Design of Turbo MIMO Wireless Communications

Zhiyuan Wu

A Thesis

in

The Department

of

Eletrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

December 2006

Your file Votre référence
ISBN: 978-0-494-30140-1
Our file Notre référence
ISBN: 978-0-494-30140-1

Abtract

Design of Turbo MIMO Wireless Communications

Zhiyuan Wu, PhD

Concordia University, 2006

Since the discovery of the significant gain in capacity provided by multiple antennas over fading channels, tremendous research and development efforts in academia and industry have been invested to multiple-input-multiple-output (MIMO) technology. However, difficulties still exist in the design of flexible MIMO transmission schemes. The overall goal of our study is to develop efficient turbo MIMO transceivers that are capacity-achieving and yet with reasonable complexity.

First, we study the design of serially concatenated MIMO transmitters. For simpler design and flexible rate-versus-performance tradeoff, conventional encoders are used before a linear space-time modulator. A joint iterative receiver based on the turbo principle is assumed that precludes the use of Tarokh's design criteria for such a concatenated system. Extending the *extrinsic* information transfer (EXIT) charts to MIMO systems, design criteria that concern both the data rate and error performance are developed for the inner space-time (ST) modulator. Based on the new design criteria, an optimal space-time linear dispersion modulation scheme is presented. In addition, the tradeoff between constellation size and symbol rate for a given data rate is discussed. Simulation results are provided to verify the new design criteria and to demonstrate the merits of the proposed coded space-time modulation.

For such a turbo MIMO transmitter, the abundance of highly developed outer decoders allows us to focus on the study on the inner multi-user detection (MUD). To start, a new parallel interference cancellation (PIC) MUD scheme is proposed. This detector can be used as a stand-alone multi-user detector in a non-iterative MIMO receiver. The new scheme employs a novel nonlinear minimum mean square error (MMSE) estimator as the soft decision device. The nonlinear estimator exploits the knowledge of symbol alphabet for refined estimates with smaller mean square error (MSE). Simulation results demonstrate that the proposed detector significantly outperforms the conventional detectors with comparable complexity. The nonlinear MMSE estimator is further extended to develop the soft-output MUD for iterative turbo MIMO receivers. The nonlinear estimation makes use of *a priori* information from the outer decoder and *preliminary estimates* gleaned from channel observation as well. With the new nonlinear MMSE estimator, a generalized SIC-MMSE scheme is proposed. Depending on the methods for generating preliminary estimates, two generalized SIC-MMSE detectors are proposed: a 2-staged SIC-MMSE and a recursive SIC-MMSE detectors. Analytical study and simulations are carried out to demonstrate the merits of the two proposed detectors.

Last, tradeoffs among the components of the transmitter that affect error performance at the receiver are discussed to provide design guidelines for practice.

Acknowledgments

I would like to express my deep gratitude to my supervisor Dr. Xiaofeng Wang for his great guidance, advice and support. Without him, I could not pursue Ph.D degree in Concordia University. I would like to thank my administrative supervisor Dr. Walaa Hamouda for his generous help in finalizing my thesis.

I would like to thank the members of my dissertation committee. Dr. Yousef R. Shayan, Dr. William E. Lynch, and Dr. Mingyuan Chen gave me invaluable comments and suggestions during my study. I am grateful to Dr. Ke Wu at Ecole Polytechnique for his kind acceptance to be the external examiner of this dissertation.

Special thanks to my wife, Zhuni Bao, for her continuous support and encouragement in my life. I would also like to express my profound appreciation to my family for their support and encouragement.

I would like to extend my sincere thanks to all professors, office stuff, and graduate students at this department for their assistance in various ways.

ZHIYUAN WU

Concordia University

December 2006

To My Family

Contents

List of Figures

Notations and Abbreviations

- $\mathbf{X}$: upper bold letter for matrix
- $\mathbf{x}$: lower bold letter for column vector
- $\mathbf{X}^H$: hermitian of $\mathbf{X}$
- $\mathbf{X}^T$: transpose of $\mathbf{X}$
- $\otimes$: Kronecker product
- diag[x]: a diagonal matrix with x on its main diagonal
- tr($\mathbf{X}$): trace of $\mathbf{X}$
- det($\mathbf{X}$): determinant of $\mathbf{X}$
- vec($\mathbf{X}$): a column vector formed by stacking the column vectors of $\mathbf{X}$ in order
- $\| \ \|$: Euclidean norm
- $\{x\}$: a set of x
- $P_r(x)$: probability of event x
- $E(x)$: expectation of x
- $I(x;y)$: mutual information between x and y
- APP: *a posteriori* probability

- AWGN: additive white Gaussian noise
- BCJR: Bahl, Cocke, Jelinik and Raviv
- BER: bit error rate
- BICM: bit-interleaved coded modulation
- BLAST: Bell-labs layered space-time
- CDF: cumulative distribution function
- CSI: channel state information
- CSTM: coded space-time modulation
- EXIT: *extrinsic* information transfer
- FER: frame error rate
- LD: linear dispersion
- LDC: linear dispersion code
- LLRs: log-likelihood-ratios
- MAP: maximum *a posteriori* probability
- MFR: matched filter receiver
- MIMO: multiple-input-multiple-output
- ML: maximum-likelihood
- MMSE: minimum mean square error
- MSE: mean square error
- MUD: multi-user detection

- PIC: parallel interference cancellation
- SIC: successive interference cancellation
- SIC-MMSE: soft-interference-cancellation minimum mean square error
- SISO: soft-input soft-output
- SNR: signal-to-noise ratio
- ST: space-time
- STBC: space-time block code
- STTC: space-time trellis code
- ST Turbo TC: space-time turbo trellis code
- TCM: trellis-coded modulation
- T-BLAST: thread BLAST
- V-BLAST: vertical BLAST
- *i.i.d.*: independently identically distributed

Chapter 1

Introduction

Over the last several decades, the demand for high reliability and bandwidth efficiency in wireless communications has experienced an unprecedented growth. With this growth, radio spectrum has been recognized as one of the very precious resources of nature. In response to the growing demand of higher bandwidth efficiency, the multiple-input multiple-output (MIMO) technology has been developed and is regarded as the most important advance in today's wireless communications. Since the advent of MIMO technology in 1995 and 1996, tremendous research and development efforts in academia and industry have been invested, and this investment is ever increasing. To date, MIMO technology has been widely used in modern wireless communication systems, such as WLAN and 3G cellular systems, and is recognized as the most important enabling technology for future wireless communication systems.

The MIMO technology makes use of multiple transmit and receive antennas to improve the data rate and error performance over fading channels. Recent results in information theory have demonstrated that the capacity of a scattering-rich MIMO channel grows linearly with respect to the minimum number of the transmit and receive antennas [1]-[3]. Furthermore, the use of multiple antennas increases the reliability via diversity [4]-[9]: in a MIMO system with N_t transmit and N_r receive

antennas, given that the path gains between individual antenna pairs are independently Rayleigh faded, the MIMO system can achieve a maximal diversity gain of $N_t N_r$.

With about a decade's fast development in MIMO technology, it is clear that today that the most promising technique for high data rate applications is the so-called turbo MIMO communications, which combines powerful and well-understood conventional binary codes with simple spatial multiplexing schemes on the basis of the *turbo principle* [10].

In this thesis, the focus is on the design of efficient turbo MIMO transceivers that are capacity-achieving and yet with reasonable complexity. The new transceivers are aimed at applications in current and near-future communications.

1.1 Previous Works

The achievable channel capacity over MIMO channels depends on the channel state information (CSI) available at the transmitter and the receiver [1]-[3]. In most applications, CSI is known or can be estimated at the receiver but unknown to the transmitter. Most existing space-time (ST) coding designs based on this assumption mainly fall into two categories: either performance-oriented schemes by exploiting the achievable spatial diversity, such as the space-time trellis codes (STTCs) [4], space-time block codes (STBCs) [5][6] and space-time turbo trellis codes (ST Turbo TCs) [7]-[9], or rate-oriented schemes by capitalizing the capacity gain of MIMO fading channels (spatial multiplexing gain), such as Bell-labs layered space-time (BLAST) architectures [11]-[13] and various linear dispersion codes (LDCs) [14]-[16].

The design of the performance-oriented schemes is aimed at lower error rate. In this category, STTCs, pioneered by Tarokh *et al* [4], are a joint processing of spatial-temporal modulation and encoding, which can be viewed as an extended two-

dimensional trellis-coded modulation (TCM) design. STTCs can achieve full diversity and large coding gain. However, their computational complexity grows exponentially with respect to the number of states and transmit antennas and they are often designed by hand. STBCs [5][6] are two-dimensional linear coding applied on a block of input symbols. It is very attractive because it can achieve full diversity gain with orthogonal design and possess significantly low complexity with linear structure. However, it suffers from smaller coding gain and often loss in data rate. ST Turbo TCs [7]-[9] can achieve very low bit error rates with a suboptimal but powerful iterative decoding algorithm. However, like STTCs, the decoding complexity grows exponentially with respect to the number of states and transmit antennas and the design is generally conducted by exhaustive search. Furthermore, how to achieve the maximum potential diversity and coding gain is still unclear for ST Turbo TCs.

The design of the rate-oriented schemes is aimed at maximizing achievable data rate. For BLAST architectures [11]-[13], the key component is the spatial multiplexing that multiplexes several data streams (often referred to as *layers*) in a way such that each stream will experience all the fading modes. In LDCs [14]-[16], the data stream is broken into sub-streams that are dispersed in different fashions over space and time. The codes are designed to optimize the mutual information between the transmit and receive signals. The linear structure of these rate-oriented schemes allows a variety of decoders including simple linear techniques. Although the rate-oriented schemes can achieve higher data rate with lower complexity, they are often designed without fully considering the error performance, owing to the difficulties in applying the Tarokh's design criteria [4] in these high-rate space-time schemes. Consequently, the error performance of these high-rate schemes is often less satisfactory.

In summary, existing space-time codes based on the above two approaches suffer from the design difficulty, performance loss or decoding complexity. In addition, they often lack the flexibility of rate-versus-performance tradeoff, which is the key

to the future success of wireless communications with the expectation of a variety of levels of quality of services. In [17], Zheng and Tse have shown that there exists an optimal tradeoff curve between diversity and multiplexing, but it is still not clear how to achieve the maximal diversity gain for a given multiplexing gain with structured codes. This has motivated a major research thrust targeted at developing practical MIMO schemes with high data rate, desirable error performance and reasonable decoding complexity to fulfill a variety of requirements in wireless communications.

Most recently, the idea behind concatenated coding schemes has been applied in MIMO communications. By combining two or more relatively simple *constituent codes*, a concatenated coding scheme [18] can achieve large coding gain with a moderately complex decoding. In addition, such a coding structure also allows flexible and simple design. A "Turbo code" first proposed in [19] can be thought of as a refined concatenated coding scheme which is capacity-achieving. With recent progress in MIMO transmission, it has been recognized that the idea of the powerful "Turbo codes" can also be applied in the MIMO system. Both parallel and serial concatenated systems have been proposed. In the parallel concatenated systems [7]-[9], the information bit stream is passed though two or more encoding processes with different permutation and then punctured and multiplexed at the transmit antennas. In serial concatenated systems such as [10], [20]-[24], some form of outer encoding is applied before space-time coding/modulation. Such a serial concatenated system is often preferable due to its simpler design and greater flexibility in rate-versus-performance tradeoff. In fact, most space-time schemes can be subsumed as an outer encoder serially concatenated before an inner ST modulator which performs space-time mapping or modulation that maps a number of input symbols onto a space-time grid before transmission. For example, a ST Turbo TC can be viewed as an outer turbo channel encoder serially concatenated before an inner V-BLAST [12] mapper.

In a serial concatenated MIMO transmitter, often, conventional encoders such

as convolutional codes, trellis-coded modulation (TCM), and turbo codes designed for single-input single-output channels can be used to provide extra redundancy and to simplify the design of the inner ST modulator. In order to decouple the correlation between outer encoding and inner ST modulation, interleaving is often applied to the encoded bits. Hence, the corresponding concatenated system is often called MIMO bit-interleaved coded modulation (BICM) [20]-[24]. Such a concatenated coding system possesses many advantages of both the conventional codes and the inner ST modulation. On the one hand, conventional outer codes can provide large *coding gain and time diversity*; on the other hand, space-time coding/modulation provides guaranteed *spatial multiplexing and diversity gains.* Together, they enable a variety of design targets in performance, bandwidth efficiency, complexity, and tradeoffs among them.

For such a turbo MIMO transmitter, the abundance of the conventional codes optimized for a variety of purposes allows us to focus on the design of inner ST modulator first. However, how to design the inner ST modulator is still unclear to us. Although any existing space-time code can be a potential candidate for the inner space-time modulation, a particular desirable choice is linear dispersion (LD) codes. This is because it subsumes many existing block codes as its special cases, allows suboptimal linear receivers with greatly reduced complexity, and provides flexible rate-versus-performance tradeoff [14]. To emphasize the "modulation" flavor of the inner ST process, we will also call such a serially concatenated system as coded space-time modulation (CSTM).

In such a CSTM system, the use of interleaver makes it impractical to evaluate the coding gain and diversity gain of the overall system based on Tarokh's criteria. Furthermore, Tarokh's criteria are developed based on maximum-likelihood (ML) decoding which is too complex to be practical for a concatenated system. At best, a joint iterative (turbo) receiver based on the *turbo principle* can be used [10]. In such

a case, Tarokh's criteria apply neither to the overall concatenated system nor to the inner ST modulation alone. Hence, it is important to study the design of the inner ST modulator when used in a concatenated system.

The optimal turbo receiver employs *a posteriori* probability (APP) detector/decoder for both the inner ST modulation and the outer code [25][26]. However, the complexity of an APP detector/decoder is prohibitively high, particularly for the inner ST modulation that often does not possess a trellis structure. For reduced complexity, suboptimal multiuser detection (MUD) schemes, such as [27]-[29], can be used as the inner ST detector in a turbo receiver for a concatenated MIMO transmission system. In the thesis, both the transmitter and receiver designs are considered.

1.2 Scope of Thesis Work

This thesis consists of seven chapters.

In Chapter 2, research backgrounds and concepts are introduced pertaining to turbo MIMO system, including MIMO channel models, performance measures and tradeoffs, linear dispersion codes (LDCs), turbo principle and iterative receiver, and the *extrinsic* information transfer (EXIT) chart technique.

In Chapter 3, we study the turbo MIMO transmitter, also referred to as coded space-time modulation (CSTM), for MIMO wireless channels. For simpler design and flexible rate-versus-performance tradeoff, conventional encoders are used before a linear space-time modulator. A joint iterative receiver based on the turbo principle is assumed that precludes the use of Tarokh's design criteria for space-time codes. Using the EXIT chart technique, design criteria that concern both the data rate and error performance are developed. These criteria are much easier to apply than the well-known Tarokh's criteria. It is shown that the use of outer encoders significantly simplifies the design of linear space-time coding/modulation. Based on the new design

criteria, an optimal space-time linear dispersion modulation scheme is presented. In addition, the tradeoff between constellation size and symbol rate for a given overall data rate is discussed. Simulation results are provided to verify the new design criteria and to demonstrate the merits of the proposed CSTM.

Before the study on the turbo MIMO receiver, stand-alone inner MUD schemes are studied in Chapter 4. In practice, the application of the optimal MUD is prohibitive due to its exponentially increasing complexity. As such, numerous suboptimal schemes with much lower complexity are developed. Nonlinear interference cancellation techniques, such as successive interference cancellation (SIC) and parallel interference cancellation (PIC), are more attractive than linear MUD schemes due to their better error performance. In this study, a new PIC multi-user detection scheme is proposed. The new scheme employs a novel nonlinear minimum mean square error (MMSE) estimator that replaces the soft decision device in the conventional partial PIC schemes. The nonlinear estimator exploits the knowledge of symbol alphabet for refined estimates with smaller mean square error (MSE). The overall computational complexity is close to that of the conventional PIC schemes. Simulation results demonstrate that the proposed detector significantly outperforms the conventional detectors.

In Chapter 5, the stand-alone, hard-output MUD developed in Chapter 4 is extended to develop soft-output MUD to be used in turbo receivers. In a turbo MIMO receiver, suboptimal soft MUD schemes are often preferred due to the prohibitively high complexity of the optimal multi-user detector. A remarkable suboptimal scheme is the so-called soft interference cancellation plus minimum mean-square error (SIC-MMSE) detection that uses nonlinear estimates, generated solely from the *a priori* information provided by the outer decoder(s), for soft interference cancellation before linear MMSE detection. Besides the *a priori* information, channel observation also provides useful information in estimating the interfering symbols. Motivated by

this observation, a nonlinear MMSE estimator is developed that takes the *a priori* information and a so-called *preliminary estimate* gleaned from channel observation together to produce estimates. With the nonlinear estimation, a generalized SIC-MMSE scheme is proposed. Based on this scheme, two new SIC-MMSE detectors are developed. One is a 2-staged detector in which the first stage is just a conventional SIC-MMSE detection used to produce preliminary estimates. The other is a recursive detector which uses the detection outputs from the *previous* iteration. Analytical study and simulation results demonstrate that the turbo receivers with the two proposed detectors significantly outperform the conventional SIC-MMSE-based receiver. Among the three detectors, the proposed recursive detector leads to the best error performance at convergence but with the lowest complexity.

In Chapter 6, tradeoff issues of the turbo MIMO transceiver are discussed. Effects on overall performance are presented for each constituent component at the transmitter, i.e., outer encoder, interleaver, constellation mapper and inner ST modulator. For a given overall data rate, tradeoffs among three important system parameters are studied: the tradeoff between the coding rate and constellation size, the tradeoff between the coding rate and the modulation symbol rate, and the tradeoff between the constellation size and the modulation symbol rate. Analytical results are presented to give design suggestions in practice.

Finally, in Chapter 7, some future works are identified and conclusions are drawn.

1.3 Contributions

The main contribution of Chapter 3 is the proposed design criteria for the inner ST modulator in turbo MIMO transmitters. Although turbo MIMO systems have been studied in numerous literatures, how to design the inner ST modulator is still unclear.

Furthermore, the existing design criteria for space-time codes, such as Tarokh's criteria based on ML decoding, can no longer be applicable to such a turbo system due to the use of random interleaver. In this study, the idea of EXIT chart, proposed for single-input single-output channels by S. ten Brink, is first extended to turbo MIMO systems. Based on the technique, new design criteria are developed for the inner ST modulator, i.e., *Capacity and Error-Performance Criteria.* With the new design criteria, an optimal scheme is proposed. For practical applications, tradeoff between the constellation size and symbol rate for a given overall data rate is also discussed. The relevant studies have been published in [40] and [41].

The main contribution of Chapter 4 is a new suboptimal MUD scheme based on parallel interference cancellation and nonlinear MMSE estimation. In multi-user communications and single-user MIMO communications, nonlinear interference cancellation is desirable due to their low complexity and good performance. The proposed scheme can be viewed as a partial PIC scheme, whose key component is a nonlinear MMSE estimator used as soft decision device. Although partial PIC schemes have been studied intensively in the past, the soft decision devices are often developed intuitively. In this study, a novel nonlinear estimator is developed by exploiting knowledge of symbol alphabet in sense of MMSE. With the nonlinear estimation, refined symbol estimates can be obtained for interference cancellation and linear detection. Simulation results demonstrate that the proposed scheme significantly outperforms the existing linear schemes and other partial PIC schemes with comparable complexity. The relevant studies have been published in [55] and [57].

The main contribution of the Chapter 5 is a generalized SIC-MMSE detector proposed for turbo MIMO receivers. The novelty of the generalized SIC-MMSE detection exploits channel observation, in addition to the *a priori* information in interference estimation. This is done by extending the nonlinear MMSE estimation in Chapter 4. Two generalized SIC-MMSE are proposed: one is a 2-staged detector

and the other is a recursive detector. Analytical study and simulation results show that the turbo receiver with the two proposed detectors significantly outperform the conventional SIC-MMSE in terms of error performance with reduced complexity. The relevant studies have been published in [66].

In Chapter 6, tradeoffs among components of a turbo MIMO transmitter are studied. Guidelines for specific designs in practice are provided.

Chapter 2

Preliminaries

In this chapter, some background concepts and terminologies pertaining to turbo MIMO systems are presented. The content in this chapter will provide the foundation for the development of subsequent chapters.

2.1 MIMO Channel Model

In wireless communications, the surrounding static and moving objects such as buildings, trees and vehicles act as reflectors so that multiple reflected waves of the transmitted signals arrive at the receive antennas from different directions with different propagation delays. When these signals are collected at the receiver, they may add constructively or destructively depending on the random phases of the signals arriving at the receiver. The amplitude and phase of the combined multiple signals vary with the relative movement of the surrounding objects in the wireless channel. The resultant fluctuation is called fading [30].

Fading, can be catalogued into flat fading also known as frequency non-selective fading and frequency selective fading. In a flat fading channel, the transmitted signal bandwidth is smaller than the coherence bandwidth of the channel. Hence, all frequency components in the transmitted signal are subjected to the same fading at-

tenuation. In a frequency selective fading channel, the transmitted signal bandwidth is larger than the coherence bandwidth of the channel, different frequency components in the transmitted signal experience different fading attenuation. As a result, the spectrum of the received signal differs from that of the transmitted signal. This is called distortion.

Fading can also be classified as fast fading or slow fading depending on how rapidly the channel changes compared to the symbol duration. If the channel can be deemed constant over a large number of symbols, the channel is said to be a slow fading channel; otherwise it is a fast fading channel [31].

In wireless communications, the envelope of the received signal can be described by *Rayleigh* distribution or *Ricean* distribution. In a no line-of-sight propagation, Rayleigh distribution is applied and the fading is called Rayleigh fading. While in a line-of-sight propagation, since there exists a dominant non-fading component, Ricean distribution is often used to model the envelope of the received signal. Thus it is called Ricean fading.

In this study, we consider applications where the channel is unknown to the transmitter but known or can be perfectly estimated at the receiver. Furthermore, our attention is limited to uncorrelated Rayleigh flat fading channel where the channel gains are independently Rayleigh faded. The research results can also be extended to other channel conditions.

Consider a MIMO channel with N_t transmit and N_r receive antennas as shown in Figure 2.1. The gain of the path from the n-th transmit antenna to the m-th receive antenna is denoted by h_{mn} with $n = 1, 2, \ldots, N_t$ and $m = 1, 2, \ldots, N_r$. The path gains are assumed to be independently complex circular symmetric Gaussian with zero mean and unit variance, that is, $h_{mn} \sim CN(0, 1)$. All the complex gains are collected to form the MIMO channel $\mathbf{H} = [h_{mn}]$. The transmitted symbols are denoted by a column vector $\mathbf{x} = [x_1, x_2, \ldots, x_{N_t}]^T$ with x_n being the symbol transmitted on

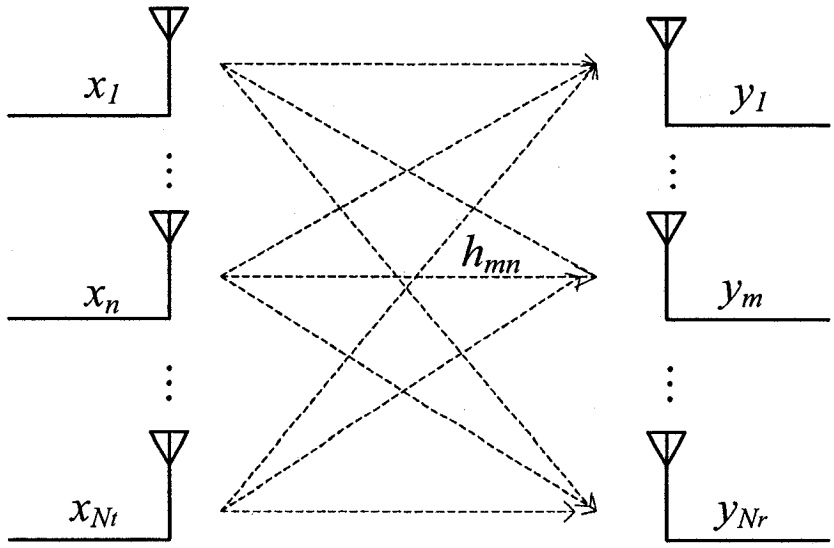


Figure 2.1: A MIMO flat-fading channel.

the n-th transmit antenna. The symbols are assumed to be mutually independent to each other with unit average symbol energy for convenience. Then the received signal vector $\mathbf{y}$ is written as

$$\mathbf{y} = \sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{x} + \mathbf{z} \tag{2.1}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_{N_r}]^T$ with y_m being the signal received at the m-th receive antenna P is the total transmitted power over the N_t transmit antennas, and $\mathbf{z}$ is the additive white Gaussian noise (AWGN). The entries of $\mathbf{z}$ are assumed to be independently identically distributed (*i.i.d.*) and symmetrical complex Gaussian random variables with zero mean and variance σ_z^2.

When the MIMO channel $\mathbf{H} = [h_{mn}]$ is known perfectly to the receiver but unknown to the transmitter, the ergodic capacity of such a MIMO channel is given by [1][2]

$$C\left(\frac{P}{\sigma_z^2}\right) = E\left[\log\det(\mathbf{I}_{N_t} + \frac{P}{N_t\sigma_z^2}\mathbf{H}\mathbf{H}^H)\right] \tag{2.2}$$

2.2 Performance Measures and Their Tradeoffs

From (2.2), at high signal-to-noise ratio (SNR), the ergodic capacity becomes [17]

$$C\left(\frac{P}{\sigma_z^2}\right) = \min(N_t, N_r)\log\left(\frac{P}{N_t\sigma_z^2}\right) + \sum_{i=|N_t-N_r|+1}^{\max(N_t,N_r)} E(\log\chi_{2i}^2) + o(1) \qquad (2.3)$$

where χ_{2i}^2 is chi-square distributed with $2i$ degree of freedom. From (2.3), we observe that at high SNR, a MIMO channel can offer $\min(N_t, N_r)$ times of capacity compared to a single-antenna channel. In addition, multiple antennas can also be used to improve the communication reliability. A "good" MIMO transmission scheme must exploit the available temporal and spatial resources as much as possible and provide a flexible tradeoff between the reliability and data rate.

2.2.1 Tradeoff Between Diversity and Multiplexing Gain

The increase of data rate in a MIMO transmission scheme can be characterized by multiplexing gain. A MIMO transmission scheme is said to achieve a multiplexing gain r if, at high SNR ρ, its transmission rate $R \approx r\log\rho$, which is r times of the capacity of a single-input single-output channel with the same SNR ρ.

The improvement in reliability of a MIMO transmission scheme is often characterized by diversity gain. There exist two definitions for diversity gain.

Zheng's spatial diversity Zheng's spatial diversity is defined as the exponent of the error performance of a scheme when the data rate is fixed against the channel capacity [17]. Denote $P_e(\rho)$ as the error probability with respect to SNR ρ of a scheme whose data rate is $R = r\log\rho$. Then Zheng's spatial diversity can be found in the asymptotic behavior, i.e.,

$$\lim_{\rho\to\infty}\frac{\log P_e(\rho)}{\log\rho} = -d(r) \qquad (2.4)$$

where $d(r)$ is the diversity order.

Tarokh's diversity Tarokh's diversity is more widely used definition, which is the exponent of the error performance of a scheme when the data rate is constant [4].

The first definition is useful when discussing the tradeoff between data rate and reliability; whereas the second definition is useful when discussing the error performance of a scheme with a fixed data rate.

It was shown in [17] that there exists a fundamental tradeoff between how much the multiplexing gain and Zheng's diversity gain can be achieved simultaneously. Specifically, a MIMO system with N_t transmit and N_r receive antennas can provide the maximum diversity gain $d_{max} = N_t N_r$ or the maximum spatial multiplexing gain $r_{max} = \min(N_t, N_r)$ but not both at the same time. Any type of gain comes at the penalty of another [32]. Therefore, in designing a MIMO transmission scheme, flexible tradeoff is important particularly for multi-rate wireless communications.

