BANDWIDTH-EFFICIENT COMMUNICATION SYSTEMS BASED ON FINITE-LENGTH LOW DENSITY PARITY CHECK CODES

A Thesis 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 and Computer Engineering
University of Saskatchewan

by

Huy G. Vu

Saskatoon, Saskatchewan, Canada

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate 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 thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised this thesis work or, in their absence, by the Head of the Department of Electrical and 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 thesis, 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 thesis.

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

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

ACKNOWLEDGMENTS

This thesis would not have existed without the help, support and love of many people. The foremost one is from my supervisor, Professor Ha Nguyen. The fruitful discussions with him gave me the insightful understanding of fundamental problems of communication engineering. His endless support encouraged me throughout the course of graduate studies. I would like to express my deepest appreciation and gratitude to him for his guidance and support during my studies.

I wish to thank my co-supervisor, Prof. David E. Dodds. His constant support, encouragement and comments helped me so much in order to complete this thesis.

I also would like to thank the other members of my advisory committee: Professors E. Shwedyk, E. J. Salt and S. B. Ko for serving on my advisory committee and for providing me with their comments and suggestions.

The financial support for this work by the University of Saskatchewan's graduate scholarship and Natural Sciences and Engineering Research Council of Canada (NSERC) research grants are gratefully acknowledged. These financial sources are important to help me keep working on my research and finish this thesis.

Through the course of my research, I have enjoyed good times, especially the seminars, with my lab-mates, Nghi Tran, Son Nguyen, Trung Bui, Hieu Nguyen and Jajun Yang here, in Saskatoon. My hearfelt thanks also go to other friends, Bernie Boos, Sivapathasundaram Sureshkumar and Tung Lai for their help, understanding and encouragement. Without them, life here would have been more difficult.

Last but not the least, my deepest love and gratitude is devoted to my family, especially my parents. This is all for them.

Look at every path closely and deliberately. Then ask yourself, and yourself alone, one question: Does this path have a heart? If it does, the path is good; if it doesn't it is of no use. *Carlos Castaneda*

ABSTRACT

Low density parity check (LDPC) codes are linear block codes constructed by pseudo-random parity check matrices. These codes are powerful in terms of error performance and, especially, have low decoding complexity. While infinite-length LDPC codes approach the capacity of communication channels, finite-length LDPC codes also perform well, and simultaneously meet the delay requirement of many communication applications such as voice and backbone transmissions. Therefore, finite-length LDPC codes are attractive to employ in low-latency communication systems. This thesis mainly focuses on the bandwidth-efficient communication systems using finite-length LDPC codes. Such bandwidth-efficient systems are realized by mapping a group of LDPC coded bits to a symbol of a high-order signal constellation. Depending on the systems' infrastructure and knowledge of the channel state information (CSI), the signal constellations in different coded modulation systems can be two-dimensional multilevel/multiphase constellations or multi-dimensional space-time constellations.

In the first part of the thesis, two basic bandwidth-efficient coded modulation systems, namely LDPC coded modulation and multilevel LDPC coded modulation, are investigated for both additive white Gaussian noise (AWGN) and frequency-flat Rayleigh fading channels. The bounds on the bit error rate (BER) performance are derived for these systems based on the *maximum likelihood* (ML) criterion. The derivation of these bounds relies on the union bounding and combinatoric techniques. In particular, for the LDPC coded modulation, the ML bound is computed from the Hamming distance spectrum of the LDPC code and the Euclidian distance profile of the two-dimensional constellation. For the multilevel LDPC coded modulation, the bound of each decoding stage is obtained for a *generalized* multilevel coded modulation, where more than one coded bit is considered for level. For both systems, the bounds are confirmed by the simulation results of ML decoding and/or the performance of the ordered-statistic decoding (OSD) and the sum-product decoding. It is demonstrated that these bounds can be efficiently used to evaluate the error performance and select appropriate parameters (such as the code rate, constellation and mapping) for the

two communication systems.

The second part of the thesis studies bandwidth-efficient LDPC coded systems that employ multiple transmit and multiple receive antennas, i.e., multiple-input multiple-output (MIMO) systems. Two scenarios of CSI availability considered are: (i) the CSI is unknown at both the transmitter and the receiver; (ii) the CSI is known at both the transmitter and the receiver. For the first scenario, LDPC coded unitary space-time modulation systems are most suitable and the ML performance bound is derived for these non-coherent systems. To derive the bound, the summation of chordal distances is obtained and used instead of the Euclidean distances. For the second case of CSI, adaptive LDPC coded MIMO modulation systems are studied, where three adaptive schemes with antenna beamforming and/or antenna selection are investigated and compared in terms of the bandwidth efficiency. For uncoded discrete-rate adaptive modulation, the computation of the bandwidth efficiency shows that the scheme with antenna selection at the transmitter and antenna combining at the receiver performs the best when the number of antennas is small. For adaptive LDPC coded MIMO modulation systems, an achievable threshold of the bandwidth efficiency is also computed from the ML bound of LDPC coded modulation derived in the first part.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

γ	The average signal-to-noise ratio
$\lambda(x)$	The variable node degree distribution
λ_i	The ith eigenvalue
$\mathbf{\Lambda}_\tau$	The covariance matrix of $\mathbf{Y}_\tau$
$\rho(x)$	The check node degree distribution
σ^2	The noise variance
Ψ	The unitary space-time constellation
$\mathbf{\Psi}_b^k$	The subsets of $\mathbf{\Psi}$, where the kth bits of the mapping labels are b
c_j	The check node j
c_k	The transmitted signal at time index k
C	The capacity of the channel
$E(R)$	The channel reliability function
$f(\mathbf{n})$	The number of error sequences of type $\mathbf{n}$
h_k	The channel coefficient at the time index k
$\mathbf{h}$	The channel coefficient vector
$\mathbf{H}_\tau$	The channel coefficient matrix at time index τ
K	The number of information bits encoded to a codeword
$\mathbf{n}$	The type of error sequences
N	The length of the codeword
N_0	The noise power spectral density
N_s	The number of modulation symbols corresponding to a codeword
N_r	The number of receive antennas
N_t	The number of transmit antennas
$\mathcal{N}(l)$	The number of codewords of weight l
$P_{l,\mathbf{n}}$	The number of possible cases of error sequences of type $\mathbf{n}$ with a weight l
P_e	The bit error probability

R	The data rate
R_0	The cutoff rate
R_c	The code rate
S	The power of a channel symbol
T	The number of time intervals with a constant fading coefficient
v_i	The variable node i
w_k	The noise at time index k
$\mathbf{w}$	The noise vector
W	The bandwidth of the channel
$\mathbf{W}_\tau$	The noise matrix at time index τ
$\underline{\mathbf{W}}$	The noise sequence
x_k	The transmitted signal at time index k
$\mathbf{x}$	The transmitted vector
$\mathbf{X}_\tau$	The transmitted matrix at time index τ
$\mathcal{X}$	The 2-dimensional constellation
$\mathcal{X}_b^k$	The subsets of $\mathcal{X}$, where the kth bits of the mapping labels are b
$\underline{\mathbf{X}}$	The transmitted sequence
y_k	The received signal at time index k
$\mathbf{y}$	The received vector
$\mathbf{Y}_\tau$	The received matrix at time index τ
$\underline{\mathbf{Y}}$	The received sequence
z	The symbol of a normalized constellation in adaptive modulation

ARQ	Automatic Repeat Request
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit Interleaved Coded Modulation
BICM-ID	Bit Interleaved Coded Modulation with Iterative Decoding

BLAST	Bell Labs Layered Space-Time
BPSK	Binary Phase Shift Keying
CDF	Cumulative Distribution Function
CSI	Channel State Information
dB	Decibel
EDGE	Enhanced Data Rates for GSM Evolution
EXIT	Extrinsic Information Transfer
FDD	Frequency Duplex Division
FFT	Fast Fourier Transform
GF	Galois Field
GMD	Generalized Minimum Distance
GSM	Global System for Mobile Communications
MAP	Maximum A Posteriori Probability
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MLC	Multilevel Coding
MRB	Most Reliable Basis
MRIP	Most Reliable Independent Positions
OFDM	Orthogonal Frequency Division Multiplexing
OSD	Ordered Statistic Decoding
PAM	Pulse Amplitude Modulation
PDF	Probability Density Function
PEG	Progressive Edge Growth
PEP	Pairwise Error Probability
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
LAPPR	Log a Posteriori Probability Ratio
LDPC	Low Density Parity Check
LLR	Log Likelihood Ratios

SISO	Soft-Input Soft-Output
SNR	Signal to Noise Ratio
SP	Sum-Product
SSP	Semi Set Partitioning
TCM	Trellis Coded Modulation
TDD	Time Duplex Division
VLSI	Very Large Scale Integrated Circuit

1. Introduction

1.1 Digital Communication Systems

The fundamental problem of communications is to reproduce at one point a digital message selected at another point as exactly as possible [1]. To solve this fundamental problem, communication engineers have designed sophisticated systems to transmit messages over hostile noisy channels. The general block diagram of a digital communication system is illustrated in Figure 1.1.

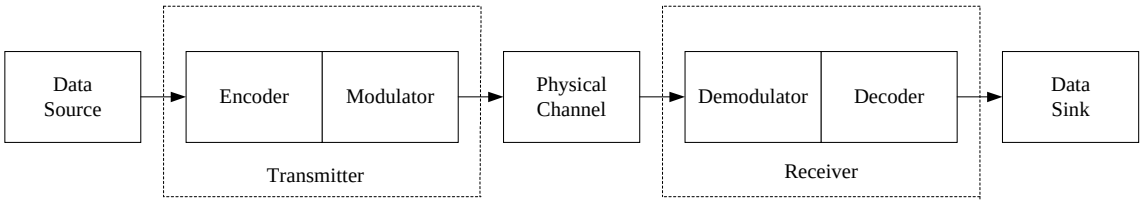


Figure 1.1 The block diagram of a digital communication system.

At the transmitter, digital messages are processed before they are sent to the physical channel. The objective of this process is twofold: (i) to choose proper signal waveforms to avoid bad effects of the physical channel and (ii) to be able to detect these waveforms easily at the receiver end. At the receiver, a computational algorithm is needed to detect the transmitted waveforms as precisely as possible with a practical complexity.

In digital communication systems, the modulator is the block that maps the information to the physical channel. Due to the continuous nature of most physical channels, the modulator needs to transform discrete waveforms to continuous waveforms that adapt to the channel. Because the physical channel attenuates the transmitted signal, creates random noise and presents interference, the continuous waveforms should be chosen to cope with

attenuation, noise and interference. For example, in mobile wireless communications, the frequency, phase and bandwidth of the modulated signal are selected properly to reduce the fading effects of the wireless channel, corruption of random noise and interference among users. In modern communications, the modulating function should be understood as any method to efficiently adapt to the channel. For example, these methods may include moving the signal to high frequency, spreading the spectrum of the signal, modulating the signal with many carriers or even sending the signal to many antennas.

The demodulator processes the received continuous waveforms that are corrupted by random factors of the physical channel. Usually, the demodulator tries to replace the received continuous waveforms by finite-dimensional vectors to enable the calculation of the decision variable based on the joint density functions of random variables [2]. In practice, matched filters are often used in digital communication systems to transform continuous waveforms to sufficient statistics, i.e., finite-dimensional vectors. These random variables or sufficient statistics are processed by computational algorithms more easily.

On the other hand, the coding process in digital communications approaches the problem digitally or in the discrete time domain. The encoder can be divided into two blocks, namely the source encoder and the channel encoder. In this research, we are only concerned with the channel encoder and shall refer to it simply as the encoder. The encoder also tries to create waveforms to transmit effectively the information message against random factors of the channel as in the modulator block, but this is done in the discrete domain. The encoder implements this task by inserting redundant information in the message in a controlled manner. At the receiver, the decoder recovers the original information from the discrete outputs of the demodulator with the help of this redundant information.

From the above discussion, it is important to emphasize that channel coding and modulation have the same objective of producing the appropriate signal waveforms to cope with the noisy channel. However, compared to modulation, coding is more flexible in terms of the performance and processing complexity tradeoff due to working in the discrete time domain. In modern communication applications, the two functions can not be separated and should be considered as an entity. It is sometimes observed that a new modulation scheme

shows a large gain for the uncoded systems, however, for coded systems, the improvement does not correspond or even is insignificant [3]. Therefore, performance evaluation of the whole digital communication system should be carried out instead of separate performance assessment of the modulation and coding functions. In this thesis, the error performance of such coded systems is studied.

1.2 Research Motivation

In his landmark paper published in 1948 [1], Shannon proved that the arbitrarily reliable transmission over a noisy channel is possible if the information rate is less than a quantity called the channel capacity. Before his work, people still believed that communications with arbitrarily small bit error probability can only be achieved with information rate approaching zero. The work of Shannon suggests that this channel capacity can be achieved by a very long random code and a maximum likelihood (ML) receiver, that is the optimum receiver. Although the existence of communication schemes signaling at the channel capacity is proven, practical solutions to achieve the Shannon limit are still open over a half of century due to the very high complexity of the maximum likelihood receiver for long block-length codes.

Recently, there has been a great improvement in coding techniques towards achieving this limit. The new approach uses pseudo-random codes and suboptimum decoders, called iterative decoders, instead of the optimum one. Due to this iterative decoding technique, the performance of coded systems is significantly improved due to the ability to increase the code length and, simultaneously, still keep a reasonable computational complexity for the receivers.

In 1993, Turbo code was invented by Berrou, Glavieux and Thitimajshima [4]. This awkward code is constructed by parallel concatenation of two convolutional codes through a random interleaver. Iterative decoding is applied in the decoder where the soft output is exchanged between the two component decoders. The performance of original Turbo code approaches the Shannon limit within 0.5 dB with this iterative decoding technique. In wake of Turbo codes, Gallager's low density parity check (LDPC) codes [5, 6] were

rediscovered and it was shown that with long block-length codes they also achieve near Shannon limit performance. Low density parity check codes are linear block codes with *sparse* parity check matrices generated randomly and these codes can also be decoded iteratively. The performance of Turbo codes and LDPC codes with a very long block-length depends on the convergence of the iterative decoding algorithms. The performance can be evaluated by either the extrinsic information transfer (EXIT) chart [7] or density evolution techniques [8].

By using density evolution to optimize the performance of LDPC codes, the authors in [8–10] found irregular LDPC codes which perform much better than Gallager's regular LDPC codes. So far, the best known error control code over additive white Gaussian noise (AWGN) channels is an irregular LDPC code in [11] whose empirical performance achieves the bit error rate (BER) of 10^{-6} within 0.04 dB of Shannon limit with a block length of 10^7. Theoretically, the code threshold is within 0.0045 dB of Shannon limit which can be reached when the block length tends to infinity.

However, this long block length is impractical for applications that require low latency such as speech or backbone transmissions [12]. For LDPC codes with short and medium block lengths, the error performance of a LDPC coded system depends on both the convergence property and the Hamming distance spectrum of the code. The convergence property determines how the performance of the iterative decoder can approach that of the maximum likelihood decoder. On the other hand, the Hamming distance spectrum of the codes determines the performance of the ML receiver. For short and medium length codes or finite-length codes, specifically high-rate codes, the error performance of ML decoder serves as a lower bound for the iterative decoder's performance. The convergence of the iterative decoder's performance to ML decoder's performance can still be observed at practical bit error rate (BER) levels for finite lengths of the codes.

Compared to Turbo codes, LDPC codes are more flexible in construction in terms of the code rate and other parameters. Moreover, there is an error floor on the error performance of Turbo codes due to the poor Hamming distance spectra of these codes. On the other hand, the error performance of LDPC codes does not clearly show an error floor because

of the good Hamming distance property of the codes. This fact makes finite-length LDPC codes a good candidate for applications that require very low bit error rate (BER), and simultaneously, low delay. Another advantage of LDPC codes is the ability to implement fully parallel decoders thanks to the mechanism of their decoding algorithm. The parallel iterative decoding algorithms of LDPC codes are easily implemented in VLSI [13]. Due to these advantages, LDPC codes are proposed for most future data applications such as wireless, wireline communications and storage systems [12]. Motivated by the successes and potential of LDPC codes and the technique of iterative decoding, this thesis studies digital communication systems based on LDPC codes.

In the design of communication systems, engineers try to achieve some targets with limited resources in order to optimize the economical aspect of the systems and to meet the demands of a given application. Usually, the targets are conflicting and there are trade-offs among them. Some common objectives of the design can be summarized as follows: (i) to obtain the highest reliability, i.e., the lowest error probability, with a limited transmission power; (ii) to simplify the processing at both the transmitter and the receiver, but still provide good performance; and (iii) to achieve the highest data rate with a constraint on the available bandwidth. As discussed before, the LDPC codes that are chosen to study provide good solutions to offer the highest error performance with the constraints on the average power, computational complexity and delay. Other constraints such as the bandwidth and data rate can be more easily manipulated by the modulation function. In practice, since the encoders and the decoders for various purposes are often packaged on VLSI with a given set of parameters [13], communication engineers need to properly choose the modulation techniques to meet the requirement of a specific system. In this thesis, a number of communication systems are investigated with LDPC codes and different modulation schemes, especially the bandwidth-efficient ones. This is important because bandwidth is a scarce and very expensive resource in wireless communications.

In all bandwidth-efficient coded communication systems, the coded bits (that are encoded by an encoder or even several encoders) are grouped and mapped to a continuous electrical signal waveform, called a symbol. The set of these symbols forms a constellation

and the number of the symbols is the size of the constellation (which is often more than two). Because a number of bits is transmitted over one signal interval instead of one bit, the bandwidth is more efficiently utilized. At the receiver, the likelihood ratios (or the soft estimations) of the coded bits are computed by the demodulation block. Evaluating the performance of the systems under ML decoding is carried out by considering the whole sequence of transmitted symbols.

For wireless communications, the channel is time-varying and experiences fading. Typically the modulation function tries to combat the bad effect of wireless channels in different ways. Various modulation techniques have been developed for different scenarios according to the parameters of the wireless channels and the application, such as the coherence time, delay, reliability requirement and the number of equipped antennas. For example, for fixed wireless channels, the channel state information (CSI) changes very slowly and can be easily estimated at the receiver and sent back to the transmitter. Based on the feedback information, the modulation block can change the constellation size, i.e., the date rate of the system, to adapt to the condition of the channel. However, for mobile wireless channels, this technique might not be applicable and other bandwidth-efficient modulation techniques such as using multiple transmit and receive antennas can be considered. When a system is equipped with multiple antennas at the transmitter and the receiver, there are many propagation paths from the transmitter to the receiver. If the distances among the transmit antennas and among the receive antennas are large enough, the fading coefficients of the paths are independent. A signal symbol in this system can be spread over independent fading paths and detected properly at the receiver. When a path is in deep fade (i.e., the path gain is very small), other paths are likely still good, hereby maintaining the quality of the transmission.

The schemes chosen to study in this thesis are for the popular scenarios of wireline and wireless communications. Since the systems based on infinite-length LDPC codes were intensively investigated in [14, 15], the bandwidth-efficient coded communication systems that are based on LDPC codes of finite block-lengths are the focus of this thesis. The main objective is to evaluate the bit error rate (BER) performance of various bandwidth-efficient

LDPC coded communication systems. BER is the most common criterion for performance evaluation of a communication system. Although frame error rate (FER) or block error rate (BLER) can also be used, they are more suitable for data applications in which automatic retransmission request (ARQ) is incorporated.

1.3 Thesis Outline

This thesis includes seven chapters. The first chapter gives an introduction to the digital communication systems, the motivation and the outline of the thesis.

In Chapter 2, the background of error control coding and LDPC codes are presented. The relevant concepts introduced in this chapter are helpful for the exposition of subsequent chapters.

The LDPC coded modulation is studied in Chapter 3. The bandwidth-efficient coded modulation scheme in this chapter is a basic scheme in which a group of coded bits of one LDPC encoder is simply mapped to a constellation symbol. This scheme is a natural approach for block coded modulation. In context of LDPC codes, this scheme corresponds to the bit-interleaved coded modulation (BICM) with convolutional codes [16,17]. The BICM scheme based on LDPC codes does not need an interleaver between the LDPC encoder and the modulator due to random construction of LDPC codes. Our main contribution in this chapter is the derivation of an upper bound for the error performance of the ML receiver. The union bound is compared with simulation results of two different iterative demodulation/decoding receivers that are based on sum-product decoding [5, 6] and the ordered statistic decoding (OSD) [18] with different mappings, constellations, and code lengths over additive white Gaussian noise (AWGN) and fading channels. For medium length codes, the performance of the iterative demodulation/sum-product decoding is close to the performance of the ML decoding at low BER, especially for the system with high-rate codes and high-order modulation. The union bound is also derived for the frequency-flat Rayleigh fading channels when the channel state information (CSI) is known at the receiver. For short length codes, the performance gap between the sum-product decoder and the ML decoder suggests the application of a better decoding algorithm. The computa-

tional complexity of OSD is relatively reduced compared with the ML decoding, although its performance is still close to the performance of ML decoding. In this work, the OSD algorithm is also considered for coded modulation systems. For a large constellation, the iteration between the demodulation and OSD decoding improves the error performance of the system. The error performance of the iterative demodulation/OSD is very close to that of ML decoding for coded modulation system with Gray mapping. Compared to iterative demodulation/sum-product decoders, the receiver based on OSD outperforms for short-length codes, although its complexity makes it impractical for medium length codes. For fading channels, the iterative receiver based on OSD also shows a performance near the ML performance for the short-length codes.

Chapter 4 studies multilevel coded modulation schemes based on LDPC codes. These systems are bandwidth-efficient coded modulation systems proposed in [19]. In these systems, the data stream is divided into substreams and these substreams are separately encoded before they are mapped to the same signal constellation. This chapter investigates the *generalized* multilevel coded modulation with multistage decoding, where finite-length LDPC codes are used as component codes and where the coded bits of each component code are mapped to more than one labelling bits of the constellation symbols. The union bound on the bit error probability for each decoding stage is derived. The bounds, obtained for both additive white Gaussian noise (AWGN) and Rayleigh fading channels, are applicable for any code rate, constellation and mapping. The tightness of the bounds is verified by simulation results of ordered statistic decoding (OSD) and sum-product decoding of LDPC codes. The bounds are useful to benchmark the performance and to select the appropriate parameters of multilevel coded modulation systems in order to provide unequal error protection of different data streams.

In Chapter 5, LDPC coded unitary space-time modulation systems are investigated. These systems use multiple transmit and multiple receive antennas (multiple-input and multiple-output or MIMO). For MIMO channels, space-time codes are widely studied with the assumption that the channel coefficients among different pairs of transmit antennas and receive antennas are independent and known to the receiver [20–22]. However, for many

mobile communication channels, the CSI of MIMO channels is difficult to estimate. In this case, a class of signal constellations, known as unitary space-time constellations, is the most suitable modulation for MIMO channels [23, 24]. In Chapter 5, error performance of non-coherent LDPC coded unitary space-time modulation systems is studied. A bound on the performance is derived for the systems built from finite-length LDPC codes and unitary space-time constellations. The analytical derivations are substantiated by simulation results of OSD and sum-product decoding.

Chapter 6 presents adaptive coded modulation systems based on LDPC codes over MIMO channels [25]. These systems are variable-rate systems in which the CSI is known at both the receiver and the transmitter. Therefore, the data rate or the bandwidth efficiency and the transmitted power can be dynamically adapted to the conditions of the wireless channel. For these adaptive modulation systems, three diversity schemes over the MIMO channel are investigated. One scheme is based on optimum beamforming at both the transmitter and the receiver. Two other schemes are based on antenna selection at the transmitter. In the scheme that is based on transmit beamforming, the bits are loaded to the parallel eigen-subchannels. The schemes based on transmit antenna selection are also chosen for investigation because these schemes reduce the required feedback information of the channel state information. The spectral efficiency of the LDPC coded system is computed based on the simulation results of the iterative demodulation/sum-product decoder and the performance bound of the ML decoder derived in this thesis. The spectral efficiency of the systems based on the ML bound is called the achievable rate threshold. This rate threshold can be achieved by a decoder whose performance approaches that of the ML decoder. The results show the superiority of the scheme with antenna selection at the transmitter and antenna beamforming at receiver when the number of antennas is small.

Finally, the conclusions are given in Chapter 7. Further research problems and topics also are presented in this chapter.

2. Background of Error Control Coding and LDPC Codes

The objective of this chapter is to introduce the basic concepts of error control coding in general and LDPC codes in particular. These concepts serve as background material for subsequent chapters. The maximum likelihood (ML) decoding and the sum-product decoding are presented for general codes and LDPC codes. Issues affecting the error performance of LPDC codes under the ML and sum-product decoders, such as the distance and convergence properties, are also discussed.

2.1 Error Control Coding

As mentioned before, the main function of digital communications is to transmit digital messages over a noisy channel as reliably as possible. Consider a channel that has a certain bandwidth W starting at zero frequency (i.e., a baseband channel), and this channel is used for a duration T. It means that arbitrary signal functions of time can be used if their spectra lie entirely within the band W and their time functions lie within the interval T. Although it is not possible to fulfill both of the above conditions on bandwidth and duration exactly, one can choose functions that are very small outside the interval T, and simultaneously, their spectra are kept within the band W. This fact is due to the well-known sampling theorem which can be stated as follows:

Theorem [26]: If a function $f(t)$ contains no frequencies higher than W Hz, it is completely determined by giving its ordinates at a series of points spaced $1/(2W)$ seconds apart.

The theorem can be intuitively interpreted as follows: If $f(t)$ contains no frequencies

higher than W, it cannot change to a substantially new value in a time less than one-half cycle of the highest frequency. Thus, any function $f(t)$ can be specified by $2TW$ real numbers or a point in a $2TW$-dimensional Euclidean space.

From the viewpoint of Euclidean signal space, one can say that a transmission scheme, designed to transmit digital messages of K information bits, is a mapping from a set of 2^K K-tuples to 2^K signal points in the $2TW$-dimensional signal space. In this way, the design of a digital communication system is simplified to a problem that can be considered to be totally in the discrete domain.

However, when the time-bandwidth product WT is large, the number of dimensions to represent $f(t)$ is very large. It means that the signal design problem is both complex at the transmitter and also very complex at the receiver since detection requires computing in the large-dimensional space. In most digital communication systems, to overcome the above complications, the signal design problem at the transmitter is separated into two steps known as channel coding and modulation as shown in Fig. 1.1. An intermediate signal space of N dimensions is used for this separation. The K-tuple is first mapped to a digital message of N bits, i.e., an N-tuple. Then, this N-tuple is mapped to a function $f(t)$ in the $2TW$-dimensional space. In practice, the second mapping is often a simple rule, hence, the performance of the system strongly depends on the first step, i.e., the channel encoder.

A binary (N, K) block code $\mathcal{C}$ is often used to denote the mapping from a K-tuple of information bits to the N-tuple, called the codeword. The code $\mathcal{C}$ can be described as:

$$\mathcal{C}(N, K) = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2^K}\},\ \mathbf{x}_i \in \{0, 1\}^N,\ 1 \leq i \leq 2^K. \tag{2.1}$$

The ratio $R_c = K/N$ is called the code rate. In the system illustrated in Fig. 1.1, the coded bits are processed by the modulation block. The simplest rule for modulation is binary phase-shift keying (BPSK), where coded bits are simply mapped to two antipodal sinusoids of duration T/N. When this BPSK modulation is employed, the pair of the modulator and the demodulator can be modeled by a discrete channel for the design of channel coding. When the physical channel is AWGN, this discrete channel is simply a discrete memoryless

channel (DMC). The error correction capability of different channel coding schemes can be evaluated over this DMC.

2.1.1 The Coding Theorem

When one considers the error performance of digital communication systems, the coding theorem due to Shannon [1] is one of the most general forms of the error performance evaluation and needs to be understood. The coding theorem can be stated as follows.

Theorem [1]: Associated with each discrete memoryless channel (DMC), there is a channel capacity $C > 0$ with the following property: For any $\epsilon > 0$ and $R < C$, there exist a coding scheme of data rate R and a decoding algorithm such that the probability of error is less than ϵ.

For a bandlimited AWGN channel, this channel capacity is computed as follows [1]:

$$C = W \log \left(1 + \frac{P}{N_0 W} \right) \tag{2.2}$$

where P is the average power of the received signal, W is the bandwidth of the channel and N_0 is the one-sided power spectral density of the noise. This capacity can be achieved by an infinite-length random code, where 2^K codewords are randomly chosen from 2^N N-tuples and N is very large. However, for practical systems, the length of the code cannot be infinite and is restricted by the delay requirement of the systems.

For codes of finite length N, a more general form of the coding theorem for the DMC can be restated as follows.

Theorem [27]: The error probability for a code of block length N is upper bounded by:

$$P_e \leq e^{-NE(R)} \tag{2.3}$$

where R is the data rate (bit/sec) that can be computed from the code rate R_c, and $E(R)$ is called the channel reliability function.

The function $E(R)$, illustrated in Fig. 2.1, is a convex, decreasing and non-negative function for $0 \leq R \leq C$. The original Shannon's coding theorem can also be obtained by observing this function. When the data rate R approaches the channel capacity C, the

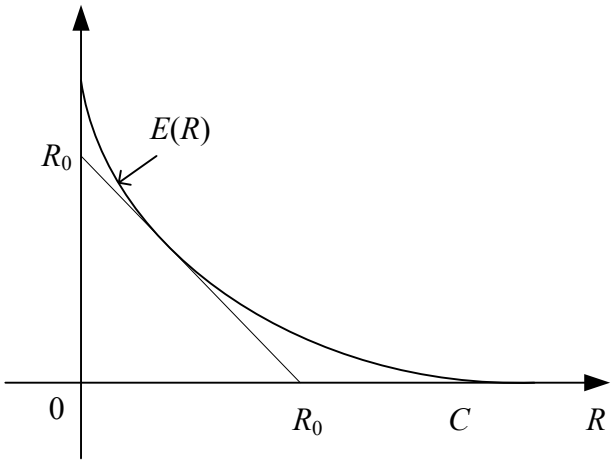


Figure 2.1 The channel reliability function $E(R)$.

reliability function decreases to zero. It follows from (2.3) that one should increase the code length N to infinity to make the error probability arbitrarily small at the data rate equal to the channel capacity. Beside the channel capacity C, some other parameters of the channel, for example the cut-off rate R_0, can be related to the reliability function $E(R)$. As illustrated in Fig. 2.1, the cut-off rate R_0 is the rate at which the tangent to $E(R)$ of slope -1 intersects the R axis. The quantity R_0 shows how difficult to approach the channel capacity C.

This general form of coding theorem is derived in [27] by averaging the performance of the ensemble of random codes. This derivation can also be used as a guideline to analyze the error performance of coded communication systems.

2.1.2 The Optimum Decoder

The coding theorem is proved with the assumption that the optimum decoder, i.e., the optimum algorithm to detect the codewords, is used. The optimum decoder is based on the maximum *a posteriori* probability (MAP) criterion to minimize the error probability of each bit (bit-MAP criterion) in the information massage. However, the decoder works with codewords, and it is often based on the codeword maximum *a posteriori* criterion (codeword-MAP criterion) whose performance is very close to the performance of the bit-MAP criterion.

Consider the binary block code $\mathcal{C}(N, K)$ as described in (2.1). Suppose that the codeword $\mathbf{x}_i$ is transmitted over a given channel and the received word is $\mathbf{y}$. The codeword-MAP decoder chooses $\mathbf{x}_i^{MAP}$ to maximize the following *a posteriori* probability:

$$\mathbf{x}_i^{MAP} = \arg\max_{\mathbf{x}\in\mathcal{C}} \Pr(\mathbf{x} = \mathbf{x}_i|\mathbf{y}) \tag{2.4}$$

The *a posteriori* probability $\Pr(\mathbf{x} = \mathbf{x}_i|\mathbf{y})$ is related to the probability density function $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)$ by Bayes' rule as [28]:

$$\Pr(\mathbf{x} = \mathbf{x}_i|\mathbf{y}) = \frac{\Pr(\mathbf{x} = \mathbf{x}_i)p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)}{p(\mathbf{y})} \tag{2.5}$$

The specific form of the density function $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)$ depends on the channel model. For example, $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)$ is a multi-dimensional Gaussian distribution if the channel is AWGN.

On the other hand, the maximum likelihood (ML) decoder identifies the codeword $\mathbf{x}_i$ that maximizes $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)$, i.e.,

$$\mathbf{x}_i^{ML} = \arg\max_{\mathbf{x}\in\mathcal{C}} p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i) \tag{2.6}$$

It is obvious from (2.4), (2.5) and (2.6) that if all the codewords are equally likely, i.e., $\Pr(\mathbf{x} = \mathbf{x}_i) = \frac{1}{2^K}$, then ML decoder and codeword-MAP decoder are equivalent. The above discussion of the optimum decoders is also proper for communication schemes in which $\mathbf{x}$ is a sequence of symbols instead of bits.

If the structure of a given code is not considered or not available, one needs to compute $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)$ for every codeword in order to implement (2.6). Note that, the number of codewords, 2^K, is very large, even when the number of information bits, K, is relatively small. Therefore, the computational complexity of the ML decoding is very high and impractical for most practical codes.

Suboptimum decoders use different algorithms to approximate the optimum decoders. These algorithms have lower computational complexity than that of the optimum decoders. Therefore, the suboptimum decoders are widely used in practice due to their economical advantage.

2.1.3 Practical Codes and Decoding Complexity

In the previous subsections, the coding theorem and the optimum decoders are discussed for random codes whose codewords are randomly chosen from the code space. However, the computational complexity of the optimum decoders is extremely high for random codes due to non-structured property of these codes. In practice, the codes are constructed by certain rules. The decoding algorithms use information from the construction rule to reduce the necessary computation to detect the codewords. Practical codes are often named by their construction rules.

From the viewpoint of code construction, the principle of coding is the addition of redundant bits. Hence, the most important parameter of a code is the code rate because it determines the amount of redundancy in the code. From the economical aspect, the code rate should be as high as possible. However, the performance of a code is reduced when the code rate increases. A trade-off between the code rate and performance needs to be decided according to the requirements of a specific application. Naturally, the most simple code, is a repetition code in which the information bits are repeated exactly. This code is good in terms of complexity, but it is bad in terms of performance because it only achieves arbitrarily small bit error probability when the rate reduces to zero by increasing the length of the codeword. Coding is time diversity, so that the longer the length of a codeword is, the better the error performance becomes. Here, the information of a bit is diverted in the time domain to every coded bit so that the information can be hidden from the attack of the random noise.

Although a code is always expected to have good performance, the decoding complexity is also an important criterion in practice, especially for long codes. For example, repetition codes are bad codes in terms of performance, but they are simple in both encoding and decoding. This is why these codes are sometimes used in practice. Random codes are good codes as Shannon proved, however, these codes are impractical because they can only be decoded by the ML decoder that compares the received sequence to every codeword. This decoder requires a computational complexity that is exponential in block length, i.e., the cost of the system is very high.

The most popular practical codes are algebraic block codes and convolutional codes. These codes can be seen in most recent communication and storage systems. However, it is very likely that the powerful pseudo-random codes will be used in the future applications [29].

Algebraic codes include linear block codes such as cyclic block codes and BCH codes. The decoding of these codes is often based on hard-decisions on the received words. They were appropriate for the era when powerful computational circuits were very expensive. The best known algebraic codes are Reed-Solomon codes, which are BCH codes based on Galois field 2^q. Recently, some soft decoding algorithms have been developed to decode Reed-Solomon codes [30–32]. However, the complexity of these algorithms is still high when the length of the codes increases.

An important subclass of block codes are the linear block codes. Since many concepts of linear block codes are used for LDPC codes, an example of a simple linear block code, the (7,4) Hamming code, is introduced next in order to illustrate these concepts.

A binary block code is linear if and only if the modulo-2 sum of two codewords is also a codeword. The codewords of the linear block code $\mathcal{C}(N, K)$ are generated by a *generator matrix* $\mathbf{G}$, where the rows of $\mathbf{G}$ are linearly independent:

$$\mathbf{x} = \mathbf{u} \cdot \mathbf{G} \tag{2.7}$$

The first practical error correcting codes were Hamming codes, invented by Hamming [33]. The generator matrix of the simplest Hamming code, the (7,4) Hamming code, is as follows:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

Using the above generator matrix, all 16 codewords in the (7,4) Hamming code are listed in Table 2.1 together with the corresponding information vectors.

Each linear block code has a *parity check matrix* $\mathbf{H}$ that is often used in the decoder.

Table 2.1 The (7,4) Hamming code.

Information $(\mathbf{u})$				Codeword $(\mathbf{x} = \mathbf{u} \cdot \mathbf{G})$						
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	1	0	0	0
0	1	0	0	0	1	1	0	1	0	0
1	1	0	0	1	0	1	1	1	0	0
0	0	1	0	1	1	1	0	0	1	0
1	0	1	0	0	0	1	1	0	1	0
0	1	1	0	1	0	0	0	1	1	0
1	1	1	0	0	1	0	1	1	1	0
0	0	0	1	1	0	1	0	0	0	1
1	0	0	1	0	1	1	1	0	0	1
0	1	0	1	1	1	0	0	1	0	1
1	1	0	1	0	0	0	1	1	0	1
0	0	1	1	0	1	0	0	0	1	1
1	0	1	1	1	0	0	1	0	1	1
0	1	1	1	0	0	1	0	1	1	1
1	1	1	1	1	1	1	1	1	1	1

The valid codewords satisfy:

$$\mathbf{x} \cdot \mathbf{H}^T = \mathbf{0} \tag{2.9}$$

The parity check matrix $\mathbf{H}$ is related to the generator matrix $\mathbf{G}$ by $\mathbf{G} \cdot \mathbf{H}^T = \mathbf{0}$. The parity check matrix of the above (7,4) Hamming code is as follows:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \tag{2.10}$$

An important parameter of block codes is the minimum Hamming distance among the codewords. For example, the minimum Hamming distance of the above (7,4) Hamming code is 3, which can be easily verified from Table 2.1. Note also that for a linear block code, the minimum Hamming distance is also the minimum Hamming weight among all codewords. Coding researchers try to design algebraic codes with a large minimum Hamming distance in order to improve the error performance of the codes under the ML decoder. However, some recent results of coding theory prove that the minimum Hamming distance is not a sufficient criterion for very long codes [29, 34, 35].

Convolutional codes are constructed based on the approach of linear systems. The transmitted bit stream is a convolution of the information bit stream with a linear filter in Galois field 2. The impulse response of this filter (also known as the shift register) can be finite or infinite, depending on whether there is a feedback connection. If there is feedback, the convolutional codes are called recursive convolutional codes. Figure 2.2 shows the diagram of a feedforward convolutional encoder of rate $1/2$ for the illustration purpose.

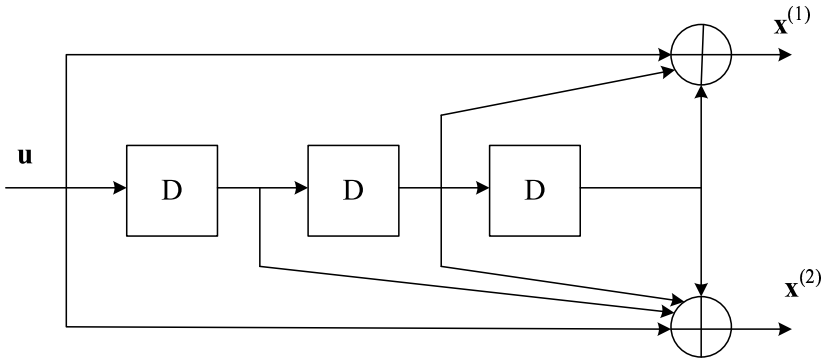


Figure 2.2 The diagram of a nonsystematic convolutional encoder with rate 1/2 and memory length 3.

Convolutional codes are often decoded by the Viterbi algorithm, which is the ML decoding algorithm implemented on a trellis diagram. Unlike block codes, a convolutional code with a given generator polynomial has a fixed minimum Hamming distance even when the length of codewords increases. Performance of the convolutional codes can only approach the Shannon limit when the memory length of the shift register increases [35]. However, the computational complexity per bit of the Viterbi decoder increases exponentially with the memory length.

A solution to increase the length of codes and still keep the low computational complexity is concatenation [36]. Before the invention of Turbo codes, the best known code is a concatenated code used for the Galileo spacecraft [29, 34]. In this code, a convolutional code of memory length 15 is used as the inner code and a Reed-Solomon code is used as the outer code. Decoding of this code requires a room full of special hardware. In the future systems, the above coding scheme shall be replaced by Turbo codes or Turbo-like codes.

2.1.4 Turbo Codes and Turbo-like Codes

The invention of Turbo codes triggered a revolution in channel coding techniques as well as digital communications. The idea of Turbo codes is now applied for many iterative communication systems. Thus, Turbo codes are described and their principles applying to Turbo-like codes are introduced in this subsection.

Turbo codes are parallel concatenated codes. The new ideas in Turbo codes are (i) using a random interleaver between the two component convolutional codes in the encoder and (ii) applying an iterative soft decoding algorithm in the decoder. With this code construction and decoding structure, Turbo codes can be decoded with linear complexity per bit and achieve a near Shannon limit performance [4].

In the decoder of original Turbo codes, the BCJR algorithm (also known as the backward-forward algorithm or symbol-by-symbol MAP algorithm) [37] is used for each convolutional code to calculate the soft *a posteriori* probability of each bit. The soft output of one component convolutional decoder, often called the soft-input soft-output (SISO) decoder, is fedback to the other component decoder. Since the complexity of the BCJR algorithm is twice the complexity of the Viterbi algorithm, it was rarely used in the convolutional decoders previously. As illustrations, Figures 2.3 and 2.4 show the block diagrams of the encoder and the decoder of the original Turbo code, respectively [4].

Due to the outstanding performance of Turbo codes, a class of Turbo-like codes has been discovered or rediscovered. They include low density parity check codes, repeat accumulated codes [38], Turbo product codes [39]. These codes have the following common properties:

- Performance is near Shannon limit.
- Decoding complexity is linear with the code length.
- Soft-decision decoding is used.
- Suboptimum, but high performance iterative decoding is used.

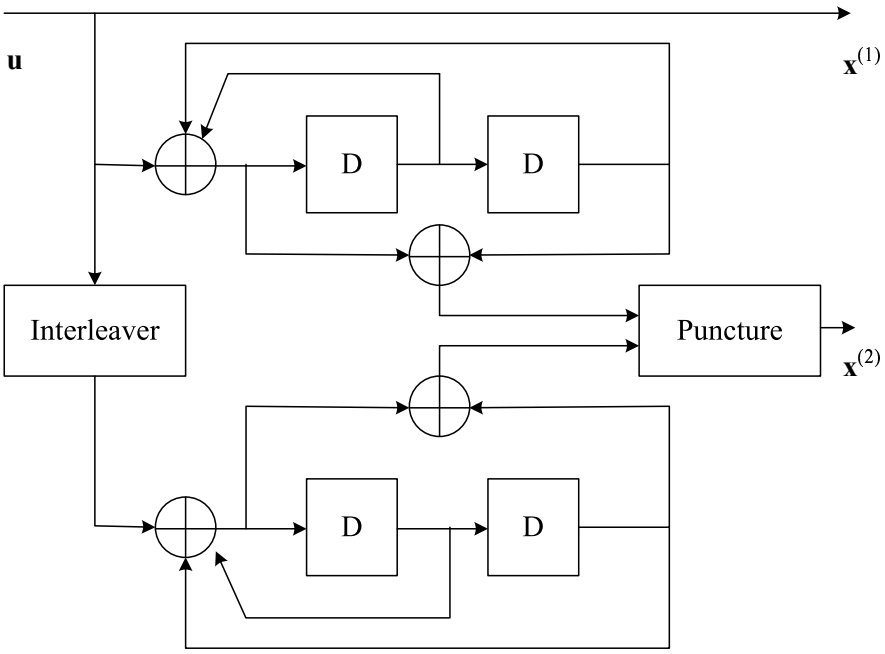


Figure 2.3 Block diagram of the original Turbo encoder.

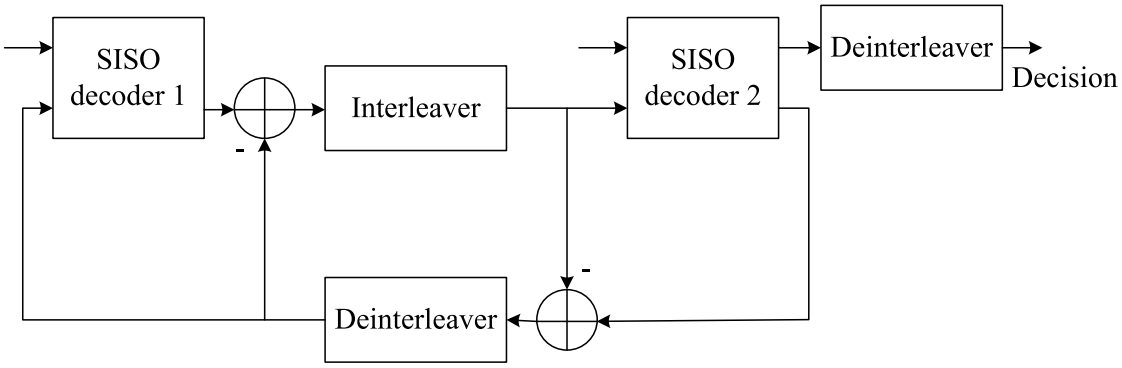


Figure 2.4 The basic structure of the Turbo decoder.

- Can be represented on a graph.

The above codes can be very long in order to approach the Shannon limit, thanks to the constant decoding complexity per bit of the suboptimum iterative decoder. Since the code-words are scattered in a huge space, the number of errors occurring due to the noise effect, which exceed the borders of the decision regions, is relatively smaller than the number of errors due to the suboptimum decoder.

By the last property, Turbo-like codes are sometime called codes on graph. Among

these codes, low density parity check codes are the most flexible in construction and they can be easily optimized to achieve good performance in various communication channels.

2.2 Low Density Parity Check (LDPC) Codes

2.2.1 Structure of LDPC Codes

Low density parity check codes (LDPC) are linear block codes with sparse parity check matrices. The original Galager's LDPC codes are called regular LDPC codes in which the number of 1's is the same in every row and every column [5]. The following matrix is the parity check matrix of a rate-1/2, length-10 regular LDPC code:

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \tag{2.11}$$

In this matrix, the number of 1's in each row is 6 and the number of 1's in each column is 3. This parity check matrix is not very spare because the code is still short. This code can be presented by a bipartite graph as in Figure 2.5. In this graph, each left node, called a variable node, represents a bit of the codeword. Each right node, called a check node, represents a parity check bit. The number of variable nodes corresponds to the number of columns in the parity check matrix **H**, while the number of check nodes corresponds to the number of rows in **H**. Edges connect the variable nodes to the check nodes according to the parity check matrix **H**.

If the number of edges emanating from a variable node is called variable node degree d_v and the number of edges emanating from a check node is called check node degree d_c, then the rate of the (d_v,d_c) regular LDPC code is $R_c = 1 - \frac{d_v}{d_c}$. The number of 1's in the parity check matrix **H** is $N \cdot d_v$, while the total number of elements in **H** is $N^2 \cdot R_c$, where N is the length of the code. Obviously, when N increases, the number of 1's increases linearly and the total number of elements increases quadratically. Hence, the parity check matrix is sparse with large N.

