\boldsymbol{s}} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d]$ and $\underline{\breve{\boldsymbol{s}}} = [\breve{\boldsymbol{s}}_1, \ldots, \breve{\boldsymbol{s}}_d]$. Also let $\underline{\boldsymbol{H}} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_d]$, where $\boldsymbol{H}_e = \mathrm{diag}(H_{e,1}, \ldots, H_{e,N})$, $1 \leq e \leq d$, represents the path gains that affect the transmitted symbol $\boldsymbol{s}_e$. More specifically, for each channel realization e, one has:

$$H_{e,i} = \sum_{l=1}^{L} h_{e,l} \exp\left(-\frac{j2\pi i(l-1)}{N}\right) = \boldsymbol{h}_e \boldsymbol{w}_i^{\top}, \tag{6.8}$$

where $\boldsymbol{h}_e = [h_{e,1}, \ldots, h_{e,L}]$ are tap gains at channel realization e and $\boldsymbol{w}_i$ is given in (6.6).

The two OFDM symbols $\boldsymbol{s}_e$ and $\breve{\boldsymbol{s}}_e$ correspond to the two rotated OFDM symbols $\boldsymbol{x}_e$ and $\breve{\boldsymbol{x}}_e$, i.e., $\boldsymbol{x}_e^\top = \boldsymbol{G}\boldsymbol{s}_e^\top$ and $\breve{\boldsymbol{x}}_e^\top = \boldsymbol{G}\breve{\boldsymbol{s}}_e^\top$. Then the PEP conditioned on $\underline{\underline{\boldsymbol{H}}}$ can be computed as follows:

$$P(\underline{\boldsymbol{s}} \to \underline{\breve{\boldsymbol{s}}}|\underline{\underline{\boldsymbol{H}}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d} d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e)}\right), \tag{6.9}$$

where $d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e)$ is the squared Euclidean distance between the two received signals corresponding to $\boldsymbol{x}_e$ and $\breve{\boldsymbol{x}}_e$ conditioned on $\boldsymbol{H}_e$ and in the absence of the AWGN. This distance is given by

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \left|\left|\boldsymbol{H}_e\boldsymbol{G}(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right|\right|^2 = \sum_{i=1}^{N}\left\|H_{e,n}\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right\|^2, \tag{6.10}$$

where $\boldsymbol{g}_i$ is the ith row of $\boldsymbol{G}$. Clearly, $\{H_{e,i}, 1 \le i \le N\}$ are correlated. Following the same procedure in [86, 91], substitute (6.8) into (6.10), one has:

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \sum_{i=1}^{N}\left\|\boldsymbol{h}_e\boldsymbol{w}_i^\top\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top\right\|^2. \tag{6.11}$$

Let $\boldsymbol{t}_{e,i} = [\boldsymbol{w}_i^\top\boldsymbol{g}_i(\boldsymbol{s}_e - \breve{\boldsymbol{s}}_e)^\top]^\top$, it then follows that:

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \sum_{i=1}^{N}\left\|\boldsymbol{h}_e\boldsymbol{t}_{e,i}^\top\right\|^2 = \boldsymbol{h}_e\boldsymbol{Q}_e\boldsymbol{h}_e^\dagger, \tag{6.12}$$

where $\boldsymbol{Q}_e = \sum_{i=1}^{N}(\boldsymbol{t}_{e,i}^\top)(\boldsymbol{t}_{e,i}^\top)^\dagger$. Furthermore, let the square root of the correlation matrix $\boldsymbol{R}_{\boldsymbol{h}}$ be $\boldsymbol{B}_{\boldsymbol{h}}$. Then the channel vector $\boldsymbol{h}_e$ can be whitened as follows [80–82]:

$$\bar{\boldsymbol{h}}_e^\top = \boldsymbol{B}_{\boldsymbol{h}}^{-1}\boldsymbol{h}_e^\top \tag{6.13}$$

where the elements $\{\bar{h}_{e,l}\}$ of $\bar{\boldsymbol{h}}_e$ are i.i.d. complex zero-mean Gaussian random variables with unit variance, i.e., $\mathcal{CN}(0, 1)$. It then follows from (6.12) that:

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \bar{\boldsymbol{h}}_e(\boldsymbol{B}_{\boldsymbol{h}}^\top)\boldsymbol{Q}_e(\boldsymbol{B}_{\boldsymbol{h}}^\top)^\dagger\bar{\boldsymbol{h}}_e^\dagger = \bar{\boldsymbol{h}}_e\bar{\boldsymbol{Q}}_e\bar{\boldsymbol{h}}_e^\dagger, \tag{6.14}$$

where $\bar{\boldsymbol{Q}}_e = (\boldsymbol{B}_{\boldsymbol{h}}^\top)\boldsymbol{Q}_e(\boldsymbol{B}_{\boldsymbol{h}}^\top)^\dagger$. Since $\bar{\boldsymbol{Q}}_e$ is a Hermitian matrix, it can be decomposed with the eigenvalue decomposition as:

$$\bar{\boldsymbol{Q}}_e = \bar{\boldsymbol{P}}_e^\dagger\bar{\boldsymbol{\Lambda}}_e\bar{\boldsymbol{P}}_e, \tag{6.15}$$

where $\bar{\boldsymbol{P}}_e$ is unitary and $\bar{\boldsymbol{\Lambda}}_e$ is a diagonal matrix whose diagonal elements $\{\bar{\lambda}_{e,l}\}$ are the eigenvalues of $\bar{\boldsymbol{Q}}_e$. Then one has:

$$d^2(\boldsymbol{x}_e, \breve{\boldsymbol{x}}_e|\boldsymbol{H}_e) = \left(\bar{\boldsymbol{h}}_e \bar{\boldsymbol{P}}_e^{\dagger}\right) \bar{\boldsymbol{\Lambda}}_e \left(\bar{\boldsymbol{P}}_e \bar{\boldsymbol{h}}_e^{\dagger}\right) = \hat{\boldsymbol{h}}_e \bar{\boldsymbol{\Lambda}}_e \hat{\boldsymbol{h}}_e^{\dagger} = \sum_{l=1}^{L} \bar{\lambda}_{e,l} \|\hat{h}_{e,l}\|^2. \tag{6.16}$$

Clearly, the elements $\{\hat{h}_{e,l}\}$ of $\hat{\boldsymbol{h}}_e$ are also i.i.d. complex Gaussian random variables with zero-mean and unit variance. Substituting (6.16) into (6.9), the PEP conditioned on $\underline{\boldsymbol{H}}$ is then given by:

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}|\underline{\boldsymbol{H}}) = P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}|\hat{\underline{\boldsymbol{h}}}) = Q\left(\sqrt{\frac{1}{2N_0} \sum_{e=1}^{d} \sum_{l=1}^{L} \bar{\lambda}_{e,l} \|\hat{h}_{e,l}\|^2}\right). \tag{6.17}$$

By invoking the Gaussian probability integral $Q(\gamma) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\gamma^2}{2\sin^2\theta}\right) \mathrm{d}\theta$ and averaging (6.17) over $\hat{\underline{\boldsymbol{h}}}$ gives

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}) = \frac{1}{\pi} \int_0^{\pi/2} \mathcal{E}_{\hat{\underline{\boldsymbol{h}}}} \left\{ \prod_{e=1}^{d} \exp\left(-\frac{\sum_{l=1}^{L} \bar{\lambda}_{e,l} \|\hat{h}_{e,l}\|^2}{4N_0 \sin^2\theta}\right) \right\} \mathrm{d}\theta. \tag{6.18}$$

Furthermore, using the equality $\mathcal{E}_g\{\exp(-\gamma g^2)\} = 1/(1+\gamma)$, where g is a Rayleigh random variable with $\mathcal{E}_g(g^2) = 1$, one has:

$$P(\underline{\boldsymbol{s}} \to \breve{\underline{\boldsymbol{s}}}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\prod_{e=1}^{d} \Delta_e \right) \mathrm{d}\theta, \tag{6.19}$$

where

$$\Delta_e = \prod_{l=1}^{L} \left(1 + \frac{1}{4N_0} \frac{\bar{\lambda}_{e,l}}{\sin^2\theta}\right)^{-1}. \tag{6.20}$$

Since it is assumed that the channel realization e changes independently, the variables $\{\Delta_e\}$ can be considered as i.i.d. random variables. Furthermore, owning to the success of decoding step as normally seen in the analysis of BICM-ID systems, the assumption of error-free feedback from the decoder to the demodulator can be made in order to analyze the asymptotic error performance. This assumption implies that one needs to consider only the pairs of OFDM symbols $\boldsymbol{s}_e$ and $\breve{\boldsymbol{s}}_e$ whose labels differ in only 1 bit. Then the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be computed by averaging

over the constellation $\boldsymbol{\Psi}$ as follows:

$$f(d,\boldsymbol{\Psi},\xi,\boldsymbol{G}) \leq \frac{1}{\pi}\int_0^{\pi/2}\left[\underbrace{\mathcal{E}_{\boldsymbol{s},\boldsymbol{p}}\left\{\prod_{l=1}^{L}\left(1+\frac{\bar{\lambda}_l(\boldsymbol{s},\boldsymbol{p})}{4N_0\sin^2\theta}\right)^{-1}\right\}}_{\gamma(\boldsymbol{\Psi},\xi,\boldsymbol{G})}\right]^d \mathrm{d}\theta, \tag{6.21}$$

where the expectation is over all pairs of OFDM symbols $\boldsymbol{s}$ and $\boldsymbol{p}$ in $\boldsymbol{\Psi}$ whose labels differ in only 1 bit. The values $\{\bar{\lambda}_l(\boldsymbol{s},\boldsymbol{p})\}$ are the eigenvalues of $\bar{\boldsymbol{Q}}(\boldsymbol{s},\boldsymbol{p})$ where

$$\bar{\boldsymbol{Q}}(\boldsymbol{s},\boldsymbol{p}) = (\boldsymbol{B}_{\boldsymbol{h}}^{\top})\boldsymbol{Q}(\boldsymbol{s},\boldsymbol{p})(\boldsymbol{B}_{\boldsymbol{h}}^{\top})^{\dagger}, \tag{6.22}$$

and

$$\boldsymbol{Q}(\boldsymbol{s},\boldsymbol{p}) = \sum_{i=1}^{N}(\boldsymbol{t}_i^{\top}(\boldsymbol{s},\boldsymbol{p}))(\boldsymbol{t}_i^{\top}(\boldsymbol{s},\boldsymbol{p}))^{\dagger} \tag{6.23}$$

with

$$\boldsymbol{t}_i(\boldsymbol{s},\boldsymbol{p}) = [\boldsymbol{w}_i^{\top}\boldsymbol{g}_i(\boldsymbol{s}-\boldsymbol{p})^{\top}]^{\top}. \tag{6.24}$$

A straightforward way to compute the expectation operation in (6.21) is as follows:

$$\gamma(\boldsymbol{\Psi},\xi,\boldsymbol{G}) = \frac{1}{Nm2^{Nm}}\sum_{\boldsymbol{s}\in\boldsymbol{\Psi}}\sum_{k=1}^{Nm}\left[\prod_{l=1}^{L}\left(1+\frac{\bar{\lambda}_l(\boldsymbol{s},\boldsymbol{p})}{4N_0\sin^2\theta}\right)^{-1}\right], \tag{6.25}$$

where $\boldsymbol{p}$ is the symbol in $\boldsymbol{\Psi}$ whose label differs in only one bit at the position k compared to the label of $\boldsymbol{s}$.

For a large value of Nm, the computation of (6.25) becomes intractable due to the huge number of OFDM symbols in $\boldsymbol{\Psi}$. For simplicity, consider the mapping $\boldsymbol{\xi}$ that is implemented independently and identically for each signal component in the two-dimensional constellation, denoted by Ω. First, by interchanging the summations in (6.25) one can write $\gamma(\boldsymbol{\Psi},\xi)$ as a sum of N terms as:

$$\gamma(\boldsymbol{\Psi},\xi,\boldsymbol{G}) = \frac{1}{N}\sum_{u=1}^{N}\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G}), \tag{6.26}$$

where

$$\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G}) = \frac{1}{m2^{Nm}}\sum_{k=(u-1)m+1}^{um}\sum_{\boldsymbol{s}\in\boldsymbol{\Psi}}\prod_{l=1}^{L}\left(1+\frac{\bar{\lambda}_l(\boldsymbol{s},\boldsymbol{p})}{4N_0\sin^2\theta}\right)^{-1} \tag{6.27}$$

is essentially obtained by averaging over all pairs of $\boldsymbol{s}$ and $\boldsymbol{p}$ whose labels differ in 1 bit at positions from $k = (u-1)m+1$ to $k = um$. Note that the inner sum in (6.27) is taken over $m2^{Nm}$ possible symbols $\boldsymbol{s}$, which might still appear computationally impossible when N and m are large. Fortunately, by independently mapping for each component of OFDM symbols, it is easy to verify that there is only one distinct component between $\boldsymbol{s}$ and $\boldsymbol{p}$ at the uth position. Hence, for a given u, one has:

$$\|\boldsymbol{g}_i(\boldsymbol{s}-\boldsymbol{p})^\top\|^2 = \|g_{i,u}(s_u - p_u)\|^2, \tag{6.28}$$

where s_u and p_u are the uth components of $\boldsymbol{s}$ and $\boldsymbol{p}$, respectively. By adding the index u, the matrix $\boldsymbol{Q}_u(\boldsymbol{s},\boldsymbol{p})$ in (6.23) is then computed as:

$$\begin{aligned}\boldsymbol{Q}_u(\boldsymbol{s},\boldsymbol{p}) &= \sum_{i=1}^{N}(\boldsymbol{t}_{u,i}^\top(\boldsymbol{s},\boldsymbol{p}))(\boldsymbol{t}_{u,i}^\top(\boldsymbol{s},\boldsymbol{p}))^\dagger \\ &= \|s_u - p_u\|^2 \sum_{i=1}^{N}\|g_{i,u}\|^2(\boldsymbol{w}_i^\top)(\boldsymbol{w}_i^\top)^\dagger. \end{aligned}\tag{6.29}$$

It then follows that:

$$\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G}) = \frac{1}{m2^{Nm}}\sum_{k=(u-1)m+1}^{um}\sum_{\boldsymbol{s}\in\boldsymbol{\Psi}}\prod_{l=1}^{L}\left(1+\frac{\bar{\lambda}_{u,l}(\boldsymbol{s},\boldsymbol{p})}{4N_0\sin^2\theta}\right)^{-1} \tag{6.30}$$

where $\{\bar{\lambda}_{u,l}(\boldsymbol{s},\boldsymbol{p})\}$ are the eigenvalues of $\bar{\boldsymbol{Q}}_u(\boldsymbol{s},\boldsymbol{p}) = (\boldsymbol{B}_{\boldsymbol{h}}^\top)\boldsymbol{Q}_u(\boldsymbol{s},\boldsymbol{p})(\boldsymbol{B}_{\boldsymbol{h}}^\top)^\dagger$. It can be seen that for a given u, $\boldsymbol{Q}_u(\boldsymbol{s},\boldsymbol{p})$ depends only on the uth components of $\boldsymbol{s}$ and $\boldsymbol{p}$. Therefore, $\bar{\lambda}_{u,l}(\boldsymbol{s},\boldsymbol{p})$ can be denoted as $\bar{\lambda}_{u,l}(s_u,p_u)$.

Observe that s_u can be any signal point in the real two-dimensional constellation Ω and p_u is also a signal point in Ω whose label differs in only 1 bit at precisely position $j = (k-(u-1)m)$ compared to that of s_u. Therefore, instead of averaging over $m2^{Nm}$ cases of $\boldsymbol{s}\in\boldsymbol{\Psi}$ as in (6.30), $\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G})$ can be computed more efficiently by averaging over $m2^m$ cases of $s_u\in\Omega$. Furthermore, $\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G})$ can be written without the subscript u for s_u and p_u as follows:

$$\gamma_u(\boldsymbol{\Psi},\xi,\boldsymbol{G}) = \frac{1}{m2^m}\sum_{s\in\Omega}\sum_{j=1}^{m}\prod_{l=1}^{L}\left(1+\frac{\bar{\lambda}_{u,l}(s,p)}{4N_0\sin^2\theta}\right)^{-1}, \tag{6.31}$$

where s and p are two signal points in Ω whose labels differ in the position j and $\{\bar{\lambda}_{u,l}(s,p)\}$ are the eigenvalues of $\bar{\boldsymbol{Q}}_u(s,p) = (\boldsymbol{B}_{\boldsymbol{h}}^{\top})\boldsymbol{Q}_u(s,p)(\boldsymbol{B}_{\boldsymbol{h}}^{\top})^{\dagger}$ with

$$\boldsymbol{Q}_u(s,p) = \|s-p\|^2 \sum_{i=1}^{N} \|g_{i,u}\|^2 (\boldsymbol{w}_i^{\top})(\boldsymbol{w}_i^{\top})^{\dagger} \tag{6.32}$$

for a given u. Therefore, the average in (6.25) can be computed much easier as follows:

$$\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s\in\Omega} \sum_{j=1}^{m} \prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_{u,l}(s,p)}{4N_0 \sin^2\theta} \right)^{-1} \right]. \tag{6.33}$$

Applying (6.33) in (6.21), the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be efficiently computed with a high accuracy via a single integral. It can be observed that $\gamma(\boldsymbol{\Psi}, \xi, \boldsymbol{G})$ is computed by essentially averaging over the real two-dimensional constellation Ω, instead of $\boldsymbol{\Psi}$. This is due to the assumption that the mapping ξ is implemented independently for each signal component in Ω. Therefore, the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G})$ can be denoted as $f(d, \Omega, \xi, \boldsymbol{G})$.

To give an insight on how to design the matrix $\boldsymbol{G}$ and good mappings for a given constellation, use the inequality $Q\left(\sqrt{2\gamma}\right) < \frac{1}{2}\exp(-\gamma)$ to approximate the function $f(d, \boldsymbol{\Psi}, \xi, \boldsymbol{G}) = f(d, \Omega, \xi, \boldsymbol{G})$ as:

$$f(d, \Omega, \xi, \boldsymbol{G}) \approx \frac{1}{2}\delta^d(\Omega, \xi, \boldsymbol{G}), \tag{6.34}$$

where

$$\delta(\Omega, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s\in\Omega} \sum_{j=1}^{m} \prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_{u,l}(s,p)}{4N_0} \right)^{-1} \right]. \tag{6.35}$$

The parameter $\delta(\Omega, \xi, \boldsymbol{G})$ then can be used to characterize the influence of the rotation matrix $\boldsymbol{G}$, the constellation Ω and the mapping rule ξ to the asymptotic BEP performance of BICM-ID in OFDM with signal space diversity. In particular, for a given constellation Ω and the mapping ξ, one would prefer the rotation matrix $\boldsymbol{G}$ that minimizes $\delta(\Omega, \xi, \boldsymbol{G})$. It can be observed that the parameter $\delta(\Omega, \xi, \boldsymbol{G})$ depends only on the magnitudes of the elements in $\boldsymbol{G}$ at a specific signal to noise ratio (SNR) (i.e., given N_0). This is due to the fact that the error bound derived earlier is for the asymptotic performance of the systems. At very high SNR, (i.e., $N_0 \to 0$), the design

criterion can be made simpler and more meaningful as follows:

$$\bar{\delta}(\Omega, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s \in \Omega} \sum_{j=1}^{m} \prod_{l=1}^{L} \frac{1}{\bar{\lambda}_{u,l}(s,p)} \right]. \tag{6.36}$$

Furthermore, for a given u, one has the following equality:

$$\begin{aligned} \prod_{l=1}^{L} \bar{\lambda}_{u,l}(s,p) &= \det(\bar{\boldsymbol{Q}}_u(s,p)) = \det(\boldsymbol{R_h}) \det(\boldsymbol{Q}_u(s,p)) \\ &= \det(\boldsymbol{R_h}) \prod_{l=1}^{L} \lambda_{u,l}(s,p), \end{aligned} \tag{6.37}$$

where $\{\lambda_{u,l}(s,p)\}$ are eigenvalues of $\boldsymbol{Q}_u(s,p)$. Thus, the above equality simplifies the design parameter in (6.36) to the following one, which is independent of the correlation matrix $\boldsymbol{R_h}$:

$$\kappa(\Omega, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{u=1}^{N} \left[\sum_{s \in \Omega} \sum_{j=1}^{m} \prod_{l=1}^{L} \frac{1}{\lambda_{u,l}(s,p)} \right]. \tag{6.38}$$

In the next section, the optimal choice of $\boldsymbol{G}$ in terms of both the error performance and the decoding complexity based on the design criterion in (6.38) shall be discussed in more details.

6.3 Optimal Rotation Matrix G and Subcarrier Grouping

This section addresses the design problem to obtain the optimal matrix $\boldsymbol{G}$ in terms of both error performance and decoding complexity for a given constellation and mapping. To do so, the ultimate error performance achieved with rotation matrix $\boldsymbol{G}$ is investigated first.

6.3.1 Optimal Rotation Matrix

By interchanging the summations in (6.38), $\kappa(\Omega, \xi, \boldsymbol{G})$ is rewritten as:

$$\kappa(\Omega, \xi, \boldsymbol{G}) = \frac{1}{N} \cdot \frac{1}{m2^m} \sum_{s \in \Omega} \sum_{j=1}^{m} \left[\sum_{u=1}^{N} \prod_{l=1}^{L} \frac{1}{\lambda_{u,l}(s,p)} \right]. \tag{6.39}$$

Next, for each pair of signal points s and p in Ω whose labels differ in only 1 bit at position j, $1 \leq j \leq m$, define the parameter $\alpha(s,p,j)$ as follows:

$$\alpha(s,p,j) = \sum_{u=1}^{N} \prod_{l=1}^{L} \frac{1}{\lambda_{u,l}(s,p)}. \tag{6.40}$$

Observe that the matrix $(\boldsymbol{w}_i^{\top})(\boldsymbol{w}_i^{\top})^{\dagger}$ is positive semi-definite. Therefore, the eigenvalues of $\boldsymbol{Q}_u(s,p)$ in (6.32) are all non-negative. Furthermore, given $\boldsymbol{Q}_u(s,p)$ with L eigenvalues $\{\lambda_{u,l}(s,p)\}$, one has the following property:

$$\sum_{l=1}^{L} \lambda_{u,l}(s,p) = \text{trace}(\boldsymbol{Q}_u(s,p)). \tag{6.41}$$

Since the diagonal elements of $(\boldsymbol{w}_i^{\top})(\boldsymbol{w}_i^{\top})\dagger$ are all one, it then follows that:

$$\sum_{l=1}^{L} \lambda_{u,l}(s,p) = L\|s-p\|^2 \sum_{i=1}^{N} \|g_{i,u}\|^2. \tag{6.42}$$

Summing up the eigenvalues $\{\lambda_{u,l}(s,p)\}$ for all $1 \leq u \leq N$ and under the power constraint of $\boldsymbol{G}$, one has:

$$\sum_{u=1}^{N}\sum_{l=1}^{L} \lambda_{u,l}(s,p) = L\|s-p\|^2 \sum_{i=1}^{N}\sum_{u=1}^{N} \|g_{i,u}\|^2 = NL\|s-p\|^2. \tag{6.43}$$

Using Cauchy inequality, it is straightforward to show that $\alpha(s,p,j)$ in (6.40) achieves the minimum value when $\lambda_{u,l} = \|s-p\|^2$ for all u and l. Therefore, one obtains the lower bound for $\kappa(\Omega,\xi,\boldsymbol{G})$ in (6.39) as follows:

$$\kappa(\Omega,\xi,\boldsymbol{G}) \geq \frac{1}{m2^m} \sum_{s\in\Omega} \sum_{j=1}^{m} \left(\|s-p\|^2\right)^{-L}. \tag{6.44}$$

Define the class of rotation matrices $\boldsymbol{G}$ that achieves the above lower bound as the optimal set of $\boldsymbol{G}$. This means that the best asymptotic error performance of BICM-ID in OFDM with signal space diversity is achieved by using any $\boldsymbol{G}$ in this class.

Assume that there exists at least one matrix $\boldsymbol{G}$ in the optimal set. Then one has the following necessary and sufficient condition:

Condition 6.1: The necessary and sufficient condition for having the optimal matrix

$\boldsymbol{G}$ is that all the eigenvalues of the matrix $\boldsymbol{Q}_u(s,p)$ in (6.32) equal to $\|s-p\|^2$ for all $1 \leq u \leq N$.

Observe that the matrix $\boldsymbol{Q}_u(s,p)$ is diagonalizable. Therefore, it has equal eigenvalues if and only if it is a diagonal matrix with the diagonal elements equal to $\|s-p\|^2$. More specifically, let define $\tilde{\boldsymbol{Q}}_u = \sum_{i=1}^{N} \|g_{i,u}\|^2 (\boldsymbol{w}_i^\top)(\boldsymbol{w}_i^\top)^\dagger$, one has:

$$\tilde{\boldsymbol{Q}}_u = \boldsymbol{I}_L, \quad \text{for all } \ 1 \leq u \leq N, \tag{6.45}$$

with $\boldsymbol{I}_L$ is identity matrix of size L. Let $\tilde{\boldsymbol{Q}}_u[a,b]$ be the element at the ath row and the bth column of an $L \times L$ Hermitian matrix $\tilde{\boldsymbol{Q}}_u$. Then $\tilde{\boldsymbol{Q}}_u[a,b]$ is given as:

$$\tilde{\boldsymbol{Q}}_u[a,b] = \sum_{i=1}^{N} \|g_{i,u}\|^2 \exp\left(-\frac{2j\pi}{N} \cdot i \cdot (a-b)\right). \tag{6.46}$$

For convenience, let $\boldsymbol{g}^u = [\|g_{1,u}\|^2, \ldots, \|g_{N,u}\|^2]$ and the $L \times N$ matrix $\boldsymbol{W}$ with the element (l,i), with $1 \leq l \leq L$, $1 \leq i \leq N$, of $\boldsymbol{W}$ is:

$$\boldsymbol{W}[l,i] = \exp\left(-\frac{j2\pi}{N} \cdot (l-1) \cdot i\right). \tag{6.47}$$

Straightforwardly, (6.45) is satisfied if and only if:

$$\boldsymbol{W}.(\boldsymbol{g}^u)^\top = [1, \boldsymbol{0}_{1\times(L-1)}]^\top, \quad \text{for all } \ 1 \leq u \leq N. \tag{6.48}$$

At this point, it is natural to ask a question that whether the optimal set of rotation matrices is empty or not? Furthermore, if the set is not empty, what should be the most suitable $\boldsymbol{G}$ in terms of the decoding complexity?

Consider the class of rotation matrix $\boldsymbol{G}$ such that all the entries $\{g_{i,u}\}$ equals in magnitude as follows:

$$\|g_{i,u}\| = \frac{1}{\sqrt{N}}, \quad \text{for all } \ 1 \leq i,u \leq N. \tag{6.49}$$

Obviously, trace $(\text{diag}(\boldsymbol{g}^u)) = 1$. Furthermore, for any value of l, $2 \leq l \leq L$, one has:

$$\sum_{i=1}^{N} \exp\left(-\frac{j2\pi}{N} \cdot (l-1) \cdot i\right) = 0. \tag{6.50}$$

It then follows that $\boldsymbol{W}.(\boldsymbol{g}^u)^\top = [1, \mathbf{0}_{1\times(L-1)}]^\top$. Therefore, there exists at least one optimal rotation matrix $\boldsymbol{G}$ in which all the entries are equal in terms of the magnitudes.

By using the above optimal matrix $\boldsymbol{G}$, the best error performance in terms of asymptotic performance of BICM-ID in OFDM with signal space diversity can be achieved. Unfortunately, if such an optimal matrix $\boldsymbol{G}$ is applied, the OFDM symbols need to be coded over all N subcarriers. Therefore, the receiver complexity becomes very high (e.g., increasing exponentially with N). In the following subsection, subchannel grouping approach is considered. It is then shown that a much simpler method can be used without sacrificing the error performance.

6.3.2 Subcarrier Grouping

Consider the subcarrier grouping with V groups $I_v = (p_{v,1}, \ldots, p_{v,F})$, $1 \le v \le V$. Then a $F \times F$ rotation matrix $\boldsymbol{\Theta}$ with elements $\{\theta_{f,c}\}$, $1 \le f, c \le F$ is applied to each group. Without loss of generality, the following assumption can be made:

$$\begin{cases} p_{v,f} < p_{v,c} \text{ when } f < c \\ p_{v,1} < p_{u,1} \text{ when } v < u \end{cases} \tag{6.51}$$

Any element of the rotation equivalent rotation matrix $\boldsymbol{G}$ in (6.2) can be represented as:

$$\begin{cases} g_{p_{v,f},p_{v,c}} = \theta_{f,c} \\ g_{p_{v,f},p_{z,c}} = 0 \quad \text{if } \ v \neq z \ \ , 1 \le v, z \le V \end{cases} \text{for } 1 \le f, c \le F. \tag{6.52}$$

For a given v, define $\boldsymbol{W}^v$ to be the $L \times F$ matrix whose (l, f) element is

$$\boldsymbol{W}^v(l, f) = \exp\left(-\frac{j2\pi}{N} \cdot (l-1) \cdot p_{v,f}\right). \tag{6.53}$$

Also, for a given c, define the matrix $\boldsymbol{\Theta}^c = [\|\theta_{1,c}\|^2, \ldots, \|\theta_{F,c}\|^2]$. To achieve the best asymptotic performance, one has the following equivalent condition to that of (6.48):

$$\boldsymbol{W}^v.(\boldsymbol{\Theta}^c)^\top = \left[1, \mathbf{0}_{1\times(L-1)}\right]^\top, \quad \text{for all } \ 1 \le c \le F. \tag{6.54}$$

Clearly, the above condition not only depends on the rotation matrix $\boldsymbol{\Theta}$ but also on the way the subgroup I_v is formed (i.e., how to choose $p_{v,f}$ for each I_v). It has been

shown in previous section that when $F = N$, there exists at least one optimal solution for $\boldsymbol{\Theta}$. However, in the view of receiver complexity, one would prefer the solution for $\boldsymbol{\Theta}$ in which the value of F is as small as possible. The following theorem provides the lower bound for F.

Theorem 6.1: The optimal solution of the rotation matrix $\boldsymbol{\Theta}$ does not exist when $F < L$.

