CODE DESIGN FOR SISO AND MIMO BLOCK-FADING CHANNELS

A Dissertation Presented

by

Yueqian Li

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Communications and Digital Signal Processing

Northeastern University
Boston, Massachusetts
April 2013
Dissertation Title: CODE DESIGN FOR SISO AND MIMO BLOCK-FADING CHANNELS

Author: Yueqian Li

Department: Electrical and Computer Engineering

Approved for Dissertation Requirement for the Doctor of Philosophy Degree

Dissertation Advisor Professor Masoud Salehi  Date

Dissertation Reader Professor Milica Stojanovic  Date

Dissertation Reader Professor Kaushik Chowdhury  Date

Dissertation Reader  Date

Department Chair Professor Ali Abur  Date

Graduate School Notified of Acceptance:

Director of the Graduate School  Date
Abstract

We study, analyze, and design communication systems for data transmission over block-fading channels. The block-fading channel is a model for communication under slowly-varying fading; where a codeword spans a few independent fading blocks. Code design strategies for block-fading channels are quite different from those for classical additive white Gaussian noise (AWGN) channels or fully interleaved fading channels.

From the expression for pairwise error probability (PEP) bound under maximum likelihood (ML) decoding, we can extract two major parameters for code design on block-fading channels; the diversity order and the coding gain. At high signal to noise ratios, the diversity order determines the slope of the codeword error probability curve, while the coding gain shifts the curve horizontally. Therefore, the diversity order is the determining factor in code design. The optimal diversity order achievable by coding scheme is upper bounded by the Singleton bound, which establishes the fundamental tradeoff between coding rate and diversity order. The family of codes which can achieve the optimal diversity order are referred as blockwise maximum distance separable (MDS) codes. The general approach for code construction on block-fading channels is to design MDS codes with large coding gain.

For single-input single-output (SISO) communication, we propose a blockwise convolutionally encoded bit-interleaved coded modulation (BC-BICM) scheme, which achieves optimal diversity order. In addition, the coding gain can be improved either by choosing a carefully designed signal labeling scheme for the BICM with iterative decoding or by using convolutional codes with longer constraint lengths. We also investigate quasi-cyclic low-density parity-check (QC-LDPC) codes for block-fading channels. With careful design, the proposed QC-LDPC codes exhibit the same good performance as their corresponding random root-LDPC codes. Moreover, the structure of the proposed QC-LDPC codes makes them suitable for efficient encoding.

For the multiple-input multiple-output (MIMO) situation, the system design should
take advantage of additional space diversity provided by multiple antennas besides the time diversity. We first study a turbo coded BICM scheme with iterative detection. The signal processing unit at the receiver employs sphere detector to achieve good performance with reduced computation complexity. To obtain the full diversity of the channel, we propose a coded space-time scheme based on modulation diversity, where the channel coding exploits the time diversity and space-time coding provides space diversity. It is shown that the proposed system achieves high throughput by transmitting at full spatial multiplexing.
First and foremost, I would like to express my sincere appreciation and gratitude to my advisor, Prof. Masoud Salehi, for his guidance and support during my doctoral studies. He has been a great source of knowledge to this work. I have learned a lot through our discussions and it is a real pleasure to work with him. I feel very lucky to have him as my PhD advisor.

I am grateful to Prof. Milica Stojanovic and Prof. Kaushik Chowdhury for accepting to be members of my thesis committee, and providing valuable feedbacks on the draft of this dissertation. I own many thanks to them for their time and effort in helping me with my research work.

During my PhD program, I took two internships, which have greatly enriched my experience and been benefit to my future career. I would like to thank those mentors who hosted me in their teams, including Dr. Sean Ramprashad, Dr. Sayandev Mukherjee and Dr. Haralabos Papadopoulos from DOCOMO Communications Labs USA and Dr. Toshiaki Koike-Akino from Mitsubishi Electric Research Laboratories (MERL).

I also would like to thank my friends at Northeastern University for their friendship. They have made my study at Northeastern University more enjoyable. In particular, I am thankful to the members of Communications and Digital Signal Processing (CDSP) center. It is a great honor for me to be part of CDSP and it is a great experience to work with those talented people in CDSP center.

Finally, I am indebted to my family. It would not have been possible to finish this doctoral dissertation without the support and encouragement from the family. I especially wish to express my gratitude to my wife, Ran, for her love and support.

This dissertation is dedicated to my parents and my wife.
Contents

Abstract iv

Acknowledgements vi

1 Introduction 1
   1.1 The Block-Fading Channel Model 3
   1.2 The Information Outage Limits 4
   1.3 Code Design 7
      1.3.1 The Pairwise Error Probability 8
      1.3.2 The Singleton Bound and MDS Code 11
      1.3.3 Coding for Block-Fading Channels 12
   1.4 MIMO Block-Fading Channels 13
      1.4.1 The MIMO Systems 13
      1.4.2 The MIMO Block-Fading Channel Model 23
   1.5 Research Contributions 26

2 A New BC-BICM Scheme for Block-Fading Channels 27
   2.1 Blockwise Convolutional Codes for Block-Fading Channels 27
   2.2 System Model 29
   2.3 BC-BICM with Improved Coding Gain 31
      2.3.1 Symbol Labeling for BC-BICM with Iterative Decoding 31
      2.3.2 Convolutional Codes with Longer Constraint Lengths 33
   2.4 Comparison between BC-BICM and BCC-BICM 34
   2.5 Conclusions 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4 Conclusions</td>
<td>86</td>
</tr>
<tr>
<td>6 An Efficient Decoding Algorithm for Concatenated RS-Convolutional Codes</td>
<td>87</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>87</td>
</tr>
<tr>
<td>6.2 The Reed-Solomon Codes</td>
<td>88</td>
</tr>
<tr>
<td>6.2.1 The RS Code Encoding</td>
<td>89</td>
</tr>
<tr>
<td>6.2.2 Decoding of RS Codes</td>
<td>90</td>
</tr>
<tr>
<td>6.3 The RS-Convolutional Concatenated Coding Scheme</td>
<td>94</td>
</tr>
<tr>
<td>6.4 A Low Complexity Decoding Algorithm</td>
<td>95</td>
</tr>
<tr>
<td>6.5 Simulation Results</td>
<td>98</td>
</tr>
<tr>
<td>6.6 Conclusions</td>
<td>100</td>
</tr>
<tr>
<td>7 Concluding remarks and future work</td>
<td>101</td>
</tr>
<tr>
<td>7.1 Concluding remarks</td>
<td>101</td>
</tr>
<tr>
<td>7.2 Future work</td>
<td>103</td>
</tr>
<tr>
<td>Bibliography</td>
<td>104</td>
</tr>
</tbody>
</table>
List of Tables

6.1 The decoding procedure for RS codes. . . . . . . . . . . . . . . . . . . . . . 93
6.2 Error patterns of flipping 3 bits. . . . . . . . . . . . . . . . . . . . . . . . . 98
# List of Figures

1.1 A codeword of length $N$ spans over $F = 2$ fading blocks. ........................................ 4  
1.2 Outage limits of a block-fading channel with $F = 2$ using complex Gaussian inputs. ............................................................................................................................................................................. 6  
1.3 Outage limits of block-fading channels with $F = 2, 4, 8,$ and 16 using complex Gaussian inputs, $R = 1$ bit/sec/Hz. ........................................................................................................................................ 7  
1.4 The transmitter structure of VBLAST. ............................................................................. 18  
1.5 Receiver block diagram for VBLAST using ZF-SIC. ....................................................... 21  
1.6 Outage probabilities for MIMO block-fading channels with assumption of Gaussian inputs and genie-aided detector. .......................................................................................................................... 25  

2.1 The encoder structure of $(5, 3, 7, 7)$ convolutional code. ........................................... 28  
2.2 Minimum blockwise Hamming distance path for a block-fading channel with $F = 8$. ......................................................................................................................................................... 28  
2.3 The transmitter structure of the BC-BICM scheme. ....................................................... 29  
2.4 The receiver structure of the BC-BICM scheme. ............................................................ 30  
2.5 16-QAM labeling schemes: Gray, SP, MSP. ................................................................. 32  
2.6 WER of the BC-BICM scheme obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel with $F = 4$. ........................................ 33  
2.7 WER of the BC-BICM scheme with different convolutional codes obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel. 34  
2.8 WER of the BCC-BICM scheme and BC-BICM scheme obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel with $F=4$. ................................................................. 35  

3.1 A rate $r_c = 1/2$ regular $(3, 6)$ root-LDPC code for a block-fading channel with $F = 2$. ......................................................................................................................................................... 40
5.4 FER performance of the coded MIMO system over a $2 \times 1$ MIMO quasi-static fading channel. .................................................. 82

5.5 FER performance of the coded MIMO system over MIMO block-fading channels with $F = 2$ fading blocks. .................................................. 83

5.6 Average number of nodes visited during the sphere decoding search in a $2 \times 2$ MIMO block-fading channel with $F = 2$ fading blocks. ............... 84

5.7 Comparison of FER performance between the coded MIMO system and the serially Turbo-coded BICM scheme with $N_t = 2$, $N_r = 2$, over MIMO block-fading channels of $F = 2$ fading blocks. .......................... 85

6.1 The block diagram RS/Conv concatenated coding scheme. ............... 94

6.2 BER performance of the concatenated RS/Conv coding system. .......... 99

6.3 Complexity comparison between proposed algorithm and Chase algorithm. . 100
Chapter 1

Introduction

Mobile communication techniques have evolved dramatically in recent years; mainly due to the increasing demand for high quality and fast wireless data service. The wireless channel is considered to be one of the most challenging channel models for reliable communications. The main impairments present in a wireless communication environment include noise, fading, and interference, which pose major challenges in designing effective and reliable communication systems for mobile users.

The deteriorating effect of fading is one of the most challenging problems encountered by the designer of wireless communication systems. The effects of fading can generally be classified into large scale fading (noticeable over distances much larger than the wavelength) and small scale fading [1]. Large scale fading is mainly due to path loss and shadowing whereas small scale fading is a result of multipath propagation of electromagnetic waves and their reflection and scattering from nearby objects. In this research, we only consider the effects of small scale fading. Because the received signal is the sum of a large number of reflections of the transmitted signal (diffuse component), by applying the central limit theorem we can conclude that the amplitude of the received signal is a Rayleigh distributed random variable. When there exists a line-of-sight component (specular component), the amplitude of the received signal is modeled as a Ricean distributed random variable.

Another property of wireless channels is that their characteristics change with time. This is mainly due to mobility of transmitter and/or receiver, as well as variations in the transmission environment. The coherence time $T_c$ of a fading channel is defined as the time duration during which the channel impulse response does not change. Therefore, the value of the coherence time of the channel characterizes how rapidly the channel changes with
CHAPTER 1. INTRODUCTION

The coherence time is inversely related to the Doppler spread \( f_d \) \[1\]

\[ T_c \approx \frac{1}{f_d}. \] (1.1)

In (1.1), \( f_d \) is the maximum Doppler shift given by \( f_d = v/\lambda = f_c v/c \), where \( v \), \( c \), \( f_c \) and \( \lambda \) are the relative velocity between transmitter and receiver, the velocity of light, the carrier frequency, and the carrier wavelength, respectively. When the coherence time is less than the symbol period, the channel is refereed as a fast fading channel. When the coherent time is larger than the symbol (or the codeword) duration, the channel is refereed as a slow fading channel. Two convenient models that are widely used in the study of wireless channels are the quasi-static and fully interleaved fading channel models. In the quasi-static fading channel model, it is assumed that the fading coefficient is constant for an entire codeword duration and changes independently from one codeword to another. For fully interleaved fading channel model, we assume that the codeword passes through a long interleaver prior to transmission and hence the codeword spans a relatively long time interval, which is much longer than the coherence time. In this case, the fading coefficients are different and independent for every transmitted bit.

In this dissertation, we investigate the problem of wireless data transmission over a more realistic fading channel model, i.e., the block-fading channel model. In this channel model, a single codeword may span a few fading blocks with independent fading coefficients. The quasi-static and fully interleaved fading channel models can be considered as the two extreme cases of the block-fading channel. The block-fading channel model is a general and convenient model for delay sensitive communications over wireless links affected by slow-varying fading. The block-fading channel model is an appropriate channel model for the following situations:

- Slow frequency hopping: A codeword is divided into segments, which are sent on different frequency bands. The purpose of slow frequency hopping is to provide diversity for transmission links. Systems that employ slow frequency hopping include Global System for Mobile communications (GSM) and Enhanced Data rates for GSM Evolution (EDGE) \[2\].

- Hybrid Automatic Repeat reQuest (HARQ) with incremental redundancy (IR): When the initial transmission fails, the subsequent retransmissions contain different coded
bits from the initial transmission. The transmissions are sent on different transmission time intervals (TTIs) with independent interference levels. Systems that employ HARQ with IR, include High Speed Packet Access (HSPA), 3GPP Long Term Evolution (LTE) [3, 4], and worldwide interoperability for microwave access (WiMAX) [5].

- Cooperative communication: A codeword is sent in parts by the original node and the cooperative partner relay nodes. As those nodes are distributed at different locations, the fading coefficients for those parts are independent [6].

The selection of communication system architecture and system parameters in general depends on channel conditions. An off-the-shelf scheme designed for additive white Gaussian noise (AWGN) channel, quasi-static fading channel, or fully interleaved channel may not perform well over a block-fading channel. In fact, the design rules for block-fading channels are quite different from those for such classic channels. In this chapter, we first provide preliminaries on the block-fading channel, including the mathematical model for the channel and information outage limits. Then, error correcting code design criteria for block-fading channels are discussed. Some existing communication systems for block-fading channels are introduced.

We also discuss how to improve the performance of the communication over block-fading channels using multiple-input multiple-output (MIMO) techniques. At the end of this chapter, we summarize the contributions of the dissertation.

1.1 The Block-Fading Channel Model

The block-fading channel was first studied in [7] and [8]. We consider the situation that a codeword with finite length is transmitted only over a few independent fading blocks. To be more specific, we assume that a codeword of length $N$ spans $F$ fading blocks. With BPSK symbols $x_l \in \{+1, -1\}$, the $l$th received signal is given by

$$y_l = \alpha_f x_l + z_l, \quad l = 1, 2, \cdots, N \text{ and } f = 1, 2, \cdots, F$$

(1.2)

where $\alpha_f$ is the complex Rayleigh fading coefficient on the $f$th fading block $f = \lceil F \cdot l/N \rceil$ with $\lceil r \rceil$ denoting the ceiling of $r$. The fading coefficients are normalized with $E[|\alpha_f|^2] = 1$. Both the real and imaginary parts of fading coefficients are Gaussian distributed with zero mean and $1/2$ variance. We assume all the fading blocks have the same length of $L$. 

Therefore, \( FL = N \). The noise is sampled from a zero mean circularly symmetric complex Gaussian variable with variance \( \sigma^2 = N_0/2 \) per each dimension. The signal-to-noise ratio (SNR) is expressed as \( E_b/N_0 \), where \( E_b \) is the energy per information bit. A simple example of a codeword in a block-fading channel with \( F = 2 \) fading blocks is illustrated in Fig. 1.1.

![Figure 1.1: A codeword of length \( N \) spans over \( F = 2 \) fading blocks.](image)

When QAM symbols from constellation \( \mathcal{X} \) of size \( 2^Q \) are used, a frame is referred as a symbol sequence of length \( N/Q \) mapped from a codeword of length \( N \). In this situation, the block length is \( L = N/(F \cdot Q) \).

### 1.2 The Information Outage Limits

The capacity of a communication channel is defined as the highest rate at which the information can be reliably transmitted over the channel. Since the channel gain of a fading channel is random, the instantaneous channel capacity is also a random variable. For a fast fading channel, we assume the channel changes fast enough and a single codeword can experience all possible realizations of the channel. In this case, the **ergodic channel capacity** is defined as the ensemble average of channel capacity over all states of the channel. The word “ergodic” infers the assumption that the time average of instantaneous channel capacity within a codeword duration equals to statistical average of channel capacity. For slow fading channels including block-fading channels, different codewords will experience different realizations of the channel. Thus, the channel capacity changes for each codeword transmission. Instead of using ergodic capacity, **information outage limit** and **outage capacity** are defined to measure the slow fading channel quality, which are detailed in the rest of this section for block-fading channels.

Due to the nonergodic nature of the block-fading channel, its Shannon capacity is zero, which means reliable transmission of information at any positive rate is impossible over this channel. There is a certain probability that the current channel condition cannot support the loading transmission rate \( R \), which describes that situation that the channel is in **outage**.
The channel is characterized by its information outage limit, also called outage probability, which is defined as

\[ P_{\text{out}} = \Pr(I(E_b/N_0, \{\alpha_f\}) < R), \]  

where \( I(E_b/N_0, \{\alpha_f\}) \) is the instantaneous mutual information between the input and the output of the channel. The value of the instantaneous mutual information depends on current SNR and channel condition \( \{\alpha_f\} \). \( R = (K/N) \times Q = r_c \times Q \) denotes the transmission rate, where \( K \) is the number of information bits and \( r_c = K/N \) is the code rate. The outage limit (1.3) is the ultimate lower bound on frame error rate (FER) for any coding scheme over block-fading channels. With real Gaussian input and a throughput of \( R \) information bits per dimension, the instantaneous mutual information is calculated by

\[ I_{G_r}(E_b/N_0, \{\alpha_f\}) = \frac{1}{F} \sum_{j=1}^{F} \frac{1}{2} \log_2 \left( 1 + 2R \frac{E_b}{N_0} |\alpha_f|^2 \right). \]  

With complex Gaussian input and a throughput of \( R \) information bits per complex dimension, the instantaneous mutual information is calculated by

\[ I_{G_c}(E_b/N_0, \{\alpha_f\}) = \frac{1}{F} \sum_{j=1}^{F} \log_2 \left( 1 + R \frac{E_b}{N_0} |\alpha_f|^2 \right). \]  

With BPSK inputs, the corresponding instantaneous mutual information does not have a closed form and it is given by [9, p. 363]

\[ I_{BP SK}(E_b/N_0, \{\alpha_f\}) = \frac{1}{F} \sum_{f=1}^{F} \frac{1}{2} \left[ g \left( \sqrt{2R \frac{E_b}{N_0}} |\alpha_f|^2 \right) + g \left( -\sqrt{2R \frac{E_b}{N_0}} |\alpha_f|^2 \right) \right], \]  

where

\[ g(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w-t)^2}{2}} \log_2 \frac{2}{1 + e^{-2\pi w^2}} dw. \]  

The above function is symmetric \( g(t) = g(-t) \) and the value of \( g(t) \) can be computed using the Gauss-Hermite quadrature method.

For general complex QAM inputs, the instantaneous mutual information can be calculated
as [2]

\[ I_{QAM}(E_b/N_0, \{\alpha_f\}) = \frac{1}{F} \sum_{f=1}^{F} I_{QAM}(E_b/N_0, \alpha_f), \]  

(1.8)

where

\[ I_{QAM}(E_b/N_0, \alpha_f) = Q - 2^{-Q} \sum_{x \in X} \mathbb{E} \log_2 \sum_{x' \in X} \exp(-|\alpha_f \sqrt{R \frac{E_b}{N_0}} (x - x')^2 + |Z|^2). \]  

(1.9)

In above equation the expectation is taken with respect to the circular complex noise \( Z \sim \mathcal{N}_C(0, 1) \) and the value of the expectation can also be evaluated by the Gauss-Hermite quadrature method.

Fig. 1.2 shows outage limits of a block-fading channel with \( F = 2 \) using complex Gaussian inputs. The channel loads are set to be \( R = 0.5, 1, 2 \) and 3 bits/complex dimension. We can see that the outage curves shift to right when the channel load \( R \) increases. We observe that all outage limit curves have the same slop of 2 at high SNR region, which is the diversity order of the channel.

![Figure 1.2: Outage limits of a block-fading channel with \( F = 2 \) using complex Gaussian inputs.](image-url)
Fig. 1.3 illustrates outage limits of block-fading channels with $F = 2, 4, 8,$ and 16. We fix the channel load at $R = 1$ bit / complex dimension and assume complex Gaussian inputs are used. For the same system load, we get steeper outage curves with more fading blocks.

![Outage limits of block-fading channels with $F = 2, 4, 8,$ and 16 using complex Gaussian inputs, $R = 1$ bit/sec/Hz.](image)

Another channel quality indicator is the outage capacity defined as the highest transmission rate that can be supported by the channel without exceeding the pre-set outage limit $\epsilon \in [0, 1)$

$$C_\epsilon = \max\{R : P_{out}(R) < \epsilon\}. \quad (1.10)$$

In this dissertation, we will only use the outage limit as performance criterion for coding system designs.

### 1.3 Code Design

The goal of channel code design is to match code structure to channel conditions. Code design criteria are different for different channel conditions. For example, over a binary
symmetric channel (BSC) with cross probability less than 0.5, a good code should have a large minimum Hamming distance. Similarly, for an AWGN channel, the design rule is to maximize the minimum Euclidean distance.

Code design criteria for block-fading channels have been discussed in various works including [2, 10, 11]. In this section, we briefly review the design rules based on asymptotic analysis. In the high SNR region, the average FER, also known as codeword error rate (WER), for the block-fading channel can be expressed as

$$P_e \approx G_c \cdot (\text{SNR})^{-d}. \tag{1.11}$$

From (1.11), we can see that the FER, in the high SNR region, is determined by two parameters, $d$ and $G_c$, where $d$ is referred to as the diversity order and $G_c$ is known as the coding gain. By taking the logarithm of both sides of (1.11), we have

$$\log(P_e) \approx -d \cdot \log(\text{SNR}) + \log(G_c). \tag{1.12}$$

The diversity order determines the asymptotic slope of the error probability curve as a function of SNR on a log-log scale, while the coding gain provides a horizontal shift in the curve.