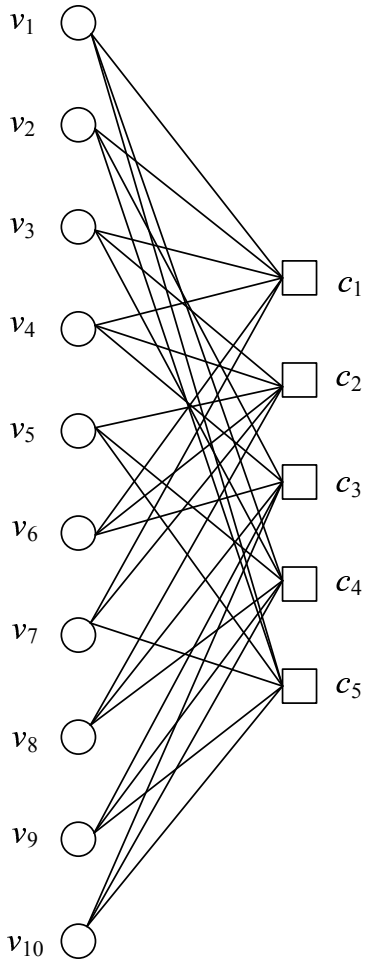


Figure 2.5 The bipartite graph of a regular (3,6) LDPC code of length 10, rate 1/2.

The sparse characteristic of the parity check matrix is important, because the number of 1's presents the number of relations between a variable node and a check node. Since the decoder uses these relations to decode, this quantity determines the complexity of the decoder.

Irregular LDPC codes are LDPC codes that have nodes with different degrees. The degrees of the variable nodes and the check nodes are chosen according to some distribution. For compact description, the degree distribution is often presented in polynomial form. The variable node degree distribution is denoted by $\lambda(x)$ and it can be expressed as:

$$\lambda(x) = \sum_{i=2}^{d_v} \lambda_i x^{i-1} \tag{2.12}$$

where, λ_i is the fraction of edges emanating from variable nodes of degree i and d_v is

the maximum variable degree of the irregular LDPC code. Note that the coefficient λ_i is associated with x^{i-1}, rather than x^i. Similarly, the check node degree distribution is denoted by $\rho(x)$ and can be expressed as:

$$\rho(x) = \sum_{i=2}^{d_c} \rho_i x^{i-1} \tag{2.13}$$

where ρ_i is the fraction of edges emanating from check nodes of degree i and d_c is the maximum check degree. For example, the degree distributions of the previous (3,6) regular LDPC code are $\lambda(x) = x^2$ and $\rho(x) = x^5$. With this presentation, the number of variable nodes of degree i of the (λ, ρ) irregular LDPC code of length N is

$$N \frac{\lambda_i / i}{\sum_{j \geq 2} \lambda / j} = N \frac{\lambda_i / i}{\int_0^1 \lambda(x) dx}. \tag{2.14}$$

The total number of edges emanating from all variable nodes is

$$E = N \sum_{i \geq 2} \frac{\lambda_i}{\int_0^1 \lambda(x) dx} = N \frac{1}{\int_0^1 \lambda(x) dx}. \tag{2.15}$$

The quantity E can also be expressed in terms of the total number of check nodes M as:

$$E = M \frac{1}{\int_0^1 \rho(x) dx}. \tag{2.16}$$

The relation between M and N is

$$M = N \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx}. \tag{2.17}$$

Assuming that all these check equations are linear independent, the design rate is equal to

$$r(\lambda, \rho) = \frac{N - M}{N} = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx}. \tag{2.18}$$

For a given length and a given degree distribution, there is a set of codes, called an *ensemble* of codes. The concentration theorem in [8, 40] proves that the performance of all codes in an ensemble with very long length is close to an average performance. Therefore, the design of LDPC codes that can approach Shannon limit is equivalent to finding a degree distribution of the ensemble.

2.2.2 Sum-Product or Belief Propagation Decoder

The sum-product decoding is also known as belief propagation decoding. The term "belief propagation" was coined by researchers in the artificial intelligence community when they studied Bayes networks where many random events are correlated by a network topology. The problem is to find the probability of an event when other events are known with given probabilities. This problem is called "inference". It is solved by calculating a certain marginal probability from the joint probability density function of all random variables corresponding to the events. For small Bayes networks, the marginalization can be easily done by a summation over states of other random variables. However, when the number of random variables is large, this solution is impossible because the number of terms in the summation grows exponentially with the number of variables. The principle of belief propagation is to transmit "belief" according to edges of the Bayes network. This way, belief propagation can compute the marginal probabilities, at least approximately, with a complexity that grows only linearly with the number of random variables in the system. Here, the artificial intelligence community prefers the term "belief" to the term "probability" because they define probability as the "degree of belief" [29, 41].

When researchers in the communications community study error control coding and digital communications they met a similar problem. They recognized that the problems of communications are solved by a variety of algorithms that have the same principles as belief propagation. They call the algorithms the sum-product algorithms which includes the forward/backward (BCJR) algorithm, the Viterbi algorithm, the iterative "turbo" decoding algorithm, the Kalman filter, and certain fast Fourier transform (FFT) algorithms [42].

The belief propagation or sum-product algorithm can be applied to decode LDPC codes. In the context of decoding LDPC codes, the sum-product algorithm is best understood as an iterative message passing algorithm on the bipartite graph. At each half of an iteration, the *extrinsic* information or the outgoing messages from a node are calculated based on the previous messages from the nodes connected to this node and passed along the edges to the other side of the bipartite graph. Here, the *extrinsic* information for the other side of the bipartite graph is understood as the additional part of information that is created from

outside of the incoming messages. The complexity of this algorithm is a function of the number of edges, so it is linear with the length of the codewords.

For purposes of this thesis, the messages are log a *posteriori* probability ratios (LAPPR) of the variable nodes passed in both directions: from the variable nodes to the check nodes and from the check nodes to the variable nodes. If x_i is the ith bit in a codeword corresponding to variable node v_i and y_i is the channel output of this bit, then the LAPPR of this bit or a soft input of the decoder is

$$\log \frac{\Pr(x_i = 0|y_i)}{\Pr(x_i = 1|y_i)}. \tag{2.19}$$

For example, when a coded bit is modulated by binary phase shift keying (BPSK) and transmitted over an AWGN channel with power spectral density of σ^2, the channel output of the ith bit is:

$$y_i = (2x_i - 1) + w_i \tag{2.20}$$

where w_i is zero-mean Gaussian random variable with variance σ^2. The conditional density distribution of the channel output is

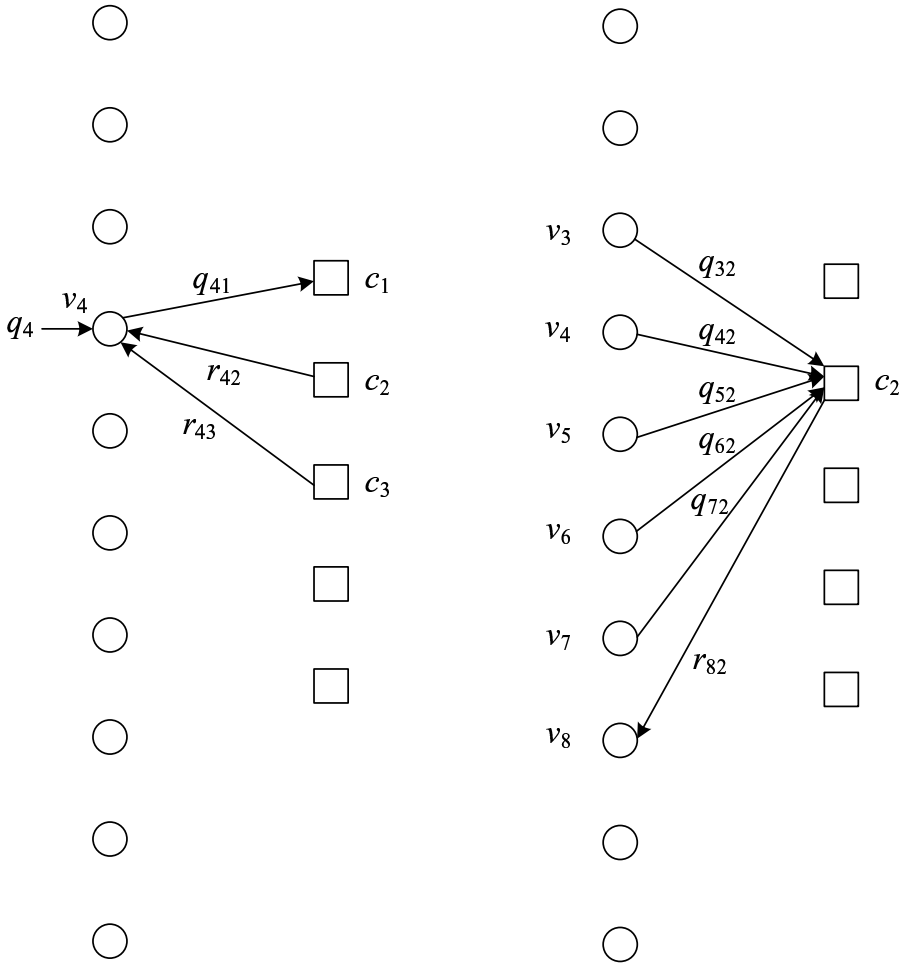
$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - 2x_i + 1)^2\right] \tag{2.21}$$

If the 0s and 1s in codewords are equally likely, then

$$\log \frac{\Pr(x_i = 0|y_i)}{\Pr(x_i = 1|y_i)} = \log \frac{p(y_i|x_i = 0)}{p(y_i|x_i = 1)} = -\frac{2}{\sigma^2} \cdot y_i. \tag{2.22}$$

The decoding procedure is based on the soft inputs and proceeds as follows. At a variable node v_i, the outgoing message q_{ij} to a given check node c_j (that connects to the variable node v_i in the bipartite graph) is a summation of the initial LAPPR q_i and the incoming messages $r_{ij'}$ to the variable node v_i, except the incoming message r_{ij} from the check node c_j. The flow of an outgoing message at a variable node and incoming messages that are used to calculate this outgoing message is illustrated in Figure 2.6-(a). At a check node c_j, the outgoing message r_{ij} to a given variable node v_i is the LAPPR of variable node v_i that is calculated from the incoming messages $q_{i'j}$ to this check node except the incoming message q_{ij} from the variable node v_i. Here, the calculation is to find the LAPPR

of a binary random variable when the LAPPRs of other binary variables are known and the summation of all binary variables is zero which is known by the parity check function. The flow of an outgoing message at a check node and incoming messages, that are used to calculate this outgoing message, is illustrated in Figure 2.6-(b).



(a) Variable node processing (b) Check node processing

Figure 2.6 The message processing at variable nodes and check nodes.

Mathematically, the steps of the belief propagation algorithm for decoding LDPC codes can be summarized as follows [5, 6, 14]:

- Initiation: For each variable node v_i $(i = 1, ..., n)$, the associated decoder input is set to the initial LAPPR, $q_i = \log \frac{\Pr(x_i=0|y_i)}{\Pr(x_i=1|y_i)}$, where x_i is the value of the ith bit in the

codeword corresponding to the variable node v_i, y_i is channel output of the ith bit. The messages from the check nodes are set to zero in the initiation.

- Variable node processing: For each variable node v_i, the outgoing message to the check mode c_j, that is connected to v_i, is given by:

$$q_{ij} = q_i + \sum_{j' \in J(i) \backslash j} r_{ij'} \tag{2.23}$$

where the set $J(i)\backslash j$ contains those check nodes connected to the variable node v_i except the check node c_j. Here, the content of a message is the soft extrinsic probability of the ith bit in the log domain, hence the summation of the log probabilities is implemented instead of a multiplication.

- Check node processing: For each check node c_j, the outgoing message to the incidental variable node v_i is given by:

$$r_{ij} = \log \frac{1 + \prod_{i' \in I(j) \backslash i} \tanh(q_{i'j}/2)}{1 - \prod_{i' \in I(j) \backslash i} \tanh(q_{i'j}/2)} \tag{2.24}$$

where, similarly, the set $I(j)\backslash i$ contains those variable nodes connected to the check node c_j except the variable node v_i. The above equation is equivalent to the procedure that calculates the *a posterior* probabilities as in [43]. The derivation of (2.24) is presented in Appendix A. A simple example, that helps to understand (2.24), is also provided in Appendix B.

- Stopping and decision: After each iteration, the *a posteriori* probability of the bit x_i is computed as

$$p_i = q_i + \sum_{j \in J(i)} r_{ij}. \tag{2.25}$$

Then each bit is estimated as

$$p_i \gtrless_{\hat{x}_i=0}^{\hat{x}_i=1} 0 \tag{2.26}$$

to form $\hat{\mathbf{x}} = [\hat{x}_1, ..., \hat{x}_n]$. If $\mathbf{H}\hat{\mathbf{x}} = \mathbf{0}$, then declare the decoding a success and claim $\hat{\mathbf{x}}$ is the decoded codeword. Otherwise, iterations are continued by the two previous steps. If the number of iterations exceeds a pre-specified limit without finding a valid codeword, declare the decoding failure.

The procedure to process the detected errors depends on specific applications. If the feedback link is available, retransmission can be done for non real-time applications. If retransmisison is impossible, the LDPC codes can be concatenated with an outer erasure-correcting code [44,45]. Otherwise, the information bits are still obtained from the systematic bits for systematic LDPC codes or by a Gaussian elimination for nonsystematic LDPC codes.

Thus, according to the above procedure, the outgoing message to a given variable node is the extrinsic LAPPR of this variable node. This is because it does not include the information that have passed though this variable node in the case of a cycle-free graph. After each iteration, the messages are more accurate because they are implicitly updated with information from more nodes. The structure of nodes, that sends information to a given variable node after a given number of iterations, shapes a tree as illustrated in Figure 2.7.

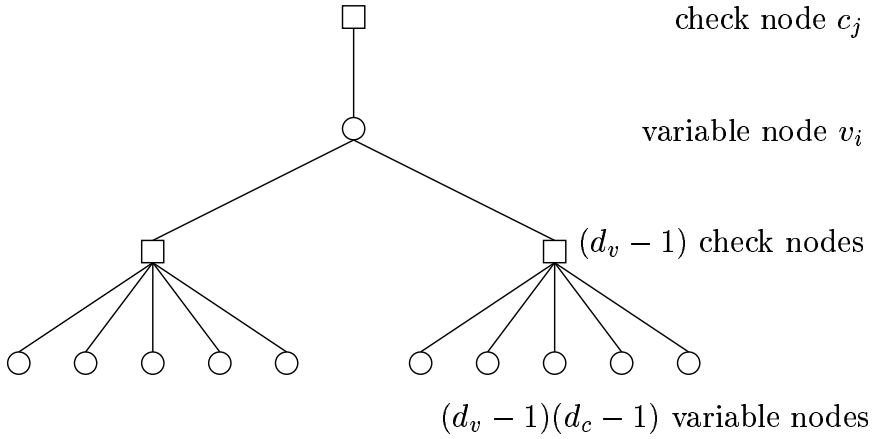


Figure 2.7 The support tree of depth 2 for a regular (d_v, d_c) LDPC code.

The sum-product algorithm can produce the exact *a posterior* probabilities of all bits if the bipartite graph defined by matrix **H** contains no cycles whose girth is longer than double of the number of iterations [6,46]. However, when the code length is small and the number of iteration is large, this may not be secured. Therefore, this sum-product algorithm is only suboptimal in this case. Coding designers often try to omit the short cycles in the parity matrix **H**. This is important in designing short LDPC codes [47].

Some suboptimum message passing algorithms of belief propagation were investigated

in [8]. These algorithms use discrete levels of the LAPPR instead of the real variables, hence the computational complexity is reduced. Of course, the error performance is degraded moderately. Therefore, the trade-off should be made when one wants to design a practical decoder.

2.2.3 Distance Property of LDPC Codes and Convergence of Sum-Product Decoder

Performance of all codes as well as LDPC codes under ML decoding depends on the Hamming distances among the codewords, i.e., the distance property of the code. When a codeword is sent over a noisy channel, the received signal can be a point outside the decision region of the original codeword. It means that the ML decoder makes an incorrect decision in this case and an error occurs. Thus, the probability of error depends on the size of the decision regions. The ML criterion is equivalent to the minimum distance criterion. It means that the ML receiver makes decision to the codeword that is closest to the received signal. Therefore, the distances among the codewords decide the performance of the ML decoder.

The distance property of Gallager's regular LDPC codes is analyzed in [5]. Figure 2.8 shows the bounds on the Hamming distance spectra of (3,6), (4,8), and (5,10) regular LDPC ensembles of length 1000. The bounds are compared with the Hamming distance spectrum of a random code which is a binomial distribution. The analysis shows that LDPC codes have good distance property, i.e., the minimum Hamming distance increases and the distance spectra of LDPC codes are close to those of the random codes when the code-length increases. Other analyzes of distance properties for irregular LDPC ensembles and other regular ensembles are introduced in [48, 49], which provides the bounds on the Hamming distance spectra of very long codes. For short and medium-length LDPC codes, searching algorithms can find the first distance terms of the codes more accurately [50, 51].

For finite-length codes, the pairs of codewords with the small Hamming distances are most easily confused by the noisy channel. Hence, the minimum Hamming distance and some first terms of the Hamming distance spectra play a major role in determining the error probability of the ML decoder. For long LDPC codes, almost all terms of the Hamming

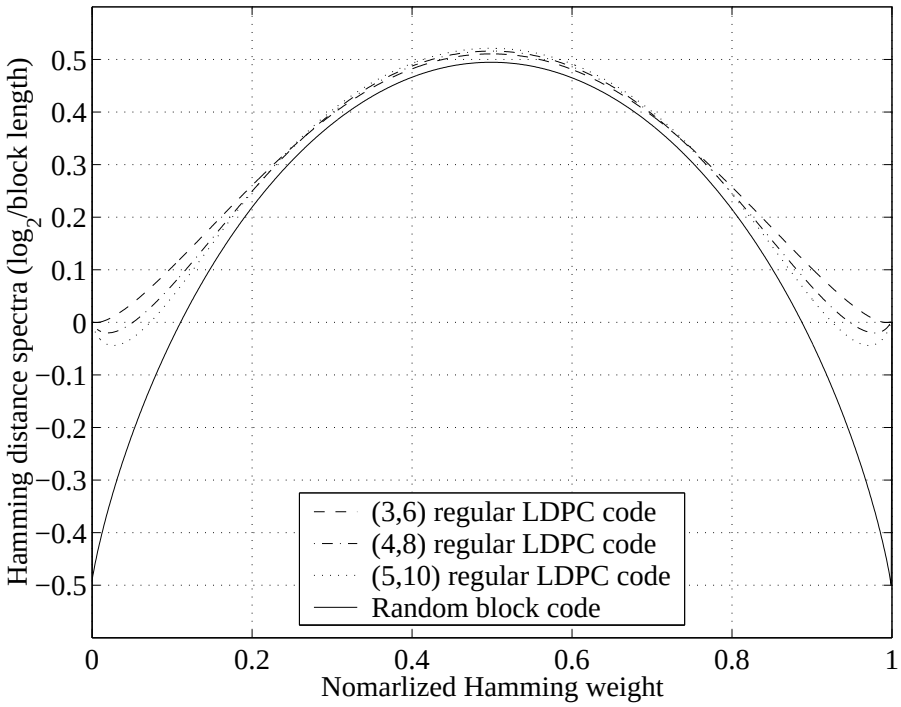


Figure 2.8 The normalized Hamming distance spectra of regular LDPC codes and a random code of length 1000 and rate 1/2.

distances contribute to the error performance of the ML decoder. Evaluation of the ML performance for long LDPC codes with binary phase shift keying (BPSK) modulation over AWGN channels is investigated by tangential sphere bounds and Gallager's bounds in [5, 52–54].

As discussed in the previous section, the LDPC codes are decoded by the sum-product decoder in practice. Therefore, the performance of these codes under the sum-product decoding needs to be evaluated. To understand how the performance of LDPC codes converges under the sum-product decoding algorithm, one needs to look at the decoding mechanism. After half of an iteration, a bit or a variable node is updated by the extrinsic probabilities from some check nodes connected to it. These check nodes get information messages from the other variable nodes from the previous iteration. Thus, after some iterations, many check nodes and variable nodes, that sent information to a given variable node, construct a tree structure, called a support tree, as shown in Figure 2.7.

If the graph is cycle-free after a given number of iterations, all nodes in the support

tree are different. Therefore, the random variables corresponding to nodes in a row of the tree can be considered as independent random variables. Thus, from the initial probability density function of the variables, the probability density functions of the variables can be obtained from this tree structure.

Assuming the support tree is cycle-free and the initial probability density functions of the variable nodes are symmetric and identical, the authors in [8] investigate the evolution of the densities due to iterations on the support tree and they call it density evolution. With the assumption of a symmetric channel, they observe the convergence of the log density of the all-zero codeword at a given noise level and state that this convergence also occurs with other codewords. Due to computational complexity, the density evolution is determined by computer.

For an AWGN channel, a given threshold of the noise level for a certain code is obtained by the density evolution. If the noise level is above this threshold, the density evolution does not converge and the sum-product algorithm cannot decode. Based on the concepts of the threshold and density evolution, the performance of very long LDPC codes is evaluated over various symmetric channels [10, 14]. By combining the density evolution technique with an optimization technique, called differential evolution, irregular LDPC code ensembles with performance very near Shannon limit threshold are found in [10, 11, 55].

It should also be mentioned here that there is an error floor observed at very low bit error rates for some specific LDPC codes, especially the algebraically constructed LDPC codes, under the sum-product decoding [56,57]. This phenomenon is due to low-weight detected errors, which are also called the "pseudo-codewords" on the graph of the code [56]. For these codes, the sum-product decoding is occasionally trapped by these neighboring pseudo-codewords and cannot converge to the true codewords. Note that, the neighboring pseudo-codewords on the graph are not the codewords with minimum Hamming weight and this floor does not relate to the maximum likelihood performance.

When the probability density functions (PDF) of received coded bits are not symmetric or even i.i.d., which often occurs in communication systems, the density evolution tech-

nique is impossible to determine due to the complexity. The density evolution progressing in a support tree is too complicated because the PDFs of the random variables are not the same (non i.i.d) and a general structure corresponding to different random variables of a support tree is different for each bit.

Observing the change of one parameter instead of the entire density function after iterations can reduce the computational complexity. The parameter can be the mean, variance, error probability or a mutual information. Among these parameters, mutual information seems to be the most accurate statistic [58]. Based on mutual information, a technique called the extrinsic information transfer (EXIT) chart [7, 59], proves to be efficient in the analysis by tracking an average mutual information.

Since mutual information is the most robust statistic, in the sense that it can be applied without change to a widest range of channels, modulations, and detectors, the EXIT chart technique can be applied not only to analyze LDPC codes but also to analyze most of the iterative coding systems such as serial concatenated codes, parallel concatenated codes, or iterative decoding and detection.

Unfortunately, the density evolution and EXIT chart techniques only correctly predict the convergence of the sum-product algorithm if the length of the LDPC code is assumed to be infinite. Analyzing the convergence of finite-length codes is still an open problem. Since this thesis focuses on finite-length LDPC codes, these techniques are not used to analyze the systems. The convergence of the sum-product decoder to the ML decoder is only observed and discussed by simulation results.

3. LDPC Coded Modulation

This chapter studies a bandwidth-efficient coded modulation system. The system includes an LDPC encoder and a constellation in a 2-dimensional space, where a signal point in the constellation is represented by the in-phase and quadrature sinusoids. Since LDPC codes are randomly constructed, this scheme is similar to the bit-interleaved coded modulation (BICM) originally proposed for convolutional codes [16,17]. The error performance of the system with different receivers is investigated in this chapter. The performance bound of the ML decoding is derived based on the Hamming distance spectrum of the LDPC code and the Euclidean distance profile of the constellation. This ML bound is then compared with the performance of the systems based on the sum-product decoding and the ordered-statistics decoding (OSD). Various parameters of the systems such as the code rate, the constellation and mapping are investigated for both AWGN and Rayleigh fading channels.

3.1 Introduction

Combination of channel coding and modulation into one entity achieves a remarkable coding gain with the same spectral efficiency compared to the scheme that treats channel coding and modulation separately. Trellis coded modulation (TCM), proposed by Ungerboeck [60, 61], is the first bandwidth-efficient coded modulation system. The main principle of TCM is based on mapping by set partitioning. More specifically, the partitions of the modulating constellation in TCM are assigned accordingly to the trellis diagram of a convolutional code in order to increase the minimum Euclidean distance of the code sequences. Various methods were developed to improve the performance of TCM, which includes multiple TCM [62] and higher-dimension TCM [63]. However, the time diversity in these methods is still limited by the constraint length of the convolutional codes. The

Turbo-TCM [64] enlarges the time diversity by using a symbol interleaver between two concatenated TCM components. Moreover, the iterative technique is used at the receiver to improve the error performance. Nevertheless, TCMs are used only with convolutional codes, and the technique cannot be applied to block codes such as LDPC codes.

Another coded modulation scheme, called multilevel coded modulation (MLC), was proposed simultaneously with TCM in [19]. This scheme uses several encoders at the transmitter and also several decoders at the receiver. This MLC scheme is capable of employing block codes such as LDPC codes. However, because the structure of MLC scheme is more complicated due to the use of many component codes, multilevel LDPC coded modulation system shall be studied in the Chapter 4 after the basic LDPC code modulation is investigated in this chapter.

More recently, another combined coding and modulation scheme, called bit interleaved coded modulation (BICM), has been proposed in [16] and investigated in detail in [17]. In these papers, the BICM scheme uses only one convolutional component code. The important feature is that a bit-wise interleaver is employed between the modulation block and the convolutional encoder. The results in [16] demonstrates the performance advantage of BICM over TCM over a Rayleigh fading channel. However, the iteration between demodulation and decoding is not considered in the receiver of [16]. Therefore, Gray mapping is still considered to be the best mapping for BICM with convolutional codes in [16, 17]. Recently, BICM schemes with iterative processing between demodulation and decoding, called BICM-ID, were studied in [65–68]. The results in these papers show that the error performance with Gray mapping only has the best performance after the first iteration. After a few iterations, the performance with other mappings is superior. In [66], the authors state that the best mapping depends on the number of iterations, the region of signal-to-noise ratio and the size of the interleaver. In [67], with a careful design of the interleaver and investigating the bit distance of 8-PSK constellation, the authors show that the semi set-partitioning (SSP) mapping performs better than both the set-partitioning and Gray mappings. The BICM-ID with QAM constellation was also investigated in [68]. The results in the paper show that a random mapping is the best among the number of mappings

investigated. More recently, the work in [69, 70] presents multi-dimensional mappings for BICM-ID and shows that a significant performance improvement is obtained with multi-dimensional mappings. Such performance advantage, however, comes at the price of a higher-complexity receiver.

BICM using LDPC codes was also investigated in [14, 43]. Although the scheme is called bit-interleaved coded modulation with LDPC codes, no interleaver is needed between the encoder and the modulator. This is due to the random construction of LDPC codes. In other words, there is a built-in interleaver in the parity check matrix of LDPC codes. Therefore, in this thesis, the bit-interleaved LDPC coded modulation is often called the LDPC coded modulation. In [14, 43], only BICM schemes with Gray mappings of signal constellations are considered. With Gray mapping, the inputs of the decoder are identical and independently distributed (i.i.d.) random variables. Therefore, the irregular LDPC codes can be optimized by the density evolution technique. BICM-ID scheme with LDPC codes and Gray mapping is considered in [43]. Different from the case of BICM with convolutional codes, the result in [43] show that the gain provided by the iterations between demodulation and decoding is negligible. It is also shown in [43] that BICM outperforms Turbo-TCM for long codeword lengths. The coded modulation system based on LDPC codes and BPSK modulation is investigated for an uncorrelated frequency-flat fading channel in [14, 71], where LDPC codes are optimized for this particular channel to achieve a better error performance.

The above performance results of LDPC coded modulation are all studied with infinite-length LDPC codes. Performance of such infinite-length LDPC coded modulation systems are usually analyzed by the convergence threshold of the sum-product (SP) decoding over different channels and with different modulation schemes.

However, very long LDPC codes can not be used in low-latency applications such as speech transmission (for example, the information blocks of 40 to 5114 bits are recommended in 3G wireless communications [72]). For finite-length LDPC codes, the performance of the sum-product decoding does not strictly converge around one threshold value of the signal-to-noise ratio. It means that the water-fall phenomenon on the error perfor-

mance can not be observed for finite-length LDPC coded communication systems. At high SNR and low BER, the performance of the iterative sum-product decoding can approach the performance of maximum likelihood (ML) decoding quite closely. Although the complexity of ML decoding is prohibitive even for short-length codes, the performance of ML decoding can still be evaluated by bounding techniques [54]. Furthermore, the performance of the systems under ML decoding shows the ultimate capability of the systems and it serves as a lower bound of the performance of sum-product decoding. For a given range of code-length, the ML bound can be used to estimate the performance of sum-product decoding.

In this chapter, the upper bounds on the bit error rate (BER) performance of LDPC coded modulation under ML decoding are studied for both AWGN and Rayleigh fading channels. To derive the bounds, the Euclidean distance spectrum of the sequences of symbols needs to be computed. Here, the framework of [73], originally proposed to compute the Euclidean distance spectrum of turbo coded systems, is applied for the LDPC coded modulation systems under consideration. The tightness of the bounds is verified by simulating ML decoding of a very short code.

Another decoding technique, called ordered-statistics decoding (OSD) [18], is also considered in this chapter. The OSD algorithm is capable of decoding linear codes such as LDPC codes. The performance of OSD is close to the performance of ML decoding for the systems using short codes and BPSK modulation. The disadvantage of the OSD algorithm is that its computational complexity is not linear in the code length . For LDPC coded modulation, the log likelihood probability ratios (LLR) are used to compute the reliability values of the coded bits and to find the closest codeword instead of the received signals as used in [18]. For the iterations between demodulation and decoding, the SISO-OSD module has to be carefully devised to produce the soft-output LLR [74].

3.2 System Description

Figure 3.1 shows the coded modulation system under consideration. In this system, an information sequence of K bits is encoded by an LDPC encoder to produce a code-

word $\mathbf{c}$ of length N bits in the code $\mathcal{C}$. Thus, the code rate is $R = K/N$. The code $\mathcal{C}$ can be regular or irregular LDPC code and the minimum Hamming distance of the code is denoted by $l_{\min}$. Then, every group of q bits in the codeword $\mathbf{c}$ is mapped by some mapping scheme μ to one signal point in the constellation $\mathcal{X}$ of size $M = 2^q$ to produce the transmitted signal $\mathbf{x} = [x_1, x_2, \cdots, x_k, \cdots, x_{N_s}]$, where $N_s = N/q$ is the number of symbols. The modulated signal $\mathbf{x}$ is transmitted over a frequency-flat fading channel. Let $\mathbf{h} = [h_1, h_2, \cdots, h_k, \cdots, h_{N_s}]$ denote the vector of the fading coefficients affecting the symbols of the transmitted signal $\mathbf{x}$. When the channel is an uncorrelated Rayleigh fading channel, the fading coefficients in $\mathbf{h}$ are independent and identically distributed (i.i.d.) random variables with Rayleigh distribution. When an AWGN channel is considered, these coefficients are set to one. After transmission over the fading channel, a noise vector $\mathbf{w} = [w_1, w_2, \cdots, w_k, \cdots, w_{N_s}]$ is added to the signal vector $\mathbf{x}$ to give the received vector $\mathbf{y} = [y_1, y_2, \cdots, y_k, \cdots, y_{N_s}]$. Thus, the kth symbol of $\mathbf{y}$ can be represented as:

$$y_k = h_k x_k + w_k \tag{3.1}$$

where w_k is a zero-mean complex Gaussian random variable with variance $N_0/2$ per dimension. Here, N_0 is one-sided power spectral density of the AWGN.

Equation (3.1) represents the discrete model of the received signal. In practice, the discrete received signal $\mathbf{y}$ is obtained by a using a pair of matched filters and each component of $\mathbf{y}$ is a complex number, i.e., a pair of real numbers. The task of the receiver is to use $\mathbf{y}$ to decide which codeword was transmitted among all 2^K possible codewords. It means that the K binary bits should be decided based on $2N_s$ real numbers contained in $\mathbf{y}$. Since the vector $\mathbf{y}$ contains all information about K information bits, it is often called the sufficient statistics of the received signal as discussed in Chapter 1.

The received vector $\mathbf{y}$ can be geometrically represented as a point in a space of $2N_s$ dimensions. For an AWGN channel, the 2^K possible transmitted vectors $\mathbf{x}$ corresponding to 2^K possible codewords can also be represented as 2^K points in this $2N_s$-dimensional space. In fact, the received point $\mathbf{y}$ is generated from one of the 2^K transmitted points and the effect of AWGN noise. The optimum receiver or the maximum likelihood (ML) receiver needs to decide one vector among 2^K possible vectors to minimize the error probability.

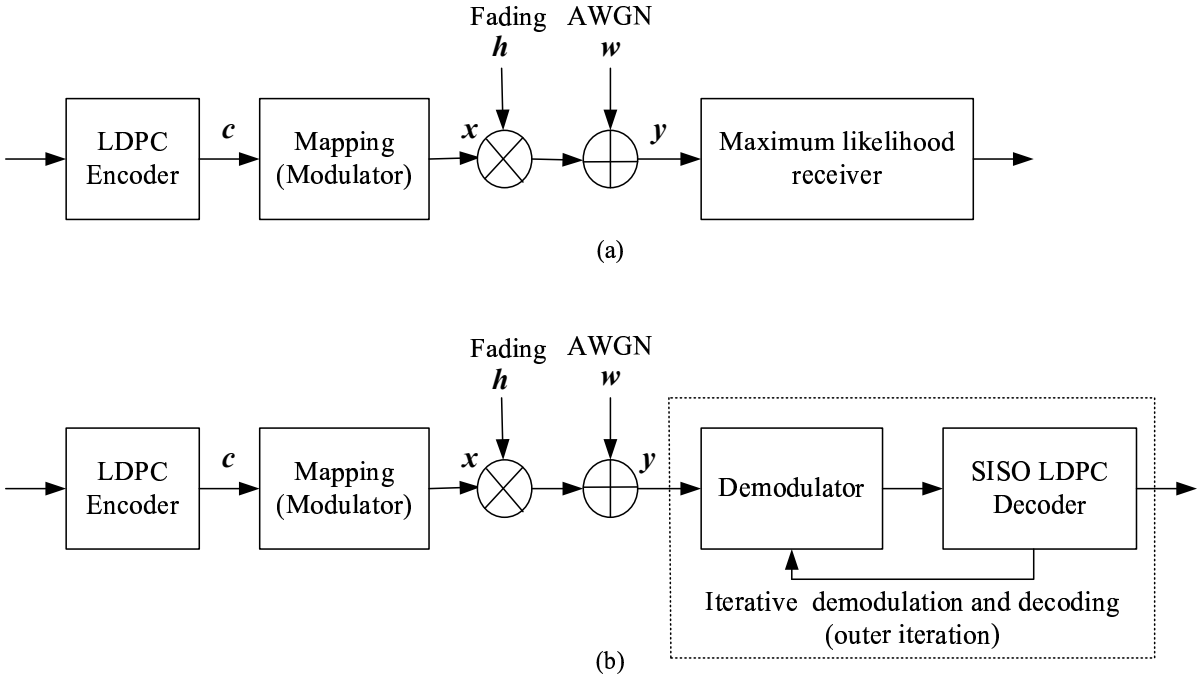


Figure 3.1 The block diagram of LDPC coded modulation systems (a) with the ML receiver, and (b) with a suboptimal receiver based on iterative demodulation/decoding.

Since the PDF of the noise is a Gaussian distribution, which is maximum at the mean value and monotonically decreases with the distance from the mean, the minimum error probability criterion turns out to the minimum distance criterion. It means that the ML receiver should select the signal vector $\mathbf{x}$ that is closest to $\mathbf{y}$ in terms of the Euclidean distance in the $2N_s$-dimensional space. For the information block length K, 2^K Euclidean distances need to be computed and compared for a given received vector $\mathbf{y}$. For a moderate length K, the computational complexity of this ML receiver is thus very large. Therefore, this ML receiver can only be implemented or simulated for very short codes. However, the error performance of this receiver can still be estimated because of its simple decision boundary. The ML receiver is assumed in this thesis when the performance bound is derived, hence, the bound is called the ML bound. In practical systems, when the length of the information sequence is large, the ML receiver is infeasible prohibited due to its computational complexity.

A practical receiver includes a demodulator and a decoder as shown in Fig. 3.1-(b). In

some practical systems, the demodulator implements hard-decision and it outputs q coded bits computed from y_k. The hard-input decoder then detects and corrects errors based on this demodulated binary sequence. However, there is information loss when this hard-decision demodulation/decoding process is used. The performance of the system can be improved by a soft-output demodulator. The outputs of this demodulator contain coded bits together with the accuracy measures of these hard decisions. Such a demodulator is described in detail in Subsection 3.2.1. When the output of the demodulator is soft, the decoder should be a soft-input decoder. The output of the decoder can be hard-decision or even soft-decision. Following the Turbo principle in decoding Turbo codes [4, 75], iteration is often carried out between the demodulator and the decoder. To implement the iteration at the receiver of this coded modulation system, the soft-input soft-output (SISO) demodulator and SISO decoder are required. It should be noted that practical receivers, such as the iterative receivers, are suboptimum and their performance is inferior to that of the ML receiver.

In this chapter, a practical receiver that includes a SISO demodulator and a SISO LDPC decoder as shown in Fig. 3.1-(b) is also employed. The SISO demodulator computes the soft-decisions of the coded bits based on the received signal and the *a priori* probabilities from the previous iteration. Then, the soft output of the demodulator is used as the soft input of the SISO LDPC decoder to generate the new soft output or the *a posteriori* probabilities in the next iteration. The *extrinsic* information is computed by subtracting the *a priori* probabilities from the *a posteriori* probabilities and it is fedback to the SISO demodulator in the next iteration.

For the receiver that implements iterative demodulation/decoding as described above, two kinds of a SISO LDPC decoder are considered. The first SISO LDPC decoder is the sum-product decoder, which is often employed for LDPC codes. Since, the sum-product algorithm was described in detail in Chapter 2, it is not discussed again here. The second SISO LDPC decoder under consideration is the SISO ordered-statistics decoder (OSD). This decoder is applicable for any short linear block codes. A detailed description of SISO OSD algorithm based on the log-likelihood ratios (LLR) is presented in Subsection 3.2.2.

3.2.1 The Demodulator

The procedure of the SISO demodulator is summarized as follows [67]. The SISO demodulator computes *a posteriori* probabilities of coded bits $\Pr(c_k^i = b|y_k)$ based on the received signal y_k and the symbols in the constellation $\mathcal{X}$. The *a posteriori* probability of the kth bit taking the b value is simply a summation of the *a posteriori* probabilities of symbols that have label b at this kth position. The *a posteriori* probabilities of symbols $\Pr(x_k|y_k)$ can be represented by the *a priori* probabilities $\Pr(x_k)$, the *a priori* probability density $p(y_k)$ and the conditional *a priori* probability density $p(y_k|x_k)$ based on the Bayes's theorem. The *a posteriori* probabilities for coded bits are written as:

$$\Pr(c_k^i = b|y_k) = \sum_{x_k \in \mathcal{X}_b^i} \Pr(x_k|y_k) = \frac{\sum_{x_k \in \mathcal{X}_b^i} p(y_k|x_k) \Pr(x_k)}{p(y_k)} \tag{3.2}$$

where c_k^i is the coded bit of codeword $\mathbf{c}$ that is mapped to the ith bit of the kth symbol x_k, and b can be two alternative values 0 or 1. The variable $\mathcal{X}_b^i$ denotes the subset of $\mathcal{X}$ which contains the symbols whose ith labeling bit equals b.

At the first-iteration demodulation, the *a priori* probabilities $\Pr(x_k)$ of all the symbols are equal, i.e., these *a priori* probabilities are $1/M$, where M is the size of the constellation. From the second-iteration demodulation, the *a priori* probability $\Pr(x_k)$ of a particular symbol is a product of the *a priori* probabilities of the coded bits mapped to this symbol. These *a priori* probabilities of the coded bits are computed from the previous iteration. That is,

$$\Pr(x_k) = \prod_{j=1}^{q} \Pr(c_k^j = c^j(x_k); I) \tag{3.3}$$

Here, $c^j(x_k) \in \{0, 1\}$ is the value of the jth bit of the label of x_k, which is determined by the mapping μ.

The output of the demodulator for the second iteration is the *extrinsic* probabilities of the coded bit, i.e., these probabilities are additional information and exclude the *a priori* probabilities that are supplied by the previous iteration. Mathematically, these *extrinsic*

probabilities for the second iteration of the demodulator can be computed as follows:

$$\begin{aligned}\Pr(c_k^i = b; O) &= \frac{\Pr(c_k^i = b|y_k)}{\Pr(c_k^i = b; I)} \\ &= \frac{\sum_{x_k \in \mathcal{X}_b^i} p(y_k|x_k) \Pr(x_k)}{\Pr(c_k^i = b; I) p(y_k)} \\ &= \frac{\sum_{x_k \in \mathcal{X}_b^i} \left[p(y_k|x_k) \prod_{j \neq i} \Pr(c_k^j = c^j(x_k); I) \right]}{\Pr(c_k^i = b; I) p(y_k)}\end{aligned} \tag{3.4}$$

For an uncorrelated Rayleigh fading channel, if the receiver knows the channel state information (CSI), the conditional probability density function $p(y_k|x_k)$ is computed as follows:

$$p(y_k|x_k) = \frac{1}{\pi N_0} \exp\left[-\frac{|y_k - h_k x_k|^2}{N_0} \right] \tag{3.5}$$

where, recall that, $N_0/2$ is variance of white noise in each dimension. Note that, x_k and y_k are complex numbers while the fading coefficient h_k is a real number (due to the assumption of perfect CSI, the phase is perfectly recovered). For an AWGN channel, h_k is set to one. The log likelihood probability ratio (LLR) can be computed as:

$$\mathrm{LLR}(c_k^i) = \log \frac{\Pr(c_k^i = 0; O)}{\Pr(c_k^i = 1; O)} \tag{3.6}$$

Thus, the output of the demodulator is calculated by (3.4), (3.5) and (3.6). Note that, the PDF $p(y_k)$ in (3.4) is often a complicated function, but it is cancelled in the calculation of the LLR outputs in (3.6).

3.2.2 The Soft-Input Soft-Output OSD Decoder

The ordered-statistics decoder (OSD) is one of the reliability-based decoders for linear block codes [76]. The first algorithm in this type of decoders is the generalized minimum distance (GMD) decoder devised in [36]. This decoding approach was later generalized to Chase's algorithms [77]. The reliability-based decoders also compute and compare the distances from candidate codewords to the received signal as in the ML decoder in order to select the best codeword. However, the reliability of the received coded bits is used to restrict the search region of the reliability-based decoders. Instead of searching over all codewords as in the case of ML decoding, the reliability-based decoders only make

decision on a set of the most possible candidates. These candidates surround the most reliable candidate. In this way, the complexity of reliability-based decoders is relatively reduced compared with ML decoding, but the performance is still close to the performance of ML decoding for moderate code lengths. Different types of reliability-based decoding use different methods to choose the first candidate and to extend the search region. For coded modulation systems, the soft-outputs of the demodulator are used to determine the reliability orders of the candidates as well as the cost functions of the candidates.

The hard-output OSD algorithm makes decisions on a set of codewords devised from the most reliable independent positions (MRIP) of the received signal and the error patterns added to these MRIPs. The hard-output OSD algorithm was proposed in [18] for a BPSK modulation system. The MRIPs are ordered and the cost function of a codeword is directly computed based on the received signal y_k of BPSK modulation. In this section, the MRIPs are determined and the cost function of a codeword is calculated based on the output LLRs of the demodulator.

The soft-output OSD decoder [74] is a two-stage decoder. The first stage is a hard-output OSD decoder. In the second stage, the soft-output of each position is obtained by a joint estimation between the decided codeword and the closest codeword to the received signal with an opposite value at this position. This codeword is found by the OSD stages working on punctured codes of the original code. The details of the hard-output OSD stage and the soft-output reprocessing stage are briefly presented next.

The Hard-Output Ordered-Statistics Decoding Stage

First, the LLR sequence corresponding to a codeword is re-ordered in decreasing values of reliability. The code corresponding to this ordering is referred to as $\tilde{C}$. The K most reliable independent positions (MRIPs) form an information set called the most reliable basis (MRB). An initial codeword $\mathbf{c}_0$ is constructed by making hard decisions of K MRIPs and re-encoding these K bits. Then $\mathbf{c}_0$ is reprocessed as follows [18]:

1. For $0 \leq j \leq i$, $\binom{K}{j}$ candidate codewords are constructed by adding all possible error patterns of weight j to the K MRIPs of $\mathbf{c}_0$ and re-encoding.

2. The decoding cost of each constructed codeword is computed by a product of bit probabilities or a sum of the corresponding LLRs.

3. The candidate with the best cost is chosen by the decoder.

The above decoding procedure is called order-i reprocessing of the hard-output OSD-i decoder. The decoding cost of a candidate codeword $\mathbf{c}$ can be computed as follows:

$$L_{\mathbf{c}} = \sum_{l=1}^{N} (-1)^{c_l} LLR(c_l) \tag{3.7}$$

If the cost function is calculated based on the cost of the initial codeword $\mathbf{c}_0$, the computational complexity can be reduced by a summation of the costs of bits at positions of 1 in the error patterns only.

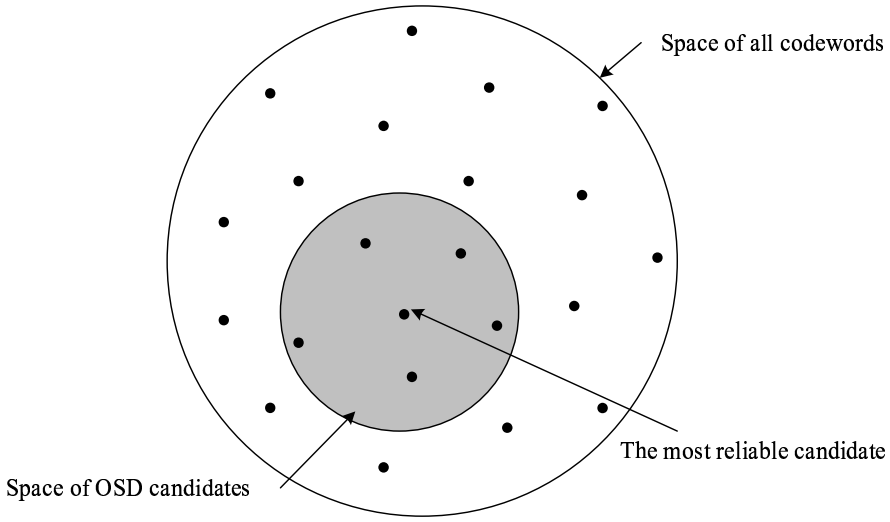


Figure 3.2 The space of OSD candidates.

Figure 3.2 illustrates the space of OSD candidate codewords. The OSD decoder finds the best codeword in terms of the cost function among these candidates. The OSD candidates surround the most reliable candidate, i.e., the initial codeword. The computational complexity of OSD algorithm depends on the order i of OSD and the length K of the information sequence. The OSD algorithm of order i requires processing of a total of

$$1 + \binom{K}{1} + \cdots + \binom{K}{i} \tag{3.8}$$

candidates to make a decision. The OSD algorithm with order K is the maximum likelihood decoding, which requires the processing of 2^K codewords.

The Soft-Output Reprocessing Stage

The soft-output of a coded bit is computed based on a joint estimation of two codewords, namely, the most likely codeword $\mathbf{c}_{ML}$ in the code C and the most likely codeword $\mathbf{c}_j$ in the subset $C(j)$ of codewords that are different to $\mathbf{c}_{ML}$ at the jth position. That is,

$$L_j = (-1)^{c_{ML,j}} \left[L(\mathbf{c}_{ML}) - L(\mathbf{c}_j) \right] \tag{3.9}$$

The above expression means that the soft estimation of the LLR by this SISO OSD module is equivalent to that of the max-log-MAP algorithm [78].