Proof: Assume that there exist an optimal solution when $F < L$, i.e., $F \leq (L-1)$. Construct the square matrix $\boldsymbol{W}_F^v$ with size $F \times F$ including F rows from the second row to the $(F+1)$th row of $\boldsymbol{W}^v$. More specifically, the (q, f) element is $\boldsymbol{W}_F^v(q, f) = \exp\left(-\frac{j2\pi}{N} \cdot q \cdot p_{v,f}\right)$, $1 \leq q, f \leq F$. It follows from (6.54) that:

$$\boldsymbol{W}_F^v.(\boldsymbol{\Theta}^c)^\top = \mathbf{0}_{F\times 1}. \tag{6.55}$$

The determinant of $\boldsymbol{W}_F^v$ is computed as:

$$\det\left(\boldsymbol{W}_F^v\right) = \left[\prod_{f=1}^{F} \exp\left(-\frac{j2\pi}{N} \cdot p_{v,f}\right)\right] \det\left(\boldsymbol{W}_{F-1}^v\right), \tag{6.56}$$

where

$$\boldsymbol{W}_{F-1}^v = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{v,1} & x_{v,2} & \dots & x_{v,F} \\ x_{v,1}^2 & x_{v,2}^2 & \dots & x_{v,F}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{v,1}^{F-1} & x_{v,2}^{F-1} & \dots & x_{v,F}^{F-1} \end{pmatrix}, \tag{6.57}$$

with $x_{v,f} = \exp\left(-\frac{j2\pi}{N} \cdot p_{v,f}\right)$, $1 \leq f \leq F$. Observe that $\boldsymbol{W}_{F-1}^v$ is Vandermonde matrix of order F. Its determinant can be computed in a particularly simple form as follows [92]:

$$\det\left(\boldsymbol{W}_{F-1}^v\right) = \prod_{f,q:f>q} (x_{v,f} - x_{v,q}). \tag{6.58}$$

Since $1 \leq p_{v,q} < p_{v,f} \leq N$, one has $x_{v,f} - x_{v,q} \neq 0$. Therefore, the determinant of $\boldsymbol{W}_{F-1}^v$ and $\boldsymbol{W}_F^v$ can not be 0. It then follows that the only solution for (6.55) is $\boldsymbol{\Theta}^c = \mathbf{0}_{1\times F}$, which contradicts the condition that $\text{trace}\left(\text{diag}(\boldsymbol{\Theta}^c)\right) = 1$ in (6.54). The theorem is thus proved.

The result stated in Theorem 6.1 is predicable, since there are in total of L independent fading channels. Now, the optimal solution of subcarrier grouping and rotation matrix $\boldsymbol{\Theta}$ for the case of $F = L$ will be given explicitly. Clearly, $F = L$ is the most desirable value in terms of the decoding complexity. Furthermore, another condition is N div $L = 0$, which can be easily met by picking up a suitable number of subcarriers N.

Now consider the case $F = L$. It is clear that the matrix $\boldsymbol{W}^v$ in (6.53) equals $\boldsymbol{W}^v_{F-1}$ in (6.57). Let $\boldsymbol{b} = \left[1, \mathbf{0}_{1\times(F-1)}\right]$. The necessary and sufficient condition for the existence of optimal solution in (6.54) then becomes:

$$\boldsymbol{W}^v_{F-1}.(\boldsymbol{\Theta}^c)^\top = \boldsymbol{b}^\top, \quad \text{for all } \ 1 \le c \le F. \tag{6.59}$$

The above system is exactly the linear Vandermonde system. Since the determinant of Vandermonde matrix $\boldsymbol{W}^v_{F-1}$ is not 0, there is a unique solution for $\boldsymbol{\Theta}^c$. Therefore, it is expected that $\{\boldsymbol{\Theta}^c\}$ are identical for all $1 \le c \le F$. The solution of (6.59) is closely related to Lagrange's polynomial interpolation formula, which is generally presented in [93]. First, construct the set of polynomial of degree $(F-1)$ $\{P_{v,f}(x)\}$, $1 \le f \le F$, as follows:

$$P_{v,f}(x) = \prod_{q=1,q\neq f}^{F} \frac{x - x_{v,q}}{x_{v,f} - x_{v,q}} = \sum_{k=1}^{F} A^v_{f,k} x^{k-1}, \tag{6.60}$$

where $A^v_{f,k}$ is the coefficient corresponding to the order $(k-1)$ of $P_{v,f}(x)$. The solution of (6.59) is just the matrix inverse times the right-side $\boldsymbol{b}^\top$, which is [93]:

$$\Theta^c_f = \sum_{k=1}^{F} A^v_{f,k} b_k, \tag{6.61}$$

where Θ^c_f and b_k are the fth and the kth elements of $\boldsymbol{\Theta}^c$ and $\boldsymbol{b}$, respectively. Since the vector $\boldsymbol{b} = \left[1, \mathbf{0}_{1\times(F-1)}\right]$, the solution of (6.59) is simply:

$$\Theta^c_f = A^v_{f,1}, \tag{6.62}$$

where $A^v_{f,1}$ can be computed from (6.60) as:

$$A^v_{f,1} = (-1)^{F-1} \prod_{q=1,q\neq f}^{F} \frac{x_{v,q}}{x_{v,f} - x_{v,q}}. \tag{6.63}$$

Therefore, one has:

$$\Theta_f^c = (-1)^{F-1} \prod_{q=1, q\neq f}^{F} \frac{x_{v,q}}{x_{v,f} - x_{v,q}}. \tag{6.64}$$

Each term in the above product can be computed as:

$$\begin{aligned} \frac{x_{v,q}}{x_{v,f} - x_{v,q}} &= \frac{\exp\left(-\frac{j2\pi}{N} \cdot p_{v,q}\right)}{\exp\left(-\frac{j2\pi}{N} \cdot p_{v,f}\right) - \exp\left(-\frac{j2\pi}{N} \cdot p_{v,q}\right)} \\ &= \frac{-1}{1 - \exp\left(-\frac{j2\pi}{N} \cdot (p_{v,f} - p_{v,q})\right)} \\ &= \frac{-\exp\left(\frac{j\pi}{N} \cdot (p_{v,f} - p_{v,q})\right)}{2j \sin\left(\frac{\pi}{N}(p_{v,f} - p_{v,q})\right)}. \end{aligned} \tag{6.65}$$

It then follows that:

$$\Theta_f^c = \left(\frac{1}{2j}\right)^{F-1} \left[\frac{\exp\left(\frac{j\pi}{N} \cdot \left(F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,q}\right)\right)}{\prod_{q\neq f} \sin\left(\frac{\pi}{N}(p_{v,f} - p_{v,q})\right)}\right]. \tag{6.66}$$

Since $\Theta_f^c = \|\theta_{f,c}\|^2$, a necessary condition for Θ_f^c in (6.66) is that it is a real number. In the following, two cases of F will be considered separately.

- F is odd: It is then straightforward to see that Θ_f^c is real if and only if:

$$\sin\left(\frac{\pi}{N} \cdot \left(F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,q}\right)\right) = 0. \tag{6.67}$$

  Therefore,

$$F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,q} = k_f N, \tag{6.68}$$

  with k_f is an integer. Consider for any f and q such that $1 \leq f < q \leq F$. It then follows that:

$$F \cdot p_{v,q} - F \cdot p_{v,f} = k_q N - k_f N. \tag{6.69}$$

  Thus,

$$p_{v,q} - p_{v,f} = (k_q - k_f)N/F = (k_q - k_f)V. \tag{6.70}$$

  Combining with the first condition in (6.51), one has the following unique form of the elements of I_v for a given v:

$$p_{v,f} = p_{v,1} + V(f-1), \quad 1 \leq f \leq F. \tag{6.71}$$

Using the second condition in (6.51), the necessary condition for the existence of optimal solution related to subcarrier groups is:

$$p_{v,f} = v + V(f-1) \tag{6.72}$$

for all $1 \leq v \leq V$ and $1 \leq f \leq F$.

- F is even: It is then straightforward to see that Θ_f^c is real if and only if:

$$\cos\left(\frac{\pi}{N} \cdot \left(F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,q}\right)\right) = 0. \tag{6.73}$$

Therefore,

$$F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,q} = N/2 + k_f N. \tag{6.74}$$

Similar to the case of odd F, it is easy to show that the elements of subcarrier groups also satisfy the condition in (6.72).

From the above results, it is obvious that (6.72) is the necessary condition for the existence of the optimal solution in terms of both subcarrier grouping and applying the rotation matrix $\mathbf{\Theta}$. Furthermore, it is shown in Appendix D that when (6.72) is satisfied, $\Theta_f^c = 1/F$. Therefore, one ends up with the following unique jointly optimal solution of both subcarrier grouping and rotation matrix $\mathbf{\Theta}$:

$$\begin{cases} p_{v,f} = v + V(f-1) \\ \|\theta_{f,c}\| = \frac{1}{\sqrt{F}} \end{cases} \quad \text{for } 1 \leq v \leq V \text{ and } 1 \leq f, c \leq F. \tag{6.75}$$

It should be mentioned that the subcarrier grouping in (6.75) is similar to the result in [82] for uncoded OFDM, which was only specifically chosen. On the other hand, here, the jointly optimal solution for both subcarrier grouping and the rotation matrix $\mathbf{\Theta}$ in (6.75) was presented thoroughly and explicitly for coded OFDM systems.

When $F > L$, it is simple to see that the subcarrier grouping and the rotation matrix in (6.75) also guarantee the optimal solution. Such a solution, however, might not be unique.

For iterative decoding systems, one also needs to study the performance of the system after the first iteration (i.e., the performance of BICM). This is because such performance influences the convergence behavior of BICM-ID. To this end, the analysis for the worst case of the PEP for uncoded OFDM systems with subcarrier grouping can also be used, as similar to the previous chapters. Recent results in [77] indicated that one can implement the subcarrier grouping in (6.75) and the LCP-A rotation matrix $\mathbf{\Theta}$ in (4.22)[3] for each group in conjunction with QAM constellations to achieve the maximum diversity and coding gains for uncoded OFDM systems. Therefore, for the system considered in this chapter, the use of LCP-A $\mathbf{\Theta}$ and the subcarrier grouping in (6.75) would be optimal with respect to both the asymptotic error performance and the convergence behavior.

6.4 Illustrative Results

In this section, analytical and simulation results are provided to confirm the analysis carried out in the previous sections. The QPSK modulation scheme with anti-Gray mapping rule is employed. It should be mentioned here that Gray and anti-Gray mappings are the only two available mappings for QPSK constellation. Anti-Gray mapping is chosen simply due to its superiority in providing a better error performance in iterative systems. This fact can also be confirmed for the coded OFDM systems under consideration by comparing the parameters $\kappa(\Omega, \xi, \boldsymbol{G})$ in (6.38) for the two mappings when the optimal rotation matrix and subcarrier grouping are applied.

All the systems are simulated with $N = 64$ subcarriers, which is similar to that considered in HiperLan/2 standard [90]. Unless specified otherwise, a rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ is applied as an outer code, along with a random interleaver of length 9,600 coded bits. Each point in the BER curves is simulated with 10^7 to 5×10^8 coded bits. The following first two subsections present the results for the case that the channel gains are generated independently for each OFDM symbol and the optimal MAP demodulator is implemented. The last

[3]It should be noted that $\mathbf{\Theta}$ is of size $F \times F$.

subsection studies the error performances of the proposed systems in a real channel environment by considering Channel Model A of HiperLan/2 [90]. Due to the large number of channel taps specified in this channel model, the high complexity of the optimal MAP demodulator makes it impractical to implement. To overcome this difficulty, a suboptimal SISO demodulator based on the vector Gaussian approximation (GA) developed in [51] is employed.

6.4.1 I. I. D. Channels

This subsection considers the i.i.d. channel model, in which the correlation matrix $\boldsymbol{R_h} = 1/L \cdot \boldsymbol{I}_L$.

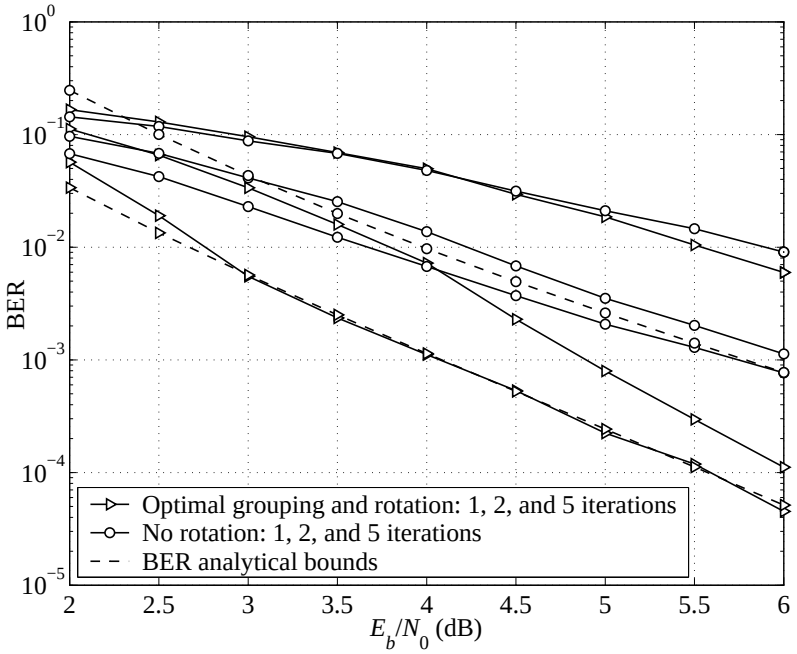


Figure 6.2 Performance of BICM-ID in OFDM systems: Optimal subcarrier grouping and rotation is compared to that without rotation.

Figure 6.2 compares the BER performance with 1, 2 and 5 iterations of the system without signal space diversity and to that of the system employing optimal grouping and rotation matrix when $L = 2$ channel taps. The rotation matrix $\boldsymbol{\Theta}$ of size $F \times F$, $F = L$, is chosen by using the construction in (4.22) with $\alpha_1 = \alpha = \exp\left(j\frac{2\pi}{4F}\right)$ and $\alpha_f = \alpha \exp\left(j\frac{2\pi(f-1)}{F}\right), 1 \leq f \leq F$. It can be seen that performances of both systems

converge to the analytical BER bounds. A significant coding gain is also achieved by the system employing signal space diversity.

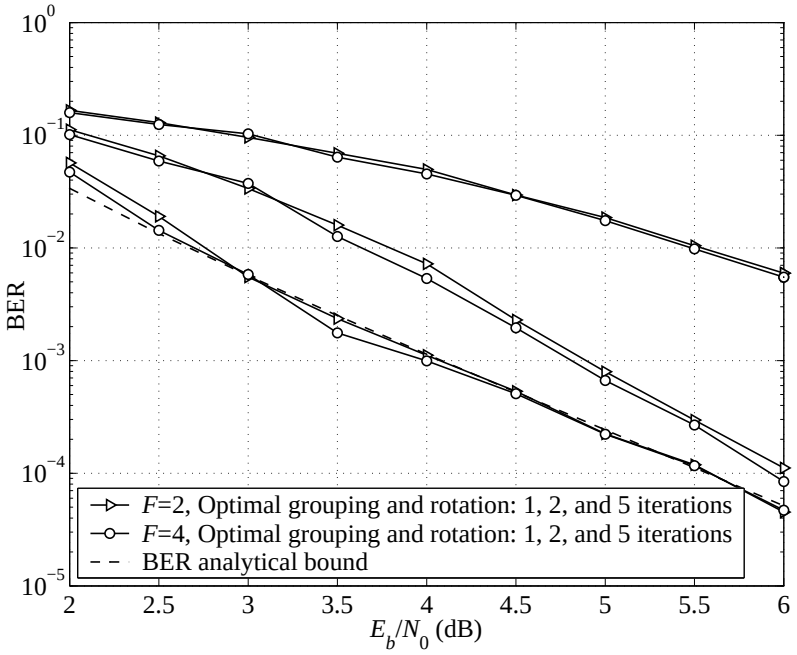


Figure 6.3 Performance of BICM-ID in OFDM systems: Optimal subcarrier grouping and rotation for $F = 2$ and $F = 4$.

In Fig. 6.3, the error performances after 1, 2 and 5 iterations for systems employing optimal grouping for the cases of $F = 2$ and $F = 4$ are presented. The two rotation matrices are also constructed using (4.22). It can be observed that there is almost no difference between the performances of the two systems at any iteration and they all converge to the same BER analytical bound. This results confirms the analysis for the optimal grouping and rotation matrices with different values of F. Clearly, $F = L = 2$ is preferred in terms of receiver complexity.

Figure 6.4 illustrates the performance of various grouping approaches. Besides the optimal one obtained in section 6.3, two more different approaches are considered, namely natural grouping with $I_v = (2v - 1, 2v)$ and average grouping with $I_v = (v, N + 1 - v)$. The error performances of systems employing the same 2×2 rotation matrix $\mathbf{\Theta}$ as in the previous case with 1 and 5 iterations are shown. Obviously, the error bounds of all systems are very tight, which makes them very useful tool

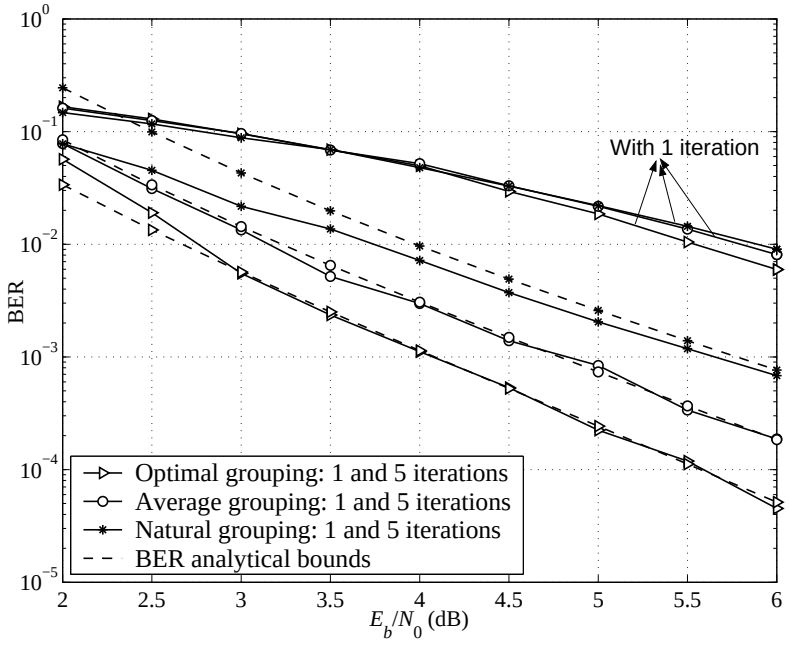


Figure 6.4 Performance of BICM-ID in OFDM systems with various carrier grouping approaches.

to predict the error performance of BICM-ID in OFDM with signal space diversity. Though not explicitly shown here for the brevity of presentation, examining the error bounds plotted over a wider range of SNR reveals that, compared to the averaged and natural groupings, the use of the optimal grouping results in coding gains as high as 1dB and 2.2dB at the BER level of 10^{-4}, respectively.

Figure 6.5 demonstrates the effect of the rotation matrix $\mathbf{\Theta}$ where the optimal grouping is applied. Shown in the figure are the BER performance after 1 and 5 iterations of systems using different choices of $\mathbf{\Theta}$. These choices include (i) an optimal matrix $\mathbf{\Theta}$ in (4.22); (ii) a randomly-generated $\mathbf{\Theta}$:

$$\mathbf{\Theta} = \begin{pmatrix} 0.6771 - 0.7985i & -0.0182 - 0.0936i \\ -0.0175 - 0.7708i & 0.2115 + 0.5056i \end{pmatrix}, \tag{6.76}$$

and (iii) the optimal rotation matrix $\mathbf{\Theta}_\theta$ that maximizes the minimum product distance for real unitary transformation with a full signal space diversity [17, 46, 60, 61],

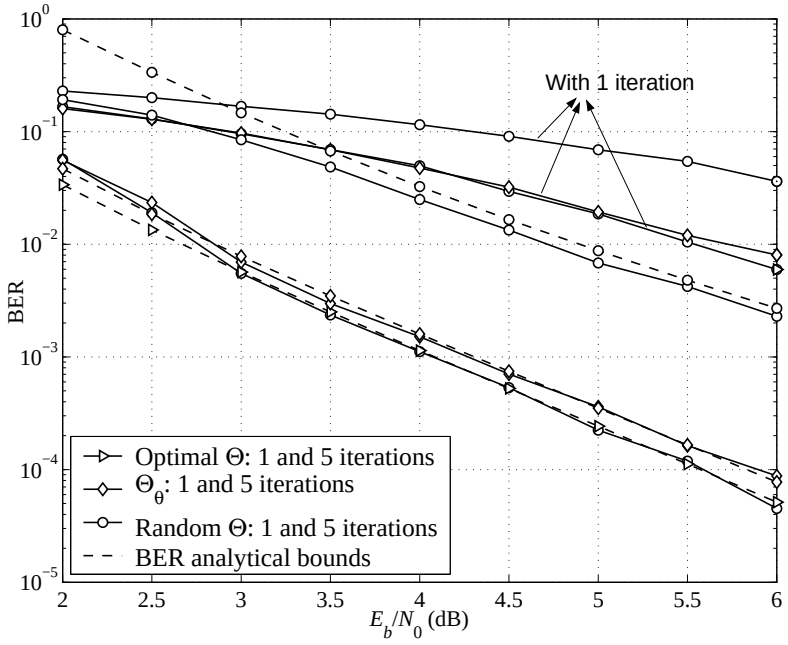


Figure 6.5 Performance of BICM-ID in OFDM systems: Optimal subcarrier grouping with various rotation matrices $\mathbf{\Theta}$.

given as:

$$\mathbf{\Theta}_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \text{ where } \tan(2\theta) = 2. \tag{6.77}$$

Also plotted in this figure are the error bounds for all the systems under consideration (the broken lines). Observe that, compared to the random $\mathbf{\Theta}$ and $\mathbf{\Theta}_\theta$, the use of the optimal rotation matrix results in coding gains as high as 0.25dB and 2.6dB at the BER level of 10^{-4}, respectively. The results clearly agree with the analysis on the jointly optimum subcarrier group and rotation matrix $\mathbf{\Theta}$. Note also that the error bounds are tight for all the systems.

Finally, Fig. 6.6 shows the error performances of two systems employing optimal carrier grouping and rotation matrices with $L = 2$ and $L = 4$. In the case of $L = 4$, the 4×4 rotation matrix $\mathbf{\Theta}$ is chosen from (4.22). Clearly, with optimal grouping and rotation matrix, the use of BICM-ID in OFDM with signal space diversity successfully exploit the advantage of multipath diversity in channel environment with the richer scattering effect (i.e., corresponding to a larger number of channel taps).

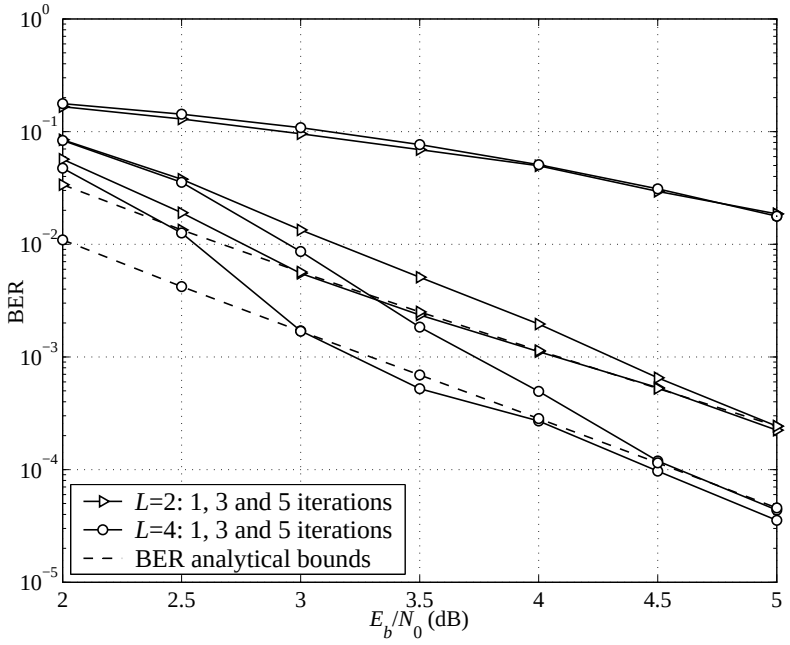


Figure 6.6 Error performances for systems with $L = 2$ and $L = 4$: Optimal carrier grouping and rotation matrices are employed.

6.4.2 Correlated Channels

Figure 6.7 first compares the BER performance with 1, 2 and 5 iterations of systems over the i.i.d. channel and a correlated channel with number of channel taps $L = 2$. For the correlated channel, the channel correlation matrix is assumed to be:

$$\boldsymbol{R_h} = \frac{1}{L} \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \tag{6.78}$$

For both systems, the same optimal carrier grouping and rotation matrix $\boldsymbol{\Theta}$ are employed. It can be seen that the BER performances over the correlated channel significantly degrade at any iterations. Observe that the asymptotic error bound is also very tight for correlated channel. The slopes of the asymptotic error performances of both systems are, however, the same, which confirms that both systems fully exploit the maximum diversity order $L = 2$. By examining the error bounds over a wider range of SNR, it is observed that performance degradation over the correlated channel transfers to about an additional 1dB required to achieve the BER level of 10^{-4}, which is quite noticeable. Though not shown here due to space limit, the effects of subcarrier

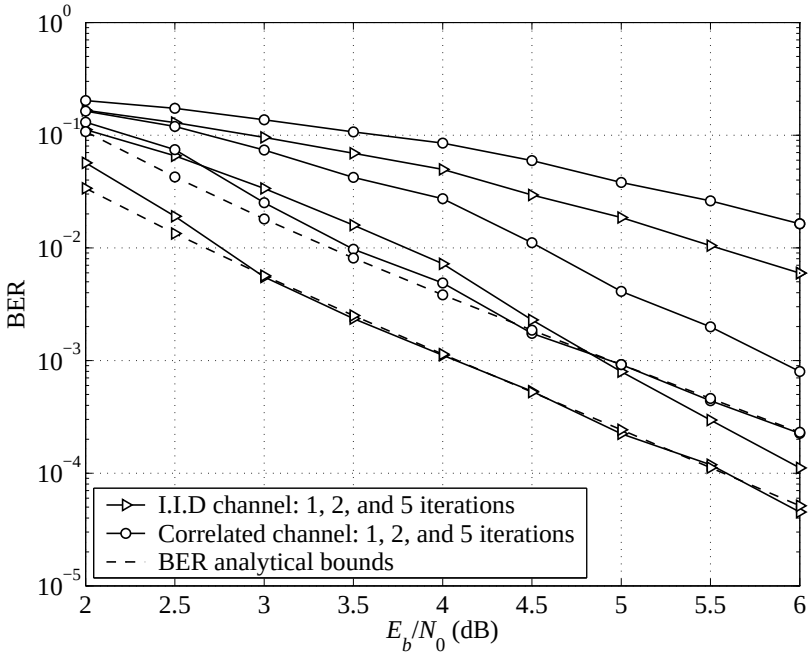


Figure 6.7 Error performances for systems operating over i.i.d. and correlated channels with $L = 2$: Optimal carrier grouping and rotation matrices are employed in both cases.

grouping and rotation matrix over correlated channels are very similar to that over the i.i.d. channels.

6.4.3 Channel Model A in HiperLan/2

In Channel Model A of HiperLan/2 [90], the carrier frequency is set at 5.2 GHz and the mobile's velocity is 3 m/s, which results in the Doppler frequency $f_d = 52$ Hz. Each OFDM symbol duration is $T_s = 4$ μs, corresponding to the normalized Doppler frequency $F_d = 2.08 \times 10^{-4}$. There are 8 channel taps (i.e., $L = 8$) where the variances in the tap order are given as {0.4505, 0.3467, 0.1283, 0.0522, 0.0102, 0.0077, 0.0029, 0.0014} [82, 90]. The gains of the channel taps are modeled as independent circularly symmetric complex Gaussian random variables and they are generated according to the Jakes' model. This means that there is correlation among channel taps over time. As mentioned earlier, the large number of channel taps implies a very high complexity of the optimal MAP demodulator in the systems with rotation. To reduce

the receiver complexity of such systems, a suboptimal SISO demodulator based on the vector Gaussian approximation (GA) in [51] shall be used. In the case of the systems without rotation, the MAP demodulator is still employed.

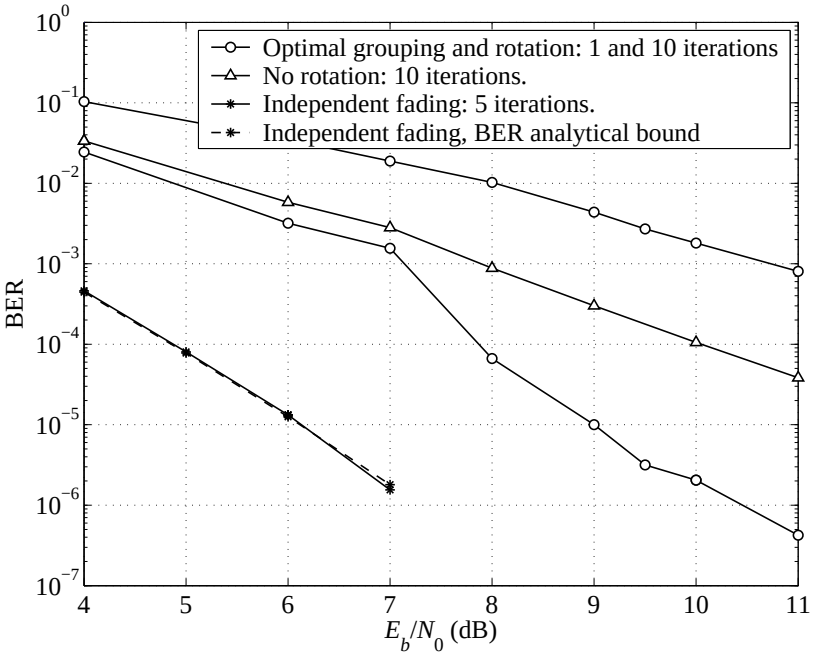


Figure 6.8 Error performance comparisons with Channel Model A in HiperLan/2.

Figure 6.8 compares the error performances of the proposed system implementing the optimal rotation and subcarrier grouping after 1 and 10 iterations and that of the system without rotation after 10 iterations. Due to the high correlation of the channel taps across multiple OFDM symbols, the random interleaver of length 204,800 coded bits is applied, which corresponds to a frame of 1,600 OFDM symbols. The same outer code as used in the previous subsections is applied. To serve as the reference, the error performance after 5 iterations and the analytical error bound of the proposed system when the channel taps are generated *independently* across OFDM symbols are provided. As can be seen from Fig. 6.8, implementing the optimal rotation and grouping significantly improves the error performance of the conventional system that does not have constellation rotation after 10 iterations. For example, an SNR gain of about 2dB is realized by the proposed system at the BER level of 10^{-4}. Compared to the case of independent fading across OFDM symbols, the error performance is

significantly poorer because of the high correlation of the channel taps over time present in Channel Model A of HiperLan/2.