2.2.2 Error Performance and Design Criteria for Fixed Data Rate

Pairwise error probability is often used to analyze the error performance of a MIMO transmission scheme with a given data rate. The pairwise error probability between a codeword pair $(\mathbf{X}, \hat{\mathbf{X}})$ is the probability that a decoder decides in favor of $\hat{\mathbf{X}}$ when in fact $\mathbf{X}$ is transmitted [4].

Let $x_i(n)$ be the symbol to be transmitted at time n over transmit antenna i,

and collect a block of transmitted symbols (a codeword) as

$$\mathbf{X} = \begin{pmatrix} x_1(1) & x_1(2) & \dots & x_1(L) \\ x_2(1) & x_2(2) & \dots & x_2(L) \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_t}(1) & x_{N_t}(2) & \dots & x_{N_t}(L) \end{pmatrix} \tag{2.5}$$

Then with ML decoding, the upper bound of the pairwise error probability of $(\mathbf{X}, \hat{\mathbf{X}})$ in the case of Rayleigh fading, is given by [4]

$$P_r(\mathbf{X}, \hat{\mathbf{X}}) \leq \left(\prod_{i=1}^{N_t} \frac{1}{1 + \frac{E_s}{4N_0}\lambda_i} \right)^{N_r} \tag{2.6}$$

where E_s is the average symbol energy, N_0 is the noise power spectral density, and λ_i is the i-th non-zero eigenvalue of matrix $\mathbf{A} = (\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^H$. At high SNR, the above upper bound can be simplified as

$$P_r(\mathbf{X}, \hat{\mathbf{X}}) \leq \left(\prod_{i=1}^{r} \lambda_i \right)^{-N_r} \left(\frac{E_s}{4N_0} \right)^{-rN_r} \tag{2.7}$$

where r is the rank of matrix $\mathbf{A}$. It can be seen from (2.7) that the pairwise error probability decreases exponentially as SNR increases. The exponent rN_r in the error probability of (2.7) determines the slope of the error probability curve as SNR increases and is called Tarokh's diversity gain. It can also be observed that a coding gain of $\left(\prod_{i=1}^{r} \lambda_i \right)^{\frac{1}{r}}$ is achieved.

Based on the above observations, two design criteria were obtained by Tarokh *et al.* in [4]:

The Rank Criterion: In order to achieve the maximum diversity $N_t N_r$, the difference matrix $(\mathbf{X} - \hat{\mathbf{X}})$ has to be full rank for all pairs of $\mathbf{X}$ and $\hat{\mathbf{X}}$. If $(\mathbf{X} - \hat{\mathbf{X}})$ has minimum rank r over the set of two pairs of distinct codewords, then a diversity

of rN_r is achieved.

The Determinant Criterion: Suppose that a diversity gain of rN_r is our target, then the design goal is making the minimum product $\left(\prod_{i=1}^{r} \lambda_i\right)$ as large as possible over all pairs of distinct codewords $\mathbf{X}$ and $\hat{\mathbf{X}}$. If a diversity gain of $N_t N_r$ is the design target, then the minimum product $\prod_{i=1}^{r} \lambda_i$ must be maximized over all pairs of distinct codewords. Note that the coding gain $\prod_{i=1}^{r} \lambda_i$ is the absolute value of the sum of determinants of all $r \times r$ principal co-factors of $\mathbf{A}$ taken over all pairs of codewords $\mathbf{X}$ and $\hat{\mathbf{X}}$.

2.3 Linear Dispersion Code

At very high data rate and with a large number of antennas, many of existing space-time codes, such as space-time trellis codes [4] and space-time block codes [5][6], suffer from complexity or performance difficulties. For instance, the number of states in the trellis codes of [4] grows exponentially with either the rate or the number of transmit antennas. The block codes of [5][6] suffer from rate and error performance loss as the number of antennas grows. As such, many linear dispersion (LD) codes are proposed [14][15][16], which possess many of the coding and diversity advantages with simple decoding and flexible linear structure.

The LD codes break the data stream into sub-streams, each sub-stream is dispersed over space and time and then the sub-streams are combined linearly. Suppose that there are N_t transmit antennas, and N_r receiver antennas, and T symbol intervals during which the channel is constant and known to the receiver. The transmitted signal over N_t transmit antennas during the T symbol intervals can be written as a $N_t \times T$ matrix $\mathbf{X}$. The data stream is broken into L sub-streams and $x_1, \ldots, x_L$ are the complex symbols chosen from a complex constellation of size 2^Q, such as QPSK and 16QAM. The transmitted signal $\mathbf{X}$ of the linear dispersion code with a rate

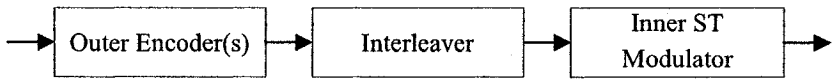


Figure 2.2: System diagram of the serial concatenated transmitter.

$R = L/T \cdot Q$ is

$$\mathbf{X} = \sum_{q=1}^{L} x_q \mathbf{M}_q \tag{2.8}$$

where $\mathbf{M}_q$ is the $N_t \times T$ complex dispersion matrix associated with symbol x_q.

The design of LD codes depends crucially on the choice of the paraments T, L and the dispersion matrices $\{\mathbf{M}_q\}$. Nonlinear information-theoretic criterion is applied to optimize the choice of $\{\mathbf{M}_q\}$. Namely, the mutual information between the transmitted symbols $\{x_q\}$ and the received signal. The criterion is very important for achieving high data rate with multiple antennas because of the limitation of linearity and often small T for reasonable complexity. LD codes, however, are not capable of providing large coding gain. Hence, LD codes are rate-oriented without fully considering the error performance.

The above features of linear dispersion coding make it the preferred choice as the inner space-time modulator of a serial concatenated MIMO transmitter as shown in Fig. 2.2. In such a transmitter, the inner LD code provides spatial multiplexing gain with guaranteed spatial diversity while the outer encoder provides large coding gain and temporal diversity.

2.4 Turbo Principle and Iterative Receiver

To describe the "Turbo principle", we use the serial concatenated transmission system as shown in Fig. 2.2.

The turbo principle was originally proposed for decoding parallel concatenated turbo codes [19] but now has been successfully applied in joint detection/decoding

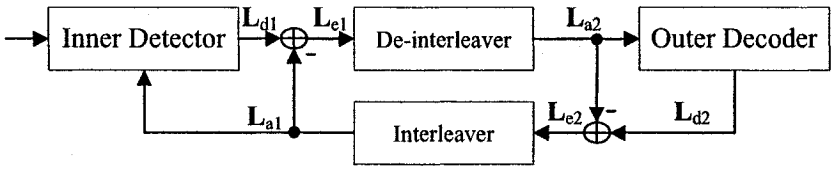


Figure 2.3: System diagram of the turbo receiver.

and other areas. Basically, if the transmitter of a communication system can be modelled as the combination of several components (arranged in serial or parallel fashion) with random interleaver in between, e.g. the system in Fig. 2.2, the turbo principle is applicable as shown in Fig. 2.3. In this figure and following discussions, variables with subscript "1" are associated with the inner detector and variables with subscript "2" are associated with the outer decoder, and subscripts "a (or A)", "e (or E)" and "d (or D)" stand for *a priori*, *extrinsic*, and *a posteriori*, respectively.

In Fig. 2.3, soft information in the form of confidence levels (*extrinsic* information) is exchanged between the SISO inner detector and the SISO outer decoder in an iterative fashion. In each iteration, the soft inner detector calculates the *a posteriori* information based on the channel observation (*intrinsic* information) and the *a priori* information generated by permuting the *extrinsic* information from the soft outer decoder during the last iteration. After cancelling the input *a priori* information, the *a posteriori* information becomes the *extrinsic* information of the soft inner ST detector, which will be de-interleaved and then used as the *a priori* information for the soft outer decoder. The outer decoder calculates the *a posteriori* information, which after cancelling the *a priori* information, becomes the *extrinsic* information for next iteration. After a sufficient number of iterations, neither the inner ST detector nor the outer decoder can benefit from the exchange of the *extrinsic* information any longer and this phenomenon is often called "convergence". Once convergence is reached, the outer decoder will perform hard decision on the *a posteriori* information to generate the decoded information bits. Normally, the exchanged information is

often quantified in the form of log-likelihood ratios (LLRs). The bit LLRs can be calculated as

$$L(b) = \log\left[P_r(b=1)/P_r(b=0)\right] \tag{2.9}$$

Below, we briefly introduce the turbo receiver for the concatenated turbo MIMO system. For the inner detector, the *extrinsic* information is gleaned from the received signals (*intrinsic* information) and the *a priori* information fed back from the outer decoder. Various MUDs can be used for the inner detection. The optimal detector calculates exactly the *a posteriori* LLR for each bit and then generate *extrinsic* LLR by cancelling the *a priori* LLR.

We assume that the channel observation $\mathbf{y}$ is given by equation (2.1), each transmitted symbol x_i in $\mathbf{x}$ is taken from a complex constellation of size 2^Q, such as 2^Q-QAM, i.e., $x_i \in \Omega \equiv \{\Omega_m | m = 0, 1, \ldots, 2^Q - 1, Q \geq 1\}$, and the *a priori* LLRs associated with bits in the transmitted symbol vector $\mathbf{x}$ is $\mathbf{L}_{a1} = [L_{a1,1}, L_{a1,2}, \ldots, L_{a1,N_tQ}]$. Then, the optimal SISO MUD calculates the *a posteriori* LLR of the j-th bit in symbol x_i as

$$L_{d1,(i-1)Q+j} = \log \frac{P_r(b_{(i-1)Q+j} = 1 \mid \mathbf{y}, \mathbf{L}_{a1})}{P_r(b_{(i-1)Q+j} = 0 \mid \mathbf{y}, \mathbf{L}_{a1})} \tag{2.10}$$

Denote $J_{\mathbf{x}}$ as the non-zero indices of bits $\{b_k\}$ associated with the transmitted symbol vector $\mathbf{x}$, i.e., $J_{\mathbf{x}} \doteq \{k | b_k \neq 0, 1 \leq k \leq N_tQ\}$. Denote $\mathbf{X}^+$ as the set of 2^{N_tQ-1} symbol vectors with the j-th bit in symbol x_i equal to 1, i.e., $b_{(i-1)Q+j} = 1$, while $\mathbf{X}^-$ as the set of 2^{N_tQ-1} symbol vectors with the j-th bit in symbol x_i equal to 0, i.e., $b_{(i-1)Q+j} = 0$. Then, (2.10) can be further written as

$$L_{d1,(i-1)Q+j} = \log \frac{\sum\limits_{\mathbf{x}\in\mathbf{X}^+} \exp\left[\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_z^2}\right) + \sum\limits_{j\in J_{\mathbf{x}}} L_{a1,j}\right]}{\sum\limits_{\mathbf{x}\in\mathbf{X}^-} \exp\left[\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_z^2}\right) + \sum\limits_{j\in J_{\mathbf{x}}} L_{a1,j}\right]} \tag{2.11}$$

With the *a posteriori* LLR in (2.11), the *extrinsic* LLR can be calculated by cancelling

the corresponding *a priori* LLR, i.e.,

$$L_{e1,(i-1)Q+j} = L_{d1,(i-1)Q+j} - L_{a1,(i-1)Q+j} \tag{2.12}$$

From (2.11), the calculation of *a posteriori* LLR involves the evaluation of the Euclidean distances between all possible transmitted symbol vectors $\mathbf{x}$ and $\mathbf{y}$. Hence, the complexity grows exponentially with respect to the number of bits in $\mathbf{x}$. This is often impractical even for applications with moderate data rate. As a results, various suboptimal SISO detectors are developed. Two notable suboptimal detection schemes with much reduced complexity are described below.

SIC-MMSE Detection: The soft interference cancellation with minimum mean-square-error (SIC-MMSE) detector uses nonlinear estimation and linear minimum mean-square-error (MMSE) detection to provide soft outputs that are essentially the MMSE estimates of the symbols under given *a priori* information [27][28].

List-based Sphere Decoder: The list-based sphere decoder calculates the *a posteriori* probability using signal points in a sphere of a certain radius centered around the ML signal points [10][33]. Improved methods are available, such as [34].

For the outer code, the *a posteriori* information is computed based on the *a priori* information from the inner detection and the trellis structure of the outer code. The optimal outer decoder algorithm will be introduced in the following subsection 2.4.1 and other suboptimal outer decoder algorithms are also available, such as [35][36].

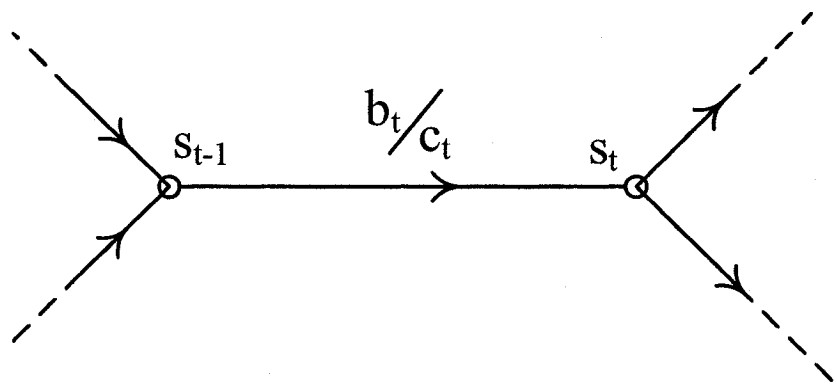


Figure 2.4: Transition process at time instant t.

2.4.1 The Optimal Outer Decoder: the BCJR/MAP Decoder

Maximum *a posteriori* (MAP) algorithm, first proposed in 1974 by Bahl, Cocke, Jelinik and Raviv [26], is also called BCJR algorithm in memory of its designers. As discussed above, the outer decoder takes the de-interleaved *extrinsic* information from the inner detector as *a priori* input to compute the *a posteriori* probability and then generate *extrinsic* information by cancelling the *a priori* information.

Here, we present a BCJR/MAP decoding algorithm for a k/n rate code with 2^m-state trellis (constraint length m). The encoding operation can be modelled as a discrete time finite-state one-order Markov process. This process can be graphically represented by state and trellis diagram as shown in Fig. 2.4. This figure shows the transition process at time instant t, which changes its state from s_{t-1} to s_t with message symbol input b_t and coded symbol output c_t. Note, message symbol b_t consists of k bits and coded symbol c_t consists of n bits.

We assume that one coding frame lasts T time instants and the encoder starts with state 0 and ends with state 0. Denote the *a priori* information of one coding frame by $\mathbf{L}_{a2}$. The *a posteriori* probability $P_r(c_t = v|\mathbf{L}_{a2})$ on the basis of *a priori* $\mathbf{L}_{a2}$

can be calculated as

$$P_r(c_t = v \mid \mathbf{L}_{a2}) = \sum_{(l',l)\in B_v} \frac{P_r(s_{t-1} = l', s_t = l, \mathbf{L}_{a2})}{P_r(\mathbf{L}_{a2})} \tag{2.13}$$

where B_v is the set of state transitions $s_{t-1} = l' \rightarrow s_t = l$ that output coded symbol v ($v \in \{0, 1, \ldots, 2^n - 1\}$). Let N_v be the number of state transitions in the set B_v. Since the probability $P_r(\mathbf{L}_{a2})$ is a common factor and is to be cancelled in calculation of LLRs, we focus on the calculation of the joint probability $P_r(s_{t-1} = l', s_t = l, \mathbf{L}_{a2})$, which will be referred to as $\sigma_t(l', l)$. Denote $\mathbf{L}_{a2}^{(t_1,t_2)}$ as the subset of *a priori* information associated with time instant from t_1 to t_2. In order to compute the joint probability $\sigma_t(l', l)$, we define the following probabilities.

Forward recursive probability

$$\alpha_t(l) = P_r(s_t = l, \mathbf{L}_{a2}^{(1,t)}) \tag{2.14}$$

Backward recursive probability

$$\beta_t(l) = P_r(\mathbf{L}_{a2}^{(t+1,T)} \mid s_t = l) \tag{2.15}$$

Branch probability

$$\gamma_t^x(l', l) = P_r(c_t = x, s_t = l, \mathbf{L}_{a2}^{(t,t)} \mid s_{t-1} = l') \tag{2.16}$$

where $x \in \{0, 1, \ldots, 2^n - 1\}$.

Now, the joint probability $\sigma_t(l', l)$ can be expressed as

$$\sigma_t(l', l) = \alpha_{t-1}(l')\beta_t(l) \sum_{x=0}^{2^n-1} \gamma_t^x(l', l) \tag{2.17}$$

The forward recursive probability $\alpha_t(l)$ defined in (2.14) can be computed recursively as

$$\alpha_t(l) = \sum_{l'=0}^{2^m} \alpha_{t-1}(l') \sum_{x=0}^{2^n-1} \gamma_t^x(l', l) \tag{2.18}$$

for $t = 1, 2, \ldots, T$.

For $t = 0$, the boundary conditions are

$$\alpha_0(l) = \begin{cases} 1 & \forall l = 0 \\ 0 & \forall l \neq 0 \end{cases} \tag{2.19}$$

The backward recursive probability $\beta_t(l)$ defined in (2.15) can be computed recursively as

$$\beta_t(l) = \sum_{l'=0}^{2^m} \beta_{t+1}(l') \sum_{x=0}^{2^n-1} \gamma_{t+1}^x(l, l') \tag{2.20}$$

for $t = T-1, \ldots, 1, 0$.

For $t = T$, the boundary conditions are

$$\beta_T(l) = \begin{cases} 1 & \forall l = 0 \\ 0 & \forall l \neq 0 \end{cases} \tag{2.21}$$

From the definition of the branch probability $\gamma_t^x(l', l)$ in (2.16), it can be obtained from the corresponding *a priori* information. From the encoder trellis, for given state transition (l', l), if the output coded symbol is v, then the corresponding branch probability at time instant t is

$$\gamma_t^x(l', l) = \begin{cases} \frac{1}{N_v} P_r^a(c_t = v) & x = v \\ 0 & x \neq v \end{cases} \tag{2.22}$$

where $P_r^a(c_t)$ is the *a priori* probability regarding coded symbol c_t, obtained from *a priori* information $\mathbf{L}_{a2}^{t,t}$, which contains n coded bits. Note, if there is no transition from state l' to l, then the corresponding branch probability $\gamma_t^x(l', l)$ is 0.

The *a posteriori* LLR associated with the i-th bit $c_{t,i}$ in coded symbol c_t can be computed as

$$L_{d2}(c_{t,i}) = \log \frac{\sum\limits_{c_t \in C_i^1} P_r(c_t \mid \mathbf{L}_{a2})}{\sum\limits_{c_t \in C_i^0} P_r(c_t \mid \mathbf{L}_{a2})} \tag{2.23}$$

where C_i^1 is the set of n-bit symbols with i-th bit equal to 1 and C_i^0 is the set of n-bit symbols with i-th bit equal to 0. The above *a posteriori* LLR can be expressed as the summation of two parts as

$$L_{d2}(c_{t,i}) = L_{e2}(c_{t,i}) + L_{a2}(c_{t,i}) \tag{2.24}$$

where $L_{a2}(c_{t,i})$ is the *a priori* LLR fed back from the inner detector and $L_{e2}(c_{t,i})$ is the *extrinsic* LLR obtained by the outer decoder.

Similarly, *a posteriori* LLRs of bits in message symbol b_t can be computed according to the encoder trellis and *a priori* information. At the final iteration, hard decision will perform on the *a posteriori* LLRs of bits contained by message symbols $\{b_t\}$.

From above equations, we see that $\alpha_t(l)$ and $\beta_t(l)$ drop toward zero exponentially. In order to obtain a numerically stable algorithm, scaling on these parameters is often applied as the computation proceeds.

Let $\tilde{\alpha}_t(l)$ denote scaled version of $\alpha_t(l)$. Initially, $\alpha_1(l)$ is computed according to equation (2.18), and we set $\hat{\alpha}_1(l) = \alpha_t(l)$, $\tilde{\alpha}_1(l) = q_1\hat{\alpha}_1(l)$ and $q_1 = 1/\sum_{l=0}^{2^m-1} \hat{\alpha}_1(l)$.

For $t \geq 2$, $\tilde{\alpha}_t(l)$ is computed as

$$\begin{aligned} \hat{\alpha}_t(l) &= \sum_{l'=0}^{2^m} \tilde{\alpha}_{t-1}(l') \sum_{x=0}^{2^n-1} \gamma_t^x(l', l) \\ \tilde{\alpha}_t(l) &= q_t \hat{\alpha}_t(l) \end{aligned} \tag{2.25}$$

with $q_t = 1/\sum_{l=0}^{2^m-1} \hat{\alpha}_t(l)$. Similarly, let $\tilde{\beta}_t(l)$ denote scaled version of $\beta_t(l)$. Initially, $\beta_{T-1}(l)$ is computed according to equation (2.20), and we set $\hat{\beta}_{T-1}(l) = \beta_{T-1}(l)$,

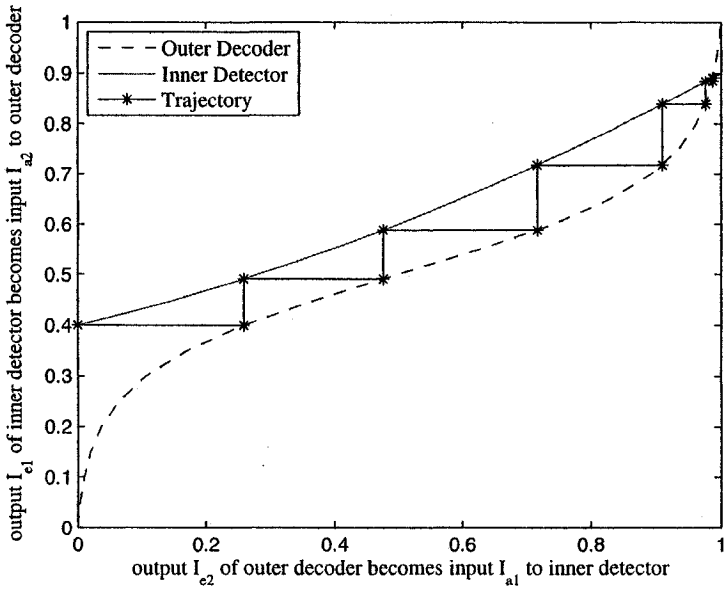


Figure 2.5: An example of the EXIT chart.

$\tilde{\beta}_{T-1}(l) = h_{T-1}\hat{\beta}_{T-1}(l)$ and $h_{T-1} = 1/\sum_{l=0}^{2^m-1}\hat{\beta}_{T-1}(l)$.

For $t \leq T-2$, $\tilde{\beta}_t(l)$ is computed as

$$\begin{aligned} \hat{\beta}_t(l) &= \sum_{l'=0}^{2^m} \tilde{\beta}_{t+1}(l') \sum_{x=0}^{2^n-1} \gamma_{t+1}^{x}(l', l) \\ \tilde{\beta}_t(l) &= h_t \hat{\beta}_t(l) \end{aligned} \tag{2.26}$$

with $h_t = 1/\sum_{l=0}^{2^m-1}\hat{\beta}_t(l)$.

2.5 Extrinsic Information Transfer (EXIT) Chart

The *extrinsic* information transfer (EXIT) chart technique, proposed by S. ten Brink [37][38], is a powerful technique for studying the convergence behavior of the turbo receiver. Without a rigorous proof, simulation results show that the EXIT chart accurately predicts the convergence behavior of a turbo receiver for *large* interleaving depth.

The idea of the EXIT chart is to study the iterative behavior by solely looking at the input/output relations of the constituent soft devices. In the EXIT chart, the exchange of *extrinsic* information between soft devices is measured in terms of *mutual information.* For a given soft device, denote the mutual information between a bit and its *a priori* LLR as *a priori* information $I_a = I(b; L_a^b)$ and similarly the mutual information between a bit and its *extrinsic* LLR as *extrinsic* information $I_e = I(b; L_e^b)$. For a well-designed detector or decoder, the *extrinsic* information I_e of one soft device in the turbo receiver is a non-decreasing function with respect to the *a priori* information I_a. This function is called EXIT function $T : I_e = T(I_a)$. In the same chart, the EXIT functions associated with the soft devices are plotted together as shown in Fig. 2.5: the *extrinsic* information of the inner detector I_{e1} becomes the *a priori* information of the outer decoder I_{a2}, while the *extrinsic* information of the outer decoder I_{e2} becomes the *a priori* information of the inner detector I_{a1}. The exchange process of *extrinsic* information is visualized as a detection/decoding trajectory in the EXIT chart. As such, the EXIT chart predicts the error-performance behaviors of the inner detection and the outer decoding, such as convergence point, iterations for convergence, etc.

Chapter 3

Design of Turbo MIMO Transmitter

3.1 Introduction

To achieve the promised capacity over MIMO channels[1][2], various transmission schemes have been proposed, such as [4]-[16]. However, these existing schemes, either performance-oriented or rate-oriented, can not fulfill the flexible requirements in performance, bandwidth efficiency, complexity, and tradeoffs among them. Additionally, they often suffer from design difficulty, performance loss and/or decoding complexity. With the applications of "turbo codes" [19] in MIMO systems, it has been realized that the turbo MIMO transmission is the most promising technique [20]-[24]., The turbo MIMO transmission system is a serial concatenation of conventional outer binary code(s) and an inner ST modulation. The turbo MIMO transmission is general and includes many existing schemes as special cases. It is flexible in providing both diversity and multiplexing gains. In this chapter, our goal is to develop turbo MIMO transmission schemes that are flexible in rate-versus-performance tradeoff and yet with reasonable decoding complexity.

The existence of well-designed conventional codes allows us to focus on the design of inner ST modulator. However, how to design the inner ST modulator is still unclear to us. To emphasize the "modulation" flavor of the inner ST process, such a concatenated system is referred to as the coded space-time modulation (CSTM). In such a CSTM system, the LD code is preferable as the inner ST modulation due to its generality, simplicity and flexibility. To decouple the correlation between inner ST modulation and outer encoding, a random interleaver is applied to the encoded bits. Hence, the corresponding concatenated system is often called MIMO bit-interleaved coded modulation (BICM) [20]-[24]. The use of interleaver makes ML decoding impractical due to the prohibitively high computational complexity. At best, a joint iterative (turbo) detection/decoding receiver based on the turbo principle can be employed with reduced computation. In such a case, Tarokh's criteria based on ML decoding apply neither to the overall concatenated system nor to the inner space-time modulation alone. Hence, it is important to study the design of the inner space-time modulator when used in a concatenated system.

Using the idea of the extrinsic information transfer (EXIT) chart pioneered by S. ten Brink [37][38], we consider the design of the inner LD space-time modulator in the CSTM system under the assumption of a turbo receiver. Although the EXIT chart technique developed for single-input single-output systems has been used in the study of some specific MIMO systems such as [22], it cannot be directly used for more general cases. Unlike in single-input single-output systems, the outputs of the inner ST modulator often have different statistics. By extending the existing EXIT chart technique to MIMO transmission systems, new design criteria are to be developed for the design of inner space-time modulation when used in a CSTM system. It is shown that the inner space-time modulator shall (a) maximize the average mutual information between a bit and the received signals and (b) minimize the pair-wise error performance of the codeword pairs that differ at only one symbol within a

modulation block. Criterion (a) is similar to those used in the design of existing high-rate schemes [14]-[16] while (b) is unique for CSTM. These two criteria concern the channel capacity and error performance, respectively, and together reflect a joint optimization of both data rate and error rate. It is worthy of noting that criterion (b) is much simpler to apply than Tarokh's criteria set since the later requires an optimization over all the possible codeword pairs.

In section 3.2, preliminaries including the system model are introduced and the joint iterative (turbo) receiver is briefly reviewed. In section 3.3, the EXIT chart technique for MIMO transmission system is introduced and two observations concerning the iterative receiver are provided. In section 3.4, new design criteria are proposed and an optimal LD ST modulator is provided. Several design examples are provided to demonstrate the merits of the proposed CSTM and to verify the proposed design criteria as well. In addition, the tradeoff between constellation size and symbol rate for a given data rate is discussed. In section 3.5, the associated simulation results are provided to demonstrate the merits of our design. Finally, conclusions are drawn in section 3.6.