To describe steps of the soft-output order-i reprocessing, the following definitions are required. The error pattern $\mathbf{e}(j)$ of $\tilde{C}$ is an error pattern that has $e_l(j) = 1$ for $l = j$ and $e_l(j) = 0$ otherwise, $1 \leq l \leq K$. The values of $e_l(j)$'s for $K \leq l \leq N$ are determined by the generator matrix of $\tilde{C}$. The code $\tilde{C}(j)$ is a punctured code of $\tilde{C}$ at the jth position. The codeword $\mathbf{c}_{ML}(j)$ of $\tilde{C}$ is a codeword obtained by complementing position-j in $\mathbf{c}_{ML}$, i.e., $\mathbf{c}_{ML}(j) = \mathbf{c}_{ML} \oplus \mathbf{e}(j)$.

The steps of the soft-output order-i reprocessing can now be summarized as follows [74]:

1. Find the most likely codeword $\mathbf{c}_{ML}$ by the hard-output ordered-statistic decoding stage.

2. For each bit-j of the most reliable basis (MRB) $(1 \leq j \leq K)$:

   - Setup the first soft output values of the least reliable positions (LRPs) in the support of $\mathbf{e}_j$ based on $\mathbf{c}_{ML}$ and $\mathbf{c}_{ML}(j)$.
   - Find the most likely codeword $\mathbf{c}(j)$ in $\tilde{C}(j)$ by a hard-output order-i reprocessing.
   - Compute L_j based on $\mathbf{c}_{ML}$ and $\mathbf{c}(j)$.

- Update the soft output values of the LPRs in the support of $\mathbf{c}_{ML} \oplus \mathbf{c}(j)$ with L_j.

3. For each bit-j of the LPRs $(K+1 \leq j \leq N)$, choose the smallest soft output value associated with each bit-j.

Note that, the generator matrix of $\tilde{C}(j)$ needs to be properly derived according to the value of the jth bit of $\mathbf{c}_{ML}$ in order to generate the exact punctured code of $\tilde{C}$. Moreover, the soft-outputs should be re-ordered again to correspond to the original positions.

3.3 Performance Bound of the ML Decoding

In this section, an upper bound on the error performance of the system with ML receiver is derived. The block diagram of this system is shown in Fig. 3.1-(b). Although this system might not be practically implemented, its error performance is of interest since it provides the performance limit for any suboptimum receivers such as the receivers based on the sum-product decoding or OSD.

3.3.1 Performance Bound for AWGN channels

For the system in Fig. 3.1-(b), the ML decoding makes a decision to the nearest signal of $\mathbf{y}$. The demapping of the chosen signal to the binary bits is then performed. Note that the ML decision is carried out in $\{(\mathbb{R}^2)^{N_s}\}$ signal space. A codeword error occurs when the received signal $\mathbf{y}$ exceeds the boundary of the decision region of the transmitted signal $\mathbf{x}$. The bit errors can be computed easily from the codeword errors. Thus the bit error probability depends on the locations of the signals $\mathbf{x}$ in $\{(\mathbb{R}^2)^{N_s}\}$. It is very difficult, if not impossible, to exactly determine the locations of the transmitted signals because LDPC codes are constructed randomly. Fortunately, the union bounding techniques only need to know the Euclidean distance *spectrum* of the transmitted signals to approximately calculate the error probability. Note that, the error probability depends not only on the minimum distance of the signals, but also on the distance spectrum of the signals. Therefore, in order to apply the union bounds, the Euclidean distance spectrum of the transmitted signals needs to be computed first.

In [73], the authors developed a procedure to compute the Euclidean distance spectrum

of Turbo coded modulation systems in order to determine the union bound on the BER. This method can also be applied for the systems based on linear, randomly-constructed codes such as LDPC codes. Such an application is considered to evaluate the BER performance of LDPC coded modulation systems. There are some important differences when the method of [73] is applied to the systems considered in this section. With a given parity check matrix $\mathbf{H}$, an LDPC code can be encoded in different ways [39]. This is because different generator matrices can be found for a given parity check matrix. Therefore, it can be assumed that the positions of the systematic bits (information bits) are randomly distributed in a codeword. With this assumption, the bound of the coded bit error probability, rather than the bound of the information bit error probability, shall be derived. In fact, simulation results also show that the error probabilities of the information bits and coded bits are approximately the same. When the union bound of the coded bit error probability is computed, the Hamming weight spectrum of LDPC codes is used instead of the weight enumerator function of Turbo codes as in [73].

The bit error probability when a codeword $\mathbf{c}$ is selected (i.e., the signal $\mathbf{x}$ is transmitted) can be computed as follows:

$$P_{e,\mathbf{c}} = \sum_{\mathbf{c}' \neq \mathbf{c}} \frac{W_{\mathbf{c},\mathbf{c}'}}{N} \Pr\left(\mathbf{y} \in \Lambda_{\mathbf{x}'} | \mathbf{x}\right) \tag{3.10}$$

where $\mathbf{c}'$ is another codeword in $\mathcal{C}$, $W_{\mathbf{c},\mathbf{c}'}$ is the Hamming distance between the codewords $\mathbf{c}$ and $\mathbf{c}'$ and $\Lambda_{\mathbf{x}'}$ is the decision region associated with signal $\mathbf{x}'$ of the codeword $\mathbf{c}'$. Generally, the codewords are chosen equally likely, then the average bit error probability of the system is simply given by:

$$P_e = \frac{1}{2^{NR}} \sum_{\mathbf{c} \in \mathcal{C}} P_{e,\mathbf{c}} \tag{3.11}$$

The union bound can then be written as:

$$P_e \leq \frac{1}{2^{NR}} \sum_{\mathbf{c} \in \mathcal{C}} \sum_{\mathbf{c}' \neq \mathbf{c}} \frac{W_{\mathbf{c},\mathbf{c}'}}{N} \Pr(\mathbf{c} \rightarrow \mathbf{c}') \tag{3.12}$$

where the pairwise error probability $\Pr(\mathbf{c} \rightarrow \mathbf{c}') \geq \Pr\left(\mathbf{y} \in \Lambda_{\mathbf{x}'} | \mathbf{x}\right)$ is the probability of deciding on the codeword $\mathbf{c}'$ given that the codeword $\mathbf{c}$ was selected at the transmitter side. Corresponding to each pair of codewords, there is an error sequence $\mathbf{e}$. In essence, (3.12)

is a sum of the pairwise codeword error probabilities weighted by the Hamming distances of the codewords. Since LDPC codes are linear codes, the error sequences are also the codewords. Therefore, the Hamming weight spectra of error sequences are exactly the Hamming weight spectra of the code.

Due to the randomization in constructing of LDPC codes, it is difficult to determine the error performance of a given code. Instead of considering a specific code, the error performance bound is averaged over the code ensemble in which a code is a permutated version of another code in this ensemble. It means that the columns of a parity check matrix are permuted to create the parity check matrixes of other codes. In fact, this ensemble is only a subset of the LDPC ensemble of given check node and variable node degree distributions. This procedure is necessary in order to apply combinatoric techniques. This is similar to the case of Turbo codes, where the error performance bound is averaged over all random interleavers, which essentially correspond to different codes of a Turbo code ensemble. Thus, the codes of this ensemble are defined by the same graph topology and have the same Hamming distance spectrum. After averaging (3.12), the average union bound can be written as follows:

$$\overline{P}_e \leq E\left[\frac{1}{2^{NR}}\sum_{\mathbf{c}\in\mathcal{C}}\sum_{\mathbf{c}'\neq\mathbf{c}}\frac{W_{\mathbf{c},\mathbf{c}'}}{N}\Pr(\mathbf{c}\rightarrow\mathbf{c}')\right] \tag{3.13}$$

For computing the pairwise codeword error probabilities, the squared Euclidean distance should be known for each codeword pair or each error sequence. The squared Euclidean distance between two codewords are the squared Euclidean distance between the two symbol sequences corresponding to the two codewords. This Euclidean distance between the two codewords is simply the sum of the contributed Euclidean distances between two symbols in each modulation plane. These two symbols are determined by the corresponding bit groups and the mapping rule. Therefore, if there is no difference in the bit groups, these two symbols are identical and the Euclidean distance between them is zero. In other words, such bit groups do not contribute to the Euclidean distance between the two symbol sequences.

Consider a given error sequence, which is the modulo-2 sum of two codewords. The Hamming weight of an error sequence is the number of ones in the sequence. For each error sequence, only pairs of symbols that correspond to the bit groups containing errors (i.e., bits 1) contribute to the total Euclidean distance. Recall that the number of error sequences with a given weight is determined by the Hamming weight spectrum of the code. However, the positions of the errors are not determined because of the randomization in creating LDPC codes. Error sequences of a given weight can be classified to some types of error sequences according to the number of bit groups with the same weights. For example, in the case of 8-PSK modulation, the weight of a bit group of an error sequence (called error-bit group) can be 0, 1, 2 or 3. Note that, the bit group of weight 0 does not contribute to the total Euclidean distance.

Let n_i, $0 < i < q$, denote the number of error-bit groups of weight i. An error sequence can then be described by the parameters n_i. Let $\mathbf{n}$ be the vector that contains these parameters. Then $\mathbf{n}$ can be referred to as the type of error sequences. Note that, each type of error sequences can be mapped to different Euclidean distances [73]. Thus, corresponding to each type $\mathbf{n}$ of error sequences, the corresponding Euclidean distance $D_{\mathbf{n}}$ is a random variable.

After classifying the error sequences, the union bound can be rewritten as:

$$\overline{P}_e \leq \sum_{l=l_{\min}}^{N} \sum_{n_1=0}^{N_s} \cdots \sum_{n_q=0}^{N_s} \frac{l}{N} \overline{f}(\mathbf{n}) E\left[Q\left(\sqrt{\frac{D_{\mathbf{n}}^2}{N_0/2}}\right)\right] \tag{3.14}$$

Here, $f(\mathbf{n})$ is the number of error sequences of type $\mathbf{n}$ corresponding to a given transmitted codeword and $\overline{f}(\mathbf{n})$ is the expected value of this function over the ensemble. The expected total number of error sequences of type $\mathbf{n}$ is $2^{NR}\overline{f}(\mathbf{n})$. The function $f(\mathbf{n})$ is computed as follows:

$$\overline{f}(\mathbf{n}) = \mathcal{N}(l) \frac{P_{l,\mathbf{n}}}{\binom{N}{l}} \tag{3.15}$$

In the above equation, $\mathcal{N}(l)$ is the number of codewords of weight l, which is also the number of error sequences of weight l for any transmitted codeword. The quantity $P_{l,\mathbf{n}}$ is the number of possible cases of error sequences of type $\mathbf{n}$ with a weight l. This quantity

can be computed by the following equation:

$$P_{l,\mathbf{n}} = \begin{cases} \binom{N_s}{n_0, n_1, \ldots, n_q} \prod_{i=1}^{i=q} \binom{q}{i}^{n_i}, & \text{if } \sum_{i=1}^{i=q} i n_i = l \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

Here, the multinomial coefficient $\binom{N_s}{n_0, n_1, \ldots, n_q} = \frac{N_s!}{n_0! n_1! \ldots n_q!}$ is the number of possible locations of certain type of errors among N_s channel symbols. For the product term, each coefficient $\binom{q}{i}$ is the number of possible error patterns of a bit group of size q with i errors. Furthermore, the number of terms in the summation of (3.14) can be reduced by smaller upper limits. For example, one can use $\lfloor (l - \sum_{k=0}^{i-1} k n_k)/i \rfloor$ as the upper limit for n_i if all the n_k with $k < i$ are known.

To compute the average union bound over the random variable $D_{\mathbf{n}}$, two simplifying assumptions are needed. The first assumption is that all N_s channel symbol errors are independent of each other. The second assumption is that all the points in the constellation are equally likely to be used. Now, the total distance $D_{\mathbf{n}}$ can be computed from the single symbol error distances D_k that are also random variables as:

$$D_{\mathbf{n}}^2 = \sum_{k=1}^{n_1+\cdots+n_q} D_k^2 \tag{3.17}$$

Due to the first assumption, the probability mass function (PMF) of $D_{\mathbf{n}}$ can be computed from the PMFs of D_k. The PMFs of D_k for a given error-bit group can be computed from the mapping and the constellation by the second assumption. Thus, the bound can be rewritten as

$$\overline{P}_e \leq \sum_{l=l_{\min}}^{N} \sum_{n_1=0}^{N_s} \cdots \sum_{n_q=0}^{N_s} \sum_{j=1}^{k_{\mathbf{n}}} \frac{l}{N} \overline{f}(\mathbf{n}) p_{\mathbf{n},j} Q\left(\frac{\Delta_{\mathbf{n},j}}{\sqrt{2N_0}}\right) \tag{3.18}$$

where $p_{\mathbf{n},j} = P[D_{\mathbf{n}}^2 = \Delta_{\mathbf{n},j}^2]$, $j = 1, 2, \ldots, k_{\mathbf{n}}$. Here, $k_{\mathbf{n}}$ is the number of distinguishable Euclidean distances and each distinguishable Euclidean distance has a probability $p_{\mathbf{n},j}$.

An example is given next to illustrate the procedure of calculating $\Delta_{\mathbf{n},j}$ and $p_{\mathbf{n},j}$. Consider 8-PSK modulation and Gray mapping as shown in Fig. 3.3. Then the random variable D_k can take on one of several values $d_{i,h}$ with probabilities $p_{i,h}$, respectively. Here,

i, $0 \leq i \leq 3$, is the number of errors in one bit group and h, $0 \leq h \leq h_{\max}$, is the number of possible Euclidean distances. If there is one error (i.e., $i = 1$) in a bit group, D_k can take on two values $d_{1,1} = \sqrt{2-\sqrt{2}}\sqrt{E_s}$ and $d_{1,2} = \sqrt{2+\sqrt{2}}\sqrt{E_s}$ with probabilities $p_{1,1} = 8/12 = 2/3$ and $p_{1,2} = 4/12 = 1/3$, respectively. If there are two errors ($i = 2$) in one bit group, D_k can be $d_{2,1} = 2\sqrt{E_s}$ and $d_{2,2} = \sqrt{2}\sqrt{E_s}$ with probabilities $p_{2,1} = 1/3$ and $p_{2,2} = 2/3$, respectively. Finally, when all three bits are in error ($i = 3$), the value of D_k is $d_{3,1} = \sqrt{2+\sqrt{2}}\sqrt{E_s}$ with certainty. Thus, the PMFs of all D_k's can be computed. The random variable $D^2_{\mathbf{n}}$ thus can take on the following values:

$$\Delta^2_{\mathbf{n},j} = \sum_{i=1}^{q} \sum_{h=1}^{h_{\max}} n_{i,h} d^2_{i,h} \tag{3.19}$$

where $n_{i,h}$ is the number of distances $D_k = d_{i,h}$. Note that $0 \leq n_{i,h} \leq n_i$ and $\sum_{h=0}^{h_{\max}} n_{i,h} = n_i$. The probability of $D^2_{\mathbf{n}} = \Delta^2_{\mathbf{n},j}$ with a given set of $n_{i,h}$ is simply:

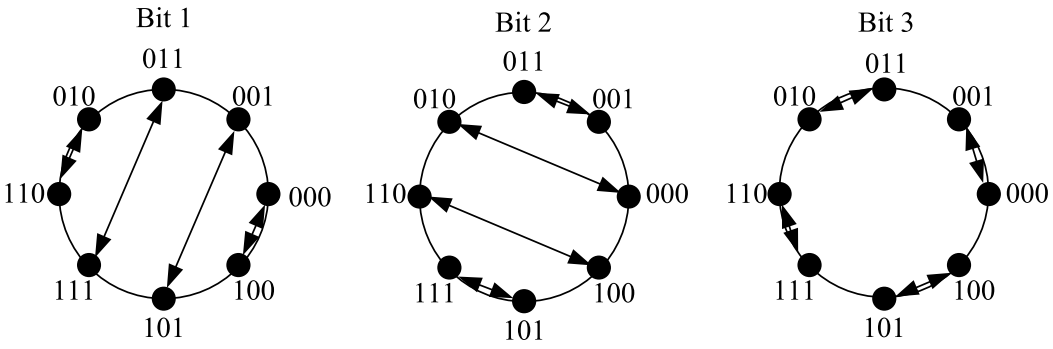
$$p_{\mathbf{n},j} = \prod_{i=1}^{q} \binom{n_i}{n_{i,0}, \ldots, n_{i,h}} \prod_{h=0}^{h_{\max}} p_{i,h}^{n_{i,h}} \tag{3.20}$$

The above procedure shows all the necessary computations to determine the union bound in (3.18).

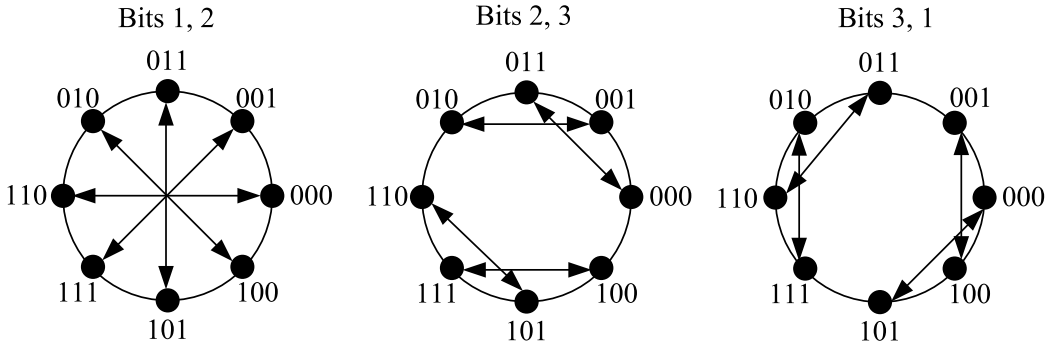
Before closing this section, it is appropriate to discuss the tightness of the union bound. For the LDPC coded modulation systems under consideration, the pair of modulator and demodulator and the channel form a discrete memoryless channel (DMC) for the coding and decoding blocks. The union bound of the ML receiver is considered to be tight when the rate R_s (bits/symbol) of the system is above the cut-off rate R_0 (bits/symbol) of this discrete channel [73, 79].

Here, the cutoff rate R_0 is defined as the rate at which the tangent to the reliability function $E(R)$ of slope -1 intersects the R axis as depicted in Fig. 2.3. The cutoff rate can partly describe $E(R)$ because it shows how difficult it is to approach the channel capacity. For an AWGN channel, the cutoff rate R_0 corresponding to a given constellation can be computed by the following equation [80, 81]:

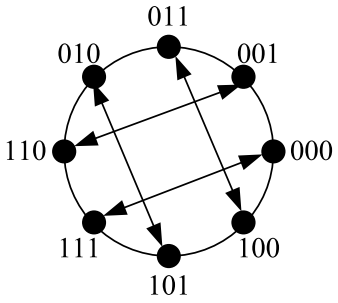
$$R_0 = -\log_2 \left(\min_{\{q_i\}} \sum_{i=1}^{M} \sum_{j=1}^{M} q_i q_j e^{-|x_i - x_j|^2/4N_0} \right) \tag{3.21}$$



(a) One error in a bit group



(b) Two errors in a bit group



(c) Three errors in a bit group

Figure 3.3 The Euclidean distances of 8-PSK constellation with Gray mapping.

where $\{x_i\}$ are the symbols of the constellation $\mathcal{X}$. These symbols are transmitted over an AWGN channel with probabilities $\{q_i\}$. When the modulation symbols are used with the same probabilities, (3.21) reduces to:

$$R_0 = -\log_2 \left[\sum_{i=1}^{M} \sum_{j=1}^{M} \left(\frac{1}{M} \right)^2 e^{-|x_i - x_j|^2 / 4N_0} \right] \tag{3.22}$$

Thus, a value $[E_b/N_0]^{(*)}$ can be numerically computed by solving (3.22) when the cutoff rate R_0 is assigned to the data rate R_s . The union bound is considered as a tight bound when the signal-to-noise ratio is above this $[E_b/N_0]^{(*)}$. For example, for the coded modulation scheme with code rate $1/2$ and 8-PSK modulation, this $[E_b/N_0]^{(*)}$ value is 3.28 dB.

3.3.2 Performance Bounds for Fading Channels

The union bound of the BEP for a fading channel is different from that for an AWGN channel due to the difference in computing the pairwise error probability (PEP). This means that Equation (3.13) can still be applied for fading channels if the pairwise error probability in this equation can be computed for fading channels. Such a computation of the pairwise error probability can be carried out similarly to that for a trellis coded modulation (TCM) system.

The derivation of the exact PEP for a fading channel is obtained by averaging the Gaussian probability integral over the PDFs of the fading coefficients [82]. When the channel-state information is perfectly known at the receiver, the conditional PEP of deciding $\mathbf{x}'$ when indeed $\mathbf{x}$ was transmitted is given by

$$\Pr(\mathbf{x} \to \mathbf{x}'|\mathbf{h}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{k\in\omega} h_k^2|x_k' - x_k|^2}\right) \tag{3.23}$$

where ω is the set of all k for which $x_k' \neq x_k$. Using the following alternative expression of the Q-function [83]:

$$Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) \mathrm{d}\theta \tag{3.24}$$

Equation (3.23) can be rewritten as:

$$\Pr(\mathbf{x} \to \mathbf{x}'|\mathbf{h}) = \frac{1}{\pi}\int_0^{\pi/2} [f(\theta)]^{d^2(\mathbf{x},\mathbf{x}')}\, d\theta \tag{3.25}$$

where

$$f(\theta) = \exp\left\{-\frac{1}{4N_0\sin^2\theta}\right\} \tag{3.26}$$

and

$$d^2(\mathbf{x},\mathbf{x}') = \sum_{k\in\omega} h_k^2|x_k' - x_k|^2 \tag{3.27}$$

Then, the unconditional pairwise error probability is determined as

$$\begin{aligned}\Pr(\mathbf{x}\to\mathbf{x}') &= E_{\mathbf{h}}\left\{\frac{1}{\pi}\int_0^{\pi/2}\left[f(\theta)\right]^{d^2(\mathbf{x},\mathbf{x}')}\,d\theta\right\}\\ &= \frac{1}{\pi}\int_0^{\pi/2}\prod_{k\in\omega}E_{h_k}\left\{\left[f(\theta)\right]^{h_k^2|x_k'-x_k|^2}\right\}d\theta\end{aligned} \tag{3.28}$$

For a Rayleigh fading channel, averaging over the pdf of the normalized Rayleigh distribution of h_k yields:

$$E_{h_k}\left\{\left[f(\theta)\right]^{h_k^2|x_k'-x_k|^2}\right\} = \frac{1}{1+\frac{E_s|x_k'-x_k|^2}{4N_0\sin^2\theta}} \tag{3.29}$$

Thus, the pairwise error probability for a Rayleigh fading channel is [82]:

$$\Pr(\mathbf{x}\to\mathbf{x}') = \frac{1}{\pi}\int_0^{\pi/2}\prod_{k\in\omega}\frac{1}{1+\frac{E_s|x_k'-x_k|^2}{4N_0\sin^2\theta}}d\theta \tag{3.30}$$

When the above PEP is substituted in (3.18) in place of the Q-function, the union bound on the BER is given as follows:

$$\overline{P}_e \le \sum_{l=l_{\min}}^{N}\sum_{n_1=0}^{N_s}\cdots\sum_{n_q=0}^{N_s}\sum_{j=1}^{k_{\mathbf{n}}}\frac{l}{N}\overline{f}(\mathbf{n})p_{\mathbf{n},j}\left[\frac{1}{\pi}\int_0^{\pi/2}\prod_{i=1}^{|\omega|}\frac{1}{1+\frac{E_s d_{\mathbf{n},j,i}^2}{4N_0\sin^2\theta}}d\theta\right] \tag{3.31}$$

Note that the size of the set ω is precisely the number of channel symbols in error, i.e., $|\omega| = \eta = \sum_i n_i$. The distances $d_{\mathbf{n},j,i}$ are the Euclidean distances of the 2-D constellation that contribute to the Euclidean distance of the symbol sequence. Thus these distances correspond to the distances $d_{i,h}$ in (3.19).

To reduce the computational complexity of the union bound when the number of terms in the summation is large, the pairwise error probability can be further bounded as follows [84]:

$$\Pr(\mathbf{x}\to\mathbf{x}') \le \mathcal{K}(\eta, x_{\min})\prod_{k\in\omega}\frac{1}{1+\frac{E_s}{4N_0}|x_k'-x_k|^2} \tag{3.32}$$

where $x_{\min} = \sqrt{\frac{d_{\min}^2/4N_0}{1+d_{\min}^2/4N_0}}$ with $d_{\min}^2$ is the minimum squared Euclidean distance of the constellation, and

$$\mathcal{K}(\eta, x_{\min}) = \frac{1}{2^\eta(\eta-1)!}\sum_{j=1}^{\eta}\frac{a_j^{(\eta)}}{(1+x)^j} \tag{3.33}$$

The coefficients $a_j^{(\eta)}$ are computed by the following recursive relations [84]:

For $\eta = 1$, $a_1^{(1)} = 1$

For $\eta \geq 2$, $a_1^{(\eta)} = a_2^{(\eta)} = (2\eta - 3)(2\eta - 5) \cdots 3 \cdot 1$

$a_j^{(\eta)} = 2\left(a_{j-1}^{(\eta)} - (\eta - 1)a_{j-2}^{(\eta)}\right)$,

$j = 3, 4, \ldots, \eta$

This bound on the pairwise error probability is very close to the exact pairwise error probability for a Rayleigh fading channel and does not involve the integral.

3.4 Numerical Results

3.4.1 AWGN Channels

Numerical results in this subsection illustrates the derived ML bound for an AWGN channel. The bound is also compared with the simulation results of the error performance of various receivers. First, a very short LDPC code is chosen to verify the accuracy of the derived bound. This LDPC code is a regular $(3, 6)$ code with rate $1/2$ and a length of 24 bits. The Hamming weight spectrum of this code can be easily determined by searching all 2^{12} codewords and it is listed in Appendix C. Simulation of the ML decoding scheme can also be implemented for this very short length. Both QPSK and 8-PSK modulations with Gray mappings are used.

The union bounds and simulation results over an AWGN channel are presented in Fig. 3.4. As can be seen from this figure, the union bounds are consistent with the simulation results, especially at the high signal-to-noise ratio region. The small difference between the simulation result and the bound is merely due to the statistical deviation associated with the simulation. The bound is considered to be a tight bound when the data rate of the system is above the cut-off rate of the channel. Here, the data rate of 1.5 bit/symbol of the system with 8-PSK modulation corresponds to the signal-to-noise ratio of $[E_b/N_0]^{(*)} = 3.28$ dB. This means that the union bound is considered as a tight bound in the range of the signal-to-noise ratio above 3.28 dB, which can be verified by the results shown in Fig 3.4.

Computation of the union bounds is generally quite complicated because of the large number of terms in the first summation of (3.18). However, it is expected that only the first

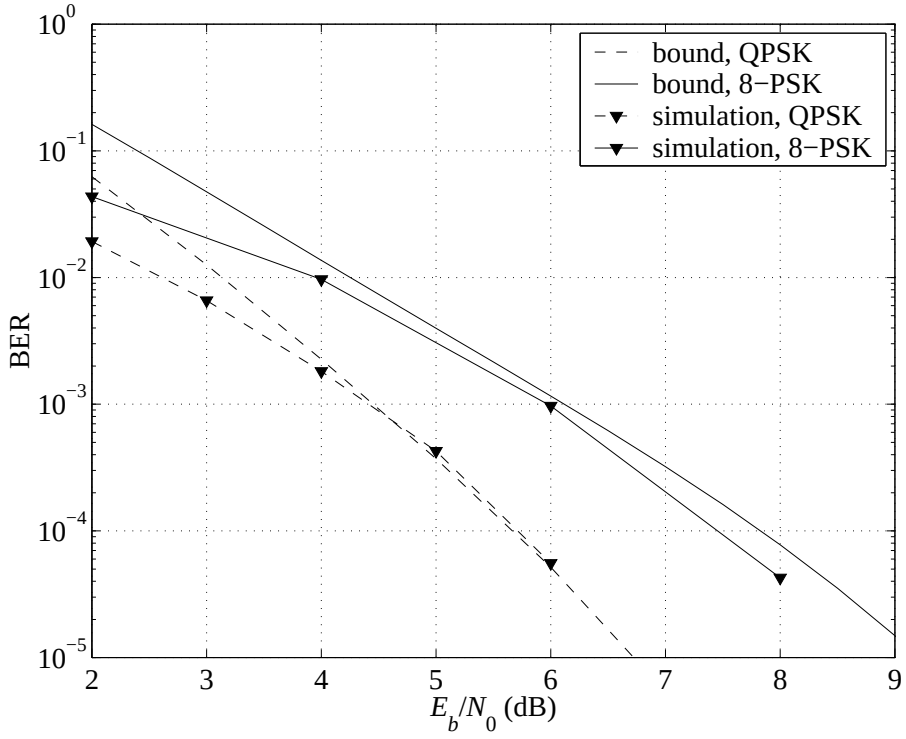


Figure 3.4 The bounds and simulation results for the performance of the ML decoding over an AWGN channel: A regular $(3, 6)$ LDPC code of rate 1/2 and length 24 bits, Gray mapping.

few terms contribute mainly to the union bounds due to their small Hamming distances and the corresponding Euclidean distances. The union bounds computed using different numbers of terms in the first summation of (3.18) are shown in Fig. 3.5. Here, the system employs a regular (3,6) LDPC code of rate 1/2 and a length of 72 bits, an 8-PSK modulation and Gray mapping. The Hamming distance spectrum of this code, listed in Appendix C, is also exhaustively searched over the set of all codewords. It is found that the minimum Hamming distance of this code is 6 with only one nearest codeword. As can be seen from Fig. 3.5, the union bound does not change significantly at high signal-to-noise ratio when more terms are added to the summation. This observation suggests that to reduce the computational complexity of the derived bound for systems employing longer codes, only the first several terms are sufficient instead of using the complete Hamming spectra of the codes.

Fig. 3.6 presents the union bounds for systems employing this (3,6) regular LDPC code and 8-PSK constellation with Gray mapping and semi-set partitioning (SSP) mapping [67].

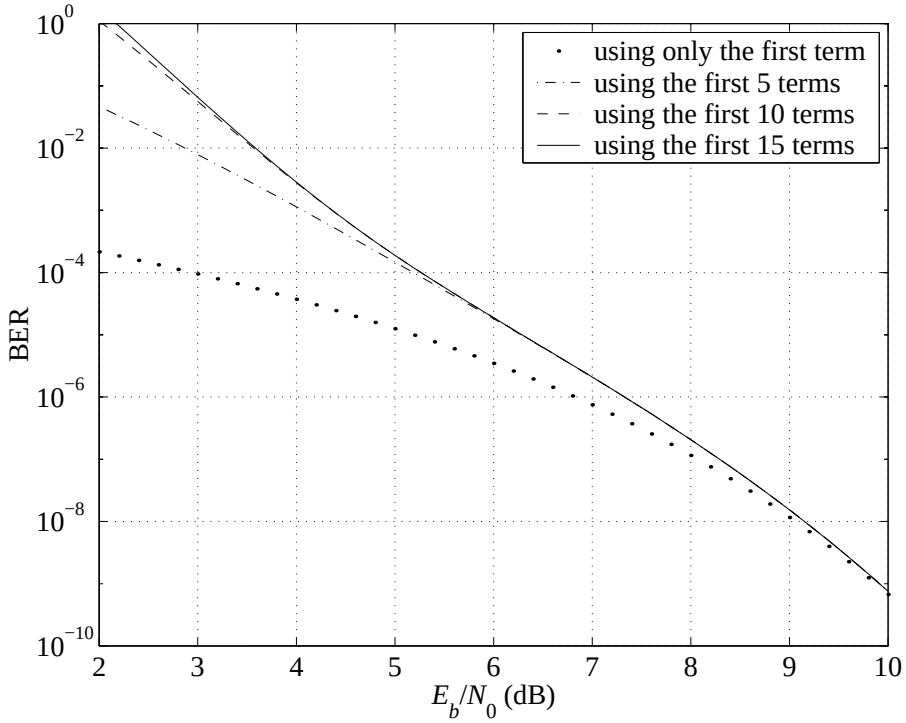


Figure 3.5 The bounds over an AWGN channel: A regular (3,6) LDPC code of rate 1/2 and length 72, 8-PSK modulation and Gray mapping.

The union bounds in Fig. 3.6 show that ML performance of SSP mapping is better than that with Gray mapping at low SNR, whereas Gray mapping outperforms SSP mapping at high SNR. This result is due to the different Euclidean distance spectra created by these two mappings, although they are both created from the same Hamming distance spectrum. In particular, with the same Hamming distance of an error pattern, Gray mapping generally transforms to a bigger Euclidean distance compared to SSP mapping. Due to the characteristic of the Q-function, a few smallest Euclidean distance terms, including the minimum Euclidean distance one, mainly contribute to the performance at high SNR. Therefore, the small number of larger Euclidean distance terms produced by Gray mapping gives a better performance at high SNR. At low SNR, the terms with bigger Euclidean distances become dominant because the Q-function gives similar values at this range of SNR.

The effect of different mappings to the error performance of a coded modulation system is also investigated in [67] for the bit-interleaved coded modulation (BICM) scheme with iterative demodulation and decoding using convolutional codes. It is observed in [67] that,

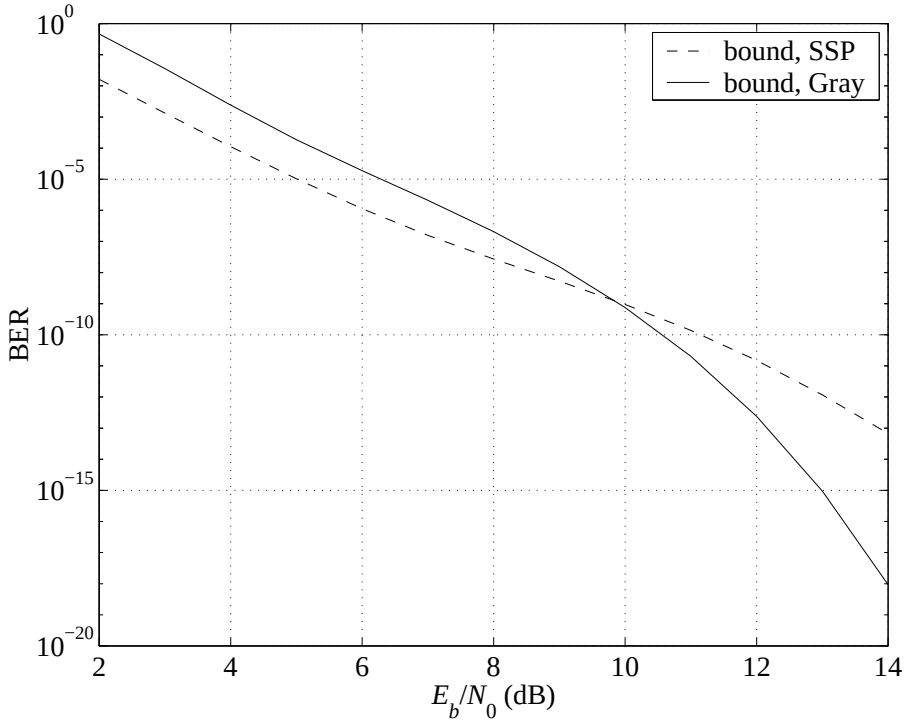


Figure 3.6 The bounds on the performance of the ML decoding over an AWGN channel: A regular $(3, 6)$ LDPC code of rate 1/2 and length 72 bits, 8-PSK modulation, SSP and Gray mappings.

for 8-PSK constellation, the error performance of the systems employing the set partitioning and semi-set partitioning (SSP) mappings is superior to Gray mapping. It is important, however, to point out here that such observation is only applicable when the iterative demodulation and decoding scheme is performed at the receiver of the BICM systems (hence, they are generally referred to as BICM-ID systems). The coded modulation system considered in this chapter is the same as the BICM scheme, with the only exception that an LDPC code is used instead of a convolutional code. It is therefore of interest to investigate the iterative demodulation and decoding scheme and different mappings for the practical LDPC coded modulation systems.

Figure 3.7 presents the simulation results of the iterative demodulation and decoding for the regular (3,6) LDPC code of rate 1/2 and length 72. The 8-PSK constellation with Gray and SSP mappings are considered. Here, two kinds of SISO LDPC decoders, namely the sum-product (SP) decoder and the ordered-statistic decoder (OSD), are considered for iteration between the demodulator and the decoder. The soft-output of these SISO decoders

is used as the *a priori* probabilities of the next outer iteration as described in Section 3.2.

For the sum-product decoding with 20 iterations (i.e., the inner iterations), the results show that the iterative demodulation and decoding (i.e., the outer iterations) does not improve the error performance of the system with Gray mapping. The error performance of the system using SSP mapping and sum-product decoder is improved and almost saturated after only two iterations. It is observed from Fig. 3.7 that the error performance of the iterative demodulation and SP decoding scheme with Gray mapping is superior to the error performance of the scheme with SSP mapping. When comparing the union bounds of the ML receiver and the simulation results of the iterative SP receiver, the performance gaps are about 3.5dB and 7dB for the systems using Gray and SSP mappings, respectively. These wide performance gaps between the ML scheme and the iterative SP scheme are due to the suboptimality of the iterative demodulation and decoding, especially when applied to short LDPC codes. This is the reason for studying decoding schemes such as the OSD in order to approach the error performance of the ML receiver. Of course, the trade-off between performance and complexity should also be considered.

For the coded modulation systems with OSD, simulation results show that the error performance of these systems is superior to that of the system based on SP decoders by about 3dB and 2dB with Gray and SSP mappings, respectively. The performance of the OSD decoder is very close to the performance bound of the ML scheme for Gray mapping. However, the performance of the OSD decoder with the SSP mapping cannot approach that of the ML decoding. This is because the OSD for a coded modulation system with SSP mapping is very suboptimum. This phenomenon encountered with the suboptimum receivers is similarly observed for BICM-ID systems with convolutional codes [65, 67]. In BICM-ID, the system using a mapping that has a better asymptotic performance, has the worse convergence property. Although iterative demodulation/decoding is implemented, the improvement due to iterations is negligible for the system with 8-PSK modulation.

It should also be mentioned here that the iterative demodulation and decoding of LDPC coded systems with the inclusion of an interleaver between the encoder and the modulator was also investigated. However, it was observed that such additional interleaver does not

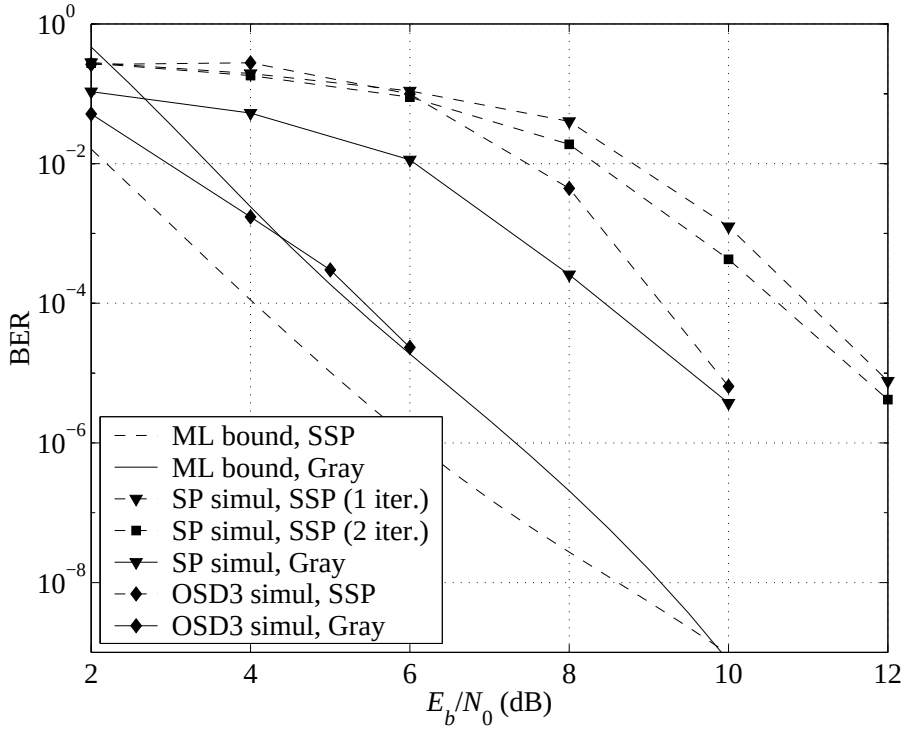


Figure 3.7 Comparison of the ML bound and the performances of the OSD and SP decoders: regular (3,6) LDPC code of length 72 and rate 1/2, 8-PSK modulation, AWGN channel.

influence the error performance of the systems. This fact can be explained by the "interleaver inside" characteristic of the LDPC codes.

Finally, the union bound and simulation results of the iterative demodulation and decoding are also investigated for a medium-length and high-rate LDPC code. Specifically, a (3,6) regular LDPC code of length 495 bits and an information block of 433 bits constructed in [85] is considered. The terms with the smallest Hamming distances of this code are provided in [50, 85] and also listed in Appendix C. As can be seen from Fig. 3.8, the error performance of this suboptimum decoding scheme is closer to the error performance of the ML decoding scheme for this longer code. Here, the union bound in Fig. 3.8 is considered to be tight only above the cutoff-rate point of 6.97dB. Thus, the ML bound can be used to accurately predict the performance of the sum-product receiver in the high-signal-noise ratio range. The bound computed by the numerical method is much faster and more efficient than performing computer simulation. Therefore, the parameters of the system can be conveniently and appropriately chosen by using the bound. Note that, the OSD decoders

are not applicable at this length of the LPDC code due to their very high computational complexity.

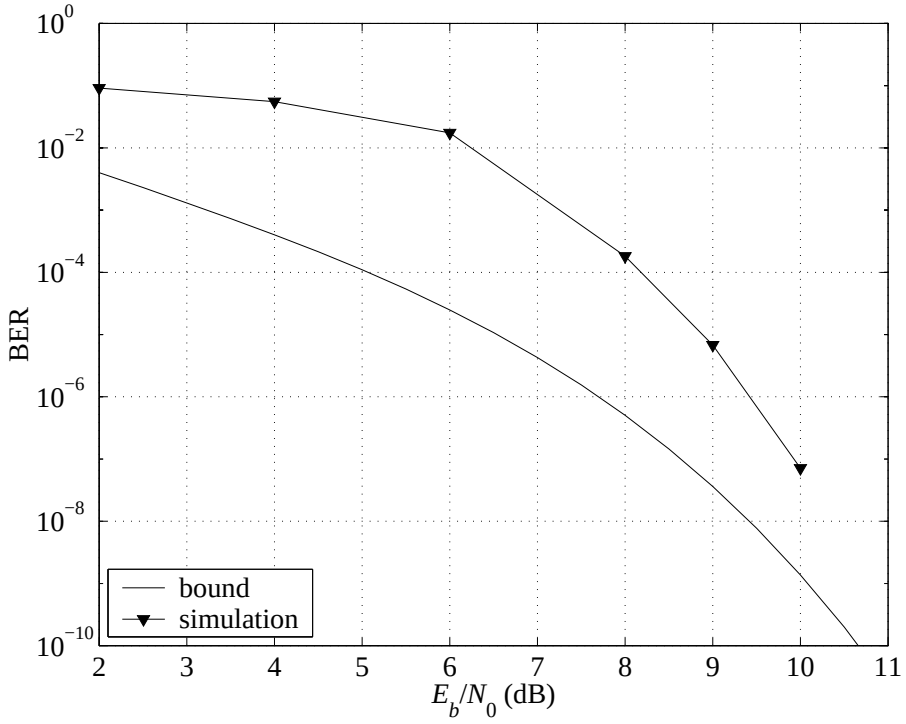


Figure 3.8 The bound of the ML decoding and simulation result of the sum-product decoding over an AWGN channel: A regular $(3, 6)$ LDPC code of rate 433/495 and length 495 bits, 8-PSK constellation, Gray mappings.

3.4.2 Fading Channels

In this subsection, the union bound for a Rayleigh fading channel with perfect CSI is investigated. Again, the very short LDPC code of length 24 bits is chosen to verify the accuracy of the bound. Fig. 3.9 shows that the simulation result of the ML decoding is also consistent with the union bound derived for the Rayleigh fading channel.

Next, computer simulation is also implemented for the iterative demodulation and decoding systems based on the sum-product decoding and OSD. The simulation results together with the ML bounds are shown in Fig. 3.10. Observe that the simulation results show a gain of 0.8 dB due to the iterations between the demodulation and the sum-product decoding for the system with SSP mapping. The error performance of the iterative sum-product scheme with Gray mapping is still superior to that with SSP mapping. The perfor-

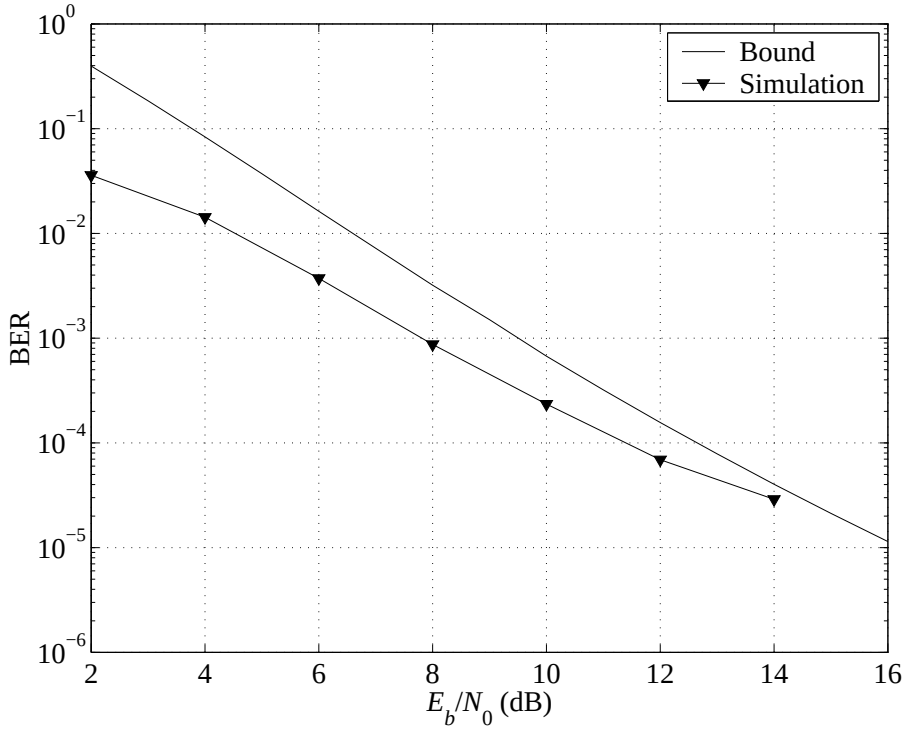


Figure 3.9 The bound and simulation result for the ML decoding over a Rayleigh fading channel: A regular $(3, 6)$ LDPC code of rate 1/2 and length 24 bits, 8-PSK modulation and Gray mapping.

mance gap between systems with the two mappings is about 3dB at the BER level of 10^{-5}. At this BER level, the performance gaps between the iterative sum-product scheme and the ML scheme are about 6dB and 10dB for the systems using Gray and SSP mappings, respectively.