7. Multi-Dimensional Subcarrier Mapping for Bit-Interleaved Coded OFDM with Iterative Decoding[1]

This chapter extends the investigation in the application of BICM-ID in OFDM systems by combining the ideas of subcarrier grouping studied in Chapter 6 and multi-dimensional (multi-D) mapping introduced in Chapter 3.

In Chapter 3, it was demonstrated that the performance of BICM-ID over *frequency non-selective* fading channels with multi-D mapping together with a simple convolutional code can achieve a significant gain over a classical BICM-ID system. On the other hand, by investigating bit-interleaved coded modulation (BICM) in OFDM with iterative decoding (BI-COFDM-ID) employing the standard convolutional codes and a conventional mapping rule, it was shown in Chapter 6 that SSD can be implemented over subcarrier groups, each having at least L elements, to achieve the same optimum performance as in the case of spreading information over all N subcarriers, where L is the number of channel taps. Compared to the system without SSD, a significant coding gain can be achieved by the system that implements SSD. The trade-off is the higher receiver complexity in order to perform the soft-input soft-output demodulation of a higher-dimensional constellation.

Given the impressive results with respect to subcarrier grouping and signal space diversity previously obtained for BI-COFDM-ID and the performance advantage of multi-dimensional mapping in general, it is natural to study BI-COFDM-ID sys-

[1] A part of this chapter was presented in [94, 95]

tems in which both the multi-dimensional mapping and signal space diversity can be performed over a group of subcarriers. In particular, this chapter proposes a multi-dimensional subcarrier mapping scheme over groups of L subcarriers each. The proposed scheme offers more flexibility for the mapping design to optimize the system performance with the same receiver complexity as compared to the BI-COFDM-ID systems that employ SSD in Chapter 6. Interestingly, it shall be shown that the subcarrier mapping and grouping leads to a significant performance improvement over the systems in Chapter 6, even without using SSD.

The remainder of the chapter is organized as follows. The system model of BI-COFDM-ID with optional use of SSD, and the implementation of multi-dimensional subcarrier mapping are described first. A tight bound on the *asymptotic* error performance of the proposed scheme is then derived and used to establish the design criterion that characterizes the effects of subcarrier mapping and grouping on the system performance. It is then shown that their effects are independent of each other, and hence subcarrier mapping and grouping can be independently optimized to achieve the best performance. Furthermore, it is proved that, for the proposed scheme, the optimal subcarrier mapping and grouping are similar to the optimal multi-dimensional mapping structure of BICM-ID in a *frequency non-selective* Rayleigh fading environment studied in Chapter 3, and the optimal subcarrier grouping for OFDM with SSD studied in Chapter 6, respectively. Analytical and simulation results are provided to show that the combined optimal subcarrier mapping and grouping can successfully exploit the multi-path diversity and, at the same time offer remarkable coding gains without the need of implementing SSD as compared to the systems in Chapter 6.

7.1 System Model

The block diagram of BI-COFDM-ID with multi-D mapping is shown in Fig. 7.1. The information sequence $\underline{\boldsymbol{u}}$ is encoded into a coded sequence $\underline{\boldsymbol{c}}$ by a convolutional code. The coded sequence $\underline{\boldsymbol{c}}$ is then interleaved by a bit-wise interleaver to become the interleaved sequence $\underline{\boldsymbol{v}}$. Consider an OFDM system with N subcarriers and assume

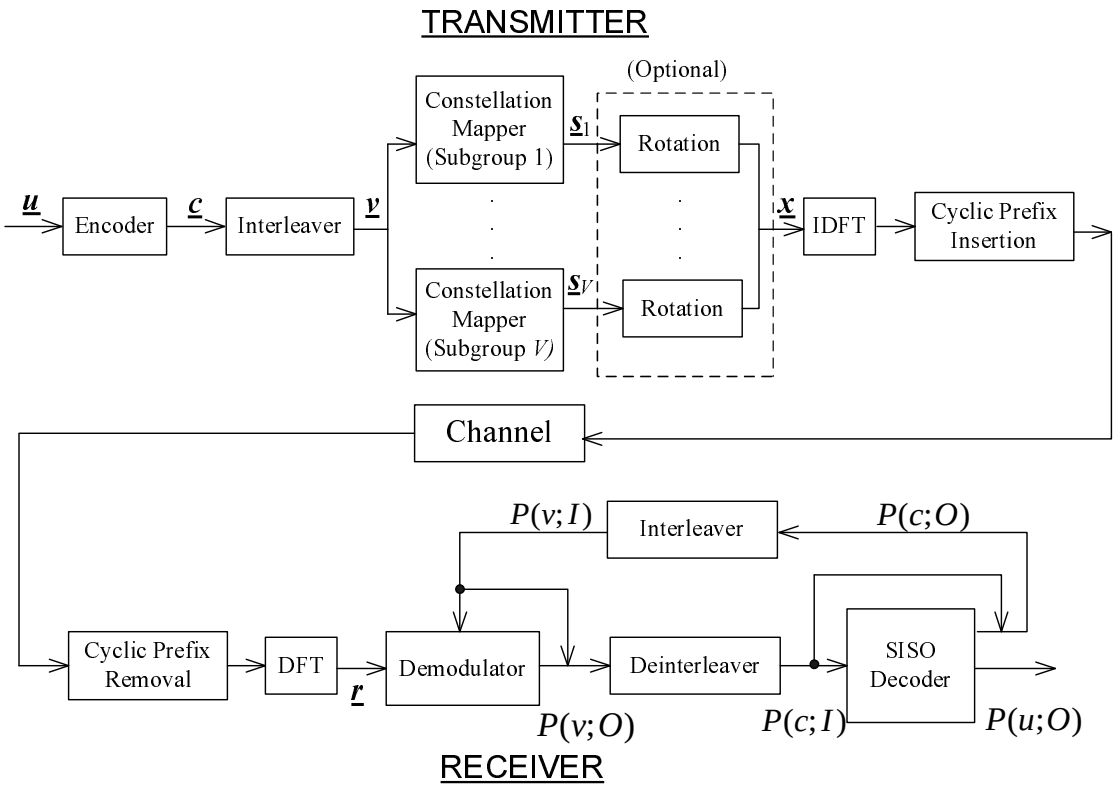


Figure 7.1 Block diagram of a BICM-ID for OFDM with multi-D mapping over subgroups of carriers.

that $N = VL$, where L is the number of channel taps in the representation of the *frequency selective* fading channel. Similar to Chapter 6, let $\mathcal{I} := \{1, \ldots, N\}$ be the set of N carrier indices. Divide N carriers into V nonintersecting subsets $\{I_1, \ldots, I_V\}$, where each subset I_v, $1 \leq v \leq V$, includes L carrier indices $I_v = (p_{v,1}, \ldots, p_{v,L})$.

Each group of mN interleaved coded bits is divided into V subgroups, where each subgroup contains mL bits. These mL bits are then mapped to a sequence of L complex symbols $\boldsymbol{s}_v = [s_{p_{v,1}}, s_{p_{v,2}}, \ldots, s_{p_{v,L}}]$, where each element $s_{p_{v,l}}$, $1 \leq l \leq L$, is a symbol in a complex 1-D constellation Ω of size $M = 2^m$. In a conventional system considered in Chapter 6, a complex 1-D mapping rule ξ would be used to map each group of m coded bits to one signal point in Ω. Here, by mapping mL coded bits simultaneously to $\boldsymbol{s}_v$, a more general mapping rule ζ is implemented in a complex L-D constellation Φ of size M^L. Clearly, the mapping rule ξ is just one special case

of ζ. SSD is then applied to each group I_v to obtain a sequence $\boldsymbol{x}_v$ of length L as $\boldsymbol{x}_v^\top = \boldsymbol{\Theta}\boldsymbol{s}_v^\top$, where $\boldsymbol{\Theta}$ can be any $L \times L$ matrix whose complex elements $\theta_{l,c}$ satisfies the power constraint $\sum_{l=1}^{L}\sum_{c=1}^{L} \|\theta_{l,c}\|^2 = L$. The V sequences $\{\boldsymbol{x}_v\}$ form the rotated sequence $\boldsymbol{x}$.

In Fig. 7.1, the blocks inside the dashed lines are optional, i.e., it might not be necessary to perform SSD (or rotation) in the system. This case corresponds to setting the rotation matrix $\boldsymbol{\Theta}$ equal to the identity matrix $\boldsymbol{I}_L$. This option will be discussed in detail in the next sections. It should also be mentioned that in order to take the advantages of frequency diversity, it is assumed that $L > 1$. It is not surprising to see that if $L = 1$ then Φ becomes Ω and the techniques of multi-D mapping and/or SSD are not necessary.

Let $\boldsymbol{s}$ be formed by concatenating V sequences $\{\boldsymbol{s}_v\}$, $v = 1, \ldots, V$. In the most general view, one can consider $\boldsymbol{s}$ as a signal point in a complex N-D constellation $\boldsymbol{\Psi}$, which has M^N signal points in total. The mapping rules ξ and ζ implemented in Ω and Φ mentioned before can also be seen as special cases of the more general mapping rule χ from Nm coded bits to a signal point in $\boldsymbol{\Psi}$. The relationship between $\boldsymbol{s}$ and $\boldsymbol{x}$ is $\boldsymbol{x}^\top = \boldsymbol{G}\boldsymbol{s}^\top$, where the elements of the $N \times N$ matrix $\boldsymbol{G}$ can be determined from the elements of $\boldsymbol{\Theta}$ as:

$$\begin{cases} g_{p_{v,l},p_{v,c}} = \theta_{l,c} \\ g_{p_{v,l},p_{z,c}} = 0 \quad \text{if } v \neq z, \quad 1 \leq v, z \leq V \end{cases} \quad \text{for } 1 \leq l, c \leq L. \tag{7.1}$$

In the discussion hereafter, each group of coded bits carried by a signal point is called the label of that signal point, where the signal point can be in either Ω, Φ or $\boldsymbol{\Psi}$.

Similar to Chapter 6, the sequence $\boldsymbol{x}$ is processed with IDFT and then inserted with a cyclic prefix of length L_p. With proper DFT processing and cyclic prefix removal at the receiver, the equivalent received signal can be written as $\boldsymbol{r}^\top = \boldsymbol{H}\boldsymbol{G}\boldsymbol{s}^\top + \boldsymbol{n}^\top$, where each entry n_i of the noise vector $\boldsymbol{n} = [n_1, \ldots, n_N]$ is $\mathcal{CN}(0, N_0)$. The matrix $\boldsymbol{H} = \text{diag}(H_1, \ldots, H_N)$ contains the correlated fading coefficients in its diagonal,

where $H_i = \boldsymbol{h}\boldsymbol{w}_i^{\top}$, $1 \leq i \leq N$, with

$$\boldsymbol{w}_i = \left[\exp\left(-j\frac{2\pi i 0}{N}\right), \ldots, \exp\left(-j\frac{2\pi i(L-1)}{N}\right)\right]. \tag{7.2}$$

The channel vector $\boldsymbol{h} = [h_1, \ldots, h_L]$ contains the channel gains of all the L taps. For a Rayleigh fading channel, each h_l, $l = 1, \ldots, L$, is also modeled as a complex circularly symmetric Gaussian random variable. Similar to Chapter 6, it is assumed that the channel gains remain constant within one OFDM symbol and change independently from one OFDM symbol to the next. Furthermore, within each OFDM symbol, the general case of correlated channel taps is considered by assuming that the channel vector $\boldsymbol{h}$ has a full-rank correlation matrix $\boldsymbol{R}_{\boldsymbol{h}} = E(\boldsymbol{h}^{\dagger}\boldsymbol{h})$, whose square root matrix is $\boldsymbol{B}_{\boldsymbol{h}}$ [80–82].

As shown in Fig. 7.1, the receiver includes the soft-input soft-output (SISO) demodulator and the SISO channel decoder. The optimal soft-output demodulator, i.e., the maximum *a posteriori* probability (MAP) demodulator [7,96], is implemented for each group of subcarriers I_v in both the proposed BI-COFDM-ID with multi-D mapping and BI-COFDM-ID with SSD studied in Chapter 6. Therefore, they have the same receiver complexity. The SISO channel decoder uses the MAP algorithm. The demodulator and the channel decoder exchange the extrinsic information of the coded bits through an iterative process. The total *a posteriori* probabilities of the information bits can be computed to make the hard decisions at the output of the decoder after each iteration.

7.2 Performance Evaluation and Design Criterion

In this section, given a complex L-D constellation Φ, the asymptotic bit error probability of BI-COFDM-ID with a multi-D subcarrier mapping ζ is evaluated by assuming error-free feedback from the decoder to the demodulator.

The union bound on the bit error probability for a system using a rate-k_c/n_c convolutional code, a complex N-D constellation $\boldsymbol{\Psi}$, a mapping rule χ, and a general

rotation matrix $\boldsymbol{G}$ can be written in a general form as:

$$P_b \leq \frac{1}{k_c} \sum_{d=d_H}^{\infty} c_d f(d, \boldsymbol{\Psi}, \chi, \boldsymbol{G}). \tag{7.3}$$

By evaluating the PEP of two codewords and invoking the assumption of error-free feedback from the decoder to the demodulator, $f(d, \boldsymbol{\Psi}, \chi, \boldsymbol{G})$ can be computed as in Chapter 6:

$$f(d, \boldsymbol{\Psi}, \chi, \boldsymbol{G}) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[\underbrace{\mathcal{E}_{\boldsymbol{s},\boldsymbol{p}} \left\{ \prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_l(\boldsymbol{s}, \boldsymbol{p})}{4N_0 \sin^2 \theta} \right)^{-1} \right\}}_{\gamma(\boldsymbol{\Psi}, \chi, \boldsymbol{G})} \right]^d \mathrm{d}\theta, \tag{7.4}$$

where the expectation is over all pairs of OFDM symbols $\boldsymbol{s}$ and $\boldsymbol{p}$ of $\boldsymbol{\Psi}$ whose labels differ in only 1 bit. The quantities $\{\bar{\lambda}_l(\boldsymbol{s}, \boldsymbol{p})\}$ in (7.4) are the eigenvalues of $\bar{\boldsymbol{Q}}(\boldsymbol{s}, \boldsymbol{p}) = (\boldsymbol{B}_{\boldsymbol{h}}^{\top}) \boldsymbol{Q}(\boldsymbol{s}, \boldsymbol{p}) (\boldsymbol{B}_{\boldsymbol{h}}^{\top})^{\dagger}$, where $\boldsymbol{Q}(\boldsymbol{s}, \boldsymbol{p}) = \sum_{i=1}^{N} (\boldsymbol{t}_i^{\top}(\boldsymbol{s}, \boldsymbol{p}))(\boldsymbol{t}_i^{\top}(\boldsymbol{s}, \boldsymbol{p}))^{\dagger}$, $\boldsymbol{t}_i(\boldsymbol{s}, \boldsymbol{p}) = [\boldsymbol{w}_i^{\top} \boldsymbol{g}_i(\boldsymbol{s} - \boldsymbol{p})^{\top}]^{\top}$ and $\boldsymbol{g}_i$ is the ith row of $\boldsymbol{G}$. A straightforward way to compute the expectation in (7.4) is as follows:

$$\gamma(\Psi, \chi, \boldsymbol{G}) = \frac{1}{Nm 2^{Nm}} \sum_{\boldsymbol{s} \in \Psi} \sum_{k=1}^{Nm} \left[\prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_l(\boldsymbol{s}, \boldsymbol{p})}{4N_0 \sin^2 \theta} \right)^{-1} \right], \tag{7.5}$$

where $\boldsymbol{p}$ is the symbol in $\boldsymbol{\Psi}$ whose label differs in only one bit at the position k compared to the label of $\boldsymbol{s}$.

Since the mapping rule ζ is implemented independently for each subcarrier group of OFDM symbols, instead of averaging over 2^{Nm} signal points of $\boldsymbol{s} \in \boldsymbol{\Psi}$ as in (7.5), it suffices to average over Lm signal points of Φ. That is,

$$\gamma(\boldsymbol{\Psi}, \chi, \boldsymbol{G}) = \frac{1}{V} \sum_{v=1}^{V} \gamma_v(\boldsymbol{\Psi}, \chi, \boldsymbol{G}). \tag{7.6}$$

where

$$\gamma_v(\boldsymbol{\Psi}, \chi, \boldsymbol{G}) = \frac{1}{(Lm) 2^{Lm}} \sum_{\boldsymbol{s}_L \in \Phi} \left[\prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_{v,l}(\boldsymbol{s}_L, \boldsymbol{p}_L)}{4N_0 \sin^2 \theta} \right)^{-1} \right] \tag{7.7}$$

with $\boldsymbol{s}_L$ and $\boldsymbol{p}_L$ are the two signal points in Φ corresponding to the mapping rule ζ, whose labels differ in only 1 bit. The parameters $\bar{\lambda}_{v,l}(\boldsymbol{s}_L, \boldsymbol{p}_L)$ in (7.7) are the

eigenvalues of $\bar{\boldsymbol{Q}}_v(\boldsymbol{s}_L, \boldsymbol{p}_L) = (\boldsymbol{B}_{\boldsymbol{h}}^{\top})\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L)(\boldsymbol{B}_{\boldsymbol{h}}^{\top})^{\dagger}$, where

$$\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L) = \sum_{l=1}^{L} \|\boldsymbol{\Theta}_l(\boldsymbol{s}_L - \boldsymbol{p}_L)^{\top}\|^2 (\boldsymbol{w}_{p_{v,l}}^{\top})(\boldsymbol{w}_{p_{v,l}}^{\top})^{\dagger}. \tag{7.8}$$

Observe that the average in (7.6) depends Φ, ζ and $\boldsymbol{\Theta}$. It is therefore more meaningful to rewrite $\gamma(\boldsymbol{\Psi}, \chi, \boldsymbol{G})$ and $f(d, \boldsymbol{\Psi}, \chi, \boldsymbol{G})$ in (7.4) as $\gamma(\Phi, \zeta, \boldsymbol{\Theta})$ and $f(d, \Phi, \zeta, \boldsymbol{\Theta})$, respectively. Apply (7.6) in (7.4), the function $f(d, \Phi, \zeta, \boldsymbol{\Theta}) = f(d, \boldsymbol{\Psi}, \chi, \boldsymbol{G})$ can be efficiently computed with a single integral.

Using the inequality $Q\left(\sqrt{2\gamma}\right) < \frac{1}{2}\exp(-\gamma)$, one obtains the following design criterion:

$$\delta(\Phi, \zeta, \boldsymbol{\Theta}) = \frac{1}{V(Lm)2^{Lm}} \sum_{v=1}^{V} \sum_{\boldsymbol{s}_L \in \Phi} \left[\prod_{l=1}^{L} \left(1 + \frac{\bar{\lambda}_{v,l}(\boldsymbol{s}_L, \boldsymbol{p}_L)}{4N_0} \right)^{-1} \right]. \tag{7.9}$$

The parameter $\delta(\Phi, \zeta, \boldsymbol{\Theta})$ can be used to characterize the influence of the mapping rule ζ, the subcarrier grouping $\{I_v\}$ and the rotation matrix $\boldsymbol{\Theta}$ to the asymptotic error performance. In particular, for a given constellation, one would prefer the mapping rule, the subcarrier grouping and the rotation matrix that lead to the minimum value of $\delta(\Phi, \zeta, \boldsymbol{\Theta})$. Furthermore, similar to Chapter 6, the design criterion can be made simpler and more meaningful at high SNR and independent of the correlation matrix $\boldsymbol{R}_{\boldsymbol{h}}$ as:

$$\kappa(\Phi, \zeta, \boldsymbol{\Theta}) = \frac{1}{V(Lm)2^{Lm}} \sum_{v=1}^{V} \sum_{\boldsymbol{s}_L \in \Phi} \left[\det \boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L)\right]^{-1}. \tag{7.10}$$

From (7.10), it is expected that the optimal solution for the mapping rule ζ and subcarrier grouping $\{I_v\}$ results in the full-rank matrix $\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L)$, or equivalently, the full diversity L for all pairs of $\boldsymbol{s}_L$ and $\boldsymbol{p}_L$ in Φ whose labels differ in only 1 bit. It was already shown in Chapter 6 that the full diversity order L can be easily achieved when the complex 1-D single-symbol mapping ξ is employed together with the following subcarrier grouping and the rotation matrix:

$$\begin{cases} p_{v,l} = v + V(l-1) \\ \|\theta_{l,c}\| = \frac{1}{\sqrt{L}} \end{cases} \quad \text{for } 1 \le v \le V \text{ and } 1 \le l, c \le L \tag{7.11}$$

Different from Chapter 6, it shall be shown that in addition to the full diversity order L, the use of multi-D mapping provides very impressive coding gains without

the need of implementing the rotation $\mathbf{\Theta}$. To this end, the optimal choices of $\{I_v\}$ to apply the mapping ζ and the mapping rule ζ itself are studied in the next section.

7.3 Optimal Subcarrier Grouping $\{I_v\}$ and Mapping ζ

7.3.1 Optimal Grouping I_v for a given ζ

For a given subgroup I_v and two signal points $\boldsymbol{s}_L$ and $\boldsymbol{p}_L$, the determinant of $\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L)$ in (7.8) is examined by considering $L \times L$ square matrix $\boldsymbol{D}_v$, where each element $D_{c,l}$, $1 \leq c, l \leq L$, is given as:

$$D_{c,l} = d_l \exp\left(-j\frac{2\pi}{N} \cdot (c-1) \cdot p_{v,l}\right) \tag{7.12}$$

with $d_l = \|\mathbf{\Theta}_l(\boldsymbol{s}_L - \boldsymbol{p}_L)^\top\|$, $1 \leq l \leq L$. One can verify that $\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L) = \boldsymbol{D}_v\boldsymbol{D}_v^\dagger$. It then follows that $\det\left(\boldsymbol{Q}_v(\boldsymbol{s}_L, \boldsymbol{p}_L)\right) = \|\det(\boldsymbol{D}_v)\|^2$. The determinant of $\boldsymbol{D}_v$ can be computed as $\det(\boldsymbol{D}_v) = \left(\prod_{l=1}^{L} d_l\right)\det(\boldsymbol{U}_v)$, where $\boldsymbol{U}_v$ is obtained from $\boldsymbol{D}_v$ by setting all the parameters $d_l = 1$. Since $\det(\boldsymbol{U}_v)$ depends only on the subgroup I_v, one would prefer I_v that leads to the maximum value of $\|\det(\boldsymbol{U}_v)\|$, regardless of the choices of Φ, ζ and $\mathbf{\Theta}$. Since $\text{trace}(\boldsymbol{U}_v\boldsymbol{U}_v^\dagger) = L^2$, $\|\det \boldsymbol{U}_v\|^2 \leq L^L$ [82], with the equality holds when the element of I_v satisfies the condition in (7.11), i.e., $p_{v,l} = v + V(l-1)$.

Therefore, one obtains the following lower bound on $\kappa(\Phi, \zeta, \mathbf{\Theta})$ in (7.10):

$$\kappa(\Phi, \zeta, \mathbf{\Theta}) \geq \frac{1}{L^L(Lm)2^{Lm}} \sum_{\boldsymbol{s}_L \in \Phi} \frac{1}{D(\boldsymbol{s}_L, \boldsymbol{p}_L, \mathbf{\Theta})} = \eta(\Phi, \zeta, \mathbf{\Theta}), \tag{7.13}$$

where, as before, $\boldsymbol{s}_L$ and $\boldsymbol{p}_L$ are the two signal points in Φ whose labels differ in only 1 bit and $D(\boldsymbol{s}_L, \boldsymbol{p}_L, \mathbf{\Theta}) = \prod_{l=1}^{L} d_l^2 = \prod_{l=1}^{L} \|\mathbf{\Theta}_l(\boldsymbol{s}_L - \boldsymbol{p}_L)^\top\|^2$ is the squared product distance between the two rotated sequences $\boldsymbol{x}_L$ and $\boldsymbol{y}_L$ of $\boldsymbol{s}_L$ and $\boldsymbol{p}_L$, respectively. Equation (7.13) suggests that for a given mapping rule ζ on Φ and the rotation matrix $\mathbf{\Theta}$, the subcarrier grouping $\{I_v\}$ in (7.11) is the optimal choice.

7.3.2 Optimal Mapping ζ and Rotation Matrix $\mathbf{\Theta}$.

When the subcarrier grouping in (7.11) is chosen, the optimization problem is now related to the parameter $\eta(\Phi, \zeta, \mathbf{\Theta})$ on the right hand side of (7.13). Observe that

$\eta(\Phi, \zeta, \boldsymbol{\Theta})$ is obtained by averaging over the squared product distances of all the signal points $\boldsymbol{x}_L$ and $\boldsymbol{y}_L$ in the rotated constellation Φ_{r} of Φ whose labels differ in only 1 bit. For a given constellation Φ and the rotation matrix $\boldsymbol{\Theta}$, searching techniques can be used to find the most suitable mapping ζ. For example, the binary switching algorithm (BSA) in [40] and the algorithm related to the quadratic assignment problem (QAP) in [36] are good candidates. These methods, however, only provide a locally optimal mapping ζ. Furthermore, they quickly become intractable when dealing with a large constellation, such as the multi-D constellation Φ with a large L.

For a given complex 1-D constellation Ω from which the multi-D constellation Φ is built, the joint optimization problem in designing ζ and the rotation matrix $\boldsymbol{\Theta}$ is even much more complicated because of a huge number of variables. For this reason, attention is restricted to the following constraints, which makes the problem at hand simpler:

i) The complex 1-D QPSK constellation is considered.

ii) The unitary rotation matrix $\boldsymbol{\Theta}$ is employed. Even though it is a small family of the general rotation $\boldsymbol{\Theta}$, the unitary matrix has been studied extensively, especially in connection with SSD techniques [17,55,82]. It is not hard to verify that the optimal rotation matrix considered in Chapters 4, 5, and 6 is also a unitary matrix.

Since QPSK constellation is assumed, the multi-D constellation Φ becomes a hypercube in a complex L-D space [25]. Normalize the energy per QPSK symbol to be 2 and denote the coordinates of QPSK symbols as $[+1, +1]$, $[+1, -1]$, $[-1, +1]$, and $[-1, -1]$.

First, consider the simplest case of $\boldsymbol{\Theta}$, which is $\boldsymbol{\Theta} = \boldsymbol{I}_L$ (i.e., no rotation). In this case, the parameter $\eta(\Phi, \zeta, \boldsymbol{I}_L)$ is similar to the design criterion for multi-D mapping of BICM-ID over a fast fading channel in Chapter 3. Therefore, the optimal mapping proposed in [25] can be readily applied here to achieve the minimum $\eta(\Phi, \zeta, \boldsymbol{I}_L)$. More

specifically, applying the mapping rule in [25], it can be verified that the minimum value of $\eta(\Phi, \zeta, \boldsymbol{I}_L)$ is given by:

$$\begin{aligned} \eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L) &= \frac{1}{L^L}\frac{1}{2L}\left[8^{-L} + (2L-1)\cdot 8^{-L+1}\cdot 4^{-1}\right] \\ &= \frac{8^{-L}}{2\cdot L^{L+1}}\cdot(4L-1). \end{aligned} \tag{7.14}$$

As an example, Table 7.1 shows the optimal mapping rule for the case $L = 2$ [25].

Table 7.1 The optimal 2-D mapping which achieves the minimum value of $\eta(\Phi, \zeta, \boldsymbol{I}_L)$ for the case of no rotation.

Signal	Label	Signal	Label
$[-1,-1,-1,-1]$	0000	$[-1,+1,-1,-1]$	1100
$[-1,-1,-1,+1]$	1001	$[-1,+1,-1,+1]$	0101
$[-1,-1,+1,-1]$	1010	$[-1,+1,+1,-1]$	0110
$[-1,-1,+1,+1]$	0011	$[-1,+1,+1,+1]$	1111
$[+1,-1,-1,-1]$	0111	$[+1,+1,-1,-1]$	1011
$[+1,-1,-1,+1]$	1110	$[+1,+1,-1,+1]$	0010
$[+1,-1,+1,-1]$	1101	$[+1,+1,+1,-1]$	0001
$[+1,-1,+1,+1]$	0100	$[+1,+1,+1,+1]$	1000

For BI-COFDM-ID with SSD in Chapter 6, where the mapping rule ζ becomes the complex 1-D mapping ξ, it is shown that:

$$\eta(\Phi, \zeta, \boldsymbol{\Theta}) \geq \frac{1}{m2^m}\sum_{s\in\Omega}\sum_{j=1}^{m}\left(\|s-p\|^2\right)^{-L}, \tag{7.15}$$

where s and p are the two signal points in the 1-D constellation Ω whose labels differ in only 1 bit. The equality in (7.15) holds when the rotation matrix $\boldsymbol{\Theta}$ and the subcarrier grouping in (7.11) are used. When Ω is a QPSK constellation, it can be shown that the best mapping turns out to be the anti-Gray mapping, which results in the following minimum value of $\eta(\Phi, \zeta, \boldsymbol{\Theta})$:

$$\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta}) = \frac{1}{2\cdot 8^L}(2^L + 1). \tag{7.16}$$

From (7.14) and (7.16), define the following ratio:

$$R = \frac{\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta})}{\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)} = \frac{(2^L + 1) \cdot L^{L+1}}{4L - 3}. \tag{7.17}$$

Observe that when $L > 1$, $R \gg 1$ and a significant coding gain (or one can say mapping gain) can be achieved by the multi-D mapping[2]. Note that the gain is possible even without using the rotation $\boldsymbol{\Theta}$.

From the above evaluations and discussion, it is of interest to further investigate the influence of the unitary matrix $\boldsymbol{\Theta}$ to the asymptotic performance of BI-COFDM-ID with multi-D subcarrier mapping. To this end, the lower bound on $\eta(\Phi, \zeta, \boldsymbol{\Theta})$ for any mapping ζ and unitary rotation $\boldsymbol{\Theta}$ is derived next.