1.3.1 The Pairwise Error Probability

We begin by considering a block-fading channel with independent Ricean fading coefficients on each fading block. We further assume the energy of the fading coefficient is normalized, i.e., $E(|\alpha_f|^2) = S^2 + 2\sigma^2 = 1$, where $S^2$ and $2\sigma^2$ is the specular and the diffuse components, respectively. For this channel model, the Rice factor is defined as $K_{\text{Rice}} = \frac{S^2}{2\sigma^2}$. When $K_{\text{Rice}} = 0$, we have a Rayleigh fading channel with $2\sigma^2 = 1$. When $K_{\text{Rice}} \to \infty$, we have a Gaussian channel with $S^2 = 1$. We assume that the message $m$ consisting of binary coded bits of length $N$ is mapped onto the symbol sequence $\bar{x}(m)$ and is transmitted over the channel spanning $F$ independent fading realizations. The conditional pairwise error probability (PEP) between two codewords $\bar{x}(m)$ and $\bar{x}(m')$ under maximum likelihood (ML) decoding can be expressed as

$$P_e (\bar{x}(m) \to \bar{x}(m')|\{\alpha_f\}) = Q \left( \sqrt{\frac{F}{2N_0} \sum_{f=1}^{F} |\alpha_f|^2 d_f^2 (\bar{x}(m), \bar{x}(m'))} \right), \tag{1.13}$$
where
\[ d_f^2(\bar{x}(m), \bar{x}(m')) = \sum_{l=1}^{L} |x_{(f-1)l+1}(m) - x_{(f-1)l+1}(m')|^2 \] (1.14)
is the squared Euclidean distance between the portions of two codewords on fading block \( f \). The \( Q \) function in (1.13) is referred as Gaussian tail probability function and defined as
\[ Q(x) = P[N(0,1) > x] = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp(-\frac{t^2}{2}) dt. \]

By averaging the channel conditional pairwise error probability with channel distribution, we obtain the average pairwise error probability
\[ P_e(\bar{x}(m) \rightarrow \bar{x}(m')) \leq \frac{1}{2} \prod_{f=1}^{F} \frac{K_{Rice} + 1 + \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0}}{K_{Rice} + 1 + \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0}} \exp \left( -\frac{K_{Rice} \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0}}{K_{Rice} + 1 + \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0}} \right). \] (1.15)

A detailed derivation of the average pairwise error probability over the distribution of \( \{\alpha_f\} \) random variables by using moment-generating function of Chi-squared distribution is given in [10]. A similar result for the fully interleaved Ricean fading channel can be found in [9, Sec. 14.4]. We examine two special cases of Rayleigh block-fading channel and AWGN channel by setting \( K_{Rice} = 0 \) and letting \( K_{Rice} \to \infty \).

- \( K_{Rice} = 0 \): average PEP for Rayleigh block-fading channel
\[ P_e(\bar{x}(m) \rightarrow \bar{x}(m')) \leq \frac{1}{2} \prod_{f=1}^{F} \frac{1}{1 + \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0}} \] (1.16)

- \( K_{Rice} \to \infty \): average PEP for AWGN channel
\[ P_e(\bar{x}(m) \rightarrow \bar{x}(m')) \leq \frac{1}{2} \prod_{f=1}^{F} \exp \left( -\frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0} \right) = \frac{1}{2} \exp \left( -\frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0} \right), \] (1.17)

where \( d_f^2(\bar{x}(m), \bar{x}(m')) \) is the squared Euclidean distance between symbol sequences mapped from two codewords.
CHAPTER 1. INTRODUCTION

To obtain the average PEP over the codewords, we can apply the union bound

\[
P_e \leq \sum_m \sum_{m' \neq m} P(\bar{x}(m))P_e(\bar{x}(m) \rightarrow \bar{x}(m'))
\]

\[
\leq \sum_m \sum_{m' \neq m} \frac{1}{2N} \max_{m, m'} \{P_e(\bar{x}(m) \rightarrow \bar{x}(m'))\}
\]

\[
= (2^N - 1) \cdot \max_{m, m'} \{P_e(\bar{x}(m) \rightarrow \bar{x}(m'))\}
\]

(1.18)

In (1.15), the PEP bound consists of the product of \(F\) terms. For a fading block with partial Euclidean distance \(d_f(\bar{x}(m), \bar{x}(m')) = 0\), the \(f\)th term in the product is 1. In this situation, the portion of the codeword sent on the \(f\)th block will not contribute to discriminate two codewords. We define a new parameter – the blockwise Hamming distance, \(d_{FH}(m, m')\), as the number of blocks, on which the partial Euclidean distance \(d_f(\bar{x}(m), \bar{x}(m'))\) is nonzero. To simplify the asymptotic performance at high SNR region, we can assume that

\[
\frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{4N_0} \gg K_{Rice} + 1.
\]

The bound in (1.15) can be approximated as

\[
P_e(\bar{x}(m) \rightarrow \bar{x}(m')) \approx \frac{1}{2} \prod_{f=1}^d \frac{K_{Rice} + 1}{4N_0} \exp(-K_{Rice})
\]

\[
= \frac{1}{2} \left[4(K_{Rice} + 1) \exp(-K_{Rice})\right] \frac{d_{FH}^d(m, m')}{\prod_{f=1}^d d_f^2(\bar{x}(m), \bar{x}(m'))} \left(\frac{E_s}{N_0}\right)^{-d_{FH}^d(m, m')},
\]

(1.20)

where \(d_f(\bar{x}(m), \bar{x}(m'))\) is nonzero and \(E_s\) is the average symbol energy. (We assume that the partial Euclidean distances are nonzero for the first \(d_{FH}^d(m, m')\) blocks in (1.20) to simplify the expression.) By substituting (1.20) into (1.18), we can see that the diversity order of the coding system is the minimum of \(d_{FH}^d(m, m')\) over all pairs of codewords. The following term in (1.20)

\[
\Gamma_d(\bar{x}(m), \bar{x}(m')) = \prod_{f=1}^d \frac{d_f^2(\bar{x}(m), \bar{x}(m'))}{E_s}
\]

(1.21)

is defined as the product distance, which results in a parallel shift of the asymptotic PEP curve.
CHAPTER 1. INTRODUCTION

The diversity order is clearly the determining parameter in the design of error correcting codes for block-fading channels. The error correcting codes designed for a block-fading channel are expected to exploit the limited diversity orders that the channel provides and at the same time achieve good coding gain.

1.3.2 The Singleton Bound and MDS Code

For a block code $C$ with codeword length $N$, information length $K$ and the minimum distance $d$, the Singleton bound states that

$$d \leq N - K + 1 = N(1 - r_c) + 1.$$  \hspace{1cm} (1.22)

The Singleton bound can be modified to bound blockwise Hamming distance for code design in block-fading channels.

**Theorem 1.3.1** (Singleton Bound). The achievable diversity order on a block-fading channel with $F$ blocks is bounded by the Singleton bound [10, 11]:

$$d \leq \lfloor F(1 - r_c) \rfloor + 1,$$  \hspace{1cm} (1.23)

where $r_c$ ($\leq 1$) is the code rate and $\lfloor x \rfloor$ represents the the largest integer smaller than or equal to $x$.

Equation (1.23) clearly reflects the optimal trade-off between code rate $r_c$ and diversity order $d$.

**Proof.** Let $L$ denote the length of a fading block. It is convenient to view $L$ symbols over a fading block together as a super-symbol $X^L$ [7]. With this interpretation of super-symbols, the analysis of block-fading channels is reduced to the analysis of a non-binary block code with symbols in the form of $X^L$ with a fixed codeword length $F$. The diversity order of the coding system is equal to the minimum Hamming distance among all the codewords. Furthermore, all traditional bounding techniques can be applied to analyze the coding system for the block-fading channels. \hfill $\square$

The class of error correcting codes, which can achieve the Singleton bound on block-fading channels, are referred to as blockwise maximum distance separable (MDS) codes.
CHAPTER 1. INTRODUCTION

However, most existing MDS codes cannot be used on block-fading channels directly, because the codeword length is constrained to be $F$ with typical value of $2 \sim 8$, which is generally too short for a codeword.

It has been shown in [10, 12, 13] that the maximal diversity order $F$ is achieved by outage limit using Gaussian inputs; the optimal diversity order given by the Singleton bound is achieved by outage limit using a finite size QAM constellation.

It is interesting to find out the highest rate code which can achieve a diversity order of $d = F$. By the fact that no two codewords can have the same super-symbol in one fading block, we obtain the maximum number of the codewords to be $|X|^L$, where $|X|$ is the size of the modulation set. Therefore, the highest rate is achieved by employing all these codewords and the corresponding code rate is

$$r_c = \frac{\text{info bits}}{\text{coded bits}} = \frac{\log_2 |X|^L}{F \times L \times \log_2 |X|} = \frac{1}{F}. \tag{1.24}$$

This result was noted in [10]. We notice that the simplest code to achieve $d = F$ is a rate $r_c = 1/F$ repetition code.

1.3.3 Coding for Block-Fading Channels

Many coding schemes along with tools to analyze them have been proposed for block-fading channels. In addition to early presentation of block-fading channel model in [7, 8], the channel model was further studied in [14] alongside other fading channel models. In [11, 15] upper and lower bounds for performance of block codes over block-fading channels with ML decoding and perfect channel state information are derived. In [10], both linear block codes and trellis codes are considered for block-fading channels. In particular, a computer search method is developed to find non-systematic convolutional codes with different rates and modulation constellations, which can achieve the Singleton bound and provide the best product distance. Further results on convolutional code search for block-fading channels can be found in [16]. The blockwise concatenated coded modulation scheme for block-fading channels is proposed in [2]. The proposed coded modulation scheme achieves the Singleton bound and performs close to outage limits with iterative belief-propagation decoding for any codeword length. Both regular and irregular parallel turbo codes are shown to achieve the Singleton bound with the help of certain multiplexers in [17, 18]. Low density parity check (LDPC) codes with root connections are proposed for block-fading channels in [19].
CHAPTER 1. INTRODUCTION

1.4 MIMO Block-Fading Channels

Communication systems designed for block-fading channels are expected to achieve both high diversity order and good coding gain [10]. Since the diversity order is limited by channel intrinsic diversity orders in block-fading channels, we consider employing MIMO techniques to provide additional spatial diversity. Another important advantage of using multiple antennas is to enable high rate transmission through spatial multiplexing. Diversity coding can be used together with spatial multiplexing, where transmission reliability is in tradeoff with the system throughput.

1.4.1 The MIMO Systems

In this part, we briefly introduce some classical MIMO techniques. We consider a communication system with \( N_t \) transmit and \( N_r \) receive antennas. The MIMO channel matrix can be expressed as

\[
H = \begin{bmatrix}
    h_{1,1} & h_{1,2} & \cdots & h_{1,N_t} \\
    h_{2,1} & h_{2,2} & \cdots & h_{2,N_t} \\
    \vdots & \vdots & \ddots & \vdots \\
    h_{N_r,1} & h_{N_r,2} & \cdots & h_{N_r,N_t}
\end{bmatrix},
\]

where \( h_{i,j} \) is the Rayleigh fading coefficient on the link between transmit antenna \( j \) and receive antenna \( i \). The path gain of each link is normalized to be 1. We assume a codeword can span a total of \( F \) channel realizations in time domain. Assuming space-time code \( x \in \mathbb{C}^{N_t \times T} \) is sent, the channel input-output relation can be written as

\[
y = Hx + n,
\]

where \( y \in \mathbb{C}^{N_r \times T} \) is the matrix of received symbols and \( n \in \mathbb{C}^{N_r \times T} \) is the matrix of zero mean circularly symmetric complex Gaussian noise with variance \( N_0/2 \). The signal-to-noise ratio (SNR) is defined as \( E_b/N_0 \), where \( E_b \) is the transmit energy of an information bit.

Capacity of MIMO Channels

The capacity of MIMO channel Gaussian for single user communications was discussed by Telatar in 1999 [20]. More comprehensive results on single-user and multiuser MIMO capacity are provided by Goldsmith et al. in [21]. In this part, the basics of MIMO
channel capacity for single user case are reviewed. To simplify the representation, we assume a single use of the deterministic MIMO channel. Following the Telatar’s approach, we can decompose MIMO channel into a set of parallel SISO channels by singular value decomposition (SVD) of channel matrix

$$H = UΛV^H,$$  \hfill (1.27)

where $U$ and $V$ are two complex unitary matrices, $Λ$ is an $N_r \times N_t$ rectangular diagonal matrix with real nonnegative sigular values $λ_i$ on the diagonal. By left-multiplying $U^H$ to both sides of (1.27), we have

$$\tilde{y} = Λ\tilde{x} + \tilde{n},$$  \hfill (1.28)

where $\tilde{y} = U^Hy$, $\tilde{x} = V^Hx$ and $\tilde{n} = U^Hn$. The new noise vector $\tilde{n}$ is still white and its power is unchanged, as $U$ is a unitary matrix. The number of the independent parallel channels is determined by the rank $k$ of the channel matrix $H$. When the channel state information (CSI) is available at the transmitter, the capacity is achieved by waterfilling the available transmit power $P$ over those parallel channels with power gains given by $λ_i^2$ \cite{22}. The resulting channel capacity is given by

$$C_{CSIT}^{MIMO} = \sum_{i=1}^{k} \log_2 \left( 1 + \frac{P_i λ_i^2}{N_0} \right),$$  \hfill (1.29)

where $P_i = (μ - N_0 λ_i^2)^+$ is the power filled into channel $i$. The water level $μ$ is chosen to satisfy the total power constraint $P = \sum_{i=1}^{k} P_i$.

**Capacity without CSIT At High SNR** At the high SNR region, we have a huge amount of power. In this situation, we will use all available parallel channels and it is asymptotically optimal to fill equal amounts of power to those channels. The channel capacity can be expressed as

$$C_{CSIT}^{MIMO} \approx \sum_{i=1}^{k} \log_2 \left( 1 + \frac{Pλ_i^2}{kN_0} \right).$$  \hfill (1.30)

The condition number (eigenvalue spread) of the channel matrix $H$ is defined as the ratio between the largest and smallest eigenvalue values ($\max(λ_i)/\min(λ_i)$). The channel is said
CHAPTER 1. INTRODUCTION

to be well-conditioned, if the condition number is close to 1. In this situation, the capacity is achieve by equal power allocation. If the channel is full rank, we have \( k = \min\{N_t, N_r\} \) and the MIMO capacity linearly increases as \( k \cdot \log(SNR) \) at the high SNR region. This result is refereed as degree-of-freedom gain in [23].

**Capacity with CSIT At Low SNR** At the low SNR region, we have a small amount of power and the optimal strategy is to fill all available power to the channel with largest channel gain. The resulting channel capacity is

\[
C_{MIMO}^{CSIT} \leq \log_2 \left( 1 + \frac{P \max\{\lambda_i^2\}}{N_0} \right). \tag{1.31}
\]

For small value \( x \), we have the approximation \( \log_2(1 + x) \approx x \log_2 e \) by Taylor expansion around 0. And the capacity in (1.31) can be approximated by

\[
C_{MIMO}^{CSIT} \approx \frac{P \max\{\lambda_i^2\}}{N_0} \log_2 e = \max\{\lambda_i^2\} \cdot SNR \cdot \log_2 e. \tag{1.32}
\]

The MIMO system provides a power gain of \( \max\{\lambda_i^2\} \) with low SNR [23].

**Capacity without CSIT** When channel state information or statistical properties of the channel is not available at the transmitter, we will allocate equal amount of power to each transmit antenna. The capacity of MIMO channel without CSI at the transmitter is

\[
C_{MIMO}^{no-CSIT} = \log_2 \det \left( I_{N_r} + \frac{P}{N_t N_0} H H^H \right). \tag{1.33}
\]

Similarly, we decompose \( H H^H = U \Lambda^2 U^H \). The capacity in (1.33) can the expressed as

\[
C_{MIMO}^{no-CSIT} = \log_2 \det \left( I_{N_r} + \frac{P}{N_t N_0} \Lambda^2 \right) = \sum_{i=1}^{k} \log_2 \left( 1 + \frac{P \lambda_i^2}{N_t N_0} \right), \tag{1.34}
\]

where \( k \) is the rank of the channel matrix \( H \).
Space-Time Block Codes

Space-time block codes (STBC) based on orthogonal designs are able to achieve full space diversity with low-complexity receiver at the expense of low rate (no more than 1 symbol per channel use) [24]. The very first and simplest STBC design with two transmit antennas and a single receive antennas was invented by Alamouti in 1998 [25]. The transmitted codeword of Alamouti scheme is

\[
x = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{bmatrix} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix},
\]  

(1.35)

where \( s_1 \) and \( s_2 \) are symbols selected from any signal constellation. We assume the MIMO channel matrix

\[
H = [h_1 \ h_2]
\]

(1.36)
is constant for two consecutive channel use. The received symbols for two transmission are

\[
\begin{align*}
y_1 &= h_1 s_1 + h_2 s_2 + n_1 \\
y_2 &= -h_1 s_2^* + h_2 s_1^* + n_2.
\end{align*}
\]

(1.37)
The maximum-likelihood (ML) decoding of Alamouti scheme can be decoupled into two independent parts

\[
\begin{align*}
\hat{s}_1 &= \arg \min_{s_1} \left( (|h_1|^2 + |h_2|^2 - 1) |\hat{s}_1| + |\hat{s}_1 - y_1 h_1^* - y_2^* h_2| \right)^2 \\
\hat{s}_2 &= \arg \min_{s_2} \left( (|h_1|^2 + |h_2|^2 - 1) |\hat{s}_2| + |\hat{s}_2 - y_1^* h_1 h_2 + y_2 h_1| \right)^2.
\end{align*}
\]

(1.38)

Following the similar decoding procedure above, we actually decode each symbol separately for general STBC codes. Therefore the ML decoding complexity for STBC codes grows linearly, instead of exponentially, with the number of transmit antennas. It can be shown that the received signal SNR is the sum of SNRs on two links

\[
SNR = \frac{(|h_1|^2 + |h_2|^2) E_s}{N_0},
\]

(1.39)

where \( E_s \) is average transmitted symbol energy of the signal constellation [26].

The Alamouti coding scheme can be generalized with more than two transmit antennas based on orthogonal designs [24]. Two attractive features of STBC are:
• low-complexity ML decoding: each symbol can be decoded separately;

• the ability to provide full diversity: the code based on orthogonal designs satisfies rank criterion [24].

Spatial Multiplexing and Receiver Architectures

Instead of utilizing the multiple transmit antennas for additional diversity gain, we may increase the transmission rate by sending different data streams on different transmit antennas. The has been shown the capacity of MIMO system increases linearly with the number of transmit antennas, when the number of receive antennas is equal to the number of transmit antennas [27]. Layered space-time codes allow spatial multiplexing to increase the transmission rate with reduced diversity, based on the general framework of Bell Laboratories Layered Space-Time (BLAST) architectures [28]. However, it is not possible to have a simpler ML decoder to detect each symbol separately as STBC codes. Sub-optimal and low-complexity receiver designs are necessary for spatial multiplexing to work in practice. We take vertical BLAST (V-BLAST) as an example and briefly discuss several receiver architectures for spatial multiplexing.

The transmitter structure of VBLAST is draw in Fig. 1.4. First, the bit stream is mapped into symbol stream. Then, the symbol stream is demultiplexed into $N_t$ substreams and emitted from the corresponding transmit antenna. Assuming there is no additional temporal channel coding, each receive antenna will receive all $N_t$ symbols from all transmit antennas at the receiver. The input and output relation for a single transmission can be represent as

$$y = Hx + n,$$

where $x$ is the transmitted symbol vector of size $N_t$ and $y$ is the received signal vector of size $N_r$.

A detector which is capable to suppress the interference and separate those signals should be employed. The algorithms are pretty much the same as those developed for multiuser detection (MUD) [29].

**ML Detector** The ML detector processes all $N_t$ symbols jointly by searching over all possible transmitted symbol vectors. With circularly symmetric complex Gaussian noise $n$,
the ML detector choose the symbol vector, which minimizes the following metric

$$\hat{x} = \arg \min_{\hat{x}} \| H\hat{x} - y \|^2_F,$$

(1.41)

where $\| \cdot \|_F$ is Frobenius norm. The ML detector is optimal in terms of minimizing the probability of decoding a codeword in error. When the constellation size is $2^M$, the number of candidate vectors is $2^{MN_t}$, which grows exponentially with the number of transmit antennas. Therefore, it is impractical to perform ML decoding with a large number of transmit antennas. Suboptimal low-complexity linear equalization techniques can be used to distinguish symbols.

**Zero-Forcing (ZF) Detector**  The zero-forcing equalizer tries to find the inverse of the channel matrix, which completely eliminates the intersymbol interference (ISI). In case that $H$ is full rank square matrix, we have $W_{ZF} = H^{-1}$. For general situation, we can use Moore-Penrose pseudo-inverse (left inverse) of the channel as the Zero-forcing coefficients

$$W_{ZF} = H^+ = (H^HH)^{-1}H^H,$$

(1.42)

where $H^H$ denotes the conjugate transpose of the channel matrix $H$. Multiplying $H^+$ with 1.40 results in

$$\tilde{x} = W_{ZF}y = H^+y = y + H^+n.$$

(1.43)

After the zero-forcing equalization, we could decoder each symbol separately by finding the nearest point in the constellation to the corresponding element in $\tilde{x}$. The ZF equalization requires $N_r \geq N_t$ to work. As shown in (1.43), the disadvantage of using zero-forcing
equalization is to enhance the noise power by a factor of
\[ H^+ (H^+)^H = \left( (H^H H)^{-1} H^H \right) \left( (H^H H)^{-1} H^H \right)^H = (H^H H)^{-1}. \] (1.44)

When noise power is small (near zero), ZF is a good choice to suppress ISI. A more balanced linear filter is a linear minimum mean square error (MMSE) equalizer, which provides a trade-off between suppressing the interference and enhancing the noise.

**Linear MMSE detector** In general, the MMSE detector provides a conditional mean estimation for \( x \) based on observation \( y \)

\[ \tilde{x} = E [x|y], \] (1.45)

which minimize mean square error (MSE) \( E [(\tilde{x} - x)^2] \). However, it is usually difficult to calculate the close-form MMSE detector. One solution is to restrict the estimator to be linear, i.e., \( \tilde{x} = W y \). And the linear MMSE detector equalizer tries to find the coefficients \( W_{MMSE} \) to minimize following term

\[ E \left[ (W_{MMSE} y - x)(W_{MMSE} y - x)^H \right]. \] (1.46)

The linear MMSE filter coefficients can be derived by orthogonality principle

\[
\begin{align*}
E [(\tilde{x} - x)y^H] &= E [(W_{MMSE} y - x)y^H] = 0 \\
E [(\tilde{x} - x)] &= E [(W_{MMSE} y - x)] = 0,
\end{align*}
\] (1.47)

which is actually the necessary and sufficient condition for optimality in general MMSE sense. As the mean values of \( x \) and \( y \) are both zero vectors, we ignore the second condition of the orthogonality principle. From the first condition, we get

\[ W_{MMSE} R = H^H \cdot P_x, \] (1.48)

where \( P_x = E[x_i x_i^*] \) is the average transmit symbol power and \( R \) is \( N_r \times N_r \) autocorrelation matrix of received vector \( y \)

\[ R = E \left[ y y^H \right] = [H H^H \cdot P_x + I_{N_r} \cdot N_0]. \] (1.49)
Therefore, the linear MMSE estimator

$$
W_{MMSE} = P_x H^H R^{-1} = H^H \left( H H^H + I_{N_t} \cdot \frac{N_0}{P_x} \right)^{-1} = \left[ H^H H + I_{N_t} \cdot \frac{N_0}{P_x} \right]^{-1} H^H
$$

where the equivalent form in the last step is obtained using Woodbury matrix identity.