The iterative OSD scheme is also investigated with 8-PSK constellation and Gray and SSP mappings. The error performance of the iterative demodulation/decoding system based on the OSD decoder is better than that of the system based on the SP decoder by about 4.5 dB and 4 dB with Gray mapping and SSP mapping at the BER of 10^{-5}, respectively. The performance of the system with Gray mapping closely approaches the performance of the ML scheme. Note that, the upper bound for the Rayleigh fading channel, however, is not as tight as the bound for the AWGN channel. Particularly, the demodulation/decoding iteration improves the performance of the system with 8-PSK modulation and SSP mapping by about 1 dB over a Rayleigh fading channel.

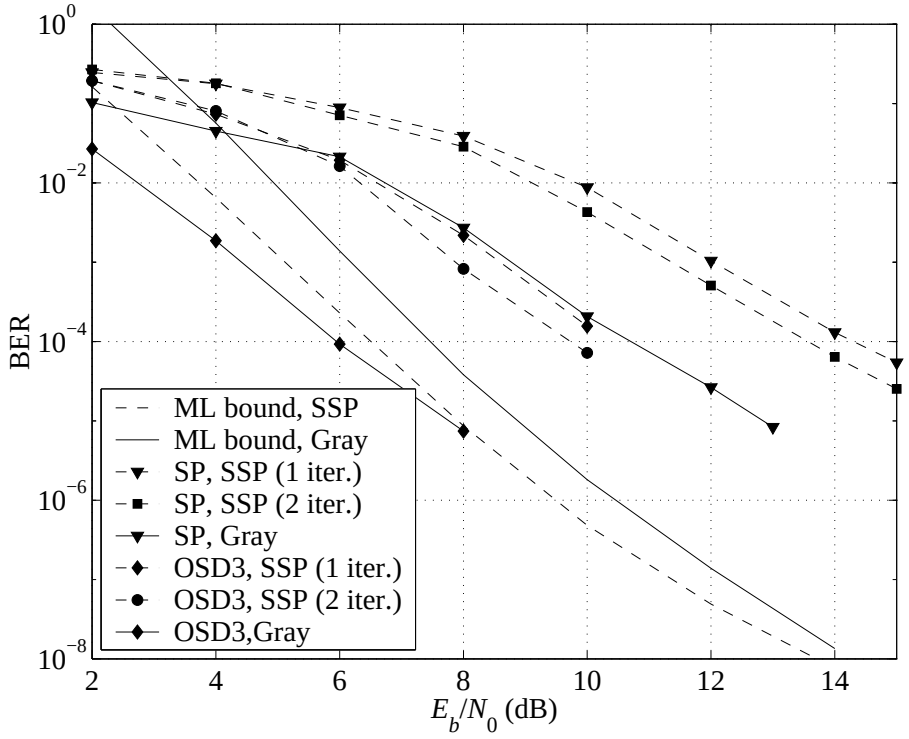


Figure 3.10 Comparison between the ML bound and simulation results of the OSD and SP decoders: Regular (3,6) LDPC code of length 72 and rate 1/2, 8-PSK modulation, a Rayleigh fading channel.

Finally, the coded modulation system using an irregular LDPC code and 16-QAM modulation is also considered. An irregular LDPC code of rate $1/2$ and length 200 bits is taken from [50]. The variable node distribution of this irregular LDPC code is $\lambda(x) = 0.31570x + 0.26758x^2 + 0.41672x^6$ and the check node distribution is $\rho(x) = 0.4381x^5 + 0.5619x^6$. Note that, these parameters of the irregular LDPC codes are defined in Section 2.2. The smallest Hamming-distance terms of this LDPC code are provided in [50]. The union bound of the ML decoding and the error performance of the iterative demodulation and decoding are shown in Fig. 3.11. At the BER of 10^{-5}, the performance gap between the iterative scheme and the ML scheme is about 2 dB for the system using the Gray mapping. At this code length, the computational complexity of the OSD algorithm is still very high for the simulation purpose.

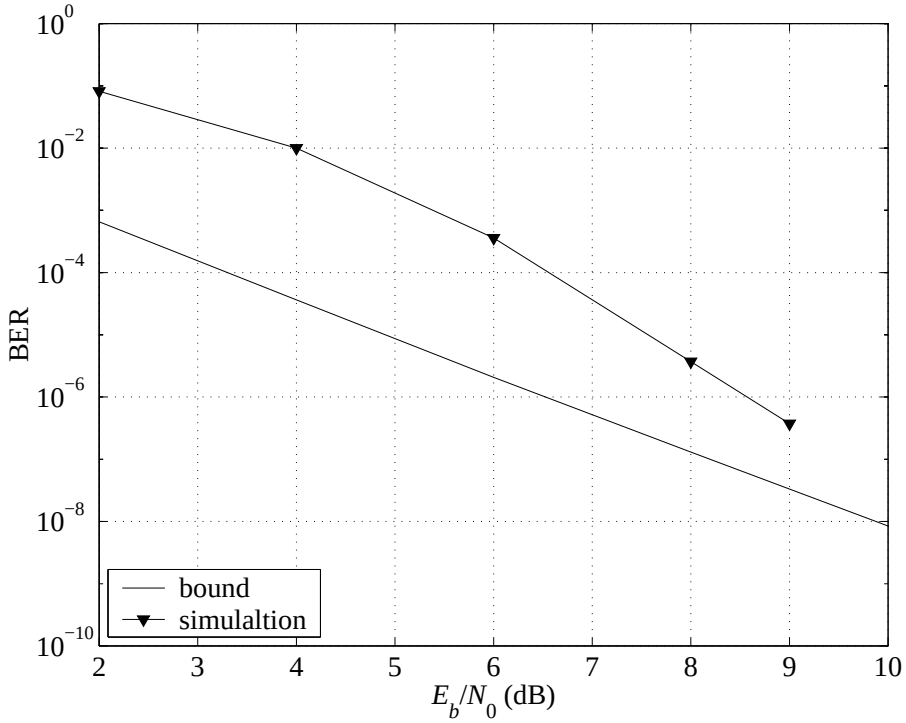


Figure 3.11 The bound of the ML decoding and simulation result of the sum-product decoding over a Rayleigh fading channel with perfect CSI: Irregular LDPC code of rate 1/2 and length 200 bits, 16-QAM modulation and Gray mapping.

3.5 Conclusions

In this chapter, the union bounds for the bit error probabilities of LDPC coded modulation systems with ML decoding were derived for both AWGN and Rayleigh fading channels. The tightness of the bounds was verified by computer simulation of the systems employing very short LDPC codes. Simulation of the iterative demodulation and decoding schemes based on the sum-product decoding and OSD were also carried out to compare with the performance of the ML receiver. LDPC coded modulation systems employing different constellations, mappings and LDPC codes were also studied. It was found that when the short LDPC codes and Gray mapping are employed, the error performance of the system using the OSD can be very close to the error performance of the ML scheme. Thus, the OSD is a good candidate for the systems using short codes. The derived ML bound can be used to benchmark the error performance of these practical systems. For medium-length codes, the performance of the sum-product decoding approaches the ML bound at very low

bit error rates or very high signal-to-noise ratios. Therefore, the bound can also be useful to estimate the error performance of communications systems that require low delay and very low BER, such as in the backbone communications.

4. Multilevel LDPC Coded Modulation

This chapter considers another bandwidth-efficient coded modulation scheme, called multilevel coded modulation. In particular, the system under consideration is a *generalized* multilevel coded modulation with multistage decoding, where finite-length LDPC codes are used as component codes and where the coded bits of each component code are mapped to more than one labelling bits of the constellation symbols. The performance bound on the bit error probability for each decoding stage is derived based on the ML criterion. These ML bounds, obtained for both AWGN and Rayleigh fading channels, are applicable for any code rate, constellation and mapping. As examples, performance of the systems using two component codes, non-uniform 16-QAM constellations and different mappings is presented. These systems are appropriate for providing unequal error protection of different data streams. Error performance of the ordered statistic decoding and sum-product decoding of LDPC codes is also investigated and compared to the derived bound.

4.1 Introduction

As discussed in the previous chapter, the invention of trellis coded modulation (TCM) [60, 61] proved the advantage of combining coding and modulation. However, TCM can only be used with convolutional codes, and it cannot be applied to block codes such as LDPC codes. Another scheme for bandwidth-efficient coded modulation is bit-interleaved coded modulation (BICM) [16, 17], which is a natural approach for pseudo-random codes such as LDPC codes. The performance of finite-length LDPC coded systems, i.e., the BICM scheme based on LDPC codes, was investigated in detail in the previous chapter.

In this chapter, another coded modulation scheme, known as multilevel coded modula-

tion (MLC) is investigated. This MLC scheme was proposed by Imai and Hirakawa [19] independently and simultaneously to TCM. This scheme is based on the combination of many component codes and one signal constellation. In this system, the data stream is divided into substreams. These substreams are separately encoded before they are mapped to the same signal constellation. Each substream, or level, in this system can use a convolutional code, a block code such as LDPC code or is even not coded. Here, the main idea of multilevel coding is to protect each bit position in the modulating symbols, called a level, by a separate binary component code. At the receiver, each code is decoded individually starting from the lowest level and it is referred to as a decoding stage. The input of the decoder in each stage takes into account the decision of the previous stage. This procedure is called multi-stage decoding. Compared to TCM, MLC is more flexible because any code, e.g. block codes, convolutional codes or concatenated codes, with certain code rates, can be used as component codes. In multi-stage decoding, the code constraints at higher levels are not taken into account in decoding the lower level. Due to this simplification, the performance of the MLC scheme is degraded compared to the maximum likelihood decoding [86]. This disadvantage can be alleviated by choosing individual code rates properly.

As mentioned above, the MLC systems often employ multistage decoding in which the decoded bits from the lower-level decoding stages are also used to decode bits in the higher levels. Therefore, the error probability of the lower-level bits should be significantly smaller than that of the higher levels in order to minimize error propagation. This means that the component codes employed for the lower levels are often more powerful than the codes used in the higher levels. This fact also gives multilevel coded modulation the capability of unequal error protection. Furthermore, even when the same component code is used in all levels, unequal error protection can still be achieved by the use of non-uniform constellation and appropriate signal mapping. Coded modulation systems with unequal error protection capability are attractive in applications that require different qualities of service (i. e., bit error rate (BER) performance) for different classes of information (such as multimedia applications that include voice and video data) [87].

For a *conventional* multilevel coded modulation, the coded bits of each component code are mapped to only one particular position of the constellation labels. When the constellation size is large, the number of required component codes is thus high, making the system less flexible and more complicated. Systems that allow coded bits from one component code to be mapped to more than one position of the constellation labels were first investigated in [87] with two convolutional codes used as component codes. Error performance of this *generalized* multilevel coded modulation is only studied by simulation in [87] and no analytical results for the bit error probabilities (BEPs) are provided.

Multilevel LDPC coded modulation has also been investigated in [88], but with infinite-length LDPC component codes. The authors of [88] optimize the parameters of very long LDPC codes by the density evolution technique to achieve the capacity of an additive white Gaussian noise (AWGN) channel. However, this code length is impractical for systems that require a short delay.

In this chapter, the *generalized* multilevel LDPC coded modulation is investigated with finite-length LDPC codes. The union bounds of the bit error probabilities is obtained for the *generalized* multilevel coded systems that are built from LDPC component codes and decoded by multistage decoding. Compared to the ML bound in the previous chapter, the derivation of the bound in this chapter shares some similarity in the steps of derivation. Specifically, the ML bounds are also computed based on the Euclidean distance spectra corresponding to the decoding stages. These Euclidean distance spectra are derived from the Hamming distance spectra of the LDPC codes and the Euclidian distance profile of the constellation. However, due to the different structure of the multilevel coded modulation scheme, the Euclidian distance profile of the constellation is different from that of the scheme in the previous chapter. For decoding levels, these Euclidian distance profiles of the constellation for each level are different and strongly depend on the mapping. Especially, the mapping can play an important role in the distribution of total transmitted power among the levels. Therefore, the mapping and parameters of the constellation can be used to adjust the error performance of each level.

It should also be mentioned here that the union bound for the *conventional* multilevel

coded modulation using convolutional codes is considered in [89] where the error propagation from the lower stages to the higher stages is also taken into account. Due to the effect of the error propagation, the bounds presented in [89] are very loose. In contrast, the *generalized* multilevel coded modulation under consideration concentrates on the scenario that the bit error probability of the lower stage is much smaller than that of the higher stage and error propagation can be practically ignored. This is also a common assumption in performance analysis of the multistage decoding [90, 91] and can be met in our systems by properly choosing the parameters of the component codes and/or the constellation and mapping.

Performance of specific MLC systems using two component LDPC codes, uniform and non-uniform 16-QAM constellations and different mappings is presented in detail in this chapter. However, the framework of analysis can be generalized to other MLC systems based on LDPC codes or other pseudo-random codes such as Turbo codes or Turbo product codes. Two kinds of LDPC decoding, namely the order-statistics decoding (OSD) and the sum-product decoding, are investigated by computer simulation. The algorithms of the OSD and the sum-product decoding were presented in Chapter 3 and Chapter 2, respectively. The performance of the OSD algorithm is close to the ML performance for short LDPC codes. Hence, simulation results of this OSD algorithm are useful to verify the ML bound. For practical MLC systems with medium-length LDPC codes, the performance of the system using the sum-product decoding is compared to the ML bound.

4.2 System Model

Consider a *generalized* multilevel coded system, illustrated in Fig. 4.1, which employs two component LDPC codes, an M-ary constellation $\mathcal{X}$ ($M = 2^q$) and the multistage decoding. Let $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$ denote the first and second component codes, whose code rates are $R^{(1)}$ and $R^{(2)}$; and code lengths are $N^{(1)}$ and $N^{(2)}$, respectively. Generally, the lengths $N^{(1)}$ and $N^{(2)}$ might not be equal. The encoded bit stream of the first component code is mapped to the first q_1 bits of the labels of the constellation and the second encoded bit stream is mapped to the last q_2 bits of the labels, where $q_1 + q_2 = q$. The mapping rule

is denoted by $\mu : \{0,1\}^q \rightarrow \mathcal{X}$. The transmitted sequence $\mathbf{x}$ is made up of symbols from $\mathcal{X}$. For each codeword $\mathbf{c}_m^{(1)}$ of code $\mathcal{C}^{(1)}$, there is a corresponding transmitted sequence $\mathbf{x}$ of length $N^{(1)}/q_1$ symbols that contains $N^{(1)}$ coded bits of codeword $\mathbf{c}_m^{(1)}$ and $N^{(1)}q_2/q_1$ coded bits of the second component code $\mathcal{C}^{(2)}$. These coded bits of component code $\mathcal{C}^{(2)}$ can belong to several codewords of $\mathcal{C}^{(2)}$, depending on the length $N^{(2)}$. Similarly, each codeword $\mathbf{c}_m^{(2)}$ of code $\mathcal{C}^{(2)}$ corresponds to a transmitted sequence $\mathbf{x}$ of length $N^{(2)}/q_2$. As before, the received signal at the time index k over a fading channel can be presented as follows:

$$y_k = h_k x_k + w_k \tag{4.1}$$

where the symbols x_k, y_k, h_k and w_k have the same meaning as in Chapter 1.

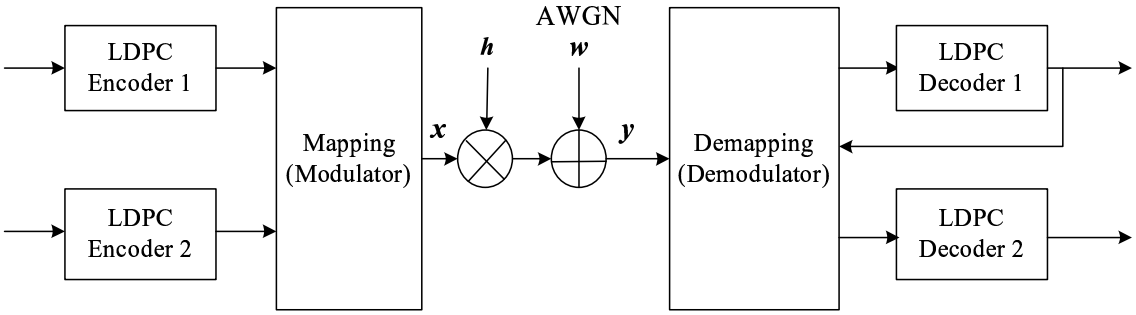


Figure 4.1 Block diagram of *generalized* multilevel LDPC coded modulation with multistage decoding.

The system employs multistage decoding, which includes two stages. The first stage decodes the code $\mathcal{C}^{(1)}$ based on the soft-outputs of the demapping block. The demapping procedure is carried out by assuming that the last q_2 bits of one symbol label are equally likely. This is because these q_2 bits of the mapping labels belong to code $\mathcal{C}^{(2)}$, which has not been decoded. The soft-input decoder of an LDPC code can be the ordered-statistic decoder (OSD) [18] or the sum-product (SP) decoder [5, 6].

When the channel state information is perfectly estimated, the likelihood ratios of coded bits (or the soft-outputs of the demapping block for Stage 1) corresponding to the kth received symbol can be mathematically computed as:

$$L(c_{k,i}^{(1)}|y_k) = \frac{P(c_{k,i}^{(1)} = 0|y_k)}{P(c_{k,i}^{(1)} = 1|y_k)} = \frac{\sum_{x_k \in \mathcal{X}_{i,0}} p(y_k|x_k)}{\sum_{x_k \in \mathcal{X}_{i,1}} p(y_k|x_k)} \tag{4.2}$$

where $c_{k,i}^{(1)}$ is the coded bit of codeword $\mathbf{c}_m^{(1)}$ at time index k, that is mapped to the ith position of the symbol label. Note that, i can be from 1 to q_1 for code $\mathcal{C}^{(1)}$. The sets $\mathcal{X}_{i,0}$ and $\mathcal{X}_{i,1}$ are the subsets of constellation $\mathcal{X}$ in which the ith bit of the symbol label is 0 and 1, respectively. Similar to the system model in Chapter 2, the conditional probability density function $p(y_k|x_k)$ is given by Equation (3.5).

For the second stage, since the code $\mathcal{C}^{(1)}$ is already decoded, the demapping procedure is different from that of the first stage. The hard decision of decoder 1, assumed to be correct, restricts the number of possible transmitted symbols in the constellation. If there is an error in the hard decision of decoder 1, this error will affect to the input of decoder 2. Because of this possible error propagation, the code $\mathcal{C}^{(1)}$ is often chosen to have a much better correction capability than that of code $\mathcal{C}^{(2)}$. The soft-output of the demapping block for Stage 2 can be computed as follows:

$$L(c_{k,i}^{(2)}|\mathcal{C}^{(1)}, y_k) = \frac{P(c_{k,i}^{(2)} = 0|\mathcal{C}^{(1)}, y_k)}{P(c_{k,i}^{(2)} = 1|\mathcal{C}^{(1)}, y_k)} = \frac{\sum_{x_k \in \mathcal{X}_{i,0,\mathcal{C}^{(1)}}} p(y_k|x_k)}{\sum_{x_k \in \mathcal{X}_{i,1,\mathcal{C}^{(1)}}} p(y_k|x_k)} \tag{4.3}$$

Here, $\mathcal{X}_{i,0,\mathcal{C}^{(1)}}$ and $\mathcal{X}_{i,1,\mathcal{C}^{(1)}}$ are subsets of symbols of $\mathcal{X}$ in which the first q_1 bits are determined by the decoder of $\mathcal{C}^{(1)}$ and the ith bit of the symbol label is 0 and 1, respectively.

The soft-outputs of the demapping blocks are used as the soft-inputs of the LDPC decoders. As mentioned in previous chapters, decoding an LDPC code can be carried out with the ordered statistic decoder (OSD) [18] or the sum-product (belief propagation) decoder [5, 6]. Which decoding algorithm to use is mainly determined by the length of the LDPC code. If the code is short, the computational complexity of OSD is practical and the performance of the OSD is close to that of the maximum likelihood (ML) decoder. However, when the code length is medium or large, the OSD becomes impractical and the sum-product decoding should be used.

4.3 Performance Bounds for an AWGN Channel

In this section, the union bounds of BEPs for Stage 1 and Stage 2 are first derived for the AWGN channel. The extension to flat fading channels is presented in Section 4.4.

4.3.1 Performance Bound for Stage 1

Consider a received sequence **y** of length $N_s^{(1)} = N^{(1)}/q_1$ that corresponds to codeword $\mathbf{c}_m^{(1)}$ of $\mathcal{C}^{(1)}$ and other coded bits of $\mathcal{C}^{(2)}$. This received sequence is due to the transmission of the symbol sequence $\mathbf{x}$. Each element of **x** is a symbol of the constellation $\mathcal{X}$. It might first appear that the total number of possible sequences is $M^{N^{(1)}/q_1} = 2^{qN^{(1)}/q_1}$. However, the number of possible sequences is much smaller because of the constraints of codes $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$. If the length of both $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$ is $N^{(1)}$, the number of possible sequences is $2^{(R^{(1)}+R^{(2)})N^{(1)}}$. Here, $(R^{(1)} + R^{(2)})N^{(1)}$ is the number of information bits carried by one symbol sequence. If the lengths of the two codes are different, the number of possible sequences is also close to $(R^{(1)} + R^{(2)})N^{(1)}$ depending on the chosen frame length. The constraints imply that demapping of the sequence $\mathbf{x}$ should yield valid codewords of $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$.

When there is no a priori information from the code $\mathcal{C}^{(2)}$ as in the case of Stage 1 decoding, the selection of the last q_2 bits of labels is carried out by not taking into account the constraint of code $\mathcal{C}^{(2)}$. Thus, the number of possible sequences is $2^{(R^{(1)}+1)N^{(1)}}$, which is bigger than that for the case when the constraints of two codes are taken into account. This also means that the Euclidean distances among possible sequences are reduced.

Based on the received sequence $\mathbf{y}$, the maximum likelihood (ML) decoding rule for Stage 1 makes the decision to a symbol sequence that is closest to $\mathbf{y}$ in terms of the Euclidean distance. Note that, this ML decision rule for Stage 1 is not equivalent to ML decoding when both component codes are taken into account simultaneously. The decoded codeword $\mathbf{c}_m^{(1)}$ of code $\mathcal{C}^{(1)}$ is obtained by demapping this symbol sequence. The error performance of the ML rule is the ultimate performance limit for any other decoding algorithm such as the sum-product or OSD decoders.

ML decoding without the constraint of code $\mathcal{C}^{(2)}$ makes an error when the received signal $\mathbf{y}$ exceeds the boundary of the decision region of the transmitted sequence $\mathbf{x}$. Thus, the conditional frame error probability is a sum of the conditional error probabilities corre-

sponding to $2^{(R^{(1)}+1)N^{(1)}}$ decision regions of all possible transmitted sequences. That is:

$$P_{s,\mathbf{x}} = \sum_{\mathbf{x}' \neq \mathbf{x}} \Pr(\mathbf{y} \in \Lambda_{\mathbf{x}'} | \mathbf{x}) \tag{4.4}$$

where $P_{s,\mathbf{x}}$ is the frame error probability when the symbol sequence $\mathbf{x}$ is transmitted, and $\Pr(\mathbf{y} \in \Lambda_{\mathbf{x}'}|\mathbf{x})$ is the probability that the received sequence $\mathbf{y}$ belongs to the decision region of $\mathbf{x}'$ given that $\mathbf{x}$ is transmitted. All possible sequences are chosen equally likely, thus the union bound of the bit error probability for Stage 1 can be written as:

$$P_e^{(1)} \leq \frac{1}{2^{(R^{(1)}+1)N^{(1)}}} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \frac{W_{\mathbf{x},\mathbf{x}'}^{(1)}}{N^{(1)}} \Pr(\mathbf{x} \rightarrow \mathbf{x}') \tag{4.5}$$

where $\Pr(\mathbf{x} \rightarrow \mathbf{x}')$ is the pairwise error probability when $\mathbf{x}$ is transmitted, but $\mathbf{x}'$ is decided at the receiver. The parameter $W_{\mathbf{x},\mathbf{x}'}^{(1)}$ is the Hamming distance of two codewords $\mathbf{c}_m^{(1)}$ and $\mathbf{c}_{m'}^{(1)}$ of $\mathcal{C}^{(1)}$ obtained by demapping the two sequences $\mathbf{x}$ and $\mathbf{x}'$, respectively. Note that, there are many pairs of sequences $\mathbf{x}$ and $\mathbf{x}'$ with $W_{\mathbf{x},\mathbf{x}'}^{(1)} = 0$, since there are many pairs that are demapped to the same codeword of $\mathcal{C}^{(1)}$. This is due to the fact that the code constraint of $\mathcal{C}^{(2)}$ is not take into account. This means that many pairwise error probabilities do not contribute to the bit error probability of Stage 1 decoding.

As discussed in Chapter 3, for LDPC codes, the error performance is averaged over the permuted code ensemble, where one code in this ensemble is a randomly permuted version of another code. The average union bound can thus be written as follows:

$$\overline{P}_e^{(1)} \leq E\left[\frac{1}{2^{(R^{(1)}+1)N^{(1)}}} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \frac{W_{\mathbf{x},\mathbf{x}'}^{(1)}}{N^{(1)}} \Pr(\mathbf{x} \rightarrow \mathbf{x}')\right] \tag{4.6}$$

Following the technique in Chapter 3, the bound can be computed more conveniently by classifying the pairwise error probabilities into groups with the same characteristics and then averaging over these groups.

The summation of two codewords $\mathbf{c}_m^{(1)}$ and $\mathbf{c}_{m'}^{(1)}$ is an error sequence $\mathbf{e}$. For a linear block code such as LDPC codes, the number of error sequences with a given weight is determined by the Hamming weight spectrum of the code $\mathcal{C}^{(1)}$. However, due to the random construction of LDPC codes, the positions of the errors are also random over the length of

a codeword. Similar to Chapter 3, error sequences of a given weight can be classified into different types of error sequences $\mathbf{n}^{(1)}$ according to the number of bit groups with the same weights. Here, the ith element of $\mathbf{n}^{(1)}$ determines the number of error bit groups of weight i. For the example of 16-QAM constellation, each group of two coded bits of $\mathcal{C}^{(1)}$ is mapped to the first two bits of the 4-bit labels of the constellation and each group of two coded bits of $\mathcal{C}^{(2)}$ is mapped to the last two bits of these 4-bit labels. The weight of a bit group of an error sequence can then be 1 or 2.

After classifying error sequences, the average union bound can be rewritten as follows:

$$\overline{P}_e^{(1)} \leq \sum_{l=l_{\min}^{(1)}}^{N^{(1)}} \sum_{n_1^{(1)}=0}^{N_s^{(1)}} \cdots \sum_{n_{q1}^{(1)}=0}^{N_s^{(1)}} \frac{l}{N^{(1)}} B\overline{f}(\mathbf{n}^{(1)}) E\left[Q\left(\frac{D_{\mathbf{n}^{(1)}}}{\sqrt{2N_0}}\right)\right] \tag{4.7}$$

where $f(\mathbf{n}^{(1)})$ is the number of error sequences of type $\mathbf{n}^{(1)}$ corresponding to a given transmitted sequence $\mathbf{x}$ and $\overline{f}(\mathbf{n}^{(1)})$ is the expected value of this function over the ensemble of code $\mathcal{C}^{(1)}$. The expected total number of error sequences of type $\mathbf{n}^{(1)}$ is $2^{R^{(1)}N^{(1)}}\overline{f}(\mathbf{n}^{(1)})$. Note that $2^{(R^{(1)}+1)N^{(1)}}$ is the number of possible sequences of $\mathbf{x}$ without considering the constraint of code $\mathcal{C}^{(2)}$. For a given transmitted sequence $\mathbf{x}$, there are $2^{N^{(1)}}-1$ sequences $\mathbf{x}'$, whose demapped sequences are only different at the bit positions of the code $\mathcal{C}^{(2)}$ and all of them correspond to only one error sequence. Furthermore, among these $2^{N^{(1)}}-1$ sequences, there is a number of sequences whose corresponding Euclidean distances are small enough compared to the smallest Euclidean distance. In (4.7), the coefficient B is the effective error multiplicity that represents this number of sequences. This coefficient B is often called the effective number of the nearest neighbors [89, 92]. Determining this coefficient for the systems under consideration is quite different compared to the coded modulation systems that employ only one component code and will be discussed later.

The function $\overline{f}(\mathbf{n}^{(1)})$ is computed as in Equations (3.15) and (3.16). Here, the parameters of Stage 1 decoding $\overline{f}(\mathbf{n}^{(1)})$, $P_{l,\mathbf{n}}^{(1)}$, $\mathcal{N}^{(1)}(l)$ and $N^{(1)}$ are substituted for $\overline{f}(\mathbf{n})$, $P_{l,\mathbf{n}}$, $\mathcal{N}(l)$ and N, respectively.

To compute the union bound by averaging over the random variable $D_{\mathbf{n}^{(1)}}$, it is assumed that all $N_s^{(1)}$ channel symbol errors are independent of each other and all the signal points

in the constellation are equally likely to be used. Then, the total distance $D_{\mathbf{n}^{(1)}}$ can be computed from the single symbol error distances $D_k^{(1)}$ that are also random variables as $(D_{\mathbf{n}^{(1)}})^2 = \sum_{k=1}^{N_s^{(1)}} \left(D_k^{(1)}\right)^2$. Thus, the union bound can be rewritten as

$$\overline{P}_e^{(1)} \leq \sum_{l=l_{\min}^{(1)}}^{N^{(1)}} \sum_{n_1^{(1)}=0}^{N_s^{(1)}} \cdots \sum_{n_{q_1}^{(1)}=0}^{N_s^{(1)}} \sum_{j=1}^{k_{\mathbf{n}^{(1)}}} \frac{l}{N^{(1)}} B\overline{f}(\mathbf{n}^{(1)}) p_{\mathbf{n}^{(1)},j} Q\left(\frac{\Delta_{\mathbf{n}^{(1)},j}}{\sqrt{2N_0}}\right) \tag{4.8}$$

where $p_{\mathbf{n}^{(1)},j} = P[D^2_{\mathbf{n}^{(1)},j} = \Delta^2_{\mathbf{n}^{(1)},j}]$, $j = 1, 2, \ldots, k_{\mathbf{n}^{(1)}}$. Here, $k_{\mathbf{n}^{(1)}}$ is the number of distinguishable Euclidean distances and each of them has the probability $p_{\mathbf{n}^{(1)},j}$.

The probability mass function (PMF) of $D_{\mathbf{n}}^{(1)}$ can be computed from the PMFs of $D_k^{(1)}$ with the assumption that all $N_s^{(1)}$ channel symbol errors are independent of each other. The PMFs of $D_k^{(1)}$ for a given error bit group can be computed from the mapping and the constellation by the assumption that all the signal points in the constellation are equally likely to be used. Then the random variable $D_k^{(1)}$ can take on one of several values $d_{i,h}^{(1)}$ with probabilities $p_{i,h}^{(1)}$, respectively. Here, i, $1 \leq i \leq q_1$, is the number of errors in the bit group of code $\mathcal{C}^{(1)}$, and h, $0 \leq h \leq h_{\max}$, is the number of possible Euclidean distances. Note that, for a given symbol in a 2-D constellation, there are 2^{q_2} Euclidean distances that correspond to each error pattern of a bit group. However, only the smallest Euclidean distance among these 2^{q_2} distances is taken into account. Other distances are considered in a latter procedure that computes the effective number of nearest neighbors. The random variable $D^2_{\mathbf{n}^{(1)}}$ can take on the following values:

$$\Delta^2_{\mathbf{n}^{(1)},j} = \sum_{i=1}^{q_1} \sum_{h=1}^{h_{\max}} n_{i,h}^{(1)} (d_{i,h}^{(1)})^2 \tag{4.9}$$

where $n_{i,h}^{(1)}$ is the number of distances $D_k^{(1)} = d_{i,h}^{(1)}$. Note that $0 \leq n_{i,h}^{(1)} \leq n_i^{(1)}$ and $\sum_{h=0}^{h_{\max}} n_{i,h}^{(1)} = n_i^{(1)}$. The probability of $D^2_{\mathbf{n}^{(1)}} = \Delta^2_{\mathbf{n}^{(1)},j}$ with a given set of $n_{i,h}^{(1)}$ is simply:

$$p_{\mathbf{n}^{(1)},j} = \prod_{i=1}^{q_1} \binom{n_i^{(1)}}{n_{i,0}^{(1)}, \ldots, n_{i,h}^{(1)}} \prod_{h=0}^{h_{\max}} {p_{i,h}^{(1)}}^{n_{i,h}^{(1)}} \tag{4.10}$$

To illustrate the procedure to calculate $\Delta_{\mathbf{n},j}$ and $p_{\mathbf{n},j}$, the example of 16-QAM system with Gray mapping in Fig. 4.2 is again considered. The random variable $D_k^{(1)}$ can take on

one of several values $d_{i,h}^{(1)}$. If there is one error in a bit group of 2 bits ($i = 1$), two error patterns 10 and 01 are possible. In Fig. 4.2, only Euclidean distances corresponding to two symbols of labels 1011 and 1100 are drawn for error pattern 10. The Euclidean distance profile of one half of 16 symbols is the same as that of the symbol with label 1011 and the Euclidean distance profile of another half of symbols is the same as that of the symbol labelled with 1100. There are 4 Euclidean distances from one symbol to four other symbols with the same error pattern of a bit group because two other bits of code $\mathcal{C}^{(2)}$ have not been determined. The smallest Euclidean distance among these four distances is d_0. For the error pattern 01, the two smallest Euclidean distances are d_0 and $3d_0$. Thus, distances $d_{1,0}^{(1)}$ and $d_{1,1}^{(1)}$ are d_0 and $3d_0$, respectively. The corresponding probabilities $p_{1,0}^{(1)}$ and $p_{1,1}^{(1)}$ are computed by dividing the number of symbols with the same smallest Euclidean distance to the total number of symbols. That yields $p_{1,0}^{(1)} = 3/4$ and $p_{1,1}^{(1)} = 1/4$. The Euclidean distances and their corresponding probabilities are computed similarly for the case of 2 errors in the bit group of code $\mathcal{C}^{(1)}$. That yields $d_{2,0}^{(1)} = 2d_0$ and $p_{2,0}^{(1)} = 1$. The Euclidean distances $d_{i,h}$ with a standard 16-QAM constellation and Gray mapping are tabulated in Table 4.1. Another mapping, called embedded mapping and shown in Fig. 4.4, is also considered. The Euclidean distances $d_{i,h}$ for this embedded mapping is also tabulated in Table 4.1 for comparison.

Now, the effective error coefficient B in (4.8) is discussed. For conventional multi-level coded modulation systems, this coefficient B is set to b^η, which is the number of possible sequences with the same Euclidean distance and the error pattern compared to the transmitted sequence [89]. Here, b is the number of neighboring symbols in the 2-D constellation and η is the number of error symbols. When the symbols in the 2-D constellations have different numbers of neighbors, the effective error coefficient B should take into account the proportion of the symbols that have the same number of neighboring symbols [92]. For example, if half the symbols have one neighboring symbol and the other half have two neighboring symbols, the effective error coefficient is computed as $B = \sum_{i=0}^{\eta}(0.5)^{\eta-i}1^{\eta-i}(0.5)^i 2^\eta = (\frac{3}{2})^\eta$. In the cases of asymmetric constellations, when the effect of other neighbors is comparable to that of the nearest neighboring symbol, the

two closest neighboring symbols are taken into account. However, these two symbols are only counted as $(1 + L)$ symbols with the likelihood ratio $L = \exp\{-E_s(D_2^2 - D_1^2)/4N_0\}$ [91]. Here, D_1 and D_2, where $D_2 > D_1$, are the Euclidean distances from the two closest symbols of the 2-D constellation and E_s is the average symbol energy. For the generalized multilevel coded modulation systems under consideration, there are more than one labeling bit used by each component code. Hence the neighboring symbols should only be counted for each smallest Euclidean distance with a given error pattern of a bit group. However, only possible symbols corresponding to this error pattern are considered. For an 16-QAM constellation and Gray mapping shown in Fig. 4.2, there are 4 possible symbols corresponding to an error pattern, but there is always one nearest symbol among these 4 symbols. Hence, the effective error coefficient B is simply set to $1^\eta = 1$ in this case. In general, this effective error coefficient B depends on the constellation and mapping.

The above procedure shows all the necessary computations to determine the union bound in (4.8) for Stage 1 decoding.

4.3.2 Performance Bound for Stage 2

The derivation of the union bound for Stage 2 is similar to that of Stage 1. First, the bit error probability of Stage 2 is bounded by the summation of pairwise error probabilities. Then, the average union bound is computed over the ensemble of permuted versions of code $\mathcal{C}^{(2)}$. The pairwise error probabilities are then classified into groups with the same characteristics. Averaging over these groups of pairwise error probabilities yields a closed-form expression of the union bound on the BEP based on the Hamming distance spectrum of the code and the Euclidean distance profile of the constellation. It is important to note, however, that code $\mathcal{C}^{(1)}$ has been decoded in this case. The decoded bits of $\mathcal{C}^{(1)}$ are assumed to be correct and yield a different Euclidean distance profile of the constellation for Stage 2 decoding. The detailed steps to compute the union bound for Stage 2 decoding are provided in Appendix D.

Compared to the union bound for Stage 1, the union bound for Stage 2 does not include the effective error coefficient B. Another difference is the PMF of the single symbol error

distance $D_k^{(2)}$. The possible distances $d_{1,h}^{(2)}$ and $d_{2,h}^{(2)}$ for $D_k^{(2)}$ are illustrated in Fig. 4.3 for 16-QAM with Gray mapping. Recall that, when the first two label bits of code $\mathcal{C}^{(1)}$ are determined, the 16-QAM constellation simplifies to a constellation with 4 symbols. Due to the symmetry of 16-QAM and Gray mapping, this simplified constellation is 4-PAM with Gray mapping for every possible coded bits of $\mathcal{C}^{(1)}$. Therefore, the number of distances $d_{i,h}^{(2)}$ is small in this case. If there is one error in the bit group of code $\mathcal{C}^{(2)}$, $D_k^{(2)}$ can take on two values $d_{1,1}^{(2)} = d_0$ and $d_{1,2}^{(2)} = 3d_0$ with probabilities $p_{1,1}^{(2)} = 3/4$ and $p_{1,2}^{(2)} = 1/4$, respectively. If there are two errors in the bit group, $D_k^{(2)}$ can only be $d_{2,1}^{(2)} = 2d_0$ with $p_{2,1}^{(2)} = 1$. In Fig. 4.3, one half of symbols in 16-QAM constellation with Gray mapping have the same Euclidean distance profile, and for simplicity, the distance profiles of only two symbols are indicated. The Euclidean distances $d_{i,h}$ for both stages with the standard 16-QAM constellation and Gray mapping and embedded mapping are tabulated in Table 4.1 for comparison.

Table 4.1 Euclidean distances and the corresponding probabilities for the standard 16-QAM constellation.

		Gray mapping Euclidean distance index (h)				Embedded mapping Euclidean distance index (h)					
		1		2		1		2		3	
Stage	i	d	p	d	p	d	p	d	p	d	p
1	1	d_0	$3/4$	$3d_0$	$1/4$	d_0	$1/2$	$2d_0$	$1/2$	0	0
	2	$2d_0$	1	0	0	$\sqrt{2}d_0$	$1/4$	$\sqrt{5}d_0$	$1/2$	$2\sqrt{2}d_0$	$1/4$
2	1	d_0	$3/4$	$3d_0$	$1/4$	d_0	1	0	0	0	0
	2	$2d_0$	1	0	0	$\sqrt{2}d_0$	1	0	0	0	0

4.3.3 Constellations and Mappings

The derivations in the previous subsections can be applied to arbitrary constellations and mappings. The 16-QAM with Gray mapping was used as an example. Gray mapping is well known to be the best mapping for uncoded data. Recent research results show that Gray mapping is, however, not the best mapping for bit-interleaved convolutionally coded modulation with iterative decoding (BICM-ID) [67]. Nevertheless, Gray mapping is still found to be the best mapping for BICM systems with finite-length block codes such as LDPC codes [93]. The criterion to evaluate different mappings in these systems is the bit

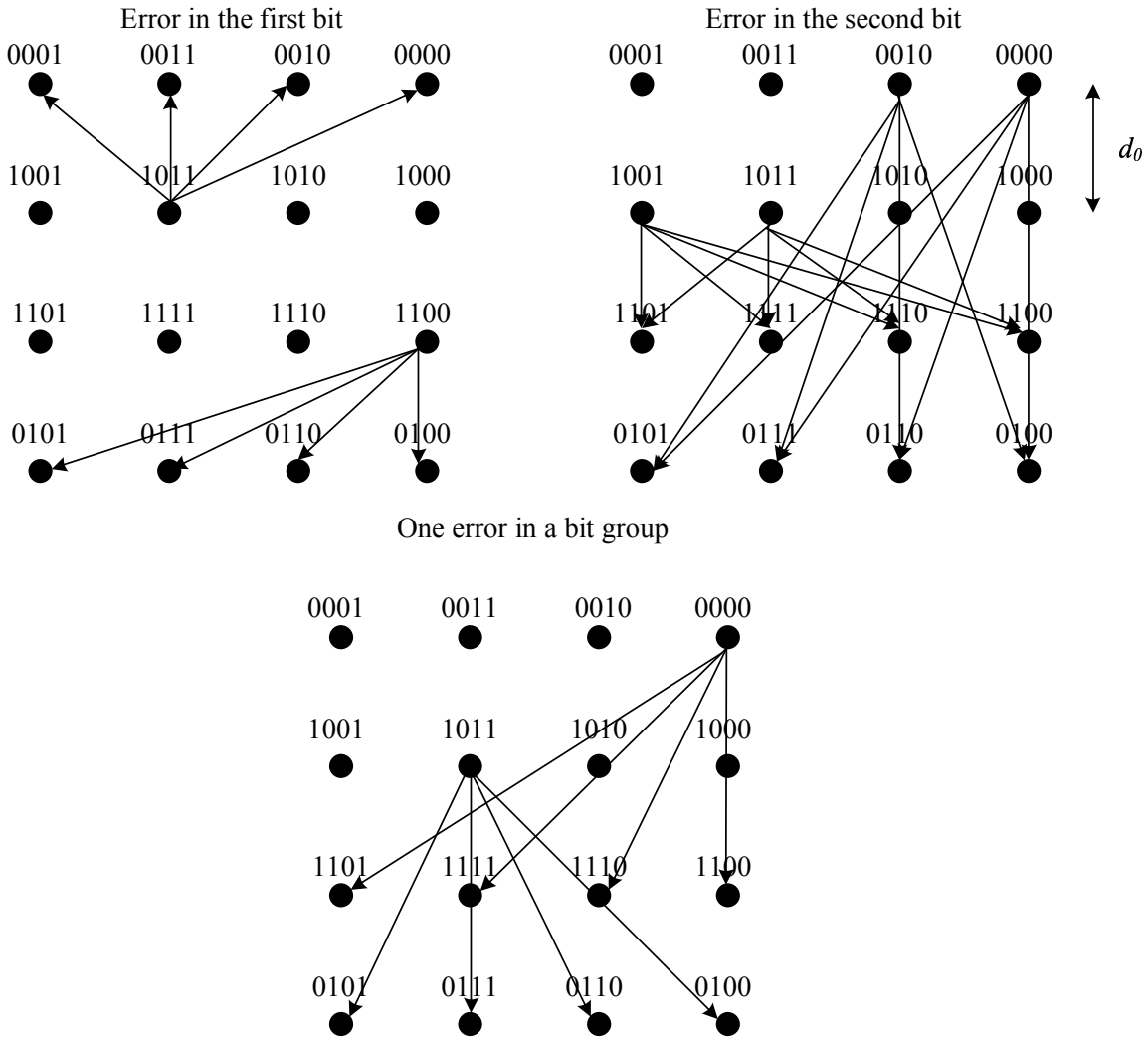


Figure 4.2 The Euclidean distances of 16-QAM constellation with Gray mapping for Stage 1.

error probability because only one component code is employed.

For our multilevel coded systems, there is no universal criterion to measure the performance. The average BEP, computed as $P_e = \left[R^{(1)}q_1P_e^{(1)} + R^{(2)}q_2P_e^{(2)}\right]/(R^{(1)}q_1+R^{(2)}q_2)$, is dominated by the worse BEP among the BEPs of different stages. Different mappings provide various and flexible modes of operations for multilevel coded systems. Each mode can give different BEPs to serve applications with different required qualities of service. However, since the number of available mappings is small, the flexibility in adjusting the

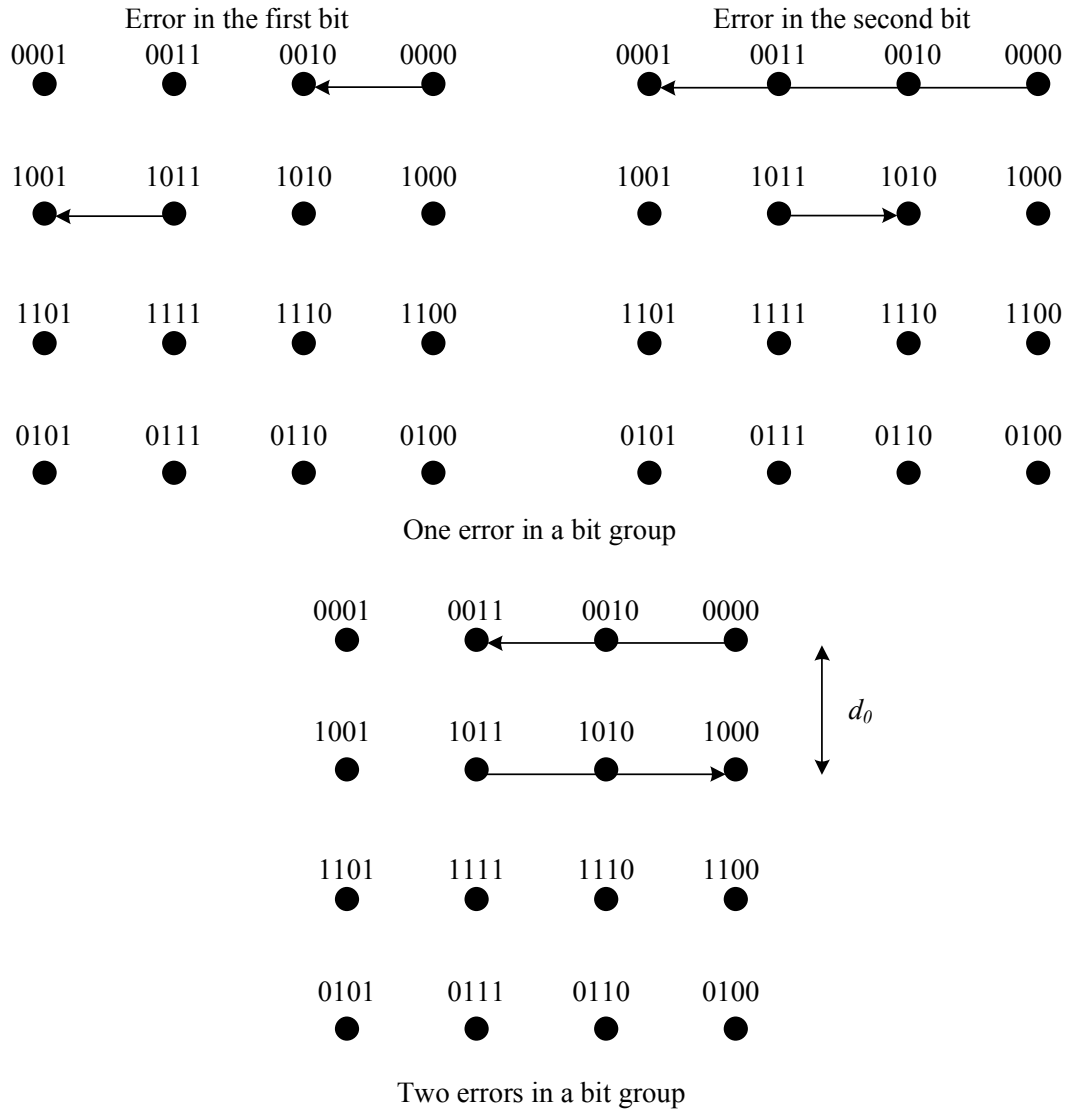


Figure 4.3 The Euclidean distances of 16-QAM constellation with Gray mapping for Stage 2.