3.2 Preliminaries

In this section, some background concepts pertaining to a serial concatenated MIMO system are presented. As mentioned before, we consider a serially concatenated MIMO transmission system with LDCs as the inner ST modulator under the assumption of a suboptimal but powerful iterative (turbo) receiver. The system model is given in section 3.2.1 and the joint iterative (turbo) receiver is briefly reviewed in section 3.2.2.

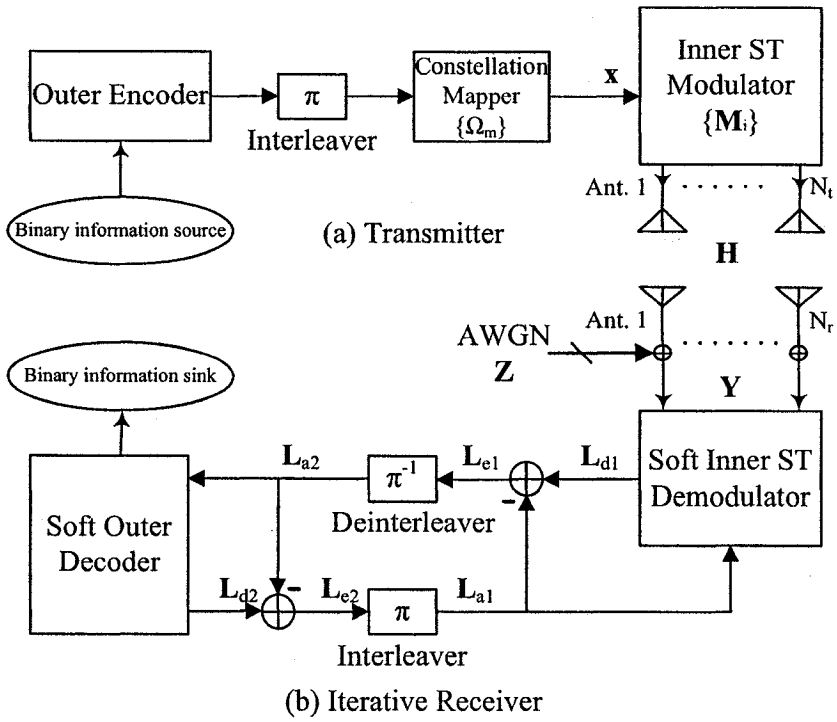


Figure 3.1: System block diagram for CSTM.

3.2.1 System Model

In this study, a block fading channel model is assumed where the channel keeps constant in one modulation block but may change from block to block. That is, the channel is not necessarily constant within a coding frame which often consists of a large number of modulation blocks. Furthermore, the channel is assumed to be a Rayleigh flat fading channel with N_t transmit and N_r receive antennas. Let's denote the complex gain from transmit antenna n to receiver antenna m by h_{mn} and collect them to form an $N_r \times N_t$ channel matrix $\mathbf{H} = [h_{mn}]$, known perfectly to the receiver but unknown to the transmitter. The entries in $\mathbf{H}$ are assumed to be *i.i.d.* and symmetrical complex Gaussian random variables with zero mean and unit variance.

The CSTM under investigation is a serial concatenation of an outer encoder and an inner ST modulator as shown in Figure 3.1(a), which subsumes many MIMO transmission schemes as its special cases. In a CSTM system, the information bits are

first encoded, shuffled by an interleaver and then mapped into symbols. After that, the symbol stream is parsed into blocks of length L. The symbol vector associated with one modulation block is denoted by $\mathbf{x} = [x_1, x_2, \ldots, x_L]^T$ with $x_i \in \Omega \equiv \{\Omega_m | m = 0, 1, \ldots, 2^Q - 1, Q \geq 1\}$ (i.e., a complex constellation of size 2^Q, such as 2^Q-QAM). The average symbol energy is assumed to be 1, i.e., $\frac{1}{2^Q} \sum_{m=0}^{2^Q-1} |\Omega_m|^2 = 1$. Each block of symbols will be mapped by the inner ST modulator to a dispersion matrix of size $N_t \times T$ and then transmitted over the N_t transmit antennas over T channel uses. The system model in Figure 3.1(a) is often called bit-interleaved coded modulation (BICM) [20]-[24].

As mentioned before, we will consider LD ST modulation for various reasons. An LD ST modulator is defined by its L $N_t \times T$ dispersion matrices $\mathbf{M}_i = [\mathbf{m}_{i1}, \mathbf{m}_{i2}, \ldots, \mathbf{m}_{iT}]$ and the corresponding output matrix of one modulation block is given by

$$\mathbf{X} = \sum_{i=1}^{L} \mathbf{M}_i x_i \tag{3.1}$$

With a constellation of size 2^Q, the data rate of the inner space-time modulator is $R_m = Q \cdot L/T$ bits per channel use and the data rate of the overall concatenated system is $R = R_c R_m$ bits per channel use, where $R_c \leq 1$ is the coding rate of the outer encoder.

Hence, one can adjust symbol rate L/T, constellation size Q, and coding rate R_c to meet different requirements on data rate and error performance. Since the inner ST modulation is linear, suboptimal linear receivers can be used for detection [14]. It can also be observed that the space-time mapping schemes used in the existing layered space-time architectures, e.g., [11][13], are LD modulation. Hence the proposed CSTM with LD ST modulation subsumes existing layered space-time schemes as special cases.

At the receiver, the received signals associated with one modulation block can

be written as

$$\mathbf{Y} = \sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{X} + \mathbf{Z} = \sqrt{\frac{P}{N_t}}\mathbf{H}\sum_{i=1}^{L}\mathbf{M}_i x_i + \mathbf{Z} \tag{3.2}$$

where $\mathbf{Y}$ is a complex matrix of size $Nr \times T$ whose (m,n)-th entry is the received signal at receive antenna m and time instant n, $\mathbf{Z}$ is the additive white Gaussian noise (AWGN) matrix with *i.i.d.* symmetrical complex Gaussian elements with zero mean and variance σ_z^2, and P is the average energy per channel use at each receive antenna. Let vec() be the operator that forms a column vector by stacking the columns of a matrix and define $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{z} = \text{vec}(\mathbf{Z})$, and $\mathbf{m}_i = \text{vec}(\mathbf{M}_i)$, then (3.2) can be rewritten as

$$\mathbf{y} = \sqrt{\frac{P}{N_t}}\overline{\mathbf{H}}\mathbf{G}\mathbf{x} + \mathbf{z} = \sqrt{\frac{P}{N_t}}\widetilde{\mathbf{H}}\mathbf{x} + \mathbf{z} \tag{3.3}$$

where $\overline{\mathbf{H}} = \mathbf{I}_T \otimes \mathbf{H}$ with $\otimes$ as the Kronecker product operator and $\mathbf{G} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_L]$ will be referred to as the modulation matrix. Since the average energy of the signal per channel use at a receive antenna is assumed to be P, we have $\text{tr}(\mathbf{G}\mathbf{G}^H) = N_t T$. Denoting $\tilde{\mathbf{h}}_i = \overline{\mathbf{H}}\mathbf{m}_i$ as the i-th column vector of $\widetilde{\mathbf{H}}$, the above equation can also be written as

$$\mathbf{y} = \sqrt{\frac{P}{N_t}}\sum_{i=1}^{L}\tilde{\mathbf{h}}_i\mathbf{x}_i + \mathbf{z} \tag{3.4}$$

3.2.2 Joint Iterative (Turbo) Detection/Decoding Receiver

The computational complexity of an optimal MAP/ML-based receiver can be impractical for a CSTM system due to the use of the interleaver between the inner ST modulator and the outer encoder. At best, an iterative joint decoding and detection receiver, as illustrated in Figure 3.1(b), can be employed, which has been described in Chapter 2.

Normally, log-likelihood-ratios (LLRs) are used in information exchange. The bit LLRs can be calculated as $L^b = \log\frac{P_r(b=1)}{P_r(b=0)}$. Below, we consider the *extrinsic* LLRs

for the inner ST demodulator and the outer decoder, respectively.

1) The *extrinsic* LLR of the inner ST demodulator

Given the *a priori* bit LLRs $\mathbf{L}_{a1}^{b} = [L_{a1,1}^{b}, L_{a1,2}^{b}, \ldots, L_{a1,LQ}^{b}]$ from the outer decoder at the last iteration and the corresponding channel observation $\mathbf{y}$, using MAP criterion, the *extrinsic* LLR of the j-th bit in symbol x_i can be calculated as

$$L_{e1,(i-1)Q+j}^{b} = \underbrace{\log \frac{P_r(b_{(i-1)Q+j} = 1|\mathbf{y}, \mathbf{L}_{a1}^{b})}{P_r(b_{(i-1)Q+j} = 0|\mathbf{y}, \mathbf{L}_{a1}^{b})}}_{L_{d1,(i-1)Q+j}^{b}} - L_{a1,(i-1)Q+j}^{b} \tag{3.5}$$

This demodulator is also referred to as APP demodulator.

For future discussion, we also present the computation of LLRs on symbols as well. Since symbols are taken from a constellation of size 2^Q, each symbol has a $(2^Q - 1)$-tuple whose m-th element is the logarithm of the ratio of the probability of the symbol taking value Ω_m over that of the symbol taking value Ω_0, i.e., $L^s(x = \Omega_m) = \log \frac{P_r(x=\Omega_m)}{P_r(x=\Omega_0)}$.

Given the *a priori* symbol LLRs $\mathbf{L}_{a1}^{s} = [L_{a1,1}^{s}, L_{a1,2}^{s}, \ldots, L_{a1,L}^{s}]$ and the channel observation $\mathbf{y}$, one can compute the *extrinsic* LLR of symbol x_i in the block using the MAP criterion as

$$\begin{aligned} L_{e1,i}^{s}(\Omega_m) &= \underbrace{\log \frac{P_r(x_i = \Omega_m|\mathbf{y}, \mathbf{L}_{a1}^{s})}{P_r(x_i = \Omega_0|\mathbf{y}, \mathbf{L}_{a1}^{s})}}_{L_{d1,i}^{s}(\Omega_m)} - L_{a1,i}^{s}(\Omega_m) \\ &= \log \frac{\sum_{\mathbf{x}_I \in \Gamma} \exp\left[\left(-\frac{\|\mathbf{y} - \sqrt{\frac{P}{N_t}}\widetilde{\mathbf{H}}_I \mathbf{x}_I - \sqrt{\frac{P}{N_t}}\tilde{\mathbf{h}}_i \Omega_m\|^2}{\sigma_z^2}\right) + \left(\sum_{k \in \Xi} L_{a1,k}^{s}(x_k)\right)\right]}{\sum_{\mathbf{x}_I \in \Gamma} \exp\left[\left(-\frac{\|\mathbf{y} - \sqrt{\frac{P}{N_t}}\widetilde{\mathbf{H}}_I \mathbf{x}_I - \sqrt{\frac{P}{N_t}}\tilde{\mathbf{h}}_i \Omega_0\|^2}{\sigma_z^2}\right) + \left(\sum_{k \in \Xi} L_{a1,k}^{s}(x_k)\right)\right]} \end{aligned} \tag{3.6}$$

where $\widetilde{\mathbf{H}}_I$ is the matrix by removing the i-th column from $\widetilde{\mathbf{H}}$ in (3.3), $\mathbf{x}_I = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_L$ Γ is the set of $2^{Q(L-1)}$ column symbol vectors and Ξ is the set of indices of $\mathbf{x}_I$, i.e.

$\Xi = \{k | 1 \leq k \leq L \text{ and } k \neq i\}$.

Note, the *a priori* symbol LLRs $L^s_{a1,i}$ are calculated from bit LLRs $L^b_{a1,i}$ from the outer decoder, i.e.,

$$L^s_{a1,i}(\Omega_m) = \sum_{j \in \Upsilon_m} L^b_{a1,j} \tag{3.7}$$

where Υ_m is the set of indices of the bits in Ω_m equal to 1. Then from (3.5) and (3.6), the relationship between the bit LLRs and the corresponding symbol LLRs can be found as

$$L^b_{e1,(i-1)Q+j} = \underbrace{\log \frac{\sum_{\Omega^+_m \in \Omega_{b_j=1}} \exp\left[L^s_{e1,i}(\Omega^+_m) + L^s_{a1,i}(\Omega^+_m)\right]}{\sum_{\Omega^-_m \in \Omega_{b_j=0}} \exp\left[L^s_{e1,i}(\Omega^-_m) + L^s_{a1,i}(\Omega^-_m)\right]}}_{L^b_{d1,(i-1)Q+j}} - L^b_{a1,(i-1)Q+j} \tag{3.8}$$

where $\Omega_{b_j=1}$ is the set of constellation points whose j-th bit in Ω_m equal to 1 while $\Omega_{b_j=0}$ is the set of constellation points whose j-th bit equal to 0. From (3.8), the APP inner ST demodulator can be implemented by adding a so-called de-mapper following the symbol-level APP demodulator as defined in (3.6). This process is shown in Fig. 3.2. This de-mapper calculates the *extrinsic* bit LLR using the output *extrinsic* LLR of the corresponding symbol from the symbol-level demodulator and the *a priori* LLRs of the bits in that symbol. This observation will be useful in the later development of this study. In addition to the optimal APP algorithm, other linear suboptimal methods are also available for reduced complexity such as [27][28].

2) The *extrinsic* LLR of the outer decoder

Unlike the inner ST modulator, the computation of *extrinsic* LLRs is only based on the input *a priori* LLRs from the inner ST demodulator in such a CSTM

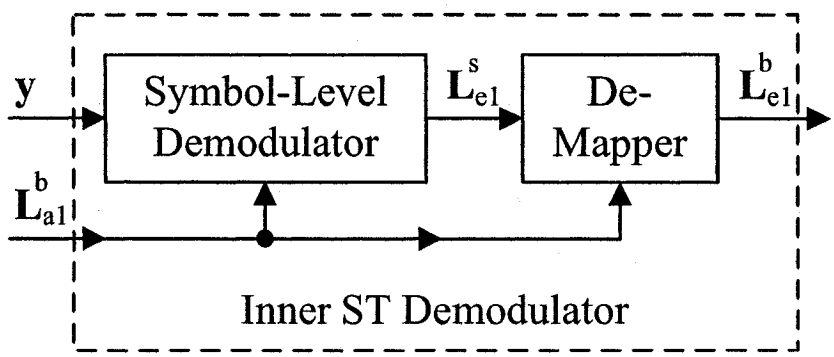


Figure 3.2: System block diagram for Inner ST demodulator.

system. The *extrinsic* bit LLRs can be computed as

$$L_{e2,i}^{b} = \underbrace{\log \frac{P_r(b_i = 1|\mathbf{L}_{a2}^{b})}{P_r(b_i = 0|\mathbf{L}_{a2}^{b})}}_{L_{d2,i}^{b}} - L_{a2,i}^{b} \tag{3.9}$$

where $\mathbf{L}_{a2}^{b}$ is the given *a priori* bit LLRs. The associated optimal MAP algorithm has been developed by Bahl *et al* in [26], which has been described in Chapter 2.

3.3 EXIT Chart For MIMO Transmission

In this section, by means of the EXIT chart, we examine the convergence behavior of the iterative detection/decoding procedure for a CSTM system as described in the last section.

In Figure 3.3, an EXIT chart of a CSTM for given channel $\mathbf{H}$ is illustrated as an example. In this CSTM, each LD modulation block has two input BPSK symbols corresponding to two bits, b_1 and b_2. The exchange of *extrinsic* information is measured as the mutual information between the LLRs and the corresponding bits as described in Section 2.5. For instance, the *extrinsic* information of the outer decoder is $I_{e2} = I(b; L_{e2})$.

It is important to notice that the *extrinsic* LLRs from the inner ST demodula-

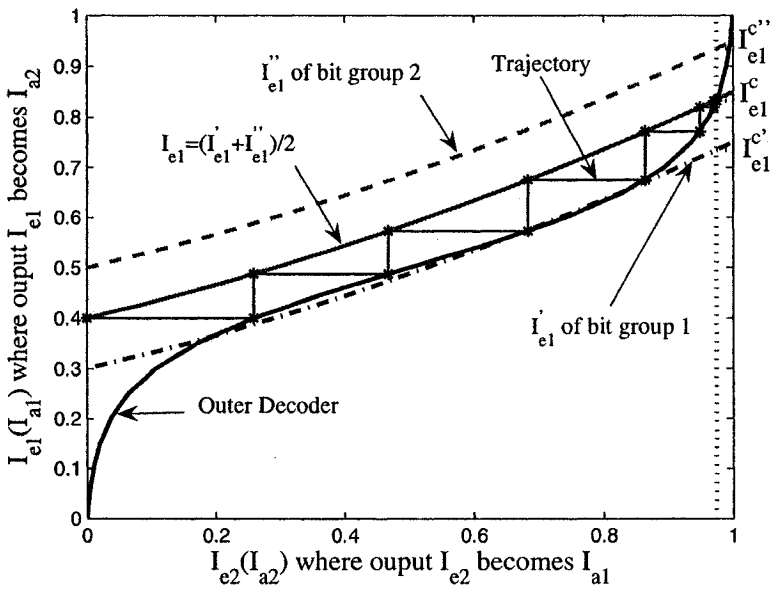


Figure 3.3: An example EXIT chart of turbo receiver.

tor may have *different* statistics, depending on the channel $\mathbf{H}$, the dispersion matrices $\{\mathbf{M}_i\}$ as shown in (3.1) and constellation pattern as well. Hence, different bits may have different EXIT functions. In Fig. 3.3, $I_{e1}^{'}$ and $I_{e1}^{''}$ corresponds to b_1 and b_2, respectively, while I_{e1} is their average, i.e., $I_{e1} = \left(I_{e1}^{'} + I_{e1}^{''}\right)/2$. As such, the EXIT of the outer decoder I_{e2} is a function of the two *a priori* inputs, i.e., $I_{e2} = T_{e2}(I_{e1}^{'}, I_{e1}^{''})$. In the later discussion, it will be shown that I_{e2} solely depends on the average *a priori* inputs and thus has only one curve, i.e., $I_{e2} = T_{e2}(I_{e1})$. Furthermore, the illustrated trajectory of the *extrinsic* information exchange happens between the EXIT curves of I_{e1} and I_{e2}.

The following observations are useful for the analysis of the MIMO transmission system. In the other literatures, such as [22], often apply these observations impliedly.

Observation 1: If the interleaver is large and random and the constraint length is sufficiently large, the outer decoder yields the same amount of *extrinsic* information for all the bits and the exact amount depends on the average *a priori* information of the inputs.

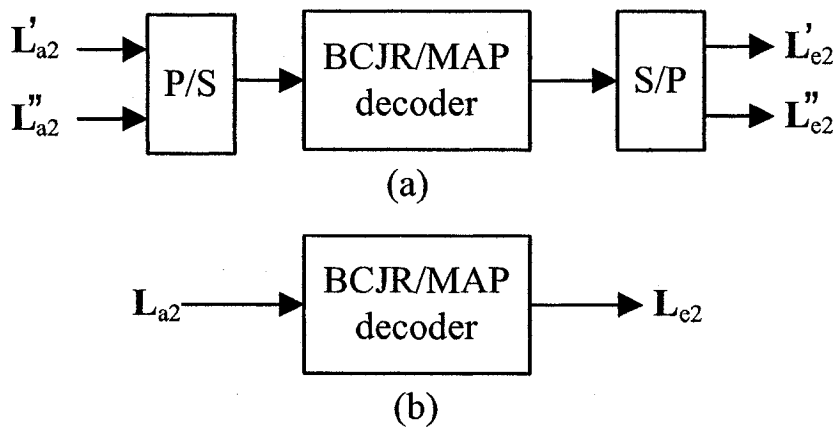


Figure 3.4: BCJR/MAP decoding with different *a priori* LLR inputs.

Remarks: The above observation tells that I_{e2} in the example shown in Figure 3.3 is the same for all the *extrinsic* outputs and is a function of $I_{e1} = (I'_{e1} + I''_{e1})/2$. This is caused by the following reasons. If the constraint length is large, the output *extrinsic* bit LLRs will depend, of about equal degree, on the *a priori* information of a large number of its neighbors. Furthermore, if the interleaver is large and random, these neighbors are evenly distributed among the input groups. Hence, different bits tend to have the same amount of *extrinsic* information. In addition, a specific neighbor (e.g., the bit immediately next to the bit of concern at the decoder output) may be associated with any of the input groups with equal probability. Actually, the trellis structure of the outer code averages the *a priori* inputs and thus outputs *i.i.d. extrinsic* LLRs. Consequently, the *a priori* information of any neighbor is the average *a priori* information of all the input groups and, hence, in Figure 3.3, $I_{e2} = T_{e2}\left((I'_{e1} + I''_{e1})/2\right)$.

To verify the above hypothesis, two BCJR/MAP decoders as depicted in Figure 3.4 were set up. For the decoder (a) in Figure 3.4(a), two streams of independent *a priori* bit LLRs with mutual information $I(b; L'_{a2})$ and $I(b; L''_{a2})$, respectively, were sent to a BCJR/MAP decoder and then two corresponding streams of *extrinsic* LLRs were calculated by the decoder. The BCJR/MAP decoder (b) in Figure 3.4(b) only

had one stream of *a priori* LLRs with mutual information $I(b; L_{a2})$ equal to the average *a priori* information of bit LLRs in BCJR/MAP decoder (a), i.e.,

$$I(b; L_{a2}) = \frac{1}{2}\left[I(b; L'_{a2}) + I(b; L''_{a2})\right] \tag{3.10}$$

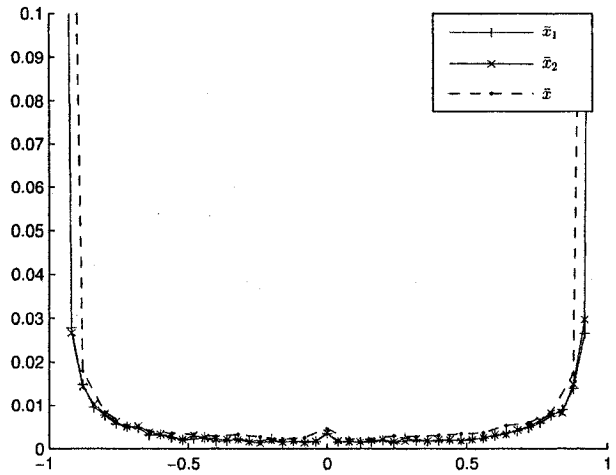
Since there is a one-to-one correspondence between a bit LLR L and the corresponding soft output $\bar{x}$ measured as

$$\bar{x} = \frac{\exp(L) - 1}{\exp(L) + 1} \tag{3.11}$$

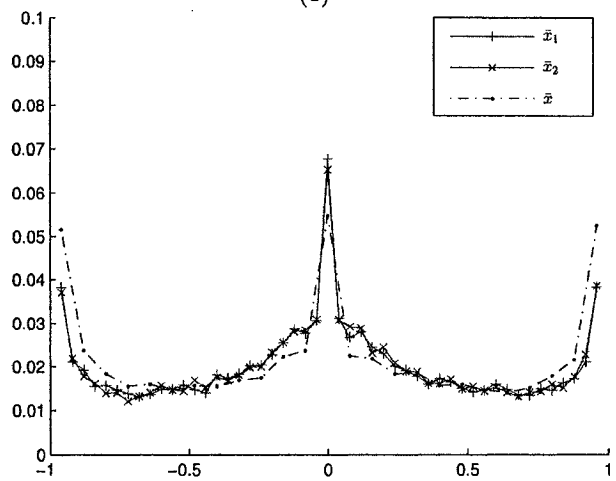
mutual information between the transmitted bit and the corresponding LLR is equal to mutual information between the bit and the corresponding $\bar{x}$, i.e., $I(b; L_e) = I(b; \bar{x}(L_e))$. Hence, instead of using *extrinsic* bit LLRs directly, histograms of soft output $\bar{x}$ were generated to reflect the statistics nature of corresponding *extrinsic* LLRs for convenience. In the simulation, a convolutional code with coding rate 1/2 and constraint length 5 were used. Following the Gaussian assumption in [38], the *a priori* bit LLRs were generated from BPSK symbols corrupted by zero-mean AWGN with different variances. The simulation results are shown in Figure 3.5. As can be seen from the figure, the histograms of the two *extrinsic* soft output streams (i.e., $\bar{x}_1$ and $\bar{x}_2$) corresponding to the two input LLR streams (i.e., L'_{e2} and L''_{e2}) for decoder (a) coincide with each other and they also coincide with the histogram of the *extrinsic* soft output $\bar{x}$ for decoder (b). In summary, the simulation results suggest that the *extrinsic* information output of the outer BCJR/MAP decoder only depends on the *average a priori* information input.

Observation 2: (a) The convergence point is where the *average* EXIT curve of the inner ST demodulator and the EXIT curve of the outer decoder meet. (b) The *a posteriori* LLRs associated with different bits in a modulation block sent to the final decision device may have different statistics determined by their own EXIT curves.

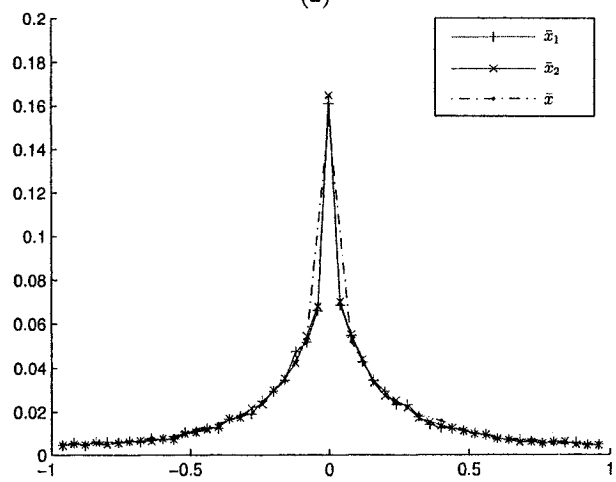
Remarks: This observation is just a direct result from Observation 1. For



(1)



(2)



(3)

Figure 3.5: Histograms of $\bar{x}$ for the BCJR/MAP decoder (a) and (b).
Note: "$\bar{x}_1$", "$\bar{x}_2$" and "$\bar{x}$" are the soft outputs corresponding to $L_e^{'}$, $L_e^{''}$ in BCJR/MAP decoder (a) and L_e in BCJR/MAP decoder (b), respectively.

instance, at convergence in Figure 3.3, let I_{e2}^{c} denote the value of the *extrinsic* information of the outer decoder, $I_{e1}^{c'}$ and $I_{e1}^{c''}$ the values of *extrinsic* information corresponding to the two bit groups, b_1 and b_2, of the inner demodulator, we have the following relationship

$$\begin{aligned} I_{e2}^{c} &= T_{e2}((I_{e1}^{c'} + I_{e1}^{c''})/2) \\ I_{e1}^{c'} &= T_{e1}'(I_{e2}^{c}) \\ I_{e1}^{c''} &= T_{e1}''(I_{e2}^{c}) \end{aligned} \tag{3.12}$$

where $I = T(x)$ indicates that I is a function of x. Denote *a posteriori* LLRs associated with b_1 and b_2 at convergence as

$$\begin{aligned} L_{d2}' &= L_{e2}^{c'} + L_{e1}^{c'} \\ L_{d2}'' &= L_{e2}^{c''} + L_{e1}^{c''} \end{aligned} \tag{3.13}$$

where $L_{e1}^{c'}$ and $L_{e1}^{c''}$ correspond to $I_{e1}^{c'}$ and $I_{e1}^{c''}$ respectively, and both $L_{e2}^{c'}$ and $L_{e2}^{c''}$ correspond to I_{e2}^{c}. In (3.13), $L_{e2}^{c'}$ and $L_{e2}^{c''}$ have the *same* statistics as predicted by Observation 1, but $L_{e1}^{c'}$ and $L_{e1}^{c''}$ may have *different* statistics. Hence, if I_{e2}^{c} is sufficiently large, the bit error rate will be mainly determined by the *extrinsic* information of the individual bits from the inner ST demodulator. In other words, the "weakest" *extrinsic* LLRs from the inner ST detection will dominate the error rate.