**Decision Feedback Equalization (DFE)** In contrast to ML detector, which possesses symbols jointly, the aforementioned two linear equalization approaches multiply the received vector with a matrix and then decodes each symbol separately. These symbol-by-symbol style detectors reduce the system processing complexity at the cost of reduced performance. A decision feedback equalizer improves the performance of a linear detector by using previous detector decisions to eliminate the interference.

If the symbols are processed successively, the previous detected symbols can be subtracted from received signal to aid the detection of the next symbol. The method is referred as successive interference cancellation or serial interference cancellation (SIC), which was originally proposed for multi-user detection (MUD) [30]. The problem of SIC is the excessive detection delay, as the algorithm processes each layer serially. Multistage parallel interference cancellation (PIC) was proposed as an alternative for SIC to reduce the processing delay, in which the interference is simultaneously removed from all layers [31].

The original detection method for VBLAST is based on SIC with optimal ordering [28]. The algorithm first decodes the strongest layer. Assuming the successful detection of this strongest layer, the interference effect of this layer can be canceled from all of the receiver equations. Then, the second strongest layer is detected and its effects are removed from all equations before the detection of the next layer. The process continues until the last layer is detected. The block diagram of ZF-SIC equalizer is depicted in Fig. 1.5.

The algorithm starts with ordering the signals by their post-detection SNR. For ZF equalization, the post-detection SNR for layer-$i$ can be expressed as

$$
SNR_i = \frac{P_i}{N_0 |\bar{w}_i|^2}, \quad (1.51)
$$

where $|\bar{w}_i|^2$ is the norm of $i$th row vector from ZF equalization matrix $W_{ZF}$ (1.42). For equal power case, the ordering is dependent on the value of norm of nulling vector $\bar{w}_i$, which acts as a noise enhance factor in ZF equalization. After $(l-1)$ layers have been processed, their interference will be removed from the received signal. We have new input-output
Figure 1.5: Receiver block diagram for VBLAST using ZF-SIC.
relation at the $l$th stage

$$y_l = H_l x_l + n = y - I_l = [h_{l}, h_{l+1}, \cdots, h_{N_t}] x_l + n,$$ (1.52)

where $x_l$ contains $(N_t - (l - 1))$ undetected signals and $I_l$ is the reconstructed interference from detected signals. We need to recalculate ZF equalization matrix according to (1.52) and start the process again. The ZF-SIC detection algorithm can be summarized as following 3 steps:

- Ordering by post-detection SNR values;
- Interference nulling by ZF equalization;
- Interference reconstruction and cancellation.

It is interesting to point out that the diversity order achieved by each layer is different by its detection order. For ZF equalization, it is required that $N_r$ should be no less than $N_t$. The diversity gain is $(N_r - N_t + 1)$ for the first detected layer and $N_r$ for the last detected layer. The reason of diversity gain difference is that ZF equalization uses receive diversity to nullify the interference.

**Precoding Techniques for MIMO**

If channel state information is available at the transmitter side through feedback, precoding can be performed combined with spatial multiplexing (multi-layer beamforming) to maximize the MIMO system throughput. The major difference between precoding and conventional beamforming (smart antennas) is that the precoding enables multi-layer transmission and beamforming supports only signal-layer transmission.

For point-to-point communication, singular value decomposition (SVD) precoding is the optimal following the similar analysis for capacity of single user deterministic MIMO channel [20]. Multiple streams are precoded with a unitary matrix and emitted from all available transmit antennas. At the receiver side, another unitary matrix is multiplied to received vector to detected all transmitted symbols. In this circumstance, each non-zero singular value act as a parallel channel without interference and can support an independent data stream. In addition, the channel capacity is achieved with water filling power allocation.

The precoding technique for multi-user communication is adopted by multi-user MIMO (MU-MIMO) to support multiple access, which is an extended concept of space-division
multiple access (SDMA). It is known that non-linear Dirty Paper Coding (DPC) is optimal precoding, which can pre-cancel known interference at the transmitter [32, 33]. As the complexity with non-linear precoding is high, it is preferable to adopt linear precoding strategies in practical applications [34].

Due to their attractive performance, MIMO techniques have been adopted by almost all modern wireless communication standards, including UMTS HSPA and LTE [3, 4], IEEE 802.16e WiMAX [5] and IEEE 802.11n WLAN [35]. We will consider several MIMO techniques and corresponding signal processing techniques in our research to take advantage of their diversity gain and spatial multiplexing.

1.4.2 The MIMO Block-Fading Channel Model

In this model, the channel matrix $H$ is constant over $L$ transmissions and changes independently for next $L$ transmissions. We assume a codeword can span a total of $F$ channel realizations in time domain. We also assume that the transmit antennas and the receive antennas are separated far enough in space, so that the fading coefficients between each transmit-receive pair are independent.

The Channel Model

At time $t$, a vector $x(t) = [x_1(t), x_2(t), \cdots, x_{N_t}(t)]$ is transmitted over the $N_t$ transmit antennas. We assume the symbol energy is normalized to 1, i.e., $E(|x_j(t)|^2) = 1$. A vector $y(t) = [y_1(t), y_2(t), \cdots, y_{N_r}(t)]$ is received at the $N_r$ receive antennas. The received vector $y(t)$ can be expressed as

$$y(t) = H(t)x(t) + n(t), \quad (1.53)$$

where $n(t) = [n_1(t), n_2(t), \cdots, n_{N_r}(t)]$ is the zero mean circularly symmetric complex Gaussian noise vector with each element of variance $N_0/2$ per dimension. The system signal to noise ratio (SNR) is measured by $E_b/N_0$, where $E_b$ is the transmit energy per information bit.

We follow the assumption in [13] that when we detect the signal from the transmit antenna $j$ at receive antenna $i$, a genie provides the knowledge of symbols from other transmit antennas. In this way, we can convert MIMO channel into a set of single-input
single output (SISO) non-interfering parallel block-fading channels,

\[ y_{i,j}(t) = h_{i,j}(t)x_j(t) + n_i(t). \]  

(1.54)

This assumption is valid when we are able to detect interferences from other transmit antennas and cancel them successfully. With this assumption, we can apply both coding schemes and analysis method designed for SISO block-fading channels to the MIMO cases. For example,

**The Outage Limits**

The Shannon capacity of the nonergodic MIMO block-fading channel is zero. As in SISO case, we characterize the channel by its information outage limit, which is defined as

\[ P_{out} = \Pr \left( I(E_b/N_0, \mathbf{H}) < R \right), \]  

(1.55)

where \( I(E_b/N_0, \mathbf{H}) \) is the instantaneous mutual information between the input and the output of the MIMO channel and \( R = r_c \times M \times N_t \) is the system loading rate, where \( r_c \) is the coding rate and \( 2^M \) is the modulation constellation size.

Assuming the CSI is unknown at the transmitter side, the power is equally allocated to each fading block. The instantaneous mutual information is calculated as

\[ I_G(E_b/N_0, \mathbf{H}) = \sum_{j=1}^{N_t} \left[ \frac{1}{F} \sum_{k=1}^{F} \log_2(1 + \text{SNR}_{k,j}) \right], \]  

(1.56)

where

\[ \text{SNR}_{k,j} = r_c \times M \times (E_b/N_0) \sum_{i=1}^{N_r} |h_{i,j}(L \times k)|^2. \]  

(1.57)

In (1.56), we assume complex Gaussian input is used and Shannon capacity formula is adopted. The SNR value is calculated according to a practical communication system with certain coding and modulation configuration as shown in (1.57). Several outage limits with versus \( E_b/N_0 \) are drawn in Fig.1.6. The spectral efficiency is \( R = 1 \) bits/transmission for \( N_t = 1 \) and \( R = 2 \) bits/transmission for \( N_t = 2 \). The Fig.1.6 shows that we will get curves with steeper slop, when the number of antennas or the number of fading blocks increases. The season is that the available diversity order determines the slope of the outage probability
curve with Gaussian input in the high SNR region. For $F = 2$, the maximum achievable diversity order is just 2 in the SISO case. Using $2 \times 2$ MIMO, we can achieve a diversity order of 8 over an $F = 2$ block-fading channel. For the case $N_t = 1, N_r = 1, F = 2$ and the case $N_t = 2, N_r = 1, F = 1$, the outage limits are the same.

![Figure 1.6: Outage probabilities for MIMO block-fading channels with assumption of Gaussian inputs and genie-aided detector.](image)

The achievable diversity order $d$ by a coding scheme for MIMO block-fading channels is upper-bounded by the modified Singleton bound

$$d \leq N_r (\lfloor FN_t(1 - r_c) \rfloor + 1) .$$

(1.58)

The internal part $\lfloor FN_t(1 - r_c) \rfloor + 1$ in (1.58) can be viewed as the diversity order provided by a SISO block-fading channel with $F \times N_t$ independent fading blocks. The receive diversity order $N_r$ is achieved by maximum ratio combining.
1.5 Research Contributions

In this dissertation, we focus on the communication system designs for data transmission over block-fading channels. The block-fading channel model is a channel model for delay sensitive communication over slow-varying wireless link, where a codeword experiences a few independent fading blocks. We review code design criteria for the block-fading channel through the analysis of PEP performance bound under ML decoding. Two major parameters are diversity order and coding gain. Therefore, the code design approach is to find blockwise MDS codes with good coding gain.

Our first design is a BC-BICM scheme for SISO block-fading channels. Compared with BCC-BICM scheme, the proposed scheme also achieves the Singleton bound but has less delay in both encoding and decoding. We can improve the the coding gain of BC-BICM scheme through proper design of signal mapping scheme or using a strong convolutional code. Then, we study a QC-LDPC scheme, whose parity check matrices are designed using circulant matrices. The QC-LDPC achieves the same good performance as its corresponding root-LDPC scheme. The structure of the QC-LDPC code enables efficient encoding.

For MIMO cases, we investigated both serial and parallel concatenated turbo coded bit-interleaved coded modulation (BICM) schemes. The two schemes are compared in terms of performance in MIMO block-fading channels. In addition, a sphere decoder is used as a practical low-complexity signal processing solution, when the number of transmit antennas or the size of modulation constellation is large. The sphere decoder offers better performance compared with the group suppression technique. However, the achievable diversity order is not guaranteed with serial and parallel concatenated turbo coded BICM schemes. The systems designed for SISO case cannot be used directly in MIMO block-fading channels to achieve optimal diversity order because of the interference introduced by multiple transmitting antennas. To solve the diversity problem, a coded MIMO system based on modulation diversity technique is proposed. The scheme achieves both full spatial multiplexing and full diversity of $N_tN_rF$, i.e., the product of the number of transmit antennas $N_t$, receive antennas $N_r$ and fading blocks $F$. The drawback of modulation diversity is the increased complexity at the detector. We show how to employ a sphere decoder in the proposed system to reduce the signal processing complexity.
Chapter 2

A New BC-BICM Scheme for Block-Fading Channels

In this chapter, we introduce and study a blockwise convolutional coded bit-interleaved coded modulation (BC-BICM) scheme for the class of block-fading channels. The proposed BC-BICM scheme can be viewed as a simplified version of the blockwise concatenated convolutional coded bit-interleaved coded modulation (BCC-BICM) scheme [2], which achieves the Singleton bound as long as the blockwise partitioned outer code is a MDS code. Furthermore, we improve the coding gain by choosing better symbol mapping rules with iterative detection of BICM [36, 37] or using stronger convolutional codes. Simulations results are provided to confirm the effectiveness of our approach.

2.1 Blockwise Convolutional Codes for Block-Fading Channels

Non-recursive convolutional codes for block-fading channels are studied in [10, 16], where MDS convolutional codes with large product distance (coding gain) are designed using computer search techniques. We illustrate the search principle by an example. The coding rate of the example convolutional code is $r_c = 1/4$ with generator $G = (5, 3, 7, 7)$ in octal form. The encoder structure is shown in Fig. 2.1. The rate-1/4 code has 4 output taps and the coded bits from those taps are blockwisely partitioned according to the number of fading blocks. We assume that BPSK modulation is employed. For a block-fading channel
of 8 fading blocks, two input information bits will generate 8 coded bits, which are sent to 8 fading blocks respectively as illustrated in Fig. 2.1.

Figure 2.1: The encoder structure of (5, 3, 7, 7) convolutional code.

Because convolutional codes are linear, the distance property of a convolutional code can be studied by using the all-zero codeword as a reference. The trellis diagram of the code is illustrated in Fig. 2.2.

Figure 2.2: Minimum blockwise Hamming distance path for a block-fading channel with $F = 8$.

For a block-fading channel of $F = 8$ fading blocks, the maximum achievable diversity is 7 with a rate $r_c = 1/4$ code, which is given by the Singleton bound (1.23). However, the example code only provides a diversity order of $d_H[F] = 6$. The path with minimum blockwise Hamming distance is illustrated by red dotted arrows. For comparison, we also show the path corresponding to minimum Hamming distance (minimum free distance) by
blue solid arrows. From the example, we can see that the minimum Hamming distance path and the minimum blockwise Hamming distance are not necessarily the same path. The Hamming distance property of the convolutional code can be characterized by its transfer function [38]. The minimum blockwise Hamming distance can only be found using computer search, because it is also depended on the number of fading blocks. Convolutional codes with good blockwise Hamming distance for different values of $F$ are found by computer search [10, 16].

2.2 System Model

The transmitter structure of the BC-BICM scheme is shown in Fig. 2.3. The blockwise convolutional code encoder and the symbol mappers are serially concatenated with $F$ interleavers in between. The coded bits from the encoder are sent to each fading blocks by a blockwise partitioner.

For example in case of $F = 2$, a sequence of 8 coded bits $\{c(0), c(1), \cdots, c(7)\}$ is sent over the two fading blocks. The coded bits $c(0), c(2), c(4), c(6)$ are sent in the first block and $c(1), c(3), c(5), c(7)$ are sent in the second block. Since we use an individual bit-level interleaver on each fading block, the blockwise Hamming distance of the outer convolutional code is unchanged after interleaving. To obtain the the optimal diversity orders, we need to use a MDS convolutional code as the outer code. Such codes can be found by computer search and are tabulated in [10, 16]. Note that we need to choose MDS convolutional codes associated with BPSK modulation from these table because of bit-level interleaving is adopted at the transmitter.

At the receiver side, a standard belief-propagation (BP) iterative decoding algorithm is employed. The details of the iterative decoding algorithm for convolutional coded BICM
The channel decoder uses the BCJR algorithm [39] and the symbol demapper uses maximum a posteriori (MAP) algorithm. The extrinsic information is exchanged between the symbol demapper and the decoder via de-interleaver/interleaver, just like the decoding of process serially concatenated turbo codes. The receiver structure is illustrated in Fig. 2.4.

![Figure 2.4: The receiver structure of the BC-BICM scheme.](image)

The \( i \)th received signal from a fading block with coefficient \( \alpha_f \) can be expressed as

\[
y_i = \alpha_f x_i + n_i,
\]

where \( n_i \) is the noise. The transmitted symbol \( x_i \) is from a signal constellation \( \mathcal{X} \) of size \( 2^M \) and each symbol conveys \( M \) coded bits. Let us denote coded bits by \( c = [c^1, c^2, \cdots, c^M] \).

The a posteriori log-likelihood ratio (LLR) value for the \( l \)th coded bit can be calculated as

\[
\lambda(c^l) = \log \frac{\Pr[c^l = 1|y_i, \alpha_f]}{\Pr[c^l = 0|y_i, \alpha_f]} = \log \frac{\Pr[c^l = 1, y_i|\alpha_f]}{\Pr[c^l = 0, y_i|\alpha_f]}
\]

\[
= \log \sum_{c: c^l = 1} \frac{\Pr[y_i, c|\alpha_f]}{\Pr[y_i, c|\alpha_f]} = \log \sum_{c: c^l = 1} \Pr[y_i|c, \alpha_f] \Pr(c)
\]

\[
= \log \sum_{c: c^l = 0} \Pr[y_i|c, \alpha_f] \Pr(c).
\]

(2.2)

Under the condition of perfect interleaving, we have \( \Pr(c) = \prod_{j=1}^{M} \Pr(c^j) \), which indicates
those coded bits are independent. In this situation, (2.2) can be expressed as

$$\lambda(c^l) = \log \frac{\sum_{c'=1}^{c} \Pr[y_i|c, \alpha_f] \prod_{j=1}^{M} \Pr(c^j)}{\sum_{c'=0}^{c} \Pr[y_i|c, \alpha_f] \prod_{j=1}^{M} \Pr(c^j)}.$$  

(2.3)

Equation (2.2) shows the \textit{a posteriori} LLR value of coded bit $c^l$. The \textit{a priori} LLR value from convolutional code decoder is

$$\text{LLR}(c^l) = \log \frac{\Pr(c^l) = 0}{\Pr(c^l) = 1}.$$  

(2.4)

The extrinsic information $e(c^l)$ generated by symbol demapper is calculated by subtracting (2.4) from (2.2)

$$e(c^l) = \log \frac{\sum_{c'=1}^{c} \Pr[y_i|c, \alpha_f] \Pr(c)}{\sum_{c'=0}^{c} \Pr[y_i|c, \alpha_f] \Pr(c)} - \log \frac{\Pr(c^l) = 0}{\Pr(c^l) = 1},$$  

(2.5)

which is the input to the BCJR decoder.

\section*{2.3 BC-BICM with Improved Coding Gain}

In this section, we propose two methods to improve the coding gain of the BC-BICM scheme. In the first method, we improve the performance of the modulation part by employing better symbol labeling schemes. Then, we improve the performance of coding part by using convolutional codes with longer constraint lengths. Simulation results are provided to demonstrate the effectiveness of these methods.

\subsection*{2.3.1 Symbol Labeling for BC-BICM with Iterative Decoding}

Gray mapping is the best labeling method for non-iterative decoding of convolutionally coded BICM, because Gray labeling has the fewest number of nearest neighbors for the decision of each bit in each position regardless of its position in the coded sequence. However, the performance of Gray mapping is not as good when used in an iterative decoding scheme.
With an optimized signal labeling scheme, the feedback soft bits can enlarge the conditional intersignal Euclidean distance, which leads to a larger coding gain [36, 37, 40].

\[ R \]

\[ I \]

\( 1111 \) 1011 0011 0111

\( 1100 \) 1000 0010 0100

\( 1101 \) 1001 0011 0101

(a) Gray mapping
(Minimum SED = 1)

\( 1000 \) 1010 0010 0100

\( 0100 \) 0010 0000 0010

\( 0111 \) 0011 0001 0001

(b) SP mapping
(Minimum SED = 1)

\( 0100 \) 0010 0000 0010

\( 0111 \) 0011 0001 0001

\( 0110 \) 0011 0001 0001

(c) MSP mapping
(Minimum SED = 2)

Figure 2.5: 16-QAM labeling schemes: Gray, SP, MSP.

In the high SNR region, we can assume that the demapper receives ideal soft information about the coded bits from the feedback provided by the decoder. Therefore, the demapper can decide the value of each coded bit by considering only two signal points in the signal constellation. Hence the error performance is dominated by the conditional intersignal squared Euclidean distance (SED). The general labeling design rule is to find a constellation with optimized minimum conditional intersignal SED. Computer search results of optimized signal labeling for 8-PSK and 16-QAM constellations are given in [36, 40]. The 16-QAM constellation with Gray, set partitioning (SP), and modified set partitioning (MSP) mapping rules are shown in Fig.2.5. The minimum conditional intersignal SED is 1 for both Gray and SP labeling schemes. The minimum conditional intersignal SED is 2 for the MSP labeling scheme.

Fig. 2.6 shows the word error rate performance of the BC-BICM scheme with Gray, SP and the MSP mapping rules. The system employs a \( r_c = 1/4 \) convolutional code (3,1,1,1) and 16-QAM BICM. The three signal labeling schemes for the 16-QAM constellation can be found in [37]. The information length is \( K = 1024 \) and the number of fading blocks is \( F = 4 \). Simulation results are obtained with 10 iterations using the belief propagation decoding algorithm.

For comparison, the outage probability curve with Gaussian input and a system load
CHAPTER 2. BC-BICM SCHEME

Figure 2.6: WER of the BC-BICM scheme obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel with $F = 4$.

of $R = 1$ bit/sec/Hz is also plotted in Fig. 2.6. All three simulated systems attain maximum diversity order of 4 with the same MDS (3,1,1,1) convolutional code and 16-QAM constellation is employed. The system with Gray mapping has the least coding gain and the one with MSP mapping has the largest. There is about 6 dB SNR difference between the system with Gray mapping and the system with the MSP mapping at a WER of $10^{-3}$. By employing better mapping, significant coding gains are obtained without extra cost of increasing system complexity. We will discuss how to further improve the coding gain by increasing the constraint lengths of convolutional codes in the next section. The tradeoff is between performance and processing complexity (more memory).

2.3.2 Convolutional Codes with Longer Constraint Lengths

We can further improve the coding gain of the BC-BICM by using convolutional codes with longer constraint lengths.

In Fig. 2.7, we compare the performance of BC-BICM with convolutional codes of two different constraint lengths. The two solid lines represent the WER performance curves...
using the 2-state convolutional code (3,1,1,1). The two dashed lines depict the corresponding results when the 16-state convolutional code (37,33,31,25) is used. Both convolutional codes are rate-1/4 non-systematic codes. It is worth noting that both of the convolutional codes used in these simulations are MDS codes and all the other system configurations are similar to those used in Fig. 2.6. We have used only two kinds of signal labeling here, Gray mapping and SP mapping. From Fig. 2.7, we can see the performance of BC-BICM is enhanced by employing better convolutional codes for both Gray and SP labeling schemes.