BEPs of different stages can be restricted. Though the BEPs can also be adjusted by changing the error control coding, such a method is often complicated and expensive due to the need of having many encoders and decoders. Fortunately, for some mappings, the constellation can also be adjusted to become a nonuniform constellation or asymmetric constellation in order to properly change the BEP of each stage. The distance parameters of a constellation are continuous and they are difficult to optimize by simulation. Here, the union bounds become very useful to chose the parameters of the constellation.

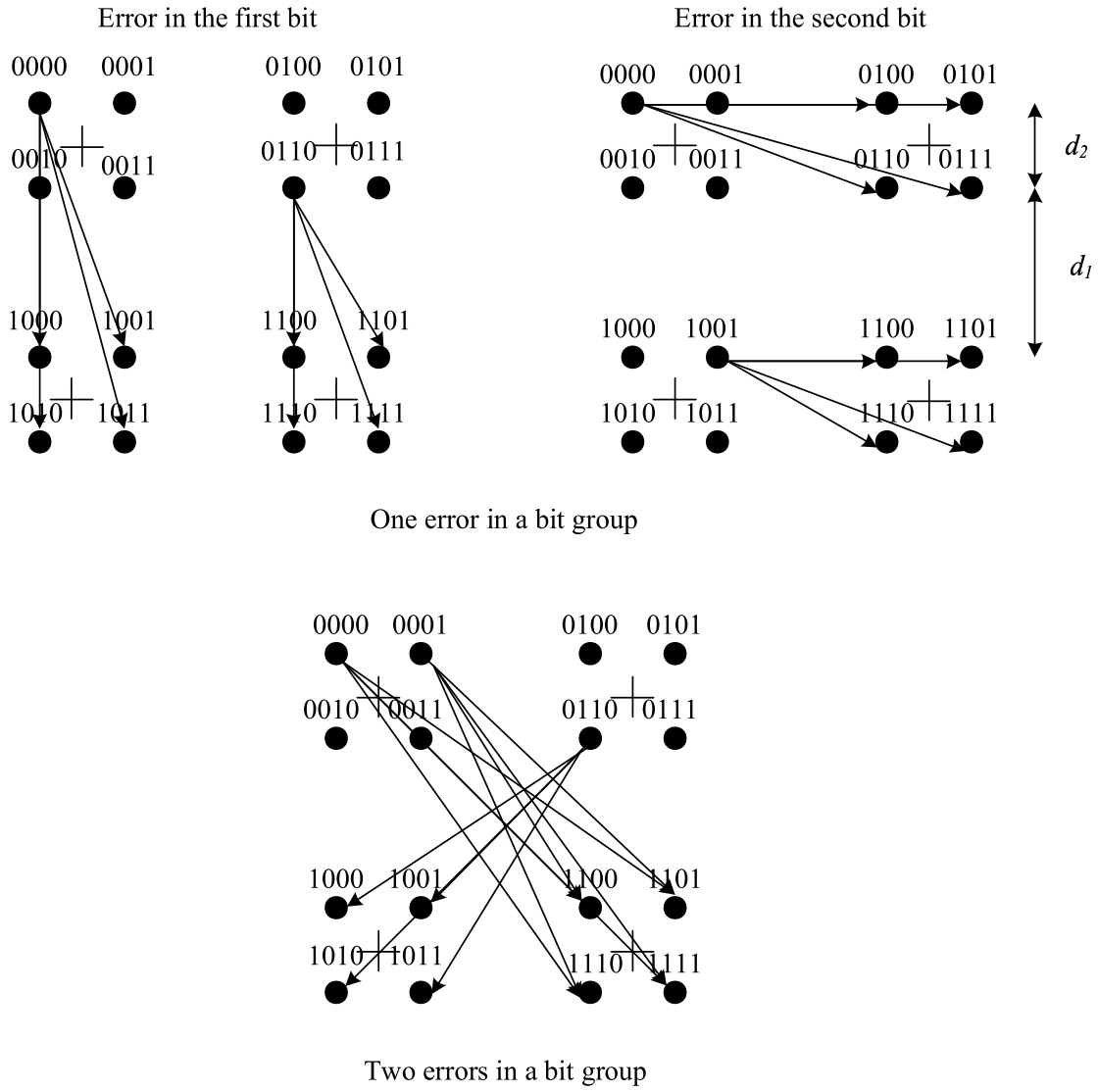


Figure 4.4 The Euclidean distances of nonuniform 16-QAM constellation with embedded mapping for Stage 1.

As mentioned earlier, another mapping, called embedded mapping, of 16-QAM constellation is also considered. When this embedded mapping is employed, a nonuniform 16-QAM constellation can be used. Figs. 4.4 and 4.5 show the nonuniform 16-QAM constellation with the embedded mapping. The minimum distance among the blocks of 4 symbols in the four quadrants of the 2-D space is d_1 and the minimum distance within these blocks is d_2. This mapping is called embedded mapping because 4 symbols in each block can be considered as a QPSK constellations for Stage 2 and these QPSK constellations are embedded to a virtual bigger QPSK constellation for Stage 1. This virtual QPSK

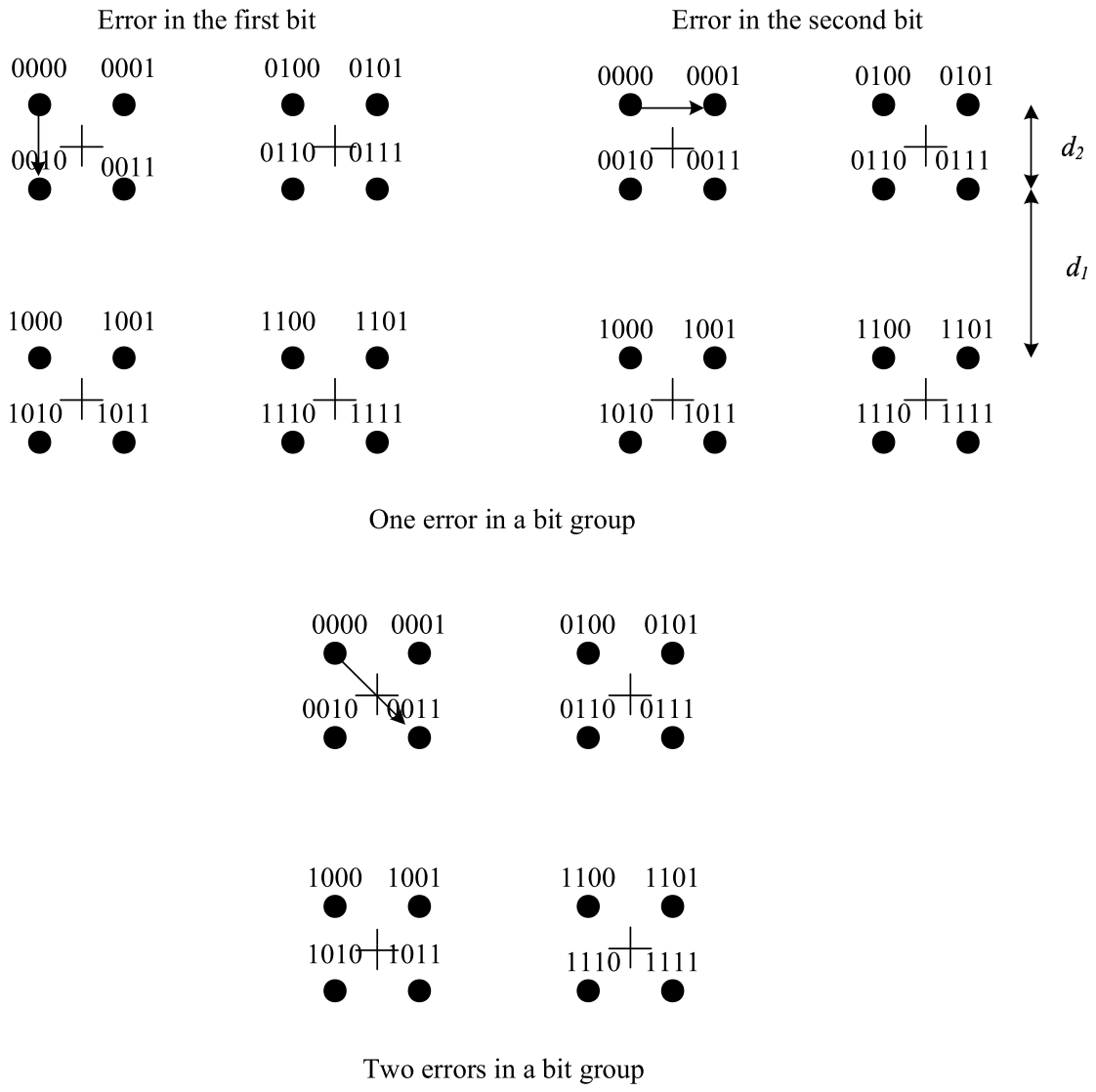


Figure 4.5 The Euclidean distances of nonuniform 16-QAM constellation with embedded mapping for Stage 2.

constellation is indicated by the four crosses in Figs. 4.4 and 4.5. When the average energy E_s is given, this nonuniform 16-QAM constellation is determined by one parameter $\rho = d_1/d_2$. The distances d_1 and d_2 are given by $d_1 = 2\rho\sqrt{E_s}/\sqrt{\rho^2 + (\rho+2)^2}$ and $d_1 = 2\sqrt{E_s}/\sqrt{\rho^2 + (\rho+2)^2}$. Note that, when the distance d_1 increases, the distance d_2 decreases. It can be intuitively predicted that when d_1 increases and d_2 decreases, error performance of Stage 1 improves and error performance of Stage 2 deteriorates. This fact can also be recognized by observing the Euclidean distances corresponding to the error patterns in Table 4.2. Note that, when the ratio d_1/d_2 equals one, the nonuniform 16-QAM

constellation becomes the standard 16-QAM constellation.

Table 4.2 Euclidean distances and the corresponding probabilities for nonuniform 16-QAM constellation with embedded mapping.

		Euclidean distance index (h)					
		1		2		3	
Stage	i	d	p	d	p	d	p
1	1	d_1	$1/2$	d_2	$1/2$	0	0
	2	$\sqrt{2}d_1$	$1/4$	$\sqrt{d_1^2+(d_1+d_2)^2}$	$1/2$	$\sqrt{2}(d_1+d_2)$	$1/4$
2	1	d_2	1	0	0	0	0
	2	$\sqrt{2}d_2$	1	0	0	0	0

4.4 Performance Bounds for Fading Channels

As pointed out in Chapter 3, the union bound of the BEP for a fading channel are different from that for an AWGN channel due to the difference in computing the pairwise error probability (PEP). The derivation of the exact PEP for a fading channel is obtained by averaging the Gaussian probability integral over the PDFs of the fading coefficients [82] and it is given by Equation (3.30)

When one substitutes the above PEP into (4.8) and (D.4) in the place of the Q-function, the union bounds on the BEPs for Stage 1 and Stage 2 decoding are given as follows:

$$\overline{P}_e^{(1)} \leq \sum_{l=l_{\min}^{(1)}}^{N^{(1)}} \sum_{n_1^{(1)}=0}^{N_s^{(1)}} \cdots \sum_{n_{q_1}^{(1)}=0}^{N_s^{(1)}} \sum_{j=1}^{k_{\mathbf{n}^{(1)}}} \frac{l}{N^{(1)}} B\overline{f}(\mathbf{n}^{(1)}) p_{\mathbf{n}^{(1)},j} \left[\frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{|\omega|} \frac{1}{1+\frac{E_s d^2_{\mathbf{n}^{(1)},j,i}}{4N_0 \sin^2\theta}} d\theta \right] \tag{4.11}$$

and

$$\overline{P}_e^{(2)} \leq \sum_{l=l_{\min}^{(2)}}^{N^{(2)}} \sum_{n_1^{(2)}=0}^{N_s^{(2)}} \cdots \sum_{n_{q_2}^{(2)}=0}^{N_s^{(2)}} \sum_{j=1}^{k_{\mathbf{n}^{(2)}}} \frac{l}{N^{(2)}} \overline{f}(\mathbf{n}^{(2)}) p_{\mathbf{n}^{(2)},j} \left[\frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{|\omega|} \frac{1}{1+\frac{E_s d^2_{\mathbf{n}^{(2)},j,i}}{4N_0 \sin^2\theta}} d\theta \right] \tag{4.12}$$

where ω is the set of all k for which there is a symbol error. Note that the size of the set ω is precisely the number of channel symbols in error, i.e., $|\omega| = \eta = \sum_i n_i$. The distances $d_{\mathbf{n}^{(1)},j,i}$ and $d_{\mathbf{n}^{(2)},j,i}$ are the Euclidean distances of the 2-D constellation that contribute to the Euclidean distance of the symbol sequence. Thus these distances correspond to the distances $d_{i,h}^{(1)}$ and $d_{i,h}^{(2)}$ in (4.9) and (D.5), respectively. Finally, to reduce the computational

complexity of the union bounds when the number of terms in the summation is large, the pairwise error probability can be bounded as in (3.32) [84].

4.5 Illustrative Results

This section presents analytical results based on the bounds obtained in the previous sections. Simulation results of the MLC system using the OSD are also provided to confirm the analytical results. The simulation of the MLC system employing the sum-product decoding also shows how the performance of the sum-product decoding converges to the ML bound.

4.5.1 AWGN Channels

First, short LDPC codes are chosen in order to verify the tightness of the union bounds. Two LDPC codes are selected, a (3,6) regular code and an irregular code each each with the same length of 72 bits and code rates of $1/2$ and $1/3$, corresponding to 36 and 24 bits of information blocks, respectively. The rate-$1/2$ code is used for the first level, while the rate-$1/3$ code is used for the second level. With this short code length, the Hamming distance spectra of these codes are exactly determined by exhaustive search and listed in Appendix C. Moreover, with this code length, the ordered-statistic decoding (OSD) can be used. Here, OSDs are implemented based on ordering the likelihood ratios of the demapping bits as discussed in Chapter 3.

The union bounds and simulation results over an AWGN channel are presented in Fig. 4.6 for the systems using the standard 16-QAM with Gray and embedded mappings. Note that for multilevel coded modulation with many component codes, the signal to noise ratio (SNR) is usually expressed as E_s/N_0 as opposed to E_b/N_0 in the coded modulation with only one component code. The reason is that the average symbol energy cannot be clearly divided among component codes. Observe that the bounds for Gray mapping are very tight over a wide range of SNR for both Stage 1 and Stage 2. The bound for Stage 1 in the case of embedded mapping is tight, however, there is a small gap of about 0.5 dB at the BER of 10^{-5} between the bound for Stage 2 and simulation results.

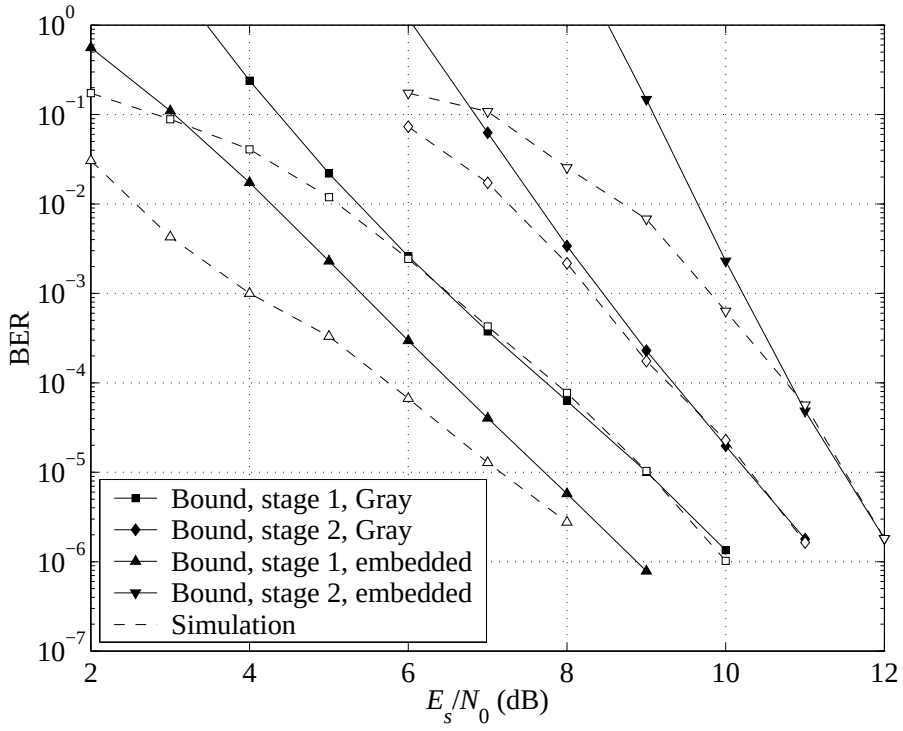


Figure 4.6 Bounds and simulation results with OSD decoding over an AWGN channel: $(3, 6)$-regular LDPC codes of rates $1/2$ and $1/3$, a length of 72 bits, standard 16-QAM constellation, Gray and embedded mappings.

The fact that the tightness of the bounds depends on the employed mapping can be explained as follows. The union bound is a sum of PEPs that are always larger than the error probabilities computed by the decision region of the codewords. When the PEPs are computed over heavily overlapped regions, the bound becomes loose. A close examination of the Euclidean distances of 16-QAM constellation reveals that, with Gray mapping, for many symbols in the constellation, their two nearest symbols corresponding to the cases of 1 bit error in a bit group (which mainly contribute to the union bound) are opposite in direction. On the other hand, with embedded mapping, these two nearest symbols are in orthogonal direction. It is conceivable that orthogonal directions of the nearest symbols lead to larger intersections of the regions in computing the PEPs compared to the opposite directions. Therefore it makes the union bounds generally looser for embedded mapping compared to Gray mapping. For example, in Fig. 4.2, the two nearest neighboring symbols of 1011 are 0011 and 1111, which correspond to one error in a bit group. These two symbols are in opposite directions from the symbol 1011. In contrast, the symbol 0011 in Fig. 4.4

has two nearest neighboring symbols of 1001 and 0110. The directions of these symbols from 1011 are orthogonal.

It can also be observed from Fig. 4.6 that for Gray mapping the performance of Stage 1 is superior to that of Stage 2 by about 1 dB at the BER of 10^{-5}. Simulation indicates that this performance gap is large enough to ignore the error propagation from Stage 1 to Stage 2. This performance gap is solely due to the different Hamming distance spectra of the chosen component codes. Note that the performance gap is not caused by the Euclidean distance distribution in the constellation, because the two first bits and the two last bits of the Gray labelling are symmetric for 16-QAM constellation.

When the embedded mapping is employed, the performance of Stage 1 improves by about 1 dB at the BER of 10^{-5} compared to that of Gray mapping. This is due to the better minimum Euclidean distances in the constellation as can be seen from Table 4.1. However, the performance of Stage 2 with embedded mapping is inferior to that of Gray mapping by about 1 dB at the BER of 10^{-5}. This observation shows that the average symbol energy is redistributed between the 2 bit streams by the mapping rule.

Recall that, for embedded mapping, when Stage 1 is perfectly decoded, the demapping procedure for Stage 2 is only carried out with 4 symbols in one quadrant of the 2-D signal space. Hence, this mapping is flexible to adjust the performance of multilevel coded modulation. Fig. 4.7 shows the bounds and simulation results for nonuniform 16-QAM constellation with the embedded mapping. The parameters d_1 and d_2 of the constellation are also indicated in the figure. Observe from Fig. 4.7 that the performance of Stage 1 improves and the performance of Stage 2 deteriorates when the ratio d_1/d_2 increases. When d_1/d_2 is set to 2, the performance of Stage 1 is improved by about 1 dB and the performance of Stage 2 degrades by about 1.5 dB compared to the standard 16-QAM. Thus, by changing the constellation, the reliability of different data streams can be flexibly adjusted to meet different requirements.

A generalized multilevel coded modulation system using an LDPC code of length 200 coded bits is considered next. This code length is practical for applications such as voice

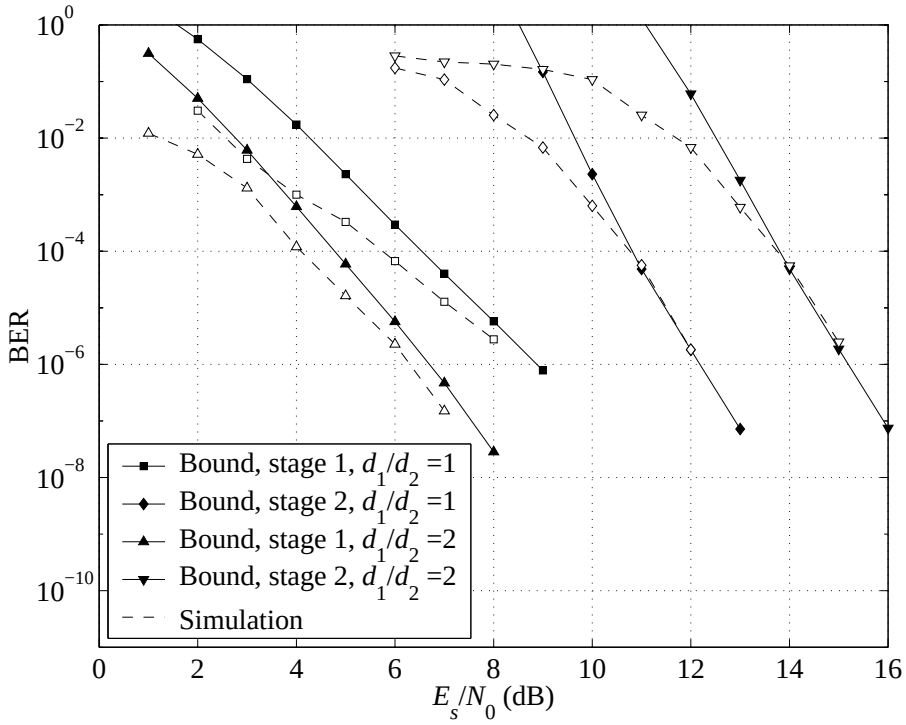


Figure 4.7 Bounds and the simulation results with OSD decoding over an AWGN channel: $(3, 6)$-regular LDPC codes of rates $1/2$ and $1/3$, a length of 72 bits, nonuniform 16-QAM constellations, embedded mapping.

or control signals that require a low delay. This LDPC code was used in Chapter 3 and it is an irregular LDPC code constructed by the progressive-edge-growth (PEG) method [50]. Irregular LDPC codes created by this method have a better convergence property compared to regular LDPC codes. The smallest terms of the Hamming distance spectrum of this code are provided in [50]. At this code length, OSD is impossible to implement. Instead, the sum-product (SP) decoding is used with 20 iterations. Fig. 4.8 shows the bounds and simulation results of SP decoding for system employing nonuniform 16-QAM constellation and embedded mapping. Although the same code is applied for both stages, the embedded mapping still guarantees that the BEP of Stage 1 is low enough not to affect the BEP of Stage 2. For this code length, the performance of SP decoders does not converge to the performance of the ML decoder at low SNR range. Therefore, the performance curves of the simulation results in Fig. 4.8 lie above the upper bounds of the ML decoding rule. The performance of the SP decoder is closer to the performance of the ML decoder at higher SNR and simulation results are quite close the union bounds over this higher SNR range.

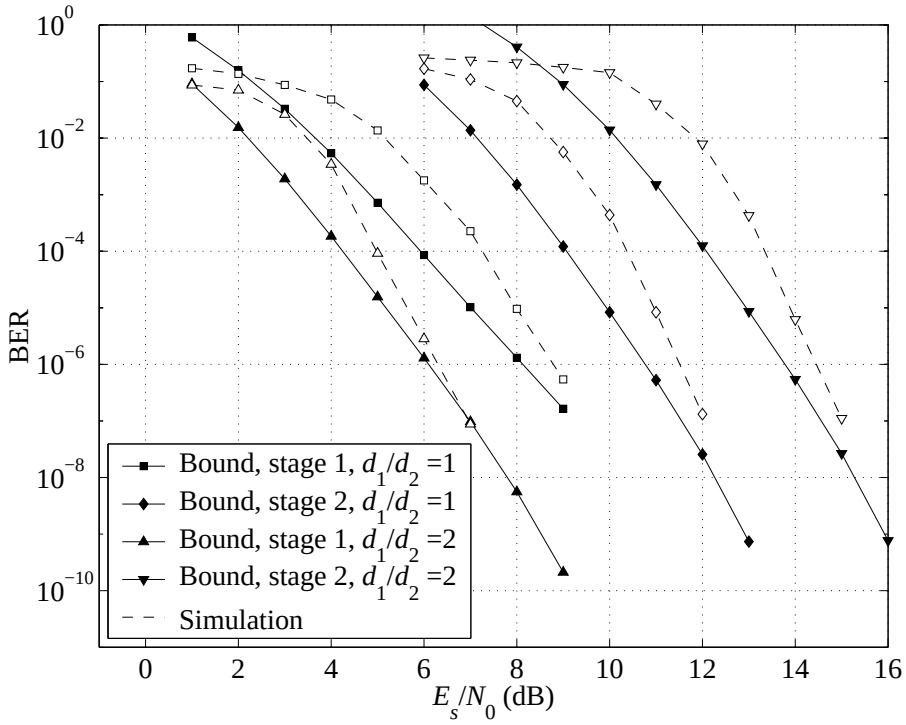


Figure 4.8 Bounds and simulation results with sum-product decoding over an AWGN channel: An irregular LDPC code of rate $1/2$ and length 200 bits, nonuniform 16-QAM constellations and embedded mapping.

4.5.2 Fading Channels

Fig. 4.9 presents the union bounds and simulation results of the generalized multilevel coded modulation with nonuniform 16-QAM and embedded mapping. The OSD is implemented for regular LDPC codes of length 72 and code rates $1/2$ and $1/3$, which are used for Stage 1 and Stage 2, respectively. It can be observed that the union bounds are also tight for the Rayleigh fading channel. The bounds for Stage 2 decoding seem to be tighter than the bounds for Stage 1 decoding. This observation is similar to the case of an AWGN channel. Compared to the performance over an AWGN channel, the performance over the Rayleigh fading channel degrades about 3.5 dB for Stage 1 decoding and about 3.0 dB for Stage 2 decoding, at the BER level of 10^{-5}.

The performance of SP decoding is also investigated for the fading channel. Again, the irregular LDPC code of length 200 coded bits is employed for both stages. Fig. 4.10 shows the union bounds and simulation results of the system with nonuniform 16-QAM

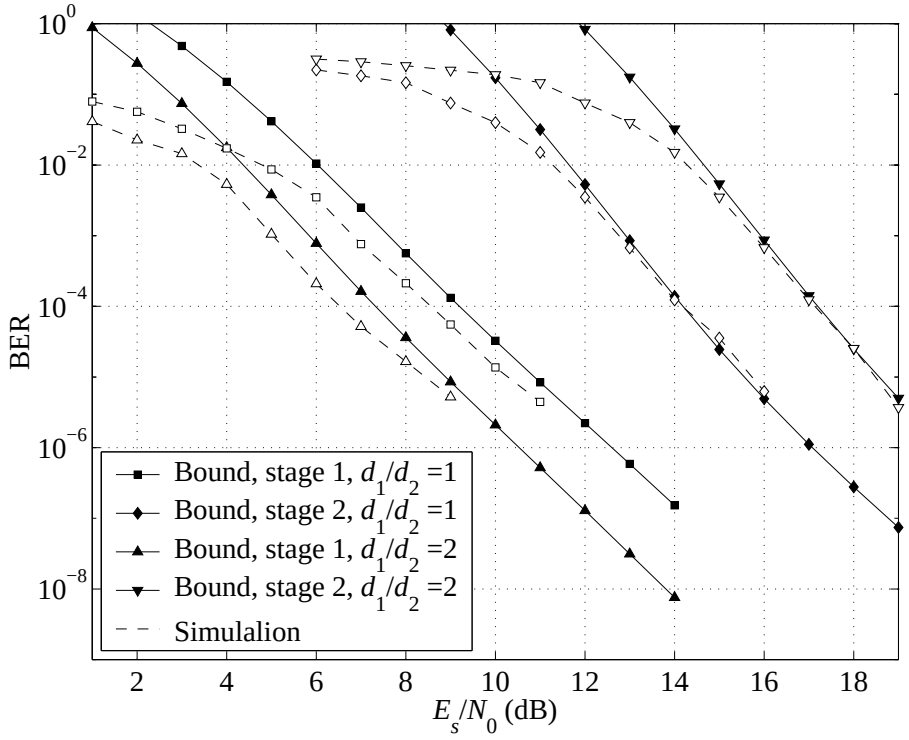


Figure 4.9 Bounds and simulation results with OSD decoding over a Rayleigh fading channel: A $(3, 6)$-regular LDPC code of rate $1/2$ and length 72 bits, nonuniform 16-QAM constellations and embedded mapping.

and embedded mapping. It can be seen that there are performance gaps of about 1.0 dB and 1.5 dB at BER level of 10^{-6} between the simulation results and the union bounds for Stage 1 and Stage 2, respectively. Similar to the case of an AWGN channel, this is due to the convergence property of the SP algorithm. It appears that with medium-length codes the SP decoders do not converge well over the Rayleigh fading channel. Hence, in general, the gaps between the bounds and the simulation results in the case of a Rayleigh fading channel are larger than that in the case of an AWGN channel at the practical BER levels between 10^{-6} and 10^{-7}.

4.6 Conclusions

Error performance of *generalized* multilevel coded modulation systems with multistage decoding was analyzed in this chapter. In these systems, the encoded bit stream of each level is mapped to a group of labelling bits instead of only one bit as in the case of *conventional* multilevel coded modulation. The union bounds of the bit error probabilities are

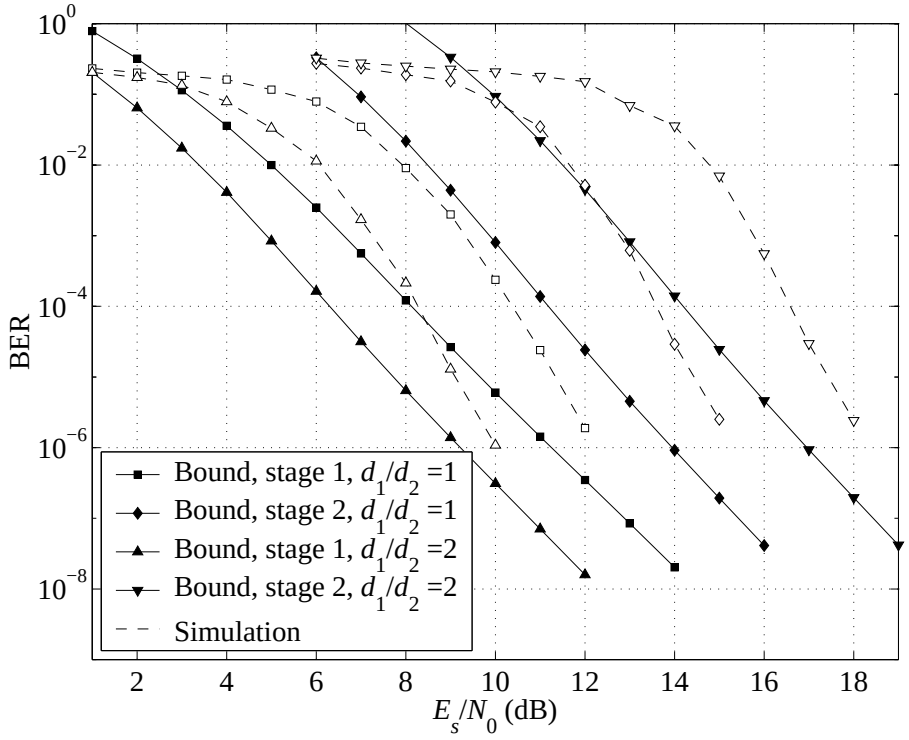


Figure 4.10 Bounds and simulation results with sum-product decoding over a Rayleigh fading channel: A regular LDPC code of rate $1/2$ and length 200 bits, nonuniform 16-QAM constellations and embedded mapping.

derived for different stages. The bounds are useful to evaluate the error performance of the systems having different parameters such as code rates, constellations and mapping rules. The bounds are shown to be tight by verifying against the simulation results of the ordered statistic decoding of short LDPC codes. For the sum-product decoding, there are gaps between the bounds and simulation results at low signal-to-noise ratio (SNR) range. However, at high SNR range, the bounds are seen to be close enough to the performance of the sum-product decoding. Thus, the bounds can be used to benchmark the performance of finite-length LDPC coded modulation systems. Although the chapter focuses on systems with 16-QAM constellations and two LDPC component codes, systems employing arbitrary constellations/mappings and more than two component codes can be similarly analyzed by the same framework.

5. LDPC Coded Unitary Space-Time Modulation

This chapter considers finite-length LDPC coded unitary space-time modulation systems. These systems are equipped with multiple antennas at both the transmitter and the receiver. The design and performance analysis are carried out for unknown channel state information (CSI) at the receiver. The structure of these LDPC coded unitary space-time modulation systems is similar to that of the bandwidth-efficient coded modulation studied in Chapter 3, where a group of coded bits is mapped to one symbol of the unitary space-time constellation. Here, the signal symbols of the unitary constellation are spread over several time intervals and multiple antennas instead of the in-phase and quadrature axes as in the case of the coded modulation systems in Chapter 3. In this chapter, the performance bound on the bit error probability is derived for any code rate, unitary space-time constellation and mapping. The tightness of the bound is verified by simulation results of the ordered statistic decoding (OSD). This bound is also useful to benchmark the error performance of LDPC coded unitary space-time modulation systems that employ sum-product decoding.

5.1 Introduction

Future applications of wireless data transmission require high data rates as well as high quality of services such as high reliability and low delay. It is a challenging task to design wireless communications systems to meet this demand. In order to achieve this goal, new communications techniques are needed to combat the physical limitations of wireless channels such as fading and interference. Recently, multiple-input multiple-output (MIMO) systems, that employ antenna arrays at both the transmitter and the receiver, have attracted a great interest in both the research community and industry. When a system is equipped with multiple antennas at the transmitter and the receiver, there are many paths

from the transmitter to the receiver. If the distances among the transmit antennas and the receive antennas are large enough, the fading coefficients of the paths are independent. A signal can thus be simultaneously transmitted over independent fading paths and detected with a high reliability at the receiver. This is because when one path is in deep fade, the other paths are likely to be good. The transmitted signals at different transmit antennas can properly cooperate to exploit the diversity effect of the MIMO channel or to be conveniently detected at the receiver. Compared to a single-transmit antenna and single-receive antenna system, the MIMO systems can provide significant improvement in terms of both the data rate, i.e., the bandwidth efficiency, and the reliability of the communications.

Based on the assumption that the channel state information (CSI) is perfectly estimated at the receiver, information theoretical results show that the capacity of MIMO systems is much higher than that of single-antenna systems [94]. Assuming that the channel coefficients among different pairs of the transmit and receive antennas are independent and known to the receiver, efficient designs of Bell Labs layered space-time (BLAST) schemes, space-time trellis codes and space-time block codes are introduced in [95], [20, 22], and [21], respectively. The V-BLAST (vertical BLAST) scheme treats two-dimensional signal symbols as independent signals and divide the symbol stream to antennas for transmission. Hence, the V-BLAST scheme increases the date rate or the bandwidth efficiency of the MIMO system. In other words, the V-BLAST scheme increases the multiplexing gain [96] of the MIMO systems. On the other hand, the D-BLAST (diagonal BLAST) scheme repeats the signal symbol over the transmit antennas. Therefore, the symbols contain redundancy for detection at the receiver and the error performance of the MIMO system is improved. For quasi-static Rayleigh fading channels, the space-time codes are designed over the time interval of constant fading coefficients. The two-dimensional symbols and their complex conjugates are transmitted over antennas and follow an order that is determined by a matrix for a space-time block code or a trellis for a space-time trellis code.

Space-time codes in general and BLAST schemes in particular are the diversity schemes implemented over space and time. For uncoded systems, the structure of these schemes decides the trade-off between multiplexing and diversity gains of a MIMO system [96, 97].

The concepts of multiplexing and diversity gains are similar to the code rate and error performance of error control coding. Hence, the problem of diversity and multiplexing trade-off for a MIMO system is similar to the trade-off between the error exponent and the reliability function as discussed in Chapter 3 [96].

Space-time codes and BLAST schemes are often designed, evaluated and compared to each other without the use of an error control coding. This means that the information bits are spread over a short-time interval in these schemes. However, when powerful error control coding is employed, the difference on the performance of these diversity schemes is significantly decreased [3].

The above multiple-antenna modulation schemes are discussed in the context of known CSI at the receiver, i.e., with the assumption that the fading coefficients are perfectly estimated. Unfortunately, for many mobile communication systems, the fading coefficients change too fast due to large Doppler shifts. For example, a wireless channel for a mobile traveling at 60-mi/h and with carrier frequency of 1.9 GHz has a coherence time period of about 3 ms. With a symbol rate of 30 kHz, a fresh fade occurs about every 100 symbols. Moreover, if the training sequences, which have no information content, are used to estimate the channel state information, there is a significant waste of the valuable spectrum.

Therefore, another design approach is introduced in [23, 24] for the systems without the knowledge of the channels coefficients at both the transmitter and the receiver. In these papers, the authors show that a class of signal constellations, known as unitary space-time constellations, is the most suitable modulation format for MIMO channels without CSI at the receiver or the transmitter. The advantages offered by these constellations motivated research work on the designs of unitary space-time constellations in order to optimize the system performance while reducing the processing complexity [98–100]. In particular, the design criterion used in [98–100] is based on the pairwise error probability, originally derived in [24] for uncoded unitary space-time modulation. This pairwise error probability for such non-coherent systems is computed in the complex Grassmanian space, which is a counterpart to the Euclidean space in the case of coherent systems.

The results on the error performance of unitary space-time modulation [24, 98–100] are mainly investigated for uncoded systems. In order to improve the performance of the systems, channel coding, especially the powerful pseudo-random codes such as Turbo codes and LDPC codes [5], should be considered. Although the authors in [101] investigate the unitary space-time modulation with Turbo codes, only simulation is carried out. The error performance of bit-interleaved coded unitary space-time modulation with iterative demodulation and decoding (BICM-ID) was recently studied in [102]. The work in [103] studies an LDPC coded unitary space-time OFDM system over a broadband mobile channel, but again only by means of computer simulation.

As discussed, analytical results for the maximum likelihood (ML) bound are not available for a system that combines an unitary space-time modulation and a pseudo-random code. In this chapter, the ML bound shall be derived for finite-length LDPC coded unitary space-time modulation systems. For finite-length LDPC codes, the error performance of the sum-product decoder converges to that of the ML decoder at practical BER levels. Therefore, the ML bound is useful to benchmark the error performance of finite-length LDPC coded unitary modulation systems. Here, the usefulness of the bound is illustrated by simulation results of LDPC coded unitary systems employing sum-product decoding and ordered-statistic decoding (OSD). These algorithms were presented in Chapter 2 and Chapter 3, respectively.

5.2 System Model

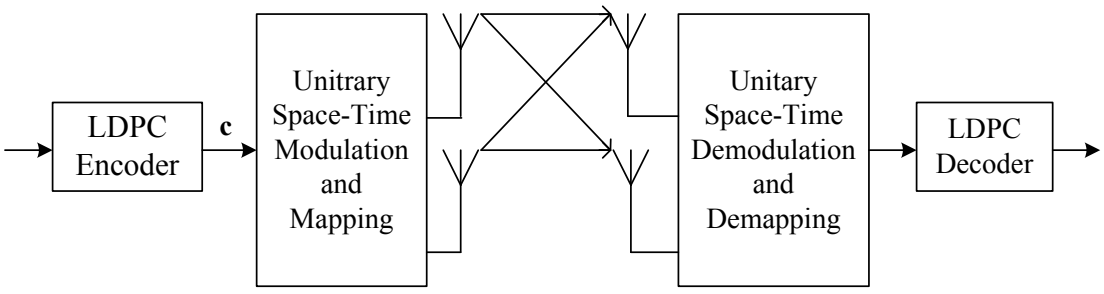


Figure 5.1 Block diagram of LDPC coded unitary space-time modulation system.

The block diagram of LDPC coded unitary space-time modulation system is illustrated in Fig. 5.1. An information sequence is first encoded into a codeword $\mathbf{c}$ using an LDPC

code of rate R_c and length N_c bits. Then every group of q bits in $\mathbf{c}$ is mapped by a mapping μ to one of $L = 2^q$ matrices $\mathbf{\Psi}(l)$, $l = 1, 2, \ldots, L$. Here $\mathbf{\Psi}(l)$, $1 \leq l \leq L$, are $T \times N_t$ unitary matrices which constitute the constellation $\mathbf{\Psi}$, N_t is the number of transmit antennas and T is the number of signalling intervals. The mapping μ from q coded bits to a signal point in $\mathbf{\Psi}$ can be a Gray mapping in which the labels of the closest signal points measured in terms of chordal distance differ in only one bit, or any other mapping. Thus, the structure of the transmitter of this system is similar to that of the coherent coded modulation system studied in Chapter 3. Here, the unitary constellation over space and time is employed instead of a two-dimensional constellation of in-phase and quadrature axes as in Chapter 3.

The system is equipped with N_r receive antennas. At the symbol time index τ, the $T \times N_r$ matrix $\mathbf{Y}_\tau$ of the received signals corresponding to the transmitted signal $\mathbf{X}_\tau \in \mathbf{\Psi}$ can be written as:

$$\mathbf{Y}_\tau = \sqrt{\frac{\gamma T}{N_t}} \mathbf{X}_\tau \mathbf{H}_\tau + \mathbf{W}_\tau \tag{5.1}$$

where $\mathbf{H}_\tau$ is the $N_t \times N_r$ matrix of the fading coefficients whose entries are zero-mean complex Gaussian random variables with unit variance. It means that the magnitude of the fading coefficients are Rayleigh random variables. The $T \times N_r$ matrix $\mathbf{W}_\tau$ represents the additive white Gaussian noise whose entries are also complex Gaussian random variables with zero mean and unit variance. The normalization factor in (5.1) ensures that the average signal-to-noise ratio (SNR) at each receive antenna is γ, independent of N_t. The spectral efficiency for the system is therefore qR_c/T information bits per second per Hertz (bits/s/Hz). In the above mathematical model, the fading coefficients are assumed to be constant over a block of T signalling intervals but change independently over blocks [22]. The channel coefficients, however, are unknown at both the transmitter and the receiver.

Conditional on the transmitted signal matrix $\mathbf{X}_\tau$, the probability density function (PDF) of the received signal is given by [23, 24].

$$p(\mathbf{Y}_\tau | \mathbf{X}_\tau) = \frac{\exp\left(-\text{tr}\{\mathbf{\Lambda}_\tau^{-1} \mathbf{Y}_\tau \mathbf{Y}_\tau^\dagger\}\right)}{\pi^{TN_r} \left[\det(\mathbf{\Lambda}_\tau)\right]^{N_r}} \tag{5.2}$$

where "tr" denotes the trace function, $(\cdot)^\dagger$ is conjugate transpose and $\mathbf{\Lambda}_\tau = \mathbf{I}_T + (\gamma/N_t)\mathbf{X}_\tau \mathbf{X}_\tau^\dagger$ is the covariance matrix of $\mathbf{Y}_\tau$.

The receiver employs the maximum *a posteriori* probability (MAP) soft-output space-time demodulator. The likelihood ratios of the transmitted bits, which are the soft outputs of the demodulator, can be computed as follows [101]:

$$L(c_{\tau,k}) = \frac{\sum_{\mathbf{\Psi}(l)\in\mathbf{\Psi}_1^k} \frac{\exp\left(-\mathrm{tr}\left\{\mathbf{\Lambda}(l)^{-1}\mathbf{Y}_\tau\mathbf{Y}_\tau^\dagger\right\}\right)}{[\det(\mathbf{\Lambda}(l))]^{N_r}}}{\sum_{\mathbf{\Psi}(l)\in\mathbf{\Psi}_0^k} \frac{\exp\left(-\mathrm{tr}\left\{\mathbf{\Lambda}(l)^{-1}\mathbf{Y}_\tau\mathbf{Y}_\tau^\dagger\right\}\right)}{[\det(\mathbf{\Lambda}(l))]^{N_r}}} \tag{5.3}$$

where $L(c_{\tau,k})$ is the likelihood ratio of the kth bit ($1 \le k \le q$) of the τth transmitted symbol. The covariance matrices $\mathbf{\Lambda}(l)$ are computed for each space-time symbol as $\mathbf{\Lambda}(l) = \mathbf{I}_T + (\gamma/N_t)\mathbf{\Psi}(l)\mathbf{\Psi}^\dagger(l)$. Also $\mathbf{\Psi}_1^k$ and $\mathbf{\Psi}_0^k$ are subsets of the symbols of $\mathbf{\Psi}$ such that the kth bits of their mapping labels are 1 and 0, respectively. It should be pointed out that (5.3) is applicable for general space-time constellations and the unitary structure of the transmitted signals is not taken into account. Furthermore, the soft-output demodulator based on (5.3) is generally very complex since it requires the computations of the inverses and determinants of the covariance matrices $\mathbf{\Lambda}(l)$. When the unitary space-time modulation is employed, one has $\mathbf{\Psi}^\dagger(1)\mathbf{\Psi}(1) = \mathbf{\Psi}^\dagger(2)\mathbf{\Psi}(2) = \cdots = \mathbf{\Psi}^\dagger(L)\mathbf{\Psi}(L) = \mathbf{I}_{N_t}$, and the likelihood ratios can be computed much simpler as follows [101]:

$$L(c_{\tau,k}) = \frac{\sum_{\mathbf{\Psi}(l)\in\mathbf{\Psi}_1^k} \exp\left(\mathrm{tr}\left\{\frac{1}{1+N_t/\gamma T}\mathbf{Y}_\tau^\dagger\mathbf{\Psi}(l)\mathbf{\Psi}^\dagger(l)\mathbf{Y}_\tau\right\}\right)}{\sum_{\mathbf{\Psi}(l)\in\mathbf{\Psi}_0^k} \exp\left(\mathrm{tr}\left\{\frac{1}{1+N_t/\gamma T}\mathbf{Y}_\tau^\dagger\mathbf{\Psi}(l)\mathbf{\Psi}^\dagger(l)\mathbf{Y}_\tau\right\}\right)} \tag{5.4}$$

The likelihood ratios computed by (5.4) are fed to the LDPC decoder. In this chapter, two kinds of LDPC decoders, the ordered statistic decoder (OSD) [18,74,104] and the sum-product (SP) decoder [5, 6], are considered. Performances of systems with these two kinds of decoders are investigated by computer simulation. The simulation results are compared to the performance bound, which is derived based on the maximum likelihood criterion. The algorithms of these decoders are described in detail in Chapters 2 and 3.

5.3 Unitary Space-Time Constellation and Mapping

For the general case, a symbol of a space-time constellation can be any matrix of size $T \times N_t$. However, information theoretical results in [23,24] show that the unitary space-time constellations, which are othornormal with respect to time among the transmit antennas, are

most suitable to achieve the capacity of the non-coherent fading channel. Mathematically, this means that a space-time symbol $\mathbf{\Psi}(l)$ should satisfy $\mathbf{\Psi}^{\dagger}(l)\mathbf{\Psi}(l) = I_{N_t}$. Thus, in contrast to being freely located in a $T \times N_t$-dimensional complex signal space in the general case, the symbols of a unitary space-time constellation lie on a manifold of this $T \times N_t$-dimensional complex space. Note that, since the unitary space-time symbols are of equal energy, the surface of the energy constraint's sphere includes the manifold that the unitary space-time symbols can be located on. This manifold is similar to a curve on the surface of a 3-dimensional sphere.