Additionally, two turbo schemes as described in Figure 3.6 were set up to verify the above observation. A channel is said to be a "slow" fading channel if it keeps constant in a coding frame but changes from frame to frame independently. In the simulation, a slow Rayleigh fading channel with $N_t = N_r = 2$ was assumed; 200 QPSK symbols per coding frame and outer encoders with coding rate of 1/2 were applied. The corresponding frame error rate (FER) comparison of the two schemes

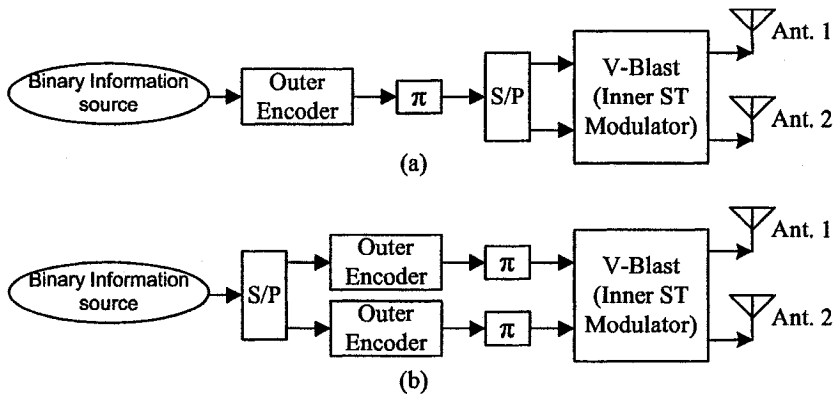


Figure 3.6: Two turbo schemes with V-Blast as the inner ST modulator.
Note: All the outer encoders in the two schemes are identical.

is presented in Figure 3.7. For the turbo V-Blast scheme (a), even if one of the EXIT curves of the inner ST demodulator can not meet the EXIT curve of the outer decoder at high *extrinsic* information for a given channel, e.g. the dash-dot line in Figure 3.3, it can still achieve good error performance if the *average* EXIT curve of the inner ST demodulator meets the EXIT curve of the outer decoder at high *extrinsic* information. That is the reason why the turbo V-Blast scheme (a) outperforms turbo V-Blast scheme (b) in the Figure 3.6.

To this end, we are ready to consider the design of inner ST modulation. As illustrated in Figure 3.3, from *a priori* information perspective, the EXIT curve of a powerful conventional outer code often presents two flat plateaus corresponding to small and large *a priori* information feedback at the two ends, respectively, and a sharp cliff corresponding to moderate *a priori* information feedback in the middle. This is caused by the trellis structure of the outer code. On the contrary, the EXIT curves of the inner ST demodulator are close to a straight line due to its linear structure.

By Observation 1, to let the trajectory of the iterative receiver snake through the bottleneck in the middle and thus reach the second plateau region of the I_{e2} curve, one can seek to maximize $\frac{1}{LQ}E(\sum_{i=1}^{L}\sum_{j=1}^{Q} I(b_{(i-1)Q+j}; L^{b}_{e1,(i-1)Q+j}))$, where the expectation

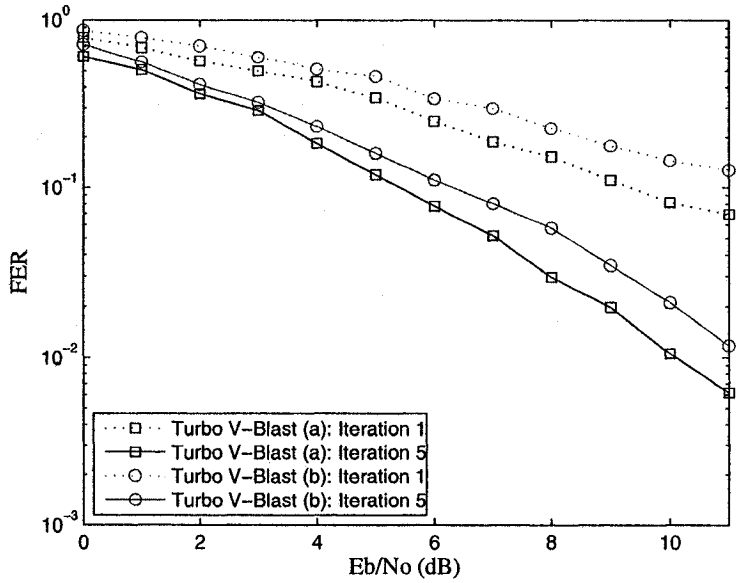


Figure 3.7: FER comparison of the two turbo V-Blast schemes under a slow Rayleigh flat fading channel.

$E()$ is taken over the channel $\mathbf{H}$. As shown in Fig. 3.2 in section 3.2.2, the *extrinsic* LLR of a bit can be calculated using the *extrinsic* LLR of its associated symbol and the *a priori* LLRs of the other bits in the same symbol. Given a constellation and *a priori* information, maximizing the sum *extrinsic* information of the bits in symbol x, i.e., $\sum_{j=1}^{Q} I(b_j; L^b_{e1,j})$, can be well approximated by maximizing the *extrinsic* information of the symbol $I(x; L^s_{e1})$. In summary, we seek to maximize $\frac{1}{LQ} E(\sum_{i=1}^{L} I(x_i; L^s_{e1,i}))$ instead of $\frac{1}{LQ} E(\sum_{i=1}^{L} \sum_{j=1}^{Q} I(b_{(i-1)Q+j}; L^b_{e1,(i-1)Q+j}))$.

By Observation 2, one should minimize the outage probability of the *a posteriori* information (i.e., $I(b; L^b_{d2})$) for each bit. Let $I_{e1}(a)$ denote the *extrinsic* information on one bit in the inner ST modulation block when the input $I_{a1} = a$ in the following discussion. From (3.13), when I^c_{e2} is large, i.e., the convergence point is located at the second plateau of the outer decoder, I^c_{e2} will change little regardless of the *a priori* information I_{e1}. Hence, minimizing the outage probability of the *a posteriori* information $I(b; L^b_{d2})$ can be approximated by minimizing the outage prob-

ability of the *extrinsic* information I_{e1} at convergence, or equivalently, maximizing $P_r(I_{e1}(I_{e2}^c) \geq \rho)$ for a certain value ρ. Noting that when the trajectory of the iterative receiver reaches the second plateau, $I_{e2}^c \approx 1$, we seek to maximize $P_r(I_{e1}(1) \geq \rho)$ for each bit. Again, for a given constellation, minimizing the outage probability of *extrinsic* information on bits can be approximated by minimizing the outage probability of *extrinsic* information on symbols. Let $I_{e1}^{(i)}(a) = \frac{1}{Q} I(x_i; L_{e1,i}^s)$ denote the normalized *extrinsic* information on symbol x_i for given *a priori* $I_{a1} = a$ in the following discussion.

We have two optimization problems concerning the design of the inner ST modulator:

1. maximizing the *average extrinsic* information per bit for any given *a priori* information, i.e., maximizing $I_{e1}(I_{a1}) = \frac{1}{L} E(\sum_{i=1}^{L} I_{e1}^{(i)}(I_{a1}))$;

2. minimizing the outage probability of the *extrinsic* information of any symbol in the modulation block at perfect *a priori* information, i.e. maximizing $P_r(I_{e1}^{(i)}(1) \geq \rho)$ for any $i = 1, 2, \ldots, L$.

In general, optimization problem 1 is difficult to solve due to the unknown statistics of the *a priori* LLRs. However, the EXIT curves of the inner ST demodulator are monotonically increasing and close to a straight line. As such, in order to avoid early convergence of EXIT functions of the inner demodulator and outer decoder, we seek to maximize the *average extrinsic* information with no *a priori* information, i.e., the starting point $I_{e1}(0)$, and the *average extrinsic* information with perfect *a priori* information as well, i.e., the ending point $I_{e1}(1)$. The area A under the EXIT function I_{e1} of the inner ST demodulator corresponds approximately to $A \approx C(\Omega)/Q$ [39], where $C(\Omega)$ is the channel capacity of the constellation Ω for given $\mathbf{H}$, dispersion matrices $\{\mathbf{M}_i\}$ and SNR P/σ_z^2. Hence, high $I_{e1}(0)$ usually corresponds to low $I_{e1}(1)$ and vice versa. Namely, tradeoff exists between $I_{e1}(0)$ and $I_{e1}(1)$. How-

ever, Optimization problem 2, i.e., maximizing $P_r(I_{e1}^{(i)}(1) \geq \rho)$, usually implies a high value of $I_{e1}(1)$. Hence, to achieve good error performance at convergence and, at the same time, let the trajectory pass though the narrow tunnel in the middle and thus reach the second plateau region of the outer decoder, we propose

- maximize the channel capacity of the constellation Ω for given $\widetilde{\mathbf{H}}$;
- maximize $P_r(I_{e1}^{(i)}(1) \geq \rho)$.

Note, maximization of the channel capacity of the constellation Ω for given $\widetilde{\mathbf{H}}$ is a necessary condition for optimization problem 1, i.e., maximizing $I_{e1}(I_{a1})$ for given I_{a1}.

3.4 Design of Inner ST Modulation

In this section, we consider the design of the inner ST modulator by maximizing the channel capacity of the constellation Ω for given $\widetilde{\mathbf{H}}$ and maximizing $P_r(I_{e1}^{(i)}(1) \geq \rho)$ for any i as described.

3.4.1 Maximizing Channel Capacity

To make the optimization problem tractable and independent of constellation, we assume *i.i.d.* Gaussian inputs. The results can be used as design guidelines for practical input symbols drawn from finite alphabets.

Under the assumption of *i.i.d.* Gaussian inputs, the channel capacity of $\widetilde{\mathbf{H}}$ is given by

$$\begin{aligned} \tilde{C}(\frac{P}{\sigma_z^2}, \widetilde{\mathbf{H}}) &= \frac{1}{T} \log \left[\det \left(\mathbf{I}_{N_r T} + \frac{T}{L} \frac{P}{N_t \sigma_z^2} \widetilde{\mathbf{H}} \widetilde{\mathbf{H}}^H \right) \right] \\ &= \frac{1}{T} \log \left[\det \left(\mathbf{I}_L + \frac{T}{L} \frac{P}{N_t \sigma_z^2} \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \right) \right] \end{aligned} \tag{3.14}$$

Concerning the channel capacity of $\widetilde{\mathbf{H}}$, we have the following theorem.

Theorem 1: The channel capacity of $\widetilde{\mathbf{H}}$ is maximized if and only if the modulation matrix $\mathbf{G}$ satisfies

$$\mathbf{G}\mathbf{G}^H = \frac{L}{N_t T}\mathbf{I}_{N_t T} \tag{3.15}$$

Proof: See Appendix A. ■

Noting that matrix $\mathbf{G}$ is of size $N_t T \times L$, equation (3.15) holds only if $L \geq N_t T$. For simpler complexity of detection, we will only consider $L = N_t T$ in the sequel. In this case, (3.15) implies that

$$\mathrm{tr}(\mathbf{M}_m^H \mathbf{M}_n) = \begin{cases} 1, & \forall\, m = n \\ 0, & \forall\, m \neq n \end{cases} \tag{3.16}$$

Substituting (3.16) into (3.14), the corresponding maximum achievable capacity by inner ST modulation is

$$\tilde{C}_o = \log\left[\det\left(\mathbf{I}_{N_r} + \frac{P}{N_t \sigma_z^2}\mathbf{H}^H\mathbf{H}\right)\right] \tag{3.17}$$

which is exactly the capacity of the original channel when the channel state information is unknown at the transmitter but perfectly known at the receiver [1][2]. This shows that the inner ST modulation preserves the channel capacity only when the rate is at least N_t symbols per channel use.

3.4.2 Minimizing the Outage Probability of the Extrinsic Information

Theorem 2: The probability $P_r(I_{e1}^{(i)}(1) \geq \rho)$ is maximized only when $\mathbf{M}_i\mathbf{M}_i^H$ is full rank with identical nonzero eigenvalues for all i.

Proof: See Appendix B. ■

Interestingly, the set of dispersion matrices satisfying the conditions in Theorem 2 also optimize the pairwise error performance. With perfect feedback ($I_{a1} \rightarrow 1$)

for the symbols other than the symbol of concern, say x_i, the detection is based on the following observation obtained by perfectly cancelling interference from other symbols within the same block, i.e.,

$$\mathbf{y}_i = \mathbf{y} - \sqrt{\frac{P}{N_t}} \sum_{l \neq i} \tilde{\mathbf{h}}_l x_l = \sqrt{\frac{P}{N_t}} \tilde{\mathbf{h}}_i x_i + \tilde{\mathbf{z}} \tag{3.18}$$

Since $\tilde{\mathbf{h}}_i = \text{vec}(\mathbf{H}\mathbf{M}_i)$, then the modified Euclidean distance between a pair of transmitted symbols differing at position i in a modulation block is

$$d^2(x_i, \hat{x}_i|\mathbf{H}) = |x_i - \hat{x}_i|^2 \sum_{m=1}^{N_r} \mathbf{h}_m^H \mathbf{M}_i \mathbf{M}_i^H \mathbf{h}_m \tag{3.19}$$

where $\mathbf{h}_m^H$ is the m-th row vector of $\mathbf{H}$. Following the similar procedure by Tarokh *et al* in [4], it can be readily found that, to minimize the error probability of x_i, one needs to maximize the rank of $\mathbf{M}_i$ and the product of the nonzero eigenvalues of $\mathbf{M}_i\mathbf{M}_i^H$ just like the *Rank and Determinant criteria* in [4]. When $\text{tr}(\mathbf{M}_i\mathbf{M}_i^H)$ is fixed, the maximum coding gain is achieved when all the eigenvalues of $\mathbf{M}_i\mathbf{M}_i^H$ are equal.

In summary, we have the following two criteria for the design of the LD ST modulator in a CSTM system [40].

Capacity Criterion: The symbol rate of the LD ST modulator must be N_t symbols per channel use. Furthermore, the dispersion matrices shall be chosen such that their F-norms to be 1 and the trace of the Hermitian product of any pair of distinct dispersion matrices to be 0.

Error-Performance Criterion: For the best error performance, the dispersion matrices $\mathbf{M}_i\mathbf{M}_i^H$ for any i must be full rank with identical eigenvalues.

If the modulation block length T is N_t or greater, full rank can be easily guaranteed; while if T is less than N_t, full rank is impossible. *Error-Performance Criterion* also suggests that the minimum modulation block length shall be N_t.

3.4.3 Design Examples

To demonstrate our design criteria, three inner ST modulation design examples are provided below. In all the three schemes, $T = N_t$ and $L = N_t^2$. Let

$$\mathbf{P} = \begin{pmatrix} \mathbf{0}_{1\times(N_t-1)} & 1 \\ \mathbf{I}_{N_t-1} & \mathbf{0}_{(N_t-1)\times 1} \end{pmatrix} \tag{3.20}$$

denote the permutation matrix of size N_t that, when applied to the left, will cyclically shift a column vector of size N_t down by 1, $\mathbf{F} = [f_{mn}]$ as the DFT matrix of size $N_t \times N_t$ with (m, n)-th entry

$$f_{mn} = \frac{1}{\sqrt{N_t}} \exp(2\pi j(m-1)(n-1)/N_t) \tag{3.21}$$

and

$$\mathbf{S} = \begin{pmatrix} 1 & \mathbf{0}_{1\times(N_t-1)} \\ \mathbf{0}_{(N_t-1)\times 1} & \mathbf{0}_{(N_t-1)\times(N_t-1)} \end{pmatrix} \tag{3.22}$$

as a selection matrix of size N_t. The dispersion matrices of the three schemes are listed below.

Scheme 1: This is an optimal design [41]. Its dispersion matrices are

$$\mathbf{M}_{(k-1)N_t+i} = \text{diag}[\mathbf{f}_k]\mathbf{P}^{-(i-1)} \tag{3.23}$$

for $k = 1, 2, \ldots, N_t$ and $i = 1, 2, \ldots, N_t$, where $\mathbf{f}_k$ denotes the k-th column vector of $\mathbf{F}$.

Scheme 2: The full-diversity and full-rate scheme is proposed in [16]. Its dispersion matrices are

$$\mathbf{M}_{(k-1)N_t+i} = \beta^{k-1}\mathbf{P}^{-(k-1)}\alpha^{i-1}\text{diag}[\mathbf{f}_k] \tag{3.24}$$

for $k = 1, 2, \ldots, N_t$ and $i = 1, 2, \ldots, N_t$, where α and β are constellation dependent and are chosen to guarantee full diversity [16].

Scheme 3: The T-BLAST scheme is first introduced in [13]. The associated dispersion matrices are

$$\mathbf{M}_{(k-1)N_t+i} = \mathbf{P}^{(k-1)+(i-1)}\mathbf{S} \cdot \mathbf{P}^{-(i-1)} \tag{3.25}$$

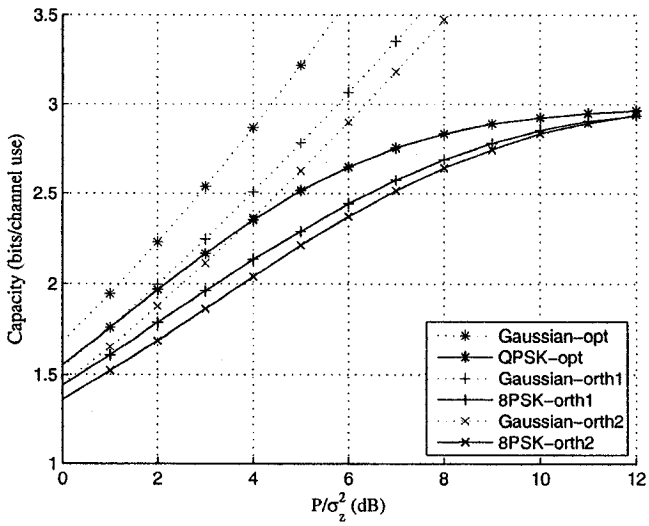
for $k = 1, 2, \ldots, N_t$ and $i = 1, 2, \ldots, N_t$.

It can be readily checked that although all the three schemes satisfy *Capacity Criterion* and preserve the original MIMO channel capacity, only the first two schemes satisfy *Error-Performance Criterion.* Without outer encoder, Scheme 2 achieves full diversity gain with appropriate α and β and hence outperforms Scheme 1. However, the presence of an outer code, the two schemes shall perform closely.

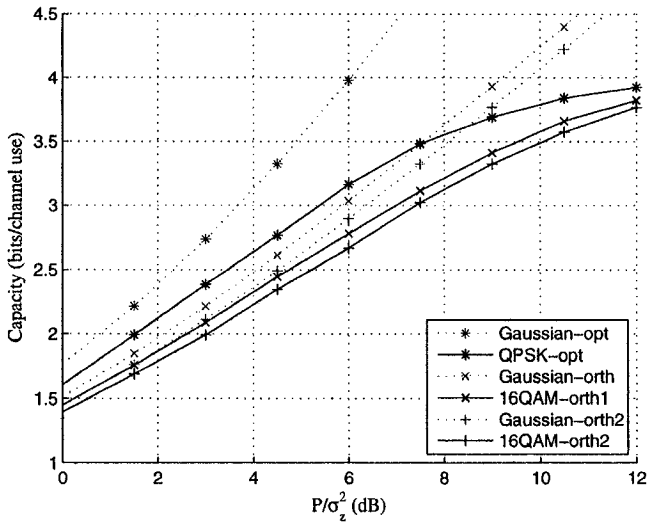
3.4.4 Trade-Off Between Constellation Size and Modulation Symbol Rate

For a given inner ST modulation rate in bit $R_m = LQ/T$, there exists a trade-off between constellation size 2^Q and symbol rate L/T. From Theorem 1, the channel capacity of $\widetilde{\mathbf{H}}$ is maximized only when $L/T \geq N_t$. In fact, it can also be shown that the channel capacity of $\widetilde{\mathbf{H}}$ is monotonically increasing with respect to L/T. Hence, for a given R_m, the minimal integer satisfying $Q \geq R_m/N_t$ shall be selected. If $R_m < N_t Q$ with the chosen Q, a subset of dispersion matrices shall be selected from an optimal design, e.g., Scheme 1.

To verify this design method, simulation was set up over a system with $N_t = N_r = T = 2$ under the Rayleigh ergodic flat fading channel. In this simulation, we compared the channel capacity of Scheme 1 which is an optimal design, Alamouti scheme [5] and an orthogonal scheme by selecting a pair of orthogonal dispersion



(a). $R_m = 3$



(b). $R_m = 4$

Figure 3.8: Channel capacity versus P/σ_z^2 for $R_m = 3$ and $R_m = 4$ bits per channel use,respectively.

Note: Dash curves correspond to Gaussian inputs and solid lines correspond to specific constellations, such as QPSK, 8PSK and 16QAM. "opt", "orth1" and "orth2" correspond to Scheme 1, Alamouti scheme, the orthogonal scheme (a orthogonal subset of Scheme 1), respectively.

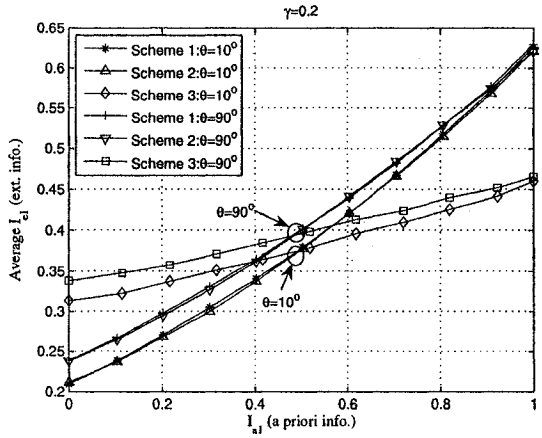
matrices $\{\mathbf{M}_i\}$ from Scheme 1. The corresponding numerical results are presented in Figure 3.8 when the modulation rate R_m is 3 and 4 bits per channel use, respectively. Note, for the case of $R_m = 3$, the optimal design selects a subset of $L = 3$ dispersion matrices from the full set of 4 dispersion matrices of Scheme 1 and uses QPSK constellation, and the other orthogonal schemes use $L = 2$ orthogonal dispersion matrices and 8PSK constellation. For the case of $R_m = 4$, the optimal design uses the full set of $L = 4$ dispersion matrices of Scheme 1 and QPSK constellation, and the other orthogonal schemes use the corresponding set of dispersion matrices as $R_m = 3$ and 16QAM constellation. As can be seen from these figures, the capacity difference between the optimal and the other two schemes grow more significantly as R_m increases.

In summary, for a given inner ST modulation rate, one seeks to choose a constellation size as small as possible till the modulation rate in symbol reaches N_t.

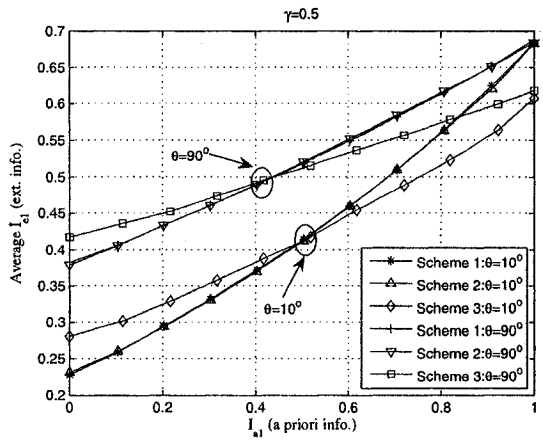
3.5 Simulation Results and Discussions

The EXIT characteristics of the three design examples as well as their error performance will be compared in this section to demonstrate our design criteria. Instead of prohibitive MAP detection as (3.5), the SIC-MMSE detection algorithm [27][28] was used as the inner demodulator in the simulation.

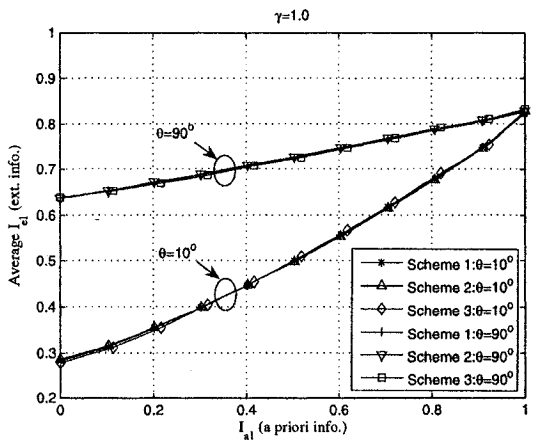
In Figure 3.9, the EXIT characteristics of the three schemes were compared under several fixed channel conditions. In the figure, $E_b/N_0 = 0dB$ (E_b is the transmitted power per bit here), $N_t = N_r = 2$, and QPSK constellation were applied and the channel coefficient matrix was assumed to have the form $\mathbf{H} = \begin{pmatrix} 1 & \gamma\cos\theta \\ 0 & \gamma\sin\theta \end{pmatrix}$. This type of $\mathbf{H}$ is the simplest MIMO channel whose interference among the transmitted symbols is controlled by γ and θ. Results were obtained for various values of magnitude γ and angle θ. In the EXIT charts, the magnitude γ determines the



(a). $\gamma = 0.2$



(b). $\gamma = 0.5$



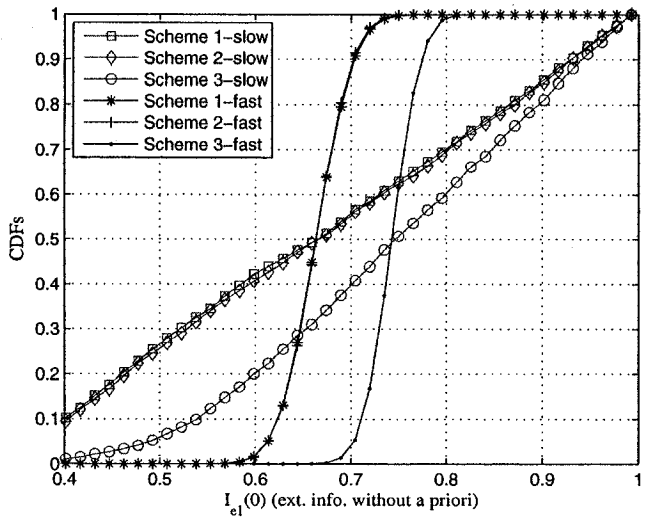
(c). $\gamma = 1.0$

Figure 3.9: EXIT characteristics of the three schemes under various channel conditions **H**.

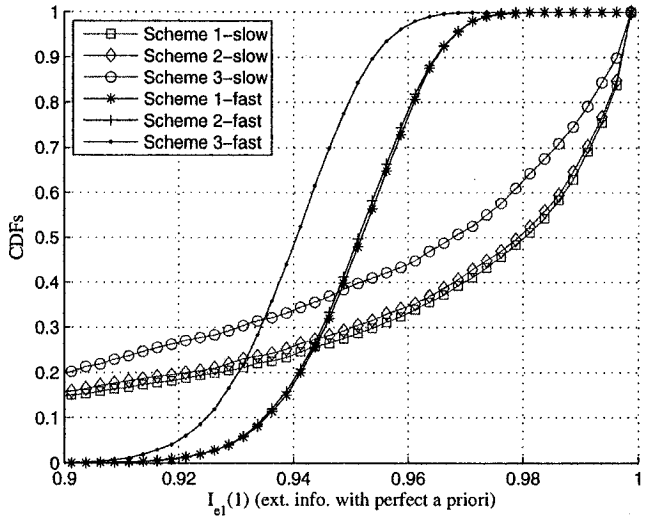
ending point of the transfer function of the inner ST demodulator, while the angle θ affects the starting point when the magnitude γ is fixed. As can be seen, for given $\mathbf{H}$, the area under the EXIT curve of a specific scheme is similar. This is expected since they all satisfy *Capacity Criterion*. However, since Scheme 3 does not comply with *Error-Performance Criterion*, under most channel conditions, it has smaller average *extrinsic* information $I_{e1}(1)$ than Scheme 2 and 3, or, its outage probability of the *extrinsic* information is larger. Also can be seen, if the starting point $I_{e1}(0)$ is large enough to let the trajectory of the turbo receiver pass though the bottleneck for each scheme, the first two schemes significantly outperform the third scheme, particularly when the gain of the second transmit antenna is significantly smaller than that of the first transmit antenna.