Figure 2.7: WER of the BC-BICM scheme with different convolutional codes obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel.

2.4 Comparison between BC-BICM and BCC-BICM

The BC-BICM scheme can be viewed as a simplified version of the BCC-BICM scheme proposed in [2]. If the BCC-BICM scheme uses a convolutional code as the outer code and a trivial rate-1 identity code as the degenerate inner code, it becomes the BC-BICM scheme.

In Fig. 2.8, the simulated BCC-BICM scheme uses a convolutional code (3,1,1,1) as the outer code and a single memory accumulator (2/3) as the inner code. The information length
K is 1024 and the number of fading blocks is $F = 4$. Simulation results are obtained with 10 iterations. We have also included BC-BICM performance plots with the same information length and the same number of decoding iterations. We can see that there is about 1 dB difference between BC-BICM and BCC-BICM when both systems use the (3,1,1,1) outer convolutional code. However, BC-BICM scheme has a simple system architecture and low decoding complexity with the same outer code. If we use a more powerful 16-state (37, 33, 31, 25) outer convolutional code for the BC-BICM scheme, it can achieve the same good performance of the BCC-BICM scheme. Note that there is not much performance improvement, when the BCC-BICM scheme employs an outer convolutional code with longer constraint lengths [2].

![Figure 2.8: WER of the BCC-BICM scheme and BC-BICM scheme obtained by BP decoding of 10 iterations over independent Rayleigh block-fading channel with $F=4$.](image)

With the same outer code, the performance of the BC-BICM scheme is inferior to the BCC-BICM scheme. However, the proposed BC-BICM scheme has a simpler system structure and its decoding complexity is lower compared to the BCC-BICM scheme with the same outer code. Compared with the BC-BICM scheme, there are two layers of coding and
interleaving with the BCC-BICM transmitter, which will cause more delays in both encoding and decoding processes. Furthermore, the BC-BICM scheme can achieve the optimal diversity-rate tradeoff by using a MDS convolutional code and the coding gain of BC-BICM can be improved by either employing better labeling or using stronger convolutional codes.

It should be noted that turbo coded BICM schemes have the best performance when used with Gray mapper [37, 41]. Although other mapping methods offer larger coding gain than Gray mapping with iterative detection, however, the powerful concatenated coding contributes the larger part of the coding gain in the BCC-BICM scheme. Because the Gray mapping provides the best estimation in the first iteration, the decoder can generate relatively better feedbacks for the demapper which results in better decisions. As a result, the system with Gray mapping achieves good performance with the help of large coding gain provided by the decoder.

Another difference between BC-BICM and BCC-BICM schemes becomes apparent when more sophisticated codes are used. It has been shown in [2] that using more sophisticated outer codes does not result in much coding gain. In fact, simple repetition codes are good enough to be used as outer codes in the BCC-BICM scheme. However, as we have shown in Section 2.3.2 when BC-BICM codes are employed, using better convolutional codes increases coding gain.

### 2.5 Conclusions

In this chapter, we proposed a BC-BICM scheme for block-fading channels. Similar to the BCC-BICM scheme, the BC-BICM scheme achieves the optimal diversity-rate tradeoff by employing MDS convolutional codes. Compared with the BCC-BICM scheme, our scheme has a simpler structure and less delays in encoding and decoding processes. In addition, for this coding method we can obtain improved coding gain by using better signal labeling or convolutional codes with longer constraint lengths. The optimized BC-BICM scheme can achieve comparable performance as the BCC-BICM scheme. Simulation results are provided to confirm the benefits of the proposed coding scheme.
Chapter 3

Quasi-Cyclic LDPC Code Design for Block-Fading Channels

In [19], Boutros et al. proposed a family of low-density parity-check codes (LDPC), called root-LDPC codes, for nonergodic block-fading channels. The root-LDPC codes are shown to achieve full diversity on block-fading channels and perform close to the outage limit when decoded using the iterative belief-propagation (BP) decoding. In the bipartite graph representation of root-LDPC codes, a subset of connections are deterministically selected to guarantee full diversity for information bits and the rest of connections are generated randomly.

LDPC codes with randomly generated parity check matrices generally have good performance, but they lack enough structure to facilitate efficient encoding methods. For practical applications, it is suggested to use structured LDPC codes, which is the main motivation of this research. In this paper, instead of using random matrices, we construct parity check matrices of root-LDPC codes by tiling circulant matrices, i.e., by designing quasi-cyclic low-density parity-check (QC-LDPC) codes. The generator matrices of QC-LDPC codes are in systematic-circulant (SC) form with the requirement that the parity check matrices are full rank. We will describe how to construct full rank parity check matrices for QC-LDPC codes. The memory requirements for encoding of QC-LDPC codes are greatly reduced and the encoder can be implemented using simple shift registers [42]. With proper design, the QC-LDPC codes can perform as good as randomly generated root-LDPC codes over block-fading channels.
3.1 Linear Block and LDPC Codes

3.1.1 Linear Block Codes

A \(C(N, K, M)\) linear block code can be specified by either its parity check matrix \(H\) of dimensions \(M \times N\) or its generator matrix \(G\) of dimensions \(K \times N\), where \(M = N - K\) is the number of parity check bits in a block.

Because any linear block code has a systematic equivalent, without loss of generality we only consider systematic linear block codes. In a systematic code, a codeword \(c\) can be divided into two parts \(c = [u \, p]\), where \(u\) of length \(K\) represents the information bits and \(p\) of length \(M\) denotes parity check bits. Accordingly, the parity check matrix \(H\) can be represented as \(H = [A \mid B]\), where \(A\) is an \(M \times K\) matrix and \(B\) is an \(M \times M\) matrix, similarly \(G = [I \mid C]\), where \(I\) is a \(K \times K\) identity matrix and \(C\) is a \(K \times M\) matrix.

3.1.2 LDPC Codes

LDPC codes were first invented by Gallager in 1960 in his doctoral dissertation [43]. However, these codes were mostly ignored by the coding community for more than 30 years. It was not until 1996 that MacKay and Neal rediscovered LDPC codes [44]. It has been demonstrated that near capacity performance can be achieved by LDPC codes with iterative message passing decoding [45,46]. LDPC codes offer an attractive alternative to turbo codes in many communication system standards.

LDPC codes are linear block codes with sparse parity check matrices \(H\), meaning the number of zeros in each row and column of \(H\) is far more than the number of ones. Besides matrix representations, the structure of an LDPC code can be graphically represented by a bipartite graph (Tanner graph) introduced by Tanner [47]. Short cycles in the Tanner graph deteriorate the performance of LDPC codes when iterative belief propagation decoding is used. The smallest cycle in the bipartite graph is called the girth of the code. If the girth is small, the information from a bit will loop back to itself in a few iterations, which would violate the principle of belief propagation.

Most design approaches of LDPC codes are based on parity check matrices without considering encoding difficulties. Even if the parity check matrices are sparse, the generator matrices are not necessarily sparse. In such situations, the encoding will involve a large number of addition operations. To address this problem, structured LDPC codes [42,48] with efficient encoding method are proposed for practical applications instead of random
generated LDPC codes. One such class of codes is the class of quasi-cyclic low-density parity check codes (QC-LDPC codes) described in the next section.

### 3.1.3 QC-LDPC Codes

A **circulant matrix** is a square matrix in which each row is a right cyclic shift of the previous row and the first row is a cyclic shift of the last row. A circulant matrix can be totally characterized by its first row. If a circulant matrix has weight \( w = 1 \) per row, it is called a **circulant permutation matrix**. The **all-zero matrix**, also called the **null matrix**, is a circulant matrix with all its elements equal to zero. If a parity check matrix \( H \) consists of \( m \times n \) circulant submatrices of dimensions \( s \times s \) with \( M = ms \) and \( N = ns \), the resulting linear block code will be a quasi-cyclic (QC) code with a period of \( n \). In such a code, the \( n \)-bit shift of any codeword is another codeword. This can be easily demonstrated by dividing a codeword \( \tau \) into \( n \) tuples of size \( 1 \times s \), where right cyclic shifts of each tuple produces another codeword. The general form of the parity check matrix of QC-LDPC codes composed of circulant permutation matrices and null matrices can be written as

\[
H = [A | B] = \begin{bmatrix}
I(a_{1,1}) & I(a_{1,2}) & \cdots & I(a_{1,n}) \\
I(a_{2,1}) & I(a_{2,2}) & \cdots & I(a_{2,n}) \\
\vdots & \vdots & \ddots & \vdots \\
I(a_{m,1}) & I(a_{m,2}) & \cdots & I(a_{m,n})
\end{bmatrix},
\]

(3.1)

where \( I(a_{i,j}) \) is an \( s \times s \) circulant permutation matrix with \( a_{i,j} \in \{1, 2, \ldots, s\} \) indicating the location of “1” in the first row of \( I(a_{i,j}) \) and \( I(0) \) represents a null matrix. For example, if the codeword length is \( N = 1200 \) and the size of the circulant matrices is \( s = 75 \), then there exists \( m = 8 \) blocks along each column and \( n = 16 \) blocks along each row in \( H \).

The most attracting property of QC-LDPC codes is that the quasi-cyclic structure of QC-LDPC codes enables low-complexity encoding, which will be discussed in Section 3.4. These codes also have memory efficient encoding and decoding algorithms. In addition, QC-LDPC codes with proper design perform as well as random LDPC codes [42]. The QC-LDPC codes have been adopted by IEEE 802.16e WiMAX [5] and IEEE 802.11n WLAN [35].
3.2 QC-LDPC Code Design for Block-Fading Channels

Although LDPC codes tend to be universally good over many channel models, neither random LDPC codes nor ML-designed blockwise MDS LDPC codes can achieve the optimal diversity order when suboptimal BP iterative decoding is used. Boutros et al. addressed this problem by designing a family of root-LDPC codes for block-fading channels [42], which can achieve full diversity order under iterative decoding.

In this chapter, we present a method to design root-LDPC codes using circulant submatrices without affecting the crucial root connections. The resulting structured QC-LDPC codes perform similar to the root-LDPC codes with random submatrices if they are carefully designed. Additionally, the QC structure makes efficient encoding of the QC-LDPC codes possible.

3.2.1 Root-LDPC Codes

![Figure 3.1: A rate $r_c = 1/2$ regular $(3, 6)$ root-LDPC code for a block-fading channel with $F = 2$.](image)

Figure 3.1: A rate $r_c = 1/2$ regular $(3, 6)$ root-LDPC code for a block-fading channel with $F = 2$. 

Root-LDPC codes employ degree-1 root connections which link each information bit to the bits transmitted on other fading blocks. Full diversity is guaranteed for all information bits with the help of root connections under message passing iterative decoding. Therefore, an information bit can be detected by either the initial observation of it or by the extrinsic information from the bits transmitted during other fading blocks. The details of the diversity analysis for root-LDPC codes over block-fading channels can be found in [19]. The rate of root-LDPC codes is fixed at $R = 1/F$, which is the highest possible rate to achieve full diversity over nonergodic block-fading channels.

The construction of a root-LDPC code is illustrated by an example in Fig. 3.1. This is a rate $R = 1/2$ regular $(3, 6)$ root-LDPC code for channel with $F = 2$ fading blocks. $N/2$ information bits are divided evenly into two groups, denoted by $\{1i\}$ and $\{2i\}$. The first group of information bits $\{1i\}$ are transmitted on fading block-1 and the second group $\{2i\}$ on fading block-2. Similarly, $N/2$ parity bits are partitioned into $\{1p\}$ and $\{2p\}$ and transmitted on the corresponding two fading blocks. The parity check matrix $H$ in Fig. 3.1(b) can be obtained by mapping connections of the bipartite Tanner graph in Fig. 3.1(a). Except for the 1’s and 0’s explicitly shown in parity check matrix $H$ as root connections, other parts of $H$ can be generated randomly.

### 3.2.2 Construction of QC-LDPC Codes

The parity check matrix $H$ of the LDPC code is split into two submatrices $A$ and $B$ as shown in Fig. 3.1(b). For a codeword $c = (u, p)$, we have $Hc^T = Au^T + Bp^T = 0$. For binary codes, the parity bits can be obtained from

$$p^T = (B)^{-1}A\bar{u}^T.$$  \hspace{1cm} (3.2)

As long as we can obtain $B^{-1}$, the encoding can be accomplished by using (3.2) with linear complexity as a function of the length of parity check bits. However, there are two main concerns with obtaining $(B)^{-1}$ in practical applications:

1. $(B)^{-1}$ can be difficult to compute or $B$ may not be invertible.

2. Although $H$ is sparse, but $(B)^{-1}$ is not generally sparse and large memory may be required to store it.
QC-LDPC codes provide a solution to both problems. We can devise a parity check matrix $H$ with certain structure to ensure $B$ is full rank. Furthermore, it can be shown that when $B^{-1}$ exists, the generator matrix $G$ of QC-LDPC codes will be in quasi-cyclic (QC) form \cite{42}, which greatly decreases the required memory to store $G$.

We only consider tiling circulant permutation matrices and null matrices of size $s \times s$ in designing the parity check matrices of QC-LDPC codes. Based on the design rules in \cite{42} and \cite{49}, we impose the following requirements on the parity check matrices of QC-LDPC codes:

1. **Sparsity**: The weight $w$ per row of the circulant matrices is small compared with its size $s$.

2. **Full rank**: $B$ is full rank, i.e., $\text{rank}(B) = M = N - K$.

3. **Girth**: There are no length-4 cycles in $H$.

Condition 1) guarantees that the resulting parity check matrix $H$ is sparse. Condition 2) facilitates efficient encoding, which will be discussed in details later. Condition 3) improves code performance.

A construction algorithm for designing QC-LDPC Codes:

- The two all-zero submatrices in the original $H$ are unchanged in order to keep the root connections.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3_2}
\caption{Illustration of construction of QC-root-LDPC Codes.}
\end{figure}
- Other parts of $H$ are modified into the QC form as shown in Fig. 3.2, where each $I(a_{ij})$ is a circulant permutation matrix and each $0$ is a null matrix.

- The value of $a_{ij}$ is a random variable uniformly distributed over \{1, 2, \ldots, s\}.

- The two identity matrices located in the top-left and bottom-right corners of matrix $A$ are replaced by matrices with circulant permutation submatrices on the diagonal. This change will not affect the root connections, as each information node still has degree-1 connection to the bits on the other block.

- Two blocks in each row of both $H_{1i}$ and $H_{2i}$ are filled with zeros to keep the code regular. These blocks can be filled with circulant permutation matrices in a general design.

- The diagonal part of new $H_{1p}$ and $H_{2p}$ are tiled with the null matrices in order to meet the full rank requirement.

- We simply discard current set of $\{a_{ij}\}$ and regenerate them if either of followings happens.
  1. Either $H_{1p}$ or $H_{2p}$ is singular.
  2. $H$ has girth-4 loops.

We will discuss how to check and optimize girth of QC-LDPC codes in the next section. The rank check is carried out by Gaussian elimination. We do not need to check matrix $B$ as a whole. Instead the rank check is done on $H_{1p}$ and $H_{2p}$ individually, which reduces the computational complexity. If only circulant permutation matrices are used in $H_{1p}$ and $H_{2p}$, we will always have redundant rows. The reason is that the sum of rows in a circulant permutation matrix is an all-one vector. However, inserting null matrices along the diagonal of $H_{1p}$ and $H_{2p}$ does not guarantee $H_{1p}$ and $H_{2p}$ are full rank with random generated $a_{i,j}$'s. Insertion of null matrices makes it possible to meet the full rank requirement.

This construction can be generalized to put other random root-LDPC codes with rate $R = 1/F$ into QC forms.

### 3.2.3 QC-LDPC Codes with Large Girth

The structure of QC-LDPC codes also facilitate the design of codes with desired girth. In Fig. 3.3, we show an example of length-4 cycle. A closed path of length-2g is formed by
traversing through $2g$ circulant permutation submatrices in $H$ while moving along rows or columns alternatively. The submatrices passed through must be different and the resulting closed path can be represented by ordered indices of those submatrices as

$$(i_1, j_1), (i_2, j_1), (i_2, j_2), \ldots (i_g, j_g), (i_1, j_g).$$

(3.3)

Defining

$$\Delta_{i_k, i_{k+1}}(j_k) = a_{i_k, j_k} - a_{i_{k+1}, j_k},$$

(3.4)

a necessary and sufficient condition [48, Theorem 2.1] for the code to have girth of at least $2(g + 1)$ is that

$$\sum_{k=1}^{v} \Delta_{i_k, i_{k+1}}(j_k) \neq 0 \mod s$$

(3.5)

for all $v$, $2 \leq v \leq g$, all $i_k$, $1 \leq i_k \leq m$, all $i_{k+1}$, $1 \leq i_{k+1} \leq m$, and all $j_k$, $1 \leq j_k \leq n$, with $i_1 = i_{v+1}$, $i_k \neq i_{k+1}$ and $j_k \neq j_{k+1}$. Although we also insert null matrices in $H$, the above condition can still be applied with a slight modification that neither $a_{i_k, j_k}$ nor $a_{i_{k+1}, j_k}$ can be 0 in (3.4).
After a QC-LDPC code is randomly generated, we can check the girth of the code using (3.5). We can regenerate $a_{i,j}$’s until a code of desired girth is found. Generally speaking, a large value of $s$ will speed up the code search, as there is more freedom for choosing $a_{i,j}$’s.

### 3.3 Simulation Results

In this section, we study the performance of the proposed QC-LDPC codes when used over a Rayleigh block-fading channel with $F = 2$ independent fading blocks. All LDPC codes simulated here are (3, 6) regular LDPC codes with a rate of $R = 1/2$. BPSK modulation is used throughout our simulations. The outage limits with Gaussian input and BPSK input are drawn as a thick solid line and a thick dashed line respectively in each figure for reference.

![Figure 3.4: FER performance comparison for QC root LDPC codes, random root LDPC codes, random LDPC codes and ML designed LDPC codes over a block-fading channel with $F = 2$.](image)

In Fig. 3.4, we compare the FER performance between proposed QC-root-LDPC codes and random root-LDPC codes. The codeword length $N$ is 1200 and size of circulant $s$ is
75. Standard sum-product algorithm is employed at the decoder with 20 iterations. The performance of randomly generated LDPC codes and ML-designed LDPC codes [19] are also drawn. From the simulation results, we can see that QC-root-LDPC codes and random root-LDPC codes have the same FER performance, which is close to the outage limit of the block-fading channels. In fact, both QC-root-LDPC codes and random-root-LDPC codes achieve the full diversity order of the channel, while randomly generated LDPC codes and ML-designed LDPC codes fail to. The detailed analysis of why ML-designed LDPC codes fail to achieve full diversity with BP iterative decoding can be found in [19].

![Figure 3.5: FER performance comparison of QC-root-LDPC codes with girth $2g = 4, 6$ and $8$ over a block-fading channel with $F = 2$.](image)

In Fig. 3.5, we compare the FER performance of QC-LDPC codes with different girths. The codeword length $N$ is 1200 with $s = 75$. We increase the iteration number to 30 in this case. QC-LDPC codes with girth-6 and girth-8 are generated randomly using the construction method we described in Section 3.2. QC-LDPC code with girth-4 is designed by setting all the circulant permutation submatrices in its parity check matrix $H$ to be identity matrices, which is the worst case with QC-LDPC codes. In this situation, it has the highest possible number of length-4 cycles. We produce this girth-4 code in order to
illustrate that length-4 cycles degrade system performance. In spite of bad coding gain, girth-4 QC-LDPC code can still achieve the full diversity of the channel as girth-6 and girth-8 QC-LDPC do. We can also notice that increasing of girth from 6 to 8 has little improvement in performance. Similar observations were made in [48].

![Figure 3.6: FER and BER performance of QC-root-LDPC codes with codeword length $N$ = 160, 1200 and 8000 over a block-fading channel with $F = 2$. The number of iterations is 20.](image)

Fig. 3.6 shows FER and bit error rate (BER) performance of QC-LDPC codes with different codeword lengths. Three codeword lengths $N$ = 160, 1200 and 8000 are used and corresponding submatrix block sizes are $s = 10, 75$ and 250. So there are 16 submatrices in each row and each column. The resulting LDPC codes are $(3, 6)$ regular LDPC codes with a rate of $R = 1/2$ in all three cases. We observe that QC-LDPC codes with different length have similar error performance.
3.4 Efficient Encoding Method

The generator matrix $G$ in systematic form can be written as

$$
G = [I | C] = \begin{bmatrix}
I & 0 & \cdots & 0 & C_{1,1} & C_{1,2} & \cdots & C_{1,m} \\
0 & I & \cdots & 0 & C_{2,1} & C_{2,2} & \cdots & C_{2,m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I & C_{k,1} & C_{m,2} & \cdots & C_{k,m}
\end{bmatrix},
$$

(3.6)

where $C_{i,j}$ is a submatrix of size $s \times s$. As $HG^T = A + BC^T = 0$, for binary codes we have

$$
C = (B^{-1}A)^T = A^T(B^{-1})^T.
$$

(3.7)

Since the inverse, sum and product of circulant matrices are themselves circulant, $(B^{-1})^T$ consists of $m \times m$ circulant matrices of size $s \times s$

$$
(B^{-1})^T = \begin{bmatrix}
D_{1,1} & D_{1,2} & \cdots & D_{1,m} \\
D_{2,1} & D_{2,2} & \cdots & D_{2,m} \\
\vdots & \vdots & \ddots & \vdots \\
D_{m,1} & D_{m,2} & \cdots & D_{m,m}
\end{bmatrix}.
$$

(3.8)

The block entries $\{C_{i,j}\}$ in the generator matrix $G$ can be calculated as

$$
C_{i,j} = \sum_{q=1}^{m} I(a_{q,i})D_{q,j} \\
= I(a_{1,i})D_{1,j} + I(a_{2,i})D_{2,j} + \cdots + I(a_{m,i})D_{m,j}.
$$

(3.9)

From this equation, we can see that $C_{i,j}$ is also a circulant matrix. The memory requirement for the generator matrix of QC-LDPC codes is reduced by a factor of $1/s$, since we only need to store the first row of $C_{i,j}$’s. The encoding of QC-LDPC codes can be done using shift registers, as studied in [42].
3.5 Conclusions

We have studied QC-LDPC codes for nonergodic block-fading channels. It has been shown that reconstructing the root-LDPC codes with circulant matrices will not affect root connections, which guarantee achieving full diversity. We also proposed several conditions on the QC-LDPC codes to make them have good performance and facilitate their encoding. The resulting QC-LDPC codes perform similarly to the corresponding random root-LDPC codes. Furthermore, the encoders of the proposed QC-LDPC codes use less memory and encoding can be done with a computational complexity that is linearly proportional to the number of parity bits.
Chapter 4

Turbo Coded BICM Scheme with Iterative Detection for MIMO Block-Fading Channels

In this chapter, we investigate turbo coded bit interleaved coded modulation (BICM) schemes for multiple-input multiple-output (MIMO) block-fading channels. Both serial and parallel concatenated turbo codes are considered and their performance is compared. A sphere decoder is employed to achieve good performance with reduced computation complexity when the size of the constellation or the number of transmit antenna is large. Simulation results are provided to demonstrate the effectiveness of the proposed schemes in MIMO block-fading channels.