Observe that the Euclidean distance profile of Φ is preserved in the rotated constellation Φ_{r} when the unitary rotation $\boldsymbol{\Theta}$ is employed. Based on the properties of $\boldsymbol{\Theta}$ provided in [25], it can be shown that for any rotated signal point $\boldsymbol{x}_L$, there is only 1 signal point at the largest squared Euclidean distance $8L$, and there are $2L$ signal points at the second largest squared Euclidean distance $8(L-1)+4$ to $\boldsymbol{x}_L$. Therefore, for a given $\boldsymbol{x}_L$, the possible largest and the second largest squared product distances to $\boldsymbol{x}_L$ are 8^L and $(8 - 4/L)^L$, respectively. These facts leads to the following lower bound on $\eta(\Phi, \zeta, \boldsymbol{\Theta})$ for any unitary rotation $\boldsymbol{\Theta}$ and mapping rule ζ:

$$\eta(\Phi, \zeta, \boldsymbol{\Theta}) \geq \frac{1}{2 \cdot L^{L+1}} \left[8^{-L} + (2L - 1)(8 - 4/L)^{-L}\right]. \tag{7.18}$$

Table 7.2 The parameters $\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta})$, $\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$ and the unachievable lower bound.

L	2	3	4	8
$\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta})$	0.0391	0.008	0.002	$7.7\mathrm{e}-6$
$\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$	0.0068	$1.3\mathrm{e}-4$	$1.8\mathrm{e}-6$	$6.9\mathrm{e}-15$
Unachieveable lower bound	0.0062	$1.2\mathrm{e}-4$	$1.5\mathrm{e}-6$	$5.8\mathrm{e}-15$

Unfortunately, as proved in Appendix E, there do not exist any rotation matrix $\boldsymbol{\Theta}$ with $L > 1$, and mapping rule ζ that achieve the equality in (7.18). On the other hand,

[2]It is not surprising to see that if $L = 1$ then $R = 1$, since Φ becomes Ω.

comparing $\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$ in (7.14) and the lower bound in (7.18) shows that the difference is negligible, since the dominant contributions are in the term 8^{-L}. For example, Table 7.2 compares the parameters $\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta})$, $\eta_{\min}(\Phi, \zeta, \mathbf{I}_L)$ and the unachievable lower bound in (7.18) for various values of L. Observe from Table 7.2 that both $\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$ and the lower bound are much smaller than $\eta_{\mathrm{AG}}(\Phi, \zeta, \boldsymbol{\Theta})$. Furthermore, the difference between $\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$ and the lower bound is very small, especially when L increases. These results suggest that for the proposed BI-COFDM-ID with multi-D mapping, one might not need to implement SSD. The multi-D mapping itself can fully exploit the multi-path diversity gain L and provide excellent coding gain (or mapping gain for this matter) without constellation rotation as far as the asymptotic performance is concerned.

As an example to show the advantage of the multi-D mapping without constellation rotation, the following optimal rotation matrix $\boldsymbol{\Theta}$ with $L = 2$ is considered:

$$\boldsymbol{\Theta}_{\pi/4} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \exp(j\frac{\pi}{4}) \\ 1 & -\exp(j\frac{\pi}{4}) \end{pmatrix}. \tag{7.19}$$

The above rotation results in the full-diversity modulation scheme with high coding gains in uncoded systems with SSD [17, 55, 82, 97], and it is also the optimal choice for the systems studied in Chapters 4, 5, and 6. By examining the distance profile of the rotated constellation Φ_{r}, it is observed that for any signal point in Φ_{r}, the four largest squared product distances are 36, 36, 32 and 16. Hence, the following lower bound can be obtained:

$$\eta(\Phi, \zeta, \boldsymbol{\Theta}_{\pi/4}) \geq \frac{1}{16}\left[\frac{1}{36} + \frac{1}{36} + \frac{1}{32} + \frac{1}{16}\right] = 0.0093 \tag{7.20}$$

It is quite straightforward to show that there does not exist any mapping rule that can achieve the above lower bound (the proof is omitted here for brevity of presentation). Using the BSA algorithm in [40] with 10^4 trials, the locally optimal mapping for the rotation $\boldsymbol{\Theta}_{\pi/4}$ is shown in Table 7.3, which yields $\eta(\Phi, \zeta, \boldsymbol{\Theta}_{\pi/4}) = 0.0124$. The value of $\eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L) = 0.0068$ with $L = 2$ (see Table 7.2) is obvious better than both the lower bound and the value of $\eta(\Phi, \zeta, \boldsymbol{\Theta}_{\pi/4})$ obtained by BSA mapping.

Table 7.3 The 2-D mapping obtained with BSA for the unitary rotation $\boldsymbol{\Theta}_{\pi/4}$.

Signal	Label	Signal	Label
$[-1,-1,-1,-1]$	0000	$[-1,+1,-1,-1]$	0110
$[-1,-1,-1,+1]$	1001	$[-1,+1,-1,+1]$	1100
$[-1,-1,+1,-1]$	1101	$[-1,+1,+1,-1]$	1000
$[-1,-1,+1,+1]$	0011	$[-1,+1,+1,+1]$	0101
$[+1,-1,-1,-1]$	0111	$[+1,+1,-1,-1]$	0001
$[+1,-1,-1,+1]$	1010	$[+1,+1,-1,+1]$	1111
$[+1,-1,+1,-1]$	1110	$[+1,+1,+1,-1]$	1011
$[+1,-1,+1,+1]$	0100	$[+1,+1,+1,+1]$	0010

7.4 Illustrative Results

In this section, the QPSK is adopted as the modulation scheme. Unless otherwise stated, the number of channel taps is fixed to $L = 2$ and the number of subcarriers is $N = 64$. In all simulation results, a simple rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ is employed. A random interleaver of length 25,600 coded bits is used. Each point in the BER curves is simulated with 10^7 to 1.5×10^9 coded bits. The channel gains are generated independently for each OFDM symbol and the optimal MAP demodulator is implemented for all systems.

7.4.1 I. I. D. Channels

For the i.i.d. channel model, the correlation matrix is $\boldsymbol{R_h} = 1/L \cdot \boldsymbol{I}_L$. Fig. 7.2 plots the BER performances with 1, 4, 8, and 12 iterations of the proposed system employing the optimal 2-D mapping in Table 7.1 over the optimal subcarrier grouping I_v in (7.11) and without constellation rotation. For comparison, the BER performance after 12 iterations of the conventional BI-COFDM-ID considered in Chapter 6 with subcarrier grouping I_v in (7.11) and the rotation $\boldsymbol{\Theta}_{\pi/4}$ in (7.19) together with anti-Gray mapping is plotted. Also shown in Fig. 7.2 is the BER performance after 12

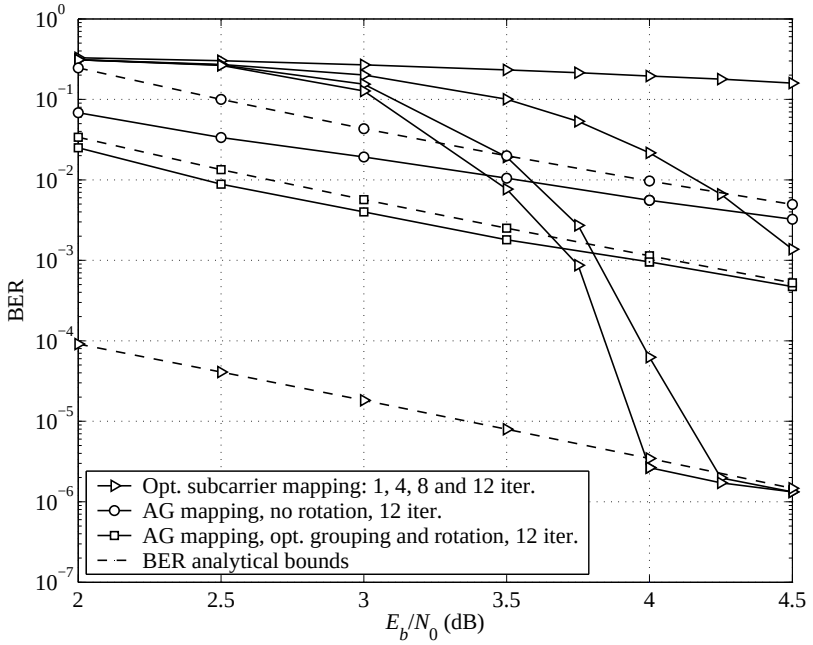


Figure 7.2 Performance of BI-COFDM-ID systems with: i) optimal multi-D subcarrier mapping without rotation; ii) anti-Gray mapping and no rotation; iii) optimal subcarrier grouping and rotation employing anti-Gray mapping (Note: AG stands for anti-Gray).

iterations of BI-COFDM-ID employing anti-Gray mapping and without rotation. The error bounds for all the systems are also provided.

It can be seen from Fig. 7.2 that a significant improvement after 12 iterations can be achieved by the proposed system compared to the BI-COFDM-ID with conventional 1-D anti-Gray mapping. Observe that the error bounds for all systems are very tight at medium and high SNR. Thought not explicitly shown here, by examining the analytical bounds over a wider range of SNR, it is observed that the proposed system offers coding gains as high as 3.7dB and 7.2dB at the BER level of 10^{-6} compared to the system with optimal subcarrier grouping/rotation and anti-Gray mapping, and the system with anti-Gray mapping and without rotation, respectively. Note that the proposed system and the system with subcarrier grouping and SSD have the same receiver complexity, since both perform the same MAP demodulation in the complex 2-D space.

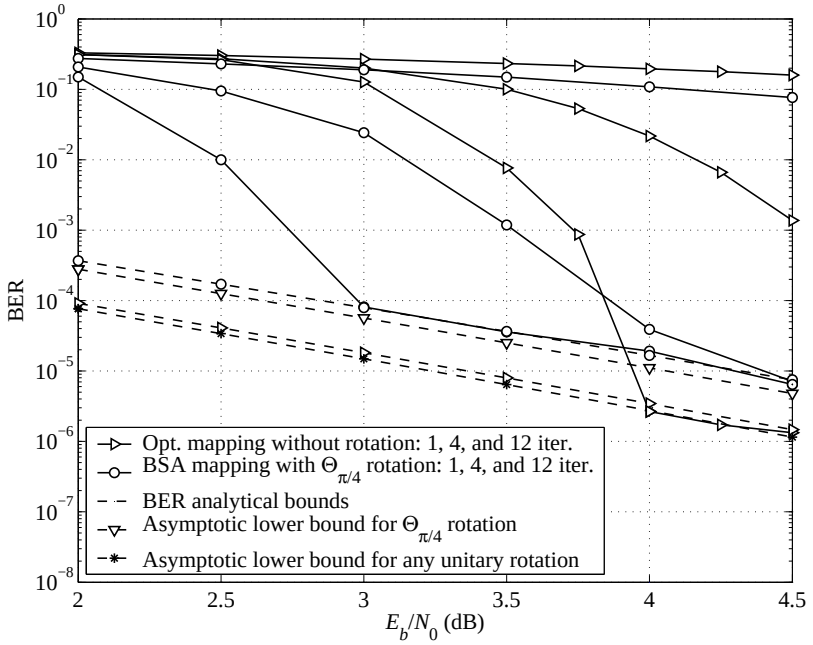


Figure 7.3 Performance of BI-COFDM-ID systems with various multi-D mappings, with and without constellation rotation.

Fig. 7.3 demonstrates the role of the rotation matrix $\mathbf{\Theta}$ on the systems employing multi-D mappings. Shown in Fig. 7.3 are the BER performances after 1, 4, and 12 iterations of the following two systems: (i) the system that employs the optimal mapping in Table 7.1 and without rotation, and (ii) the system that employs the mapping in Table 7.3 and the rotation matrix $\mathbf{\Theta}_{\pi/4}$ in (7.19). The analytical error bounds for both systems are also plotted to show the tightness of the bounds. Furthermore, the two unachievable lower bounds in (7.18) and (7.20) corresponding to any unitary rotation $\mathbf{\Theta}$ and the rotation in (7.19), respectively, are provided for comparison[3].

It can be observed from Fig. 7.3 that all the BER performances converge to the corresponding error bounds at medium and high SNR. The BER performance of the proposed BI-COFDM-ID with multi-D mapping and without rotation, closely approaches the unachievable lower bound of any unitary rotation $\mathbf{\Theta}$. The proposed system without rotation outperforms those using the rotation $\mathbf{\Theta}_{\pi/4}$ in terms of the

[3]It should be pointed out that with the bounds in (7.18) and (7.20), these lower bounds are computed from (2.20), (7.4), and (7.6).

asymptotic performance. The advantage of the system with $\mathbf{\Theta}_{\pi/4}$ rotation is its better convergence behavior, where the BER performance after 1 iteration is better than that of the system without rotation. Its BER performance also approaches to the error bound faster, as can be seen from Fig. 7.3. Examining the bounds over a wider range of SNR indicates that gap between the asymptotic performance of the system employing the optimal mapping without rotation and the unachievable lower bound for unitary rotation $\mathbf{\Theta}$ is only 0.12dB at BER level of 10^{-6}. Furthermore, compared to the system with optimal grouping, BSA mapping, and $\mathbf{\Theta}_{\pi/4}$ rotation, the coding gain is about 1dB. This again confirms the analysis made in the previous section about the unnecessarily of implementing LCP as far as minimizing the asymptotic performance is concerned.

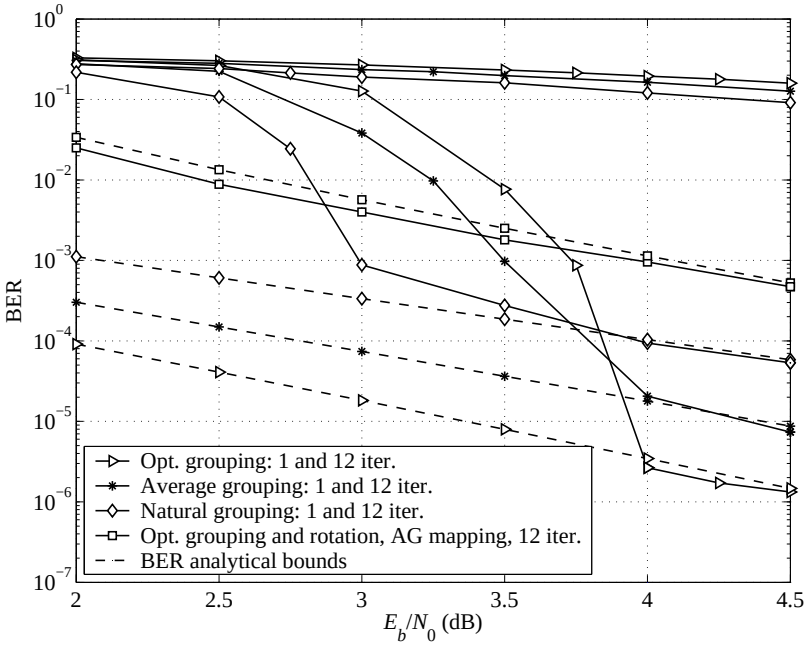


Figure 7.4 Performance of BI-COFDM-ID systems that employ the optimal multi-D mapping over various subcarrier grouping approaches and without rotation (Note: AG denotes the conventional anti-Gray mapping).

Fig. 7.4 illustrates the performance of the systems where the optimal mapping in Table 7.1 is employed over various subcarrier grouping approaches, and without constellation rotation. As similar to Chapter 6, besides the optimal subcarrier grouping in

(7.11), two more different approaches, namely natural grouping with $I_v = (2v-1, 2v)$ and average grouping with $I_v = (v, N+1-v)$, are considered. The error performances of the corresponding systems with 1 and 12 iterations are shown. Also in this figure, the error performance after 12 iterations of the conventional BI-COFDM-ID with the optimal subcarrier grouping and the rotation $\boldsymbol{\Theta}_{\pi/4}$ is also provided as a reference. Observe again that the error bounds of all the systems are very tight. It can be seen that the use of the optimal subcarrier grouping results in significant coding gains compared to the suboptimal ones. Furthermore, as can be seen from Fig. 7.4, high coding gains can still be achieved with the suboptimal subcarrier grouping methods compared to the BI-COFDM-ID that implements SSD and the conventional 1-D anti-Gray mapping.

Simulation results, which are not presented here for brevity of presentation, indicate that the proposed technique with a lower decoding complexity has a similar performance as the classical BI-COFDM using a turbo code.[4] Similar observation was made in [23] for MIMO *frequency non-selective* fading channels. Furthermore, the proposed technique can also be applied to systems that employ turbo codes and whose receiver complexity is not the main concern.

7.4.2 Correlated Channels

Fig. 7.5 compares the BER performance of various BI-COFDM-ID systems over a correlated channel, whose correlation matrix is assumed to be $\boldsymbol{R_h} = \frac{1}{L}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. As a reference, the performance bound for the proposed system with optimal 2-D mapping and optimal subcarrier grouping and without constellation rotation over the i.i.d channel is also shown. It can be seen from Fig. 7.5 that a significant improvement after 12 iterations can be achieved by the proposed system compared to the BI-COFDM-ID with conventional 1-D anti-Gray mapping, which is very similar to the case of i.i.d. channel. Due to the correlation in the channel gains, performance

[4]Note that, since each turbo-decoding iteration involves two component SISO decoders, the decoding complexity of the classical BI-COFDM using a turbo code is higher

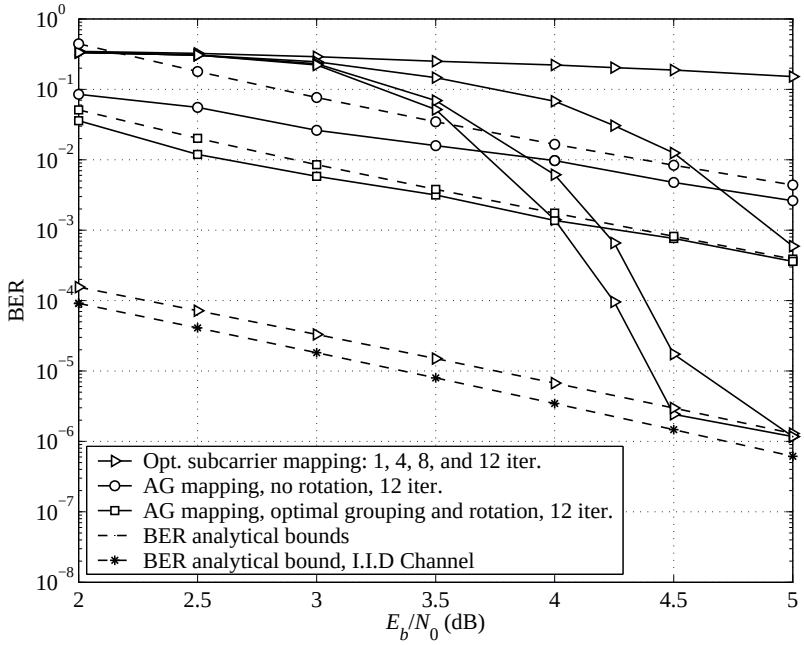


Figure 7.5 Performance of BI-COFDM-ID systems over correlated channel with: i) optimal multi-D subcarrier mapping without rotation; ii) anti-Gray mapping and no rotation; iii) optimal subcarrier grouping and rotation employing anti-Gray mapping (Note: AG stands for anti-Gray).

degradation is experienced compared to the case of the i.i.d channel.

7.4.3 Further Results for a Larger L

This subsection considers the 4-tap i.i.d channel model where the tap variances are similar to that of the pedestrian International Telecommunication Union (ITU) channel model, which are $\{0.8893, 0.0953, 0.0107, 0.0047\}$ [98]. For the consistency with analysis, it is assumed that there is no tap-correlation over time. Since $L = 4$, the multi-D mapping is employed in a complex 4-D space. Fig. 7.6 presents the BER performances with 1, 4, 8, and 12 iterations of the proposed system employing the optimal 4-D mapping and without constellation rotation. For comparison, the BER performance after 12 iterations of the conventional BI-COFDM-ID considered in Chapter 6 with the 4×4 optimal rotation matrix given in (4.22) and 1-D anti-Gray mapping is also plotted. The BER performance after 12 iterations of BI-COFDM-

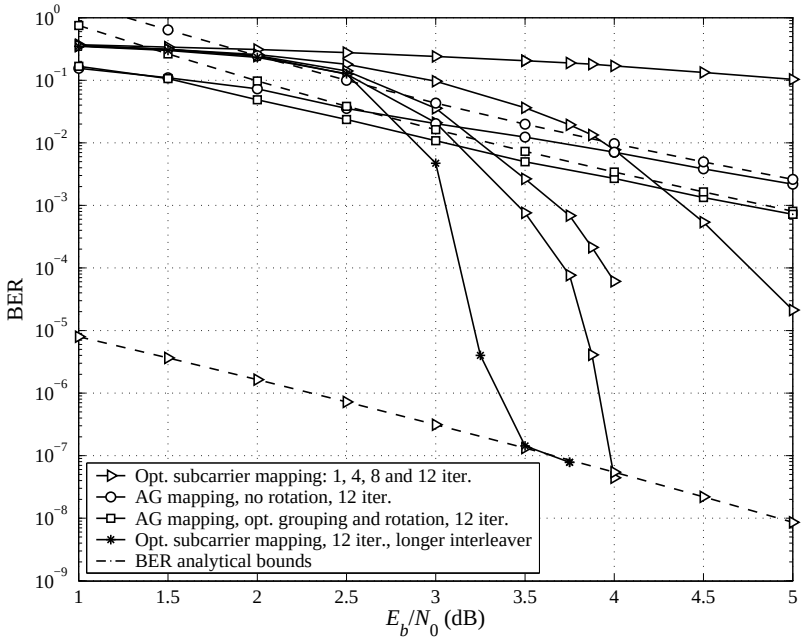


Figure 7.6 Performance of BI-COFDM-ID systems with: i) optimal multi-D subcarrier mapping without rotation; ii) anti-Gray mapping and no rotation; iii) optimal subcarrier grouping and rotation employing anti-Gray mapping for the case of $L = 4$.

ID employing anti-Gray mapping and without rotation is further provided as the reference. Similar to previous observations, it is clear from Fig. 7.6 that the proposed system provides significant mapping gains. However, it is seen that the convergence to the analytical bound happens at a lower BER level of 10^{-7}. A faster convergence behavior can be achieved by using a longer interleaver. This fact can be verified from Fig. 7.6, which shows the BER performance after 12 iterations when the interleaving length is 102,400 coded bits.

8. Coded Unitary Space-Time Modulation with Iterative Decoding: Error Performance and Mapping Design[1]

From Chapter 3 to Chapter 7, different multi-dimensional BICM-ID frameworks have been studied, where the constellation $\mathbf{\Psi}$ and the mapping rule ξ might span over either time or frequency. It was demonstrated that a multi-dimensional approach provides a more efficient and flexible design over a single-dimensional scheme. In this chapter, the application of BICM-ID is considered for a system where the constellation $\mathbf{\Psi}$ and the mapping rule ξ span over both time and space, i.e., a multiple-antenna system.

As discussed in Chapter 2, currently, major research efforts have been made to multiple-input multiple-output (MIMO) systems that employ multiple antennas at both the transmitter and the receiver. It is due to the fact that MIMO systems can provide a significant improvement for a wireless communications link, with respect to both the reliability of the link and the data rate, especially in Rayleigh fading environment. Based on the assumption of perfect channel state information (CSI) at the receiver (but not at the transmitter), the theoretical framework in [101] shows that MIMO systems can achieve a much higher channel capacity compared to that of single-antenna counterparts. Assuming perfect information of fading coefficients at the receiver, two space-time coding methods, namely space-time trellis codes and space-time block codes were introduced in [18,19] to obtain the diversity gains in both time and spatial domains. Since then, a considerable amount of work has studied

[1] A part of this chapter was presented in [99, 100]

both uncoded and coded MIMO systems, including the BICM framework, under this assumption.

Unfortunately, due to the rapidly changing conditions in fading environment, perfect knowledge of the fading coefficients might not be available at the receiver. In [102], a novel treatment to compute the capacity of the wireless channel where neither the transmitter nor the receiver knows the fading coefficients has been performed. It was shown in [102,103] that the class of unitary space-time constellations is the most suitable constellation with respect to the channel capacity for this type of channels. Furthermore, by evaluating the asymptotic union bound (AUB), it was pointed out in [104] that the use of unitary signal constellations results in the minimization of the AUB. Various unitary constellation designs have been presented [105–108]. The design criterion used in these studies was mainly based on the pairwise error probability proposed in [103] for uncoded systems, which is related to the complex Grasmannian space as opposed to the Euclidean space in the case of coherent channels.

The incorporation of channel coding in the system has also been studied in the literature. This combination was first considered in [109] where a turbo code is concatenated with the unitary signal constellation given in [105]. Simulation results in [109] indicated that the error performance can be significantly improved as compared to the uncoded systems. More recently, effects of signal mapping on the error performance of coded MIMO systems with BICM-ID were studied in [110]. However, no analytical evaluation was provided and only a few signal mappings obtained by computer search were investigated. A similar work considering trellis coded modulation (TCM) and unitary constellation was also investigated in [108]. By a clever construction of the unitary constellation based on orthogonal design, the asymptotic performance was analytically quantified in [108]. The design in [108] was limited to a number of transmit antennas $N_t \leq 4$ because of the required orthogonality. The number of block intervals of the constellation under consideration is $T = 2N_t$.

Motivated from the above observations and discussions, this chapter analytically studies a unitary bit-interleaved space-time coded modulation with iterative decoding

system. In particular, the asymptotic bit error probability (BEP) for the system under consideration is first evaluated. Similar to other BICM-ID frameworks, the evaluation is based on the assumption of ideal feedback from the channel decoder to the demodulator and can be applied for any given unitary constellation. Simulation results of the bit-error-rate performance confirms that the derived asymptotic BEP is very tight at high signal to noise ratio (SNR). Based on the asymptotic performance, design criteria for the choice of unitary constellation and mapping are also established. Using the class of unitary constellations obtained from orthogonal designs with 4-PSK and 8-PSK signal sets [108], the optimal mappings with respect to the asymptotic performances of systems with and without iterative processing are then introduced. It is shown that the use of the proposed mappings for iterative systems results in error performance that is very close to the lower bound of the asymptotic performance for any unitary constellation and mapping.

8.1 System Model

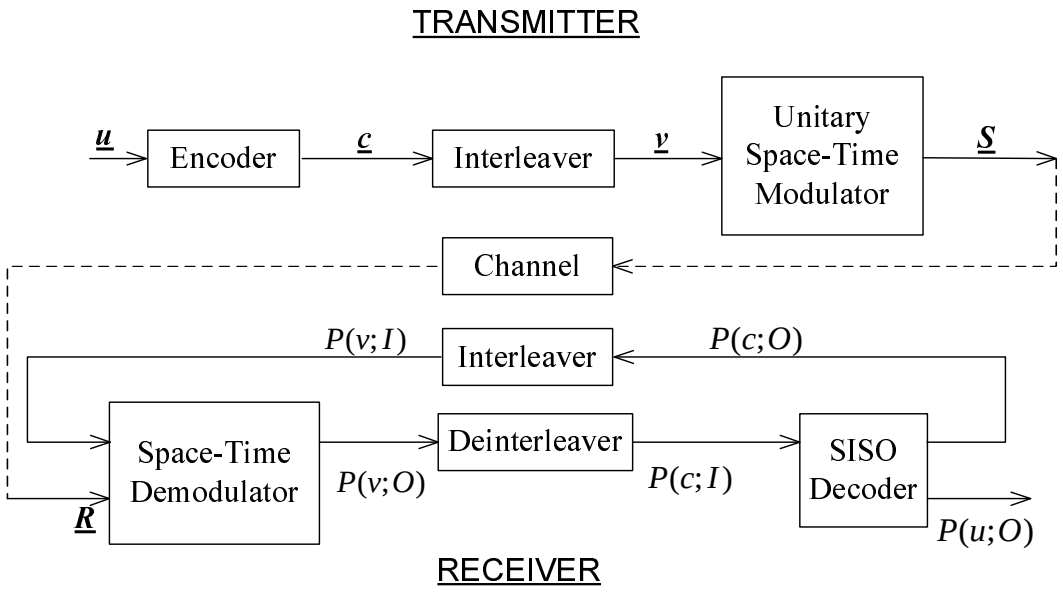


Figure 8.1 Block diagram of a unitary space-time coded modulation with iterative decoding system.

The general block diagram of a unitary space-time coded modulation system with

iterative decoding is shown in Fig. 8.1. The information sequence $\underline{\boldsymbol{u}}$ is first encoded into a coded sequence $\underline{\boldsymbol{c}}$ using a rate-k_c/n_c convolutional code. The coded sequence $\underline{\boldsymbol{c}}$ is then interleaved by a bit-wise interleaver to become the interleaved sequence $\underline{\boldsymbol{v}}$. Based on the mapping rule ξ, each block of M_c coded bits is mapped to one of K matrices $\boldsymbol{\Psi}(\kappa)$, $1 \leq \kappa \leq K$. Here, each matrix $\boldsymbol{\Psi}(\kappa)$, $1 \leq \kappa \leq K$, is a $T \times N_t$ unitary matrix that satisfies the condition $\boldsymbol{\Psi}^{\dagger}(\kappa)\boldsymbol{\Psi}(\kappa) = \boldsymbol{I}_{N_t}$, $1 \leq \kappa \leq K$, with $(\cdot)^{\dagger}$ denotes conjugate transpose and $\boldsymbol{I}_{N_t}$ is the identity matrix of size $N_t \times N_t$. Note that N_t is the number of transmit antennas, T is the number of rows of a space-time code matrix, and $M_c = \log_2 K$. The set $\{\boldsymbol{\Psi}(\kappa)\}_{\kappa=1}^{K}$ is simply referred to as the constellation $\boldsymbol{\Psi}$. The group of M_c coded bits carried by each signal point $\boldsymbol{\Psi}(\kappa)$ is the label of $\boldsymbol{\Psi}(\kappa)$. Assume that the receiver is equipped with N_r antennas. Following the same notations in [102], over a *frequency non-selective* Rayleigh fading channel, at the time index e, the $T \times N_r$ matrix $\boldsymbol{R}_e$ of received signals corresponding to the transmitted signal $\boldsymbol{X}_e$ can be written as:

$$\boldsymbol{R}_e = \sqrt{\frac{\rho T}{N_t}}\boldsymbol{X}_e\boldsymbol{H} + \boldsymbol{W}, \tag{8.1}$$

where $\boldsymbol{X}_e$ is one of K unitary matrices in the constellation $\boldsymbol{\Psi}$. The matrix $\boldsymbol{H}$ is a $N_t \times N_r$ matrix of the fading coefficients whose entries are $\mathcal{CN}(0,1)$. The matrix $\boldsymbol{W}$ is a $T \times N_r$ matrix representing additive white Gaussian noise whose entries are also $\mathcal{CN}(0,1)$. The normalization factor $\sqrt{\frac{\rho T}{N_t}}$ in (8.1) ensures that the average SNR at each receive antenna is ρ, independent of N_t. The spectral efficiency of the system is therefore $M_c k_c/(n_c T)$ information bits per channel use.