The soft outputs of the detector for the unitary space-time constellation are computed as in (5.4). Here, it can be recognized from (5.4) that only the quantity tr $\left\{\mathbf{Y}_{\tau}^{\dagger}\mathbf{\Psi}(l)\mathbf{\Psi}^{\dagger}(l)\mathbf{Y}_{\tau}\right\}$ is relevant when the hard-decision detection is carried out. More specifically, the maximum-likelihood detector for an *uncoded* unitary space-time modulation implements the following rule [98]:

$$\mathbf{\Psi}_{\mathrm{ML}} = \arg \max_{\mathbf{\Psi}(l)\in\mathbf{\Psi}} \mathrm{tr}\left\{\mathbf{Y}_{\tau}^{\dagger}\mathbf{\Psi}(l)\mathbf{\Psi}^{\dagger}(l)\mathbf{Y}_{\tau}\right\} \tag{5.5}$$

With the above ML detector, the pairwise error probability (PEP) of the uncoded unitary space-time modulation is given as follows [98]:

$$P_{l,l'} = -\frac{1}{2\pi j}\int_{-\infty}^{\infty}\frac{1}{\omega + j/2}\left(\prod_{m=1}^{N_t}\left[\frac{1+\gamma T/N_t}{(\gamma T/N_t)^2(1-d_m^2)(\omega^2+a_m^2)}\right]^{N_t}\right)d\omega. \tag{5.6}$$

where $1 \geq d_1 \geq \cdots \geq d_{N_t} \geq 0$ are the singular values of the $N_t \times N_t$ correlation matrix $\mathbf{\Psi}(l)\mathbf{\Psi}(l')$ and

$$a_m = \sqrt{\frac{1}{4} + \frac{1+\gamma T/N_t}{(\gamma T/N_t)^2(1-d_m^2)}} \tag{5.7}$$

A common and logical design criterion for unitary space-time constellations is to minimize the PEP. Let F_l and $F_{l'}$ be two N_t-dimensional planes that are spanned by $\mathbf{\Psi}(l)$ and $\mathbf{\Psi}(l')$ in the $T \times N_t$ -dimensional complex signal space. Then, one can think about the singular values d_m as the cosines of the principle angles between F_l and $F_{l'}$ and the chordal distance between the two signals $\mathbf{\Psi}(l)$ and $\mathbf{\Psi}(l')$ is $\sqrt{\sum_{m=1}^{N_t}(1-d_m^2)}$ [98]. In this way, the design criterion of an unitary space-time constellation becomes maximizing the minimum

chordal distances among the unitary symbols. The design problem is therefore related to the so-called packing in complex Grassmanian manifolds [105].

There are a few studies on the design of unitary space-time modulations based on signal processing and algebraic perspectives. Among them, the orthogonal construction introduced in [100] is a very simple but yet effective design of unitary space-time constellations. When $N_t = 2$, given the number of signal points $L = M^2$, where M is a positive integer, the construction procedure presented in [100] is as follows. For an integer number l, $1 \leq l \leq L$, define $k = (l-1)$ div M and $p = (l-1)$ mod M. Then the 4×2 unitary signal point $\mathbf{\Psi}(l)$ is defined as [100]:

$$\mathbf{\Psi}(l) = \frac{1}{2} \begin{bmatrix} 1 & -1 & \exp\left(j\frac{2\pi}{M}k\right) & -\exp\left(-j\frac{2\pi}{M}p\right) \\ 1 & 1 & \exp\left(j\frac{2\pi}{M}p\right) & \exp\left(-j\frac{2\pi}{M}k\right) \end{bmatrix}^T \tag{5.8}$$

Observe that the two elements $\exp\left(j\frac{2\pi}{M}k\right)$ and $\exp\left(j\frac{2\pi}{M}p\right)$ belong to an M-PSK constellation. As shown in [100], for two unitary signal points $\mathbf{\Psi}(l)$ and $\mathbf{\Psi}(l')$ that correspond to two pairs $[k,p]$ and $[k',p']$, the two singular values of $\mathbf{\Psi}^\dagger(l)\mathbf{\Psi}(l')$ are equal and given by,

$$d_1(l,l') = d_2(l,l') = \frac{1}{2}\sqrt{2 + \cos\frac{2\pi}{Q}(k-k') + \cos\frac{2\pi}{Q}(p-p')} \tag{5.9}$$

Hereafter, references to a signal point using $\mathbf{\Psi}(l)$ and the pair of integers $[k,p]$, where k and p are related to l as defined earlier, are interchangeable. The extension for $N_t > 2$ is also presented in [100], but only possible for $N_t = 3$ and $N_t = 4$ due to the limitation of orthogonal design.

Using the standard definition of the distance between subspaces, the chordal distance between $\mathbf{\Psi}(l)$ and $\mathbf{\Psi}(l')$ is given as [98]:

$$d^c(l,l') = \sqrt{\sum_{m=1}^{N_t}[1 - d_m^2(l,l')]} \tag{5.10}$$

As an example, the chordal distances among symbols of the unitary space-time constellation constructed with orthogonal design and QPSK are tabulated in Table 5.1.

Table 5.1 Chordal distance profile of the unitary space-time constellation constructed with orthogonal design and QPSK.

	$\mathbf{\Psi}(1)$ $[0,0]$	$\mathbf{\Psi}(2)$ $[0,1]$	$\mathbf{\Psi}(3)$ $[0,2]$	$\mathbf{\Psi}(4)$ $[0,3]$	$\mathbf{\Psi}(5)$ $[1,0]$	$\mathbf{\Psi}(6)$ $[1,1]$	$\mathbf{\Psi}(7)$ $[1,2]$	$\mathbf{\Psi}(8)$ $[1,3]$	$\mathbf{\Psi}(9)$ $[2,0]$	$\mathbf{\Psi}(10)$ $[2,1]$	$\mathbf{\Psi}(11)$ $[2,2]$	$\mathbf{\Psi}(12)$ $[2,3]$	$\mathbf{\Psi}(13)$ $[3,0]$	$\mathbf{\Psi}(14)$ $[3,1]$	$\mathbf{\Psi}(15)$ $[3,2]$	$\mathbf{\Psi}(16)$ $[3,3]$
$\mathbf{\Psi}(1)$	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1
$\mathbf{\Psi}(2)$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{2}$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$
$\mathbf{\Psi}(3)$	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	$\sqrt{2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1
$\mathbf{\Psi}(4)$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$
$\mathbf{\Psi}(5)$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$
$\mathbf{\Psi}(6)$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{2}$
$\mathbf{\Psi}(7)$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	$\sqrt{2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$
$\mathbf{\Psi}(8)$	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	1
$\mathbf{\Psi}(9)$	1	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1
$\mathbf{\Psi}(10)$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{2}$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$
$\mathbf{\Psi}(11)$	$\sqrt{2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1
$\mathbf{\Psi}(12)$	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$
$\mathbf{\Psi}(13)$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	$\sqrt{1/2}$	1	$\sqrt{3/2}$	1	0	$\sqrt{1/2}$	1	$\sqrt{1/2}$
$\mathbf{\Psi}(14)$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{2}$	1	$\sqrt{1/2}$	1	$\sqrt{3/2}$	$\sqrt{1/2}$	0	$\sqrt{1/2}$	1
$\mathbf{\Psi}(15)$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	$\sqrt{2}$	$\sqrt{3/2}$	1	$\sqrt{3/2}$	$\sqrt{3/2}$	1	$\sqrt{1/2}$	1	1	$\sqrt{1/2}$	0	$\sqrt{1/2}$
$\mathbf{\Psi}(16)$	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{3/2}$	$\sqrt{2}$	$\sqrt{3/2}$	1	1	$\sqrt{3/2}$	1	$\sqrt{1/2}$	$\sqrt{1/2}$	1	$\sqrt{1/2}$	0

Now, let $d^c_{\min}$ be the minimum chordal distance between any two different signal points in an unitary constellation. Similar to the signal constellation in the conventional Euclidean space, the mapping rule μ of a constellation $\mathbf{\Psi}$ is called Gray mapping if the labels of two signal points at chordal distance $d^c_{\min}$ differ in only 1 bit. Besides Gray mapping, another mapping, called BICM-ID mapping, is also considered in this chapter. The BICM-ID mapping is designed for bit-interleaved coded space-time modulation with iterative demodulation and decoding (BICM-ID) that relies on a convolutional code [102]. This mapping is the best in terms of minimizing the *asymptotic* performance of BICM-ID. The Gray and BICM-ID mappings for the unitary space-time constellation of 16 symbols obtained from the orthogonal design with Q-PSK are shown in Table 5.2.

5.4 Performance Bound

Due to the similarity between the structures of LDPC coded unitary space-time modulation considered here and coherent LDPC coded modulation investigated Chapter 3, the steps of derivation of the performance bounds are also similar. First, the bit error probability (BEP) is bounded by a summation of the pairwise error probabilities. Then, the

Table 5.2 Two different mappings of the unitary constellation obtained from orthogonal design and QPSK.

Gray															
$\mathbf{\Psi}(1)$	$\mathbf{\Psi}(2)$	$\mathbf{\Psi}(3)$	$\mathbf{\Psi}(4)$	$\mathbf{\Psi}(5)$	$\mathbf{\Psi}(6)$	$\mathbf{\Psi}(7)$	$\mathbf{\Psi}(8)$	$\mathbf{\Psi}(9)$	$\mathbf{\Psi}(10)$	$\mathbf{\Psi}(11)$	$\mathbf{\Psi}(12)$	$\mathbf{\Psi}(13)$	$\mathbf{\Psi}(14)$	$\mathbf{\Psi}(15)$	$\mathbf{\Psi}(16)$
0000	0001	0011	0010	0100	0101	0111	0110	1100	1101	1111	1110	1000	1001	1011	1010
BICM-ID															
$\mathbf{\Psi}(1)$	$\mathbf{\Psi}(2)$	$\mathbf{\Psi}(3)$	$\mathbf{\Psi}(4)$	$\mathbf{\Psi}(5)$	$\mathbf{\Psi}(6)$	$\mathbf{\Psi}(7)$	$\mathbf{\Psi}(8)$	$\mathbf{\Psi}(9)$	$\mathbf{\Psi}(10)$	$\mathbf{\Psi}(11)$	$\mathbf{\Psi}(12)$	$\mathbf{\Psi}(13)$	$\mathbf{\Psi}(14)$	$\mathbf{\Psi}(15)$	$\mathbf{\Psi}(16)$
0000	1001	0011	1010	1100	0101	1111	0110	1011	0010	1000	0001	0111	1110	0100	1101

average union bound is computed over the ensemble of the permuted versions of the LDPC code. The pairwise error probabilities are then classified into groups with the same characteristics. Averaging over these groups of pairwise error probabilities yields a closed-form expression of the performance bound on the BEP based on the Hamming distance spectrum of the code and the chordal distance profile of the unitary constellation. Some differences arise, however, due to the use of multiple transmit and receive antennas and non-coherent detection. It should be mentioned that although the structures of the transmitters in both systems are similar, the detection is different at the two receivers. For multiple transmit and receive antennas, the symbols of the transmitted constellation are superimposed at the receiver. Therefore, detection and performance evaluation are considered for the superimposed constellation instead of the transmitted constellation as in the case of coded modulation scheme in Chapter 3.

Consider a received sequence $\underline{\mathbf{Y}}$ of length $N_s = N_c/q$ symbols that corresponds to codeword $\mathbf{c}$ of $\mathcal{C}$. Here, the sequence $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_\tau, \ldots, \mathbf{Y}_{N_s}]$ consists of N_s received signal symbols, which results from the transmission of the symbol sequence $\underline{\mathbf{X}} = [\mathbf{X}_1, \ldots, \mathbf{X}_\tau, \ldots, \mathbf{X}_{N_s}]$, where each element of $\underline{\mathbf{X}}$ is a symbol (or signal point) of an unitary constellation $\mathbf{\Psi}$.

Based on the received sequence $\underline{\mathbf{Y}}$, the maximum likelihood (ML) decoding rule makes the decision to a symbol sequence $\underline{\widehat{\mathbf{X}}}$ whose corresponding received signal is closest to $\underline{\mathbf{Y}}$ in terms of the sum of the chordal distances. Note that, here, the ML decision rule is interpreted in the context of non-coherent detection in which the CSI is unknown at the receiver. The error performance of the ML rule is the ultimate performance limit for any

other decoding algorithm, i. e., it serves as a lower bound of the error performance of any other decoders.

The ML decoding makes an error when the received signal $\underline{\mathbf{Y}}$ exceeds the boundary of the decision region of the transmitted sequence $\underline{\mathbf{X}}$. Thus, the conditional frame error probability is a sum of the conditional error probabilities corresponding to $2^{R_c N_c}$ decision regions of all the possible transmitted sequences. Furthermore, since all the possible sequences are chosen equally likely, the union bound of the bit error probability can be written as:

$$P_e \leq \frac{1}{2^{R_c N_c}} \sum_{\underline{\mathbf{X}}} \sum_{\breve{\underline{\mathbf{X}}} \neq \underline{\mathbf{X}}} \frac{W_{\underline{\mathbf{X}},\breve{\underline{\mathbf{X}}}}}{N_c} \Pr(\underline{\mathbf{X}} \rightarrow \breve{\underline{\mathbf{X}}}) \tag{5.11}$$

where $\Pr(\underline{\mathbf{X}} \rightarrow \breve{\underline{\mathbf{X}}})$ is the pairwise error probability that $\underline{\mathbf{X}}$ is transmitted, but $\breve{\underline{\mathbf{X}}}$ is decided at the receiver. The parameter $W_{\underline{\mathbf{X}},\breve{\underline{\mathbf{X}}}}$ is the Hamming distance of two codewords $\mathbf{c}$ and $\breve{\mathbf{c}}$ of $\mathcal{C}$ obtained by demapping the two sequences $\underline{\mathbf{X}}$ and $\breve{\underline{\mathbf{X}}}$, respectively.

Averaging over the permuted LDPC code ensemble, the average union bound can thus be written as follows:

$$\overline{P}_e \leq E\left[\frac{1}{2^{R_c N_c}} \sum_{\underline{\mathbf{X}}} \sum_{\breve{\underline{\mathbf{X}}} \neq \underline{\mathbf{X}}} \frac{W_{\underline{\mathbf{X}},\breve{\underline{\mathbf{X}}}}}{N_c} \Pr(\underline{\mathbf{X}} \rightarrow \breve{\underline{\mathbf{X}}})\right] \tag{5.12}$$

Following similar analysis as in [24], if $\underline{\mathbf{X}}$ and $\breve{\underline{\mathbf{X}}}$ differ in d symbols, the PEP between the two sequences $\underline{\mathbf{X}}$ and $\breve{\underline{\mathbf{X}}}$ can then be computed as follows:

$$\Pr(\underline{\mathbf{X}} \rightarrow \breve{\underline{\mathbf{X}}}) = \Pr\left(\sum_{e=1}^{d} \mathrm{tr}(\mathbf{Y}^{\dagger}_{\tau(e)} \breve{\mathbf{X}}_{\tau(e)} \breve{\mathbf{X}}^{\dagger}_{\tau(e)} \mathbf{Y}_{\tau(e)}) > \sum_{e=1}^{d} \mathrm{tr}(\mathbf{Y}^{\dagger}_{\tau(e)} \mathbf{X}_{\tau(e)} \mathbf{X}^{\dagger}_{\tau(e)} \mathbf{Y}_{\tau(e)})\right) \tag{5.13}$$

where $\tau(e)$, $e = 1, 2, \cdots, d$, are the indices of the symbols in error.

The above PEP can be computed with a complex integral as follows [24]:

$$\Pr(\underline{\mathbf{X}} \rightarrow \breve{\underline{\mathbf{X}}}) = -\frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{1}{w + j/2} \left(\prod_{e=1}^{d} \Delta_e\right) \mathrm{d}w \tag{5.14}$$

where

$$\Delta_e = \prod_{m=1}^{N_r} \left[\frac{1 + \gamma T/N_t}{(\gamma T/N_t)^2 (1 - d^2_{m,e})(w^2 + a^2_{m,e})}\right]^{N_r} \tag{5.15}$$

In (5.15), $d_{m,e}$, $1 \leq m \leq N_r$, is the mth singular value of the $N_r \times N_r$ matrix $\breve{\mathbf{X}}^{\dagger}_{\tau(e)}\mathbf{X}_{\tau(e)}$. The parameter $a_{m,e}$ is computed from $d_{m,e}$ as in (5.7), where $d_{m,e}$ plays the role of d_m.

As mentioned before, the bound in (5.12) can be computed more conveniently by classifying the pairwise error probabilities into groups with the same characteristics and then averaging over these groups [73, 93]. To this end, consider a pair of $\underline{\mathbf{X}}$ and $\underline{\breve{\mathbf{X}}}$ that are demapped to two codewords $\mathbf{c}$ and $\breve{\mathbf{c}}$. The error sequence $\mathbf{e}$ is the modulo-2 sum of these two codewords. The number of error sequences with a given weight is determined by the Hamming weight spectrum of the code $\mathcal{C}$. However, the positions of the errors are not determined because of the randomization in creating LDPC codes. Similar to Chapter 3, error sequences of a given weight can be classified into some types of error sequences according to the number of bit groups with the same weights. Let n_i, $1 \leq i \leq q$, denote the number of error bit groups of weight i. An error sequence can be described by the parameters n_i. The vector $\mathbf{n}$, that contains these parameters n_i, can be referred to as the type of error sequences.

Then the terms of the expectation and summation in (5.12) can be grouped according to types of error sequences. The performance bound can be rewritten as follows:

$$\overline{P}_e \leq \sum_{l=l_{\min}}^{N_c} \sum_{n_1=0}^{N_s} \cdots \sum_{n_q=0}^{N_s} \frac{l}{N_c} \overline{f}(\mathbf{n}) E\left[\Omega\left(\mathbf{n}\right)\right] \tag{5.16}$$

where $f(\mathbf{n})$ is the number of error sequences of type $\mathbf{n}$ and $\overline{f}(\mathbf{n})$ is the expected value of this function over the ensemble of code $\mathcal{C}$. The expected total number of error sequences of type $\mathbf{n}$ is $2^{R_c N_c}\overline{f}(\mathbf{n})$. The function $\Omega\left(\mathbf{n}\right)$ is the average pairwise error probability corresponding to a type of error sequence. The value of this function is a random variable because the type of error sequence can correspond to different symbol error sequences that have different pairwise error probabilities.

The function $\overline{f}(\mathbf{n}^{(1)})$ can be similarly computed as in Equations (3.15) and (3.16) in Chapter 3, where the parameters of the unitary space-time modulation are substituted in the places of the parameters of the coded modulation. For example, N_c is used instead of N in Equations (3.15) and (3.16).

To compute the bound, the probability mass function (PMF) of the random variables $\Omega(\mathbf{n})$ needs to be known. Let $\Omega_k(\mathbf{n})$ denote a possible outcome of the random variable $\Omega(\mathbf{n})$ and $p_{\mathbf{n},k}$ be the corresponding probability, where k ($1 \leq k \leq \kappa_{\mathbf{n}}$) is the index of the outcome and $\kappa_{\mathbf{n}}$ is the number of possible outcomes. Each outcome $\Omega_k(\mathbf{n})$ and its corresponding probability $p_{\mathbf{n},k}$ are determined by a possible combination of erroneous symbols. The following example is given to illustrate how an outcome $\Omega_k(\mathbf{n})$ is established from the erroneous symbols.

Consider an LDPC coded unitary modulation system where the unitary space-time constellation has 16 symbols and the BICM-ID mapping is employed in Subsection 5.3. For the type $\mathbf{n} = [1\ 2\ 0\ 0]$ of error sequence, there are $d = 1 + 2 = 3$ unitary symbols in error in which 1 symbol has 1 bit in error and 2 symbols have 2 bits in errors in their labels. Figure 5.2 illustrates the transmitted sequence $\underline{\mathbf{X}}$ with this error sequence. The unitary symbols $\mathbf{X}_{\tau(1)}$, $\mathbf{X}_{\tau(2)}$ and $\mathbf{X}_{\tau(3)}$ are symbols with 1 bit in error, 2 bits error and 2 bits in error in their labels, respectively. For a 1 bit error in $\mathbf{X}_{\tau(1)}$, an erroneous symbol can be the result of making a wrong decision among 32 pairs of symbols in the 16-ary unitary constellation. For each pair of unitary symbols, there is only one singular value $d_{m,e}$ ($e = 1$) of matrix $\breve{\mathbf{X}}_{\tau(1)}^{\dagger}\mathbf{X}_{\tau(1)}$. This singular value is necessary to compute the outcome $\Omega_k(\mathbf{n})$ as indicated in (5.15) and (5.14). Fortunately, for a BICM-ID mapping, $d_{m,1}$'s take on the same value $d_{m,1}^{(1)} = 1$ for all 32 pairs of symbols. Thus, the corresponding probability for this $d_{m,1}^{(1)}$ is $p_1^{(1)} = 1$. For 2 bits in error in $\mathbf{X}_{\tau(2)}$ and $\mathbf{X}_{\tau(3)}$, 64 pairs of unitary symbols need to be considered to compute $d_{m,e}$'s ($e = 2$ or $e = 3$). However, there are 48 pairs with the same value $d_{m,e}^{(1)} = 0.5$ and 16 other pairs take the same value $d_{m,e}^{(2)} = 0.0$. Assuming that all the symbols are equally used, the corresponding probabilities of $d_{m,e}^{(1)}$ and $d_{m,e}^{(2)}$ are $p_e^{(1)} = 48/(48+16) = 3/4$ and $p_e^{(2)} = 16/(48+16) = 1/4$. For each $d_{m,e}^{(h)}$, the parameter $\Delta_e^{(h)}$ can be computed by (5.15). A possible outcome $\Omega_k(\mathbf{n})$ is determined by the set $\{\Delta_{e,k}\}$, which is created by collecting one value $\Delta_e^{(h)}$ from the erroneous symbols $\mathbf{X}_{\tau(e)}$. Thus, for $\mathbf{n} = [1\ 2\ 0\ 0]$ in the example, four possible outcomes $\Omega_k(\mathbf{n})$ are determined by four sets $\{\Delta_{e,1}\} = \left\{\Delta_1^{(1)}, \Delta_2^{(1)}, \Delta_3^{(1)}\right\}$, $\{\Delta_{e,2}\} = \left\{\Delta_1^{(1)}, \Delta_2^{(1)}, \Delta_3^{(2)}\right\}$, $\{\Delta_{e,3}\} = \left\{\Delta_1^{(1)}, \Delta_2^{(2)}, \Delta_3^{(1)}\right\}$ and $\{\Delta_{e,4}\} = \left\{\Delta_1^{(1)}, \Delta_2^{(2)}, \Delta_3^{(2)}\right\}$. Here, the number of out-

comes $\kappa_{\mathbf{n}}$ is 4. The value of $\Omega_k(\mathbf{n})$ is computed from $\{\Delta_{e,k}\}$ by (5.14) and the corresponding probability $p_{\mathbf{n},k}$ is a product of the probabilities $p_e^{(h)}$'s that contribute to the combination of $\{\Delta_{e,k}\}$. Note that, some outcomes $\Omega_k(\mathbf{n})$ can have the same value, such as $\Omega_2(\mathbf{n})$ and $\Omega_3(\mathbf{n})$ in the example. Thus, they can be grouped by summing the corresponding probabilities when the bound or the PMF of $\Omega(\mathbf{n})$ is computed by computer.

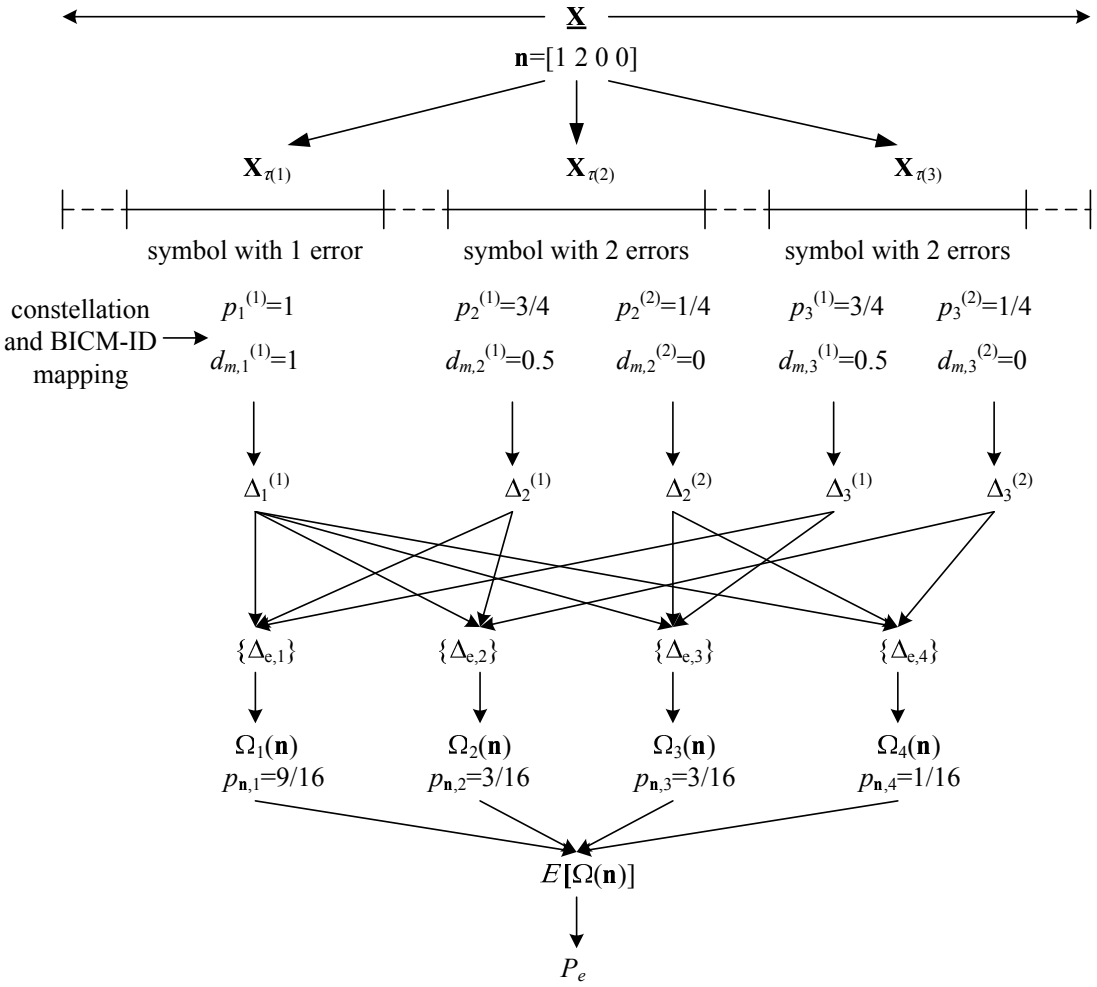


Figure 5.2 Possible outcomes of $\Omega(\mathbf{n})$, for $\mathbf{n} = [1\ 2\ 0\ 0]$: Unitary space-time constellation of 16 symbols, BICM-ID mapping.

After determining the possible outcomes $\Omega_k(\mathbf{n})$ and their corresponding probabilities $p_{\mathbf{n},k}$ of $\Omega(\mathbf{n})$, the performance bound can be rewritten as:

$$\overline{P}_e \leq \sum_{l=l_{\min}}^{N_c} \sum_{n_1=0}^{N_s} \cdots \sum_{n_q=0}^{N_s} \sum_{k=1}^{\kappa_{\mathbf{n}}} \frac{l}{N_c} \overline{f}(\mathbf{n}) p_{\mathbf{n},k} \left[-\frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{1}{\omega + j/2} \left(\prod_{e=1}^{d} \Delta_{e,k} \right) d\omega \right] \quad (5.17)$$

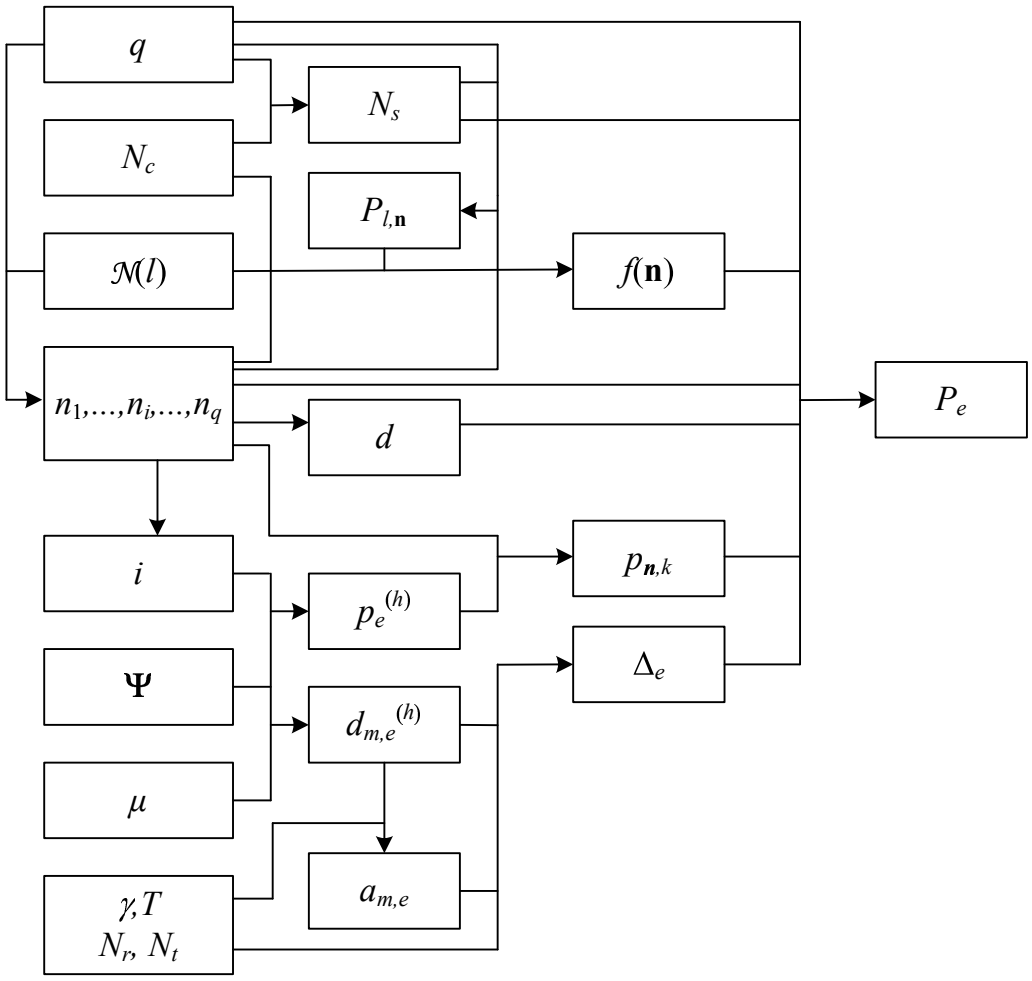


Figure 5.3 Block diagram for the computation of the bound.

The above procedure shows all the necessary computations to determine the union bound in (5.17) for the unitary space-time LDPC coded modulation systems. As a summary, the computation procedure is described with a block diagram in Fig. 5.3.

5.5 Illustrative Results

This section provides analytical and simulation results to confirm the analysis carried out in the previous section. First, a short LDPC code is chosen to verify the accuracy of the union bound. This code is a regular (3,6) LDPC code of rate $1/2$ and length 72 bits. With this short length, the Hamming weight spectrum of this code can be determined by searching over all codewords. The bound is computed based on the Hamming distances of 8 to 20. The system is equipped with 2 transmit antennas and 2 receive antennas. The unitary

space-time modulation consists of 16 symbols based on the orthogonal design with QPSK, as described in Section 5.3. Thus, the spectral efficiency of the system is 0.5 bit/sec/Hz. Since the length of the code is short, the OSD-3 decoder (order 3) is used at the receiver. Two mappings, namely Gray mapping and BICM-ID mapping, are considered. Fig. 5.4 presents the bounds and simulation results on the BER of this coded unitary space-time modulation system. For Gray mapping, the performance of the system based on OSD decoder is close to the bound at the practical BER level of 10^{-6}. For BICM-ID mapping, there is a gap of about 3 dB between the performance of the OSD and the ML bound. A similar phenomenon was observed for coherent LDPC coded modulation systems over AWGN and Rayleigh fading channels in Chapter 3. It can also be explained by the fact that OSD for coded modulation system with the BICM-ID mapping is very suboptimum due to the PDF of the likelihood ratios, i.e., the soft outputs of the demodulators, that does not make the performance converge to the ML bound at the observed BER range. This phenomenon with the suboptimum receivers is also observed for BICM-ID systems with convolutional codes in [65, 67].

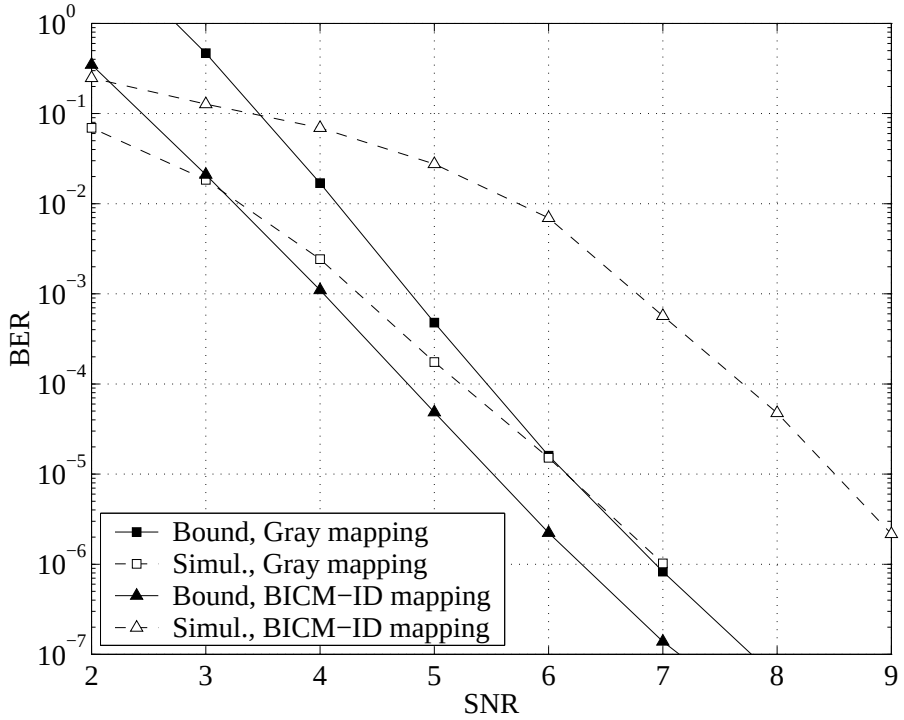


Figure 5.4 Bounds and simulation results with OSD decoding: A regular LDPC code of rate $1/2$ and length 72 bits.

A unitary space-time modulation system using an LDPC code of length 200 bits is considered next. This LDPC code is an irregular LDPC code constructed by the progressive-edge-growth (PEG) method [50]. The smallest terms of the Hamming distance spectrum of this code are provided in [50]. At this code length, the OSD decoder cannot be practically implemented or simulated. Instead, the sum-product (SP) decoding is used with 20 iterations. The same unitary constellation and Gray mapping as in the previous case are considered here. Fig. 5.5 shows the bounds and simulation results for both systems that use LDPC codes of length 72 and 200 and decoded with OSD and sum-product decoders, respectively. The two systems have the same spectral efficiency. For Gray mapping, the performance of the sum-product decoder can also approach the ML bound at low BER due to the good convergence of the sum-product algorithm over the considered range of SNR. Compared to the system using the code of length 72, the performance using the code of length 200 and the sum-product decoder is superior by 1 dB at the BER level of 10^{-6}. The bound also shows a gain of 1.8 dB at this BER level with ML decoding. Note that, the computational complexity of the sum-product decoder is much lower than that of the OSD even in the cases of short and medium code lengths.

The unitary space-time coded modulation systems are also studied with various number of receive antennas. The bounds and simulation results are presented in Fig. 5.6 with 2, 3 and 4 receive antennas. A regular LDPC code with a high rate of $433/495$ and length 495 is employed for these systems. The finite-length and high-rate codes are proper for communications systems that require both low latency and high spectral efficiency. The unitary space-time constellation of 16 symbols and Gray mapping are also employed for these systems. Fig. 5.6 shows that the performance of the systems using the sum-product decoder approaches the ML bound with only a small gap of 0.8 dB at the practical BER level of 10^{-6}. When the number of receive antennas increases from 2 to 3, the performance improves by 2 dB for both the ML decoder and the sum-product decoder. When the number of receive antennas increases from 3 to 4, there is only 1 dB gain at the BER level of 10^{-6}.

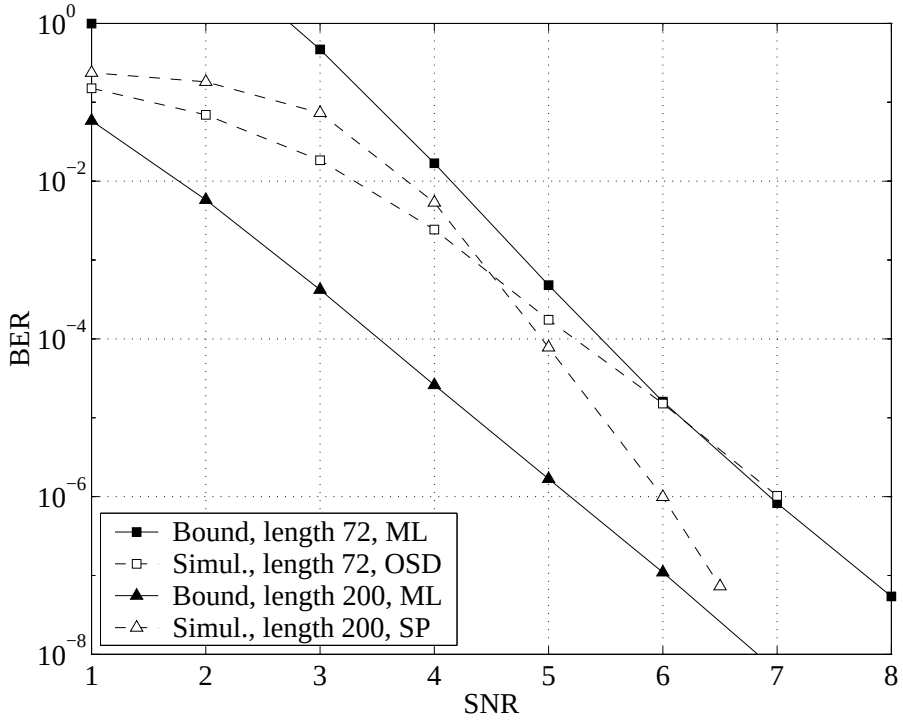


Figure 5.5 Bounds and simulation results with sum-product and OSD decoders: LDPC codes of rate $1/2$, lengths 72 and 200 bits, Gray mapping.

5.6 Conclusions

Error performance of non-coherent LDPC coded unitary space-time modulation was studied in this chapter. The performance bound was derived for the systems constructed from finite-length LDPC codes and unitary space-time constellations obtained from orthogonal design. The analytical derivations were substantiated by simulation results of OSD and sum-product decoding. Since the bounds are tight at the practical BER, these bounds can be used to benchmark the performance of LDPC coded unitary space-time modulation systems with various parameters. Although the bound is derived for the systems based on LDPC codes, the derivation framework can be applied to coded unitary space-time modulation systems with other finite-length pseudo-random codes such as Turbo codes or Turbo product codes.

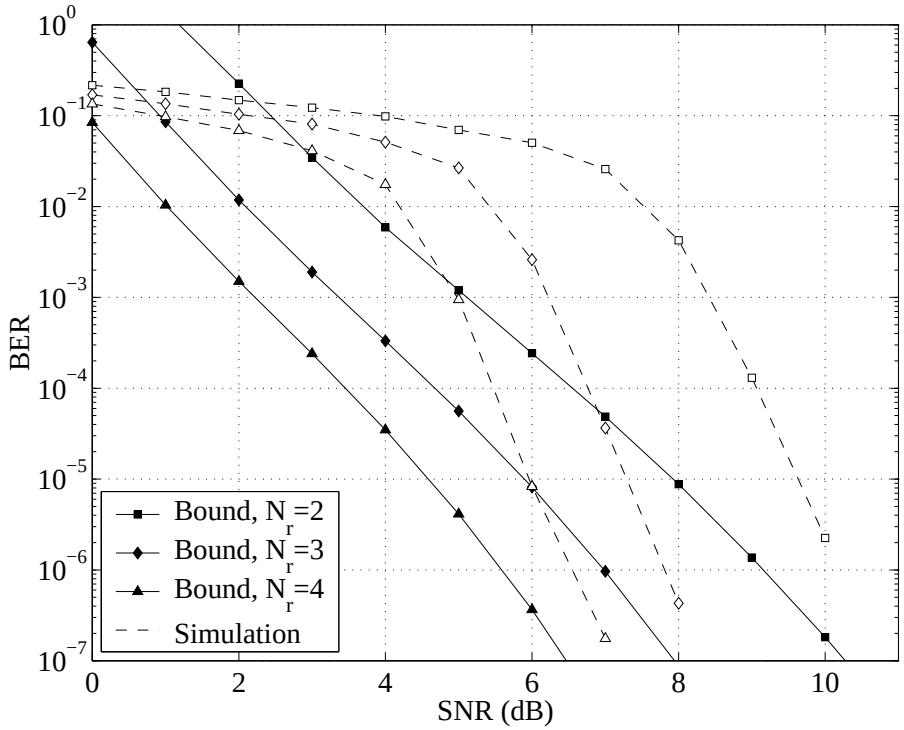


Figure 5.6 Bounds and simulation results with the sum-product decoders and various number of receive antennas (N_r): A regular LDPC code of rate $433/495$ and length 495 bits.

6. Adaptive LDPC Coded Modulation for MIMO Systems

The previous chapters study constant-rate communications systems, where the data rates are kept constant over all time. In these systems, the code rate and constellation are chosen and fixed in order to transmit a given data rate. This chapter studies variable-rate communication systems, whose constellation or code rate are changed over time to adapt to the channel condition. These variable-rate systems are commonly known as adaptive modulation systems. In particular, adaptive modulation systems with multiple transmit and receive antennas, i.e., for MIMO channels, are investigated in this chapter. One of the adaptive modulation systems is based on optimum beamforming at both the transmitter and the receiver. The other two systems are based on antenna selection at the transmitter. The advantages of these diversity schemes are discussed and their bandwidth efficiencies are compared for both uncoded and LDPC coded systems. For LDPC coded systems, the achievable threshold of spectral efficiency is computed based on the maximum likelihood bound on the error performance derived in Chapter 3.

6.1 Introduction

Wireless communications for future data applications requires high throughput, i. e., a high data rate, for a large number of users. Meanwhile, the radio transmission bandwidth is always limited and costly. Therefore, this limited bandwidth should be most efficiently utilized in future broadband wireless data networks. For many data applications such as Internet applications, data should be transmitted as fast as possible within limited bandwidth. On the other hand, these data applications do not require a constant data rate as in the case of voice applications. Therefore, it is reasonable to apply rate-variable systems for

these data applications. Moreover, for fixed wireless channels or mobile channels with low mobility speeds, the fading coefficients change slowly. This situation for wireless channels is in contrast to the high mobility speeds assumed in the previous chapters. When the fading coefficients vary slowly, it is possible to change the date rate of the systems, i. e., the modulation constellation and the code rate, in order to adapt to the channels.

To implement this adaptation, adaptive modulation techniques [25] are proposed to increase the bandwidth efficiency of these data communications systems. Recently, adaptive modulation techniques were used for the data mode of 2.5G mobile communication systems such as the general packet radio service (GPRS) or enhanced data rates for GSM evolution (EDGE) [106]. In these systems, the transmitters choose a high constellation size or a high code rate to transmit when the condition of the wireless channel is good. When the wireless channel is in deep fade, a lower constellation size or a lower code rate is selected. Here, changing the modulation constellation or code rate should still guarantee a target bit error rate for the application. In fact, adaptive modulation techniques trade the reliability of the data stream for the throughput. To vary the constellation and code rate properly, the channel state information (CSI), i.e., the channel fading coefficients, should be known at the transmitter. For two-way communications systems, the CSI is often estimated at the receiver and sent back to the transmitter via a feedback link. Of course, the delay of this feedback link affects the accuracy of the channel state information at the transmitter. The CSI can also be obtained by estimating the backward link in the time duplex division (TDD) systems [107]. Thus, for adaptive modulation systems, the CSI is considered to be known at both the transmitter and the receiver. Note that, this scenario of CSI availability is completely opposite to the scenario studied in Chapter 5, where the CSI is unknown at both the transmitter and the receiver.

As discussed in Chapter 5, employing multiple antennas at both the transmitter and the receiver is a solution to improve the error performance of the constant-rate systems. A number of diversity schemes such as space-time codes [20, 21] and Bell Labs layered space-time schemes [95] have been proposed to exploit the advantages of MIMO channels without requiring the knowledge of the CSI at the transmitter. For variable-rate systems, the

MIMO techniques can also improve the throughput or the bandwidth efficiency of wireless systems. Results of information theory show that the capacities of MIMO channels with CSI knowledge at the transmitter are much larger than that of the single-input single-output channel [94]. The capacities of MIMO channels are computed in terms of the bandwidth efficiency and are realized by adaptive modulation systems with multiple antennas.

In general, the adaptive modulation techniques can be implemented with different antenna diversity schemes at the transmitter and the receiver [108–117]. Several diversity schemes designed without the knowledge of the CSI at the transmitter such as the space-time codes [111, 118] or the vertical Bell Labs layered space-time scheme (V-BLAST) have also been adopted for systems with the CSI knowledge at the transmitter [112, 113, 115, 116]. This practice is merely due to the simplicity in implementation of these schemes at both the transmitter and the receiver. However, these schemes are not optimum in terms of throughput performance.

It was shown in [94] that the optimum power adaptation policy is implemented by water-filling over parallel channels that correspond to the eigenvalues of the channel matrix. Such an adaptation policy is generally referred to as beamforming. Thus, the bit stream should be divided into substreams in order to be separately modulated and transmitted over subchannels of eigenvalues. The capacity of a MIMO channel can be theoretically obtained using a continuous-rate adaptation. This scheme is investigated with different scenarios of power constraint, characteristics of the MIMO channel and availability of the CSI in [119–121]. In practice, adaptive systems, however, are discrete-rate systems. This means that only an integer number of bits is loaded to the subchannels. The throughput of the system is a collective result due to all the parallel channels. To maintain a given bit error rate (BER) target, the integer number of loaded bits for each subchannel should be less than the rate-threshold offered by the continuous-rate scheme. This implies that the practical throughput is below the theoretical capacity of the system.