We now consider fading channels. In Figure 3.10 (a) and (b), cumulative distribution functions (CDFs) of average *extrinsic* information I_{e1} at two extreme cases, i.e. $I_{e1}(0)$ and $I_{e1}(1)$, under Rayleigh flat fading channels are given, respectively. In the figure, a channel is said to be a "slow" fading channel if it keeps constant in a coding frame but changes from frame to frame independently. A channel is a "fast" channel if it keeps constant in a modulation block but varies from block to block independently. Note, each sampled I_{e1} was the average *extrinsic* information taken over a coding frame for given channel $\mathbf{H}$. In the simulation, $E_b/N_0 = 3dB$, $N_t = N_r = 2$, QPSK constellation were applied.

As can be seen from Figure 3.10, the statistics of $I_{e1}(0)$ and $I_{e1}(1)$ associated with the first two schemes under both slow and fast fading channels are the same as expected. For $I_{e1}(0)$, the probability $P_r(I_{e1}(0) < \rho)$ of Scheme 3 is smaller than those of the first two schemes under both slow and fast fading channels. Namely, $I_{e1}(0)$ of Scheme 3 has more chances to be higher than those of the first two schemes. For $I_{e1}(1)$, the result is reversed. Namely, $I_{e1}(1)$ of Scheme 3 has more chances to be lower than those of the first two schemes. Hence, if all the three schemes have sufficient large



(a). $I_{e1}(0)$



(b). $I_{e1}(1)$

Figure 3.10: CDF comparison of $I_{e1}(0)$ and $I_{e1}(1)$ for the three schemes.
Note: Under the different Rayleigh ergodic flat fading channels. "fast" indicates fast fading channel and "slow" indicates slow fading channel.

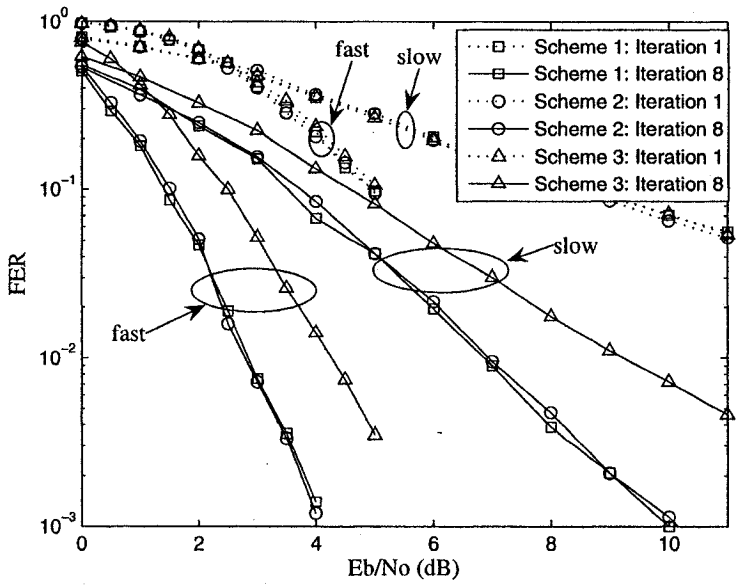


Figure 3.11: FER comparison of the three schemes under different Rayleigh ergodic flat fading channel.

Note: "fast" indicates fast fading channel and "slow" indicates slow fading channel.

$I_{e1}(0)$ to let corresponding turbo receivers pass though the bottleneck, the first two schemes outperform Scheme 3 under both slow and fast fading channels. Although Scheme 2 outperforms Scheme 1 in an uncoded system, their error performance curves in a coded system are almost indistinguishable. For both $I_{e1}(0)$ and $I_{e1}(1)$, the error-rate performance in a fast fading channel is significantly better than in a slow fading channel. This is expected because only when the channel varies significantly in a coding frame, the ergodic capacity is possible and the temporal diversity is achieved by the outer encoder.

Finally, in Figure 3.11, the frame error rate (FER) of the three schemes under different Rayleigh ergodic flat fading channels are compared. In the simulation, $N_t = N_r = T = 2$, QPSK constellation, 200 symbols per coding frame were assumed. A convolutional code with coding rate $1/2$ and constraint length 4 was used as the outer encoder. Its generator polynomials are $H^1(D) = 04, H^2(D) = 13$. In consistence with the EXIT chart analysis, the first two schemes perform indistinguishably to each other

but outperform the third scheme significantly after sufficient iterations. It is also clear that the three schemes perform closely after the first iteration since they all comply with *Capacity Criterion*. Again, the error-rate performance of all these schemes in a fast fading channel is significantly better than that in a slow fading channel.

3.6 Conclusions

A coded space-time modulation scheme with conventional outer encoder for MIMO wireless communications has been investigated. Using the EXIT chart technique, the design of the inner space-time modulator has been studied under the assumption of a joint iterative (turbo) receiver. Two design criteria are derived that relate to the channel capacity and error performance, respectively. To guarantee the convergence of the iterative receiver, the channel capacity of the inner ST modulations must be maximized. Once convergence is achieved, the error performance is optimized by maximizing the rank and determinant of the dispersion matrix of each individual symbol. The latter *Error-Performance Criterion* is much easier to apply than the well-known Tarokh's *Rank and Determinant Criteria* in [4]. The proposed two criteria together allow a complete design of the system concerning both the data rate and error performance. Specifically, it is shown that for a given inner ST modulation rate in bit, constellation size shall be minimized till the maximum symbol rate, i.e. N_t, is reached. The proposed design criteria have been verified by design examples and simulation results.

Chapter 4

Improved Stand-Alone Multi-User Detection

4.1 Introduction

In turbo MIMO communications, the abundance of existing outer decoders allows us to focus on the study on the inner MUD. To begin, we start from the stand-alone MUD, that can be used in a non-iterative receiver as in popular layered space-time architectures such as [47].

In MIMO communications [1][2], there exists interference among simultaneously transmitted symbols, which, if unattended, can severely degrade the error performance. Unfortunately, the practical applications of the optimal MUD scheme [42] is limited by its exponentially increasing complexity with respect to the number of users. As such, numerous suboptimal multi-user detectors with much reduced complexity have been developed (see [43] and the references therein).

Linear MUDs, such as the decorrelating detector and the minimum mean square error (MMSE) detector suffer from noise enhancement due to matrix inversion. This enhancement becomes significant when the cross-correlations among signature

signals are large [43][44]. Consequently, the error performance of linear MUDs is often unsatisfactory in MIMO communications where the temporary cross-correlations among spatial signatures can be very large. For this reason, nonlinear interference cancellation techniques, such as successive interference cancellation (SIC) [45]-[47] and parallel interference cancellation (PIC) [48]-[54], are desirable for MIMO communications. For instance, the V-BLAST scheme [47] uses an ordered successive interference cancellation and nulling detection scheme. This detector, which will be referred to as V-BLAST detector in the sequel, is in fact a conventional ordered SIC in which the matched-filer detection is replaced by decorrelating detection. More recently, it is found that PIC with decorrelating or MMSE detection as its first stage provides better error performance than the V-BLAST detector [51][52].

In this study, we propose a new PIC scheme [55] that is particularly suitable for MIMO communications. The proposed scheme can be viewed as a partial PIC scheme [49][50]. A key component in the proposed scheme is a nonlinear MMSE estimator that acts as the soft decision device in the conventional partial PIC schemes. The nonlinear estimator takes the decision statistics output from the linear estimator to generate a refined symbol estimate in sense of MMSE. This refinement is made possible by exploiting the knowledge of symbol alphabet. Since the nonlinear estimation involves little computation, the overall computational complexity is close to that of the conventional PIC schemes. In addition, for MIMO channels, the MMSE or decorrelating detector shall be used at the first stage. Furthermore, the proposed nonlinear MMSE estimator can also be extended to form a soft MUD scheme with *a priori* information, which can be applied in an iterative receiver. Simulation results show that the proposed detector significantly outperforms the linear MMSE detector and the other partial PIC schemes.

The rest chapter is organized as follows: In Section 4.2, system model is presented. In Section 4.3, Using the nonlinear MMSE estimator as soft decision device,

the novel partial PIC scheme is developed and some remarks on implementation of the proposed scheme are also provided. In Section 4.4, simulation results are provided to demonstrate the merits of the proposed PIC scheme. Finally, conclusions are drawn in Section 4.5.

4.2 System Model

In this study, we consider a general system model as follows. The transmitted symbols are collected in a column vector $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$. Each symbol is taken, with equal probability, from a complex constellation, i.e, $x_i \in \Omega \equiv \{\Omega_m |\ m = 0, 1, \ldots, 2^Q - 1, Q \geq 1\}$. The transmitted symbols are assumed to be mutually independent to each other. For convenience, the average symbol energy is assumed to be 1, i.e., $\frac{1}{2^Q}\sum_{m=0}^{2^Q-1} |\Omega_m|^2 = 1$. The MIMO channel is described by an $N \times M$ $(M \leq N)$ complex matrix $\mathbf{H}$. Note that the term *MIMO channel* is taken in a broad context: it may be the real physical channel as in a layered MIMO communication system or simply the collection of user signature signals as in a synchronous multi-user communication system over a Gaussian channel. It is assumed that the channel matrix $\mathbf{H}$ is known perfectly to the receiver but unknown to the transmitter. The received signal, $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$, is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{4.1}$$

where $\mathbf{z}$ is the additive white Gaussian noise (AWGN) term. The entries of $\mathbf{z}$ are assumed to be *i.i.d.* symmetrical complex Gaussian random variables with zero mean and variance σ_z^2. In the sequel, the i-th column of $\mathbf{H}$, denoted as $\mathbf{h}_i$, will be referred to as the signature signal of symbol x_i.

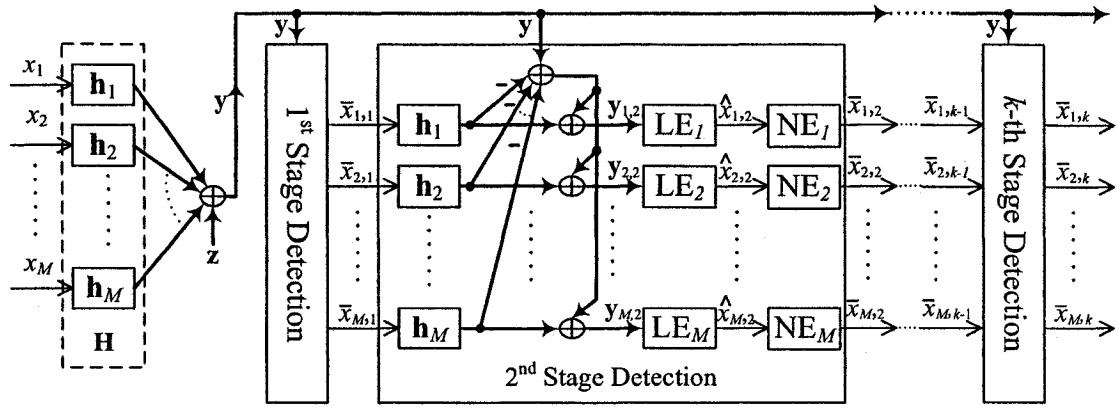


Figure 4.1: A block diagram of the proposed PIC detector.
Note: "LE" denotes the linear estimator and "NE" denotes the nonlinear MMSE estimator.

4.3 The Proposed PIC Scheme

The proposed multi-user detector is depicted in Fig. 4.1. As can be seen, it has a similar structure to that of the conventional partial PIC schemes. The main difference is that a nonlinear estimator is used in the proposed scheme to replace the soft decision device in the conventional partial PIC schemes. As such, each processing stage of the proposed detector consists of three steps: interference cancellation, linear estimation, and nonlinear estimation. In what follows, these processing steps will be described in detail.

4.3.1 Linear Estimation

Without loss of generality, we consider the estimation of one symbol, say x_i. Collect the rest of the transmitted symbols into a column vector $\mathbf{x}_I$ and their estimates from the $(k-1)$-th stage into a vector $\bar{\mathbf{x}}_{I,k-1}$. Since no estimation is performed at the beginning, $\bar{\mathbf{x}}_{I,0} = \mathbf{0}$. For symbol x_i, a "cleaner" signal is obtained by cancelling the interference due to other symbols as

$$\mathbf{y}_{i,k} = \mathbf{y} - \mathbf{H}_I \bar{\mathbf{x}}_{I,k-1} = \mathbf{h}_i x_i + \mathbf{H}_I \mathbf{n}_k + \mathbf{z} \tag{4.2}$$

where $\mathbf{H}_I = [\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, \ldots, \mathbf{h}_M]$ is the matrix obtained by removing the i-th column from $\mathbf{H}$, and $\mathbf{n}_k = \mathbf{x}_I - \bar{\mathbf{x}}_{I,k-1}$ is called the *cancellation residue* at the k-th stage. Note that the estimates $\bar{\mathbf{x}}_{I,k-1}$ are the outputs of the nonlinear MMSE estimators at the $(k-1)$-th stage and it can be verified that $E(\mathbf{n}_k) = \mathbf{0}$ if the symbol constellation is symmetric about the origin. Hence, we will assume that $\mathbf{n}_k$ has zero mean and denote its covariance matrix by $\mathbf{\Lambda}_{\mathbf{n}_k}$.

After the above interference cancellation, a linear estimator is applied and the corresponding output is given by

$$\hat{x}_{i,k} = \mathbf{w}_{i,k}^H \mathbf{y}_{i,k} = x_i + \hat{z}_{i,k}. \tag{4.3}$$

where $\hat{z}_{i,k}$ is the noise term of zero mean. Various linear estimators can be used and the corresponding $\mathbf{w}_{i,k}$ can be found as [43]

$$\mathbf{w}_{i,k} = \begin{cases} \left(\mathbf{h}_i^H \mathbf{h}_i\right)^{-1} \mathbf{h}_i & \text{for matched filter} \\ \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1} \mathbf{e}_i & \text{for decorrelator} \\ \frac{\left(\mathbf{h}_i\mathbf{h}_i^H + \mathbf{R}_{I,k}\right)^{-1}\mathbf{h}_i}{\mathbf{h}_i^H\left(\mathbf{h}_i\mathbf{h}_i^H + \mathbf{R}_{I,k}\right)^{-1}\mathbf{h}_i} & \text{for MMSE estimator} \end{cases} \tag{4.4}$$

where $\mathbf{R}_{I,k} = \mathbf{H}_I \mathbf{\Lambda}_{\mathbf{n}_k} \mathbf{H}_I^H + \sigma_z^2 \mathbf{I}$ and $\mathbf{e}_i$ is the i-th coordinate column vector. Note that the scaling factor $\frac{1}{\mathbf{h}_i^H\left(\mathbf{h}_i\mathbf{h}_i^H + \mathbf{R}_{I,k}\right)^{-1}\mathbf{h}_i}$ in the coefficient vector of the MMSE estimator $\mathbf{w}_{i,k}$ is added to ensure an unbiased estimation as indicated by (4.3). The variance of the noise term $\hat{z}_{i,k}$ can be found from (4.3) and (4.4) as

$$\hat{\sigma}_{i,k}^2 = \mathbf{w}_{i,k}^H \mathbf{R}_{I,k} \mathbf{w}_{i,k} \tag{4.5}$$

Substituting the coefficient vector for the MMSE estimator in (4.4) into (4.5), the

variance can be written as

$$\hat{\sigma}^2_{i,k,MMSE} = \frac{1}{\mathbf{h}_i^H \mathbf{R}_{I,k}^{-1} \mathbf{h}_i} \tag{4.6}$$

Denote $\lambda_{j,k}$ as the diagonal entry of the residue covariance matrix $\mathbf{\Lambda}_{\mathbf{n}_k}$ associated with symbol x_j, $j \neq i$. In addition, if $\mathbf{A}$ and $\mathbf{B}$ are invertible and $\mathbf{A} - \mathbf{B}$ is positive semi-definite, then $\mathbf{p}^H\mathbf{B}^{-1}\mathbf{p} \geq \mathbf{p}^H\mathbf{A}^{-1}\mathbf{p}$ and $\mathbf{p}^H\mathbf{A}\mathbf{p} \geq \mathbf{p}^H\mathbf{B}\mathbf{p}$ [56]. With the above fact, it can be checked from (4.5) and (4.6), if $\mathbf{\Lambda}_{\mathbf{n}_k}$ is diagonal, the output variance $\hat{\sigma}^2_{i,k}$ of all the three estimators is a monotonically increasing function with respect to the residue variance $\lambda_{j,k}$ [57]. This suggests that the variance of each residue, i.e., $\lambda_{j,k} = E(|x_j - \bar{x}_{j,k-1}|^2)$, shall be minimized at stage k in order for linear estimates with smaller MSE at stage $(k+1)$. Note, however, this observation is obtained under the assumption of mutually independent residues $x_j - \bar{x}_{j,k-1}$, or equivalently, mutually independent $\bar{x}_{j,k-1}$. In reality, they are correlated.

4.3.2 Nonlinear MMSE Estimation

Motivated by the observation in the last subsection, we seek a nonlinear MMSE estimator that uses the linear estimate in (4.3) to produce a refined estimate. With the linear estimate given in (4.3), the nonlinear MMSE estimator finds a refined estimate as $\bar{x}_{i,k} = f(\hat{x}_{i,k})$ that minimizes the MSE, i.e., $\lambda_{i,k} = E\left(|x_i - f(\hat{x}_{i,k})|^2\right)$.

To proceed, we assume that the linear estimate in (4.3) is Gaussian. This is true at the first stage, if the decorrelating estimator is used; whereas if the linear MMSE detector is used, $\hat{z}_{i,1}$ is approximately Gaussian [58]. In the subsequent stages, $\hat{z}_{i,k}$ is still approximately Gaussian based on the *central limit theorem* [31]. Let $\alpha_{i,k} = \mathrm{Re}(\hat{x}_{i,k})$ as the real part of $\hat{x}_{i,k}$ and $\beta_{i,k} = \mathrm{Im}(\hat{x}_{i,k})$ as the imaginary part of $\hat{x}_{i,k}$, i.e., $\hat{x}_{i,k} = \alpha_{i,k} + j\beta_{i,k}$, the MSE of symbol x_i at the k-th stage can be expressed

as

$$\lambda_{i,k} = E\left(|x_i - f(\hat{x}_{i,k})|^2\right)$$

$$= \frac{1}{2^Q}\int\int \sum_{m=0}^{2^Q-1} |f(\hat{x}_{i,k}) - \Omega_m|^2 \, p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right) d\alpha_{i,k} d\beta_{i,k} \tag{4.7}$$

where

$$p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right) = \exp\left(-|\hat{x}_{i,k} - \Omega_m|^2/\hat{\sigma}^2_{i,k}\right)/\pi\hat{\sigma}^2_{i,k} \tag{4.8}$$

Differentiating $\lambda_{i,k}$ with respect to f [59] and solving $\partial\lambda_{i,k}/\partial f = 0$, we obtain the optimal function f as

$$\bar{x}_{i,k} = f(\hat{x}_{i,k}) = \frac{\sum_{m=0}^{2^Q-1} \Omega_m p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right)}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right)} \tag{4.9}$$

Substituting (4.9) into (4.7), the minimum MSE is found as

$$\lambda^o_{i,k} = \min E\left(|f(\hat{x}_{i,k}) - x_i|^2\right)$$

$$= 1 - \frac{1}{2^Q}\int\int \frac{\left|\sum_{m=0}^{2^Q-1} \Omega_m p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right)\right|^2}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}^2_{i,k}\right)} d\alpha_{i,k} d\beta_{i,k} \tag{4.10}$$

Apparently, $\lambda^o_{i,k} \leq \hat{\sigma}^2_{i,k}$. That is, the MSE is further reduced at each stage by using the knowledge of symbol alphabet. In Fig. 4.2, the MSE $\lambda^o_{i,k}$ versus noise variance $\hat{\sigma}^2_{i,k}$ are plotted for various constellations. At each stage, the variances of the cancellation residue can be calculated using (4.10). The variance of $\hat{z}_{i,k}$ in (4.10) depends on the type of linear estimator used in generating the linear estimate and it can be calculated as (4.5).

Calculating $\bar{x}_{i,k}$ according to (4.9) and (4.8) requires the evaluation of exponential functions. In practical implementation, this can be accomplished by using a

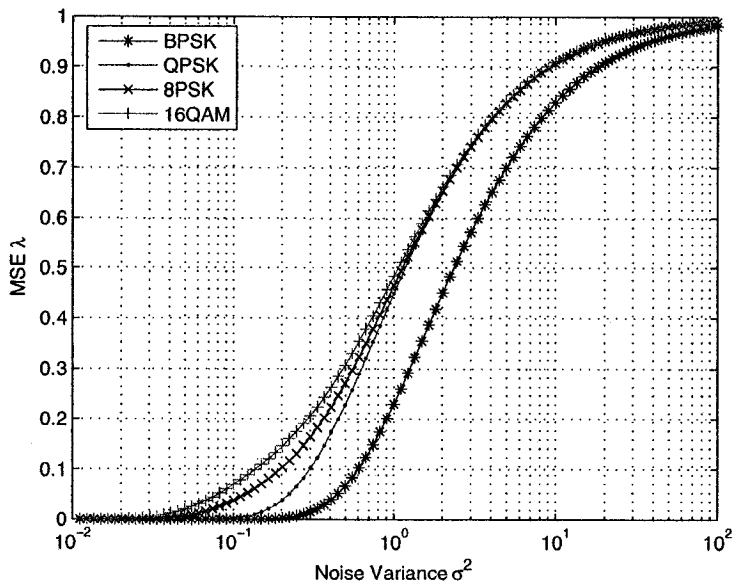


Figure 4.2: MSE of the nonlinear MMSE estimator for various constellations.

small lookup table and rewriting (4.9) as follows. Denote $d_m = |\bar{x}_{i,k} - \Omega_m|^2/\hat{\sigma}^2_{i,k}$ as the normalized Euclidean distance associated with Ω_m and assume $d_l \leq d_m$ for $\forall m \neq l$. Then (4.9) can be rewritten as

$$\bar{x}_{i,k} = \frac{\sum_{m=0}^{2^Q-1} \Omega_m \hat{p}_m}{\sum_{m=0}^{2^Q-1} \hat{p}_m} \tag{4.11}$$

where $\hat{p}_m = \exp\left[-(d_m - d_l)\right]$. Since $d_m - d_l \geq 0$, only negative exponents need to be included in the lookup table. For instance, one can have a table of 30 entries in which the n-th entry is $e^{-n/10}$ and set $\hat{p}_m = 0$ when $d_m - d_l > 3$. The error performance obtained using such a lookup table was found the same as that obtained by exactly evaluating the exponential functions. For large-sized constellations, only a few largest $\hat{p}_m$ need to be evaluated and others can be set to zero. In summary, the calculation of $\bar{x}_{i,k}$ is linear with the constellation size for it mainly involves the evaluation of the 2^Q normalized Euclidean distances.

It is interesting to note that a conditional nonlinear MMSE estimator that

minimizes the conditional MSE (i.e., $E\left(|\bar{x}_{i,k} - x_i|^2 \mid \hat{x}_{i,k}\right)$) was proposed in [60] and [61]. Apparently, the conditional nonlinear estimator also minimizes the unconditional MSE in (4.7) and, hence, leads to the same estimate as given in (4.9). However, for a constant channel $\mathbf{H}$, the conditional MSE is time-variant while the unconditional MSE in (4.10) is constant. Consequently, the linear MMSE detector is adaptive after the conditional nonlinear estimator [60] and fixed after the unconditional nonlinear estimator. This implies that the proposed detector has lower computational complexity in Gaussian or slowly fading channels. In addition, only MPSK constellations were considered in [60] while our estimator is developed for any constellations. The conditional nonlinear estimator was based on a somewhat heuristical reason: to provide a "better" symbol estimate in sense of MSE before interference cancellation. In this study, with the concept of unconditional MSE, it is proven that the use of the nonlinear estimator indeed reduces the output MSE of the following linear MMSE estimator.

4.3.3 Remarks on Implementation of the Proposed Scheme

Some remarks on the implementation of the proposed PIC detector are provided below.

Since the temporary cross-correlations among the symbols can be high for MIMO channel, the decorrelator or the linear MMSE estimator is often applied as the first-stage linear estimator rather than the matched filer. For symbol x_i at the first stage, the associated noise variance $\hat{\sigma}_{i,1}^2$ is given by

$$\hat{\sigma}_{i,1}^2 = \begin{cases} \sigma_z^2 g_{d,i} & \text{for decorrelator} \\ \sigma_z^2 g_{m,i} & \text{for MMSE} \end{cases} \tag{4.12}$$

where $g_{d,i}$ is the i-th diagonal entry of $\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$ and $g_{m,i}$ is the i-th diagonal entry of

$\left(\mathbf{H}^H\mathbf{H} + \sigma_z^2\mathbf{I}\right)^{-1}$. At the subsequent stages, since the decorrelator can no longer improve the error performance, the linear MMSE estimator or the matched filer receiver (MFR) is used for linear estimation.

As can be seen from (4.10), the calculation of residue variance $\lambda_{i,k}$ involves a double integration, which is difficult for real-time implementation. In practice, the integration can be calculated off-line and the results are stored in a table that links a value of $\hat{\sigma}_{i,k}^2$ to a value of $\lambda_{i,k}$. Alternatively, an instantaneous estimate of the variance of the cancellation residue can be used. Note that, with a given $\hat{x}_i$, the probability of x_i takes Ω_m at stage k is

$$P_m(\hat{x}_{i,k}) = \frac{p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}_{i,k}^2\right)}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{i,k}|\Omega_m, \hat{\sigma}_{i,k}^2\right)}. \tag{4.13}$$

Regarding $P_m(\hat{x}_{i,k})$ as *a priori* information independent of $f(\hat{x}_{i,k})$, an instantaneous estimate of the residue variance can be readily found as [27][28]

$$\hat{\lambda}_{i,k} = \sum_{m=0}^{2^Q-1} |\Omega_m|^2 P_m(\hat{x}_{i,k}) - |f(\hat{x}_{i,k})|^2 \tag{4.14}$$

As will be shown shortly, this approximation does not yield visible error-performance degradation. Actually, sometimes better error performance can be achieved with the instantaneous estimate in (4.14). This is possible because minimizing the MSE does not necessarily minimize the bit error rate (BER).

Since the interference is largely suppressed using the nonlinear estimates, the simple MFR can also be used as the linear estimator after the first stage. In fact, when the desired signal energy is far larger than the energy of interference, the MFR may outperform the linear MMSE detector [62].

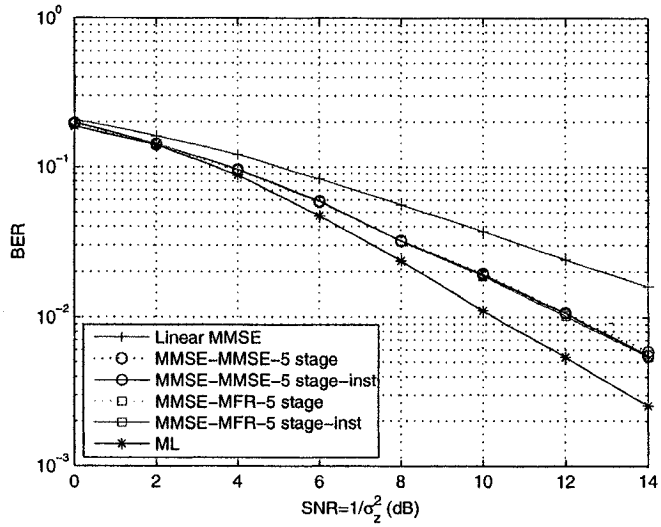
To avoid oscillation often happening in parallel interference cancellation at convergence [53][54], the following measures are taken: 1) A nonlinear estimate can be used at the next stage only if its residue variance is smaller than that at the

previous stage; otherwise, the nonlinear estimate from the previous stage will be used for the next stage. 2) Once the residue variance of an estimate is lower than a threshold, hard decision will be performed.