The receiver of the system includes the maximum *a posteriori* (MAP) soft-output space-time demodulator and the soft-input soft-output (SISO) channel decoder. The detailed algorithm for the demodulator can be followed from the general method presented in Chapter 2, with the conditional probability density function of the received signal is given as [102, 103]:

$$p(\boldsymbol{R}_e|\boldsymbol{X}_e) = \frac{\exp\left(-\mathrm{tr}\{\boldsymbol{\Lambda}_e^{-1}\boldsymbol{R}_e\boldsymbol{R}_e^{\dagger}\}\right)}{\pi^{TN_r}(\det(\boldsymbol{\Lambda}_e))^{N_r}} \tag{8.2}$$

In (8.2), "tr" denotes the trace function and $\boldsymbol{\Lambda}_e = \boldsymbol{I}_T + \left(\frac{\rho T}{N_t}\right)\boldsymbol{X}_e\boldsymbol{X}_e^{\dagger}$. The SISO channel decoder uses the MAP algorithm in [22]. Similar to other BICM-ID systems, there

is an iterative process between the space-time demodulator and the channel decoder to exchange the information. The total *a posteriori* probabilities of the information bits can be computed to make hard decisions at the output of the decoder after each iteration.

8.2 Performance Evaluation

The union bound on the bit error probability (BEP) for a rate-k_c/n_c convolutionally coded unitary space-time modulation can be re-written in a general form as in (2.20):

$$P_b \leq \frac{1}{k_c} \sum_{d=d_H}^{\infty} c_d f(d, \boldsymbol{\Psi}, \xi). \tag{8.3}$$

Similar to the analysis in the previous chapters, one needs to compute the function $f(d, \boldsymbol{\Psi}, \xi)$, which is the average pairwise error probability. This function depends on the Hamming distance d, the unitary constellation $\boldsymbol{\Psi}$ and the mapping rule ξ. In the following, the function $f(d, \boldsymbol{\Psi}, \xi)$ is computed from the pairwise error probability (PEP) of two codewords.

Also using the same notations as in the previous chapters, let $\underline{\boldsymbol{c}}$ and $\underline{\breve{\boldsymbol{c}}}$ denote the input and decoded sequences, respectively, with Hamming distance d between them and they differ in the first d consecutive bits. Furthermore, with the use of a sufficiently long interleaver, it can be assumed that each of these d different bits appears in a different block of M_c bits before mapping to the constellation symbols. Therefore, the binary sequences $\underline{\boldsymbol{c}}$ and $\underline{\breve{\boldsymbol{c}}}$ correspond to codewords $\underline{\boldsymbol{S}}$ and $\underline{\breve{\boldsymbol{S}}}$, defined as sequences of d unitary signal points scaled by $\sqrt{T}$, namely $\underline{\boldsymbol{S}} = \sqrt{T}\,[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d]$ and $\underline{\breve{\boldsymbol{S}}} = \sqrt{T}\left[\breve{\boldsymbol{X}}_1, \ldots, \breve{\boldsymbol{X}}_d\right]$. Here, $\boldsymbol{X}_e$ and $\breve{\boldsymbol{X}}_e$, $1 \leq e \leq d$, belong to the constellation $\boldsymbol{\Psi}$. Assuming the channel is memoryless and following similar analysis as in [103], the PEP between two codewords $\underline{\boldsymbol{S}}$ and $\underline{\breve{\boldsymbol{S}}}$ can be computed as follows:

$$P(\underline{\boldsymbol{S}} \rightarrow \underline{\breve{\boldsymbol{S}}}) = P\left[\sum_{e=1}^{d} \mathrm{tr}(\boldsymbol{R}_e^{\dagger} \breve{\boldsymbol{X}}_e \breve{\boldsymbol{X}}_e^{\dagger} \boldsymbol{R}_e) > \sum_{e=1}^{d} \mathrm{tr}(\boldsymbol{R}_e^{\dagger} \boldsymbol{X}_e \boldsymbol{X}_e^{\dagger} \boldsymbol{R}_e)\right]. \tag{8.4}$$

Applying the closed-form analytical expression presented in [103], the above PEP can

be computed based on a one-dimensional integral as:

$$P(\underline{\boldsymbol{S}} \to \breve{\underline{\boldsymbol{S}}}) = -\frac{1}{2\pi j}\int_{-\infty}^{\infty} \frac{1}{w+j/2}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}w, \tag{8.5}$$

where

$$\Delta_e = \prod_{m=1}^{N_t}\left[\frac{1}{1+\frac{(\rho T/N_t)^2(1-d_{m,e}^2)(w^2+1/4)}{1+\rho T/N_t}}\right]^{N_r}. \tag{8.6}$$

In (8.6), $0 \le d_{m,e} \le 1$, with $1 \le m \le N_t$, is the mth singular value of the $N_t \times N_t$ matrix $\breve{\boldsymbol{X}}_e^{\dagger}\boldsymbol{X}_e$. Since $P(\underline{\boldsymbol{S}} \to \breve{\underline{\boldsymbol{S}}})$ is real-valued, by taking the real-part of (8.5) gives [103]:

$$P(\underline{\boldsymbol{S}} \to \breve{\underline{\boldsymbol{S}}}) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{1}{w^2+1/4}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}w. \tag{8.7}$$

Owing to the success of iterative decoding steps as normally seen in BICM-ID systems, one is most interested in the asymptotic performance to which the iterations converge. Such asymptotic performance can be obtained by assuming that the iterations between the SISO decoder and the demodulator work perfectly, or equivalently, one has perfect *a priori* information of the coded bits fed-back to the demodulator. Consider the kth bit position, $1 \le k \le M_c$. With the ideal knowledge of the other coded bits carried by the transmitted symbol $\boldsymbol{X}_e$, it can be seen that the error happens only when the labels of $\boldsymbol{X}_e$ and $\breve{\boldsymbol{X}}_e$ differ at the kth bit position. Furthermore, observe that Δ_e are i.i.d. random variables. Then, instead of averaging over all pairs of codewords $\underline{\boldsymbol{S}}$ and $\breve{\underline{\boldsymbol{S}}}$, the union bound on $f(d, \boldsymbol{\Psi}, \xi)$ can be computed by averaging over the constellation $\boldsymbol{\Psi}$ as follows:

$$f(d, \boldsymbol{\Psi}, \xi) \le \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{1}{w^2+1/4}\left[E\left\{\Upsilon(\kappa,k)\right\}\right]^d \mathrm{d}w, \tag{8.8}$$

where the expectation in (8.8) denotes the average over all K signal points $\boldsymbol{\Psi}(\kappa)$, $1 \le \kappa \le K$, as well as all the bit positions, $1 \le k \le M_c$. That is,

$$E\left\{\Upsilon(\kappa,k)\right\} = \frac{1}{KM_c}\sum_{\kappa=1}^{K}\sum_{k=1}^{M_c}\Upsilon(\kappa,k). \tag{8.9}$$

The quantity $\Upsilon(\kappa, k)$ in (8.8) and (8.9) is defined for each pair (κ, k) as follows:

$$\Upsilon(\kappa, k) = \prod_{m=1}^{N_t} \left[\frac{1}{1 + \frac{(\rho T/N_t)^2 [1 - d_m^2(\kappa, j(\kappa,k))](w^2+1/4)}{1+\rho T/N_t}} \right]^{N_r}, \tag{8.10}$$

where $d_m(l, j(\kappa, k))$ is the mth singular value of $\mathbf{\Psi}^\dagger(\kappa)\mathbf{\Psi}(j(\kappa, k))$. Here, with a slight abuse of notation, $j(\kappa, k)$ returns the index of the unitary signal point in $\mathbf{\Psi}$ whose label differs at only the kth position compared to the label of $\mathbf{\Psi}(\kappa)$. Note that the expression of $\Upsilon(\kappa, k)$ is similar to that of Δ_e in (8.6). The major difference, however, is that $\Upsilon(\kappa, k)$ is associated to a specific unitary signal point $\mathbf{\Psi}(\kappa)$ and a specific label position k, rather than the time index. The above analysis thus shows that the function $f(d, \mathbf{\Psi}, \xi)$ can be easily calculated in determining the asymptotic performance.

To give an insight on how to design good mappings, the simpler Chernoff upper bound in [103] is applied to $P(\underline{\boldsymbol{S}} \to \underline{\breve{\boldsymbol{S}}})$ in (8.7) to yield:

$$P(\underline{\boldsymbol{S}} \to \underline{\breve{\boldsymbol{S}}}) \sim \frac{1}{2} \prod_{e=1}^{d} \prod_{m=1}^{N_t} \left[\frac{1}{1 + \frac{(\rho T/N_t)^2 (1 - d_{m,e}^2)}{4(1+\rho T/N_t)}} \right]^{N_r}. \tag{8.11}$$

Similar to the analysis presented earlier, the function $f(d, \mathbf{\Psi}, \xi)$ can then be approximated as:

$$f(d, \mathbf{\Psi}, \xi) \sim \frac{1}{2} \left[\delta(\mathbf{\Psi}, \xi) \right]^d, \tag{8.12}$$

where

$$\delta(\mathbf{\Psi}, \xi) = \frac{1}{K M_c} \sum_{\kappa=1}^{K} \sum_{k=1}^{M_c} \left[\prod_{m=1}^{N_t} \frac{1}{1 + \frac{(\rho T/N_t)^2 [1 - d_m^2(\kappa, j(\kappa,k))]}{4(1+\rho T/N_t)}} \right]^{N_r}. \tag{8.13}$$

Obviously, the parameter $\delta(\mathbf{\Psi}, \xi)$ characterizes the effect of the unitary constellation $\mathbf{\Psi}$ and the mapping rule ξ to the asymptotic performance of a unitary space-time coded modulation with iterative decoding. More specifically, the smaller this parameter is, the better the asymptotic performance becomes.

The design criterion can be made simpler and more meaningful with the observation that the Chernoff bound depends dominantly on the large SNR as suggested in [105]. At high SNR (i.e., $\rho \to \infty$), using the approximation $(1 + x)^{-1} \approx x^{-1}$ when

$x \to \infty$ in (8.13) and cancelling common factors, the design parameter for a given number of receive antennas, N_r, is then modified to:

$$\gamma(\mathbf{\Psi}, \xi) = \frac{1}{KM_c} \sum_{\kappa=1}^{K} \sum_{k=1}^{M_c} \left[\prod_{m=1}^{N_t} [1 - d_m^2(\kappa, j(\kappa, k))] \right]^{-N_r}. \tag{8.14}$$

Observe that, the design criterion under consideration is far different from that of a coherent system, which is mainly related to the Euclidean distances. The parameter $\gamma(\mathbf{\Psi}, \xi)$ does not depend on SNR and it needs to be minimized.

As far as the asymptotic performance is concerned, it follows from (8.14) that by using the constellation and mapping such that all the singular values $d_m(\kappa, j(\kappa, k)) = 0$, the minimum value of $\gamma(\mathbf{\Psi}, \xi) = 1$ and, consequently, the lowest asymptotic error performance is attained. This lowest asymptotic performance therefore can serve as the lower bound of the asymptotic performance for any unitary constellation and mapping scheme. As an example, the minimum value of $\gamma(\mathbf{\Psi}, \xi) = 1$ can be achieved by using the unitary constellation $\mathbf{\Psi}$ where all K signal points collapse into only two distinct unitary matrices whose all columns are orthogonal. The mapping rule ξ is implemented such that one unitary matrix corresponds to $K/2$ signal points whose labels have even Hamming weights. The other unitary matrix corresponds to the remaining $K/2$ signal points whose labels have odd Hamming weights. Unfortunately, despite the potential of giving the best asymptotic error performance, the use of this constellation and mapping will result in a very poor performance at any SNR. This is because the ambiguity in this constellation and mapping leads to a very poor error performance after the first iteration and, as a consequence, the potential asymptotic error performance can never be achieved regardless of how high the SNR can be. This issue is elaborated in detail when the performance of the system after the first iteration is considered next. Nevertheless, the above ambiguity does not happen with common and well-designed unitary constellations, such as the ones obtained from orthogonal design adopted in this chapter.

The above evaluation concentrates only on the asymptotic performance with the perfect *a priori* information of coded bits at the input of the demodulator. For systems

implementing iterative decoding and demodulation, it is also of interest to study the performance after the first iteration (i.e., without feedback from the decoder to the demodulator). The analysis based on the worst case of PEP can be used as similar to the previous chapters. However, this method cannot take into account the effect of mapping. Instead, similar to [2], with no *a priori* information about the coded bits at the input of the demodulator, the computation of the PEP $P(\underline{\boldsymbol{S}} \rightarrow \breve{\underline{\boldsymbol{S}}})$ in (8.4) is now based on the assumption that $\boldsymbol{X}_e$ and $\breve{\boldsymbol{X}}_e$ are the two "nearest" neighbors, as opposed to the assumption that $\boldsymbol{X}_e$ and $\breve{\boldsymbol{X}}_e$ are the two signal points whose labels differ in only 1 bit made earlier. This is a reasonable assumption, since at high SNR, the decoding error likely happens between the two "nearest" neighbors. Because the study in [2] considers coherent systems, the distances between signal points are simply the Euclidean distances. In the case of noncoherent systems, the analogous distance is the chordal distance [103, 105]. More specifically, the chordal distance between the two signal points $\boldsymbol{\Psi}(\kappa)$ and $\boldsymbol{\Psi}(\kappa')$ in $\boldsymbol{\Psi}$ is given as [103, 105]:

$$d^c(\kappa, \kappa') = \sqrt{\sum_{m=1}^{N_t} [1 - d_m^2(\kappa, \kappa')]}, \tag{8.15}$$

where $\{d_m(\kappa, \kappa')\}_{m=1}^{N_t}$ are the N_t singular values of $\boldsymbol{\Psi}^\dagger(\kappa)\boldsymbol{\Psi}(\kappa')$. For each signal $\boldsymbol{\Psi}(\kappa) \in \boldsymbol{\Psi}$, let I_κ^k denote the set of all the nearest signal points, in terms of the chordal distance, to $\boldsymbol{\Psi}(\kappa)$ whose labels differ at the kth position compared to the label of $\boldsymbol{\Psi}(\kappa)$. Note that the set I_κ^k generally contains more than one signal points of $\boldsymbol{\Psi}$. This is because the label of a signal point in I_κ^k might differ in positions other than the kth position compared to the label of $\boldsymbol{\Psi}(\kappa)$. Following the same steps presented before, one can obtain the following parameter that characterizes the performance of the system without iterative decoding at high SNR:

$$\varphi(\boldsymbol{\Psi}, \xi) = \frac{1}{LM_c} \sum_{\kappa=1}^{K} \sum_{k=1}^{M_c} g(\kappa, k), \tag{8.16}$$

where

$$g(\kappa, k) = \sum_{\kappa' : \boldsymbol{\Psi}(\kappa') \in I_\kappa^k} \left[\prod_{m=1}^{N_t} [1 - d_m^2(\kappa, \kappa')] \right]^{-N_r}. \tag{8.17}$$

In summary, for a given constellation $\mathbf{\Psi}$, the sensible design criterion for the systems without iterative processing is to choose the mapping ξ which results in the smallest value of $\varphi(\mathbf{\Psi}, \xi)$. Furthermore, comparing the two parameters $\varphi(\mathbf{\Psi}, \xi)$ in (8.16) and $\gamma(\mathbf{\Psi}, \xi)$ in (8.14) gives the coding gain provided by the iterative processing. For mappings that yield the same value of these two parameters, there is no benefit to run the iteration between the decoder and the demodulator. Also, for the ambiguous constellation and mapping discussed earlier, it can be verified that $\varphi(\mathbf{\Psi}, \xi) = \infty$. This implies a very poor performance after the first iteration and the asymptotic performance promised by the ambiguous constellation/mapping cannot be practically approached.

Given a unitary constellation, brute-force search can be carried out to find the optimal mappings based on the established distance criteria. However, when the size of the constellation is large, this exhaustive search becomes infeasible due to the complexity. The binary switching algorithm (BSA) proposed in [40] can then be applied. Application of BSA to find good signal mappings was recently investigated in [110] for the specific unitary constellation proposed in [105]. However, the BSA only gives locally optimal mappings in general. Furthermore, it is observed that almost all good unitary constellations presented so far in the literature are based on numerical optimization. Due to the lack of the structure of the unitary constellations, it makes the mapping design problem intractable in general.

From the above discussion, a simple but yet effective construction of unitary constellations based on orthogonal designs proposed in [108] shall be adopted to investigate the mapping problem. More specifically, by investigating properties of the singular values for the unitary constellations obtained from orthogonal designs with 4-PSK and 8-PSK signal sets [108], the optimal and good mappings for both systems, with and without iterative processing, are introduced. Interestingly, the results obtained here are inherited from the mapping designs proposed for multi-dimensional constellations of BICM-ID in coherent channels in Chapter 3 and in [25]. Furthermore, it is shown that for systems with iterative processing, using the proposed mappings

makes it possible to closely approach the lower bound on the asymptotic performance of any unitary constellation and mapping mentioned earlier.

8.3 Unitary Constellations from Orthogonal Designs and Proposed Mappings

In this section, the construction of unitary constellations from orthogonal designs and conventional M-PSK signal set proposed in [108] is adopted to investigate the mapping problem. Moreover, the cases of $M = 4$ and $M = 8$ are of particular interest.

8.3.1 Unitary Constellations from Orthogonal Designs

An important class of unitary constellations corresponds to $T = 2N_t$, which has been studied intensively in the literature. For example, reference [111] shows that given a coherence time T, the maximum number of degrees of freedom is achieved by using $T/2$ transmit antennas. Furthermore, it is demonstrated in [112] that the condition $T \geq 2N_t$ guarantees full antenna diversity. The differential unitary space-time code proposed in [112, 113] is also a special case of the unitary space-time structure with $T = 2N_t$. Other research on unitary space-time codes with $T = 2N_t$ can be found in [114] and the references therein.

The orthogonal construction introduced in [108] also concentrates on this case, with a very simple but yet effective design of unitary constellations. When $N_t = 2$, given the number of signal points $K = M^2$, where M is a positive integer, the construction procedure is as follows. For an integer number κ, $1 \leq \kappa \leq K$, define $i = (\kappa - 1)$ div M and $p = (\kappa - 1)$ mod M. Then the 4×2 unitary signal point $\mathbf{\Psi}(\kappa)$ is [108]:

$$\mathbf{\Psi}(\kappa) = \frac{1}{2} \begin{bmatrix} 1 & -1 & \exp\left(j\frac{2\pi}{M}i\right) & -\exp\left(-j\frac{2\pi}{M}p\right) \\ 1 & 1 & \exp\left(j\frac{2\pi}{M}p\right) & \exp\left(-j\frac{2\pi}{M}i\right) \end{bmatrix}^{\top}. \tag{8.18}$$

Observe that the elements $\exp\left(\pm j\frac{2\pi}{M}i\right)$ and $\exp\left(\pm j\frac{2\pi}{M}p\right)$ belong to an M-PSK constellation. As shown in [108], for two unitary signal points $\mathbf{\Psi}(\kappa)$ and $\mathbf{\Psi}(\kappa')$ that correspond to two pairs $[i, p]$ and $[i', p']$, the two singular values of $\mathbf{\Psi}^{\dagger}(\kappa)\mathbf{\Psi}(\kappa')$ are

equal and given by:

$$d_1(\kappa,\kappa') = d_2(\kappa,\kappa') = \frac{\sqrt{2+\cos\frac{2\pi}{M}(i-i')+\cos\frac{2\pi}{M}(p-p')}}{2}. \tag{8.19}$$

Hereafter, references to a signal point using $\mathbf{\Psi}(\kappa)$ and the pairs of integers $[i,p]$, where i and p are related to κ as defined earlier, are interchangeable. The extensions for $N_t > 2$ are also presented in [108], but only possible for $N_t = 3$ and $N_t = 4$ due to the limitation of orthogonal design.

With orthogonal design, all $d_m(\kappa,\kappa')$, $1 \le m \le N_t$, are equal. Thus the index m can be omitted and the chordal distance between $\mathbf{\Psi}(\kappa)$ and $\mathbf{\Psi}(\kappa')$ in (8.15) is rewritten as:

$$d^c(\kappa,\kappa') = \sqrt{N_t[1-d^2(\kappa,\kappa')]}. \tag{8.20}$$

Using (8.20), the design parameter in (8.14) can be rewritten as follows:

$$\gamma(\mathbf{\Psi},\xi) = \frac{N_t^{N_tN_r}}{KM_c}\sum_{\kappa=1}^{K}\sum_{k=1}^{M_c}[d^c(\kappa,j(\kappa,k))]^{-2N_tN_r}, \tag{8.21}$$

where $d^c(\kappa,j(\kappa,k))$ is the chordal distance between $\mathbf{\Psi}(\kappa)$ and $\mathbf{\Psi}(j(\kappa,k))$ whose labels differ at only position k. On the other hand, the parameter $\varphi(\mathbf{\Psi},\xi)$ in (8.16) can be computed similarly by averaging over $g(\kappa,k)$, which can now be given in a compact form as:

$$g(\kappa,k) = N_t^{N_tN_r}\sum_{\kappa':\mathbf{\Psi}(\kappa')\in I_\kappa^k}[d^c(\kappa,\kappa')]^{-2N_tN_r}. \tag{8.22}$$

Let $d^c_{\min}$ be the minimum chordal distance between any two different signal points in $\mathbf{\Psi}$. Similar to the constellations in the Euclidean space, the mapping rule ξ of a constellation $\mathbf{\Psi}$ is called a Gray mapping if the labels of two signal points at chordal distance $d^c_{\min}$ always differ in only 1 bit.

8.3.2 Gray Mapping

For simplicity, attention is restricted to the case of $N_t = 2$. The results for higher values of N_t are briefly presented in Appendix F.

First, let μ_G denote the conventional Gray mapping rule for M-PSK in the Euclidean signal space. Since the ith signal point in M-PSK is $\exp\left(j\frac{2\pi i}{M}\right)$, the mapping μ_G essentially relates each integer i with a label of $\log_2 M$ bits. Consider a unitary signal point $\mathbf{\Psi}(\kappa) = [i, p]$. From (8.19) and (8.20), it can be seen that there are four and only four nearest neighbors of $[i, p]$ at the distance $d^c_{\min}$ in $\mathbf{\Psi}$, which are $[i, (p \pm 1) \bmod M]$ and $[(i \pm 1) \bmod M, p]$. Thus one has the following theorem.

Theorem 8.1: A Gray mapping of a unitary constellation $\mathbf{\Psi}$ can be constructed by having the first and the last $\log_2 M$ bits of the label of each signal point $[i, p]$ in $\mathbf{\Psi}$ as the conventional Gray labels of the ith and the pth signal points of M-PSK, respectively.

Proof: The proof follows directly from the fact that the label of the $[(i\pm1) \bmod M]$th signal point in M-PSK differs in only 1 bit compared to the label of the ith signal point.

As an example, Gray mapping for the case $N_t = 2$ and $L = 16$ is shown in Table 8.1.

Table 8.1 Gray mapping of an unitary constellation obtained from orthogonal design and 4-PSK.

$\mathbf{\Psi}(1)$	$\mathbf{\Psi}(2)$	$\mathbf{\Psi}(3)$	$\mathbf{\Psi}(4)$	$\mathbf{\Psi}(5)$	$\mathbf{\Psi}(6)$	$\mathbf{\Psi}(7)$	$\mathbf{\Psi}(8)$
0000	0001	0011	0010	0100	0101	0111	0110
$\mathbf{\Psi}(9)$	$\mathbf{\Psi}(10)$	$\mathbf{\Psi}(11)$	$\mathbf{\Psi}(12)$	$\mathbf{\Psi}(13)$	$\mathbf{\Psi}(14)$	$\mathbf{\Psi}(15)$	$\mathbf{\Psi}(16)$
1100	1101	1111	1110	1000	1001	1011	1010

In the following, it is demonstrated that the above Gray mapping is most suitable for systems that do not implement iterative decoding, but it performs very poor for systems with iterative processing. To this end, the parameters $\gamma(\mathbf{\Psi}, \xi)$ and $\varphi(\mathbf{\Psi}, \xi)$ are evaluated for two separate cases of 4-PSK and 8-PSK.

4-PSK

First, consider the parameter $\varphi(\mathbf{\Psi}, \xi)$ in (8.16). For any $\mathbf{\Psi}(\kappa)$, there exists four nearest neighbors at distance $d^c_{\min} = \sqrt{1/2}$. Therefore, each of these four nearest neighbors must be in one of $M_c = 4$ sets $\{I^k_\kappa\}^{M_c}_{k=1}$. One then obtains the following lower bound on $\varphi(\mathbf{\Psi}, \xi)$ for any mapping rule:

$$\varphi(\mathbf{\Psi}, \xi) \geq N_t^{N_t N_r} (\sqrt{1/2})^{-2N_t N_r} = 16^{N_r}. \tag{8.23}$$

Furthermore, for a Gray mapping, it can be easily verified that for each $\mathbf{\Psi}(\kappa)$, $\{I^k_\kappa\}$ contains one and only one nearest signal point at the distance $d^c_{\min}$ to $\mathbf{\Psi}(\kappa)$. It follows that $\varphi(\mathbf{\Psi}, \xi) = 16^{N_r}$ for a Gray mapping, which achieves the above lower bound. Hence, it can be concluded that that the use of a Gray mapping results in the best performance for systems without iterative processing in the sense that it minimizes $\varphi(\mathbf{\Psi}, \xi)$.

To see the effect of iterative processing with a Gray mapping, it is convenient to evaluate the parameter $\gamma(\mathbf{\Psi}, \xi)$ in (8.14). Due to the properties of the nearest neighbors, it can be shown that $\gamma(\mathbf{\Psi}, \xi) = \varphi(\mathbf{\Psi}, \xi) = 16^{N_r}$ for a Gray mapping. This means that with a Gray mapping, even the perfect *a priori* information from the decoder to the demodulator is not useful. Therefore, it is unnecessary to run the iterations when a Gray mapping is employed. In other words, there should be no difference between the error performances with and without iterative processing. This fact is later confirmed by the analysis based on the *extrinsic* information transfer (EXIT) chart. Furthermore, it can be verified that with a Gray mapping, $\gamma(\mathbf{\Psi}, \xi) = 16^{N_r}$ is achieved when all the chordal distances $\{d^c(\kappa, j(\kappa, k))\}$ in (8.21) equal the minimum chordal distance $d^c_{\min}$. This also results in the maximum value of $\gamma(\mathbf{\Psi}, \xi)$ over all mapping rules for the orthogonal design based on 4-PSK. This implies that a Gray mapping provides the worst asymptotic performance for the systems with iterative decoding.

8-PSK

When orthogonal design with 8-PSK is employed, similar observations can be made due to the properties of the labels of the nearest signal points with a Gray mapping. First, since there are four nearest neighbors of any $\mathbf{\Psi}(\kappa)$ at the minimum chordal distance $d^c_{\min} = 0.3827$ and each $\mathbf{\Psi}(\kappa)$ carries M_c bits, one has the following bound for $M_c = 6$ and $N_t = 2$:

$$\varphi(\mathbf{\Psi}, \xi) > \frac{4^{N_r}}{M_c} \left[4(d^c_{\min})^{-2N_r}\right]. \tag{8.24}$$

By comparing the value of $\varphi(\mathbf{\Psi}, \xi)$ with a Gray mapping and the above bound, it is observed that $\varphi(\mathbf{\Psi}, \xi)$ is only slightly greater than the lower bound for various values of N_r. For example, with $N_r = 1$, $\varphi(\mathbf{\Psi}, \xi) = 129.6731$, whereas the lower bound is 124.3182. Therefore, it is expected that a Gray mapping results in almost the best performance for systems without iterative decoding.

To see how good Gray mapping is with iterative processing, the comparison between $\gamma(\mathbf{\Psi}, \xi)$ and $\varphi(\mathbf{\Psi}, \xi)$ is made. It is also observed that the difference is negligible for a given N_r. For example, with $N_r = 1$, one obtains $\gamma(\mathbf{\Psi}, \xi) = 126.1699$ and $\varphi(\mathbf{\Psi}, \xi) = 129.6731$, which indicates that there is only a slight improvement by performing iterative processing. Furthermore, by examining the distance profile of $\mathbf{\Psi}$, it can be seen that for any signal point $\mathbf{\Psi}(\kappa)$, there are only four distinct signal points in $\mathbf{\Psi}$ at the minimum chordal distance $d^c_{\min} = 0.3827$ and two other signal points in $\mathbf{\Psi}$ at the second minimum chordal distance $d^c_{\min,2} = 0.5412$ to $\mathbf{\Psi}(\kappa)$. This fact gives $\gamma(\mathbf{\Psi}, \xi) \leq 139.8823$ for any mapping rule when $N_r = 1$. Clearly, the upper bound on $\gamma(\mathbf{\Psi}, \xi)$ is only slightly bigger than $\gamma(\mathbf{\Psi}, \xi) = 126.1699$ for Gray mapping, indicating that a Gray mapping also performs very poor in iterative systems in terms of the asymptotic error performance.

It should be mentioned that the effects of a Gray mapping on the error performance of noncoherent systems with and without iterative processing as discussed above are very similar to the observations concerning BICM systems over coherent channels in [7].

8.3.3 Mappings for Systems with Iterative Processing

This subsection introduces novel mappings that result in very small values of $\gamma(\mathbf{\Psi}, \xi)$ for systems that implement iterative demodulation and decoding at the receiver. More specifically, optimal mappings for unitary constellations obtained from orthogonal design and 4-PSK are proposed. For orthogonal design based on 8-PSK, a novel mapping with good performance at high SNR is provided. It is conjectured that this novel mapping is the globally optimal mapping. Also for simplicity, attention is paid only to the case $N_t = 2$. Extension to higher values of N_t is briefly discussed in Appendix F.

4-PSK

When $N_t = 2$ and $K = 16$, by studying the distance profile of $\mathbf{\Psi}$, it is not hard to recognize that the structure of $\mathbf{\Psi}$ is very similar to the structure of a 4-dimensional hypercube [25], although the distances considered for the 4-dimensional hypercube in [25] are the Euclidean distances. More specifically, the following facts about chordal distances can be established for any signal point $\mathbf{\Psi}(\kappa) \in \mathbf{\Psi}$.

- *Fact 1*: There is only one signal point whose chordal distance to $\mathbf{\Psi}(\kappa)$ is $d_5^c = 0$, corresponding to the singular value $d_5 = 1$. It is $\mathbf{\Psi}(\kappa)$ itself.

- *Fact 2*: There are four signal points whose chordal distances to $\mathbf{\Psi}(\kappa)$ are equal to $d_4^c = \sqrt{1/2}$, corresponding to the singular value $d_4 = \sqrt{3/4}$. It can be verified that d_4^c is the minimum chordal distance of $\mathbf{\Psi}$.