Using antenna selection at the transmitter and antenna combining at the receiver is another approach. This diversity scheme is investigated for a constant-power variable-rate system with automatic repeat request (ARQ) in [117]. However, the variable-power

variable-rate system is not investigated and compared to the scheme based on parallel channels (i.e., beamforming). In selection diversity schemes, the bit stream is loaded to only one channel instead of many subchannels. Thus, the number of rate regions for this channel is larger than that for parallel subchannels of the beamforming scheme. This fact implies a potential of having a better loading over the beamforming scheme. Another important advantage of the schemes based on antenna selection is that less information needs to be sent back to the transmitter. In particular, only the parameters concerning one antenna need to be sent over the feedback link. For the scheme using all antennas at the transmitter, the parameters for all the antennas should be sent back. Based on the above discussion, it is of interest to study the practical throughput achieved by the adaptive modulation schemes using antenna selection and compare it to that of the beamforming scheme. This chapter studies two adaptive modulation schemes that are based on antenna selection (antenna scheduling) at the transmitter and antenna combining (or selection) at the receiver, together with the beamforming scheme.

The focus of this chapter is the investigation of the gap in throughput between these schemes for the cases of continuous-rate and discrete-rate systems. For adaptive modulation, the procedure to compute the system's parameters relies on the probability density function (PDF) of the received SNR. These PDFs of the SNRs are derived for the two antenna selection schemes mentioned earlier.

Adaptive LDPC coded modulation systems are also investigated in this chapter. Since it is not easy to vary the rate of an LDPC code, only the constellation size and signal power are varied to adapt to the channel states. The maximum-likelihood performance bound (ML bound) of the LDPC coded modulation system derived in Chapter 3 is applied to compute the parameters of the adaptive LDPC coded systems. For finite-length LDPC codes, the modulation constellation of the adaptive coded modulation remains constant over the length of codewords. Therefore, besides the advantage of low delay processing, the finite-length LDPC codes are also convenient for designing adaptive coded modulation systems since the instant bit error rate can be easily computed for a given code and constellation.

6.2 Adaptive Modulation Scheme Based on Parallel Subchannels

Consider a frequency-flat fading MIMO channel with N_t transmit antennas and N_r receive antennas. The received signal over such a channel can be represented as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{6.1}$$

In the above equation, the $N_r \times N_t$ matrix $\mathbf{H}$ represents the fading coefficients of the MIMO channel. The elements of $\mathbf{H}$ are independent complex Gaussian random variables with zero mean and unit variance. It is assumed that the channel matrix $\mathbf{H}$ is available at both the transmitter and the receiver. The vector $\mathbf{x}$ of size $N_t \times 1$ denotes the input of the channel, i.e., the transmitted signals over the antennas. The output of the channel and additive white Gaussian noise are represented by $N_r \times 1$ column vectors $\mathbf{y}$ and $\mathbf{w}$, respectively. The elements of $\mathbf{w}$ are also independent zero-mean complex Gaussian random variables with common variance $N_0/2$, where $N_0/2$ is one-sided power spectral density. Note that, compared to the channel model in the previous chapter, the transmitted and received signals in (6.1) are considered for only one time interval. The constellations are designed in the 2-dimensional space and they can be detected by the coherent demodulator. For convenience, let $m = \min(N_r, N_t)$ and $n = \max(N_r, N_t)$.

Any channel matrix $\mathbf{H}$ can be decomposed into the following form:

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger} \tag{6.2}$$

where $\mathbf{D}$ is an $N_r \times N_t$ diagonal matrix whose diagonal elements are the singular values of $\mathbf{H}$. The operation $(\cdot)^{\dagger}$ denotes conjugate transpose. The matrices $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_r}]$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{N_t}]$ are $N_r \times N_r$ and $N_t \times N_t$ unitary matrices whose columns are the left and right singular vectors of $\mathbf{H}$, respectively. In fact, the singular values $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_m}$ are square roots of the eigenvalues of the matrix $\mathbf{H}\mathbf{H}^{\dagger}$ and the columns of $\mathbf{U}$ and $\mathbf{V}$ are the eigenvectors of $\mathbf{H}\mathbf{H}^{\dagger}$ and $\mathbf{H}^{\dagger}\mathbf{H}$, respectively. Substitute (6.2) into (6.1). With simple manipulations, the following equations are obtained:

$$\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger}\mathbf{x} + \mathbf{w} \tag{6.3}$$

$$\mathbf{U}^{\dagger}\mathbf{y} = \mathbf{D}\mathbf{V}^{\dagger}\mathbf{x} + \mathbf{U}^{\dagger}\mathbf{w} \tag{6.4}$$

$$\mathbf{y}' = \mathbf{D}\mathbf{x}' + \mathbf{w}' \tag{6.5}$$

where $\mathbf{y}' = \mathbf{U}^{\dagger}\mathbf{y}$, $\mathbf{x}' = \mathbf{V}^{\dagger}\mathbf{x}$ and $\mathbf{w}' = \mathbf{U}^{\dagger}\mathbf{w}$. Thus, the MIMO channel $\mathbf{H}$ can be decomposed into m parallel eigen subchannels. The input and output of the equivalent subchannels are contained in $\mathbf{x}'$ and $\mathbf{y}'$, respectively. More importantly, the subchannel power gains are $\lambda_1, \lambda_2, \ldots, \lambda_m$. Note that, the characteristics of $\mathbf{w}'$ are kept the same as that of $\mathbf{w}$ over the unitary transformation, i.e., the random variables in both vectors $\mathbf{w}$ and $\mathbf{w}'$ are independent zero-mean Gaussian random variables with variance $N_0/2$.

If the channel matrix $\mathbf{H}$ is deterministic and the instantaneous power is constrained, the optimum power allocation is achieved by the water-filling solution over the eigen subchannels. However, here, the adaptive scheme is considered over the subchannels of the random matrix $\mathbf{H}$ and the average power is constrained. Therefore, the subchannels are considered as independent channels and the power adaptation policy is independently implemented in each channel over time. The block diagram of the adaptive LDPC coded modulation scheme based on parallel subchannels, referred hereafter as *Scheme 1*, is depicted in Figure 6.1 and described in the following.

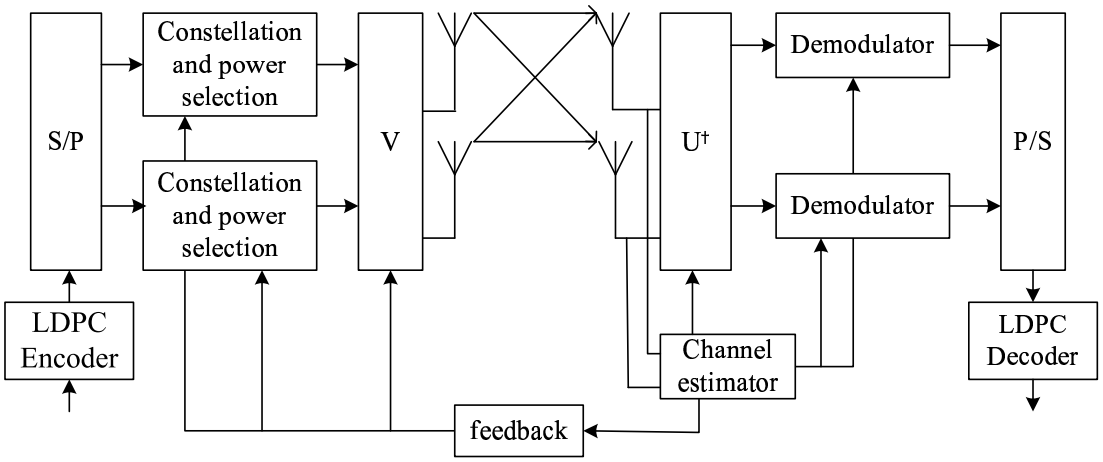


Figure 6.1 Block diagram of the adaptive modulation scheme based on parallel subchannels.

Scheme 1 (Adaptive modulation based on parallel subchannels): The information bit stream is encoded by an LDPC encoder and is divided to parallel substreams. Then, each parallel substream is mapped to a symbol in a selected constellation. The constellation and power level are chosen based on the instantaneous eigenvalues of the subchannels. Thus, the bits are loaded to the parallel subchannels of the MIMO channel. Then, the symbols

are transformed by the matrix $\mathbf{V}$ before sending to the transmit antennas. At the receiver, the received signals are re-transformed by the matrix $\mathbf{U}^{\dagger}$. The output of the transformation block is then demodulated to the information bits.

The signal model of the ith subchannel can be written as:

$$y'_i = h'_i \sqrt{S_i} z'_i + w'_i \tag{6.6}$$

where, $h'_i = \sqrt{\lambda_i}$ is the singular gain of the subchannel i. The variable w'_i is complex additive white Gaussian noise (AWGN) of zero mean and variance $N_0/2$. The variable z'_i represents a symbol of the constellation M_i. The average power of M_i is normalized to be one, while the parameter S_i determines the actual transmitted power spent for symbol z'_i. Note that, the average power for each subchannel, denoted by $\overline{S}_i$, is $1/m$ of the total average power $\overline{S}$ of the system and $\sqrt{S_i} z'_i$ is equivalent to x'_i of $\mathbf{x}'$ in (6.3). The parameters S_i and M_i are chosen based on the statistical properties of the MIMO channel.

The PDF of each eigenvalue can be written as follows [94]:

$$p_{\lambda_i}(\lambda) = \frac{1}{m} \sum_{\kappa=0}^{m-1} \frac{\kappa!}{(\kappa+n-m)!} \left[L_{\kappa}^{n-m}(\lambda) \right]^2 \lambda^{n-m} e^{-\lambda} \tag{6.7}$$

where, $L_p^K(\kappa) = \frac{1}{\kappa!} e^x x^{-K} \frac{d^{\kappa}}{dx^{\kappa}} (e^{-p} x^{K+\kappa})$ is the associated Laguerre polynomial of order κ.

Thus, the signal-to-noise ratio (SNR) of each subchannel is $\gamma_{p,i}[k] = \overline{S}_i \lambda_i[k]/(N_0 B) = \overline{\gamma}_{p,i} \lambda_i[k]$, where B is the bandwidth of the system and $\overline{\gamma}_{p,i} = \overline{S}_i/(N_0 B)$. The PDF of the SNR is:

$$p_{\gamma_{p,i}}(\gamma) = \frac{1}{m \overline{\gamma}_{p,i}} \sum_{\kappa=0}^{m-1} \frac{\kappa!}{(\kappa+n-m)!} \left[L_{\kappa}^{n-m} \left(\frac{\gamma}{\overline{\gamma}_{p,i}} \right) \right]^2 \left(\frac{\gamma}{\overline{\gamma}_{p,i}} \right)^{n-m} e^{-\frac{\gamma}{\overline{\gamma}_{p,i}}} \tag{6.8}$$

This PDF of the SNR plays an important role in the adaptation procedure that shall be presented in Section 6.4.

6.3 Adaptive Modulation Schemes Based on Antenna Selection

As discussed before, the scheme based on parallel subchannels requires the feedback of the CSI corresponding to all the antennas and it also requires a bit-loading procedure to all

the subchannels. Therefore, it is of interest to compare it with other simpler schemes that are based on antenna selection.

The two adaptive coded modulation schemes based on antenna selection (at the transmitter or at both the transmitter and the receiver) are illustrated in Figure 6.2. The description and analysis of each scheme follows.

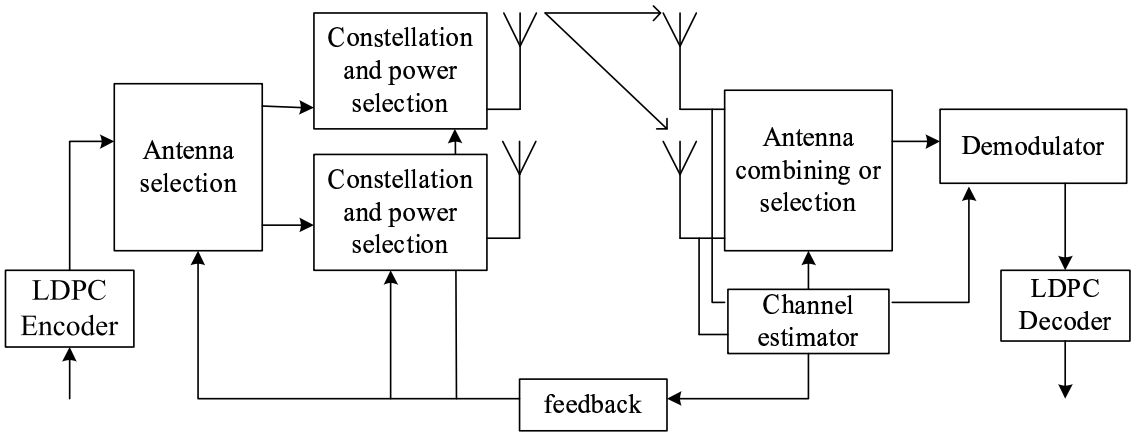


Figure 6.2 Block diagram of adaptive modulation schemes based on antenna selection.

Scheme 2 (Adaptive modulation based on antenna selection at the transmitter): The mathematical model for this scheme is quite different from the one presented in the previous subsection due to the fact that the adaptation parameters should be adjusted to only one virtual channel. When the transmit antenna i is chosen, the received signal vector can be written as follows:

$$\mathbf{y} = \mathbf{h}_i x + \mathbf{w} \tag{6.9}$$

where vectors $\mathbf{y}$ and $\mathbf{w}$ represent the receive signal and noise at the receive antennas. The vector $\mathbf{h}_i$ contains the channel fading coefficients from the ith antenna to all the receive antennas, i.e., the ith column of the matrix $\mathbf{H}$ in (6.1). The variable x represents the transmitted symbol. The maximum-ratio combiner properly co-phases and weights the received signals at the receive antennas as follows:

$$\mathbf{h}_i^\dagger \mathbf{y} = \mathbf{h}_i^\dagger \mathbf{h}_i x + \mathbf{h}_i^\dagger \mathbf{w} \tag{6.10}$$

$$y' = h' x + w' \tag{6.11}$$

where $y' = \mathbf{h}_i^\dagger \mathbf{y}$, $h' = \mathbf{h}_i^\dagger \mathbf{h}_i$ and $w' = \mathbf{h}_i^\dagger \mathbf{w}$. Thus, the multiple-antenna system in this case is equivalent to one channel. The SNR of the maximum ratio combining is:

$$\gamma_{c,i} = \overline{S}\frac{\mathbf{h}_i^\dagger \mathbf{h}_i}{N_0 B} = \sum_{j=1}^{N_r} \gamma_{i,j} \tag{6.12}$$

where $\gamma_{i,j}$ is the SNR at the jth receive antenna. Since this adaptive scheme chooses to transmit over the antenna that has the largest combining SNR at the receiver, the instantaneous channel SNR γ_c is given by

$$\gamma_c = \max_{1 \le i \le N_t} \gamma_{c,i} = \max_{1 \le i \le N_t} \sum_{j=1}^{N_r} \gamma_{i,j} \tag{6.13}$$

Next, the PDF of γ_c can be computed as follows. Each sum $\gamma_{c,i} = \sum_{j=1}^{N_r} \gamma_{i,j}$ is an independent random variable that has a Chi-squared distribution with $2N_r$ degrees of freedom, the expected value $\overline{\gamma}_{c,i} = N_r \overline{\gamma}$ and the variance $2N_r\overline{\gamma}$, where $\overline{\gamma} = \overline{\gamma}_{i,j} = E\{\gamma_{i,j}\}$ is the average SNR of every branch [106]. Thus

$$p_{\gamma_{c,i}}(\gamma) = \frac{\gamma^{N_r - 1} e^{-\gamma/\overline{\gamma}}}{\overline{\gamma}^{N_r}\Gamma(N_r)} U(\gamma) \tag{6.14}$$

The cumulative distribution function (CDF) of $\gamma_{c,i}$ is:

$$F_{\gamma_{c,i}}(\gamma) = \int_0^\gamma \frac{t^{N_r - 1} e^{-t/\overline{\gamma}}}{\overline{\gamma}^{N_r}\Gamma(N_r)} \mathrm{d}t = \frac{1}{\Gamma(N_r)} \mathcal{G}\left(N_r, \frac{\gamma}{\overline{\gamma}}\right) \tag{6.15}$$

Here, the function $\mathcal{G}$ is a lower incomplete Gamma function defined as $\mathcal{G}(\alpha, x) = \int_0^x e^{-t} t^{\alpha - 1} \mathrm{d}t$.

The CDF of γ_c is:

$$F_{\gamma_c}(\gamma) = P(\max[\gamma_{c,1}, \gamma_{c,2}, \ldots, \gamma_{c,N_t}] < \gamma) = \prod_{i=1}^{N_t} F(\gamma_{c,i}) = \frac{1}{\Gamma(N_r)^{N_t}} \mathcal{G}\left(N_r, \frac{\gamma}{\overline{\gamma}}\right)^{N_t} \tag{6.16}$$

which gives the following PDF for γ_c:

$$p_{\gamma_c}(\gamma) = \frac{N_t}{\overline{\gamma}\Gamma(N_r)^{N_t}} \mathcal{G}\left(N_r, \frac{\gamma}{\overline{\gamma}}\right)^{N_t - 1} \left(\frac{\gamma}{\overline{\gamma}}\right)^{N_r - 1} e^{-\frac{\gamma}{\overline{\gamma}}} U(\gamma) \tag{6.17}$$

Scheme 3 (Adaptive modulation based on antenna selections at both the transmitter and the receiver): In this scheme, the antenna selection block at the transmitter chooses the

antenna that has the largest gain of the corresponding wireless channel. At the receiver, selection combining is used. It means that the combiner outputs the signal of the antenna with the highest gain. Therefore, the fading coefficient of the transmitted signal is:

$$h' = \arg \max_{h_{i,j}: 1 \le i \le N_t, 1 \le j N_r} \{|h_{i,j}|\} \tag{6.18}$$

where, $|h_{i,j}|^2$ is the power gain of the channel from the ith transmit antenna to the jth receive antenna. Thus, only one channel with fading coefficient h' needs to be considered.

The PDF of the SNR, $\gamma_s = \overline{S}|h'|^2/(N_0 B)$, can be easily computed from the PDF of $\gamma_{i,j} = \overline{S}|h_{i,j}|^2/(N_0 B)$. The problem is very similar to the problem of selection combining [106]. Each transmission link is also stationary, so that every $\gamma_{i,j}$ has the same PDF $p(\gamma_{i,j})$. Note that, in the adaptive scheme under consideration, the information signal is only transmitted over one diversity branch of the channel. However, the system still knows the instantaneous SNR of each branch (via the pilot signal in the FDD systems or the backward link in the TDD systems). The PDF of $\gamma_{i,j}$ is the PDF of an exponential random variable. That is

$$p_{\gamma_{i,j}}(\gamma) = \frac{1}{\overline{\gamma}} e^{-\gamma/\overline{\gamma}} U(\gamma) \tag{6.19}$$

From (6.18), the instantaneous channel SNR is given by:

$$\gamma_s = \max_{1 \le i \le N_t, 1 \le j \le N_r} \{\gamma_{i,j}\} \tag{6.20}$$

The CDF of γ_s is therefore:

$$F_{\gamma_s}(\gamma) = P(\gamma_s < \gamma) = P\left(\max_{1 \le i \le N_t, 1 \le j \le N_r} \{\gamma_{i,j}\} < \gamma \right) = \prod_{i=1}^{N_t} \prod_{j=1}^{N_r} F_{\gamma_{i,j}}(\gamma) \tag{6.21}$$

If the average SNR $\gamma_{i,j}$ of every branch is the same and equals to $\overline{\gamma}$, the differentiation of (6.21) yields:

$$p_{\gamma_s}(\gamma) = \frac{N_t N_r}{\overline{\gamma}} \left[1 - e^{-\gamma/\overline{\gamma}}\right]^{N_t N_r - 1} e^{-\gamma/\overline{\gamma}} U(\gamma) \tag{6.22}$$

The average SNR of the scheme is:

$$\overline{\gamma}_s = \int_0^\infty \gamma p_{\gamma_s}(\gamma) d\gamma = \overline{\gamma} \sum_{i=1}^{N_t N_r} \frac{1}{i} \tag{6.23}$$

This average SNR is used in the calculation of the bandwidth efficiency of this scheme when the equivalent channel is considered for adaptive modulation in the next section.

6.4 Implementations of Adaptive Modulation Schemes

All the adaptive modulation schemes considered in the previous section are practical discrete-rate and continuous-power adaptations. This section first presents in detail how to implement these schemes for uncoded systems. The adaptive LDPC coded systems are discussed later. Here, procedures to adapt the constellation size and the transmitted power for each diversity scheme are presented. The sizes of signal constellations are powers of 2 and the transmitted power can be any real value. These important parameters of the adaptive modulation schemes dynamically change over time to adapt to the equivalent channels.

Consider the following general model of an equivalent channel:

$$y[k] = h[k]\sqrt{S[k]}z[k] + w[k] \tag{6.24}$$

where k is the time index, $y[k]$, $h[k]$, $w[k]$ represent the received signal, the channel fading coefficient and noise. Note that these variables correspond, respectively, to y' (or y'_i), h' (or h'_i) and w' in (6.6), (6.10) and (6.18). The variable $z[k]$ denotes a symbol of the constellation $M[k]$. The average power of the constellation $M[k]$ is also normalized to one. The parameter $S[k]$ determines the power at time k. Note that the variable $\sqrt{S[k]}z[k]$ corresponds to s in (6.6), (6.10).

At each time index k, the power $S[k]$ and the constellation size $M[k]$ are chosen according to the value of the instantaneous channel SNR γ_p, γ_c or γ_s, which correspond to Schemes 1, 2 and 3, respectively. The selection of $S[k]$ and $M[k]$ should be carried out to maximize the system throughput, while maintaining the required bit error probability level. Though the average bit error rate (BER) criterion can be used, here, the instantaneous bit error rate is considered in our systems.

For uncoded systems, to compute the instantaneous BER for adaptive modulation, the following upper bound for M-QAM constellation over an AWGN channel is often used [25]:

$$\mathrm{BER}(\gamma) \leq 0.2 \exp\left[\frac{-1.5\gamma}{M(\gamma)-1}\frac{S(\gamma)}{\overline{S}}\right] \tag{6.25}$$

where, γ, $S(\gamma)$ and $M(\gamma)$ can be substituted by $\gamma_p[k]$ (or $\gamma_c[k]$, $\gamma_s[k]$), $S[k]$ and $M[k]$. Here, the constellation size and the power are presented as functions of γ since they can be

adapted. Thus, the problem is to find $M(\gamma)$ and $S(\gamma)$ in order to maximize the throughput $E\{\log_2 M(\gamma)\}$. Simultaneously, the BER computed by (6.25) must be guaranteed to be below a target BER and the average power is constrained as:

$$\int_0^\infty S(\gamma)p(\gamma)d\gamma = \overline{S} \tag{6.26}$$

Note that, the PDF of γ for each adaptive scheme (γ_p, γ_c or γ_s) is provided in the previous sections.

For practical implementation, the constellation size M is an integer. The above optimization problem can only be solved numerically. Since finding the optimum solution is difficult, a suboptimum solution is presented in [106] and summarized as follows.

First, it is assumed that the constellation size M is continuous. It means that the data rate of the system can be continuously changed. Although this assumption is impractical, the throughput computed with this assumption provides the limit for the throughput of the practical scheme. Due to this assumption, M can be computed as a function of γ when the target BER is assigned to the right hand side of (6.25). That is

$$M(\gamma) = 1 + \frac{1.5\gamma}{-\ln(5P_b)}\frac{S(\gamma)}{\overline{S}} = 1 + \gamma K\frac{S(\gamma)}{\overline{S}} \tag{6.27}$$

where, $K = \frac{-1.5}{\ln(5P_b)}$. Now, the problem is to maximize:

$$E\left[\log_2 M(\gamma)\right] = \int_0^\infty \log_2\left(1 + K\gamma\frac{S(\gamma)}{\overline{S}}\right)p(\gamma)d\gamma \tag{6.28}$$

with the constraint $\int_0^\infty S(\gamma)p(\gamma)d\gamma = \overline{S}$. The power adaptation policy for the above problem is the well-known water-filling solution:

$$\frac{S(\gamma)}{\overline{S}} = \frac{1}{K}\left(\frac{1}{\gamma_K} - \frac{1}{\gamma}\right)^+ \tag{6.29}$$

with γ_K is the cutoff fade depth, obtained by solving

$$\int_{\gamma_K}^\infty \left(\frac{1}{\gamma_K} - \frac{1}{\gamma}\right)p(\gamma)d\gamma = K \tag{6.30}$$

Substitute (6.29) into (6.27) and (6.28), to obtain the instantaneous rate as:

$$\log_2 M(\gamma) = \log_2(\gamma/\gamma_K) \tag{6.31}$$

and the average spectral efficiency as:

$$\frac{R}{B} = \int_{\gamma_K}^{\infty} \log_2 \left(\frac{\gamma}{\gamma_K} \right) p\left(\gamma\right) d\gamma \tag{6.32}$$

where R is the bit rate.

Now, the case of discrete M is considered. Assume that M can only be one of N discrete values, namely $M_0 = 0, M_1 = 2, \ldots, M_{N-1} = 2^{2(N-1)}$. The constellation size is determined for a given γ by discretizing the range of the channel fade levels. It means that the range of γ is divided into N rate regions $R_j = [\gamma_{j-1}, \gamma_j)$ with $0 \leq j \leq N-1$, where $\gamma_{-1} = 0$ and $\gamma_{N-1} = \infty$. Then the constellation M_j is transmitted when $\gamma \in R_j$. The problem becomes to find the optimum boundaries of the rate regions. The computational complexity to obtain the optimum solution for this integer programming problem is very high. Therefore, a suboptimum solution is further devised based on the results of the continuous-rate case.

For the continuous-rate case, the optimum constellation size M is a linear function of γ. Hence, to find the boundaries of the rate regions, a linear function $M(\gamma) = \frac{\gamma}{\gamma_K^*}$ is defined. The coefficient γ_K^* is chosen later for optimization. For a rate region R_j ($j < 0 < N-1$), the boundaries are M_j and M_{j+1}. For the γ value in the region R_j, i.e., $M_j \leq M(\gamma) < M_{j+1}$, the constellation size M_j is chosen. Thus, the largest constellation size, which is smaller than $M(\gamma)$, is transmitted.

From (6.27), the power adaptation policy is given by:

$$\frac{S_j(\gamma)}{\overline{S}} = \begin{cases} (M_j - 1)\frac{1}{\gamma K}, & M_j < \frac{\gamma}{\gamma_K^*} \leq M_{j+1} \\ 0, & M_j = 0 \end{cases} \tag{6.33}$$

The final step is to find γ_K^* in order to maximize the following average spectral efficiency:

$$\frac{R}{B} = \sum_{j=1}^{N-1} \log_2(M_j) P\left(M_j \leq \frac{\gamma}{\gamma_K^*} \leq M_{j+1} \right) = \sum_{j=1}^{N-1} \log_2(M_j) \int_{\gamma_K^* M_j}^{\gamma_K^* M_{j+1}} p(\gamma) d\gamma \tag{6.34}$$

under the power constraint

$$\sum_{j=1}^{N-1} \int_{\gamma_K^* M_j}^{\gamma_K^* M_{j+1}} \frac{S_j(\gamma)}{\overline{S}} p(\gamma) d\gamma = 1 \tag{6.35}$$

Unfortunately, there is no closed-form solution for this problem, and a numerical technique needs to be used.

Thus, using the PDFs of the SNRs derived in the previous sections to solve the optimization problem with the objective function in (6.34) and the constraint in (6.35) yields the parameter γ_K^*. With the solution of γ_K^*, the constellation size and the transmitted power for each adaptive modulation scheme can be easily determined as described before.

For LDPC coded systems, the BER bound cannot be presented in a closed-form expression such as (6.25). Therefore, simulation results of the error performance is first utilized. The LDPC codes are often decoded by the sum-product algorithm [5]. The error performance results of LDPC coded modulation systems in Chapter 3 show that performance of the sum-product algorithm is quite close to the maximum likelihood bound at the practical BER of data communications (10^{-6}) for systems with medium-length codes. In this chapter, the performance bound for the BER of LDPC coded modulation is also used to compute the rate regions. The average spectral efficiency corresponding to these rate regions shall be referred to as "achievable" rate threshold because this rate threshold can be achieved by using a decoder that performs better than the sum-product decoder, for example, using a combination between a reliability-based decoder and a sum-product decoder [122].

The LDPC coded system under consideration varies only the modulation size. Hence, the code and the constellation do not change in each rate region. This means that the BER target is maintained by only adapting the transmitted power. The power adaptation rule in a rate region must be the channel inversion strategy. That is $S(\gamma) = \frac{\gamma_j^*}{\gamma}\overline{S}$, where γ_j^* is the received SNR required to achieve the target BER with this constellation. The value of γ_j^* is obtained from the BER performance of the coded modulation system over an equivalent AWGN channel based on simulation or bounding technique. Our optimization problem is to find $\gamma_0, \ldots, \gamma_N$ in order to maximize the average throughput $\frac{R}{B} = \sum_{j=1}^{N} R_c \log_2 M_j \int_{\gamma_{j-1}}^{\gamma_j} p(\gamma) \mathrm{d}\gamma$ with the power constraint, where R_c is the code rate. This optimization problem is a non-linear optimization. The Karush-Kuhn-Tucker condition is often not satisfied for various PDF functions of MIMO channel's SNR. Here, a numerical method was used to solve the problem.

6.5 Numerical Results

Figure 6.3 shows the average spectral efficiency of uncoded MIMO systems with different adaptive schemes for the case of continuous-rate adaptation. Although continuous-rate adaptation is impractical, these results show ultimate limits that can be achieved by the corresponding discrete-rate schemes. The BER target for numerical results in this section is set at 10^{-3}, except for the results in Fig. 6.7.

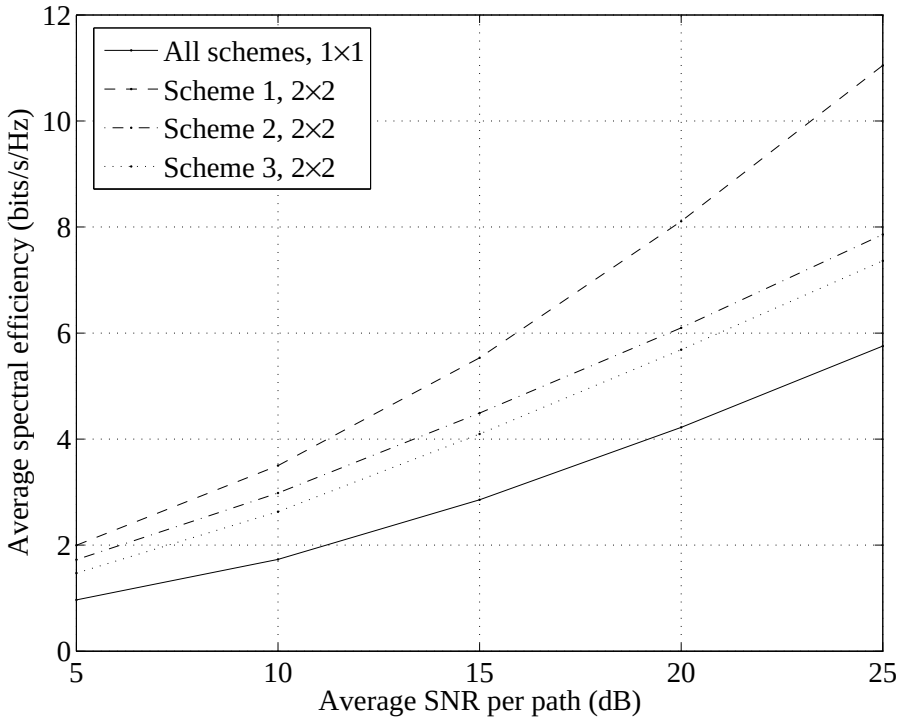


Figure 6.3 Spectral efficiency of different adaptive schemes for the case of continuous-rate.

Consider an uncoded MIMO system with two transmit antennas and two receive antennas. Figure 6.3 shows that Scheme 1 offers the highest spectral efficiency among all the schemes considered, whereas Scheme 2 is superior to Scheme 3. When the SNR increases, the gap between Scheme 1 and Scheme 2 increases. However, the gap between Schemes 2 and 3 appears to be almost the same over the whole range of SNR. Of course, the efficiency of all three schemes are identical if there are only one transmit and one receive antenna.

It is appropriate to mention here that the spectral efficiency of Scheme 1 is different from the Rayleigh channel capacity computed in [94]. Here, the spectral efficiency is

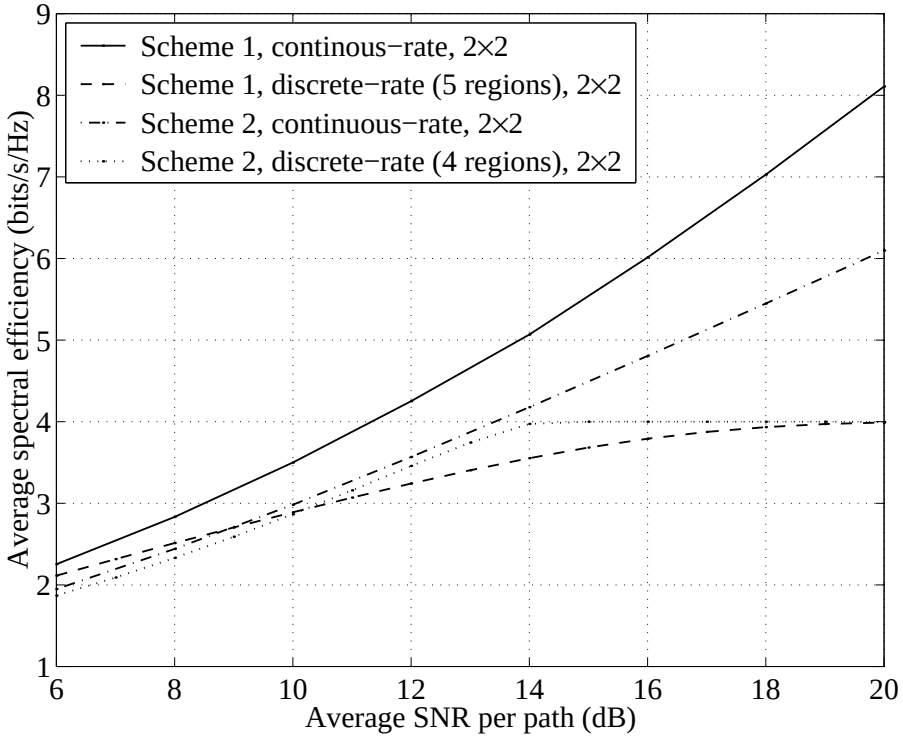


Figure 6.4 Average spectral efficiency of discrete-rate systems with Scheme 1 (3 rate regions for each subchannel, 5 rate modes for systems) and Scheme 2 (4 rate regions for the equivalent channel): 2 transmit antennas and 2 receive antennas.

computed based on the error bound of QAM and for the target BER of 10^{-3}. On the other hand, the computation of the channel capacity implies the use of an ideal code and at an arbitrarily small BER. Moreover, the channel capacity of [94] is computed for the case of no knowledge of the CSI at the transmitter. When the CSI is not available at the transmitter, the transmitted power is diverged over all directions so that only a power of $\frac{m}{N_t}$ is used for each parallel subchannel. This fact can be observed from (6.3), where only m symbols in N_t symbols of $\mathbf{x}'$ correspond to m non-zero singular values in the diagonal matrix $\mathbf{D}$.

Figure 6.5 illustrates the increment of the average spectral efficiency when the number of transmit antenna increases. All the systems are equipped with two receive antennas. Observe that for Scheme 2 and Scheme 3, the average throughput only slightly increases when increasing the number of transmit antennas. In contrast, the spectral efficiency of Scheme 1 quickly increases with the number of transmit antennas. At high SNR, the spectral efficiency of Scheme 1 tends to saturate when the number of transmit antennas is larger than

the number of receiver antennas. Recall that the number of parallel subchannels in Scheme 1 is limited by the number of transmit antennas when the number of transmit antennas is increased beyond the number of receive antennas. If there is only 1 transmit antenna, Scheme 1 and Scheme 2 are identical. Moreover, the gap between Scheme 2 and Scheme 3 seems to be a constant over both the SNR and the number of transmit antennas.

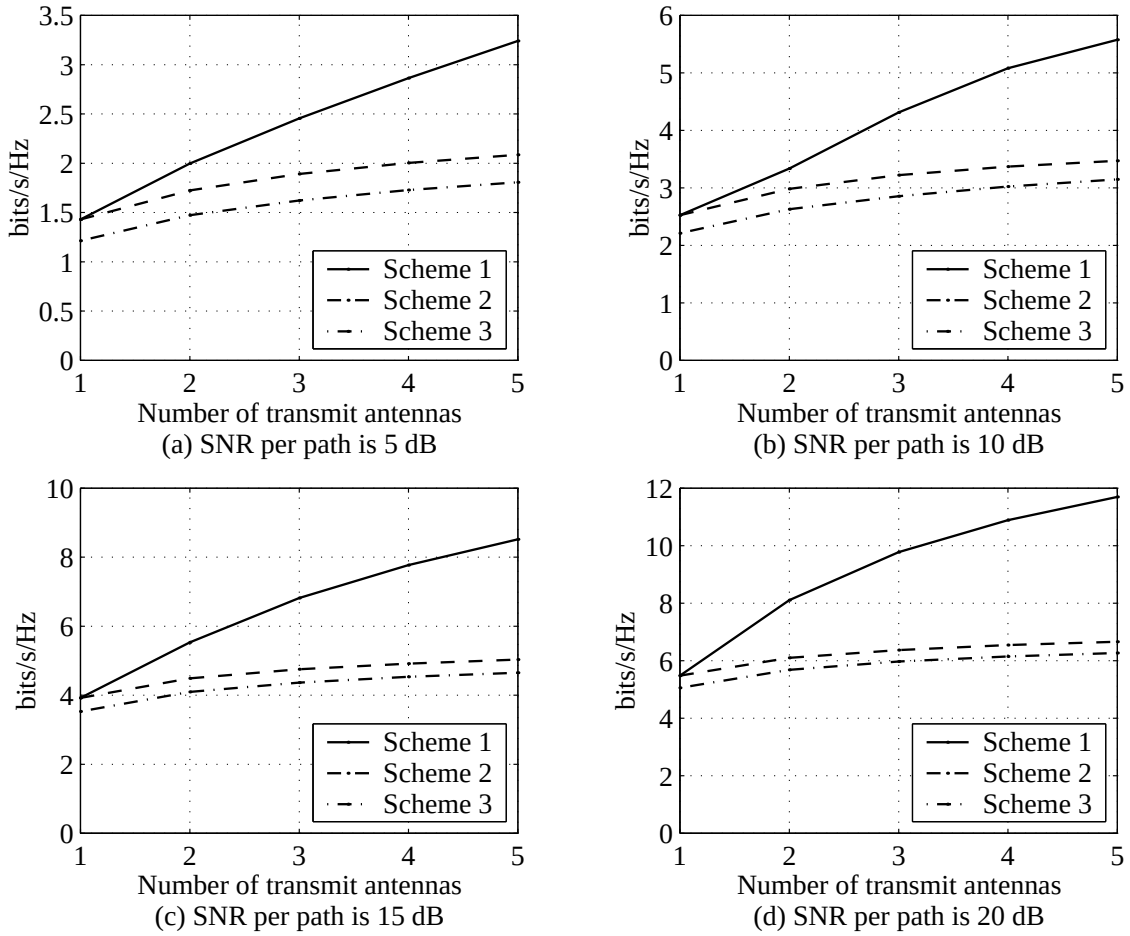


Figure 6.5 Average spectral efficiency as a function of the number of transmit antennas for continuous-rate adaptive schemes. The number of receive antenna is 2.

The average spectral efficiency for discrete-rate systems is shown in Fig. 6.4. For the ease of comparison the efficiency of continuous-rate systems is also provided. Here, all the systems use two transmit and two receive antennas. Observe that there is a significant gap in spectral efficiency between Scheme 1 and Scheme 2 for the case of continuous-rate. But, interestingly, these two schemes are comparable for the discrete-rate systems as far as the maximum data rate is concerned. At low SNR range, the spectral efficiency of Scheme 1

is higher than that of Scheme 2. However, at the SNR range closer to that of the maximum rate, the spectral efficiency of Scheme 2 is higher than that of Scheme 1. These results and observation can be explained as follows.

Recall that, the spectral efficiency of Scheme 1 is a sum of two parallel subchannels. Here, each channel can only employ 3 rate regions of 0, 1 and 2 loaded bits. For the whole system, the number of rate regions is 5 which are from 0 to 4 loaded bits. On the other hand, Scheme 2 employs 4 rate regions of 0, 1, 2 and 4 loaded bits. Thus, with the same maximum rate, the possible number of rate regions for parallel channels is always smaller than the number of rate regions in selection diversity scheme. This is precisely the reason for spectral efficiency reduction when compared to the potential capacity of the continuous-rate adaptation. Another reason is due to the different forms of the probability density functions of the SNRs in the three schemes. Fig. 6.6 shows the PDFs of the SNRs for the equivalent channels of all three adaptive schemes, where the same average SNR per path is set at 10dB. Although the sum of the continuous-rate throughput of Scheme 1's two subchannels is higher than the continuous-rate throughput of Scheme 2, the large part of continuous-rate throughput of Scheme 1 is resulted from the low SNR range that has a high density for Scheme 1's subchannels. When only an integer number of discrete-rates is employed, this part of throughput will be smaller because the power allocation will not be optimum in order to compensate for the high density of this low SNR range.

Figure 6.7 shows the gap in spectral efficiency between Scheme 1 and Scheme 2 for discrete-rate systems. The systems are either 2×2 or 3×3 MIMO systems. The gap of spectral efficiency is considered at two BER targets of 10^{-3} and 10^{-6}. For the BER target of 10^{-3}, the spectral efficiency of Scheme 1 is superior to that of Scheme 2 in the SNR range from 6 dB to 10 dB and inferior to that of Scheme 2 in the SNR range from 10 dB to 20 dB. Observe that the size of the gap between the two schemes is the same for different BER targets. However, the curves are shifted to the right for a lower BER target because it requires a higher SNR to reach the same throughput of a better reliability. When the numbers of transmit and receive antennas increase from 2 to 3, the spectral efficiency of Scheme 1 increases faster compared to Scheme 2. The SNR range, in which Scheme 2 is

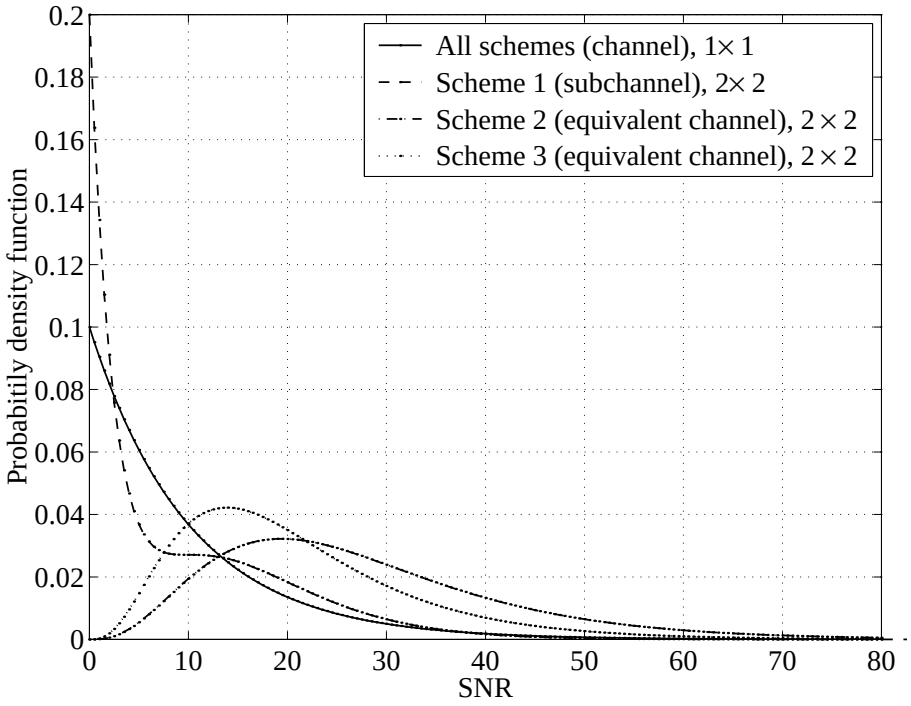


Figure 6.6 Probability density functions of the SNRs for equivalent channels in three diversity schemes with the same average SNR per path.

superior to Scheme 1, is smaller.

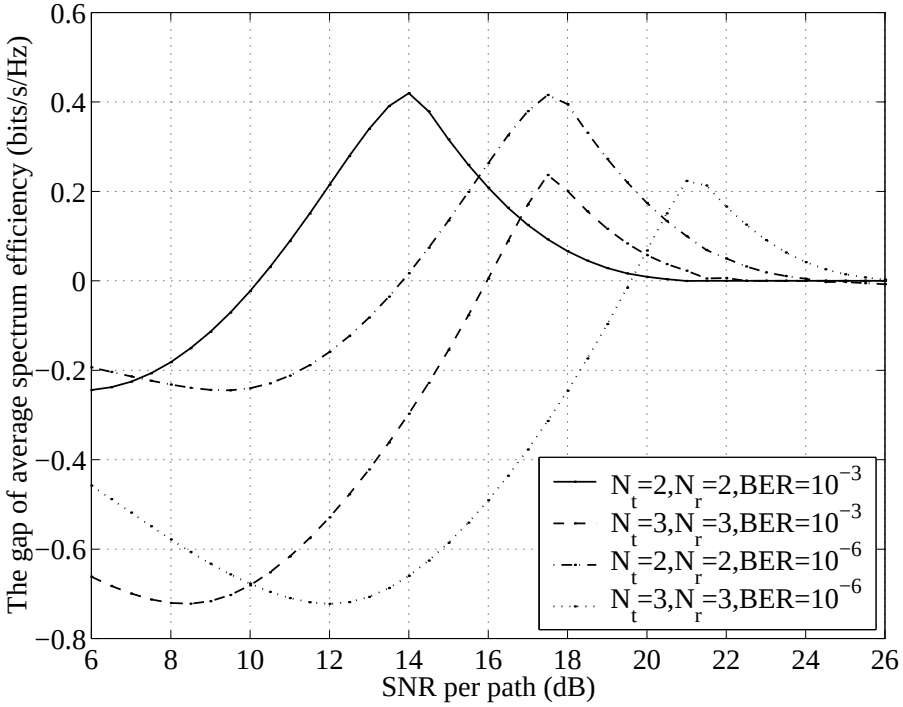


Figure 6.7 Spectral efficiency of Scheme 1 over Scheme 2 for discrete-rate systems.

Table 6.1 Required SNR of coded modulation systems with a regular LDPC code of rate 433/495 and length 495 to achieve BER of 10^{-6}.

M	Modulation	Spectral efficiency	E_b/N_0 (dB)		SNR (dB)	
			simulation	bound	simulation	bound
1	BPSK	0.85	6.3	4.7	5.59	3.99
2	QPSK	1.75	6.3	4.7	8.73	7.13
3	8PSK	2.62	9.4	7.7	13.58	11.88
4	16QAM	3.5	10.6	8.3	16.04	13.74

For adaptive coded modulation systems, a regular (3,6) LDPC code with a high rate of 433/495 and a length of 495 coded bits is employed. Because the constellation often remains constant over tens to hundreds of symbols for typical adaptive modulation systems [106], this code length is proper for practical implementation. For the target BER error of 10^{-6}, the required E_b/N_0 levels (where E_b is the average energy of an information bit) of the coded modulation system are tabulated in Table 6.1 for different constellation sizes. Note that, the channel SNR is E_s/N_0 which is equivalent to $R_c \log_2(M) E_b/N_0$. The required E_b/N_0 levels can also be obtained by maximum likelihood bound (ML bound).