It has been demonstrated in [50] that turbo coded modulation (with parallel concatenated turbo codes) is an effective communication scheme for MIMO block-fading channels without complicated space-time coding. However, no analysis is presented to show whether these codes can achieve optimal diversity. In fact, although a multiplexing method is suggested in [17] which can ensure full diversity of parallel turbo codes in block-fading channels, the general design of multiplexer which can achieve optimal diversity-rate tradeoff is not yet known. Here, we consider employing blockwise partitioned serial concatenated turbo codes used in [2] as an alternative coding scheme for parallel turbo codes in a coded modulation system. The proposed serial concatenated turbo coded bit interleaved coded modulation (BICM) scheme can archive optimal diversity order given by the Singleton bound under
the assumption of perfect interference removal. We compare the performance of serial and parallel concatenated turbo coded BICM schemes in MIMO block-fading channels.

When the constellation size or the number of transmit antennas is large, an interference group suppression algorithm is used in [50] to reduce complexity. The group suppression algorithm is of low computation complexity, but achieves only moderate performance. Especially, because a single encoder’s output is transmitted on all layers, the system performance does not benefit from successful detection of previous layers. This is the motivation to employ a sphere decoder [51, 52] to alleviate computational burden at the detector and to achieve good performance.

4.1 System Model of Turbo Coded BICM Scheme

The system model of parallel concatenated turbo coded BICM can be found in [50]. A random bitwise interleaver is inserted between the parallel turbo encoder and the modulator. However, there is no guarantee about the achievable diversity order with the parallel concatenated turbo coded BICM scheme in block-fading channels.

We propose a serial concatenated turbo coded BICM scheme. A blockwise concatenated coded BICM scheme was proposed in [2] to achieve optimal diversity-rate tradeoff in SISO block-fading channels. From the discussions on block-fading MIMO channel model in Chapter 1, we can apply the above serial concatenated turbo BICM scheme to the MIMO cases. Under the assumption of perfect multiple-antenna interference cancellation, communication schemes designed for SISO system can be applied to MIMO case.

4.1.1 Transmitter Structure in Parallel Concatenated Turbo Coded BICM

The transmitter structure of parallel concatenated turbo coded BICM is shown in Fig. 4.1. The parallel concatenated turbo codes were invented by Berrou et al. in 1993 [53]. The turbo codes illustrated in Fig. 4.1 consist two recursive systematic convolutional codes as component codes. A random interleaving is applied to the information bits, before they are passed into the second encoder. To obtain a rate \( r_c = 1/2 \) turbo code, the coded bits are punctured regularly. The internal interleaving of the turbo codes has the purpose to avoid low weight codewords. As the inputs of the second encoder is scrambled, the probability that the outputs of both recursive convolutional code encoders are low weight is rare. This means that the number of nearest-neighboring codes are reduced. Another
The function of interleaving is to make the outputs of two encoder uncorrelated. At the receiver, two convolutional code decoders can exchange hypotheses of information bits.

The encoder outputs are scrambled by a bit-level interleaver. Then those coded bits are de-multiplexed into $N_t$ streams, which are transmitted on $N_t$ transmit antennas after the symbol mapping. The function of bit-level interleaving is to spread adjacent coded bits onto different symbols. At the decoder, we assume the coded bits in a symbol carry independent information.

Figure 4.1: Block diagram of the parallel concatenated turbo coded BICM transmitter.

### 4.1.2 Transmitter Structure for Serial Concatenated Turbo Coded BICM

Concatenated codes were first introduced in form of serial concatenated codes by Forney [54]. The concatenation coding scheme is constructed by serially concatenating an inner code with an outer code. The outer code in original design is a Reed-Solomon code with long codeword length. The inner code is a convolutional code with short constraint length. The classical serial concatenated codes were used for space communications in 1970s. Inspired by the invention of “turbo codes”, new serial concatenated codes with interleavers and their corresponding iterative decoding method were studied in [55]. These codes are usually referred as serial concatenated turbo codes.

The block diagram of the serial concatenated turbo coded BICM transmitter is shown in Fig. 4.2. The output of the outer encoder is partitioned into $N_t \times F$ blocks. Each block is permuted by an individual interleaver and fed to the corresponding inner encoder. As the inner interleaver is a blockwise interleaver, the block Hamming distance of the outer code is retained. Therefore, we can employ a maximum distance separable (MDS) code as the
outer code, which enables the system to achieve the Singleton bound in the SISO case.

![Serial concatenated turbo code encoder](image)

Figure 4.2: Block diagram of the serial concatenated turbo coded BICM transmitter.

4.1.3 Iterative Detection

The receiver is composed of two parts as shown in Fig. 4.3: a signal processing unit in the front end followed by a turbo decoder. Both signal processing unit and decoder accept soft inputs and generate soft outputs. The extrinsic information is exchanged between two parts in an iterative manner.

![Signal processing unit](image)

Figure 4.3: Block diagram of the receiver with iterative detection and decoding.

The structure of turbo decoder is different between parallel concatenated turbo codes and serial concatenated turbo codes. The details of turbo code decoding algorithm are elaborated in [53]. In this section, we will focus on the signal processing part.

The goal of the signal processing is to derive soft reliability of each coded bit, which will be used as inputs of the decoder. From (1.26), the received signal at antenna $i$ can be
expressed as
\[ y_i = \sum_{j=1}^{N_t} h_{i,j} x_j + n_i, \] (4.1)
where the time index \( t \) is dropped for convenience. \( N_t \) symbols are transmitted simultaneously on \( N_t \) transmit antennas
\[ x = [x_1, x_2, \cdots, x_{N_t}]. \] (4.2)

As each symbol \( x_j \) consists of \( M \) coded bits, those \( N_t \) symbols are mapped using a vector of coded bits of length \( MN_t \)
\[ c = [c_1, c_2, \cdots, c_M, \cdots, c_{N_t}]. \] (4.3)

The a posteriori log-likelihood ratio (LLR) value for the \( l \)th coded bit \( c_l \) in \( c \) is
\[
\lambda(c_l) = \log \frac{\Pr[c_l = 1|y_1, \cdots, y_{N_t}]}{\Pr[c_l = 0|y_1, \cdots, y_{N_t}]} \sum \Pr[y_1, \cdots, y_{N_t}, c] \\
= \log \sum_{c_l=1} \Pr[y_1, \cdots, y_{N_t}, c] \\
= \log \sum_{c_l=0} \Pr[y_1, \cdots, y_{N_t}, |c]\Pr(c) \\
= \log \sum_{c_l=1} \Pr[y_1, \cdots, y_{N_t}, |c]\Pr(c). \] (4.4)

We apply following approximation in above algorithm
\[
\log(e^a + e^b) = \max(a, b) + \log(1 + e^{-|a-b|}) \approx \max(a, b). \] (4.5)

At high SNR, the correction term \( \log(1 + e^{-|a-b|}) \) can be ignored, which leads a small degradation in performance. In (4.4), the log-sum computation can be simplified by max-log approximation
\[
\log \sum_{c_l=1} \Pr[y_1, \cdots, y_{N_t}, |c]\Pr(c) \approx \max_{c_l=1} \{ \log \Pr[y_1, \cdots, y_{N_t}, |c]\Pr(c) \}. \] (4.6)
CHAPTER 4. TURBO CODED BICM

The approximation is referred as max-log-map algorithm

$$\lambda(c_l) = \max_{c_{l}=1} \{ \log \Pr[y_1, \ldots, y_{N_r}|c] \Pr(c) \} - \max_{c_{l}=0} \{ \log \Pr[y_1, \ldots, y_{N_r}|c] \Pr(c) \} \tag{4.7}$$

4.2 The Array Processing Method

The computation complexity of single coded bit LLR grows exponentially with the number of transmit antennas $N_t$ and the number of bits per symbol $M$ in (4.4). Note that even using the approximation given in (4.6), the complexity still grows exponentially, as we need to compare $2^{MN_t}$ terms instead of summing them. The front end signal processing unit becomes prohibitively complex when $N_t \times M$ is large. A low complexity group interference suppression technique (array processing) is employed to reduce the detection processing complexity in [50]. The details of the array processing detection method can be found in [56].

The array processing method first divides $N_t$ transmit antenna into $q$ small groups, with $n_1, n_2, \ldots, n_q$ antennas per group. The group interference suppression technique process each group separately by suppressing the interference from other groups. The array processing method is similar to SIC processing for VBLAST, which is briefly described in Section 1.4.1. The array processing method detects each group $G_j$ separately by suppressing signal from other groups. The method has a low processing complexity at the cost of reduced diversity order. We consider the case of spatial multiplexing in our scheme, where independent multi-layer streams are transmitted on $N_t$ transmit antennas. If ML detector is employed, the receiver can achieve a diversity order of $N_r$, i.e., full receive diversity order. With a linear detector, like ZF and linear MMSE detectors, the receiver can achieve a diversity order of $N_r - N_t + 1$. The linear detector utilizes $N_t - 1$ degrees of receive freedom to suppress $N_t - 1$ interfering signals [9, Chapter 15]. The array processing method takes the trade-off between complexity and diversity gain. The method employs ML detector among each group after canceling the interference from other group using linear detection. The diversity order for group $j$ of size $n_j$ is $N_r - (N_t - n_j)$ and it is required that $(N_r - N_t + n_j) > 0, \forall j$ for array processing method.

Without loss of generality, we can describe the array processing method by focusing on the detection process of group $G_1$. The received signal can be expressed as

$$y = Hx + n, \tag{4.8}$$
where $\mathbf{x}$ is the transmitted symbol vector of size $N_t$, $\mathbf{y}$ is the received signal vector of size $N_r$ and $\mathbf{H}$ is channel matrix of dimensions $N_r \times N_t$. The channel matrix can be divided into two parts

$$
\mathbf{H} = [\Omega_{G_1} \Lambda_{G_1}],
$$

(4.9)

where the first part has $n_1$ columns

$$
\Omega_{G_1} = \begin{bmatrix}
h_{1,1} & h_{1,2} & \cdots & h_{1,n_1} \\
h_{2,1} & h_{2,2} & \cdots & h_{2,n_1} \\
\vdots & \vdots & \ddots & \vdots \\
h_{N_r,1} & h_{N_r,2} & \cdots & h_{N_r,n_1}
\end{bmatrix},
$$

(4.10)

and the second part has the $N_t - n_1$ columns

$$
\Lambda_{G_1} = \begin{bmatrix}
h_{1,n_1+1} & h_{1,n_1+2} & \cdots & h_{1,N_t} \\
h_{2,n_1+1} & h_{2,n_1+2} & \cdots & h_{2,N_t} \\
\vdots & \vdots & \ddots & \vdots \\
h_{N_r,n_1+1} & h_{N_r,n_1+2} & \cdots & h_{N_r,N_t}
\end{bmatrix},
$$

(4.11)

To cancel the $(N_t - n_1)$ interference signals from the other $(q - 1)$ groups, we construct a nulling matrix $\Theta_{G_1}$ of dimensions $(N_r - (N_t - n_1)) \times N_r$, where we have $0 = \Theta_{G_1} \Lambda_{G_1}$. The nullity of $\Lambda_{G_1}$, i.e., the dimension of the NULL space $\mathcal{N}$ of $\Lambda_{G_1}$

$$
\dim(\mathcal{N}) = N_r - \text{rank}(\Lambda_{G_1}) \geq N_r - N_t + n_1.
$$

(4.12)

The $(N_r - N_t + n_1)$ rows in $\Theta_{G_1}$ are a set of orthonormal vectors (not necessarily unique) from the NULL space $\mathcal{N}$ of $\Lambda_{G_1}$

$$
\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{N_r-N_t+n_1}\}.
$$

(4.13)

There are several numerical methods to compute the NULL space $\mathcal{N}$ of the matrix $\Lambda_{G_1}$, including Gaussian elimination (column or row reduction), QR decomposition and singular value decomposition (SVD). We choose QR decomposition, which has good numerical
stability and moderate computation complexity. The matrix $\mathbf{A}_{G_1}$ can be factorized as

$$\mathbf{A}_{G_1} = \mathbf{Q}_{G_1} \mathbf{R}_{G_1} = \begin{bmatrix} \mathbf{Q}_{G_1}^1 & \mathbf{Q}_{G_1}^2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{G_1}^1 \\ 0 \end{bmatrix}, \quad (4.14)$$

where $\mathbf{Q}_{G_1}$ is an $N_r \times N_r$ unitary matrix and $\mathbf{R}_{G_1}$ is an $N_r \times (N_t - n_1)$ upper triangular matrix. The matrix $\mathbf{Q}_{G_1}$ consists of two parts, $\mathbf{Q}_{G_1}^1$ and $\mathbf{Q}_{G_1}^2$. The first part $\mathbf{Q}_{G_1}^1$ is composed of $(N_r - N_t + n_1)$ columns, which span the NULL space $\mathcal{N}$ of the matrix $\mathbf{A}_{G_1}$. The actual computation of QR decomposition can be performed by Householder transformation [57,58].

After obtaining the nulling matrix $\Theta_{G_1}$, we multiply both sides of (4.8) by $\Theta_{G_1}$

$$\Theta_{G_1} \mathbf{y} = \Theta_{G_1} \mathbf{H} \mathbf{x} + \Theta_{G_1} \mathbf{n}. \quad (4.15)$$

The new input-output relation can be simplified as

$$\mathbf{y}_{G_1} = \mathbf{H}_{G_1} \mathbf{x}_{G_1} + \mathbf{n}_{G_1}, \quad (4.16)$$

where

$$\begin{cases} \mathbf{y}_{G_1} = \Theta_{G_1} \mathbf{y}, \\ \mathbf{H}_{G_1} = \Theta_{G_1} \mathbf{H}, \\ \mathbf{x}_{G_1} = [x_1, x_2, \cdots, x_{n_1}]^T, \\ \mathbf{n}_{G_1} = \Theta_{G_1} \mathbf{n}. \end{cases} \quad (4.17)$$

As $\mathbf{Q}_{G_1}$ is an $N_r \times N_r$ unitary matrix, the process does not enhance noise vector. In (4.16), the interference signals from the rest $(q - 1)$ groups are removed by linear nulling operation. Then we can apply (4.4) to calculate coded bit reliabilities within group $G_1$.

### 4.3 Sphere Decoder with Max-Log MAP Performance

The array processing method achieves the low demodulation complexity by suppressing the interference from other groups. The cost of decreased complexity is the performance degradation due to reduced diversity order. Unlike ZF (inverse channel detector), the array processing method does not enhance the noise power. The method actually utilizes receive diversity to cancel interference from other groups. In addition, the group interference suppression technique was originally designed for systems with individual codes per layer.
Such system benefits from previous successfully decoded layers. The performance of the group interference suppression technique deteriorates in the case where a single code is used for all layers of the system. This motivates us to employ sphere decoder \[51, 52\] to reduce processing complexity and achieve good performance.

The basic idea of sphere decoding is illustrated in Fig. 4.4. The skewed lattice represents the distorted signal points. The sphere decoding algorithm will only search signal points that stay within the sphere of radius \(r\) and centered at received signal \(y\). However, it is difficult to set the value of radius \(r\). If \(r\) is too large, the decoding complexity will be still high, because the algorithm will still considers many points. If \(r\) is too small, we may not include any signal point in the sphere. Another difficulty about sphere decoding is how to efficiently check which point lies within the sphere. Basically, we need to calculate the distance from the received signal to signal point

\[
r^2 \geq |y - Hx|^2.
\]

(4.18)

If we calculate the above distance explicitly for each signal point in the constellation, we do not gain any complexity advantage of using sphere decoding over maximum-likelihood (ML) decoding.

Different versions of sphere decoding algorithm have been proposed in literatures \[51, 59, 60\]. Almost all of these sphere decoding algorithms are based on QR factorization of the channel matrix \(H\), which breaks the distance calculation in (4.18) into stages. The calculation is terminated at the early stage, if the partial distance exceeds the radius \(r\). The detection algorithm can be implemented as a depth first tree search.
Recently, a soft-input soft-output single tree search sphere decoder was proposed in [52]. Their algorithm is capable to incorporate a priori information and generate max-log soft estimation with reduced complexity. Assuming $N_r \geq N_t$, the algorithm starts with QR-decomposition of the channel matrix

$$H = \begin{bmatrix} Q & Q' \\ 0 & R \end{bmatrix} = QR,$$  

(4.19)

where $[Q \ Q']$ is an $N_r \times N_r$ unitary matrix and $Q$ represents its first $N_t$ columns. $R$ is a $N_t \times N_t$ upper-triangular matrix. After preprocessing, we have the modified input-output relation

$$\tilde{y} = Q^H y = Rx + Q^H n,$$  

(4.20)

where $Q^H$ denotes the conjugate-transpose of $Q$ and components in $Q^H n$ are independent complex Gaussian random variable of variance $N_0$. Then the max-log LLRs can be calculated as

$$\lambda(c_l) = \min_{c_l = -1} \left\{ \frac{1}{N_0} |\tilde{y} - Rx|^2 - \log Pr(c) \right\}$$

$$- \min_{c_l = 1} \left\{ \frac{1}{N_0} |\tilde{y} - Rx|^2 - \log Pr(c) \right\}.$$  

(4.21)

We defined MAP metric $\lambda^{MAP}$, which is associated with the MAP solution as

$$\lambda^{MAP} = \min_{\text{all } c} \left\{ \frac{1}{N_0} |\tilde{y} - Rx|^2 - \log Pr(c) \right\}.$$  

(4.22)

And the competitor metric for coded bit $c_l$ is defined as

$$\lambda^{MAP}(c_l) = \min_{c_l = -c_l^{MAP}} \left\{ \frac{1}{N_0} |\tilde{y} - Rx|^2 - \log Pr(c) \right\}.$$  

(4.23)

With above definitions, the log-MAP output can be express as

$$\lambda(c_l) = \begin{cases} 
\lambda^{MAP}(c_l) - \lambda^{MAP}, & c_l^{MAP} = 1 \\
\lambda^{MAP} - \lambda^{MAP}(c_l), & c_l^{MAP} = -1.
\end{cases}$$  

(4.24)

The distance

$$d(c) = \frac{1}{N_0} |\tilde{y} - Rx|^2 - \log Pr(c)$$  

(4.25)

in (4.22) and (4.23) can be calculated recursively using the fact that $R$ is an upper-triangular
matrix. We define distance increment \( e_i \) for step-\( i \) as

\[
e_i = \frac{1}{N_0} |\tilde{y}_i - \sum_{j=i}^{N_t} R_{i,j} x_j|^2 - \log \Pr(c_{iM+1}, \cdots, c_{(i+1)M}).
\] (4.26)

The distance \( d(c) \) can be calculated with partial distance (PD) \( d_i \) in the following recursion:

- Initialization: \( d_{Nt+1} = 0 \);
- Recursive evaluation: \( d_i = d_{i+1} + e_i \);
- End: \( d(c) = d_1 \).

With above recursive method of evaluating \( d(c) \), the detection can be transformed into a tree traversal problem. An example of sphere decoding tree for BPSK constellation and \( N_t = 4 \) transmit antennas is shown Fig. 4.5.

![Figure 4.5: Sphere decoding search tree for BPSK constellation with \( N_t = 4 \).](image)

The Tree Pruning Criterion

The sphere decoding method proposed in [52] adopts a technique known as radius reduction [61]. In this way, the search avoids the problem of choosing a proper radius \( r \). The algorithm initializes \( \lambda^{MAP} \) and \( \{\lambda^{MAP}(c_l)\}, \forall l \) to be \( \infty \), which guarantees the MAP solution for every coded bits. The method updates the radius using \( d(c) \rightarrow r^2 \), whenever a valid leaf in the
tree is reached. The depth-first search will remove a node’s and its children from further consideration, if current partial distance is already larger than the radius constraint

\[ d_i \geq r^2. \]  

(4.27)

Tree traversal is performed depth-first in sphere decoding. The search starts at the root and goes deeper in the tree until a leaf is reached or the current node is abandoned according to the general pruning criterion 4.27. The children of a node are visited in ascending order of their partial distance.

The tree traversal is confined within the subtree, which can lead to an update of either \( \lambda^{MAP} \) or any counter-hypotheses \( \lambda^{MAP}(c_l) \). Let us assume the search arrives at a node at the level-\( i \) in the tree search.

1. The counter-hypotheses \( \{\lambda^{MAP}(c_l)\} \) above level-\( i \) (towards root):
   Denote by \( c_l \) a bit in the current search branch above level-\( i \). If \( c_l = c_l^{MAP} \), the counter hypothesis \( \lambda^{MAP}(c_l) \) may be updated by searching the subtree. We denote the set of counter-hypotheses in this category as \( A_1 \).

2. The counter-hypotheses \( \{\lambda^{MAP}(c_l)\} \) below level-\( i \): As the node is not expanded yet, all counter-hypotheses \( \{\lambda^{MAP}(c_l)\} \) below-\( i \) may be updated in searching the subtree. We denote the set of counter-hypotheses in this category as \( A_2 \).

The partial distance \( d_i \) is compared with the counter-hypotheses in both \( A_1 \) and \( A_2 \). We denote the the union of the two sets as \( A = A_1 \cup A_2 \). The node’s subtree will be pruned if

\[ d_i > \max_{\lambda \in A} \lambda. \]  

(4.28)

According to the discussion above, the tree-pruning criterion ensures that we cannot provide update to any counter-hypotheses. The set \( A \) does not contain the current MAP hypothesis \( \lambda^{MAP} \), but any counter hypothesis will be larger than \( \lambda^{MAP} \). Therefore, (4.28) is a sufficient condition for tree-pruning operation.