- *Fact 3*: There are six signal points whose chordal distances to $\mathbf{\Psi}(\kappa)$ are $d_3^c = 1$, corresponding to the singular value $d_3 = \sqrt{1/2}$.

- *Fact 4*: There are four signal points whose chordal distances to $\mathbf{\Psi}(\kappa)$ equal $d_2^c = \sqrt{3/2}$, corresponding to the singular value $d_2 = 1/2$.

- *Fact 5*: There is only one signal point whose chordal distance to $\mathbf{\Psi}(\kappa)$ is $d_1^c = \sqrt{2}$, corresponding to the singular value $d_1 = 0$.

Note that the chordal distances d_h^c, $1 \leq h \leq 5$, are purposely indexed such that $d_5^c < d_4^c < d_3^c < d_2^c < d_1^c$.

Basically, the objective to obtain the mapping with a small value of $\gamma(\mathbf{\Psi}, \xi)$ can be fulfilled by labelling the signal points in $\mathbf{\Psi}$ such that two signal points at a larger chordal distance have a smaller label Hamming distance. Therefore, it follows that d_1^c and d_2^c, the two largest chordal distances of $\mathbf{\Psi}$, are of interest. With $N_t = 2$, rewrite the parameter $\gamma(\mathbf{\Psi}, \xi)$ in (8.21) as follows:

$$\gamma(\mathbf{\Psi}, \xi) = \frac{1}{K} \sum_{\kappa=1}^{K} \gamma_\kappa(\mathbf{\Psi}, \xi), \tag{8.25}$$

where the parameter $\gamma_\kappa(\mathbf{\Psi}, \xi)$ is associated with each signal point $\mathbf{\Psi}(\kappa)$, computed as:

$$\gamma_\kappa(\mathbf{\Psi}, \xi) = \frac{4^{N_r}}{M_c} \sum_{k=1}^{M_c} [d^c(\kappa, j(\kappa, k))]^{-4N_r}. \tag{8.26}$$

With the two largest chordal distances d_1^c and d_2^c, the following lower bound on $\gamma_\kappa(\mathbf{\Psi}, \xi)$ when $M_c = 4$ follows directly from Fact 4 and Fact 5:

$$\begin{aligned} \gamma_\kappa(\mathbf{\Psi}, \xi) &\geq \frac{4^{N_r}}{4} \left[(d_1^c)^{-4N_r} + 3(d_2^c)^{-4N_r} \right] \\ &= \frac{1}{4} \left[1 + 3 \left(\frac{16}{9} \right)^{N_r} \right]. \end{aligned} \tag{8.27}$$

Therefore, one obtains the following lower bound on $\gamma(\mathbf{\Psi}, \xi)$:

$$\begin{aligned} \gamma(\mathbf{\Psi}, \xi) &\geq \frac{1}{K} \sum_{\mathbf{\Psi}(\kappa) \in \mathbf{\Psi}} \left(\frac{1}{4} \left[1 + 3 \left(\frac{16}{9} \right)^{N_r} \right] \right) \\ &= \frac{1}{4} \left[1 + 3 \left(\frac{16}{9} \right)^{N_r} \right]. \end{aligned} \tag{8.28}$$

Clearly, the lower bound in (8.28) is independent of the mapping. If it exists, the mapping that yields this lower bound is the best one in terms of the asymptotic performance. Furthermore, it is easy to verify that this optimal mapping is any mapping that satisfies the following condition.

Condition 8.1: For any signal point $\mathbf{\Psi}(\kappa)$ in $\mathbf{\Psi}$, a signal point with chordal distance d_1^c to $\mathbf{\Psi}(\kappa)$ has label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$. Furthermore, there are

three other signal points with chordal distance d_2^c and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$.

As mentioned before, there is a strong similarity between the distance profiles of the unitary constellation $\mathbf{\Psi}$ and the hypercube constellation considered in [25]. In fact, the mapping that satisfies Condition 8.1 can easily be obtained by modifying the best mapping for the 4-dimensional hypercube constellation in [25]. The procedure to obtain the best mapping is stated as follows.

- *Step 1*: Consider $i = 0$. Create the constellation $\mathbf{\Theta}_1$ with 4 points $\mathbf{\Theta}_1 = \{[0,0],[0,1],[0,3],[0,2]\}$ and the corresponding labels $\xi_1 = \{(0,0),(0,1),(1,0),(1,1)\}$. Observe that this is Gray mapping with $d_{\min}^c = d_4^c$.

- *Step 2*: Construct the constellation $\mathbf{\Theta}_2$ with all signal points $\{[i,p]\}$, $0 \leq i \leq 1$, $0 \leq p \leq 3$. Each signal point $[i,p]$ is then labeled with 3 bits (i,a_1,a_2), where (a_1,a_2) is the label of $[0,p]$ in $\mathbf{\Theta}_1$. This step results in the mapping rule ξ_2 and it is also a Gray mapping with $d_{\min}^c = d_4^c$.

- *Step 3*: Modify the labeling rule ξ_2 to obtain the labeling rule ξ_3 as follows. Each element $(a_1,a_2,a_3) \in \xi_2$ corresponds to element $(b,a_1,a_2,a_3) \in \xi_3$, where $b = 0$ if $(a_1 + a_2 + a_3)$ is even; otherwise, $b = 1$. As the results of labeling rule ξ_3, each signal point in $\mathbf{\Theta}_2$ carries 4 bits.

- *Step 4*: From each element $[i,p] \in \mathbf{\Theta}_2$, construct a signal point $[\hat{i},\hat{p}]$ such that $\hat{i} = (i+2) \bmod 4$ and $\hat{p} = (p+2) \bmod 4$ to obtain the constellation $\hat{\mathbf{\Theta}}_2$. It can be verified that the distance between $[i,p]$ and $[\hat{i},\hat{p}]$ is $d_1^c = \sqrt{2}$. Each label $(b,a_1,a_2,a_3) \in \xi_3$ is then modified to $(\bar{b},a_1,a_2,a_3) \in \hat{\xi}_3$, where $\bar{b} = 1 - b$.

- *Step 5*: The constellation $\mathbf{\Psi}$ and the best mapping ξ are obtained by combining $\mathbf{\Theta}_2$ and $\hat{\mathbf{\Theta}}_2$ and their labels ξ_3 and $\hat{\xi}_3$.

Table 8.2 Chordal distance profile of the proposed optimal mapping for the unitary constellation constructed from the orthogonal design and 4-PSK.

	$\boldsymbol{\Psi}(1)$ 0000	$\boldsymbol{\Psi}(2)$ 1001	$\boldsymbol{\Psi}(3)$ 0011	$\boldsymbol{\Psi}(4)$ 1010	$\boldsymbol{\Psi}(5)$ 1100	$\boldsymbol{\Psi}(6)$ 0101	$\boldsymbol{\Psi}(7)$ 1111	$\boldsymbol{\Psi}(8)$ 0110	$\boldsymbol{\Psi}(9)$ 1011	$\boldsymbol{\Psi}(10)$ 0010	$\boldsymbol{\Psi}(11)$ 1000	$\boldsymbol{\Psi}(12)$ 0001	$\boldsymbol{\Psi}(13)$ 0111	$\boldsymbol{\Psi}(14)$ 1110	$\boldsymbol{\Psi}(15)$ 0100	$\boldsymbol{\Psi}(16)$ 1101
$\boldsymbol{\Psi}(1)=[0,0]$	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	1
$\boldsymbol{\Psi}(2)=[0,1]$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$
$\boldsymbol{\Psi}(3)=[0,2]$	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	1
$\boldsymbol{\Psi}(4)=[0,3]$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	1	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$
$\boldsymbol{\Psi}(5)=[1,0]$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	1	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$
$\boldsymbol{\Psi}(6)=[1,1]$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$
$\boldsymbol{\Psi}(7)=[1,2]$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$
$\boldsymbol{\Psi}(8)=[1,3]$	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1
$\boldsymbol{\Psi}(9)=[2,0]$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1
$\boldsymbol{\Psi}(10)=[2,1]$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$
$\boldsymbol{\Psi}(11)=[2,2]$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1
$\boldsymbol{\Psi}(12)=[2,3]$	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	1	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$
$\boldsymbol{\Psi}(13)=[3,0]$	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	1	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$
$\boldsymbol{\Psi}(14)=[3,1]$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1
$\boldsymbol{\Psi}(15)=[3,2]$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	1	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{\mathbf{3/2}}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$
$\boldsymbol{\Psi}(16)=[3,3]$	1	$\sqrt{\mathbf{3/2}}$	1	$\sqrt{1/2}$	$\sqrt{\mathbf{3/2}}$	$\sqrt{\mathbf{2}}$	$\sqrt{\mathbf{3/2}}$	1	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0

Note: The binary number under each signal (in the first row) denotes the proposed mapping. As an example, $\boldsymbol{\Psi}(4)=[0,3]$ and $\boldsymbol{\Psi}(10)=[2,1]$ are labeled with 1010, and 0010, respectively. These two labels differ in only the first bit and the chordal distance between these signals is $\sqrt{2}$.

The final results are shown in Table 8.2. It can be seen that for any signal point $\mathbf{\Psi}(\kappa)$, the signal point whose label differs in only 1 bit at the first position is at the distance $d_1^c = \sqrt{2}$ to $\mathbf{\Psi}(\kappa)$. The other three signal points whose labels differ in only 1 bit at the second, the third and the forth positions are at the distance $d_2^c = \sqrt{3/2}$ to $\mathbf{\Psi}(\kappa)$.

Observe that the value of $\gamma(\mathbf{\Psi}, \xi)$ in (8.28) for the optimal mapping is much smaller than that of a Gray mapping for a given N_r. For example, with $N_r = 1$, $\gamma(\mathbf{\Psi}, \xi) = 1.5833$ for the optimal mapping, which is very close to the minimum value of 1 and significantly smaller than the value of $\gamma(\mathbf{\Psi}, \xi) = 16$ for a Gray mapping. This suggests that a significant coding gain can be achieved by using the optimal mapping over a Gray mapping in terms of the asymptotic performance. Regarding the error performance after the first iteration, although being optimal in terms of the asymptotic performance with iterative processing, the proposed mapping performs poorer than Gray mapping after the first iteration. This is due to the large value of $\kappa(\mathbf{\Psi}, \xi)$ (e.g., $\varphi(\mathbf{\Psi}, \xi) = 36$ for $N_r = 1$) produced by the proposed mapping.

The optimal mapping for $N_t = 4$ transmit antennas is presented in Appendix F.

8-PSK

For the case of the orthogonal design with 8-PSK constellation, the globally optimal mapping in terms of minimizing the asymptotic performance is unknown. Instead, proposed in the following is a mapping that can approach a lower bound on the asymptotic performance very close. The construction of this mapping is also strongly related to the construction of a novel mapping of the multi-dimensional 8-PSK constellation proposed in Chapter 3 for coherent BICM-ID systems. Similar to Chapter 3, the basic idea behind the construction of this mapping comes from the simple fact that an 8-PSK can be decomposed into two 4-PSK signal sets. Thus, there is also a strong relationship between the orthogonal design based on 8-PSK and the one based on 4-PSK.

Upon examining the distance profile of the constellation, it can be observed that the optimal mapping, if it exists, is any mapping that satisfies the following condition.

Condition 8.2: For any signal point $\mathbf{\Psi}(\kappa)$ in $\mathbf{\Psi}$, there is one signal point with chordal distance $d_1^c = \sqrt{2}$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$. Furthermore, there are four other signal points with chordal distance $d_2^c = \sqrt{6+\sqrt{2}}/2$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$ and one signal point with chordal distance $d_3^c = \sqrt{1+1/\sqrt{2}}$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$.

Here, the chordal distances d_1^c, d_2^c and d_3^c are the three largest distances of $\mathbf{\Psi}$. Unfortunately, it is not difficult to show that there does not exist mapping that satisfies Condition 8.2. Nevertheless, a novel mapping that satisfies a weaker condition, stated as Condition 8.3 below, shall be proposed. Also, for convenience in evaluating the proposed mapping, the phrase "hypothetical mapping" is used to refer to a mapping that satisfies Condition 8.2.

Condition 8.3: For any signal point $\mathbf{\Psi}(\kappa)$ in $\mathbf{\Psi}$, there is one signal point with chordal distance $d_1^c = \sqrt{2}$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$. Furthermore, there are two other signal points with chordal distance $d_2^c = (\sqrt{6+\sqrt{2}})/2$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$ and three signal points with chordal distance $d_4^c = \sqrt{3/2}$ and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$.

The construction of the mapping that satisfies Condition 8.3 is inherited from the one in Chapter 3 and based on the optimal mapping for the orthogonal design with 4-PSK presented earlier. Specifically, for any signal point $[i, p]$, associate it with 2 bits (x_1, x_2), where $x_1 = i \bmod 2$ and $x_2 = p \bmod 2$. The corresponding decimal number of (x_1, x_2) is denoted as X, where $0 \leq X \leq 3$. Clearly, for each (x_1, x_2), one finds 16 signal points in the sub-constellation $\mathbf{\Theta}_X$, where the chordal distance profile of each sub-constellation is exactly the same with that of the unitary constellation obtained from the orthogonal design and 4-PSK. With this observation, the detailed steps for the construction of the proposed mapping are given below.

- *Step 1*: For any signal point $[i, p]$, label the first 2 bits by (x_1, x_2), where $x_1 =$

$i \bmod 2$ and $x_2 = p \bmod 2$.

- *Step 2*: Consider $\mathbf{\Theta}_0$. Since the chordal distance profile of $\mathbf{\Theta}_0$ is the same with the orthogonal design and 4-PSK, label the last 4 bits of any signal point in $\mathbf{\Theta}_0$ by the optimal mapping of the orthogonal design and 4-PSK.

  As a result, all signal points in $\mathbf{\Theta}_0$ are labeled and the properties for the last 4 bits are related to two distances $d_1^c = \sqrt{2}$ and $d_4^c = \sqrt{3/2}$.

- *Step 3*: For each signal point $[i, p] \in \mathbf{\Theta}_0$ whose label is $(0, 0, a_1, a_2, a_3, a_4)$, associate it with the signal point $[i', p'] \in \mathbf{\Theta}_X$, where $i' = (i + 3x_1 + 4x_2)$ and $p' = (p + 4x_1 + 3x_2)$. Moreover, label the signal $[i', p']$ with $(x_1, x_2, a_1, a_2, a_3, a_4)$.

  This step ensures that, as far as the last 4 bits in the labels of all the signal points in $\mathbf{\Theta}_X$ are concerned, the properties of the optimal mapping obtained for the orthogonal design and 4-PSK also hold here. Furthermore, any two signal points in two different sub-constellations whose labels differ in only 1 bit at the first or the second positions are always at the chordal distance $d_2^c = (\sqrt{6 + \sqrt{2}})/2$.

The final mapping is shown in Table 8.3, which can be verified to satisfy Condition 8.3. By the symmetry properties of this proposed mapping, it is also easy to verify that this mapping is immune to the binary-switching algorithm (BSA).[2] This means that if the BSA starts with the proposed mapping, it can never find a better one. Furthermore, simple computation shows that, when $N_r = 1$, the values of $\gamma(\mathbf{\Psi}, \xi)$ for the proposed and hypothetical mappings are 1.4436 and 1.1716, respectively, which are quite close. Compared to $\gamma(\mathbf{\Psi}, \xi) = 126.1699$ of the Gray mapping, the parameter $\gamma(\mathbf{\Psi}, \xi)$ of the proposed mapping is obviously much smaller, which suggests that a significant coding gain in terms of the asymptotic performance can be achieved with the proposed mapping. Analytical and simulation results in the next section also show that the performance the proposed mapping can closely approach a lower bound on the asymptotic performance of any constellation and mapping that corresponds to $\gamma(\mathbf{\Psi}, \xi) = 1$ and greatly outperforms the one with a Gray mapping. However, with

[2]The proof is omitted here for the brevity of presentation.

Table 8.3 The proposed mapping of the unitary constellation obtained from the orthogonal design and 8-PSK, with four sub-constellations $\mathbf{\Theta}_0$, $\mathbf{\Theta}_1$, $\mathbf{\Theta}_2$, and $\mathbf{\Theta}_3$.

$\mathbf{\Theta}_0$		$\mathbf{\Theta}_1$		$\mathbf{\Theta}_2$		$\mathbf{\Theta}_3$	
$[k,p]$	ξ	$[k,p]$	ξ	$[k,p]$	ξ	$[k,p]$	ξ
$[0,0]$	(000000)	$[0,1]$	(010010)	$[1,0]$	(101111)	$[1,1]$	(111101)
$[0,2]$	(001010)	$[0,3]$	(011011)	$[1,2]$	(100101)	$[1,3]$	(110100)
$[0,4]$	(000011)	$[0,5]$	(010001)	$[1,4]$	(101100)	$[1,5]$	(111110)
$[0,6]$	(001001)	$[0,7]$	(011000)	$[1,6]$	(100110)	$[1,7]$	(110111)
$[2,0]$	(000111)	$[2,1]$	(010101)	$[3,0]$	(100011)	$[3,1]$	(110001)
$[2,2]$	(001101)	$[2,3]$	(011100)	$[3,2]$	(101001)	$[3,3]$	(111000)
$[2,4]$	(000100)	$[2,5]$	(010110)	$[3,4]$	(100000)	$[3,5]$	(110010)
$[2,6]$	(001110)	$[2,7]$	(011111)	$[3,6]$	(101010)	$[3,7]$	(111011)
$[4,0]$	(001011)	$[4,1]$	(011001)	$[5,0]$	(100100)	$[5,1]$	(110110)
$[4,2]$	(000001)	$[4,3]$	(010000)	$[5,2]$	(101110)	$[5,3]$	(111111)
$[4,4]$	(001000)	$[4,5]$	(011010)	$[5,4]$	(100111)	$[5,5]$	(110101)
$[4,6]$	(000010)	$[4,7]$	(010011)	$[5,6]$	(101101)	$[5,7]$	(111100)
$[6,0]$	(001100)	$[6,1]$	(011110)	$[7,0]$	(101000)	$[7,1]$	(111010)
$[6,2]$	(000110)	$[6,3]$	(010111)	$[7,2]$	(100010)	$[7,3]$	(110011)
$[6,4]$	(001111)	$[6,5]$	(011101)	$[7,4]$	(101011)	$[7,5]$	(111001)
$[6,6]$	(000101)	$[6,7]$	(010100)	$[7,6]$	(100001)	$[7,7]$	(110000)

the large value of $\varphi(\mathbf{\Psi},\xi)$ (e.g., $\varphi(\mathbf{\Psi},\xi) = 297.9736$ when $N_r = 1$), this mapping provides a very poor performance after the first iteration.

8.3.4 Discussion On the Convergence Behavior of the Proposed Mappings

Up to now, it has been demonstrated analytically that the use of the proposed mappings for iterative systems will result in a significant coding gain compared to a Gray mapping, as far as the asymptotic performance is concerned. To achieve

a perfect convergence (i.e., the BER performances after a number of iterations can converge to the asymptotic one), it is pointed out in [23] that the BER measured at the input of the SISO decoder during the first iteration must be under a given threshold. This threshold varies for different error-correcting codes. In general, the powerful codes such as turbo-like codes require lower thresholds than the standard convolutional codes [23]. Due to their poor performances after the first iteration, the proposed mappings for the iterative systems need a higher SNR to achieve the lower BER thresholds of the more powerful codes to make iteration convergence happen. At a very high SNR, the asymptotic coding gains of the proposed mappings over a Gray mapping can still be observed. However, this happens at very low BER levels, which is not of practical interest. On the other hand, using a Gray mapping can achieve the practical BER level of interest (say 10^{-5} to 10^{-6}) at a lower SNR because of its superior performance after the first iteration. The above discussion suggests that a Gray mapping is the most suitable mapping when powerful codes (e.g., Turbo-like codes) are employed. This fact has been observed for systems over coherent channels using turbo codes in [23] and low-density parity-check (LDPC) codes in [115].

8.4 Analytical and Simulation Results

This section provides analytical and simulation results to confirm the analysis carried out in the previous sections. For given constellation $\mathbf{\Psi}$ and mapping rule ξ, the error bound computed from (8.3) and (8.8) with the first 20 Hamming distances of convolutional codes is provided to show the tightness of the bound. The lower bound applicable for any unitary constellation and mapping is obtained by simply setting $\gamma(\mathbf{\Psi}, \xi) = 1$ and all the singular values in (8.14) to zero. A brief comparison between systems employing the orthogonal designs and the proposed mappings and systems using the systematic design is also made. Finally, convergence analysis of the proposed mappings is provided by means of the *extrinsic* information transfer (EXIT) charts.

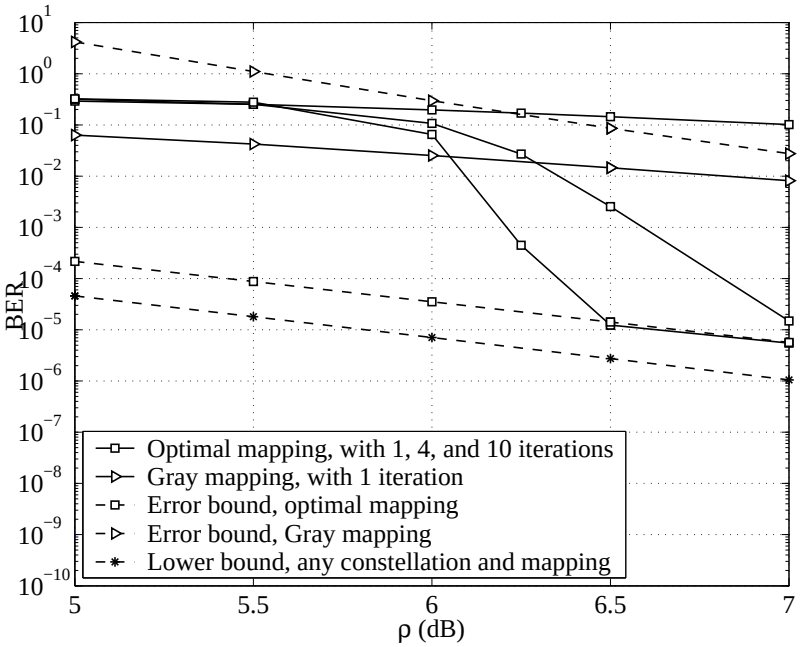


Figure 8.2 Performance of the unitary coded systems employing the orthogonal design based on 4-PSK with different mappings, $N_r = 1$.

8.4.1 Orthogonal Design with Proposed Mappings

First, start with the orthogonal design based on 4-PSK and $N_t = 2$ transmit antennas. A rate-1/2, 4-state convolutional code with generator matrix $\mathbf{g} = (5, 7)$ is used, which results in the spectral efficiency of 0.5 bit per channel use. A random interleaver of length 12,000 coded bits is implemented. Fig. 8.2 presents the BER performances of unitary space-time coded systems with the two proposed mappings. There is only one receive antenna, i.e., $N_r = 1$. For convenience, the SNR is computed as ρ (dB). More specifically, shown in Fig. 8.2 are the BER of the iterative system with the proposed optimal mapping after 1, 4 and 10 iterations. In the case of a Gray mapping, the BER is only provided with 1 iteration since it was observed that there is no improvement from the second iteration. This is expected from the analysis made in the previous section and it confirms that iterative processing is useless for a Gray mapping. The error bounds are also plotted in Fig. 8.2 to illustrate the tightness of the bounds. It can be seen that for the proposed optimal mapping, the BER converges to the error bound at a practical level of 10^{-5}. Thus, the bound is

useful to predict the error performance at high SNR. In the case of a Gray mapping, the bound is seen to underestimate the simulation result. This is only due to the poor performance of the Gray mapping over the range of SNR shown in Fig. 8.2. Though not shown here, it was observed that the bound is tight at higher SNR where the BER level of around 10^{-4} is reached.

It can also be seen that with the use of a very simple convolutional code, the system employing the proposed optimal mapping and iterative processing achieves a significant coding gain compared to that using Gray mapping, even though it offers a very poor performance after the first iteration. To gauge the performance improvement by the proposed optimal mapping, the lower bound on the asymptotic performance of any unitary constellation and mapping is also provided in Fig. 8.2. Observe that there is only a slight difference between the two systems. In particular, the performance gap is only about 0.8dB over the BER range of 10^{-5} to 10^{-6}. Though not explicitly shown here, examining the error bounds plotted over a wider range of SNR reveals that compared to the Gray mapping, the use of the proposed optimal mapping results in the coding gain of 4.8dB at the BER level of 10^{-6}. However, it should be pointed out that when a more powerful code (e.g., a Turbo-like code) is selected, Gray mapping is expected to be the most suitable mapping as discussed earlier.

Similar results can be obtained for the case of $N_r = 2$ receive antennas, which are presented in Fig. 8.3. Observe that when $N_r = 2$, the BER also converges to the bound but at a lower level of 10^{-6} as compared to the case of $N_r = 1$. Clearly, the bound is still very useful to predict the BER for the systems. The advantage of using the proposed optimal mapping is also obvious, where there is only about 0.8dB gap between the performance of the proposed optimal mapping and the general lower bound on the asymptotic performance at the BER level of 10^{-6}. Based on the error bounds plotted over a wider range of SNR, it is also observed that a coding gain of 4.5dB at BER level of 10^{-6} can be obtained with the optimal mapping compared to a Gray mapping.

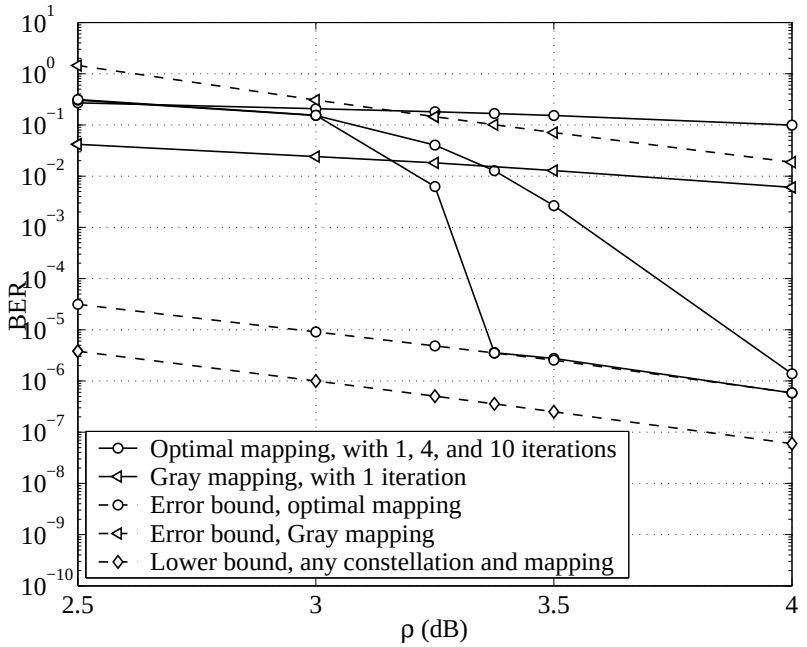


Figure 8.3 Performance of the unitary coded systems employing the orthogonal design based on 4-PSK with different mappings, $N_r = 2$.

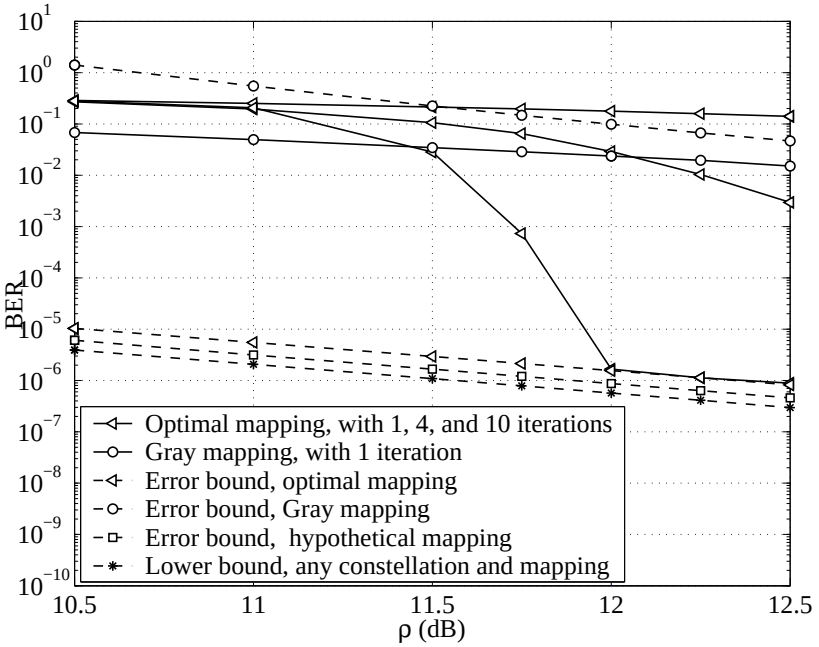


Figure 8.4 Performance of the unitary coded systems employing the orthogonal design based on 8-PSK with different mappings, $N_r = 1$.

Next, consider the case of the orthogonal design based on 8-PSK and $N_t = 2$. Simulations were carried out for systems with a rate-2/3, 4-state convolutional code

whose generator matrices are $\mathbf{g}_1 = (6, 2, 6)$ and $\mathbf{g}_2 = (2, 4, 4)$. This gives a spectral efficiency of 1 bit per channel use. A random interleaver of length 18,000 coded bits was selected. Fig. 8.4 presents the BER performance of the systems employing the Gray mapping and the mapping proposed in Table 8.3. The number of receive antennas is $N_r = 1$. The error bounds are also plotted for comparison. Similar to the case of 4-PSK, only the error performance with 1 iteration is provided for the Gray mapping, since there are only very slight improvements from the next iterations. This again confirms that iterative demodulation and decoding is useless for systems with a Gray mapping. Also in this figure, the performance of the proposed mapping is compared against the asymptotic performances of the hypothetical mapping (which satisfies Condition 8.2) and the general lower bound on the asymptotic performance of any constellation and mapping. It can be seen that the performance gaps at the BER level of 10^{-6} are only 0.4 and 0.8 dB, respectively. It can also be seen that the BER performances of the proposed mapping greatly outperform that of the Gray mapping. Based on the error bounds, it is calculated that the coding gain is about 8.5dB at the BER level of 10^{-6}.

8.4.2 Comparison of Orthogonal and Systematic Designs

As mentioned before, the mapping problem of the unitary constellation obtained from the systematic design has been studied in [110] using the BSA. More specifically, the study in [110] only concentrates on the specific unitary constellation with $N_t = 2$ and $T = 8$ introduced in [105]. It is observed in [110] that performance difference due to the use of different mapping rules is not much.