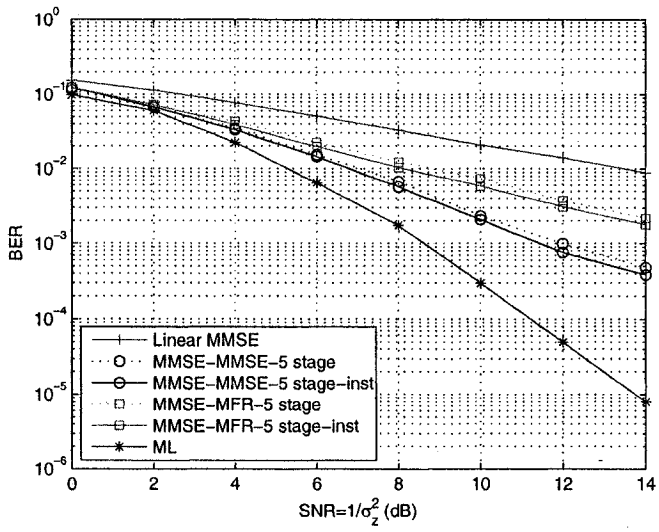
4.4 Simulation Results and Discussions

Simulation has been carried out to demonstrate the error performance of the proposed scheme. In the simulation, a fast fading MIMO channel was considered to demonstrate the average error performance of the compared schemes with random *signature signals*. That is, the entries h_{ij} in $\mathbf{H}$ were assumed to be *i.i.d.* symmetrical complex Gaussian random variables with zero mean and unit variance. The channel randomly changed from one symbol interval to another. QPSK modulation was assumed. In the proposed detector, the linear MMSE estimator or the decorrelator was applied as the first-stage linear estimator; while in the subsequent stages, the linear MMSE detector or the MFR was used. In the following figures, "MMSE", "ZF" and "MFR" stand for the linear MMSE estimator, the decorrelator and the matched filter receiver, respectively. The term right before "-" indicates the type of the linear estimator of the proposed detector at the first stage. The term after "-" indicates the type of the linear estimator of the proposed detector at the subsequent stages. The number before "stage" indicates stages of the proposed scheme and "inst" indicates the proposed scheme using instantaneous estimate of the residue variance.

In Fig. 4.3, BER performance of the proposed detector after 5 stages is compared with those of the linear MMSE detection and the optimal ML detection over a fast fading MIMO channel. In the proposed detector, the linear MMSE estimator is used as the first-stage linear estimator and the linear MMSE or the MFR are used in the subsequent stages. As can be seen from the figure, the two versions of the proposed scheme perform closely when the dimension of $\mathbf{H}$ is smaller, e.g. $N = M = 2$. Their



(a). $N \times M = 2 \times 2$



(b). $N \times M = 4 \times 4$

Figure 4.3: BER comparison of the linear MMSE detection, the ML detection, and the five-stage proposed detection.

Note: "inst" indicates the use of instantaneous estimate.

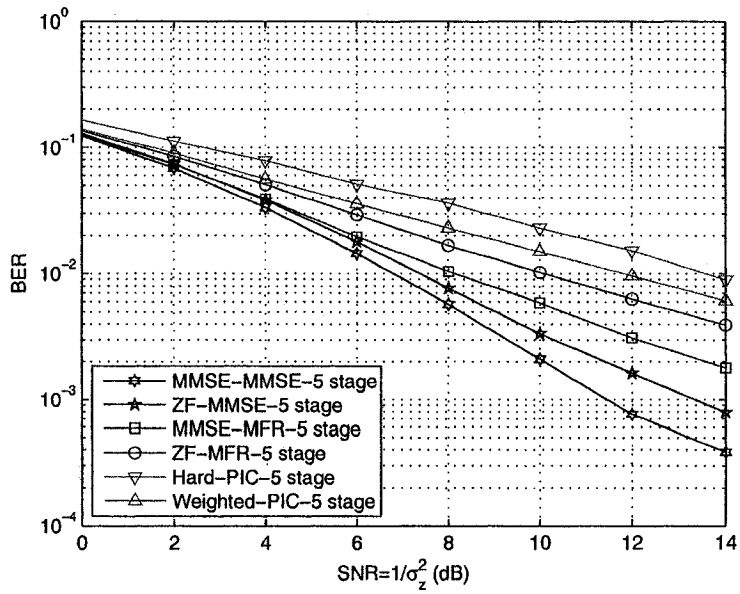
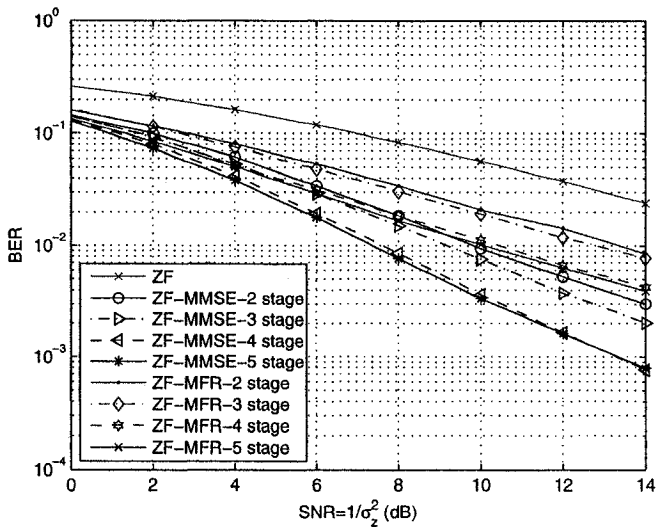


Figure 4.4: BER comparison of various proposed detectors and the other PIC schemes after 5 stages.

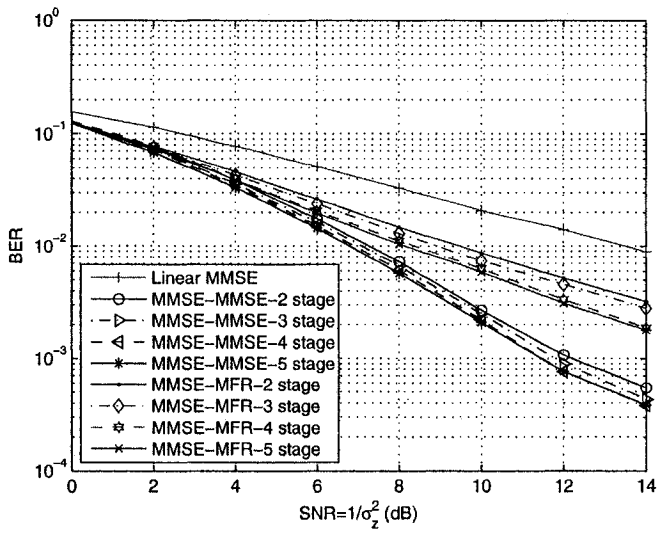
Note: "Hard-PIC" indicates the PIC scheme as [51]. "Weighted-PIC" indicates the weighted PIC scheme as [49].

difference becomes significant when the dimension becomes larger, e.g. $N = M = 4$. This is expected because the total energy of the cancellation residue before final detection is smaller in a system with less interfering symbols. In addition, from the figure, using instantaneous estimate of the residue variance as (4.14) yields slightly better error performance.

In Fig. 4.4, four versions of the proposed scheme are compared with the PIC schemes in [49] and [51] after 5 stages over a 4×4 MIMO fast fading channel. The instantaneous estimates of the residue variance was used in the proposed detectors. As can be seen from the figure, the first-stage estimator dictates the overall error performance of the proposed scheme: The better performance is achieved when the MMSE estimator is used at the first stage of the proposed detector. All the versions of the proposed scheme significantly outperform the PIC schemes in [49] and [51]. It is interesting to observe that the slope of the BER curve of the proposed scheme



(a). The decorrelator at the first stage



(b). The MMSE estimator at the first stage

Figure 4.5: Convergence of the proposed detection.

is much steeper than that of the conventional linear MMSE detector. This suggests that the proposed scheme yields larger diversity.

In Fig. 4.5, the effects of the number of iterations are studied for various proposed schemes over a 4×4 fast fading MIMO channel. The instantaneous estimates of the residue variance was used in the proposed detectors. As can be seen from the figure, the type of the linear estimator after the first stage dictates the convergence speed. That is, the proposed scheme with the linear MMSE estimator after the first stage converges faster than the counterpart with the MFR. Specifically, using linear MMSE estimator after the first stage results in faster convergence. When the linear MMSE estimator is used in all stages, 3 iterations are enough [57].

4.5 Conclusions

In this chapter, a new multi-user PIC scheme aided by a nonlinear MMSE estimation has been proposed. By exploiting the knowledge of symbol alphabet, a nonlinear MMSE estimator is developed, which uses a linear estimate to generate a nonlinear estimate with minimum mean square error. The nonlinear estimation allows more accurate regeneration of interference signal and leads to better linear estimates at next stage. The overall complexity of the proposed detector is linear with the number of users and the number of stages, which is close to that of the conventional PIC schemes. Simulation has been carried out and the results show that the proposed scheme outperforms the linear MMSE detector and other recent partial PIC schemes.

Chapter 5

Generalized SIC-MMSE Detection for Turbo MIMO Receiver

5.1 Introduction

Recently, the *turbo principle*, originally proposed for decoding parallel concatenated turbo codes[19], has found applications in various communications, such as multi-user communications[27], single-user multiple-antenna communications[10], and single-user communications with severe inter-symbol interference [28]. In these communication systems, transmitters can often be modelled as a serial concatenation of outer code(s) and an inner linear MIMO modulation, i.e., a turbo MIMO transmission system.

To decouple the correlation between the outer encoding and the inner modulation, random interleaving is often applied in between. The use of interleaver(s) makes the implementation of optimal maximum-likelihood (ML) decoding too complex to be practical. At best, joint iterative multi-user detection and decoding based on the turbo principle can be applied to achieve near-optimal solution with reasonable complexity. Such a turbo multi-user receiver employs two independent SISO devices,

the inner demodulator and the outer decoder(s), that exchange *extrinsic* information in an iterative fashion. In a turbo multi-user receiver, various multi-user detection (MUD) schemes can be used for the inner demodulation. Based on the results on MUD in Chapter 4, we study the application of MUD schemes in the turbo receiver.

Since the complexity of the optimal inner MUD grows exponentially with the number of bits in a signaling interval [10], suboptimal schemes with much reduced complexity are preferred. A popular suboptimal detector is the soft-interference-cancellation minimum-mean-square-error (SIC-MMSE) detector that uses the *a priori* information provided by the outer decoder to form estimates of interfering symbols [27][28]. Using these estimates, interference is reconstructed and then partially cancelled from the received signal. Following the interference cancellation, linear MMSE detection is performed to obtain a refined estimate of the desired symbol. The SIC-MMSE scheme has an advantage of low complexity but its error performance is often far from the optimum. In addition to the SIC-MMSE detection, other traditional MUD schemes based on interference cancellation, such as the parallel interference cancellation [49] and the serial interference cancellation [45] schemes, can also be extended to be used in a turbo multi-user receiver [63][64]. Another notable class of suboptimal MUD is the list detection/decoding based on sphere decoding. The list version of sphere decoding computes the *a posteriori* probabilities of bits using a subset of possibly transmitted symbols, called candidate list, in order for reduced complexity [10],[33]-[34]. In these list-based schemes, the key is the quality of the candidate list whose size determines the tradeoff between error performance and complexity. Much of the research effort in this front has been given to improving the candidate list. For instance, by taking into account the *a priori* information from the outer decoder, the candidate list can be improved from iteration to iteration such as in the iterative tree search (ITS) scheme [34]. The complexity of soft-output list-based decoding algorithms is often much higher than that of the schemes based on

interference cancellation. Further, the exact amount of computation of a list-based scheme varies depending on the channel condition and in the worst case can be close to that of the optimal MUD [10].

In soft-output MUD schemes based on interference cancellation such as the SIC-MMSE scheme, only the *a priori* information from the outer decoder is used to form the estimates of interfering symbols that will be used in the interference cancellation. In fact, the channel observation also provides useful information for interfering symbols. In this chapter, we propose a generalized SIC-MMSE detection scheme that exploits estimates gleaned from channel observation, in addition to decoder outputs, in estimating interfering symbols. For this purpose, a new nonlinear MMSE estimator is developed, which produces an MMSE estimate of an interfering symbol using both a *preliminary estimate* derived from channel observation and the decoder outputs associated with the symbol. Estimates of interfering symbols are then used in interference cancellation followed by linear MMSE detection. Depending on the methods of generating preliminary estimates from channel observation, we proposed two generalized SIC-MMSE detectors. In the first detector, the conventional SIC-MMSE detection is employed to generate preliminary estimates of interfering symbols, followed by the proposed nonlinear estimation and another stage of SIC-MMSE detection. This leads to a 2-staged SIC-MMSE detector. The other detector, a recursive SIC-MMSE detector, uses the detector outputs from the *previous* iteration as preliminary estimates. Analytical study and simulations are carried out to demonstrate the merits of the proposed schemes. It is shown that both of the proposed detectors significantly outperform the conventional SIC-MMSE detector. When used in a turbo receiver, they both lead to better error performance at convergence with less iterations. Furthermore, an *extrinsic* information transfer (EXIT) chart [37][38] analysis shows that the 2-staged detector outperforms the recursive detector at early iterations but underperforms at later iterations. A hybrid implementation that uses the 2-staged detector

at early iterations and the recursive detector at later iterations, provides the best error performance with significantly reduced complexity as compared to the conventional SIC-MMSE detector.

The rest chapter is organized as follows: In Section 5.2, preliminaries including the turbo MIMO transceiver and the conventional SIC-MMSE detection are presented. In Section 5.3, the generalized SIC-MMSE detection scheme is presented and the nonlinear MMSE estimator is developed to be used in the proposed scheme. In Section 5.4, two generalized SIC-MMSE detectors, the 2-staged and the recursive detectors, are developed and their convergence behavior is studied using the EXIT chart technique. Simulation results are provided in Section 5.5 to demonstrate the merits of the proposed schemes. Finally, conclusions are drawn in Section 5.6.

5.2 Preliminaries

We consider a general turbo MIMO transceiver as shown in Fig. 5.1 to demonstrate the proposed turbo MUD scheme. Note that the term *MIMO* channel is taken in a broad context: for example, it may be the real physical channel such as in a multiple-antenna communication system or the collection of user signature signals in a synchronous multi-user communication system over a Gaussian channel.

5.2.1 Turbo MIMO Transmitter

At the transmitter, the information bits are first encoded, shuffled by a random interleaver and then mapped into symbols. After that, the symbol stream is parsed into blocks of length M. Each block of symbols will be simultaneously transmitted over an $N \times M (N \geq M)$ MIMO channel as depicted in Fig. 5.1(a). This efficiently represents a standard MIMO bit-interleaved coded modulation (BICM) [20]-[24] transmitter. For multi-user systems, each user has its own channel encoder and interleaver. However,

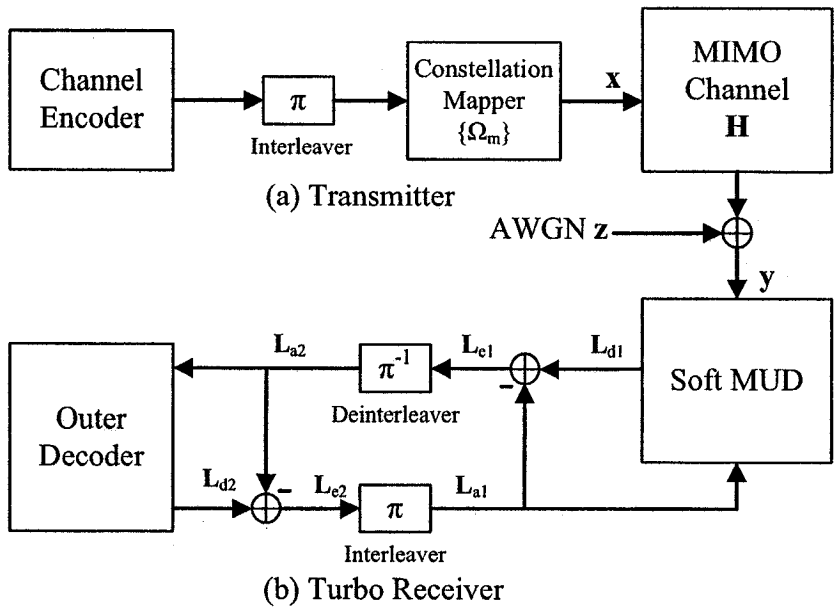


Figure 5.1: Block Diagram of A Turbo MIMO Transceiver.

this will not affect the implementation of the inner soft-output detector in the receiver. The symbol vector associated with one block is denoted by $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$ with $x_i \in \Omega \equiv \{\Omega_m | m = 0, 1, \ldots, 2^Q - 1, Q \geq 1\}$ (i.e., a complex constellation of size 2^Q, such as 2^Q-QAM). The transmitted symbols are assumed to be mutually independent to each other. For convenience, the average symbol energy is assumed to be 1, i.e., $\frac{1}{2^Q}\sum_{m=0}^{2^Q-1}|\Omega_m|^2 = 1$. The MIMO channel associated with $\mathbf{x}$ is described by an $N \times M$ complex matrix, denoted by $\mathbf{H}$. The i-th column of $\mathbf{H}$ is denoted by $\mathbf{h}_i$ and called the signature of symbol x_i. It is assumed that the channel matrix $\mathbf{H}$ is known perfectly to the receiver but unknown to the transmitter. The received signal $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$ associated with the input $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{5.1}$$

where $\mathbf{z}$ is the additive white Gaussian noise (AWGN) term. The entries of $\mathbf{z}$ are assumed to be *i.i.d.* and circularly symmetrical complex Gaussian random variables with zero mean and variance σ_z^2.

5.2.2 Turbo MIMO Receiver

The computational complexity of the optimal ML-based receiver is impractically large for such a concatenated system primarily due to the use of the interleaver. At best, an joint iterative detection and decoding receiver (turbo receiver), as illustrated in Figure 5.1(b), can be employed as presented in Chapter 2.

The optimal BCJR algorithm for the outer channel code has been developed by Bahl *et al* in [26] as presented in Chapter 2. Instead, the focus of this study is the inner soft MUD.

Given the *a priori* LLRs $\mathbf{L}_{a1}^{b} = [L_{a1,1}^{b}, L_{a1,2}^{b}, \ldots, L_{a1,MQ}^{b}]$ associated with bits in channel input $\mathbf{x}$ and the corresponding channel observation $\mathbf{y}$, the optimal MUD calculates the *extrinsic* LLR of the j-th bit in symbol x_i as

$$L_{e1,(i-1)Q+j} = \underbrace{\log \frac{P_r(b_{(i-1)Q+j} = 1|\mathbf{y}, \mathbf{L}_{a1})}{P_r(b_{(i-1)Q+j} = 0|\mathbf{y}, \mathbf{L}_{a1})}}_{L_{d1,(i-1)Q+j}} - L_{a1,(i-1)Q+j} \tag{5.2}$$

Apparently, the computational complexity of the optimal MUD grows exponentially with the number of bits in $\mathbf{x}$. Hence, suboptimal MUD schemes with reduced complexity are desirable.

5.2.3 Conventional SIC-MMSE Detection

The SIC-MMSE scheme, first proposed in [27][28] for the inner suboptimal MUD, has received a great deal of attention due to its simplicity. The SIC-MMSE detector consists of three steps: first, with the *a priori* information and knowledge of symbol alphabet, nonlinear estimates, called *a priori* estimates, are calculated for each symbol; then, for each symbol to be detected, interference is is reconstructed using the nonlinear estimates of interfering symbols and cancelled from the original received signal; last, linear MMSE estimation is performed using the relatively "cleaner" signal

to obtain a refined estimate of the symbol.

Without loss of generality, let us consider the detection of symbol x_i and collect the remaining interfering symbols $\{x_j, \forall j = 1, 2, \ldots, M, j \neq i\}$ into a column vector $\mathbf{x}_I$. Let $P_{a,j}(m)$ denote the *a priori* symbol probability that $x_j = \Omega_m$, which can be obtained from the *a priori* LLRs $\mathbf{L}_{a1}$. For x_j, an *a priori* estimate (i.e., mean) and the associated variance can be calculated using the knowledge of the symbol alphabet as [28]

$$\bar{x}_{a,j} = \sum_{m=0}^{2^Q-1} \Omega_m P_{a,j}(m) \tag{5.3}$$

$$\lambda_{a,j} = \sum_{m=0}^{2^Q-1} |\Omega_m|^2 P_{a,j}(m) - |\bar{x}_{a,j}|^2 \tag{5.4}$$

With the *a priori* estimates of the interfering symbols, interference is reconstructed and then a parallel interference cancellation is performed as

$$\mathbf{y}_{a,i} = \mathbf{y} - \mathbf{H}_I \bar{\mathbf{x}}_{a,I} = \mathbf{h}_i x_i + \mathbf{H}_I \mathbf{n}_{a,I} + \mathbf{z} \tag{5.5}$$

where $\mathbf{H}_I = [\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, \ldots, \mathbf{h}_M]$ is the matrix obtained by removing the i-th column from $\mathbf{H}$, $\bar{\mathbf{x}}_{a,I}$ is the *a priori* estimate of $\mathbf{x}_I$, and $\mathbf{n}_{a,I} = \mathbf{x}_I - \bar{\mathbf{x}}_{a,I}$. Obviously, the auto-correlation matrix of $\mathbf{n}_{a,I}$ is $E\left(\mathbf{n}_{a,I}\mathbf{n}_{a,I}^H\right) = \mathbf{\Lambda}_{a,I}$. After the above parallel interference cancellation, a linear MMSE estimator is applied to $\mathbf{y}_{a,i}$. The estimate of x_i is given by

$$\hat{x}_{a,i} = \mathbf{w}_{a,i}^H \mathbf{y}_{a,i} = x_i + \hat{z}_{a,i} \tag{5.6}$$

where $\hat{z}_{a,i}$ is the noise term of zero mean and the corresponding $\mathbf{w}_{a,i}$ can be found as [43]

$$\mathbf{w}_{a,i} = \frac{\left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_{a,I}\right)^{-1} \mathbf{h}_i}{\mathbf{h}_i^H \left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_{a,I}\right)^{-1} \mathbf{h}_i} \tag{5.7}$$

where $\mathbf{R}_{a,I} = \mathbf{H}_I \mathbf{\Lambda}_{a,I} \mathbf{H}_I^H + \sigma_z^2 \mathbf{I}$. Note that the scaling factor $\frac{1}{\mathbf{h}_i^H \left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_{a,I}\right)^{-1} \mathbf{h}_i}$ in the

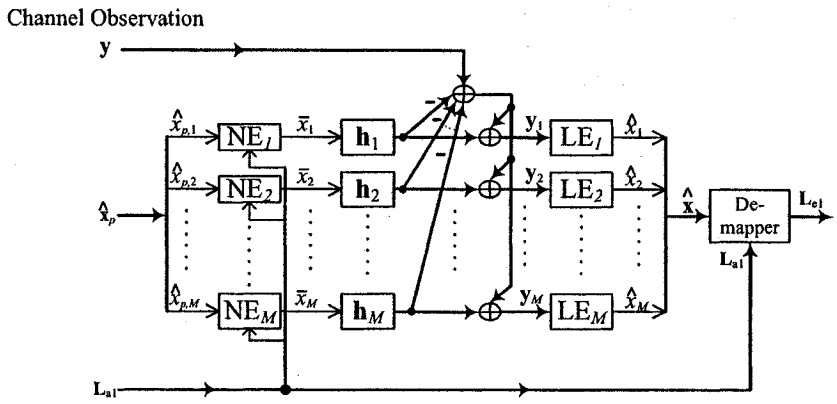


Figure 5.2: Generalized SIC-MMSE Detector.

coefficient vector of the linear MMSE estimator $\mathbf{w}_{a,i}$ is added to ensure an unbiased estimation by (5.6). The variance of the noise term $\hat{z}_{a,i}$ can be found from (5.6) and (5.7)

$$\hat{\sigma}^2_{a,i} = \mathbf{w}^H_{a,i}\mathbf{R}_{a,I}\mathbf{w}_{a,i} \tag{5.8}$$

With the estimate and the *a priori* LLRs associated with bits in symbol x_i, the *extrinsic* LLRs of the bits in x_i can be readily calculated [27][28] (also refer to equation (3.8)).

5.3 Generalized SIC-MMSE Detection

In the SIC-MMSE scheme, the nonlinear estimation (5.3) only uses the *a priori* information provided by the outer decoder. Actually, the channel observation can also be helpful in providing better estimates of interference. For instance, the outputs of a linear decorrelating detector or a linear MMSE detector applied to the channel observation are Gaussian or approximately Gaussian [58] and can be conveniently exploited by the nonlinear estimator in addition to the *a priori* information. Based on this observation, a generalized SIC-MMSE scheme is proposed as depicted in Figure 5.2. As can be seen, it has a similar structure as that of the conventional SIC-MMSE scheme. The main difference between the conventional SIC-MMSE scheme and the generalized

SIC-MMSE scheme is the nonlinear estimation. Besides the *a priori* LLRs $\mathbf{L}_{a1}$ from the outer decoder, some form of *preliminary estimates* $\{\hat{x}_{p,j}\}$ gleaned from the original channel observation are also involved in the nonlinear estimation. Note that, the generalized SIC-MMSE is also equivalent to one stage of the PIC scheme in Chapter 4 with modified nonlinear estimation. Before discussing the nonlinear estimation, some properties of the linear MMSE estimation after soft interference cancellation are described first.

5.3.1 Linear MMSE Estimation After Interference Cancellation

With the nonlinear estimate of $\mathbf{x}_I$, interference is reconstructed and then a parallel interference cancellation is performed as

$$\mathbf{y}_i = \mathbf{y} - \mathbf{H}_I\bar{\mathbf{x}}_I = \mathbf{h}_i x_i + \mathbf{H}_I\mathbf{n}_I + \mathbf{z} \tag{5.9}$$

where $\mathbf{H}_I = [\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, \ldots, \mathbf{h}_M]$ is the matrix obtained by removing the i-th column from $\mathbf{H}$, $\bar{\mathbf{x}}_I$ is the nonlinear estimate of $\mathbf{x}_I$, and $\mathbf{n}_I = \mathbf{x}_I - \bar{\mathbf{x}}_I$ is the residual interference. In the following discussion, $n_j = x_j - \bar{x}_j$ will be assumed to be zero-mean and its variance is denoted by λ_j, i.e., $\lambda_j = E(|n_j|^2)$, which is also the output mean square error (MSE) of the nonlinear estimator. To proceed, we have the following assumption.

Assumption 1: Entries of $\mathbf{n}_I$ are *independent* to each other and to the original channel noise $\mathbf{z}$ as well. As such, $\mathbf{n}_I$ has a diagonal auto-correlation matrix, i.e., $E\left(\mathbf{n}_I\mathbf{n}_I^H\right) = \mathbf{\Lambda}_I = \text{diag}[\lambda_j], \forall 1 \leq j \leq M, j \neq i$, and the cross-correlation matrix of $\mathbf{n}_I$ and $\mathbf{z}$ is an all-zero matrix, i.e., $E\left(\mathbf{n}_I\mathbf{z}^H\right) = \mathbf{0}$.

Apparently, Assumption 1 is satisfied in the conventional SIC-MMSE detector, where $\bar{\mathbf{x}}_I$ is obtained by only using the *a priori* LLRs $\mathbf{L}_{a1}$ as described in (5.3).