The decoding complexity of max-log-MAP algorithm can be greatly decreased by tree-pruning. However, the the complexity of this version of sphere decoding algorithm is not fixed. We will discuss the factors that affect decoding complexity latter in Section 4.3. Some examples of complexity results are presented in Section 4.5.
The list administration

When a valid leaf is reached in the tree traversal, the detector will update the current MAP hypothesis $\lambda^{MAP}$ and counter-hypotheses $\{\lambda^{MAP}(c_l)\}, \forall l \in [1, N_tM]$.

1. $d(c) < \lambda^{MAP}$: A new MAP hypothesis is found. If the bit $c_l$ is changing in the MAP hypothesis, the counter-hypothesis for $c_l$ should also be updated.
   - Counter-hypothesis update: $\lambda^{MAP} \rightarrow \lambda^{MAP}(c_l)$ for changing $c_l$ in MAP hypothesis $c^{MAP}$
   - MAP hypothesis $c^{MAP}$ update: $d(c) \rightarrow \lambda^{MAP}$

2. $d(c) \geq \lambda^{MAP}$: In this case, only counter-hypotheses might be updated. For the the coded bit $c_l$ in $c$, we check whether $c_l$ is the same as the bit $c_l^{MAP}$ in MAP hypothesis. If they are different, we will update counter-hypothesis for $c_l$, when $d(c) < \lambda^{MAP}(c_l)$.
   - Counter-hypothesis update: $d(c) \rightarrow \lambda^{MAP}(c_l)$
   - Two conditions are required: $c_l \neq c_l^{MAP}$ and $d(c) < \lambda^{MAP}(c_l)$.

The complexity

The visited nodes are denoted by red dots in Fig. 4.5. The complexity of the sphere decoding is a function of both the number of the visited nodes and the operations associated with each node. However, we only consider the expected number of nodes as a measure of decoding complexity in this research. The total complexity can be reduced by the tightened tree-pruning criterion, which still generates max-log outputs [52]. The decoding complexity is significantly reduced with a high SNR. Because many candidates become less probable with at high SNR and the nodes are pruned at early stage of the tree traversal. We will show simulation results in the next section.

4.4 K-Best Sphere Decoding with Fixed Complexity

In Section 4.3, we described a soft-input soft-output sphere detector, which achieves max-log MAP performance. The disadvantage of the detection method is that the complexity of the processing is not fixed. For practical application, the hardware design needs to satisfy the worst-case situation. Therefore, it is preferred to have a sphere decoder with fixed
complexity for the real-world applications. In this Section, we briefly introduce K-best sphere decoding algorithm [62].

We still assume $N_r \leq N_t$ as in previous section. The channel matrix is QR-decomposed as $H = QR$, where $R$ is an upper triangular matrix. The equivalent input-output relation of the MIMO channel can be expressed as

$$\tilde{y} = Q^H y = R x + \tilde{n},$$  \hspace{1cm} (4.29)

where $\tilde{n} = Q^H n$. The goal of MIMO detection is search the closest constellation point array

$$\hat{x} = \arg \min_{x \in \Omega_{N_t}} ||\tilde{y}||^2.$$  \hspace{1cm} (4.30)

With the help of QR decomposition, the above MIMO detection problem can be solved by a tree search. The tree-search is performed depth-first by max-log MAP sphere decoder presented in Section 4.3. The breadth-first tree-search is used by K-best sphere decoder, where the search is single direction from root to leaf. As the recursion proceeds in forward direction only, the algorithm is suitable for an implementation with parallel structure. To reduce the complexity of the detection, at most $K$ survivor nodes are kept for further exploration during the tree search process. The basic procedure of the K-best sphere is described as below.

1. At the root, we initialize level = $N_t + 1$ and set $\text{bestdist} = \{0\}$.

2. Proceed towards leaves and we have level = level-1. Expand each survivor path to $M$ nodes, where $M$ is the size of the constellation. Update $\text{partialdist}$ accordingly.

3. Select $K$ shortest distances from the set $\text{partialdist}$ as new $\text{bestdist}$.

4. Keep the history of $K$ survivors and discard the other nodes at current level.

5. If the recursion reaches level = 1, we stop the algorithm. Otherwise, go back to step 2.

With above algorithm, we will exam at most $K \times M$ leaves at the end of the tree search. Let us denote the set of leaves (candidates) found in above method as $L$. The $a \ posteriori$
log-likelihood ratio (LLR) value for the \(l\)th coded bit \(c_l\) in \(c\) is calculated over the \(L\)

\[
\lambda(c_l) = \log \frac{\Pr[c_l = 1|y_1, \ldots, y_N]}{\Pr[c_l = 0|y_1, \ldots, y_N]}
= \log \sum_{c \in L_{c_l=1}} \Pr[y_1, \ldots, y_N, c] \quad \sum_{c \in L_{c_l=0}} \Pr[y_1, \ldots, y_N, c]
= \log \frac{\sum_{c \in L_{c_l=1}} \Pr[y_1, \ldots, y_N|c] \Pr(c)}{\sum_{c \in L_{c_l=0}} \Pr[y_1, \ldots, y_N|c] \Pr(c)}
\tag{4.31}
\]

In the iterative detection, the candidate list \(L\) will be calculated only once at the first iteration. The \textit{a priori} information \(\Pr(c)\) and the \textit{a posteriori} information \(\lambda(c_l)\) will be updated at each iteration. The details of the K-best algorithm can be found in [62].

### 4.5 Simulation results

In this section, simulation results for turbo coded BICM schemes in block-fading channels are presented. We first compare the performance and complexity between sphere decoding method and array processing method with a large number of antennas. Then, we present and compare the performance between serial and parallel turbo coded BICM schemes in block-fading channels.

#### 4.5.1 Performance of sphere decoder

We consider a system with large number of transmit and receive antennas. Sphere decoder described in Section 4.3 is employed at the signal processing unit to generate estimates for coded bits, which are used as inputs for the turbo decoder. At the transmitter side, a parallel concatenated turbo coded BICM scheme is used. The configuration of the parallel concatenated turbo coded BICM follows exactly the structure of those in [50], where two systematic recursive \((1, 5/7)\) convolutional codes are employed and the parity bits are punctured regularly to get a coding rate of \(r_c = 1/2\). A random interleaver is employed between the parallel turbo encoder and modulator. The codeword length is 1024 bits for both schemes.

We compare the FER performance of the sphere decoding and array processing method
(group interference suppression) described in Section 4.2 with 10 rounds of the iterative detection in Fig. 4.6. Each codeword spans $F = 2$ independent fading blocks. The turbo code internal decoding iteration is set to be 1 to reduce the total computation at the receiver.

In group interference suppression, we partition 4 transmit antennas into two group of two antennas and partition 8 transmit antennas into three group of 3, 3 and 2 antennas. At a FER of $10^{-3}$, we can see that sphere decoder provides a gain of 5dB for the $4 \times 4$ MIMO case and a gain of 6dB for the $8 \times 8$ MIMO case. The performance loss with array processing method is caused by reduced diversity with linear processing to null the interference signals from other groups. The sphere decoding achieves max-log-map performance. However, the sphere decoding performance curve is not parallel to outage curve, which means it fails to achieve the full diversity order. The parallel concatenated turbo coded BICM scheme proposed in [50] does not guarantee achievable diversity order. This observation inspires us to explore the serial concatenated codes for MIMO block-fading channels, which is actually designed for SISO block-fading channels.

Figure 4.6: FER performance comparison between sphere decoding and array processing over a block-fading channel with $F = 2$. 
It is interesting to perform a complexity comparison between sphere decoding and group inference suppression. The complexity of sphere decoding is data dependent and the complexity of group interference suppression is fixed. We compute the complexity of sphere decoding by finding the average number of nodes visited during the tree transversal in Fig. 4.7.

![Average number of nodes visited during the sphere decoding search in a 4 x 4 MIMO block-fading channel with F = 2.](image)

The complexity of sphere decoding decreases when the number of iterations or SNR increases. For 4 x 4 MIMO channel and QPSK modulation, there are a total of 340 nodes in the sphere decoding tree. At a high SNR, we need to search about 25 nodes at the 10th iteration, which are 7% of the total nodes. With a low SNR value, we need to process 65 nodes at the 10th iteration, which are 19% of the total nodes. For group interference suppression, we divide 4 transmitter antennas into two groups, with two antennas in each group. For QPSK modulation, we need to process $4^2 = 16$ sequences in each group. The complexity of
group interference suppression is roughly \((16 \times 2)/4^4 = 12.5\%\) of MAP detection. Therefore, the complexity of sphere decoding and array processing are comparable. However, the decoding complexity of sphere decoding algorithm used here is not fixed. Some versions of fixed-complexity sphere decoding (FSD) algorithm can be found in [63,64]. These methods search only a fixed number of candidates, which achieve the performance inferior to ML decoding.

![Figure 4.8: FER performance comparison between sphere decoding (both max-log and K-best) and array processing over a 4 x 4 MIMO block-fading channel with \(F = 2\).](image)

Figure 4.8: FER performance comparison between sphere decoding (both max-log and K-best) and array processing over a 4 x 4 MIMO block-fading channel with \(F = 2\).

We compare the FER performance among max-log sphere decoder, K-best sphere detector and array processing detector in Fig. 4.8 for a 4 x 4 MIMO block-fading channel with \(F = 2\) blocks. For the max-log sphere detector and array processing detector, the parameter are set as those in previous simulation. For K-best sphere detector, we set \(K = 20\). The sphere radius is \(\infty\) and the size of candidate list is \(20 \times 4 = 80\) with K-best sphere decoder. The complexity of K-best algorithm is about \(80/4^4 = 31.25\%\) of a MAP detector. At a FER of \(10^{-3}\), we can see that max-log sphere decoder provides a gain of 2 dB over K-best sphere decoder. In the meanwhile, the average complexity of max-log sphere decoder is lower than
K-best sphere decoder in this situation.

![FER performance comparison between sphere decoding (both max-log and K-best) and array processing over a 8 × 8 MIMO block-fading channel with F = 2.](image)

The FER performance comparison of max-log sphere decoder, K-best sphere detector and array processing detector is shown in Fig. 4.9, on a 8 × 8 MIMO block-fading channel with F = 2 blocks. At a FER of $10^{-3}$, max-log sphere decoder offers a gain of 3 dB over K-best sphere decoder. We still set sphere radius to be $\infty$ for K-best sphere detector. As we have more transmit antennas in the system, the value of $K$ increases to 64. The size of candidate list is $64 \times 4 = 256$ and the detection complexity of K-best algorithm is about $256/4^6 = 0.39\%$ of a MAP detector.

### 4.5.2 Comparison between serial and parallel turbo codes

To make a fair comparison, the serial concatenated turbo codes employ a non-systematic, non-recursive, 3-memory convolutional code as the outer code and a single memory accumulator as the inner code. Therefore, there are 4 memories in both parallel turbo code and serial turbo code encoders. The codeword length is 1024 and the number of iterative
detection is 10 for both schemes. QPSK modulation is used in our simulations.

We compare frame error rate (FER) performance of parallel and serial concatenated turbo coded BICM in Fig. 4.10. The simulations are carried out in block-fading channel with $F = 2$ and $F = 4$. We use a $(11, 15)$ convolutional code as the outer code for the $N_t = 2, N_r = 2, F = 2$ case and a $(13, 11)$ convolutional code for the $N_t = 2, N_r = 1, F = 4$ case. Both outer codes for the serial concatenated turbo codes are MDS codes for SISO block fading channel [10]. Our results show that parallel turbo coded BICM outperforms serial turbo coded BICM in MIMO block-fading channels. The total available diversity order is 8 for both channel conditions. It can be seen that both schemes fail to achieve full diversity order, because there are interference signals present in the MIMO transmission. We apply use spatial multiplexing and send different data streams on different transmit antennas to achieve the high throughput at the cost of reduced spatial diversity order.

To further illustrate the effect of interference, we compare parallel and serial concatenated turbo coded BICM in a simple SISO case with $N_t = 1, N_r = 1, F = 2$ and a simple
MIMO case with $N_t = 2, N_r = 1, F = 1$. Two cases have the same outage limit. A $(13, 15)$ convolutional code is employed as the outer code for the serial concatenated turbo code. As shown in Fig. 4.11, the serial concatenated turbo coded BICM achieves the optimal diversity order of 2 in the SISO case, but parallel concatenated turbo coded BICM fails to achieve the optimal diversity order. However, both schemes fail to achieve the optimal diversity order with $N_t = 2$, where spatial interference is present. The reason is that the serial concatenated turbo code proposed in [2] is for SISO channel. A direct application of this scheme to MIMO case will not guarantee the achievable diversity order.

![Figure 4.11: FER performance for parallel and serial concatenated turbo coded BICM scheme in both SISO and MIMO block-fading channels.](image)

In the next chapter, we will explore a MIMO transmission based on modulation diversity techniques for block-fading channels. The method enables spatial multiplexing with losing any spatial diversity. The cost of the modulation diversity technique is extra system structure at the transmitter and increased possessing complexity at the receiver.
4.6 Conclusions

In this chapter, we studied turbo coded BICM schemes for MIMO block-fading channels. The research can be viewed as a supplementary work to the parallel concatenated turbo coded BICM scheme in [50]. A serial concatenated turbo coded BICM scheme is proposed. Two schemes are compared in terms of performance in MIMO block-fading channels. In addition, a sphere decoder is used as a practical low-complexity signal processing solution, when the number of transmit antennas or the size of modulation constellation is large. The sphere decoder offers better performance compared with the group suppression technique. The decoding complexity of sphere decoding is comparable to that of group suppression technique.

Our study also shows that we cannot apply schemes derived for SISO block-fading channels to MIMO cases directly to achieve full space diversity. The reason is that we need to deal with interference from different transmit antennas in the MIMO case. In the next chapter, will show a hybrid scheme based on both error correcting codes and space-time codes to achieve the full diversity order of the channel.
Chapter 5

Coded MIMO Systems with Modulation Diversity for Block-Fading Channels

We propose a new coded multiple-input multiple-output (MIMO) scheme based on modulation diversity technique for block-fading channels. The proposed scheme can achieve both full spatial multiplexing and full diversity of $N_t N_r F$, i.e., the product of the number of transmit antennas $N_t$, receive antennas $N_r$, and fading blocks $F$. The design concatenates a channel code with a space-time code. The channel code provides time diversity and coding gain, whereas the space-time code is employed to achieve space diversity. Since the modulation diversity technique enables full space diversity and full spatial multiplexing at the cost of increased processing complexity at the receiver, sphere decoding is employed to reduce the detection complexity. Simulation results are provided to demonstrate the effectiveness of the proposed schemes in MIMO block-fading channels.

It has been shown that for a coded system over a block-fading channel the maximum achievable diversity order is upper bounded by the Singleton bound [10]. If a MIMO system relies on channel coding only to obtain the full diversity order of $N_t N_r F$, the maximum code rate is $r_c = 1/N_t F$. Such a low rate is unacceptable in many applications. Space-time coding techniques can be employed to exploit partial or full space diversity of $N_t N_r$. Space-time block codes based on orthogonal designs are able to achieve full space diversity with low-complexity receiver at the expense of low rate (no more than 1 symbol per channel...
use) [24]. Layered space-time codes increase transmission rate with reduced diversity [28]. We are interested in space-time codes, which can achieve full space diversity of $N_t N_r$ with full spatial multiplexing ($N_t$ symbols per channel use), such as the threaded algebraic space-time codes discussed in [65] and the recently proposed MIMO systems with real value constellation rotations [66]. These full-diversity and full-rate space-time code designs are based on multidimensional constellation expansion (rotation), i.e., modulation diversity techniques proposed in [67], which spreads constellation points over the time and space domains. At the receiver, a joint maximum-likelihood (ML) detector over time and space domains is required to achieve modulation diversity.

We propose a coded MIMO scheme, which can achieve the full diversity order of $N_t N_r F$ and transmit at full rate over block-fading channels. The design concatenates a channel code with a space-time code. The channel code provides time diversity and coding gain, while the space-time code allows for space diversity. At the receiver, a near optimal iterative belief propagation (BP) detection algorithm is used and the extrinsic information is exchanged between the space-time signal detector and the channel decoder. Furthermore, an efficient soft-input soft-output sphere decoder is employed as signal detector to reduce the detection complexity [52].

5.1 Space-Time Coding Based on Modulation Diversity

Modulation diversity technique was originally proposed for fully-interleaved SISO fading channels [67]. The basic idea is to spread constellation points over the time domain by rotation operation, such that each original constellation point experiences independent fading coefficients. At the receiver a ML detector is equipped to exploit the modulation diversity.

The modulation diversity technique can be employed for space diversity orders in MIMO systems. The diagonal algebraic space-time (DAST) block codes based on rotated constellations and Hadamard transform can achieve full space diversity order and a throughput of 1 symbol per channel use [68]. The threaded algebraic space-time (TAST) codes achieves both full diversity order and full rate ($N_t$ symbols per channel use) [65]. The TAST scheme obtains full rate by using the space-time threading architecture proposed in [69]. The full diversity order is achieved by transmitting a scaled DAST code in each thread. The TAST coding scheme is also refereed as universal space-time coding (USTC). Both DAST scheme and TAST encode coded bits onto complex-valued QAM symbols directly. We follow a
transmission design based on real-valued rotation threads for MIMO system [66], which adopt a design similar to TAST and also achieves both full diversity and full rate.

Modulation diversity techniques and space-time codes based on real-valued rotations are introduced briefly here. Let \( \mathbf{z} = [z_1, z_2, \cdots, z_{N_t}]^T \) denotes an \( N_t \)-dimensional vector with real elements, where \( z_i \in \{ \pm 1, \pm 3, \cdots, \pm (2^M - 1) \} \) is a point in a PAM constellation of size \( 2^M \). The mapping scheme transmits \( M \) bits per two dimensions, which is equivalent to a QAM constellation of \( 2^M \) points. The rotated vector \( \mathbf{u} \) is obtained by multiplying vector \( \mathbf{z} \) by a real orthogonal rotation matrix \( \mathbf{R}_{N_t} \) of size \( N_t \) as

\[
\mathbf{u} = \mathbf{R}_{N_t} \mathbf{z}. \tag{5.1}
\]

The rotation process spreads PAM points in \( \mathbf{z} \) over all elements in \( \mathbf{u} \). So that each point will go through different fading effects. Let us denote the real PAM constellation set by \( \mathcal{Z}_{N_t} \) and the rotated constellation set by \( \mathcal{U}_{N_t} \). We have

\[
\mathcal{U}_{N_t} = \{ \mathbf{u} | \mathbf{u} = \mathbf{R}_{N_t} \mathbf{z}, \mathbf{z} \in \mathcal{Z}_{N_t} \}. \tag{5.2}
\]

Let \( \mathbf{u}_a \) and \( \mathbf{u}_b \) be two vectors from the rotated constellation set \( \mathcal{U}_{N_t} \). And \( u_{a,j} \) and \( u_{b,j} \) are \( j \)th component in vector \( \mathbf{u}_a \) and \( \mathbf{u}_a \), respectively. The following two code design criteria are employed to guarantee diversity order and good coding gain for the SISO fading channel case [67].

- **The modulation diversity order**: The minimum number of distinct components between any two rotated vectors in \( \mathcal{U}_{N_t} \).

- **The product distance**: 
  \[
  d_{p,\text{min}} = \min_{\mathbf{u}_a, \mathbf{u}_b \in \mathcal{U}_{N_t}} \prod_{j=1}^{N_t} |\Delta_{u_j}|, \tag{5.3}
  \]

  where \( \Delta_{u_j} = u_{a,j} - u_{b,j} \neq 0 \).

The optimized rotation matrices \( \mathbf{R}_{N_t} \), which enable the system to achieve full diversity and good minimum product distance can be found in [67]. For example, the optimal real rotation matrix with \( N_t = 2 \) is

\[
\mathbf{R}_2 = \begin{bmatrix} 0.5257 & -0.8507 \\ 0.8507 & 0.5257 \end{bmatrix}. \tag{5.4}
\]
For $N_t = 3$, the optimal rotation matrix is given by

$$R_3 = \begin{bmatrix} \alpha & \beta & \gamma \\ \beta & \gamma & \alpha \\ -\gamma & -\alpha & -\beta \end{bmatrix},$$

(5.5)

where

$$\alpha = \frac{1 + \omega}{1 + \omega + \omega^2}, \quad \beta = \omega \alpha \quad \text{and} \quad \gamma = -\frac{\omega}{1 + \omega} \alpha \quad \text{with} \quad \omega = 2 \cos(\frac{6\pi}{7}).$$

(5.6)

The space-time code design to minimize the maximum pairwise error probability (PEP) with ML detection follows the famous rank criterion and determinant criterion proposed for MIMO channel in [24].

- **Full rank criterion**: To ensure that the resulting space-time code achieves full space diversity with ML decoding, any difference codeword matrix $(x_a - x_b)$ has to be full rank $N_t$ (assuming $N_r \geq N_t$), where $x_a$ and $x_b$ are from space-time codeword set $X_{N_t}$.

- **Determinant criterion**: Let $D \triangleq x_a - x_b$ and $\Gamma = DD^H$. The $\Lambda_j$ is the $j$th non-zero eigenvalue of $\Gamma$, with $k = 1, 2, 3, \ldots, N_t$. The minimum determinant (5.7) over any pair of codewords should be maximized.

$$d_{p,\text{min}} = \min_{x_a, x_b \in X_{N_t}} \prod_{j=1}^{N_t} \Lambda_{x_j},$$

(5.7)

For layered MIMO transmission, the data are de-multiplexed into streams at the transmitter side and each of them is referred to as a layer [28]. A thread is defined as a layer with full spatial and temporal span [70]. Each thread is active during all transmission intervals during the time, which requires $N_t$ to be no less than the total number of threads (with complex elements). The thread switches to a different transmission antenna at each transmission interval and each thread uses $N_t$ transmission antennas equally often. The advantage of space-time threading is that each thread will experience all available diversities in both time and space domain.