The basic idea of the systematic design is to be able to construct the whole unitary constellation $\mathbf{\Psi}$ by starting from the first signal in the constellation, namely $\mathbf{\Psi}(1)$. The other signals are then obtained systematically by successively rotating $\mathbf{\Psi}(1)$ in a high-dimensional complex space. This construction can be viewed algebraically, in which the constellation is created by mapping codewords of a linear block code into complex signal matrices [105]. The construction procedure can be summarized in the

following.

Let R_n be the ring of integers modulo-n. Let $\mathcal{C}$ be a linear block code over R_n whose dimension is G, which has a total of $K = n^G$ codewords. The code $\mathcal{C}$ has a $G \times T$ generator matrix $\boldsymbol{U}$ restricted in a systematic form $\boldsymbol{U} = [\boldsymbol{I}_G \ \boldsymbol{U}']$, where the elements of the $G \times (T - G)$ matrix $\boldsymbol{U}'$ are also in R_n. Using the mapping $\phi(i) = 1/\sqrt{T}\exp(j2\pi i/n)$ with $i \in R_n$, the set of G $T \times T$ diagonal matrices $\{\boldsymbol{\Phi}_g\}_{g=1}^{G}$ with entries $\boldsymbol{\Phi}_g(t,t) = \phi(\boldsymbol{U}(g,t))$, $1 \leq t \leq T$, is then created. The matrices $\{\boldsymbol{\Phi}_g\}_{g=1}^{G}$ and the first signal point $\boldsymbol{\Psi}(1)$ are then combined to form the remaining signal points in $\boldsymbol{\Psi}$. For example, with $G = 1$ the κth signal point in $\boldsymbol{\Psi}$ is determined as $\boldsymbol{\Psi}(\kappa) = (\boldsymbol{\Phi}_1)^{\kappa-1}\boldsymbol{\Psi}(1)$, where the starting signal $\boldsymbol{\Psi}(1)$ can be chosen as the columns of a DFT matrix whose all elements have the same magnitude of $1/\sqrt{T}$ [105].

Clearly, the structure of $\boldsymbol{\Psi}$ depends largely on the generator matrix $\boldsymbol{U}$. The sensible design criterion is to maximize the minimum chordal distance $d^c_{\min}$ of $\boldsymbol{\Psi}$ [105]. As an illustrative example, consider the case $N_t = 2$ and $T = 4$. Using computer search, the unitary constellation with $K = 16$, $G = 1$ and $\boldsymbol{U} = [1, 2, 5, 12]$ is found. Then, by applying the BSA with 10^5 different trials, the following mapping is obtained: $\{\boldsymbol{\Psi}(1),\boldsymbol{\Psi}(2),\boldsymbol{\Psi}(3),\boldsymbol{\Psi}(4),\boldsymbol{\Psi}(5),\boldsymbol{\Psi}(6),\boldsymbol{\Psi}(7),\boldsymbol{\Psi}(8),\boldsymbol{\Psi}(9),\boldsymbol{\Psi}(10),\boldsymbol{\Psi}(11),\boldsymbol{\Psi}(12),\boldsymbol{\Psi}(13),\boldsymbol{\Psi}(14),$ $\boldsymbol{\Psi}(15),\boldsymbol{\Psi}(16)\}\leftrightarrow\{$0111,0000,1110,1000,0101,0010,1100,1010,0110,0001,1111,1001,0100, 0011,1101,1011$\}$. This BSA mapping yields $\gamma(\Psi, \xi) = 4.5931$. Note that the proposed optimal mapping for the orthogonal design based on 4-PSK yields $\gamma(\Psi, \xi) = 1.5833$. Although BSA generally gives only a locally optimal mapping, it is expected that the mapping obtained with 10^5 trials of BSA is very close to the globally optimal mapping.

Fig. 8.5 compares the error performance of the two systems that employ the orthogonal design/optimal mapping and systematic design/BSA mapping when $N_r = 1$. The channel code is the same rate-1/2 convolutional code mentioned earlier. More specifically, plotted in Fig. 8.5 are the BER performances with 1, 4 and 10 iterations for each system. The error bounds are also provided to show the tightness of the bounds. It can be clearly seen that the BER converges to the error bounds for

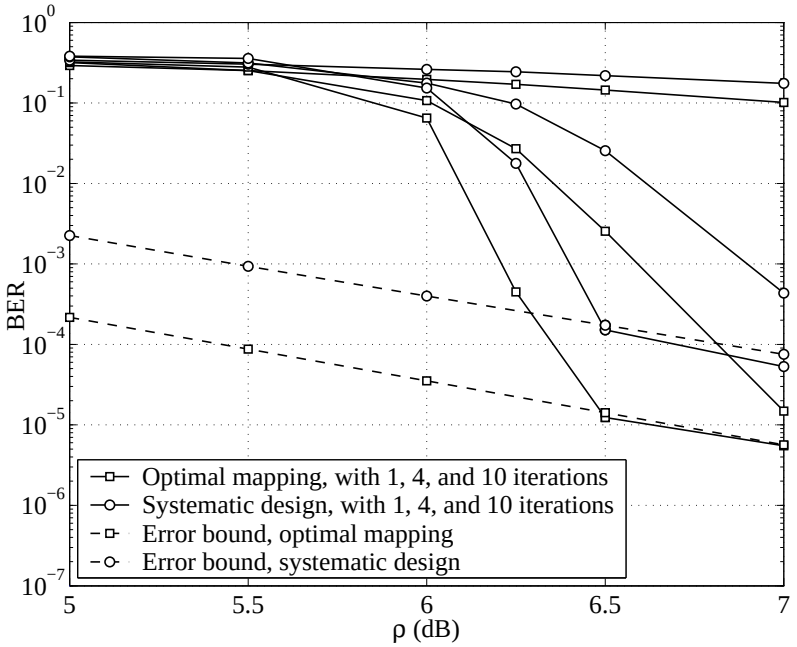


Figure 8.5 Performance comparison between the orthogonal desing/optimal mapping and the systematic design/BSA mapping, $N_r = 1$.

both systems. As expected, the use of the orthogonal design/optimal mapping offers significant coding gains over the systematic design/BSA mapping. Furthermore, it is observed that the system using the orthogonal design/optimal mapping is also preferred in terms of the convergence behavior. This observation is further confirmed by the EXIT chart analysis in the next subsection.

8.4.3 Convergence Analysis of the Proposed Mappings with EXIT Charts

The different mappings proposed for coded unitary space-time systems can also be analyzed by the *extrinsic* information transfer (EXIT) chart as introduced in Chapter 2.

Fig. 8.6 plots the EXIT charts for the system employing the orthogonal design/optimal mappings with 4-PSK, $N_r = 1$, and at $\rho = 6.5$ dB. The EXIT chart for the systematic design/BSA mapping is also provided for comparison. As can be seen, the iterative demodulation and decoding is useless for the Gray mapping as the

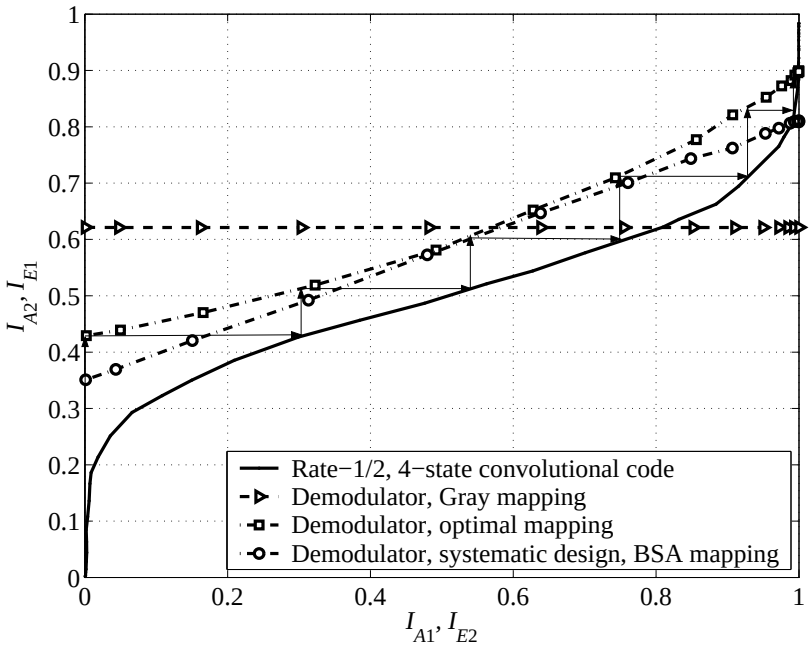


Figure 8.6 EXIT charts of the systems employing different mappings at $\rho = 6.5$ dB, $N_r = 1$ and a rate-1/2, 4-state convolutional code.

bit-wise mutual information for the Gray mapping remains almost constant for the whole range of the *a priori* information I_{A1}. This again confirms the analytical and simulation results shown in the previous sections. On the other hand, there is a significant difference in terms of the bit-wise mutual information for the optimal mapping with no *a priori* information and with the perfect *a priori* information. Based on the decoding trajectory, it takes about 6 iterations for the iterative decoding to converge, which matches well with the BER performance presented earlier. Furthermore, it is interesting to observe that the orthogonal design/optimal mapping outperforms the systematic design/BSA mapping regardless of the values of the *a priori* information. Thus, with respect to both the asymptotic performance and convergence behavior, the orthogonal design/optimal mapping is clearly preferred over the systematic design/BSA mapping.

Similar EXIT charts are shown in Fig. 8.7 (at $\rho = 12$ dB) for both the Gray and proposed mappings of the orthogonal design based on 8-PSK. Again, observe that the bit-wise mutual information for the Gray mapping is almost constant. For the

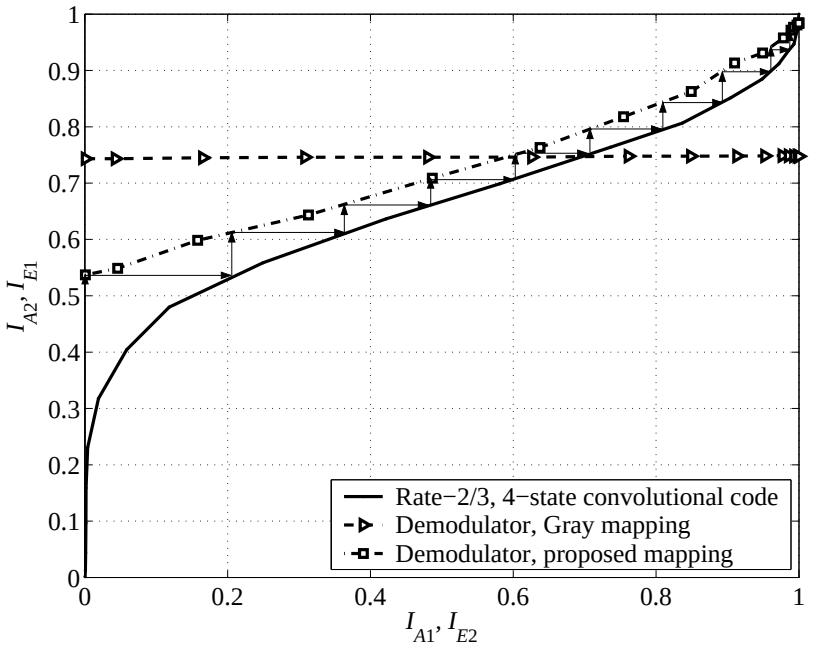


Figure 8.7 EXIT charts of the systems employing the orthogonal design based on 8-PSK with different mappings at $\rho = 12$ dB, $N_r = 1$, and a rate-2/3, 4-state convolutional code.

proposed mapping, the bit-wise mutual information approaches 1 very closely when there is a perfect feedback from the decoder. It needs about 10 iterations to achieve to the asymptotic performance, which also closely matches with the simulation results presented before.

9. Conclusion and Suggestions for Further Studies

9.1 Conclusion

This dissertation was devoted to the design and analysis of a BICM-ID system in a multi-dimensional framework combining and incorporating different diversity techniques, in order to combat fading effectively in wireless communications. The main contributions are summarized as follows:

- For BICM-ID over a single-antenna *frequency non-selective* fading channel, an algorithm to construct a good multi-D mapping of 8-PSK that significantly improves the asymptotic performance of BICM-ID systems was introduced.

- The use of signal space diversity (SSD) technique in BICM-ID over *frequency non-selective* fading that makes BICM-ID systems insensitive to the adverse effects of fading over fully interleaved *frequency non-selective* fading channels was considered. By deriving a tight error bound on the bit error probability (BEP), a design criterion that allows one to understand the individual contributions of the signal constellation, signal mapping and the role of SSD to the overall performance of the system was obtained. A simple sub-optimal but yet effective iterative receiver structure suitable for high-dimensional signal space was also developed. In essence, this research work has partially addressed a current and important research problem in wireless communications: Practically eliminate the adverse effects of fading with a low complexity receiver structure.

- The technique of SSD in BICM-ID was extended to the case of *frequency se-*

lective fading in a bandwidth-efficient bit-interleaved coded modulation with OFDM and iterative decoding (BI-COFDM-ID) system. It was shown that the use of signal space diversity within a specifically selected group of subcarriers can achieve the same maximum multipath diversity and coding gains as with all the available subcarriers. The significant contribution of the study in this work is that the problem of subcarrier grouping and rotation matrix design in SSD is jointly addressed, and as a consequence, one can easily optimize the error performance while keeping the complexity at minimum.

- By combining the ideas of multi-D mapping and subcarrier grouping, a novel power and bandwidth-efficient bit-interleaved coded modulation with OFDM and iterative decoding (BI-COFDM-ID) in which multi-D mapping is performed over a group of subcarriers for broadband transmission in a *frequency selective* fading environment was proposed. Analytical and simulation results showed that the proposed system with the combined optimal subcarrier mapping and grouping can achieve the full channel diversity without using SSD and provide significant coding gains as compared to the previously studied BI-COFDM-ID with the same power, bandwidth and receiver complexity.

- Finally, the application of BICM-ID over a multiple-input multiple-output (MIMO) system equipped with multiple antennas at both the transmitter and the receiver to exploit both time and spatial diversities, where neither the transmitter nor the receiver knows the channel fading coefficients, was studied. In particular, the problem of mapping from binary bits to signal points in a compact Grassmannian manifold used in the signal transmission over non-coherent MIMO channels was concentrated. By using the unitary constellation obtained from orthogonal design, two different mapping rules are proposed. The first mapping rule gives the most suitable mapping for systems that do not implement iterative processing, which is similar to a Gray mapping in coherent channels. The second mapping rule yields the best mapping for systems with iterative decoding. The results obtained in this work are important for both commu-

nications theories and practical applications of non-coherent MIMO systems. This is because the labelling process in a Grassmannian constellation in general is crucial but difficult, due to the absence of structure in the constellation.

9.2 Suggestions for Further Research

Although making some important contributions, the research work in this dissertation has only scratched a small tip of an iceberg and many open problems remain to be answered in order to address many existing challenges in the design of wireless communications systems. The following describes some research problems that are worthwhile to investigate

9.2.1 Extension Problems

Bit-Interleaved Coded Systems with Single-Carrier and Signal Space Diversity

As a logical but important extension of the prior work, the application of signal space diversity in bit-interleaved coded system over *frequency selective* fading channels with single carrier to achieve time, signal space, and frequency diversity, as similar to the case of multi-carrier systems can be considered. However, in contrast to the multi-carrier systems, the receiver design needs to equalize the effect of inter-symbol interference (ISI). The analysis for such a system can benefit from the results in [116] where signal space diversity for ISI channels has been investigated for uncoded systems.

Signal Space Diversity with Imperfect Channel State Information (CSI) and no CSI

In Chapters 4 and 5, it was demonstrated that by using SSD in a high-dimensional signal space, the adverse effects of fading can be practically eliminated. The model considered in Chapters 4 and 5 was limited in the sense that the channel state information (CSI) is assumed to be perfectly known at the receiver. For mobile com-

munication systems, due to the rapidly changing conditions in fading environments, perfect knowledge of the fading coefficients might not be available. As a result, one needs to consider a more realistic model, where only imperfect CSI or no ISI is known at the receiver. Therefore, it would be very interesting to investigate a novel design using SSD that also combats fading effectively for the cases of imperfect CSI and no ISI, as similar to the case of perfect CSI that was studied before.

9.2.2 More Advanced Problems

As aforementioned, diversity is an effective technique to overcome the adverse effects of fading. Among various diversity techniques, spatial diversity is of particular interest and has received a lot of attention in the literature, since it provides diversity without additional cost of increased bandwidth nor transmission time. However, due to practical limitation in putting multiple antennas at the transceivers, efforts have been devoted to a new way of realizing spatial diversity, called cooperative diversity or user cooperation diversity, in a distributed fashion. The strategy is to create a virtual antenna array by deploying multiple terminals (relay terminals) equipped with a single-antenna to assist the source terminal. As an example, by appropriately cooperating the relay and source terminals based on the idea of distributed space-time codes, it was shown in [117] that the full spatial diversity can be achieved. Furthermore, another interesting point in cooperative diversity is that many cooperative diversity schemes can be inherited from the framework of network coding [118–120]. Undoubtedly, cooperative diversity is currently one of the most important topics in communication and information research community. Therefore, it would also be very interesting to investigate a novel BICM-based structure that fully exploits diversity and achieves near-capacity in cooperative-based systems.

A. Non-Existence of The Mapping to satisfy Property 3.1

Consider any (twisted) hypercube H_X in $\mathbf{\Psi}$. Clearly, the minimum Euclidean distance between any two signal points in H_X is d_3. Hereafter, the nearest neighbor of any signal point refers to the signal point lying in the same hypercube and at the distance d_3 to the reference signal point.

Assume there is a mapping that satisfies *Property 3.1*. Consider the signal point $\boldsymbol{s} = [s_1, \ldots, s_N]$ in H_X. For each s_n, obtain the value z_n such that $s_n = (z_n+4) \bmod 8$. It then follows that the squared Euclidean distance between $\boldsymbol{z} = [z_1, \ldots, z_N]$ and $\boldsymbol{s} = [s_1, \ldots, s_N]$ is Nd_1^2. Furthermore, it can be observed that these two signal points are in the same twisted hypercube H_X. From *Property 3.1*, the labels of these two signal points must be different in only 1 bit.

In what follows, it is shown that for any two signal points in H_X, their labels differ in an even number of bits and hence, contradicting with the above observation. For a given n, let $\boldsymbol{y}$ and $\boldsymbol{p}$ denote the two signal points whose components are the same with that of $\boldsymbol{s}$, except the nth components. Instead, these nth components of $\boldsymbol{y}$ and $\boldsymbol{p}$ are given as $y_n = (z_n + 2) \bmod 8$ and $p_n = (z_n + 6) \bmod 8$, respectively. Clearly, for a given n, the two signal points $\boldsymbol{y}$ and $\boldsymbol{p}$ are in H_X and they are the only two nearest neighbors of $\boldsymbol{s}$. On the other hand, denote $\boldsymbol{t}$ and $\boldsymbol{w}$ the two signal points whose components are the same with $\boldsymbol{z}$ except the nth components. These nth components are given as $t_n = (z_n + 1) \bmod 8$ and $w_n = (z_n + 7) \bmod 8$. It then follows that:

$$||\boldsymbol{s} - \boldsymbol{t}||^2 = ||\boldsymbol{s} - \boldsymbol{w}||^2 = (N-1)d_1^2 + d_2^2 \tag{A.1}$$

Hence, the labels of $\boldsymbol{t}$ and $\boldsymbol{w}$ differ in only 1 bit compared to that of $\boldsymbol{s}$. Moreover, one has:

$$||\boldsymbol{y} - \boldsymbol{w}||^2 = ||\boldsymbol{p} - \boldsymbol{t}||^2 = (N-1)d_1^2 + d_2^2 \tag{A.2}$$

This means that for each pair of signal points $(\boldsymbol{y}, \boldsymbol{w})$ or $(\boldsymbol{p}, \boldsymbol{t})$, their labels differ in only 1 bit. It then can be concluded that the labels of $\boldsymbol{y}$ and $\boldsymbol{p}$ differ in 2 bits compared to $\boldsymbol{s}$ for a given n. This property can be observed for any signal point $\boldsymbol{s}$ and its nearest neighbors. By the symmetry of a hypercube, it can be seen that any two signal points must differ in an even number of bits. This fact contradicts with the observation of the labels between two signal points $\boldsymbol{s}$ and $\boldsymbol{z}$ made before. Therefore, there is no mapping that satisfies *Property 3.1*.

B. Immunity of The Proposed Mapping to The BSA

The purpose of the BSA is to switch any pair of signal points to lower the asymptotic performance of BICM-ID. To this end, each signal point in the constellation is assigned with a cost function. Starting with an initial random mapping, the BSA sorts the signal points in a descending order of their costs. The signal point with the highest cost is first picked up and switch to the lower-cost signal points. The lowest total cost will be taken into account. If it is smaller than the initial cost, the corresponding switching is selected and the iteration is continued. Otherwise, the signal points with the second highest cost will be considered and the switching operation is implemented. The BSA keeps running until the total cost cannot be lowered.

In what follows, the immunity of the proposed mapping to the BSA is demonstrated. To do so, a random pair of labels of two signal points in the proposed mapping will be switched. It is then shown that the total cost of a new mapping is always larger than that of the original one. It should be mentioned here that it is only of interest to consider the case $N \geq 2$. For $N = 1$, the result is obvious with the optimal SSP mapping.

Consider the quasistatic channel model with the total cost function $\hat{\delta}_B(\boldsymbol{\Psi}, \xi)$ defined in (3.23). For any signal point $\boldsymbol{s}$, the cost function associated with $\boldsymbol{s}$ is defined

as[1] :

$$\hat{\delta}_B^{\boldsymbol{s}} = \sum_{k=1}^{M_c} \hat{\delta}_B^{\boldsymbol{s},k} \tag{B.1}$$

where $\hat{\delta}_B^{\boldsymbol{s},k}$ is the inverse squared Euclidean distance between $\boldsymbol{s}$ and the signal point whose label differs in only 1 bit at position k compared to that of $\boldsymbol{s}$. Clearly, any signal point in $\boldsymbol{\Psi}$ with the proposed mapping has the same cost, which is:

$$\hat{\delta}_B^{\boldsymbol{s}} = \frac{1}{Nd_1^2} + \frac{N}{(N-1)d_1^2 + d_2^2} + \frac{(2N-1)}{(N-1)d_1^2 + d_3^2} = S \tag{B.2}$$

Following are several useful Lemmas for the proof that the proposed mapping is immune (unaffected) by the BSA.

Lemma 3.1: For a given constellation $\boldsymbol{\Psi}$ with a labelling rule ξ, denote the total cost function by $\hat{\delta}_B$. Assume the two signal points $\boldsymbol{s}$ and $\boldsymbol{z}$ have the associated cost functions $\hat{\delta}_B^{\boldsymbol{s}}$ and $\hat{\delta}_B^{\boldsymbol{z}}$, respectively. The switching operation between the labels of $\boldsymbol{s}$ and $\boldsymbol{z}$ results in a different mapping with the new cost functions[2] $\tilde{\delta}_B^{\boldsymbol{s}}$ and $\tilde{\delta}_B^{\boldsymbol{z}}$ for $\boldsymbol{s}$ and $\boldsymbol{z}$ as well as the new total cost $\tilde{\delta}_B$. Then:

$$\hat{\delta}_B - \tilde{\delta}_B = 2\left[\hat{\delta}_B^{\boldsymbol{s}} + \hat{\delta}_B^{\boldsymbol{z}} - \tilde{\delta}_B^{\boldsymbol{s}} - \tilde{\delta}_B^{\boldsymbol{z}}\right] \tag{B.3}$$

The proof of this lemma is quite straightforward. The only thing one should pay attention is the factor 2 appearing in (B.3). This is due to the fact that the switching operation also affects the associated cost functions of the other signal points whose labels differ in 1 bit compared to those of $\boldsymbol{z}$ and $\boldsymbol{s}$.

Lemma 3.2: Consider a N-D constellation $\boldsymbol{\Psi}$ constructed from 8-PSK with $N \geq 2$. Let H_X be a (twisted) hypercube in $\boldsymbol{\Psi}$. Then the maximum $2N$ squared Euclidean distances from any signal point that is not in H_X to all the signal points in H_X are distributed as follows: (i) There are 2 distances that are equal to $(N-1)d_1^2 + d_2^2$; (ii) There are $(2N-2)$ distances that are equal to $(N-2)d_1^2 + d_2^2 + d_3^2$.

[1]In this appendix the dependence of $\hat{\delta}_B(\boldsymbol{\Psi}, \xi)$ on $\boldsymbol{\Psi}$ and ξ is implicitly understood and the parameter $\hat{\delta}_B(\boldsymbol{\Psi}, \xi)$ is simply written as $\hat{\delta}_B$.

[2]It should be mentioned that $\boldsymbol{s}$ and $\boldsymbol{z}$ are not only the signal points that have the new associated cost functions due to the switching operation.

The proof of this lemma is omitted here since it can be easily obtained from the distance profile of $\mathbf{\Psi}$.

Now consider the proposed mapping with two selected signal points $\boldsymbol{s}$ in H_X and $\boldsymbol{z}$ in H_Y. The following two cases of H_X and H_Y are examined.

a) *H_X and H_Y are the same*: From the optimality for the mapping in any hypercube and the fact that the maximum squared Euclidean distance between any two hypercubes is $(N-1)d_1^2 + d_2^2$, it can be seen that $\hat{\delta}_B^{\boldsymbol{s}} \leq \tilde{\delta}_B^{\boldsymbol{s}}$ and $\hat{\delta}_B^{\boldsymbol{z}} \leq \tilde{\delta}_B^{\boldsymbol{z}}$. Therefore, $\hat{\delta}_B - \tilde{\delta}_B \leq 0$ and the BSA prefers the original mapping.

b) *H_X and H_Y are different*: First, consider $\tilde{\delta}_B^{\boldsymbol{z}}$. To compute $\tilde{\delta}_B^{\boldsymbol{z}}$, one needs to look at the $3N$ signal points whose labels differ in only 1 bit compared to $\boldsymbol{s}$ in the original mapping. Based on the distance profile of any signal point in $\mathbf{\Psi}$, one has:

$$\sum_{n=1}^{N} \tilde{\delta}_B^{\boldsymbol{z},n} \geq \frac{1}{Nd_1^2} + \frac{(N-1)}{Nd_1^2 + d_2^2} \tag{B.4}$$

Since $X \neq Y$, it follows from Lemma 2 that:

$$\sum_{n=N+1}^{3N} \tilde{\delta}_B^{\boldsymbol{z},n} \geq \frac{2}{(N-1)d_1^2 + d_2^2} + \frac{(2N-2)}{(N-2)d_1^2 + d_2^2 + d_3^2} \tag{B.5}$$

Combining (B.4) and (B.5) yields:

$$\tilde{\delta}_B^{\boldsymbol{z}} \geq \frac{1}{Nd_1^2} + \frac{(N+1)}{Nd_1^2 + d_2^2} + \frac{(2N-2)}{(N-2)d_1^2 + d_2^2 + d_3^2} = T \tag{B.6}$$

The final step is to prove the quantity S defined in (B.2) is smaller than T defined in (B.6). This is equivalent to proving the following inequality:

$$\begin{aligned} P &= \frac{1}{Nd_1^2 + d_2^2} + \frac{(2N-2)}{(N-2)d_1^2 + d_2^2 + d_3^2} \\ &\geq Q = \frac{(2N-1)}{(N-1)d_1^2 + d_3^2} \end{aligned} \tag{B.7}$$

With $d_1^2 = d_2^2 \cos^2(\pi/8) = 2d_3^2$, it can be easily verified that $P > Q$ with N=2.

When $N > 2$, using Cauchy inequality, one has:

$$\begin{aligned} P &= \frac{1}{Nd_1^2 + d_2^2} + \frac{1}{(N-2)d_1^2 + d_2^2 + d_3^2} \\ &+ \ldots + \frac{1}{(N-2)d_1^2 + d_2^2 + d_3^2} \\ &\geq \frac{(2N-1)^2}{(N-1)d_1^2 + d_2^2 + (2N-2)\left[(N-2)d_1^2 + d_2^2 + d_3^2\right]} \\ &= L \end{aligned} \tag{B.8}$$

Next, it is shown that $L \geq Q$. With a few simple mathematical manipulations, showing $L \geq Q$ is equivalent to showing that $(4N-3)/(4N-2) \geq \cos^2(\pi/8)$. This latter inequality obviously holds for any $N \geq 3$. Thus it can be concluded that $S < T$ for any $N \geq 2$. Fig. B.1 shows the difference between S and T. Observe that $(T-S)$ approaches 0 when N goes to infinity.

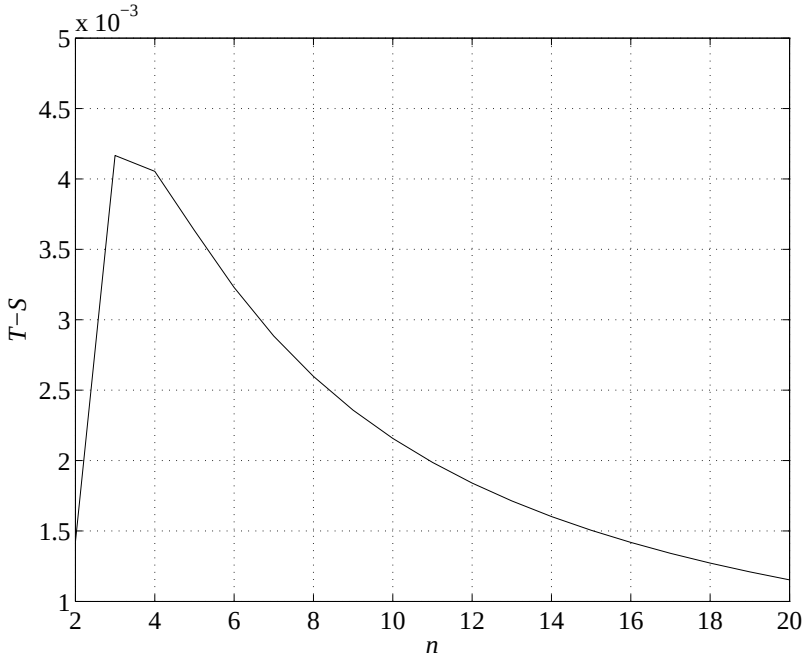


Figure B.1 The plot of $(T-S)$ versus N.