Figure 6.8 shows the practical spectral efficiency, the achievable rate thresholds and the channel capacity of an adaptive coded modulation system using 2 transmit and 2 receive antennas. Here, each subchannel of Scheme 1 employs 3 rate regions, corresponding to QPSK, BPSK and no transmission. Similar to uncoded systems, the number of possible rates are 5 when combining subchannels rates. Scheme 2 is considered with 3 rate regions, corresponding to 0, 2 and 4 loaded bits. It is observed that the spectral efficiency of Scheme 2 is superior to that of Scheme 1 over the whole considered range of SNR. This fact can be explained by a better bit loading in this SNR range of Scheme 2 for coded systems. The gaps between the achievable rate threshold (based on ML bounds) and the practical spectral efficiency (based on the sum-product decoder) are about 1.7 dB at the average spectral efficiency of 2 bits/s/Hz for both schemes. Due to the use of finite-length LDPC codes, the gap between the channel capacity and the practical spectral efficiency of both schemes is quite large.

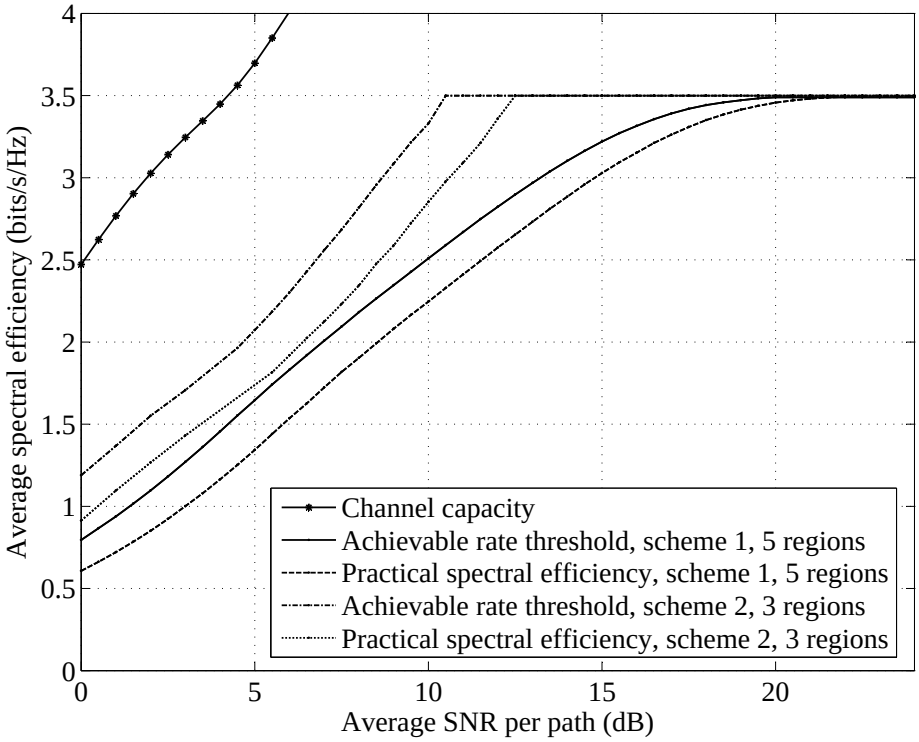


Figure 6.8 Spectral efficiency of adaptive LDPC coded systems with 2 transmit and 2 receive antennas.

6.6 Conclusions

Comparison of spectral efficiency among different adaptive antenna diversity schemes based on antenna beamforming and antenna selection has been carried out in this chapter. The average spectrum efficiency is investigated for various parameters of the systems, including the average SNR of a path, the numbers of transmit and receive antennas, and the BER target level. For continuous-rate systems, the scheme based on parallel subchannels clearly outperforms other schemes based on antenna selection. However, these two diversity techniques are comparable for practical discrete-rate systems. Over a certain range of SNR, the scheme with antenna selection at the transmitter and maximum ratio combining at the receiver is, even, superior to the beamforming scheme when the number of transmit antennas is small. Adaptive LDPC coded modulation systems also exhibit similar results because of the better bit loading in transmit antenna selection schemes. Given the reduction in feedback information and simplicity, the antenna selection schemes can be attractive candidates for systems of appropriate parameters.

7. Conclusions

7.1 Summary of Contributions

This thesis was mainly devoted to the analysis of bandwidth-efficient communication systems based on finite-length LDPC codes. In Chapters 3 and 4, two basic bandwidth-efficient LDPC coded modulation schemes, namely LDPC coded modulation and multi-level LDPC coded modulation, were investigated for the AWGN and frequency-flat Rayleigh fading channels. In these schemes, the coded bits of LDPC codes are grouped and mapped to 2-dimentional signal constellations, i.e., multiple-amplitude/multiple-phase constellations. At the receiver, three types of decoders, including the maximum likelihood (ML), the sum-product and the ordered-statistic decoders are considered. Performance of the ML receiver is analyzed by the union bound. On the other hand, the performance of the receivers based on the sum-product and ordered-statistic decoders are investigated by computer simulation. For LDPC coded modulation systems in Chapter 3, the main contribution is the derivation of the performance bound for the ML receiver. The derived bound is mainly based on the union bounding technique. However, due to the random construction of LDPC codes, this performance bound can only be obtained by averaging over the randomly permuted code ensemble. Here, combinatoric techniques are used to transform the Hamming distance spectrum of the LDPC code and the Euclidean distance profile of the constellation to the Euclidean distance spectrum of the LDPC coded modulation. This ML bound is tight for the receiver based on OSD, hence, it can be used to benchmark the LDPC coded modulation system using OSD. For LDPC coded modulation system based on the sum-product decoding, the ML bound is also close to the simulation result at low BER and for medium-length codes. The iterative demodulation/decoding are carried out

for OSD and sum-product decoder, but improvement due to this outer iteration is negligible. In Chapter 4, the derivation framework developed in Chapter 3 is extended to the generalized multilevel LDPC coded modulation systems. In these systems, more than one coded bit of one encoder is used for each level. Although the derivation of the bound for each level is similar to that of coded modulation, the computation of the Euclidean distance profile of the constellation is different due to the effect on other levels to the current level under consideration. Simulation results with both OSD and sum-product decoding show the tightness of the bounds for all decoding levels over the AWGN channel as well as the Rayleigh fading channel. Therefore, these performance bounds can be used as tools in choosing parameters of systems such as the constellation, mapping and code rate.

The second part of the thesis, including Chapters 5 and 6, focuses on LDPC coded MIMO systems that are important to further improve the bandwidth efficiency of wireless data transmission. Two scenarios of wireless channels considered are fast fading and slow fading which correspond to high and low mobility speeds of the wireless terminals. Chapter 5 studies LDPC coded unitary space-time modulation systems, where the channel state information is unknown to both the transmitter and the receiver. Therefore, the unitary space-time constellations, that are designed and optimized for MIMO systems in this scenario are employed at the transmitter. For LDPC coded unitary space-time modulation systems, the combinatoric techniques in Chapter 3 can still be applied to derive the performance bound. However, the procedure needs to transform the Hamming distance spectrum of the finite-length LDPC code to a summation of chordal distances in Grassmanian manifolds instead of the the Euclidean distance as in the case of LDPC coded modulation. The tightness of the derived bound is also confirmed by the performances of the OSD and the sum-product decoding. The situation that the CSI is known at both the transmitter and the receiver is considered in Chapter 6. In this chapter, the adaptive LDPC coded modulation technique is applied for multiple transmit and multiple receive antenna systems. Three diversity schemes based on antenna beamforming and antenna selection are investigated and compared in terms of throughput. For discrete-rate adaptive modulation, computation of bandwidth efficiencies of these diversity schemes shows a better performance for

the scheme with antenna selection at the transmitter and antenna combining at the receiver when the number of antennas is small. For adaptive LDPC coded MIMO systems, a threshold of bandwidth efficiency is determined from the ML bound of LDPC coded modulation system derived in Chapter 3. This threshold can be achieved by LDPC decoders whose performance is close to that of the ML decoder.

7.2 Suggestions for Further Studies

The derivation framework of the performance bound can be applied to many other bandwidth-efficient communication systems over various wireline as well as wireless channels. For the frequency-flat Rician or Nakagami fading channels, where the adjacent symbols are not overlapped, the difference in the derivation is mainly due to the pairwise error probabilities. For frequency-selective fading channels, the consecutive symbols are overlapped and more complicated combinatoric techniques need to be used.

As discussed in Chapter 3, the union bound is considered to be tight in the SNR range above the threshold $[E_b/N_0]^*$ corresponding to the cut-off rate. Other bounding techniques are of interest for performance evaluation at lower SNR. For coded modulation with BPSK constellation, tangential sphere bound [52] and Gallager's bound [5,53,54] are tight bounds at the low SNR range. The tangential sphere bound is also used for performance evaluation of M-PSK coded modulation scheme constructed from algebraic codes [123]. For the derivation of these bounds, all terms of the Hamming distance spectrum of the code and the corresponding Euclidean distance spectrum of the coded modulation frame need to be known. However, when these bounds are applied to coded modulation systems based on pseudo-random codes such as LDPC codes, the number of Euclidean distance terms of the coded modulation frame corresponding to a Hamming distance can be very large when the Hamming distance is big. This means that the computation of the bound is very difficult, if not impossible. An approximation method can be a solution to this problem of computational complexity.

The tangential sphere bound and Gallager's bound are developed based on the union bound, which is the simplest inequality from a larger class of the so-called Bonferroni-type

inequalities in probability theory. The bounding techniques using a high-order Bonferroni-type inequality can improve the tightness of the bound as shown in [124, 125] for coded systems with BPSK modulation. The application of high-order Bonferroni-type inequalities for coded modulation system with high-order constellation and pseudo-random codes might yield the tighter bound in a wider range of SNR. However, the bounds derived from high-order Bonferroni-type inequalities are often complicated and requires the global geometrical properties of the code.

Chapters 5 and 6 of this thesis studied the bandwidth-efficient LDPC coded MIMO communication systems, where the space diversity and time diversity are emphasized. Further studies can be carried out with frequency diversity techniques, i.e., using multi-carrier techniques such as orthogonal frequency division multiplexing (OFDM) or discrete multitone (DMT). These techniques have proved to have immunity to impulse noise and frequency-selective fading. For frequency-selective fading channels, the multi-carrier techniques reduce the rate of data stream in each frequency band to make the channel over each frequency band become a frequency flat channel. To analyze LDPC coded multi-carrier systems, the symbols sent over the multiple dimensions of the carriers should be considered together. The LDPC coded communication systems, that combine the methods of frequency and space diversities, for example MIMO-OFDM systems are also of interest to analyze.

In all bandwidth-efficient LDPC coded communication systems, a group of coded bits is mapped to a signal symbol before transmitting over the channel. One can expect that LDPC codes constructed over Galois fields of order $Q > 2$ [126] can have some advantages. For this system, a group of q coded bits is mapped to one symbol of the 2^q-QAM constellation. The application of LDPC codes in $GF(Q)$ and a proper mapping of q coded bits to 2^q-QAM constellation can improve the smallest term of the Euclidean distance spectrum of the coded modulation frame, hence improving the performance. The disadvantage of LDPC codes in $GF(Q)$ is the high decoding complexity of the sum-product algorithm. The analysis of such systems is of interest in order to compare with the performance of the coded modulation systems based on binary LDPC codes.

The practical implementation of the bandwidth-efficient LDPC coded communication systems in hardware is also a natural topic for further studies. The hardware implementation of these systems can be an application-specific integrated circuits (ASIC) or a program of hardware description language (VHDL or Verilog) for the field programmable gate arrays (FPGA) [127]. The FPGA implementation of encoder/decoder algorithms is becoming more suitable for high-speed real-time applications and easily integrated to other intellectual properties of the systems on chip (SoC). In these implementations, the algorithms of the encoder/decoder need to be appropriately discretized instead of using real numbers as in the original algorithms.

A. Proof of the Check Node Processing Equation

This appendix presents the proof of the check node processing equation in Chapter 2. To prove Equation (2.24), the following lemma is needed:

Lemma: Consider a sequence of m independent binary digits in which the lth digit is a 1 with probability P_l. Then the probability that an even number of digits are 1 is:

$$\frac{1+\prod_{l=1}^{m}(1-2P_l)}{2} \tag{A.1}$$

Proof of the Lemma: [5] Consider the function

$$\prod_{l=1}^{m}(1-P_l+P_lt) \tag{A.2}$$

Observe that if this is expanded into a polynomial in t, the coefficient of t^i is the probability of i 1's. The function $\prod_{l=1}^{m}(1-P_l-P_lt)$ is identical except that all the odd powers of t are negative. Adding these two functions, all the even powers of t are doubled, and the odd terms cancel out. Finally letting $t=1$ and dividing by 2, the result is the probability of even number of ones. But

$$\frac{\prod_{l=1}^{m}(1-P_l+P_l)+\prod_{l=1}^{m}(1-P_l-P_l)}{2}=\frac{1+\prod_{l=1}^{m}(1-2P_l)}{2} \tag{A.3}$$

Thus, the lemma is proved.

Now, when the incoming message q_{ij}, corresponding to the probabilities $P_{q_{ij}}(v_i=0)$ and $P_{q_{ij}}(v_i=1)$ come to a check node c_j, we need to compute r_{ij}, corresponding to the extrinsic probabilities $P_{r_{ij}}(v_i=0)$ and $P_{r_{ij}}(v_i=1)$, from a set of $q_{i'j}$ with $i' \subset I(j) \setminus i$, that contains variable nodes connected to check node c_j except the check node v_i. Using

the Lemma, one has:

$$P_{r_{ij}}(v_i = 0) = P(v_i = 0 | q_{i'j}, c_i = 0) \quad \text{(A.4)}$$

$$= P(\text{even number of 1s in } I(j) \setminus i) \quad \text{(A.5)}$$

$$= \frac{1 + \prod_{i' \in I(j) \setminus i} (1 - 2P_{q_{i'j}}(v'_i = 1))}{2} \quad \text{(A.6)}$$

and

$$P_{r_{ij}}(v_i = 1) = 1 - P(\text{even number of 1s in} I(j) \setminus i) \quad \text{(A.7)}$$

$$= \frac{1 - \prod_{i' \in I(j) \setminus i} (1 - 2P_{q_{i'j}}(v'_i = 1))}{2} \quad \text{(A.8)}$$

Therefore,

$$r_{ij} = \log \frac{P_{r_{ij}}(v_i = 0)}{P_{r_{ij}}(v_i = 1)} \quad \text{(A.9)}$$

$$= \log \frac{1 + \prod_{i' \in I(j) \setminus i} (1 - 2P_{q_{i'j}}(v'_i = 1))}{1 - \prod_{i' \in I(j) \setminus i} (1 - 2P_{q_{i'j}}(v'_i = 1))} \quad \text{(A.10)}$$

The derivation to (2.24), i.e.,

$$r_{ij} = \log \frac{1 + \prod_{i' \in I(j) \setminus i} \tanh(q_{i'j}/2)}{1 - \prod_{i' \in I(j) \setminus i} \tanh(q_{i'j}/2)}$$

is straightforward with $q_{i'j} = \log \frac{P_{q_{i'j}}(v'_i = 0)}{P_{q_{i'j}}(v'_i = 1)} = \log \frac{1 - P_{q_{i'j}}(v'_i = 1)}{P_{q_{i'j}}(v'_i = 1)}$ and $\tanh(q_{i'j}/2) = \frac{e^{q_{i'j}} - 1}{e^{q_{i'j}} + 1}$.

B. An Example of Check Node Processing

This appendix gives an example of the step-by-step procedure of check node processing in order to illustrate the mechanism of the belief propagation algorithm.

Assume that a check node c_1 is connected to three variable nodes v_1, v_2, v_3. When variable nodes v_1 and v_2 send messages $q_{1,1}$ and $q_{2,1}$, respectively, a message $r_{3,1}$ needs to be computed at the check node c_1 as in Figure B.1. Messages $q_{1,1}, q_{2,1}$ and $r_{3,1}$ are extrinsic LAPPRs of variable nodes v_1, v_2 and v_3, respectively.

It is difficult to imagine the mechanism of check node processing in LAPPRs. Thus, the probabilities are used to present the step-by-step procedure. The probabilities are computed from LAPPRs as follows:

- Probability of the variable node $v_1 = 0$ given by $q_{1,1}$ is $P_{q_{1,1}}(v_1 = 0) = \frac{e^{q_{1,1}}}{1+e^{q_{1,1}}}$. This is because $q_{1,1} = \log \frac{P_{q_{1,1}}(v_1=0)}{P_{q_{1,1}}(v_1=1)} = \log \frac{P_{q_{1,1}}(v_1=0)}{1-P_{q_{1,1}}(v_1=0)}$.
- Probability of the variable node $v_1 = 1$ given by $q_{1,1}$ is $P_{q_{1,1}}(v_1 = 1) = \frac{1}{1+e^{q_{1,1}}}$, since $q_{1,1} = \log \frac{P_{q_{1,1}}(v_1=0)}{P_{q_{1,1}}(v_1=1)} = \log \frac{1-P_{q_{1,1}}(v_1=1)}{P_{q_{1,1}}(v_1=1)}$.
- Probability of the variable node $v_2 = 0$ given by $q_{2,1}$ is $P_{q_{2,1}}(v_2 = 0) = \frac{e^{q}_{2,1}}{1+e^{q_{2,1}}}$, since $q_{2,1} = \log \frac{P_{q_{2,1}}(v_2=0)}{P_{q_{2,1}}(v_2=1)} = \log \frac{P_{q_{2,1}}(v_2=0)}{1-P_{q_{2,1}}(v_2=0)}$.

Table B.1 Example of Probabilities Values.

v_1	v_2	v_3
$q_{1,1}$	$q_{2,1}$	$r_{3,1}$
$P_{q_{1,1}}(v_1 = 0) = 0.7$	$P_{q_{2,1}}(v_2 = 0) = 0.4$	$P_{r_{3,1}}(v_3 = 0) =?$
$P_{q_{1,1}}(v_1 = 1) = 0.3$	$P_{q_{2,1}}(v_2 = 1) = 0.6$	$P_{r_{3,1}}(v_3 = 1) =?$

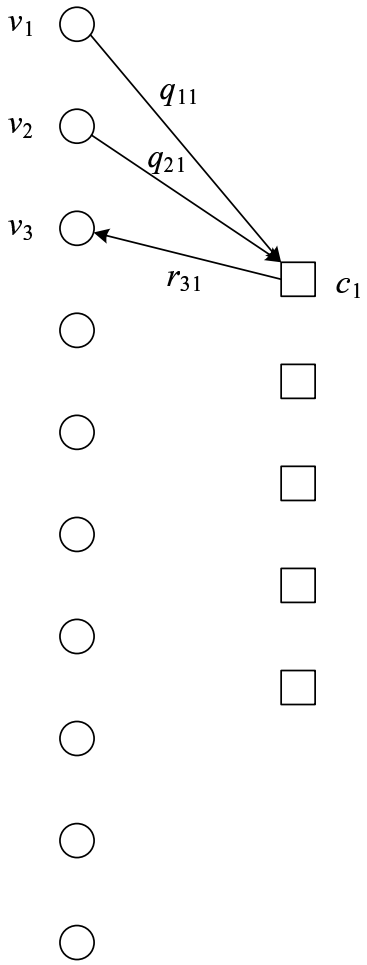


Figure B.1 Example of message processing at the check node c_1.

- Probability of the variable node $v_2 = 1$ given by $q_{2,1}$ is $P_{q_{2,1}}(v_1 = 1) = \frac{1}{1+e^{q_{2,1}}}$, since $q_{2,1} = \log \frac{P_{q_{2,1}}(v_2=0)}{P_{q_{2,1}}(v_2=1)} = \log \frac{1-P_{q_{2,1}}(v_2=1)}{P_{q_{2,1}}(v_2=1)}$

Now, for example, we have the values of $P_{q_{1,1}}(v_1 = 0), P_{q_{1,1}}(v_1 = 1), P_{q_{2,1}}(v_2 = 0), P_{q_{2,1}}(v_1 = 1)$ as in Table B.1. Our objective is to compute the probabilities $P_{r_{3,1}}(v_3 = 0), P_{r_{3,1}}(v_3 = 1)$ when $v_1 + v_2 + v_3 = 0$ by the parity check equation at check node c_1. All the possibilities that can have $v_1 + v_2 + v_3 = 0$ are listed in Table B.2.

The probability of the variable node $v_3 = 0$ given by $r_{3,1}$ is $P_{r_{3,1}}(v_3 = 0) = 0.28 + 0.18 = 0.46$. The probability of the variable node $v_3 = 1$ given by $r_{3,1}$ is $P_{r_{3,1}}(v_3 = 1) = 0.42 + 0.12 = 0.54$. Then, the message $r_{3,1} = \log \frac{P_{r_{3,1}}(v_3=0)}{P_{r_{3,1}}(v_3=1)} = \log \frac{0.46}{0.54} = -0.16$.

Table B.2 Probabilities of all cases satisfying the parity check equation.

v_1	v_2	v_3	Probability
0	0	0	$0.7 \times 0.4 = 0.28$
0	1	1	$0.7 \times 0.6 = 0.42$
1	0	1	$0.3 \times 0.4 = 0.12$
1	1	0	$0.3 \times 0.6 = 0.18$

So, $r_{3,1} = -0.16$ has been computed from $q_{1,1} = \log \frac{P_{q_{1,1}}(v_1=0)}{P_{q_{1,1}}(v_1=1)} = \log \frac{0.7}{0.3} = 0.84$ and $q_{2,1} = \log \frac{P_{q_{2,1}}(v_2=0)}{P_{q_{2,1}}(v_2=1)} = \log \frac{0.4}{0.6} = -0.41$.

C. The Hamming Distance Spectra of Several LDPC Codes

1. The regular (3,6) LDPC code of rate 1/2 and length 24:

$$\mathcal{N}(4) = 8; \qquad \mathcal{N}(14) = 972;$$
$$\mathcal{N}(6) = 52; \qquad \mathcal{N}(16) = 367;$$
$$\mathcal{N}(8) = 367; \qquad \mathcal{N}(18) = 52;$$
$$\mathcal{N}(10) = 972; \qquad \mathcal{N}(20) = 8;$$
$$\mathcal{N}(12) = 1296; \qquad \mathcal{N}(24) = 1;$$

2. The regular (3,6) LDPC code of rate 1/2 and length 72:

$$\mathcal{N}(6) = 1; \qquad \mathcal{N}(38) = 11534400347;$$
$$\mathcal{N}(8) = 28; \qquad \mathcal{N}(40) = 8314036051;$$
$$\mathcal{N}(10) = 246; \qquad \mathcal{N}(42) = 4793685822;$$
$$\mathcal{N}(12) = 2267; \qquad \mathcal{N}(44) = 2198313637;$$
$$\mathcal{N}(14) = 21536; \qquad \mathcal{N}(46) = 797876668;$$
$$\mathcal{N}(16) = 189757; \qquad \mathcal{N}(48) = 229046391;$$
$$\mathcal{N}(18) = 1463871; \qquad \mathcal{N}(50) = 52289661;$$
$$\mathcal{N}(20) = 9637162; \qquad \mathcal{N}(52) = 9637162;$$
$$\mathcal{N}(22) = 52289661; \qquad \mathcal{N}(54) = 1463871;$$
$$\mathcal{N}(24) = 229046391; \qquad \mathcal{N}(56) = 189757;$$
$$\mathcal{N}(26) = 797876668; \qquad \mathcal{N}(58) = 21536;$$
$$\mathcal{N}(28) = 2198313637; \qquad \mathcal{N}(60) = 2267;$$
$$\mathcal{N}(30) = 4793685822; \qquad \mathcal{N}(62) = 246;$$

$\mathcal{N}(32) = 8314036051;$ $\mathcal{N}(64) = 28;$

$\mathcal{N}(34) = 11534400357;$ $\mathcal{N}(66) = 1;$

$\mathcal{N}(36) = 12857549996;$ $\mathcal{N}(72) = 1;$

3. The irregular LDPC code of rate 1/3, length 72 and $\lambda(x) = x^2$, $\rho(x) = 0.5x^5 + 0.5x^6$:

$\mathcal{N}(8) = 4;$ $\mathcal{N}(36) = 2536231;$

$\mathcal{N}(10) = 14;$ $\mathcal{N}(38) = 2633226;$

$\mathcal{N}(12) = 57;$ $\mathcal{N}(40) = 2027427;$

$\mathcal{N}(14) = 281;$ $\mathcal{N}(42) = 1290789;$

$\mathcal{N}(16) = 1092;$ $\mathcal{N}(44) = 673670;$

$\mathcal{N}(18) = 4316;$ $\mathcal{N}(46) = 291107;$

$\mathcal{N}(20) = 14916;$ $\mathcal{N}(48) = 102581;$

$\mathcal{N}(22) = 46543;$ $\mathcal{N}(50) = 29497;$

$\mathcal{N}(24) = 128411;$ $\mathcal{N}(52) = 7097;$

$\mathcal{N}(26) = 313653;$ $\mathcal{N}(54) = 1387;$

$\mathcal{N}(28) = 668411;$ $\mathcal{N}(56) = 210;$

$\mathcal{N}(30) = 1230176;$ $\mathcal{N}(58) = 32;$

$\mathcal{N}(32) = 1926450;$ $\mathcal{N}(60) = 2;$

$\mathcal{N}(34) = 2549636;$

4. The irregular LDPC code of rate 1/2, length 200 and $\lambda(x) = 0.31570x + 0.26758x^2 + 0.41672x^6$, $\rho(x) = 0.4381x^5 + 0.5619x^6$:

$\mathcal{N}(8) = 2;$ $\mathcal{N}(14) = 97;$

$\mathcal{N}(9) = 1;$ $\mathcal{N}(15) = 224;$

$\mathcal{N}(10) = 5;$ $\mathcal{N}(16) = 592;$

$\mathcal{N}(11) = 12;$ $\mathcal{N}(17) = 1339;$

$\mathcal{N}(12) = 32;$ $\mathcal{N}(18) = 2964;$

$\mathcal{N}(13) = 54;$ $\mathcal{N}(19) = 6515;$

5. The regular (3,6) LDPC code of rate 433/495 and length 495:

$$\mathcal{N}(4) = 2^{5.9}; \qquad \mathcal{N}(12) = 2^{13.6};$$
$$\mathcal{N}(6) = 2^{8.2}; \qquad \mathcal{N}(14) = 2^{15.0};$$
$$\mathcal{N}(8) = 2^{10.2}; \qquad \mathcal{N}(16) = 2^{16.5};$$
$$\mathcal{N}(10) = 2^{12.1}; \qquad \mathcal{N}(18) = 2^{17.2};$$

D. Derivation of Performance Bound for Stage 2 of the MLC System

Let $\mathbf{x}$ be the symbol sequence of length $N_s^{(2)} = N^{(2)}/q_2$, which corresponds to a codeword $\mathbf{c}_m^{(2)}$ of $\mathcal{C}^{(2)}$ and other coded bits of $\mathcal{C}^{(1)}$. Consider a received sequence $\mathbf{y}$ resulted from the transmission of $\mathbf{x}$. If the coded bits of code $\mathcal{C}^{(1)}$ are perfectly known for Stage 2 decoding, the number of possible sequences $\mathbf{x}$ is $2^{R^{(2)}N^{(2)}}$. The ML rule chooses the closest $\mathbf{x}$ to the received sequence $\mathbf{y}$ among all the $2^{R^{(2)}N^{(2)}}$ possible sequences.

The union bound of bit error probability for Stage 2 under ML decoding can be written as:

$$P_e^{(2)} \leq \frac{1}{2^{R^{(2)}N^{(2)}}} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \frac{W_{\mathbf{x},\mathbf{x}'}^{(2)}}{N^{(2)}} \Pr(\mathbf{x} \rightarrow \mathbf{x}') \tag{D.1}$$

where $\Pr(\mathbf{x} \rightarrow \mathbf{x}')$ is the pairwise error probability when $\mathbf{x}$ is transmitted, but $\mathbf{x}'$ is decided at the receiver. The parameter $W_{\mathbf{x},\mathbf{x}'}^{(2)}$ is the Hamming distance of two codewords $\mathbf{c}_m^{(2)}$ and $\mathbf{c}_{m'}^{(2)}$ of $\mathcal{C}^{(2)}$ obtained by demapping the two sequences $\mathbf{x}$ and $\mathbf{x}'$, respectively. The union bound averaged over permuted LDPC ensemble can then be written as follows:

$$\overline{P}_e^{(2)} \leq E\left[\frac{1}{2^{R^{(2)}N^{(2)}}} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \frac{W_{\mathbf{x},\mathbf{x}'}^{(2)}}{N^{(2)}} \Pr(\mathbf{x} \rightarrow \mathbf{x}')\right] \tag{D.2}$$

Similar to derivation of the performance bound for State 1, the pairwise error probabilities $\Pr(\mathbf{x} \rightarrow \mathbf{x}')$ with the same type of error sequences $\mathbf{n}^{(2)}$ are grouped together to compute the union bound as:

$$\overline{P}_e^{(2)} \leq \sum_{l=l_{\min}^{(2)}}^{N^{(2)}} \sum_{n_1^{(2)}=0}^{N_s^{(2)}} \cdots \sum_{n_{q_2}^{(2)}=0}^{N_s^{(2)}} \frac{l}{N^{(2)}} \overline{f}(\mathbf{n}^{(2)}) E\left[Q\left(\frac{D_{\mathbf{n}^{(2)}}}{\sqrt{2N_0}}\right)\right] \tag{D.3}$$

Note that, the effective error coefficient does not appear for State 2. This is because the

coded bits of Stage 1 are already decided and the multiplicity of the number of nearest neighbors is simply set to one.

The function $\overline{f}(\mathbf{n}^{(2)})$ is computed as in (3.15) and (3.16). Here, the parameters corresponding to Stage 2 $\overline{f}(\mathbf{n}^{(2)})$, $P_{l,\mathbf{n}^{(2)}}$, $\mathcal{N}^{(2)}(l)$, and $N^{(2)}$ are substituted for $\overline{f}(\mathbf{n})$, $P_{l,\mathbf{n}}$, $\mathcal{N}(l)$, and N in (3.15) and (3.16), respectively.

The total distance $D_{\mathbf{n}^{(2)}}$, a random variable, can be computed from the single symbol error distances $D_k^{(2)}$ that are also random variables as $\left(D_{\mathbf{n}^{(2)}}\right)^2 = \sum_{k=1}^{N_s^{(2)}} \left(D_k^{(2)}\right)^2$. Then, the final form of the union bound for Stage 2 decoding is as follows:

$$\overline{P}_e^{(2)} \leq \sum_{l=l_{\min}^{(2)}}^{N^{(2)}} \sum_{n_1^{(2)}=0}^{N_s^{(2)}} \cdots \sum_{n_{q2}^{(2)}=0}^{N_s^{(2)}} \sum_{j=1}^{k_{\mathbf{n}^{(2)}}} \frac{l}{N^{(2)}} \overline{f}(\mathbf{n}^{(2)}) p_{\mathbf{n}^{(2)},j} Q\left(\frac{\Delta_{\mathbf{n}^{(2)},j}}{\sqrt{2N_0}}\right) \tag{D.4}$$

where $p_{\mathbf{n}^{(2)},j} = P[D^2_{\mathbf{n}^{(2)},j} = \Delta^2_{\mathbf{n}^{(2)},j}]$, $j = 1, 2, \ldots, k_{\mathbf{n}^{(2)}}$. Here, $k_{\mathbf{n}^{(2)}}$ is the number of distinguishable Euclidean distances and each of them has a probability $p_{\mathbf{n}^{(2)},j}$. The random variable $D^2_{\mathbf{n}^{(2)}}$ can take on the following values:

$$\Delta^2_{\mathbf{n}^{(2)},j} = \sum_{i=0}^{q_2} \sum_{h=1}^{h_{\max}} n_{i,h}^{(2)} (d_{i,h}^{(2)})^2 \tag{D.5}$$

The probability of $D^2_{\mathbf{n}^{(2)}} = \Delta^2_{\mathbf{n}^{(2)},j}$ with a given set of $n_{i,h}^{(2)}$ is:

$$p_{\mathbf{n}^{(2)},j} = \prod_{i=0}^{2} \binom{n_i^{(2)}}{n_{i,0}^{(2)}, \ldots, n_{i,h}^{(2)}} \prod_{h=0}^{h_{\max}} {p_{i,h}^{(2)}}^{n_{i,h}^{(2)}} \tag{D.6}$$

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Tech. J.*, vol. 27, pp. 623–656, July 1948.

[2] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. John Wiley and Sons, 1965.

[3] S. Parker, M. Sandell, M. S. Yee, Y. Sun, M. Ismail, P. Strauch, and J. McGeehan, “Space-time codes for future WLANs: principles, practice, and performance,” *IEEE Commun. Mag.*, vol. 42, pp. 96–103, Dec. 2004.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes,” in *Proc. IEEE Int. Conf. Commun.*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*. MIT Press, 1963.

[6] D. J. C. MacKay and R. M. Neal, “Near Shannon limit performance of low density parity check codes,” *Electronics Letters*, vol. 32, pp. 1645–1646, Aug. 1996.

[7] S. ten Brink, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[8] T. J. Richardson and R. L. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[9] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Speilman, “Improved low-density parity-check codes using irregular graphs,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, Feb. 2001.

[10] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, “Design capacity approach irregular low-density parity-check codes,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[11] S. Y. Chung, G. D. Forney, T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Letters*, vol. 5, pp. 58–60, Feb. 2001.

[12] T. Richardson and R. Urbanke, "The renaissance of Gallager's low-density parity-check codes," *IEEE Commun. Mag.*, vol. 41, pp. 126–131, Aug. 2003.

[13] A. J. Blanksby and C. J. Howland, "A 690-mW 1-Gb/s 1024-b, rate-1/2 low-density parity-check code decoder," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 404–412, Mar. 2002.

[14] J. Hou, *Capacity-Approaching Coding Schemes Based on Low-Density Parity-Check Codes*. PhD thesis, University of California, San Diego, California, USA, 2003.

[15] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for multi-antenna modulation and detection," *IEEE Trans. Commun.*, vol. 52, pp. 670–678, Apr. 2004.

[16] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.

[17] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[18] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *IEEE Trans. Inform. Theory*, vol. 41, p. 13791396, Sept. 1995.

[19] H. Imai and S. Hirakawa, "A new multilevel coding method using error correting codes," *IEEE Trans. Inform. Theory*, vol. 25, pp. 371–377, Feb. 1977.

[20] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[21] 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.

[22] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, Mar. 1999.

[23] 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.

[24] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[25] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power MQAM for fading channel," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.

[26] C. E. Shannon, "Communication in the presence of noise," *Proceeding of the IRE*, vol. 37, pp. 10–21, Jan. 1949.

[27] R. G. Gallager, "Simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, pp. 3–18, Jan. 1965.

[28] A. Papoulis, *Probability, Random Variable and Stochastic Processes*. McGraw-Hill, 1984.

[29] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2004.

[30] M. Sudan, "Coding theory: Tutorial and survey," in *Proceeding of the 42nd IEEE Symposium on Foundation of Computer Science*, pp. 1–17, 2002.

[31] R. Kotter and A. Vardy, "Algebraic soft decision decoding of Reed-Solomom codes," in *Proceeding of the 38th Annual Allerton Conference on Communication, Control and Computing*, pp. 625–635, 2000.

[32] A. Vardy, "Algorithmic complexity in coding theory and the minimum distance problem," in *Proceeding of Symposium on Theory of Computing*, pp. 92–109, 1997.

[33] R. W. Hamming, “Error detecting and error correcting codes,” *Bell System Tech. J.*, vol. 26, pp. 147–160, Apr. 1950.

[34] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, Mar. 1999.

[35] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. to be published, avalaible at http://lthcwww.epfl.ch/mct/index.php, 2006.

[36] G. D. Forney, *Concatenated Codes*. MIT Press, 1966.

[37] L. R. Bahl, J. Coke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[38] D. Divsalar, H. Jin, and R. J. McEliece, “Coding theorems for “Turbo-like” codes,” in *Proceeding of the 36th Allerton Conference on Communication, Control, and Computing*, pp. 201–210, 1998.

[39] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sept. 1981.

[40] T. J. Richardson and R. L. Urbanke, “Concentrate!,” in *Proceeding of Allerton Conf. on Communication, Control and Communication*, (Princeton), 1999.

[41] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Understanding belief propagation and its generlizations,” in *Technical Report, Mitsubishi Electrial Research laboratories*, 2002.

[42] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, “Factor graphs and sum-product algorithm,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[43] R. Narayanaswami, *Coded Modulation with Low Density Parity Check Codes*. Master Thesis, Texas A & M University, Texas, USA, 2001.

[44] D. J. C. MacKay and E. A. Ratzer, “Gallager codes for high rate applications,” in *Technical Report*, pp. 1–8, 2003.

[45] J. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *Proceeding of ACM SIGCOMM*, pp. 1–8, 1998.

[46] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann, 1988.

[47] Y. Mao and A. Banihashemi, "A heuristic search for good LDPC codes at short block length," in *Proc. IEEE Int. Conf. Commun.*, 2001.

[48] S. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3140–3159, Dec. 2003.

[49] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 887–908, Jan. 2002.

[50] X. Y. Hu, M. P. C. Fossorier, and E. Eleftheriou, "On computation of the minimum distance of low density parity check code," in *Proc. IEEE Int. Conf. Commun.*, pp. 767–771, 2004.

[51] C. Berrou, S. Vaton, M. Jezequel, and C. Douillard, "Computing linear distance of linear codes by error impulse method," in *Proc. IEEE Global Telecommun. Conf.*, (Taipei, Taiwan), pp. 1017–1020, Nov. 2002.

[52] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1284–1292, July 1994.

[53] I. Sason, *Upper bounds on the maximum likelihood decoding error probability for block codes and Turbo-like codes*. PhD thesis, The Technion-Israel Institute of Technology, Haifa, Israel, Sept. 2001.

[54] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3029–3051, Dec. 2002.

[55] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gausian approximation," *IEEE Trans. Inform. Theory*, vol. 5, pp. 657–670, Feb. 2001.

[56] D. Mackay and M. Postol, "Weaknesses of Margulis and Ramanujan-Margulis low-density parity-check codes," *Electronic Notes in Theoretical Computer Science*, vol. 74, 2003.

[57] T. Richardson, "Error floors of LDPC codes," in *Proceeding of 41st Allerton Conference on Communication, Control, and Computing*.

[58] A. Ashikhmin, G. Kramer, and S. ten Brink, "Estrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. summitted, pp. 1–17, Apr. 2003.

[59] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, pp. 806–808, May 1999.

[60] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence coding," in *Proc. IEEE Int. Symp. Inform. Theory*, (Ronneby, Sweden), June 1976.

[61] G. Ungerboeck, "Channel coding with multilevel/phase signal," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55–67, Jan. 1982.

[62] D. Divsalar and M. K. Simon, "Multiple trellis coded modulation (MTCM)," *IEEE Trans. Commun.*, vol. 36, pp. 410–419, Apr. 1988.

[63] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 177–195, Mar. 1987.

[64] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 206–218, Feb. 1998.

[65] S. ten Brink, J. Speidel, and R. H. Yan, "Iterative demapping for QPSK modulation," *Electronics Letters*, vol. 34, pp. 1459–1460, July 1998.

[66] S. ten Brink, J. Speidel, and R. H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf.*, pp. 579–584, 1998.

[67] X. Li, A. Chidapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8-PSK signaling," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.

[68] A. Chidapol 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.

[69] N. H. Tran and H. H. Nguyen, "Design and performance of BICM-ID systems with hypercube constellations," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1169–1179, May 2006.

[70] N. H. Tran and H. H. Nguyen, "Signal mappings of 8-ary constellations for bit-interleaved coded modulation with iterative decoding," *IEEE Transactions on Broadcasting*, vol. 52, pp. 92–99, Jan. 2006.

[71] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity check codes on Rayleigh fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 924–934, May 2001.

[72] 3GPP TS 25. 212, *Technical Specification Group Radio Access Network: Multiplexing and channel coding*. available at http://www.3gpp.org, 2002.

[73] T. Duman and M. Salehi, "Performance bounds for turbo-coded modulation systems," *IEEE Trans. Commun.*, vol. 47, pp. 511–521, Apr. 1999.

[74] M. P. C. Fossorier and S. Lin, "Soft-input soft-output decoding of linear block codes based on ordered statistics," in *Proc. IEEE Global Telecommun. Conf.*, pp. 2828–2833, 1998.

[75] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *International Symposium on Turbo Codes*, (Brest, France), pp. pp.1–11, 1997.

[76] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, 2nd Edition, 2004.

[77] D. Chase, "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inform. Theory*, vol. 18, pp. 172–182, Jan. 1972.

[78] L. Papke and P. Robertson, "Improved decoding with SOVA in a parallel concatenated (Turbo-code) scheme," in *Proc. IEEE Int. Conf. Commun.*, (Dallas), pp. 102–106, June 1996.

[79] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. Commun.*, vol. 51, pp. 2011–2018, Dec. 2003.

[80] E. Biglieri, D. Divsalar, P. MacLane, and M. Simon, *Introduction to Trellis-Coded Modulation and Applications*. Macmillan Publishing Company, 1991.

[81] J. L. Massey, "Coding and modulation in digital communication," in *Zurich Seminar on Digital Communications*, pp. E2(1)–E2(4), 1974.

[82] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1736–1740, Sept. 1997.

[83] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Military Commun. Conf.*, (Boston, MA), pp. 25.5.1–25.5.5, 1991.

[84] S. Slimane and T. Le-Ngoc, "Tight bounds on error probability of coded modulation schemes in Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 121–130, Feb. 1995.

[85] D. J. C. Mackay, *Encyclopedia of Sparse Graph Codes*. Available at http://www.inference.phy.cam.ac.uk/mackay/codes/data.html, 2004.

[86] U. Wachsmann, R. F. H. Ficher, and J. B. Huber, "Multilevel codes: Theoretical concepts and pratical design rules," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1361–1391, July 1999.

[87] L. F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, pp. 1439–1449, Oct. 1993.

[88] J. Hou, P. Siegel, L. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2141–2155, Sept. 2003.

[89] Y. Kofman, E. Zehavi, and S. Shamai, "Performance analysis of a multilevel coded modulation system," *IEEE Trans. Commun.*, vol. 42, pp. 299–312, Oct. 1994.

[90] R. H. M. Zaragoza, M. P. C. Fossorier, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding - Part I: Symmetric constellations," *IEEE Trans. Commun.*, vol. 48, pp. 204–213, Feb. 2000.

[91] M. Isaka, M. P. C. Fossorier, R. H. M. Zaragoza, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding - Part II: Asymmetric constellations," *IEEE Trans. Commun.*, vol. 48, pp. 774–786, May 2000.

[92] H. Herzberg, "On the spectrum of distances of a multilevel code, decoded by a multistage decoder," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1736–1740, Sept. 1997.

[93] H. G. Vu, H. H. Nguyen, and D. E. Dodds, "Error performance of coded modulation systems based on LDPC codes," in *Proc., the 6th IEEE International Workshop on Signal Processing Advances for Wireless Communications*, (New York, USA), pp. 131–135, June 2005.

[94] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun. Related Technol.*, vol. 10, pp. 585–596, Nov. 1999.

[95] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell System Tech. J.*, vol. 1, pp. 41–59, autumn 1996.

[96] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[97] D. Tse and P. Viswanath, *Fundamental of Wireless Communications*. Cambridge University Press, 2005.

[98] 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.

[99] D. Argawal, T. J. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channel," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2618–2626, Sept. 2001.

[100] W. Zhao, G. Leus, and G. B. Gannakis, "Orthogonal design of unitary constellations for uncoded and trellis-coded noncoherent space-time systems," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1319–1327, June 2004.

[101] I. Bahceci and T. M. Duman, "Combined Turbo coding and unitary space-time modulation," *IEEE Trans. Commun.*, vol. 50, pp. 1244–1249, Aug. 2002.

[102] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Coded unitary space-time modulation with iterative decoding: Error performance and mapping design," to appear in *IEEE Transactions on Communications*.

[103] N. Yoshimochi, T. Hiramoto, A. Mizuki, C.-J. Ahn, and I. Sasase, "LDPC coded unitary space-time modulated OFDM system in broadband mobile channel," *Electronics Letters*, vol. 39, pp. 994–995, June 2003.

[104] H. G. Vu, H. H. Nguyen, and D. E. Dodds, "Iterative ordered-statistic decoding for LDPC coded modulation systems," in *Proc., 18th Annual Canadian Conference on Electrical and Computer Engineering*, (Sakatoon, Canada), pp. 185–188, May 2005.

[105] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc. Packings in Grassmanian spaces," *Exper. Math.*, vol. 5, pp. 139–159, 1996.

[106] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[107] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, pp. 1268–1278, Apr. 2006.

[108] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Commun. Mag.*, pp. 526–536, June 2005.

[109] G. L. Stuber, "Broadband MIMO-OFDM wireless communication," *Proc. IEEE*, pp. 271–294, Feb. 2004.

[110] Z. Zhou and B. Vucetic, "MIMO systems with adaptive modulation," in *Proc. IEEE Veh. Technol. Conf.*, pp. 765–769, 2004.

[111] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1626–1636, Sept. 2004.

[112] F. Kharrat-Kammoun, S. Fontenelle, and J. J. Boutros, "Accurate approximation of QAM error probability on quasi-static MIMO channel and its application to adaptive modulation," submitted to *IEEE Trans. on Inform. Theory*, 2004.

[113] Y. D. Kim, I. Kim, J. Choi, J. Y. Ahn, and Y. H. Lee, "Adative modulation for MIMO systems with V-BLAST detection," in *Proc. IEEE Veh. Technol. Conf.*, (Jeju, Korean), pp. 1074–1078, Apr. 2003.

[114] P. Xia, S. Zhou, and G. B. Giannakis, "Multi-antenna adaptive modulation with beamforming based on bandwidth-constrained feedback," *IEEE Trans. Commun.*, vol. 53, pp. 526–536, Mar. 2005.

[115] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system," *IEEE Trans. Commun.*, vol. 49, pp. 1307–1311, Aug. 2001.

[116] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) in a noise-limited cellular system," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 226–234, Apr. 2002.

[117] G. Femenias, "SR ARQ for adaptive modulation systems combined with seclection transmit diversity," *IEEE Trans. Commun.*, vol. 53, pp. 998–1006, June 2005.

[118] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Process.*, vol. 52, pp. 202–213, Jan. 2004.

[119] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[120] A. Pandharipande, "Adaptive modulation for MIMO-OFDM systems," in *Proc. IEEE Veh. Technol. Conf.*, (Los Angeles), pp. 1266–1270, May 2004.

[121] Y. H. Pan, K. B. Letaief, and Z. Cao, "Dynamic spatial subchannel allocation with adaptive beamforming for MIMO/OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 53, pp. 2097–2017, Nov. 2004.

[122] M. P. C. Fossorier, "Iterative reliability-based decoding of low-density parity check codes," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 908–917, May 2001.

[123] H. Herzberg and G. Poltyrev, "The error probability of M-ary PSK block coded modulation schmemes," *IEEE Trans. Commun.*, vol. 44, pp. 427–433, Apr. 1996.

[124] S. Yousefi and A. K. Khandani, "Generalized tangential sphere bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2810–2815, Nov. 2004.

[125] S. Yousefi, "Performance evaluation of maximum-likelihood decoded linear binary block codes in AWGN interference using Bonferroni-type and geometrical bounds," *IEE Proceedings Communications*, vol. 153, pp. 219–232, Apr. 2006.

[126] M. Davey and D. MacKay,"Low density parity check codes over GF(q)," *IEEE Commun. Letters*, vol. 2, pp. 165–167, June 1998.

[127] M. D. Celetti, *Advanced Digital Design With the Verilog HDL*. Prentice Hall, 2004.