After the parallel interference cancellation, a linear MMSE estimator is applied to $\mathbf{y}_i$. The soft linear *extrinsic* MMSE estimate of x_i is given by

$$\hat{x}_i = \mathbf{w}_i^H \mathbf{y}_i = x_i + \hat{z}_i \tag{5.10}$$

where $\hat{z}_i$ is the noise term. The MMSE estimation coefficient vector $\mathbf{w}_i$ can be found as [43]

$$\mathbf{w}_i = \frac{\left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_I\right)^{-1} \mathbf{h}_i}{\mathbf{h}_i^H \left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_I\right)^{-1} \mathbf{h}_i} \tag{5.11}$$

where $\mathbf{R}_I$ is the auto-correlation matrix of interference and noise, i.e.,

$$\mathbf{R}_I = E\left[\left(\mathbf{H}_I \mathbf{n}_I + \mathbf{z}\right)\left(\mathbf{H}_I \mathbf{n}_I + \mathbf{z}\right)^H\right] \tag{5.12}$$

With Assumption 1, $\mathbf{R}_I$ is reduced to

$$\mathbf{R}_I = \mathbf{H}_I \mathbf{\Lambda}_I \mathbf{H}_I^H + \sigma_z^2 \mathbf{I} \tag{5.13}$$

Note that the scaling factor $\frac{1}{\mathbf{h}_i^H \left(\mathbf{h}_i \mathbf{h}_i^H + \mathbf{R}_I\right)^{-1} \mathbf{h}_i}$ of the linear MMSE estimator $\mathbf{w}_i$ in (5.11) is added to ensure an unbiased estimation. The noise term $\hat{z}_i$ is approximately Gaussian [27][28], whose variance can be found from (5.10) and (5.11) as

$$\hat{\sigma}_i^2 = \mathbf{w}_i^H \mathbf{R}_I \mathbf{w}_i = \frac{1}{\mathbf{h}_i^H \mathbf{R}_I^{-1} \mathbf{h}_i} \tag{5.14}$$

Regarding to the output variance $\hat{\sigma}_i^2$, we have the following proposition.

Proposition 1: Under Assumption 1, the output variance $\hat{\sigma}_i^2$ of the linear MMSE estimator is a monotonically increasing function with respect to the variance of the input residual interference λ_j, $\forall j \neq i$.

Proof: As can be seen from (5.14), the variance $\hat{\sigma}_i^2$ is a monotonically decreasing function with respect to $\mathbf{h}_i^H \mathbf{R}_I^{-1} \mathbf{h}_i$ with $\mathbf{R}_I$ given by (5.13). From [56], if $\mathbf{A}$ and $\mathbf{B}$

are invertible and $\mathbf{A}-\mathbf{B}$ is positive semi-definite, then $\mathbf{p}^H\mathbf{B}^{-1}\mathbf{p} \geq \mathbf{p}^H\mathbf{A}^{-1}\mathbf{p}$. With this and the fact that $\mathbf{\Lambda}_I$ is diagonal, it can be checked that the output variance $\hat{\sigma}_i^2$ is a monotonically increasing function with respect to λ_j. ■

The above proposition suggests that when used in a generalized SIC-MMSE detector, the optimal nonlinear estimator shall minimize the output MSE λ_j.

5.3.2 Nonlinear MMSE Estimation

According to Proposition 1, we seek a nonlinear estimator that minimizes the output MSE. The optimal nonlinear MMSE function is described in the following proposition.

Proposition 2: Suppose $\hat{x}_{p,j}$ is an observation on x_j with conditional probability density function (pdf) $p\left(\hat{x}_{p,j}|\Omega_m\right)$. Based on the observation $\hat{x}_{p,j}$, the mean square error (MSE), i.e., $\lambda_j = E\left(|f(\hat{x}_{p,j}) - x_j|^2\right)$, is minimized by the following function

$$\bar{x}_j = f(\hat{x}_{p,j}) = \frac{\sum_{m=0}^{2^Q-1} \Omega_m p\left(\hat{x}_{p,j}|\Omega_m\right) P_{a,j}(m)}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{p,j}|\Omega_m\right) P_{a,j}(m)} \tag{5.15}$$

where $\{P_{a,j}(m)\}$ is the *a priori* probabilities $P_r(x_j = \Omega_m)$

Proof: Let $\alpha_j = \mathrm{Re}(\hat{x}_j)$ as the real part of $\hat{x}_{p,j}$ and $\beta_j = \mathrm{Im}(\hat{x}_{p,j})$ as the imaginary part of $\hat{x}_j$, i.e., $\hat{x}_{p,j} = \alpha_j + j\beta_j$, the MSE of symbol x_j can be expressed as

$$\begin{aligned}\lambda_j &= E\left(|f(\hat{x}_{p,j}) - x_j|^2\right) \\ &= \int\int \sum_{m=0}^{2^Q-1} |f(\hat{x}_{p,j}) - \Omega_m|^2 \, p\left(\hat{x}_{p,j}|\Omega_m\right) P_{a,j}(m) d\alpha_j d\beta_j\end{aligned} \tag{5.16}$$

Differentiating λ_j with respect to f and solving $\partial\lambda_j/\partial f = 0$, we obtain the optimal function f in (5.15). ■

Substituting (5.15) into (5.16), the associated minimum MSE is found as

$$\lambda_j^o = 1 - \int\int \frac{\left|\sum_{m=0}^{2^Q-1} \Omega_m p\left(\hat{x}_{p,j}|\Omega_m\right) P_{a,j}(m)\right|^2}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{p,j}|\Omega_m\right) P_{a,j}(m)} d\alpha_j d\beta_j \tag{5.17}$$

In the following discussion, the observation $\hat{x}_{p,j}$ on x_j is assumed to be Gaussian with variance $\hat{\sigma}_{p,j}^2$, i.e.,

$$p\left(\hat{x}_{p,j}|\Omega_m\right) = \exp\left(-|\hat{x}_{p,j} - \Omega_m|^2/\hat{\sigma}_{p,j}^2\right)/\pi\hat{\sigma}_{p,j}^2 \tag{5.18}$$

It can be readily checked that the optimal nonlinear MMSE estimator has the following properties.

- The output MSE is smaller than or equal to the input noise variance, i.e., $\lambda_j \leq \hat{\sigma}_{p,j}^2$.
- With symmetric constellation, $E\left(f(\hat{x}_{p,j}) - x_j\right) = E(n_j) = 0$. That is, the MSE λ_j is also the variance of $n_j = x_j - \bar{x}_j$.
- Given the *a priori* probabilities $\{P_{a,j}(m)\}$, the residual variance λ_j is a monotonically increasing function with respect to the variance of $\hat{x}_{p,j}$, i.e., $\hat{\sigma}_{p,j}^2$.

Moreover, as the variance $\hat{\sigma}_{p,j}^2 \rightarrow \infty$ or, equivalently, at the absence of $\hat{x}_{p,j}$, (5.15) is reduced to the *a priori* estimate (i.e. mean) in the conventional SIC-MMSE detector as described in (5.3). Hence, whenever there exists a preliminary observation $\hat{x}_{p,j}$ with finite variance $\hat{\sigma}_{p,j}^2$, we have $\lambda_j < \lambda_{a,j}$, where $\lambda_{a,j}$ is given by (5.4). With the above facts, we have the following proposition.

Proposition 3: Under Assumption 1, the output noise variance of a generalized SIC-MMSE detector is less than that of the conventional SIC-MMSE detector for any given *a priori* probabilities $\{P_{a,j}(m)\}$.

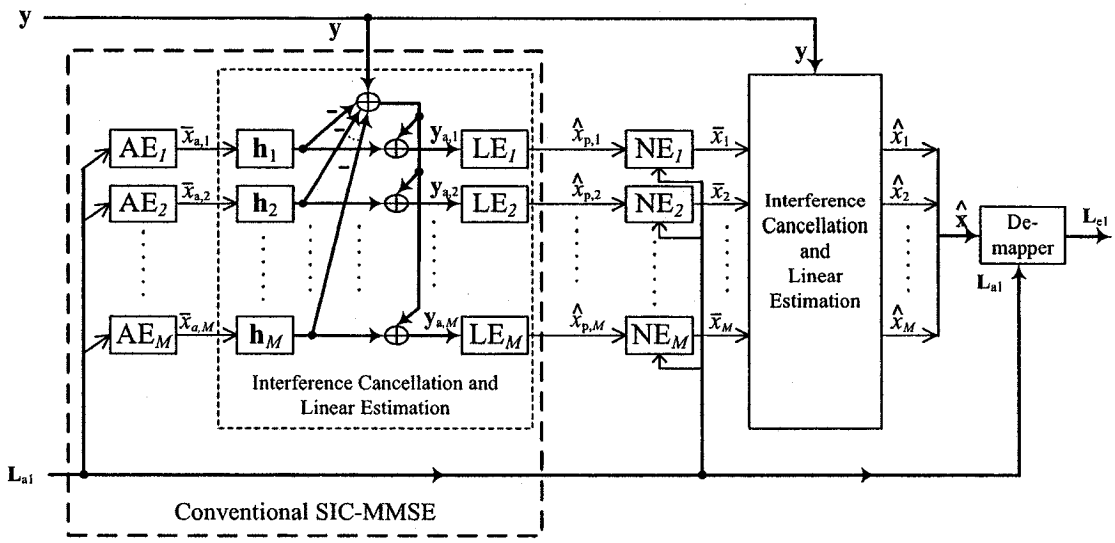


Figure 5.3: A block diagram of the proposed 2-staged soft MUD.
Note: "LE" denotes the linear MMSE estimator, "AE" denotes the *a priori* estimator and "NE" denotes the nonlinear MMSE estimator.

Proof: By Proposition 1, the output variance of the linear MMSE detector is a monotonically increasing function with respect to input variance. This with the fact that $\lambda_j < \lambda_{a,j}$ proves the proposition. ■

This proposition suggests that a generalized SIC-MMSE detector with preliminary estimates $\hat{\mathbf{x}}_p$ always outperforms the conventional SIC-MMSE detector if Assumption 1 is satisfied.

5.4 Proposed SIC-MMSE Schemes

Based on the above generalized SIC-MMSE model, we propose two methods for generating the preliminary estimates $\{\hat{x}_{p,j}\}$. Accordingly, we have two new soft-output MUD schemes, which will be described in the sequel.

5.4.1 2-Staged SIC-MMSE Detection

The preliminary estimates, $\{\hat{x}_{p,j}\}$, which are inputs to the proposed nonlinear estimator, can be obtained by using a conventional SIC-MMSE detector. As such, a 2-staged detector is proposed as depicted in Fig. 5.3. At the first stage, preliminary estimates $\hat{\mathbf{x}}_p$ are generated by a conventional SIC-MMSE detector. As mentioned before, the soft output of the first stage detection, $\hat{x}_{p,j}$ is approximately Gaussian and its variance can be found by substituting (5.4), (5.11) and (5.12) into (5.14). Then, a generalized SIC-MMSE detector follows with $\{\hat{x}_{p,j}\}$ as its preliminary estimate. Specifically, for symbol x_j, the optimal nonlinear estimation is performed using the *a priori* probabilities $\{P_{a,j}(m)\}$ from the outer decoder and the output of the first-stage detection, $\hat{x}_{p,j}$, according to (5.15) and (5.18). The outputs of the nonlinear estimators, $\{\bar{x}_j\}$, are then used for interference cancellation as in (5.9). Finally, the linear MMSE estimation is performed. The coefficient vector of the linear MMSE estimator at the second stage can be found by substituting (5.12) into (5.11). Note, however, due to the use of the nonlinear estimators, $\mathbf{R}_I$ in (5.12) is difficult to evaluate. In practice, we assume that entries of $\bar{\mathbf{x}}_I$ are independent to each other and to the channel noise $\mathbf{z}$ as well. That is, Assumption 1 is satisfied. Hence, $\mathbf{R}_I$ is reduced to (5.13) with diagonal $\mathbf{\Lambda}_I$, whose diagonal entries can be calculated using (5.17).

As can be seen, the calculation of the MSE in (5.17) involves a double integration, which is difficult for real-time implementation. Alternatively, an instantaneous estimate of the MSE can be used. Note that, with a given $\hat{x}_{p,j}$, the probability of x_j taking Ω_m is

$$P_m(\hat{x}_{p,j}) = \frac{p\left(\hat{x}_{p,j}|\Omega_m, \hat{\sigma}^2_{p,j}\right) P_{a,j}(m)}{\sum_{m=0}^{2^Q-1} p\left(\hat{x}_{p,j}|\Omega_m, \hat{\sigma}^2_{p,j}\right) P_{a,j}(m)}. \tag{5.19}$$

Hence, an instantaneous estimate of the MSE can be readily found as [27][28]

$$\hat{\lambda}_j = \sum_{m=0}^{2^Q-1} |\Omega_m|^2 P_m(\hat{x}_{p,j}) - |f(\hat{x}_{p,j})|^2 \tag{5.20}$$

It is shown in [57] that this approximation does not yield visible error performance degradation. In what follows, the instantaneous estimate of the MSE, i.e., $\hat{\lambda}_j$ in (5.20), will always be used. Additionally, the implementation issues will be treated in the same way as subsection 4.3.3.

By Proposition 3, if nonlinear estimates $\{\bar{x}_j\}$, or equivalently preliminary estimates $\{\hat{x}_{p,j}\}$, satisfy Assumption 1, the 2-staged SIC-MMSE detector certainly outperforms the conventional SIC-MMSE detector.

Note, however, since the preliminary estimates $\{\hat{x}_{p,j}\}$ are gleaned from the channel observation, they are correlated to each other and to the channel noise $\mathbf{z}$ as well. Still, simulation results show that the 2-staged SIC-MMSE detector significantly outperforms the conventional SIC-MMSE detector, particularly at early iterations. The reasons are as follows. At early iterations, when the *a priori* information is small, significant reduction on the output MSE of the nonlinear estimator is expected due to the first-stage linear MMSE detection. That is, the diagonal entries of the true auto-correlation matrix $\mathbf{\Lambda}_I$ are significantly reduced, leading to smaller MSE at the output of the second-stage detection. At later iterations, the amount of the *a priori* information becomes large. Since the *a priori* LLRs of different symbols are uncorrelated to each other [65], the correlations among $\{\bar{x}_i\}$ become small, so do the correlations between $\{\bar{x}_i\}$ and the channel noise $\mathbf{z}$.

It is worth noting that the computational complexity of the 2-staged detector is approximately twice as large as that of the conventional SIC-MMSE detector. However, as can be seen from the simulation results, the turbo receiver based on the 2-staged detector needs fewer iterations to reach convergence as compared to the receiver based on the conventional SIC-MMSE detector. Since the complexity of the outer decoder is much higher than the inner detection, less iterations suggest less overall complexity.

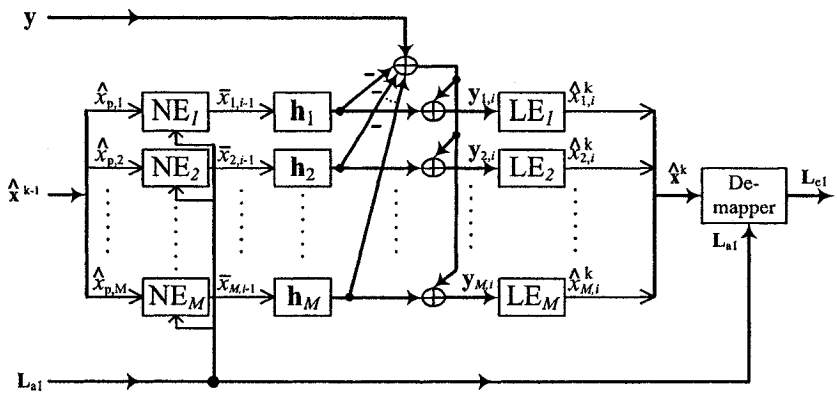


Figure 5.4: A block diagram of the proposed recursive soft MUD at iteration k.
Note: The superscript indicates the iteration. "LE" denotes the linear estimator and "NE" denotes the nonlinear MMSE estimator.

5.4.2 Recursive SIC-MMSE Detection

Naturally, the detector outputs from the *previous* iteration can also be used as the preliminary estimates in the generalized SIC-MMSE detection, i.e., at the k-th iteration, $\hat{\mathbf{x}}_p = \hat{\mathbf{x}}^{k-1}$ (the superscript $k-1$ indicates the $(k-1)$-th iteration). This leads to a *recursive* detector [66] as depicted in Fig. 5.4 for iteration k. In the proposed recursive detector, the preliminary estimates $\hat{\mathbf{x}}_p$ from the *previous* iteration are used in conjunction with the decoder outputs to generate refined nonlinear estimates as described in (5.15) and (5.18). With the nonlinear estimates $\bar{\mathbf{x}}_I$, interference is cancelled and then linear MMSE estimation is performed using (5.9), (5.11) and (5.13).

Apparently, the computational complexity of the recursive detector is comparable to that of the conventional SIC-MMSE detector. As predicted by Proposition 3, with the same number of iterations, the turbo receiver with the recursive SIC-MMSE detection outperforms the receiver with the conventional SIC-MMSE detection.

5.4.3 Convergence Behaviors

In this subsection, we compare the convergence behaviors of the turbo receivers that use the conventional SIC-MMSE detector and the proposed detectors. The EXIT

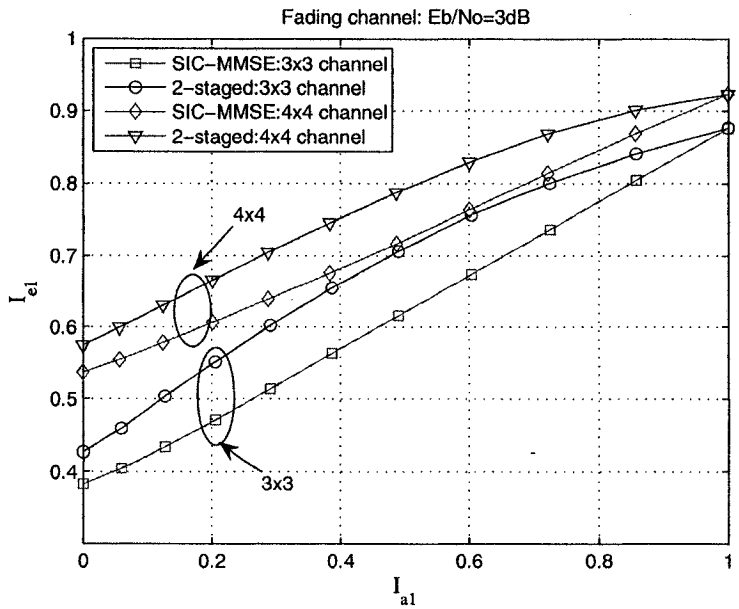


Figure 5.5: EXIT curves of the conventional SIC-MMSE detector and the 2-staged detector.

chart technique, pioneered by S. ten Brink in [37][38], will be used. The EXIT chart is a powerful tool for studying the iterative behavior of a turbo receiver by solely looking at the input/output relations of its constituent soft devices. For a given soft device, denote the mutual information between a bit and its *a priori* LLR as $I_a = I(b; L_a)$ and similarly the mutual information between a bit and its *extrinsic* LLR as $I_e = I(b; L_e)$. The EXIT function $T : I_e = T(I_a)$ for a given channel and constellation mapping is often a non-decreasing function. Plotting the two EXIT functions in one chart together, the trajectory of the turbo receiver can be predicted.

For the conventional SIC-MMSE and the 2-staged SIC-MMSE detector, the *extrinsic* information I_{e1} is only a function of the *a priori* information I_{a1} if the channel and constellation mapping are fixed. Further, the *extrinsic* LLRs of the outer decoder are approximately Gaussian [65]. With these facts, the EXIT functions of the conventional SIC-MMSE and the 2-staged detectors can be readily generated via Monte-Carlo simulations. In Figure 5.5, the EXIT curve of the 2-staged detector

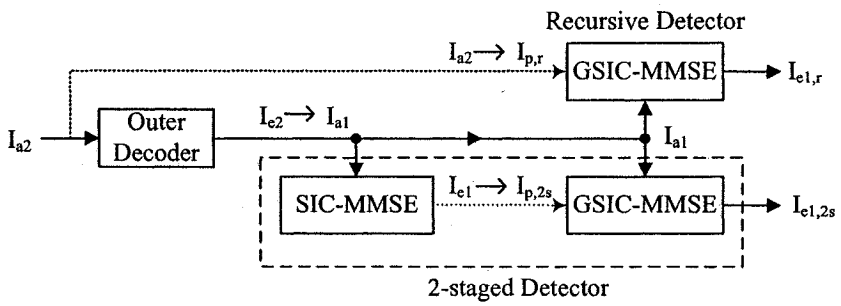


Figure 5.6: Turbo receivers for comparison of the two proposed detectors.
Note: "SIC-MMSE" denotes the conventional SIC-MMSE detector; "GSIC-MMSE" denotes the generalized SIC-MMSE detector.

is plotted together with that of the conventional SIC-MMSE detector for 3×3 and 4×4 fading channels. Note that the inner MUD is performed symbol interval by symbol interval. It has the same EXIT function regardless of the channel variation rate. In the figure, $E_b/N_0 = 3dB$ and QPSK constellation are assumed. As can be seen from the figure, the EXIT curve of the 2-staged detector is consistently higher than that of the conventional SIC-MMSE detector except when $I_{a1} = 1$. Hence, the 2-staged detector leads to higher convergence point, regardless of the outer decoder used. It is also evident that, for a given error performance, the 2-staged detector requires less iterations.

For the recursive detector, however, its outputs $\hat{\mathbf{x}}^k$ at iteration k depend not only on the *a priori* LLR inputs $\mathbf{L}_{a1}$ but also on the detector outputs $\hat{\mathbf{x}}^{k-1}$ at *previous* iteration. That is, I_{e1}^k is not only a function of I_{a1} but also a function of I_{e1}^{k-1}. For a given decoder, its EXIT function can be derived, which is however of little practical use due to its dependency on the outer decoder. Instead, we compare the *extrinsic* information of the recursive SIC-MMSE detector $I_{e1,r}$ with that of the 2-staged detector $I_{e1,2s}$ for a given decoder output $I_{e2}(I_{a1})$ as shown in Fig. 5.6.

Proposition 4: Given an outer decoder and its *extrinsic* information $I_{e2}(I_{a1})$,

we have

$$\begin{cases} I_{e1,r} < I_{e1,2s} & \text{if} I_{a1}(= I_{e2}) < I_{a1}^{c} \\ I_{e1,r} \geq I_{e1,2s} & \text{else} \end{cases} \tag{5.21}$$

where I_{a1}^{c} is the *a priori* information of the inner detector at convergence in the turbo receiver that uses the conventional SIC-MMSE detector and the given outer decoder.

Proof: For given *a priori* information I_{a1}, channel condition, and constellation mapping, the *extrinsic* information of a generalized SIC-MMSE detector is a non-decreasing function with respect to the *preliminary* information $I_p = I(b; \hat{x}_{p,j})$. As can be seen from Fig. 5.6, for the recursive detector, the preliminary information is

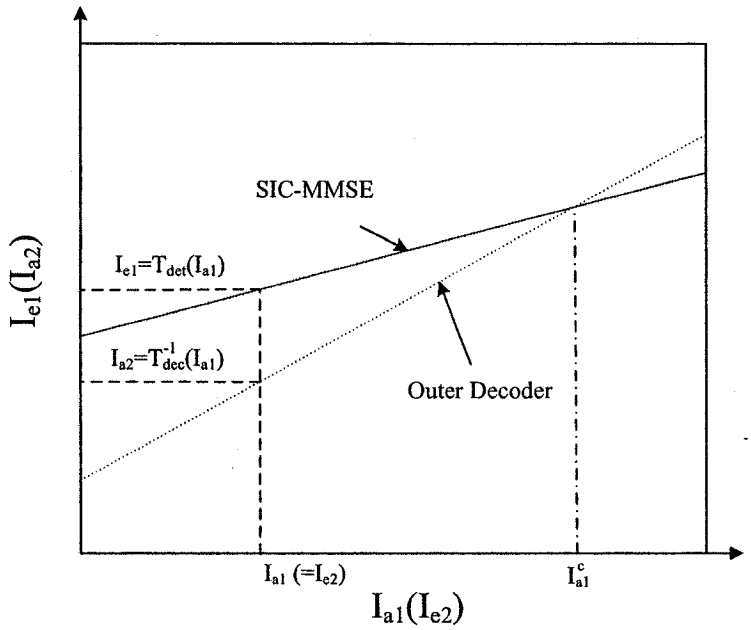
$$I_{p,r} = I_{a2} = T_{dec}^{-1}(I_{a1}) \tag{5.22}$$

where $T_{dec}^{-1}()$ is the inverse function of the EXIT function of the outer decoder. For the 2-staged detector, the preliminary information is

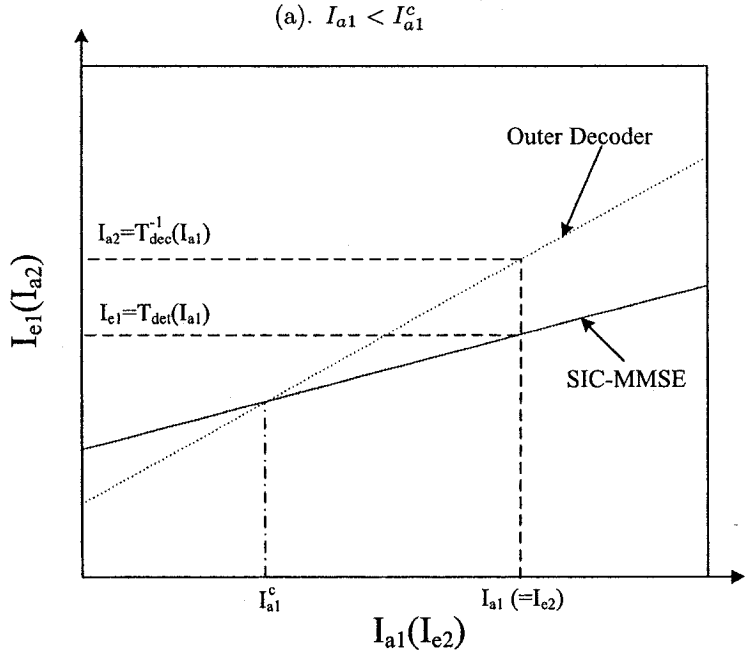
$$I_{p,2s} = I_{e1} = T_{det}(I_{a1}) \tag{5.23}$$

where $T_{det}()$ is the EXIT function of the conventional SIC-MMSE detector. From (5.22) and (5.23), it suffices to compare $T_{dec}^{-1}(I_{a1})$ and $T_{det}(I_{a1})$ in order to compare $I_{e1,2s}$ and $I_{e1,r}$. As can been seen from Fig. 5.7, $T_{dec}^{-1}(I_{a1}) < T_{det}(I_{a1})$ when $I_{a1} < I_{a1}^{c}$ and $T_{dec}^{-1}(I_{a1}) > T_{det}(I_{a1})$ otherwise. With this fact, (5.21) is proven. ■

Proposition 4 shows that the 2-staged detector outperforms the recursive detector before the convergence point of the conventional SIC-MMSE-based turbo receiver but under-performs after. Furthermore, by Proposition 3, the recursive SIC-MMSE detector consistently outperforms the conventional SIC-MMSE detector and, hence, yields higher convergence point. This with Proposition 4, shows that the recursive SIC-MMSE detector leads to a convergence point higher than that of the 2-staged SIC-MMSE detector.



(a). $I_{a1} < I^c_{a1}$



(b). $I_{a1} > I^c_{a1}$

Figure 5.7: The preliminary information of the proposed detector for a given *a priori* information I_{a1}.

Noting that the complexity of the recursive detector is comparable to that of the conventional SIC-MMSE detector, while the complexity of the 2-staged detector is about twice of theirs. Since the 2-staged detector performs the best at early iterations but is outperformed by the recursive SIC-MMSE detector at later iterations. These suggest a hybrid implementation in which the 2-staged detector is used at early iterations and the recursive SIC-MMSE detector is used at later iterations. Such a hybrid implementation yields the same convergence as that of the recursive detector but with less computation.