In [66], a threaded space-time code of dimension $N_t \times N_t$ is constructed by $2N_t$ threads as shown in (5.8). The resulting layered space-time code transmits $MN_t$ bits per channel use, which has the same spectral efficiency as full spatial multiplexing MIMO system with a $2^M$-QAM constellation. A scaling number is introduced for each thread to guarantee full
The thread $\mathbf{u}_i$ is constructed by

$$\mathbf{u}_i = \phi_i \mathbf{R}_{N_i} \mathbf{z}_i,$$

(5.9)

where $\phi_i$’s are referred as Diophantine numbers

$$\{\phi_1 = 1, \phi_2 = \phi^{1/N_i}, \ldots, \phi_{N_i} = \phi^{(N_i-1)/N_i}\},$$

and $\phi_i = \phi_{i-N_i}$, for $i \geq N_i + 1$. It has been shown that $\phi = e^{j\pi/6}$ for $N_t = 2$ and $\phi = e^{j\pi/12}$ for $N_t = 3$ yield optimal coding gain and ensure the full rank property [65]. An example of threaded layering scheme is illustrated in Fig.5.1. The numbers are used to denote threads and there are a total of 6 real threads in the example for $N_t = 3$.

```
1 + j4  6 + j3  5 + j2
2 + j5  1 + j4  6 + j3
3 + j6  2 + j5  1 + j4
```

Figure 5.1: The threaded layers in space-time code scheme with $N_t = 3$.

The modulation diversity achieves the full diversity order at the cost of increasing processing complexity at the receiver side, as it actually encapsulate information about all bits in the original PAM vector in every single rotated constellation point. In order to achieve maximum diversity order, ML detection at the receiver is required whose complexity increases exponentially with the constellation size index $M$ and square of the number of transmission antennas $N_t^2$. Therefore, decoding a system with more than 4 transmission
antennas using any mapping scheme would be impractical with original ML decoding algorithm. Low complexity ML detector can be implemented in the form of sphere decoder, which was discussed in Section 4.3.

5.2 System Model of Coded MIMO System Based on Modulation Diversity

The proposed coded MIMO system consists of two parts: an outer channel code and an inner space-time code. The two parts are serially concatenated by a bit-level blockwise interleaver, which scrambles the the coded bits for each fading block separately. Therefore, the distance property is reserved after blockwise interleaving. The outer channel code is selected to be a maximum distance separable (MDS) code for SISO block-fading channels, which provides time diversity. The inner space-time code is responsible for space diversity. A more general design is possible as discussed in [71]. A low rate channel code can be used to collect partial space diversity and alleviate the processing burden of space-time code detector. However, a more complicated design is needed to enable the outer channel code to exploit space diversity. We only consider the system splitting time and space diversity collection duties to two parts.

5.2.1 Transmitter Structure

The block diagram of the transmitter structure is shown in Fig. 5.2.

![Transmitter structure of the coded MIMO system.](image)

We choose serial concatenated turbo codes proposed in [2] as the channel code, which consists of an outer MSD non-systematic convolutional code and an inner repeat accumulate (RA) code. The output of the outer encoder is partitioned into $F$ blocks, one for each fading
block in the time domain. A blockwise interleaver is inserted between the channel encoder and space-time code encoder, so the blockwise Hamming distance of the outer channel code is retained after interleaving. We use the space-time code presented in Section 5.1 as the inner code to exploit space diversity. The space-time unit encodes $MN_t^2$ coded bits into a space-time code $x$ with dimensions $N_t \times N_t$ as shown in (5.8) at each operation.

### 5.2.2 Receiver with Iterative Detection

The receiver is composed of two parts as shown in Fig. 5.3; a signal processing unit in the front end followed by a channel code decoder. Both parts accept soft inputs and generate soft outputs. The extrinsic information is exchanged between two parts in an iterative manner.

As each space-time symbol $x$ consists of $N_t^2 M$ coded bits, the signal processing unit has to perform a joint ML detection on those bits. Let $c$ of length $N_t^2 M$ be the coded bits

$$c = [c_1, c_2, \cdots, c_M, \cdots, c_{N_t^2 M}].$$

(5.10)

Assuming a maximum a posteriori (MAP) detector is employed, the a posteriori log-likelihood ratio (LLR) value for the $l$th coded bit $c_l$ in $c$ is

$$\lambda(c_l) = \log \frac{\Pr[c_l = 1|y]}{\Pr[c_l = 0|y]} = \log \frac{\sum_{c_{c_l} = 1} \Pr[y|c] \Pr(c)}{\sum_{c_{c_l} = 0} \Pr[y|c] \Pr(c)}.$$  

(5.11)

Thanks to bit-level interleaver, we can assume that the a priori probabilities $\Pr(c_l)$ from
channel decoder are statistically independent. The extrinsic information $\xi(c_l)$ fed back to channel decoder is

$$\xi(c_l) = \log \frac{\sum_{c: c_l = 1} \left( \Pr[y|c] \prod_{m \neq l} \Pr(c_m) \right)}{\sum_{c: c_l = 0} \left( \Pr[y|c] \prod_{m \neq l} \Pr(c_m) \right)}.$$  \hspace{1cm} (5.12)

In (5.11) and (5.12), we have

$$\Pr[y|c] = K \exp \left( -\frac{\|y - Hx\|_F^2}{N_0} \right),$$ \hspace{1cm} (5.13)

where $x$ is the space-time code mapped from $c$, $K$ is a constant and $\| \cdot \|_F$ is Frobenius norm.

The soft-input soft-output channel code decoding algorithm can be found in [2], which also includes an iterative process of extrinsic information exchange between outer convolutional code decoder and inner RA code decoder. To reduce the complexity of total decoding complexity, the inner iteration number is set to be 1.

The detection complexity of a single coded bit grows exponentially with $N_t^2 M$ in (5.11). The log-sum computation in (5.11) can be simplified by max-log approximation without much performance loss

$$\log \sum_{c: c_l = 1} \Pr[y|c] \Pr(c) \approx \max_{c: c_l = 1} \log \Pr[y|c] \Pr(c).$$ \hspace{1cm} (5.14)

However, the search candidates complexity of (5.14) still grows exponentially with $N_t^2 M$. Compared with a V-BLAST system in [28] with the same settings, the number of candidate space-time codes $x$ increases from $2^{N_t M}$ to $2^{N_t^2 M}$. The increased detection complexity with diversity modulation is the cost of getting diversity gain. A soft-input soft-output sphere decoder [51,52] can be employed to reduce processing complexity without performance loss.

In order to apply the existing sphere decoding algorithms to the proposed MIMO system, some transforms are needed. We start with the case with two transmit antennas. From
(5.8), the space-time code can be expressed as
\[
x = [x_1, x_2] = \begin{bmatrix} u_{1,1} + ju_{3,1} & u_{4,2} + ju_{2,2} \\ u_{2,1} + ju_{4,1} & u_{1,2} + ju_{3,2} \end{bmatrix}.
\]

Similarly, we represent channel matrix, noise matrix and received signal matrix in (1.26) in using column vectors as \(H = \begin{bmatrix} h_1, h_2 \end{bmatrix}, \ n = \begin{bmatrix} n_1, n_2 \end{bmatrix}\) and \(y = \begin{bmatrix} y_1, y_2 \end{bmatrix}\). The input-output relation of the channel can be rewritten as
\[
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}.
\]

Considering the structure of space-time code given in (5.15), we can further transform the term \(\begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\) in (5.15) to
\[
\begin{bmatrix} h_1 & 0 & h_2 & 0 & h_3 & 0 & h_4 & 0 \\ 0 & h_2 & 0 & h_3 & 0 & h_4 & 0 & h_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix},
\]
where \(h_3 = jh_1\) and \(h_4 = jh_2\). The real thread \(u_i\) is obtained by \(u_i = \phi_i R_2 z_i\) in (5.9). Therefore, we can take PAM vector \(\{z_i\}\) as inputs and rewrite (5.15) as
\[
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = H' \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix},
\]
where \(H'\) is the new equivalent channel matrix of dimensions \(2N_r \times 8\)
\[
H' = \begin{bmatrix} \phi_1 R_2 & 0 & 0 & 0 \\ \phi_2 R_2 & 0 & 0 & 0 \\ 0 & \phi_1 R_2 & 0 & 0 \\ 0 & 0 & \phi_2 R_2 & 0 \end{bmatrix}.
\]
The above transformations can be easily generalized to system with any number of transmit antennas \( N_t \). The general equivalent channel input-output relation can be expressed as

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_{N_t}
\end{bmatrix}
= \mathbf{H}'
\begin{bmatrix}
    z_1 \\
    z_2 \\
    \vdots \\
    z_{2N_t}
\end{bmatrix}
+ \begin{bmatrix}
    n_1 \\
    n_2 \\
    \vdots \\
    n_{N_t}
\end{bmatrix},
\]  

(5.17)

where \( \mathbf{H}' \) is the general equivalent channel matrix with dimensions \( N_t N_r \times 2N_t^2 \).

Most existing sphere decoding algorithms like [51, 52] can be directly applied to (5.17). For MIMO block-fading channels, the equivalent channel matrix \( \mathbf{H}' \) only needs to be computed once per fading block.

### 5.3 Simulation Results

In this section, we present simulation results of the proposed coded MIMO systems over block-fading channels.

Let us first consider a \( 2 \times 1 \) MIMO quasi-static Rayleigh fading \( (F = 1) \) channel. We intend to show that full space diversity can be achieved by the proposed coded scheme with this channel. The size of PAM constellation is two, i.e., BPSK modulation. The system throughput is 4 bits per channel use. The outer channel code is a blockwise concatenated convolutional code (BCCC) proposed in [2], which consists of a 1/2-rate (5, 7) non-systematic non-recursive convolutional code (MDS code for \( F = 2 \)) and a memory-1 RA code. The codeword length is 1024 for all of our simulation results in this chapter. The BCCC code is designed to achieve maximum diversity over SISO block-fading channels. After channel coding, the coded bits are interleaved and then encoded by the space-time code using Equation (5.8). For comparison, we consider a MIMO system using a BCCC channel code to exploit space diversity, where the coded bits of the BCCC encoder are first partitioned blockwise, then QPSK modulated and sent on two transmit antennas. We also plot the performance of the BCCC code over a SISO block-fading channel with \( F = 2 \), which has the same intrinsic channel diversity as a \( 2 \times 1 \) quasi-static fading MIMO channel. We compare the FER performance of the proposed MIMO space-time scheme based on rotated constellation (Rotated ST) with BCCC code in Fig. 5.4. The simulation results
Figure 5.4: FER performance of the coded MIMO system over a $2 \times 1$ MIMO quasi-static fading channel.
are obtained with a MAP detector as show in (5.11) and 10 rounds of iterative detection-decoding. We can see that the proposed coded MIMO scheme achieves full diversity of 2 with the help of space-time code based on modulation diversity technique. The BCCC code can achieve full diversity in the SISO case, but fails to achieve full diversity in the MIMO case. The reason is that BCCC can be used for space diversity based on the assumption that interferences can be fully removed at the detector, which is an ideal situation. To achieve space diversity with ML detection, the rank criterion must be satisfied [24]. We also note that the system throughput of the SISO case is 2 bits per channel use, which is half of the MIMO case. Spacial multiplexing gain and space diversity gain are two attractive achievements with MIMO systems.

![FER performance of the coded MIMO system over MIMO block-fading channels with F = 2 fading blocks.](image)

Figure 5.5: FER performance of the coded MIMO system over MIMO block-fading channels with $F = 2$ fading blocks.

In the next step, we test the proposed system in MIMO block-fading channels with $F = 2$ fading blocks. The channel code settings are the same as those in the previous case. The space-time code needs to be adjusted according to the number of transmit antennas. We simulate 3 MIMO settings: $2 \times 1$, $2 \times 2$ and $3 \times 3$. In the first MIMO setting, a MAP
signal detector is used. In the second and third settings, a soft-input soft-output sphere detector is used. A system throughput of 4 bits per channel use is achieved for $N_t = 2$ and 6 bits per channel use for $N_t = 3$. The number of iterations is 15 for all cases. Both simulated FER performance and their corresponding outage limits with Gaussian inputs (by curves using the same line styles without marker) are drawn in Fig. 5.5. We can see that all three FER curves are parallel to their corresponding outage limits, which means all of them achieve full time and space diversity.

![Figure 5.6: Average number of nodes visited during the sphere decoding search in a $2 \times 2$ MIMO block-fading channel with $F = 2$ fading blocks.](image)

For the simulation with sphere decoder, we have adopted the soft-input soft-output single tree search algorithm proposed in [52], which is capable to incorporate a priori information and generate max-log estimation. The signal detection can be transformed into a tree search problem and a tree pruning criterion is provided to avoid searching the entire tree. The average number of visited nodes during the detection process can be used as a measure of decoding complexity. Assuming $N_r \geq N_t$, there are a total of $\sum_{i=1}^{2N_t^2} 2^i = 510$ nodes
with $N_t = 2$ and PAM constellation of size two. We show average number of nodes visited for $2 \times 2$ MIMO block-fading channel with $F = 2$ in Fig. 5.6. We can see that the signal processing cost has been greatly reduced by using sphere decoder. The detection complexity of sphere decoder is a function of the SNR and the number of iterations. The detection complexity decreases, as the SNR or the number of iteration increases. At a SNR of 8dB and 15 iterations, an average of 21 nodes need to be processed, which are just 4.1% of the total nodes. We also note that at low SNR, the detection complexity is almost unchanged with different number of iterations, which means the feedbacks from channel decoder are unreliable and do not provide much help in signal separation.

![Figure 5.7: Comparison of FER performance between the coded MIMO system and the serially Turbo-coded BICM scheme with $N_t = 2$, $N_r = 2$, over MIMO block-fading channels of $F = 2$ fading blocks.](image)

We already show the diversity advantage of proposed coded space-time scheme based on modulation diversity over the BCCC scheme in Fig.5.4, where the total available diversity order is 2. Now we consider a $2 \times 2$ MIMO system over a block-fading channel of $F = 2$ blocks. In this situation, the total available diversity order is 8. We compare the proposed
scheme with the corresponding scheme without modulation diversity. In both schemes, we use BCCC codes, where a 1/2-rate (5, 7) non-systematic non-recursive convolutional code is used as the outer code and a memory-1 RA code is used as inner code. As we discussed in Chapter 4, the serially concatenated BCCC is originally designed for SISO block-fading channels, which is not guaranteed to achieve full diversity order in MIMO channels. From Fig. 5.7, we observe the scheme with modulation diversity has only marginal advantage in performance over the scheme without modulation diversity when total available diversity order is 8. The reason is that when there are more available diversity (≥8), the system without modulation diversity has high probability to obtain a high diversity order and its performance is close to the system with full diversity order. The systems designed for block-fading channels show superior performance by achieving full diversity order for the situation, where the total available diversity is limited (≤4).

5.4 Conclusions

In this chapter, we proposed a coded MIMO scheme for block-fading channels. The proposed scheme consists of a channel code and a space-time code. The channel code is selected to be a MDS code designed for SISO block-fading channels and the function of the channel is to provide time diversity. The space-time code is designed based on modulation diversity technique, which allows full spatial multiplexing MIMO transmission and achieves full space diversity at the same time. The drawback of modulation diversity is the increased complexity at the detector. We also show how to apply a sphere decoder to the proposed system to reduce the signal processing complexity. Simulation results have verified the effectiveness of our designs over MIMO block-fading channels. Here, we use channel coding to achieve time diversity and space-time coding for space diversity. We would like to mention that a general design is also possible, in which the processing complexity can be traded for code rate.
Chapter 6

An Efficient Decoding Algorithm for Concatenated RS-Convolutional Codes

6.1 Introduction

The code design criterion for error correcting codes over block-fading channels is to find block-wise maximum distance separable (MDS) codes with large coding gain. The Reed-Solomon (RS) codes are an important class of MDS codes, which were invented by Reed and Solomon in 1960 [72]. The RS codes cannot be directly applied to block-fading channels, because we consider fading channel with only a few fading blocks \( F \leq 8 \) and the codeword length of RS codes is usually longer than that. For block-fading channels with number of fading blocks \( F = 4, 6 \) and 8, lengthening (adding information symbols) and shortening (removing information symbols) of RS codes are proposed in [10]. However, in this chapter we focus on standard RS codes design and their decoding. The results can be applied to lengthened and shortened RS codes for block-fading channels.

Concatenated schemes using Reed-Solomon codes as the outer code and convolutional codes as the inner code have been widely used for deep space communications as well as digital video broadcasting systems [73]. Convolutional codes can sweep channel noise, while RS codes can combat burst errors which are caused by the decoding process of convolutional codes when a Viterbi decoder is employed. Using this method, the concatenated
coding schemes can achieve high coding gains.

The concatenated Reed-solomon/Convolutional (RS/Conv) coding scheme has been around for a long time, but using iterative decoding schemes based on turbo principle for decoding these codes is relatively new. The main reason is that the soft input soft output (SISO) decoding algorithms for RS code are usually complicated, especially when codeword length is long. In 1994, Pyndiah et al. developed a simple and effective SISO decoder for RS codes based on the Chase algorithm. Pyndiah’s algorithm has been used for turbo decoding of both product [74] and serial concatenated codes [75].

In this chapter, we improve the decoding algorithm for the concatenated RS/Conv scheme investigated in [75] by adopting Kaneko’s algorithm for RS decoding [76]. This idea is inspired by [77], in which the authors discussed how to reduce decoding complexity of turbo product codes. The proposed algorithm reduces the computational complexity by employing the stopping criteria during the Chase decoding process. While conventional Chase algorithm generates constant number of codeword candidates, we expect to examine fewer number of codeword candidates using Kaneko’s stopping criteria. The major advantage of the proposed algorithm is lower computational complexity without sacrificing performance loss.

6.2 The Reed-Solomon Codes

The RS code is probably the most widely used traditional error control scheme in the coding society. After their invention in 1960s, the RS codes were employed in numerous applications in digital media and systems due to their excellent performance in combating random and burst symbol errors. In this section, we provide a brief introduction to RS codes. More detailed description of RS codes can be found in [38, 78].

The RS codes have been employed by various digital storage schemes for error protection. They are adopted in compact disc (CD) standard in the form of cross-interleaved Reed-Solomon (CIRS) coding, which is used to recover the information loss caused by scratches on the disc surface [79]. The RS codes are also found with various machine-readable barcode systems including popular quick response (QR) code [80]. With the error correction provide by RS codes, the machine can still decode a barcode with damage up to certain level. Other than consumer electronics, the RS codes are used in communication system as well, including digital video broadcasting and deep space communications.
Let's consider an \((n, k, d)\) RS code constructed over Galois field \(GF(q)\), where \(n\) is the codeword symbol length, \(k\) is the information symbol length and \(d\) is the minimum symbol Hamming distance. There are a total of \(q = 2^m\) elements in Galois field \(GF(2^m)\) and each element can be represented using \(m\) bits.

The standard RS code has a codeword length of \(n = 2^m - 1\). The minimum Hamming distance is \(d = n - k + 1\), as RS codes are MDS codes. The code can detect up to \(d - 1 = n - k\) erroneous symbols and correct up to \(t = \lfloor (n - k)/2 \rfloor\) erroneous symbols. By design, one can select information length to be \(k = 2^m - 1 - 2t\) and the minimum distance of the code is \(d = 2t + 1\). The code can correct any single burst error of length \(m(t - 1) + 1\) bits.

### 6.2.1 The RS Code Encoding

RS codes are cyclic codes, which can be generated using a generator polynomial of degree \((n - k)\) over \(GF(2^m)\).

\[
G(x) = x^{n-k} + g_{n-k-1}x^{n-k-1} + \cdots + g_1x + g_0, \quad (6.1)
\]

where \(g_i, \ i \in [0, n - k - 1]\) is an element from \(GF(2^m)\). The encoding of RS codes can be done in either a systematic way or a non-systematic way.

- For non-systematic encoding, we simply multiply the information symbol polynomial \(U(x)\) by the generator polynomial \(G(x)\)

\[
C(x) = U(x)G(x), \quad (6.2)
\]

where \(U(x) = u_k x^{k-1} + \cdots u_1 x + u_0\), with \(g_i \in GF(2^m), \ 0 \leq i < k - 1\).

- For systematic encoding, the codeword is obtained by

\[
C_{\text{sys}}(x) = x^{n-k}U(x) - [x^{n-k}U(x) \mod G(x)], \quad (6.3)
\]

It is easy to verify \(C_{\text{sys}}(x)\) is a valid codeword, as \(G(x)\) divides \(C_{\text{sys}}(x)\). The first \(k\) symbols in \(C_{\text{sys}}(x)\) are raw information symbols and the remaining \(n - k\) symbols are parity-check symbols.

The generator polynomials of RS codes are the minimal polynomials of \(\alpha^i, \ b \leq i \leq b+2t-1\), where \(\alpha\) is the primitive element of Galois field \(GF(q)\) and \(b\) is an integer; usually \(b = 0\).
or \( b = 1 \). The primitive element \( \alpha \) of \( GF(q) \) is a generator of the field, where \( q \) is the smallest value such that \( \alpha^{q-1} = 1 \). The minimal polynomial of \( \beta \in GF(q) \) is defined as the polynomial of the lowest degree over \( GF(q) \) that has \( \beta \) as its root. Therefore, the generator of RS codes can be expressed as

\[
G(x) = \prod_{i=b}^{b+2t-1} (x - \alpha^i). \tag{6.4}
\]

Thanks to the above special structure, there exist computationally efficient decoding algorithms for RS codes.

### 6.2.2 Decoding of RS Codes

RS codes are non-binary Bose-Chaudhuri-Hocquenghem (BCH) codes and therefore any decoding algorithm for BCH codes can be applied to decode RS codes [38, 78]. Here we briefly describe the well-known Berlekamp-Massey algorithm, which is used to find the error locations [81]. For non-binary code, the error value can be any non-zero element in \( GF(q) \).

Compared with binary BCH codes, the additional stage required for decoding RS codes is to find the error magnitudes, which can be done by using Forney algorithm. This step is not required by binary BCH codes, because we can simply flip the bit at each error location.