Putting all the above results together, one concludes that $\hat{\delta}_B^{\mathbf{z}} \leq \tilde{\delta}_B^{\mathbf{z}}$. Similarly, one can show that $\hat{\delta}_B \leq \tilde{\delta}_B$, and therefore the BSA prefers the original mapping.

For the case of fast fading channel, it can be shown in the same manner that the

proposed mapping is also immune to the BSA algorithm. However, for brevity of the presentation, such a proof is omitted here.

C. Proof of Lemma 5.2

Let $x \geq 0$ and $y \geq 0$ and $\theta \in (0,1)$. Then one has:

$$\begin{aligned} & h(\theta x + (1-\theta)y) = \int_0^\infty z(\theta x + (1-\theta)y, t)v(t)\mathrm{d}t \\ \leq & \int_0^\infty z^\theta(x,t)z^{1-\theta}(y,t)v(t)\mathrm{d}t \\ \leq & \left(\int_0^\infty z(x,t)v(t)\mathrm{d}t\right)^\theta \cdot \left(\int_0^\infty z(x,t)v(t)\mathrm{d}t\right)^{1-\theta} \\ = & h(x)^\theta \cdot h(y)^{1-\theta}, \end{aligned} \tag{C.1}$$

where the first inequality is due to the log-convexity of $z(y,t)$ and the second inequality is the Holder's inequality [75]. This completes the proof.

D. Optimal Values Of Θ_f^c

When (6.72) is satisfied, one has:

$$F \cdot p_{v,f} - \sum_{q=1}^{F} p_{v,c} = N(f - (F+1)/2). \tag{D.1}$$

Furthermore,

$$p_{v,f} - p_{v,q} = V(f-q). \tag{D.2}$$

Hence, Θ_f^c in (6.66) is rewritten as follows:

$$\Theta_f^c = \left(\frac{1}{2j}\right)^{F-1} \left[\frac{\exp\left(j\pi(f - \frac{F+1}{2})\right)}{\prod_{q\neq f} \sin\left(\frac{\pi}{F}(f-q)\right)}\right]. \tag{D.3}$$

After some manipulations, the product inside the square brackets is computed as:

$$\begin{aligned} &\prod_{q\neq f} \sin\left(\frac{\pi}{F}(f-q)\right) \\ &= (-1)^{F-f}\left(\prod_{i=1}^{F-1} \sin\left(\frac{\pi}{F}i\right)\right) = \frac{(-1)^{F-f}\cdot F}{2^{F-1}} \end{aligned} \tag{D.4}$$

where the last equality follows from the finite product of sin functions. It then follows that:

$$\Theta_f^c = \frac{1}{F}\left(\frac{1}{j}\right)^{F-1} (-1)^{F-f} \exp\left(j\pi\left(f - \frac{F+1}{2}\right)\right). \tag{D.5}$$

Furthermore, for any $1 \leq f \leq F$, it is easy to see that:

$$\left(\frac{1}{j}\right)^{F-1} (-1)^{F-f} \exp\left(j\pi\left(f - \frac{F+1}{2}\right)\right) = 1. \tag{D.6}$$

Therefore,

$$\Theta_f^c = \frac{1}{F}. \tag{D.7}$$

E. Non-Existence of $\boldsymbol{\Theta}$ to Achieve The Lower Bound in (7.18)

Consider any two signal points $\boldsymbol{s}_L = (s_L^1, \ldots, s_L^L)$ and $\boldsymbol{p}_L = (p_L^1, \ldots, p_L^L)$ in Φ that are at the Euclidean distance $8 \cdot L$, where s_L^l and p_L^l, $1 \leq l \leq L$, can be any QPSK symbols. It can be verified that $a_l = (s_L^l - p_L^l)$ can only take on one of the four values in $A = \{a, -a, a^\star, -a^\star\}$ with $a = 2 + j2$. For a given l with $\boldsymbol{\Theta}_l = [\theta_{l,1}, \ldots, \theta_{l,L}]$ be the lth row of $\boldsymbol{\Theta}$, one has the following necessary condition that makes the bound in (7.18) hold true: $\|\boldsymbol{\Theta}_l(\boldsymbol{s}_L - \boldsymbol{p}_L)^\top\| = \|a_1\theta_{l,1} + \ldots a_L\theta_{l,L}\| = \|a\|$. Since $\boldsymbol{\Theta}$ is unitary, there exists at least one non-zero element in $\boldsymbol{\Theta}_l$, say $\theta_{l,1}$. By considering three cases: i) $a_1 = a, a_2 = a, a_l = a$ for $l > 2$; ii) $a_1 = a, a_2 = -a, a_l = -a$ for $l > 2$; and iii) $a_1 = a, a_2 = a^\star, a_l = a^\star$ for $l > 2$, it can be verified that:

$$\theta_{l,2} + \theta_{l,3} + \ldots + \theta_{l,L} = 0. \tag{E.1}$$

Also for the three above cases, substitute a_2 by its conjugate, one obtains:

$$a^\star\theta_{l,2} + a(\theta_{l,3} + \ldots + \theta_{l,L}) = 0. \tag{E.2}$$

From (E.1) and (E.2), it follows that $\theta_{l,2} = 0$. Similarly, it can be proved that all the elements of $\boldsymbol{\Theta}_l$ equal 0 except $\theta_{l,1}$. Therefore, one has the following necessary condition for the unitary matrix $\boldsymbol{\Theta}$ to achieve the lower bound in (7.18):

Condition: There is only one non-zero element in any row of $\boldsymbol{\Theta}$. Furthermore, two non-zero elements in two different rows must be in two different columns. This is to ensure that there is no zero-column in $\boldsymbol{\Theta}$.

From the above condition, it can be seen that the rotation matrix $\boldsymbol{\Theta}$ preserves the squared product distances of Φ in Φ^{r}, which results in the minimum value

$\eta_{\min}(\Phi, \zeta, \boldsymbol{\Theta}) = \eta_{\min}(\Phi, \zeta, \boldsymbol{I}_L)$. Therefore, the lower bound in (7.18) is not achievable.

F. Optimal Mapping of the Orthogonal Design with 4-PSK and N_t =4

With $N_t = 4$, there are totally $K = 64$ signal points of the orthogonal design $\mathbf{\Psi}$ based on M-PSK, where $M = 4$. First, represent any integer κ, $1 \leq \kappa \leq K$, by a set of 3 integers $[i, p, q]$ as:

$$\kappa = iM^2 + pM + q + 1, \tag{F.1}$$

where $0 \leq i, p, q \leq M - 1$. Each signal point $\mathbf{\Psi}(\kappa)$ is then given by [108]:

$$\mathbf{\Psi}(\kappa) = \sqrt{\frac{1}{6}} \begin{pmatrix} 1 & -1 & -1 & 0 & r^i & -r^{-p} & -r^{-q} & 0 \\ 1 & 1 & 0 & 1 & r^p & r^{-i} & 0 & r^{-q} \\ 1 & 0 & 1 & -1 & r^q & 0 & r^{-i} & -r^{-p} \\ 0 & -1 & 1 & 1 & 0 & -r^q & r^p & r^i \end{pmatrix}^{\top} \tag{F.2}$$

where $r = \exp(j2\pi/M)$. It is observed in [108] that for any two unitary signal points $\mathbf{\Psi}(\kappa)$ and $\mathbf{\Psi}(\kappa')$, the four singular values of $\mathbf{\Psi}^{\dagger}(\kappa)\mathbf{\Psi}(\kappa')$ are all equal to

$$d(\kappa, \kappa') = \sqrt{\frac{3 + \cos\frac{2\pi(i-i')}{M} + \cos\frac{2\pi(p-p')}{M} + \cos\frac{2\pi(q-q')}{M}}{6}}. \tag{F.3}$$

It can be verified easily that the minimum chordal distance $d^c_{\min}$ of $\mathbf{\Psi}$ is achieved when $i' = i$, $p' = p$ and $q' = (q + 1) \bmod M$. Therefore, a Gray mapping rule can be implemented easily by assigning the conventional Gray labels of M-PSK signal set to the integers i, p and q, as done in Section 8.3.2.

To come up with the optimal mapping for the iterative system, the chordal distance profile of $\mathbf{\Psi}$ needs to be examined. It is observed that there is a strong similarity of $\mathbf{\Psi}$ and the 6-dimensional hypercube constellation considered in [25]. More specifically, the following facts about chordal distances can be established for any $\mathbf{\Psi}(\kappa) \in \mathbf{\Psi}$.

- *Fact 1*: There is only one signal point $[(i+2) \bmod M, (p+2) \bmod M, (q+2) \bmod M]$ with the maximum chordal distance $d_1^c = 2$ to $\mathbf{\Psi}(\kappa)$, corresponding to the singular value 0.

- *Fact 2*: There are six signal points with the second maximum chordal distance $d_2^c = \sqrt{\frac{10}{3}}$ to $\mathbf{\Psi}(\kappa)$, corresponding to the singular value $\sqrt{\frac{1}{6}}$.

Hence, it follows that the optimal mapping is any mapping that satisfies the following condition.

Condition 8.4: For any signal point $\mathbf{\Psi}(\kappa)$ in $\mathbf{\Psi}$, a signal point with chordal distance d_1^c to $\mathbf{\Psi}(\kappa)$ has label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$. Furthermore, there are five other signal points with chordal distance d_2^c and label Hamming distance 1 to $\mathbf{\Psi}(\kappa)$.

An algorithm to construct the mapping that satisfies Condition 8.4 is provided next. This algorithm is modified from the construction of the optimal mapping for $N_t = 2$ presented in Section 8.3.3 and the algorithm to obtain the optimal mapping of the 6-dimensional hypercube employed in BICM-ID over the coherent channels in [25]. The algorithm includes the following three simple steps.

- *Step 1*: Consider the sub-constellation $\mathbf{\Theta}_1 = \{(i,p,q)\}$ where $i \in \{0,1\}$ and $0 \leq p,q \leq 3$. Label each $[i,p,q]$ by 5 bits (k,a_1,a_2,a_3,a_4) where (a_1,a_2) and (a_3,a_4) are the labels of the pth and the qth signal points of 4-PSK with a conventional Gray mapping. It can be verified that this labeling rule of $\mathbf{\Theta}_1$ is a Gray mapping.

- *Step 2*: Modify the Gray labeling rule as follows. Each label (k,a_1,a_2,a_3,a_4) is changed to (b,k,a_1,a_2,a_3,a_4), where $b = 0$ if $(k+a_1+a_2+a_3+a_4)$ is even; otherwise, $b = 1$. The resulting labeling rule is denoted by ξ_1, where each signal point in $\mathbf{\Theta}_1$ carries 6 bits.

- *Step 3*: For every $[i,p,q] \in \mathbf{\Theta}_1$, construct the signal point $[\hat{i},\hat{p},\hat{q}]$ such that $\hat{i} = (k+2) \bmod 4$, $\hat{p} = (p+2) \bmod 4$ and $\hat{q} = (q+2) \bmod 4$. This construction

Table F.1 The proposed optimal mapping for the orthogonal design with 4-PSK and $N_t = 4$.

$[i,p,q]$	ξ	$[i,p,q]$	ξ	$[i,p,q]$	ξ	$[i,p,q]$	ξ
$[0,0,0]$	(000000)	$[0,0,1]$	(100001)	$[0,0,2]$	(000011)	$[0,0,3]$	(100010)
$[0,1,0]$	(100100)	$[0,1,1]$	(000101)	$[0,1,2]$	(100111)	$[0,1,3]$	(000110)
$[0,2,0]$	(001100)	$[0,2,1]$	(101101)	$[0,2,2]$	(001111)	$[0,2,3]$	(101110)
$[0,3,0]$	(101000)	$[0,3,1]$	(001001)	$[0,3,2]$	(101011)	$[0,3,3]$	(001010)
$[1,0,0]$	(110000)	$[1,0,1]$	(010001)	$[1,0,2]$	(110011)	$[1,0,3]$	(010010)
$[1,1,0]$	(010100)	$[1,1,1]$	(110101)	$[1,1,2]$	(010111)	$[1,1,3]$	(110110)
$[1,2,0]$	(111100)	$[1,2,1]$	(011101)	$[1,2,2]$	(111111)	$[1,2,3]$	(011110)
$[1,3,0]$	(011000)	$[1,3,1]$	(111001)	$[1,3,2]$	(011011)	$[1,3,3]$	(111010)
$[2,0,0]$	(101111)	$[2,0,1]$	(001110)	$[2,0,2]$	(101100)	$[2,0,3]$	(001101)
$[2,1,0]$	(001011)	$[2,1,1]$	(101010)	$[2,1,2]$	(001000)	$[2,1,3]$	(101001)
$[2,2,0]$	(100011)	$[2,2,1]$	(000010)	$[2,2,2]$	(100000)	$[2,2,3]$	(000001)
$[2,3,0]$	(000111)	$[2,3,1]$	(100110)	$[2,3,2]$	(000100)	$[2,3,3]$	(100101)
$[3,0,0]$	(011111)	$[3,0,1]$	(111110)	$[3,0,2]$	(011100)	$[3,0,3]$	(111101)
$[3,1,0]$	(111011)	$[3,1,1]$	(011010)	$[3,1,2]$	(111000)	$[3,1,3]$	(011001)
$[3,2,0]$	(010011)	$[3,2,1]$	(110010)	$[3,2,2]$	(010000)	$[3,2,3]$	(110001)
$[3,3,0]$	(110111)	$[3,3,1]$	(010110)	$[3,3,2]$	(110100)	$[3,3,3]$	(010101)

Note: Each signal point is represented by a string of 3 integers, $[i,p,q]$, and the corresponding label ξ is a binary 6-tuple.

gives the constellation $\hat{\mathbf{\Theta}}_1$. The label of each $[\hat{i},\hat{p},\hat{q}] \in \hat{\mathbf{\Theta}}_1$ is then assigned with $(1-b,k,a_1,a_2,a_3,a_4) \in \hat{\xi}_1$, where $(b,k,a_1,a_2,a_3,a_4) \in \xi_1$ is the label of the corresponding signal point $[i,p,q] \in \mathbf{\Theta}_1$. The final optimal mapping ξ of $\mathbf{\Psi}$ is obtained by combining $\mathbf{\Theta}_1$ and $\hat{\mathbf{\Theta}}_1$ and their labels.

The final results obtained by the above algorithm are shown in Table F.1.

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel," *IEEE Trans. on Commun.*, vol. 40, pp. 873–883, May 1992.

[2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[3] IEEE Standard, "IEEE 802.11a Wireless LAN Medium Access Control and Physical Layer Specifications: High-Speed Physical Layer in the 5GHz Band," in *IEEE std 802.11a-1999*.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correction coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun.*, pp. 1064–1070, 1993.

[5] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 1, pp. 169–171, Nov. 1997.

[6] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 715–724, Apr. 1999.

[7] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. on Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.

[8] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 944–957, May 2001.

[9] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demmaping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf.*, pp. 579–584, 1998.

[10] S. ten Brink, J. Speidel, and R. H. Yan, "Iterative demapping for QPSK modulation," *Electronics Letters*, vol. 34, pp. 1459–1460, July 1998.

[11] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information chart," *AEU Int. J. Electron. Commun*, vol. 54, pp. 389–398, Sept. 2000.

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

[13] S. Lin and D. J. Costello, *Error Control Coding*. Prentice Hall, 2004.

[14] R. G. Gallager, *Low Density Parity-Check Codes*. MIT Press, Cambridge, MA, 1963.

[15] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, Mar. 1999.

[16] K. Boulle and J. C. Belfiore, "Modulation schemes designed for the Rayleigh channel," in *Proc. CISS'92*, pp. 288–293, March, 1992.

[17] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1453–1467, July 1998.

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

[19] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[20] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. on Commun.*, vol. 36, pp. 1004–1011, Sept. 1988.

[21] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[22] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan 1997.

[23] N. Gresset, J. Boutros, and L. Brunel, "Multidimensional mappings for iteratively decoded BICM on multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3337–3346, Sept. 2005.

[24] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[25] N. H. Tran and H. H. Nguyen, "Design and performance of BICM-ID systems with hypercube constellations," *IEEE Trans. on Wireless Commun.*, vol. 5, pp. 1169–1179, May 2006.

[26] A. Martinez, A. G. Fàbregas, and G. Caire, "Error Probability Analysis of Bit-Interleaved Coded Modulation," *IEEE Trans. Inform. Theory*, vol. 52, pp. 262–271, Jan. 2006.

[27] P.-C. Yeh, S. A. Zummo, and W. E. Stark, "Error Probability of Bit-Interleaved Coded Modulation in Wireless Environments," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 722–728, Mar. 2006.

[28] T. Clevorn, S. Godtmann, and P. Vary, "BER Prediction using EXIT Charts for BICM with Iterative Decoding," *IEEE Commun. Letters*, vol. 10, pp. 49–51, Jan. 2006.

[29] N. H. Tran and H. H. Nguyen, "A novel multi-dimensional maping of 8-PSK for BICM-ID," *IEEE Trans. on Wireless Commun.*, vol. 6, pp. 1133–1142, Mar. 2007.

[30] N. H. Tran and H. H. Nguyen, "A Novel Multi-Dimensional Mapping of 8-PSK for BICM-ID," in *Proc. IEEE International Conference on Communications (ICC)-Wireless Communications Symposium,* Istanbul, Turkey, pp. 5004–5009, June 2006.

[31] T. Clevorn, S. Godtmann, and P. Vary, "PSK versus QAM for iterative decoding of bit-interleaved coded modulation," in *Proc. IEEE Global Telecommun. Conf.*, pp. 341–345, November-December, 2004.

[32] N. H. Tran and H. H. Nguyen, "Signal mappings of 8-ary constellation for bit-interleaved coded modulation with iterative decoding," *IEEE Trans. on Broadcasting*, vol. 52, pp. 92–99, Mar. 2006.

[33] N. H. Tran and H. H. Nguyen, "Signal mappings of 8-ary constellations for BICM-ID systems over a Rayleigh fading channel," *IEICE Trans. Commun.*, vol. E88-B, pp. 4083–4086, Oct. 2005.

[34] J. Tan and G. L. Stüber, "Analysis and design of symbol mappers for iteratively decoded BICM," *IEEE Trans. on Wireless Commun.*, vol. 4, pp. 662–672, Mar. 2005.

[35] Y. Huang and J. A. Ritcey, "Improved 16-QAM constellation labeling for BI-STCM-ID with the Alamouti scheme," *IEEE Commun. Letters*, vol. 9, pp. 153–155, Feb. 2005.

[36] Y. Huang and J. A. Ritcey, "Optimal constellation labeling for iteratively decoded bit-interleaved spacetime coded modulation," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1865–1871, May 2005.

[37] F. Simoens, H. Wymeersch, H. Bruneel, and M. Moeneclaey, "Multi-dimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling," *IEEE Commun. Letters*, vol. 9, May 2005.

[38] S. Bäro, "Turbo detection for MIMO systems: bit labeling and pre-coding," *European Transactions on Telecommunications*, vol. 15, pp. 343–350, Aug. 2004.

[39] Y. Li and X.-G. Xia, "Constellation mapping for space-time matrix modulation with iterative demodulation/decoding," *IEEE Trans. on Commun.*, vol. 53, pp. 764–768, May 2005.

[40] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 7, pp. 593–595, Dec. 2003.

[41] H. Schulze, "System design for bit interleaved coded QAM with iterative decoding in a Rician fading channel," *European Transactions on Telecommunications*, vol. 14, pp. 119–129, Apr. 2003.

[42] J. Shu Lin, Daniel J. Costello, *Error Control Coding.* Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632, 1983.

[43] N. H. Tran and H. H. Nguyen, "Multi-dimensional mappings of M-ary constellations for BICM-ID systems," in *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 124–127, May, 2005.

[44] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with Signal Space Diversity," *IEEE Trans. on Wireless Commun.*, vol. 6, pp. 1732–1742, May 2007.

[45] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with Signal Space Diversity," in *Proc. IEEE International Conference on Communications (ICC)-Communication Theory Symposium,* Istanbul, Turkey, pp. 1712–1717, June 2006.

[46] E. Viterbo, *Full diversity rotations.* Available at http://www1.tlc.polito.it/ viterbo/rotations/rotations.html.

[47] G. Wang, H. Liao, H. Wang, and X.-G. Xia, "Systematic and optimal cyclotomic lattices and diagonal space-time block code designs," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3348–3360, Dec. 2004.

[48] A. Chindapol and J. A. Ritcey, "Bit-interleaved coded modulation with signal space diversity in Rayleigh fading," in *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference*, pp. 1003–1007, October, 1999.

[49] Z. A. Khan and B. S. Rajan, "Bit and coordinate interleaved coded modulation," in *Proc. IEEE Global Telecommun. Conf.*, pp. 1595–1599, November-December, 2000.

[50] E. Viterbo and J. Boutros, "A universal lattice decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1639–1642, July 1999.

[51] Y. Li, X.-G. Xia, and G. Wang, "Simple iterative methods to exploit the signal space diversity," *IEEE Trans. on Commun.*, vol. 53, pp. 32–38, Jan. 2005.

[52] L. Ping, "Approximate MMSE-APP estimation for linear systems with binary inputs," *IEEE Commun. Letters*, vol. 9, pp. 172–174, Feb. 2005.

[53] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "On Error Performance of Communications Systems with Signal Space Diversity over Keyhole Nakagami-m Fading Channels," submitted to *IEEE Trans. on Information Theory*, Mar. 2007.

[54] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 502–517, Mar. 1996.

[55] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. on Wireless Commun.*, vol. 2, pp. 294–309, Mar. 2003.

[56] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. IEEE Int. Symp. Inform. Theory*, (Ulm, Germany), p. 374, 1997.

[57] X. Ma and L. Ping, "Coded modulation using superimposed binary code," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3331–3343, Dec. 2004.

[58] T. F. Coleman and C. V. Loan, *Handbook for matrix computations*. Philadelphia : SIAM, 1988.

[59] A. Matache, C. Jones, and R. D. Wesel, "Reduced complexity mimo detectors for ldpc coded systems," in *Proc. IEEE Military Commun. Conf.*

[60] E. Bayer-Fluckiger, F. Oggier, and E. Viterbo, "New algebraic constructions of rotatated Z^n-lattice constellations for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 50, pp. 702–714, Apr. 2004.

[61] P. Dayal and M. Varanasi, "An optimal two transmit antenna space-time code and its stacked extensions," in *Asilomar Conference on Signals, Systems and Computers, Monterey, CA*, November, 2003.

[62] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "BICM-ID with Signal Space Diversity over Cascaded Rayleigh Fading Channels," to appear in *IEEE Trans. on Commun.*

[63] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Application of Signal Space Diversity in BICM-ID over Cascaded Rayleigh Fading Channels," in *Proc. IEEE International Conference on Communications (ICC)- Wireless Communications Symposium,* Glasgow, Scotland, pp. 4011–4016, June 2007.

[64] W. Honcharenko, H. L. Bertoni, and J. L. Dailing, "Bilateral averaging over receiving and transmitting areas for accurate measurements of sector average signal strength inside building," vol. 43, pp. 508–512, May 1995.

[65] V. Erceg, S. Fortune, J. Ling, A. J. Rustako, and R. A. Valenzuela, "Comparisons of a computer-based propagation prediction tool with experimental data collected in urban microcellular environments," *IEEE J. Select. Areas in Commun.*, vol. 15, pp. 677–684, May 1997.

[66] A. S. Akki and F. Haber, "A statistical model of mobile-to-mobile land communication channel," *IEEE Trans. Veh. Technol.*, vol. 35, pp. 2–7, Feb. 1986.

[67] I. Z. Kovacs, *Radio Channel Chracterization for Private Mobile Radio Systems*. Ph.D. Thesis, Aalborg University, Denmark, 2002.

[68] M. Uysal, "Maximum achievable diversity order for cascaded Rayleigh fading channels," *Electronics Letters*, vol. 41, pp. 1289–1290, Nov. 2005.

[69] M. Uysal, "Diversity analysis of space-time coding in cascaded Rayleigh fading channels," *IEEE Commun. Letters*, vol. 10, pp. 165–167, Mar. 2006.

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

[71] H. Shin and J. H. Lee, "Performance analysis of space-time block codes over keyhole Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 351–361, Mar. 2004.

[72] S. Sanayei, A. Hedayat, and A. Nosratinia, "Space-time codes in keyhole channels: Analysis and design," *IEEE Trans. on Wireless Commun.*, vol. 6, pp. 2006–2011, 2007.

[73] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*. Dover Publications, Inc., New York, 1965.

[74] R. Pereira, "Matrix-theoretical derivations of some results of Borcea-Shapiro on hyperbolic polynomials," *C. R. Acad. Sci. Paris, Ser. I 341*, 2005.

[75] D. S. Mitrinovíc, *Analytic inequalities*. Berlin; New York: Springer-Verlag, 1970.

[76] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Bit-interleaved coded OFDM with signal space diversity: Subcarrier grouping and rotation matrix design," *IEEE Trans. Signal Process.*, vol. 55, pp. 1137–1149, Mar. 2007.

[77] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Subcarrier Grouping for OFDM with Linear Constellation Precoding Over Multipath Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 3607–3613, Sept. 2007.

[78] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Optimum Subcarrier Grouping and Rotation Matrix for Coded OFDM with Modulation Diversity," in *Proc. IEEE International Symposium on Information Theory (ISIT),* Seattle, USA, pp. 1384–1388, July 2006.

[79] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Asymptotic Performance of Coded OFDM with Modulation Diversity and Iterative Decoding," in *Proc. 64th IEEE Vehicular Technology Conference (VTC),* Montreal, Canada, pp. 1–5, 2006.

[80] Z. Liu, Y. Xin, and G. B. Giannakis, "Space-Time-Frequency coded OFDM over frequency-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, pp. 2465–2476, Oct. 2002.

[81] Z. Wang and G. B. Giannakis, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 707–720, Mar. 2003.

[82] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for OFDM with maximum multipath diversity and coding gains," *IEEE Trans. on Commun.*, vol. 51, pp. 416–427, Mar. 2003.

[83] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Linear block precoding for OFDM systems based on maximization of mean cutoff rate," *IEEE Trans. Signal Process.*, vol. 53, pp. 4691–4696, Dec. 2005.

[84] U. Sripati, B. S. Rajan, and V. Shashidhar, "Full-diversity Group-Space-Time-Frequency (GSTF) codes from Cyclic codes," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 2080–2084, Sept. 2005.

[85] S. Gowrisankar and B. S. Rajan, "A Rate-one Full-diversity Low-complexity

Space-Time-Frequency Block Code (STFBC) for 4-Tx MIMO-OFDM," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 2090–2094, Sept. 2005.

[86] I. Lee, A. M. Chan, and C.-E. W. Sundberg, "Space-time bit-interleaved coded modulation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 52, pp. 820–825, Mar. 2004.

[87] D. Rende and T. F. Wong, "Bit-interleaved space-frequency coded modulation for OFDM systems," *IEEE Trans. on Wireless Commun.*, vol. 4, pp. 2256–2266, Sept. 2005.

[88] Z. Hong and B. L. Hughes, "Robust space-time codes for broadband OFDM systems," in *Proc. IEEE Wireless Commun. and Networking. Conf.*, pp. 105–108, Mar. 2002.

[89] Z. Wang, S. Zhou, and G. B. Giannakis, "Joint coding-precoding with low-complexity turbo-decoding," *IEEE Trans. on Wireless Commun.*, vol. 3, pp. 832–842, May 2004.

[90] ETSI Normalization Commitee, "Channel models for HIPERLAN/2 in different indoor scenarios," in *Doc. 3ER10085B, Eur. Telecommun. Standards Inst.,*, (Sophia-Antipolis, France).

[91] B. Lu and X. Wang, "Space-time code design in OFDM systems," in *Proc. IEEE Global Telecommun. Conf.*, pp. 1000–1004, Nov. 2000.

[92] D. Sharpe, *Rings and Factorization.* Cambridge University Press, 1987.

[93] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C.* Cambridge University Press, 1992. Available online at http://www.library.cornell.edu/nr/cbookcpdf.html.

[94] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Multi-dimensional Subcarrier Mapping for Bit-Interleaved Coded OFDM with Iterative Decoding," *IEEE Trans. Signal Process.*, vol. 55, pp. 5772–5781, Dec. 2007.

[95] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Multi-dimensional subcarrier mapping for bit-interleaved coded OFDM with iterative decoding," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC),* Hong Kong, Mar. 2007.

[96] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on mutiple-antenna channel," *IEEE Trans. on Commun.*, vol. 51, pp. 389–399, Mar. 2003.

[97] X. Giraud, E. Boutillon, and J. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 938–952, May 1997.

[98] ITU-R Recommendation M.1225, "Guidelines for evaluation of radio transmission technologies for IMT-2000, 1997,"

[99] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Coded unitary space-time modulation with iterative decoding: Error performance and mapping design," *IEEE Trans. on Commun.*, vol. 55, pp. 703–716, Apr. 2007.

[100] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Tight Error Bound for Coded Unitary Space-Time Modulation," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC),* Las Vegas, USA, pp. 1893–1898, April 2006.

[101] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

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

[103] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[104] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space-time receiver," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2383–2399, Sept. 2001.

[105] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1962–1973, Sept. 2000.

[106] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for noncoherent space-time communcation on the block-Rayleigh-fading channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1186–1194, May 2002.

[107] D. Agrawal, T. J. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2618–2626, Sept. 2001.

[108] W. Zhao, G. Leus, and G. B. Giannakis, "Orthogonal design of unitary constellations for uncoded and trellis-coded noncoherent space-time systems," *IEEE Trans. Inform. Theory.*

[109] I. Bahceci and T. M. Duman, "Combined tubo coding and unitary space-time modulation," *IEEE Trans. on Commun.*, vol. 50, pp. 1244–1249, Aug. 2002.

[110] Y. Li and X.-G. Xia, "Constellation mapping for space-time matrix modulation with iterative demodulation/decoding," *IEEE Trans. on Commun.*, vol. 53, pp. 764–768, May 2005.

[111] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[112] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[113] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. on Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[114] G. Han, "Space time coding with multiple antenna systems," *Ph.D. Thesis,* Graduate Programs in Mathematics, Notre Dame, Indiana, 2004.

[115] J. Wu and H.-N. Lee, "Best mapping for LDPC coded modulation on SISO, MIMO and MAC channels," in *Proc. IEEE Wireless Commun. and Networking. Conf.*, pp. 2428–2431, Mar. 2004.

[116] R. Schober, L. H. J. Lampe, W. H. Gerstacker, and S. Pasupathy, "Modulation diversity for frequency-selective fading channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2268–2276, Sept. 2003.

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

[118] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, July 2000.

[119] S. Y. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 371–381, Feb. 2003.

[120] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, pp. 782–795, Oct. 2003.