Let us denote the received vector using \( R(x) = C(x) + E(x) \), where \( C(x) \) is the transmitted codeword and \( E(x) \) is the error polynomial. The error polynomial \( E(x) \) can be expressed as

\[
E(x) = e_{j_1}x^{j_1} + e_{j_2}x^{j_2} + \cdots + e_{j_v}x^{j_v}, \tag{6.5}
\]

where \( e_{j_i} \) is error magnitude and \( j_i \) is the error location for \( 1 < i \leq v \). By design, the generator polynomial has \( \alpha^i, \ b \leq i \leq b+2t-1 \), as its roots. Therefore, we have \( C(\alpha^i) = 0 \) for \( b \leq i \leq b+2t-1 \). The syndromes are calculated by evaluating \( R(x) \) at \( x = \alpha^i \) with \( b \leq i \leq b+2t-1 \). The resulting syndrome polynomials are expressed as

\[
\begin{align*}
S_1 &= R(\alpha^b) = e_{j_1}\alpha^{bj_1} + e_{j_2}\alpha^{bj_2} + \cdots + e_{j_v}\alpha^{bj_v} \\
S_2 &= R(\alpha^{(b+1)}) = e_{j_1}\alpha^{(b+1)j_1} + e_{j_2}\alpha^{(b+1)j_2} + \cdots + e_{j_v}\alpha^{(b+1)j_v} \\
&\vdots \\
S_{2t} &= R(\alpha^{(b+2t-1)}) = e_{j_1}\alpha^{(b+2t-1)j_1} + e_{j_2}\alpha^{(b+2t-1)j_2} + \cdots + e_{j_v}\alpha^{(b+2t-1)j_v}.
\end{align*} \tag{6.6}
\]

If syndromes are all zeros, the codeword is considered to be received correctly. Otherwise,
we need to find the error locations and error values by investigating the syndromes. The error locator polynomial is defined as

$$\sigma(x) = \prod_{k=1}^{v} (1 - \alpha^{j_k} x) = 1 + \sigma_1 x^1 + \cdots + \sigma_{v-1} x^{v-1} + \sigma_v x^v. \quad (6.7)$$

Finding the zeros of $\sigma(x)$ are equivalent to obtaining the error locations. We have $\sigma(\alpha^{-j_k}) = 0$, where $j_k$ is one of the error locations. The $\sigma(\alpha^{-j_k})$ can be expanded as

$$0 = 1 + \sigma_1 \alpha^{-j_k} + \cdots + \sigma_{v-1} \alpha^{-(v-1)j_k} + \sigma_v \alpha^{-v(j_k)}. \quad (6.8)$$

In the next step, we relate syndromes and error locations without error values.

1. Multiplying both sides of (6.8) by $e^{j_k \alpha^{bj_k+vj_k}}$, we obtain

$$0 = e^{j_k \alpha^{bj_k+vj_k}} + \sigma_1 e^{j_k \alpha^{bj_k+(v-1)j_k}} + \cdots + \sigma_v e^{j_k \alpha^{bj_k}}. \quad (6.9)$$

2. Summing (6.9) from $k = 1$ to $v$, we have

$$0 = \sum_{k=1}^{v} e^{j_k \alpha^{bj_k+vj_k}} + \sum_{k=1}^{v} \sigma_1 e^{j_k \alpha^{bj_k+(v-1)j_k}} + \cdots + \sum_{k=1}^{v} \sigma_v e^{j_k \alpha^{bj_k}}. \quad (6.10)$$

3. From (6.6), we can express the syndrome $S_{v+1}$ as

$$S_{v+1} = \sum_{k=1}^{v} e^{j_k \alpha^{bj_k+vj_k}} \text{ for } 0 \leq v \leq 2t - 1. \quad (6.11)$$

We get following relation between error location polynomial coefficients and syndromes by substituting (6.11) into (6.9)

$$0 = S_{v+1} + \sigma_1 S_v + \cdots + \sigma_v S_1. \quad (6.12)$$

If we multiply both sides of (6.8) by $e^{j_k \alpha^{bj_k+(v+1)j_k}}$ in step 1 of above process, we obtain

$$0 = S_{v+2} + \sigma_1 S_{v+1} + \cdots + \sigma_v S_2. \quad (6.13)$$
By repeating the process, we get a set of equations (key equations)

\[
\begin{bmatrix}
S_1 & S_2 & \cdots & S_v \\
S_2 & S_3 & \cdots & S_{v+1} \\
\vdots & \vdots & \ddots & \vdots \\
S_v & S_{v+1} & \cdots & S_{2v-1}
\end{bmatrix}
\begin{bmatrix}
\sigma_v \\
\sigma_{v-1} \\
\vdots \\
\sigma_1
\end{bmatrix}
= 
\begin{bmatrix}
-S_{v+1} \\
-S_{v+1} \\
\vdots \\
-S_{v+v}
\end{bmatrix}.
\]  

(6.14)

The these equations can be efficiently solved using the Berlekamp-Massey algorithm, which finds an error locator polynomial (6.7) of minimal degree. The Berlekamp-Massey algorithm uses an iterative procedure. The detailed description of the algorithm and examples can be found in [78].

At the step-\(i\), we already have an error locator polynomial \(\sigma^i(x)\) of degree \(l_i\)

\[
\sigma^i(x) = 1 + \sigma^i_1x^1 + \cdots + \sigma^i_{l_i}x^{l_i},
\]  

(6.15)

which satisfies the following set of equations

\[
\sum_{j=0}^{l_i} S_{k-j}\sigma^i_j = 0, \ S_{k-j} \in \{S_1, S_2, \cdots, S_i\}.
\]  

(6.16)

The discrepancy at the step-\(i\) is calculated as

\[
d_i = S_{i+1} + \sigma^i_1 S_i + \cdots + \sigma^i_{l_i} S_{i-1-i+1}.
\]  

(6.17)

- If \(d_i = 0\), we do not need to update the error locator polynomial. Therefore, we get \(\sigma^{(i+1)}(x) = \sigma^{(i)}(x)\) and \(l_{i+1} = l_i\).

- If \(d_i \neq 0\), we need to compensate the discrepancy. Let \(\sigma^{(m)}(x)\) be the intermediate solution and \((m - l_m)\) is maximal. We update the error locator polynomial as

\[
\sigma^{(i+1)}(x) = \sigma^{(i)}(x) + d_i d_m^{-1} x^{i-m} \sigma^{(m)}(x)
\]

\[
l_{i+1} = \max\{l_i, l_m + i - m\}.
\]  

(6.18)
The Berlekamp-Massey algorithm starts with
\[
\begin{align*}
\sigma^{(-1)}(x) &= 1, \quad l_{-1} = 0, \quad d_{-1} = 1 \\
\sigma^{(0)}(x) &= 0, \quad l_0 = 0, \quad d_{-1} = S_1
\end{align*}
\] (6.19)

And the process updates error locator polynomial iteratively until \( i \geq l_{i+1} + t - 1 \) or \( i = 2t - 1 \).

After obtaining error locator polynomial coefficients \( \sigma_k \), we need to find error locations \( \alpha_{jk} \), which are roots of (6.7). The error locations can be found by Chien search [82], which simply evaluates (6.7) with each non-zero element \( \beta = \alpha_i \in GF(q) \). If \( \sigma(\alpha_i) = 0 \), we will accept \(-i\) as a valid error location.

The last step of decoding RS codes is to find errors values in (6.5), which is achieved by Forney algorithm. For the details of Forney algorithm, one can read [38, 78]. The partial syndrome polynomial is defined as
\[
S(x) \triangleq 1 + S_1 x + S_2 x^2 + \cdots + S_2^t x^{2t}.
\] (6.20)

And the error evaluator polynomial is defined as
\[
\Lambda(x) = \sigma(x) S(x) \mod x^{2t+1}.
\] (6.21)

The error value \( e_{ji} \) is calculated as
\[
e_{ji} = \left( \alpha^{ji} \right)^{2-b} \frac{\Lambda(\alpha^{-ji})}{\sigma'(\alpha^{-ji})},
\] (6.22)

where \( \sigma'(x) \) is the derivative of \( \sigma(x) \) with respect to \( x \). With the error values, we can recover the original codeword using \( C(x) = R(x) - E(x) \).

The decoding 4 steps of RS codes are summarized in Table 6.1.

### Table 6.1: The decoding procedure for RS codes.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>compute the syndromes ( S_j ) using (6.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>obtain error locator polynomial ( \sigma(x) ) of (6.7): Berlekamp-Massey algorithm</td>
</tr>
<tr>
<td>Step 3</td>
<td>find the error locations: Chien search</td>
</tr>
<tr>
<td>Step 4</td>
<td>calculate error values: Forney algorithm</td>
</tr>
</tbody>
</table>

In the next section, an introduction on RS/Conv concatenated coding scheme and the
related iterative decoding algorithm used in [75] is presented.

6.3 The RS-Convolutional Concatenated Coding Scheme

Concatenated coding was first proposed by Forney in 1966 [54] in form of RS/Conv serially concatenated coding scheme. The most famous concatenated coding scheme is probably the turbo code, which was invented as parallel concatenated convolutional codes in 1993 [53]. In this section, we provide an introduction to the RS/Conv concatenated coding scheme and an iterative decoding algorithm for it.

The block diagram of the RS/Conv concatenated coding scheme is shown in Fig. 6.1. At the transmitter side, the RS and the convolutional encoder are connected by an interleaver. The interleaving scheme has a depth of $I$, which means that $I$ RS codewords are grouped as the inputs to the encoder of convolutional codes. The interleaving is done at the bit level. At the receiver side, standard iterative turbo decoding is performed. The extrinsic information is exchanged between the two decoders. The convolutional decoder employs the BCJR algorithm and the RS decoder employs Pyndiah’s SISO algorithm [74,75].

![Figure 6.1: The block diagram RS/Conv concatenated coding scheme.](image)

The Pyndiah’s algorithm is a soft-input soft-output based on the Chase decoding algorithm [83]. The Chase decoding algorithm is a soft-input hard-output candidate-generating algorithm. In Chase algorithm, the a certain number of least reliable bits in received vector $\mathbf{Y}$ are inverted to generate a set of test patterns $\mathbf{Y}^m$. Those test patterns are fed into a hard-input hard-output decoder to obtain corrected patterns $\mathbf{X}^m$, i.e., codeword candidates. The decoded codeword is selected as the codeword candidate with the least soft distance to
Let $D$ be the hard decision codeword, found by the Chase algorithm, which has the least distance to the received vector $Y$ among all candidates. For a bit $d_j$ in $D$, the soft reliability is calculated as

$$r_j = \frac{M^{\min(-d_j)} - M^{\min(d_j)}}{4} d_j$$

$$= \frac{|Y - C|^2 - |Y - D|^2}{4 d_j},$$

(6.23)

where $C$ is the competing codeword which is at the minimum distance from $Y$ among all candidates with $c_j = -d_j$. If at this step we cannot find $C$, $r_j$ is assigned a precalculated value $\beta(k) \in [0, 1]$, which depends on the number of iteration $k$. Because the BCJR decoder uses log-likelihood ratio (LLR) values, the information provided by the RS decoder should be converted to LLR values which for an AWGN channel is given by

$$r_j^{LLR} = 2 \times r_j / \sigma^2,$$

(6.24)

where $\sigma^2$ is the variance of the noise. Note that $r_j$ is the soft output of the decoder, we need to subtract the a priori information from $r_j$ to generate the extrinsic information. It is possible that we cannot find the codeword $D$ is our search. If this happens, we simply set the extrinsic information generated by the RS decoder to be zero. To increase the reliability during the first iterations, a precalculated scaling parameter $\gamma(k) \in [0, 1]$ for the extrinsic information is also introduced.

### 6.4 A Low Complexity Decoding Algorithm

The proposed algorithm is a modified version of Pyndiah’s soft output algorithm for linear block codes. We add Kaneko’s stopping criterion during the candidate generation process. Once we find the most likely codeword, the algebraic decoding is halted to reduce the delay and complexity.

Let us consider a RS code with codeword length $n$, information length $k$, and minimum distance $d = n - k + 1$, denoted as a $(n, k, d)$ code. The received vector is represented as

$$Y = \{y_0, y_1, \cdots, y_{n-1}\},$$

(6.25)
where each element consists of $q$ components

$$y_i = \{y_{i,0}, y_{i,1}, \ldots, y_{i,q-1}\}. \quad (6.26)$$

For $q$-ary RS codes, there exist a total of $2^q$ possible symbols. The reliability that $y_i$ is equal to a certain symbol $j$ is calculated as

$$\theta_{i,j} = \ln \frac{p(y_i|j)}{\sum_{l \neq j} p(y_i|l)}, \quad l \in \{0, 1, \ldots, 2^q - 1\}. \quad (6.27)$$

Let $\theta_{1st}^i, \theta_{2nd}^i$ denote the largest and the second largest components in

$$\theta_i = \{\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,2^{q-1}}\}. \quad (6.28)$$

The hard decision sequence $Y^H$ of $Y$ is obtained by

$$y_{i,j}^H = \begin{cases} 
0, & y_{i,j} \leq 0, \quad i = 0, 1, \ldots, n - 1 \\
1, & y_{i,j} > 0, \quad j = 0, 1, \ldots, q - 1.
\end{cases} \quad (6.29)$$

Let $X$ be a legitimate codeword in the codeword table. $S_X$ denotes the set of positions where $X$ and $Y^H$ are equal and $S_X^c$ is the complement of $S_X$. The likelihood metric $L(Y, X)$ is calculated as

$$L(Y, X) = \sum_i \theta_{i,x_i} = \sum_i \theta_{i,y_i^H} - \sum_{i \in S_X^c} \left( \theta_{i,y_i^H} - \theta_{i,x_i} \right). \quad (6.30)$$

Since we know $Y$, the above metric only depends on

$$l(Y, X) = \sum_{i \in S_X^c} \left( \theta_{i,y_i^H} - \theta_{i,x_i} \right) = \sum_{i \in S_X^c} \left( \theta_{1st}^i - \theta_{i,x_i} \right). \quad (6.31)$$

After receiving $Y$, we first rearrange its elements in ascending order

$$\theta_{1st}^{S_Y(0)} - \theta_{2nd}^{S_Y(0)} \leq \theta_{1st}^{S_Y(1)} - \theta_{2nd}^{S_Y(1)} \leq \cdots \leq \theta_{1st}^{S_Y(n-1)} - \theta_{2nd}^{S_Y(n-1)}.$$ \quad (6.32)

The basic idea of Kaneko’s algorithm is to search for the codeword $X$ which satisfies the
following relation

\[ l(Y, X) < \sum_{i=0}^{d-m_0-1} (\theta_{S_Y(i)}^{1st} - \theta_{S_Y(i)}^{2nd}), \]  

(6.33)

where \( m_0 = ||S_{X_0}^c|| \) is the size of set \( S_{X_0}^c \) and \( X_0 \) is the algebraic decoder output when \( Y^H \) is the input. It can be proved that the codeword satisfying the inequality (6.33) is the most likely codeword [76]. If we do not find \( X_0 \), we can use \( m = ||S_X^c|| \) instead of \( m_0 \) in (6.33). We test each legitimate codeword we have found during the decoding process using this inequality. Once the most likely codeword is found, the algebraic decoding is halted and we start computing the soft outputs using Pyndiah’s soft output algorithm.

Kaneko’s algorithm performs true maximum-likelihood-decoding (MLD) if we do not limit the number of generated candidate codewords. However, we set a maximum number of test error patterns in our decoding process as in the Chase algorithm because when the signal-to-noise ratio (SNR) is low, we need to generate a large set of candidates. In this situation, finding the most likely codeword takes a long time. It should also be noted that inequality (6.33) is a sufficient, and not a necessary, condition that \( X \) is the ML codeword. We may miss the most likely codeword and continue to test all possible error patterns. In this situation, we decode unnecessary error patterns.

The proposed algorithm includes three steps.

1) Generate error patterns.

2) Perform algebraic decoding using each error pattern and test the decoder output according to (6.33). The test tells whether the most likely codeword is found. If the most likely codeword is found or maximum number of tests is reached, continue to step 3); Otherwise, we go back to step 2) to decode the next error pattern.

3) Calculate soft outputs using Pyndiah’s soft output algorithm.

If the maximum number of tests is reached, the system will perform exactly as the Chase algorithm does. Otherwise, we will find the most likely codeword in the process and save processing time. In the next section, we will demonstrate that there is no performance loss although we generate fewer candidates in the proposed algorithm.

The error patterns are generated using natural binary coding (NBC) in our work. An example of 8 error patterns is listed in Table 6.2. The bits are sorted by their reliabilities.
Table 6.2: Error patterns of flipping 3 bits.

<table>
<thead>
<tr>
<th></th>
<th>bit 2</th>
<th>bit 1</th>
<th>bit 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Error 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Error 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Error 3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Error 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Error 5</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Error 6</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Error 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

in ascending order. Bit 0 is the most unreliable bit in the codeword. We can further reduce
the complexity by employing the error pattern generating method proposed in [84].

6.5 Simulation Results

We use a (31,27,5) RS code as the outer code and a rate-1/2 recursive systematic convolu-
tional (RSC) code with generator polynomials (23, 35) as the inner code. The interleaving
depth $I$ is set to be 6. The soft output generating parameters used by Pyndiah’s algorithm
$\{\beta(k)\} = \{0.4, 0.7, 0.9, 1\}$ and $\{\gamma(k)\} = \{0.6, 0.8, 1, 1\}$ are the same as those used in [75].
The maximum number of error patterns for RS decoding is set to be 16. The simulation
results are obtained with AWGN channel.

In Fig. 6.2, we show the bit error rate (BER) performance of the concatenated coding
scheme with the proposed low complexity decoding algorithm. Compared with the results
in [75], we can see that the new algorithm archives nearly the same performance. The
reason is that even if we do not find the competing codeword $C$ in our candidates, the
soft reliability calculation using a predetermined value $\beta(k)$ is good enough to estimate
the extrinsic information. We also observe that the performance improvement is marginal
beyond 4 iterations.

The complexity comparison plots are shown in Fig. 6.3. As the algebraic decoding for
RS code is the computationally demanding step, the computation complexity is measured
by the number of algebraic decoding performed during the decoding process. The curves
with legends of iteration orders are used to denote the complexity of the new algorithm.

The Chase algorithm has constant complexity of 1 regardless of the SNR or the order
of the iteration, since it generates constant number of error patterns. The complexity of
Figure 6.2: BER performance of the concatenated RS/Conv coding system.
the proposed algorithm is reduced as the SNR or the iteration order increases, which means that we can test fewer error patterns at the low BER region. At the low right corner of Fig. 6.3, we can see the complexity of proposed algorithm converges to be 1/16 of the original algorithm. In this situation, we only need to test one error pattern, which is the all zero error pattern.

Figure 6.3: Complexity comparison between proposed algorithm and Chase algorithm.

6.6 Conclusions

In this work, we proposed a low complexity decoding algorithm for the concatenated RS/Conv coding scheme. The proposed algorithm can be viewed as an improved version of Pyndiah’s algorithm with Kaneko’s stop criteria. The algorithm greatly reduces the decoding complexity at the low BER region, while barely causing any degradation in performance. The proposed decoding algorithm can be applied to lengthened and shortened RS codes for block-fading channels.
Chapter 7

Concluding remarks and future work

7.1 Concluding remarks

The focus of this dissertation is code design for block-fading channels. We first explored code design criteria for block-fading channels. Two main factors on performance of codes over block-fading channels are diversity order and coding gain. The Shannon capacity of block-fading channel is zero; therefore, we employed the outage limit as the ultimate PER lower bound for any communication systems over block-fading channels. In the first part of the research, the focus was on code design for SISO block-fading channels. In the second part of the research, the transmission schemes for MIMO block-fading channels were presented.

We proposed a BC-BICM scheme for block-fading channels. The proposed scheme is a simplified version of BCC-BICM scheme. The optimal diversity order is achieved by employing a blockwise MDS outer code. In addition, the coding gain can be improved either by choosing a carefully designed signal labeling scheme for the BICM with iterative decoding or by using convolutional codes with longer constraint lengths. Compared with BCC-BICM scheme, these coding schemes have smaller delays in both encoding and decoding processes. Simulation results show that the optimized BC-BICM scheme offers similar performance as the BCC-BICM scheme in block-fading channels but with a simpler system architecture.

We also presented QC-LDPC codes for block-fading channels. A construction method for designing QC-LDPC codes was presented. With careful design, the proposed QC-LDPC codes exhibit the same good performance as their corresponding random root-LDPC codes.
with iterative belief-propagation decoding. Moreover, the structure of the proposed QC-LDPC codes makes them suitable for efficient encoding.

Majority of modern wireless communication systems are equipped with multiple antennas. The two main purposes of using MIMO transmission is to improve the robustness of the communication by introducing space diversity and to increase transmission rate by spatial multiplexing. Since the diversity order is limited by channel intrinsic diversity order, i.e., the number of fading blocks in the block-fading channel, we consider employing multiple antennas to provide additional diversity. We investigated turbo coded BICM schemes for MIMO block-fading channels. Both serial and parallel concatenated turbo codes were considered and their performance was compared. A sphere decoder was employed to achieve good performance with reduced computational complexity when the size of the constellation or the number of transmit antenna is large.

The turbo BICM scheme, however, cannot provide full diversity order. The interference due to using multiple transmission antennas prevents serial concatenated turbo codes to achieve the full diversity order in MIMO situation. We proposed a new coded MIMO scheme based on modulation diversity technique for block-fading channels. The proposed scheme can achieve both full spatial multiplexing and full diversity of \( N_t N_r F \), i.e., the product of the number of transmit antennas \( N_t \), receive antennas \( N_r \) and fading blocks \( F \). The design concatenates a channel code with a space-time code. The channel code is responsible for providing time diversity and the space-time code is employed to achieve space diversity. Since the modulation diversity technique enables full space diversity and full spatial multiplexing at the cost of increased processing complexity at the receiver, sphere decoding is employed to reduce detection complexity.

The RS codes are well-known MDS codes and they can be modified for block-fading channels. An efficient decoding algorithm was proposed for concatenated coding schemes that use a RS code as the outer code and a convolutional code as the inner code. We compared the proposed decoding algorithm with Pyndiah’s algorithm in terms of both computational complexity and performance. The proposed decoding algorithm reduces the decoding complexity by adopting Kaneko’s decoding algorithm without suffering a performance loss.
7.2 Future work

In our work, we investigated code design for the case, where channel state information is available only at the receiver side. For MIMO system, precoding and transmitter side beam-forming techniques can be employed to improve system performance, if channel state information is known at the transmitter side. The channel information at the transmitter side is a valid assumption for some modern communication architectures. For example, in the LTE system with time division duplexing (TDD) where up-link and down-link share the same frequency channel, the channel information at the transmitter can be obtained through feedback from the receiver.

Also, we modeled the fading coefficients on each block as independent Rayleigh distributed random variables. The performance will be degraded, if the fading coefficients on different blocks are correlated. It is interesting to investigate the performance loss in this situation.
Bibliography


[75] O. Aitsab and R. Pyndiah, “Performance of concatenated Reed-Solomon/convolutional codes with iterative decoding,”