5.5 Simulation Results and Discussion

In this section, simulation results will be provided to verify previous analytical results and to demonstrate the merits of the proposed schemes. In the following simulations, a channel is said to be a "quasi-static" fading channel if it keeps constant in a coding frame but changes from frame to frame independently, while a channel is a "fast" fading channel if it keeps constant in one symbol interval but varies from signal interval to signal interval independently. S-random interleavers [72] with the maximum S-factor depending on the coding frame length were employed. Since the natural mapping was proved to be the optimum for BICM systems [67], QPSK constellation with natural mapping was employed.

First, we compare the trajectories of the turbo receivers that use conventional SIC-MMSE, the 2-staged and the recursive detectors, respectively. To do so, we assumed a constant 3×3 channel defined in the form as

$$\mathbf{H}\mathbf{H}^H = \frac{3}{1+\alpha+\alpha^2}\begin{pmatrix} 1 & \alpha^{\frac{1}{2}}\rho & \alpha\rho^2 \\ \alpha^{\frac{1}{2}}\rho & \alpha & \alpha^{\frac{3}{2}}\rho \\ \alpha\rho^2 & \alpha^{\frac{3}{2}}\rho & \alpha^2 \end{pmatrix} \tag{5.24}$$

where $\rho = \cos\frac{\pi}{4}$ and $\alpha = 0.3$. In the simulation, $E_b/N_0 = 1.5dB$ and the coding frame length was set to be 10002 bits. A convolutional code with coding rate 1/2 and constraint length 3 is used as the outer code. Its generator polynomials are $H^1(D) = [0100]_B$ and $H^2(D) = [1011]_B$. The constraint length of 3 was chosen to ensure a less "steep" slope of the EXIT function of the outer decoder for better illustration. The trajectories of the three "free-running" turbo receivers [38] are provided in Fig. 5.8. Also plotted in the figure are the EXIT functions of the outer decoder, the conventional SIC-MMSE and the 2-staged detectors. The EXIT functions of the conventional SIC-MMSE and the 2-staged detectors are obtained by randomly generating Gaussian distributed LLRs as the *a priori* input. Noting that the output symbol estimates of the generalized SIC-MMSE detectors are Gaussian random variables. Instead of using the consistent Gaussian bit LLRs [38], symbols corrupted by Gaussian noise were used to generate bit LLRs, which then were used as the input of the outer decoder. By doing so, the effects of the constellation mapping are also taken into account. As can be seen, the trajectories of the conventional SIC-MMSE-based turbo receiver and the turbo receiver with the 2-staged detector match well with the corresponding EXIT curves at early iterations. Notable discrepancy between the actual trajectories and the EXIT chart predictions can be seen at later iterations. This is expected because as the iterative process proceeds, the correlations among *extrinsic* LLRs increase due to finite interleaver size [38]. Specifically, the turbo receivers with the two proposed detectors take 6 iterations to converge. The decoder *extrinsic* information is about 0.9319 and 0.9330 for the 2-staged detector and the recursive detector, respectively. The conventional SIC-MMSE-based receiver takes at least 8 iterations to converge with decoder *extrinsic* information at about 0.8783. It can also be seen that the recursive detector yields higher *extrinsic* information than the 2-staged detector after conventional SIC-MMSE-based receiver reaches convergence. This agrees with Proposition 4.

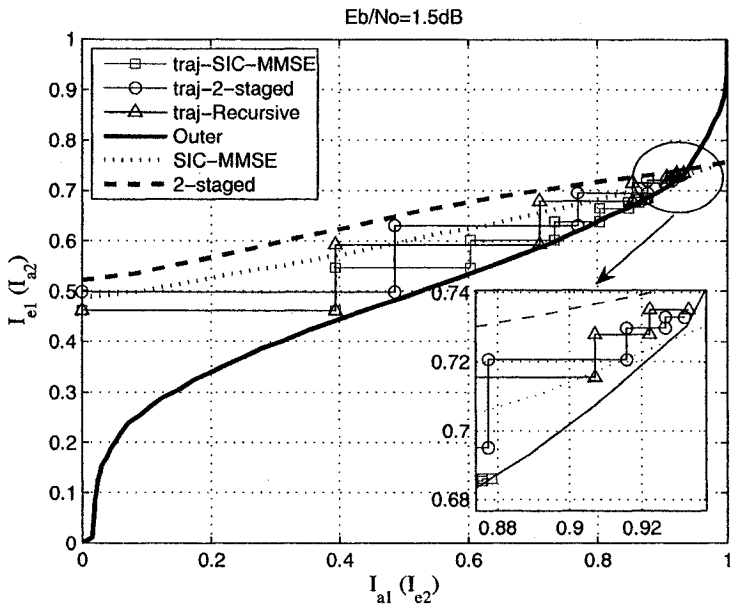


Figure 5.8: Trajectories of three schemes for a constant channel.
Note: "traj" denotes trajectory.

The error-rate performance of the three turbo receivers over fading channels are compared in Fig. 5.9 for quasi-static fading channels and in Fig. 5.10 for fast fading channels. In these simulations, a convolutional code with a coding rate 1/2 and constraint length 5 is used as the outer encoder. Its generator polynomials are $H^1(D) = [101011]_B, H^2(D) = [111101]_B$. In order to simulate a large number of frames, the coding frame length was set to be 200 for 4×4 channels and 198 for 3×3 channels.

In Fig. 5.9, the frame error rate (FER) curves of the three schemes over 3×3 and 4×4 quasi-static fading channels are plotted. As can be seen, both of the proposed receivers require 5 iterations for the 3×3 channel or 8 iterations for 4×4 channel to converge, less than the conventional SIC-MMSE-based receiver. At FER=0.002, the performance gain of the proposed 2-staged receiver is about $0.9dB$ over the 3×3 channel and $1.0dB$ over the 4×4 channel; while the performance gain of the proposed recursive receiver is about $1.3dB$ over the 3×3 channel and $1.1dB$ over the 4×4

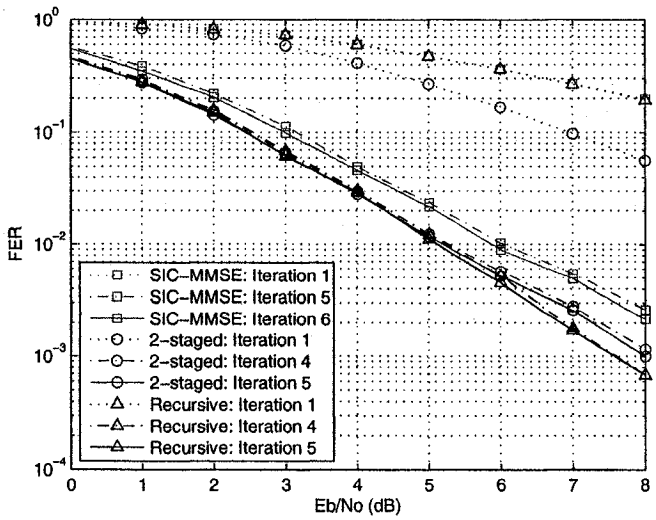
channel.

In Fig. 5.10, the bit-error rate (BER) curves of the three schemes over fast fading channels are plotted. Again, the two proposed receivers converge with less iterations, i.e., 5 iterations. At BER=2×10^{-5}, each of the two proposed turbo receivers has a performance gain about $0.9dB$ over the 3×3 channel and $0.6dB$ over the 4×4 channel.

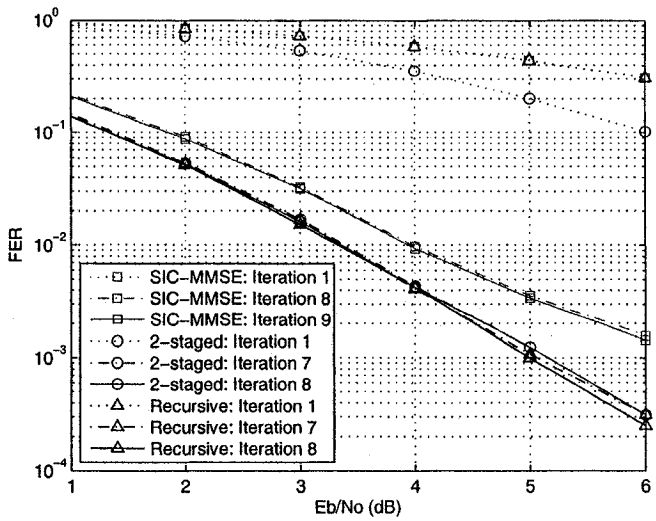
In summary, the simulation results show that the turbo receivers with the proposed detectors converge at higher point with less iterations. Additionally, the turbo receiver with the recursive detector outperforms the turbo receiver with the 2-staged detector with much reduced complexity.

5.6 Conclusions

A generalized SIC-MMSE detector scheme has been proposed to exploit the information about the interference existing in channel observation. A nonlinear MMSE estimator has been developed in the proposed generalized SIC-MMSE detection. The nonlinear estimator combines the information provided by the outer decoder and the channel observation as well in estimating interfering symbols. Depends on the methods for generating preliminary estimates from the channel observation, two generalized SIC-MMSE detectors were proposed. They are the 2-staged SIC-MMSE and the recursive SIC-MMSE detectors. Analysis study including an EXIT chart analysis indicates that both of the proposed detectors significantly outperform the conventional SIC-MMSE detector. Furthermore, a hybrid implementation that uses the 2-staged detector at early iterations and the recursive detector at later iterations yields the best error performance with least iterations. Simulation results have also been provided to verify the analytical results.

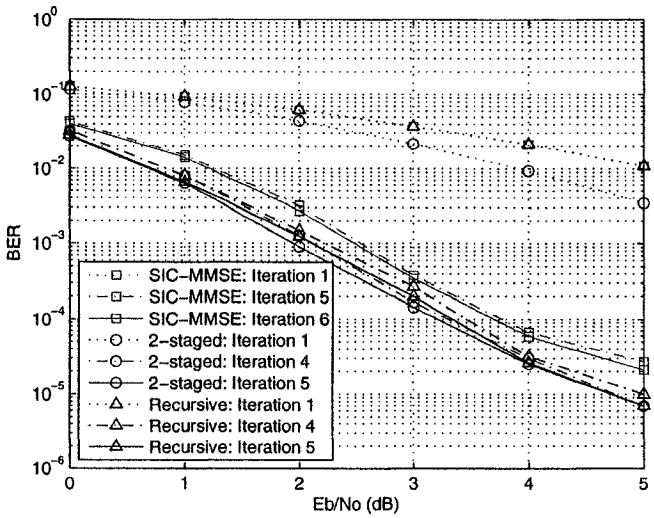


(a). 3×3 channel

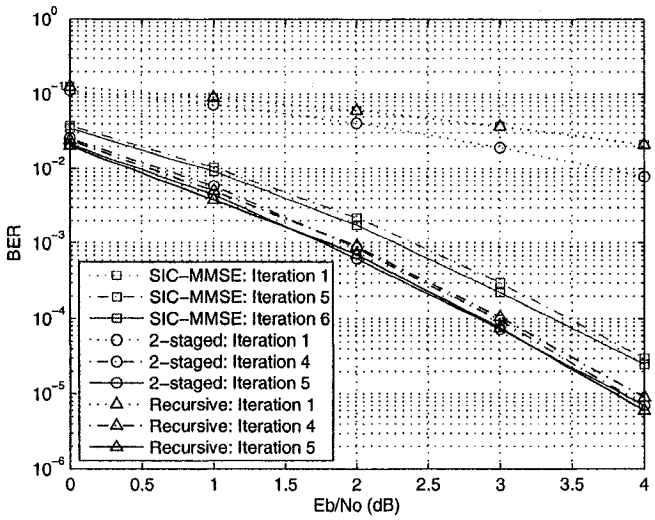


(b). 4×4 channel

Figure 5.9: Error-rate performance comparison of the three schemes under quasi-static Rayleigh flat fading channels.



(a). 3×3 channel



(b). 4×4 channel

Figure 5.10: Error-rate performance comparison of the three schemes under fast Rayleigh flat fading channels.

Chapter 6

Design Tradeoffs in Turbo MIMO Transmitter

6.1 Introduction

In the previous chapters, we have investigated the design of turbo MIMO transmitters and receivers, respectively. Here, we discuss the implementation issues of the proposed turbo MIMO transmitter, particularly, tradeoffs among system parameters.

The turbo MIMO transmitter in Chapter 3 consists of four components, i.e., outer encoder, interleaver, constellation mapper and inner ST modulator, as shown in Fig. 6.1. Note, this generalized model subsumes many existing MIMO transmission schemes as its special cases. In such a turbo transmitter, each component affects the overall error performance by providing various protections.

Outer encoder: The outer encoder is used to provide coding gain and time diversity.

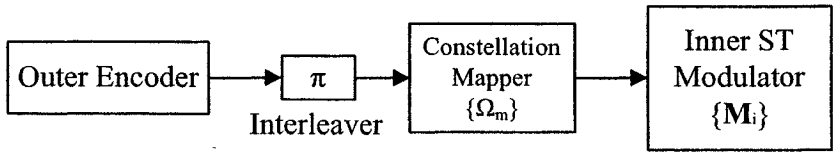


Figure 6.1: Tradeoffs in the turbo transmitter.

Often a conventional binary code with a trellis structure is preferred. The error performance of such a binary code can be predicted by the *minimum free distance* [31], which, for a given constraint length, largely depends on *coding rate* together with generator polynomials. The coding rate of the outer encoder is $R_c \leq 1$.

Interleaver: The use of the interleaver is to decouple the correlation between the outer encoding and the inner ST modulation. The interleaver avoids low-weight codewords by shuffling the coded bits properly. As such, a "good" interleaver should separate the neighboring coded bits from the outer encoder as far as possible. This requires a large interleaving depth. Good interleavers are available, e.g., the S-random interleaver in [72]. Note, the interleaver does not affect the overall data rate but it affects the decoding delay at the turbo receiver.

Constellation mapper: The constellation mapper maps several bits into one symbol. For a given SNR, the minimum Euclidean distance between transmitted symbols depends on the constellation size 2^Q. Note, for iterative receiver, the constellation mapping pattern also affects the error performance [67].

Inner ST modulator: The inner ST modulator is defined by its L $N_t \times T$ dispersion matrices $\{\mathbf{M}_i\}$. For the turbo MIMO transmitter, design criteria for the inner ST modulator are proposed in Chapter 3: *Capacity Criterion* and *Error-Performance Criterion.* These two criteria correspond to the channel capacity and error performance, respectively, and together reflect a joint optimization of both data rate and error rate. With a constellation of size 2^Q, the modulation bit rate of the inner modulator is $Q \cdot L/T$ bits per channel use. The modulation symbol rate is defined as L/T. Note, a subset of dispersion matrices should be selected from the optimal design if the bit rate of the inner modulator is smaller than $N_t Q$ with chosen constellation size 2^Q.

From Chapter 3, the overall data rate of the turbo MIMO system is $R = R_c R_m = R_c \cdot Q \cdot L/T$. In reality, the choice of the overall data rate will be set according to the channel capacity for given SNR region (refer to equation (2.2) and (2.3) in Chapter 2). For a fixed overall data rate R, rate allocation among the outer encoder, the constellation mapper and the inner ST modulator can affect the overall error performance at the turbo receiver. Hence, it is important to study rate tradeoffs among these rate-related components.

6.2 Design Tradeoffs

Below, tradeoffs among the coding rate R_c, constellation size 2^Q, and the symbol rate L/T are discussed for a turbo MIMO transceiver with N_t transmit antennas and N_r receiver antennas. In the following discussions, instead of exhaustive error-rate simulations, the technique of the EXIT chart [37][38] is applied.

6.2.1 Coding Rate R_c vs. Constellation Size 2^Q

To evaluate the tradeoff between coding rate R_c and constellation size 2^Q, the corresponding EXIT characteristics are studied first.

The EXIT function of an ideal code of rate R_c with infinite constraint length would be given by

$$I_{e2} = \begin{cases} 0 & 0 \leq I_{a2} < R_c \\ 1 & R_c \leq I_{a2} \leq 1 \end{cases} \tag{6.1}$$

Apparently, practical codes of finite constraint length would show similar behavior. EXIT functions of several codes with various coding rates and constraint lengths are plotted in Fig. 6.2. For coding rate $R_c = 1/2$, the generator polynomials for constraint length 3 and 4 are $H^1(D) = [101]_B, H^2(D) = [111]_B$, and $H^1(D) = [1101]_B, H^2(D) = [1111]_B$, respectively. For coding rate $R_c = 1/3$, the generator polynomials for con-

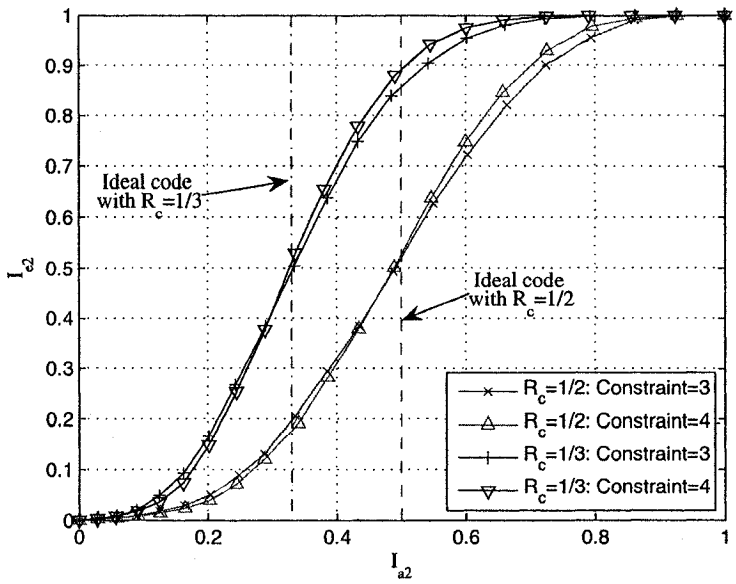
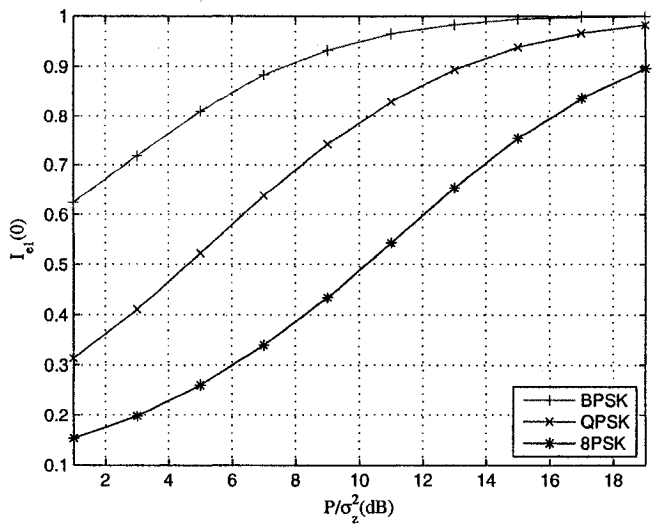


Figure 6.2: EXIT functions of the outer decoders with various coding rates and constraint lengths.

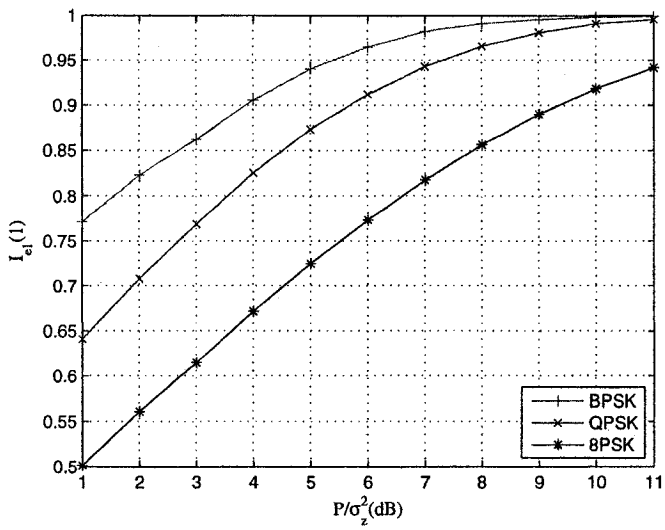
Note: "Constraint" indicates the constraint length.

straint length 3 and 4 are $H^1(D) = [101]_B, H^2(D) = [111]_B, H^3(D) = [111]_B$, and $H^1(D) = [1011]_B, H^2(D) = [1101]_B, H^3(D) = [1111]_B$, respectively. It is interesting to note that the inflection point of each outer code is approximately its coding rate. Specifically, the inflection point of $R_c = 1/2$ is about 0.5, while the inflection point of $R_c = 1/3$ is about 0.33. It is also interesting to note, for a given coding rate, a larger constraint length often indicates a sharper cliff in the middle, i.e., closer to the EXIT function of the ideal code.

For the inner MUD, the *extrinsic* outputs at two extreme cases, i.e., with no *a priori* information $I_{e1}(0)$, and with perfect *a priori* information $I_{e1}(1)$, are studied as in Chapter 3. Here, Scheme 1 with $L = 4$ and $N_t = N_r = T = 2$ in Chapter 3 is used as the inner modulation. Its symbol rate is $L/T = 2$. In Fig. 6.3, the *extrinsic* outputs at these two extreme cases are plotted for various constellations over fading channel. In the figure, P/σ_z^2 is the SNR per channel use at each receive antenna, defined in equation (3.2) of Subsection 3.2.1.



(a). with no *a priori* information



(b). with perfect *a priori* information

Figure 6.3: *Extrinsic* outputs at two extreme cases for various constellations.

With the above EXIT characteristics, one can evaluate the tradeoff between coding rate R_c and constellation size 2^Q. For instance, given overall rate $R = 2$, we compare these two designs using Scheme 1. One design uses the outer code with coding rate $R_c = 1/2$ and constraint length 4, and QPSK constellation. The other uses the outer code with coding rate $R_c = 1/3$ and the same constraint length, and 8PSK constellation. One can predict from Fig. 6.2 and 6.3 that, after $P/\sigma_z^2 = 3dB$, the second design can easily pass though the cliff of its outer decoder in the middle and reach the second plateau (right) and thus outperforms the first design due to its more powerful outer code with $R_c = 1/3$. On the other hand, when SNR is less than $2dB$, the second design will not converge at the right plateau and, hence, the first design is preferred Using the 2-staged scheme as the inner MUD, the EXIT charts of these two designs are compared for $P/\sigma_z^2 = 2dB$ and $P/\sigma_z^2 = 3dB$ in Fig. 6.4. As can be seen from the figure, the above prediction is confirmed.

In summary, given overall data rate R, modulation symbol rate L/T, and SNR, a pair of coding rate R_c and constellation size 2^Q shall be chosen such that

- $R_c \cdot Q = R \cdot T/L$
- $I_{e1}(0)$ is approximately equal to or slightly smaller than R_c.

6.2.2 Coding Rate R_c vs. Modulation Symbol Rate L/T

From Theorem 1 in Chapter 3, it is shown that the channel capacity of $\widetilde{\mathbf{H}}$ is monotonically increasing with respect to L/T. Hence, for given overall data rate R and constellation size 2^Q, the maximal $L/T = N_t$ shall be chosen to maximize the capacity. For instance, $R = 1$ and QPSK constellation are assumed. The following two designs are compared. One design uses Scheme 1 with $L = 3$ and $T = 2$ and the outer code with coding rate $R_c = 1/3$ and constraint length 3. The other design uses the Alamouti scheme (i.e., $L = 2$ and $T = 2$) and the outer code with coding

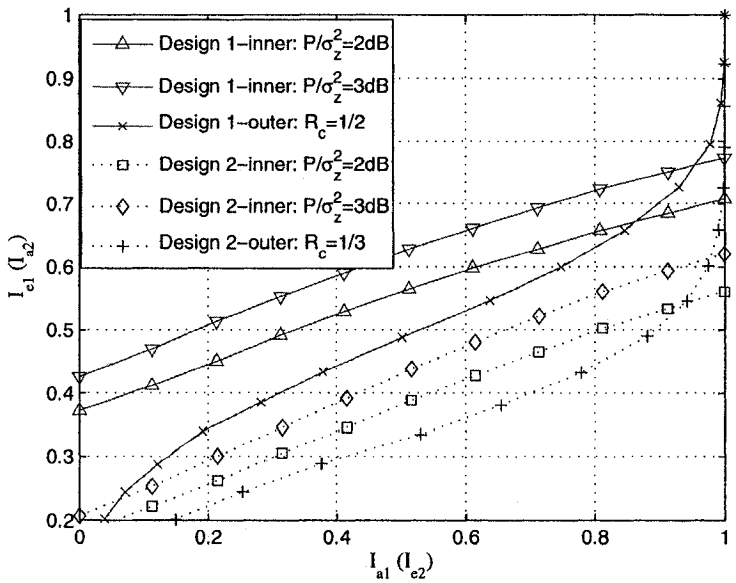


Figure 6.4: EXIT charts for the two designs

rate $R_c = 1/2$ and constraint length 3. Note, for given $L = 2$ and $T = 2$, the inner modulator reduces to an orthogonal design. Since the EXIT function of the recursive scheme in Chapter 5 can not be obtained directly, EXIT functions of the SIC-MMSE and 2-staged schemes are provided for design reference. In Fig. 6.5, the EXIT charts of the two designs are provided. As can be seen, the EXIT curves of the inner MUD for the second design are higher than the corresponding curves for the first design but all of them fall at the same point when *a priori* information is perfect, i.e., the same $I_{e1}(1)$. Still, the first design outperforms the second one due to the more powerful outer code with $R_c = 1/3$.

In summary, for given overall data rate R and constellation size 2^Q, it is preferred to choose the maximum symbol rate L/T so that the corresponding coding rate R_c is minimized.

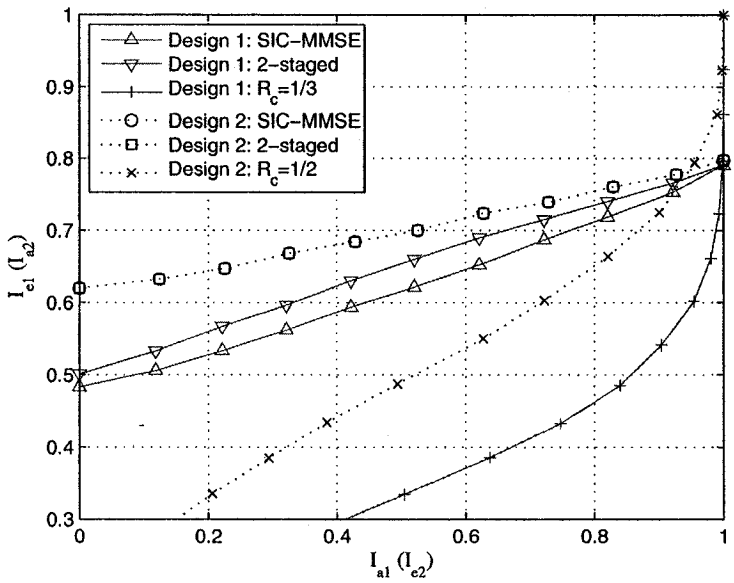
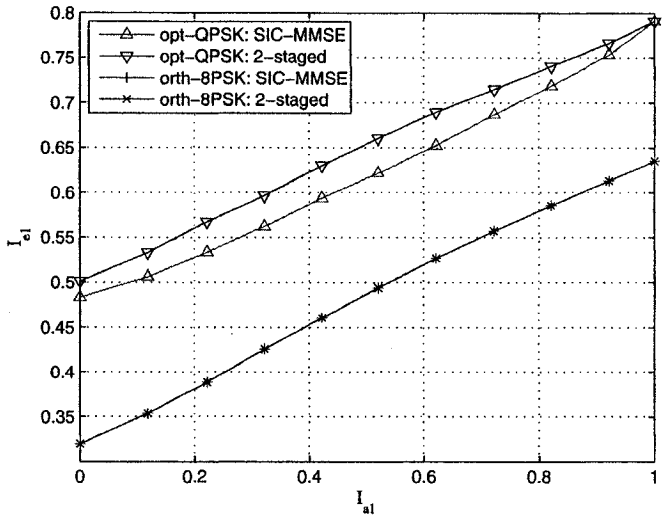


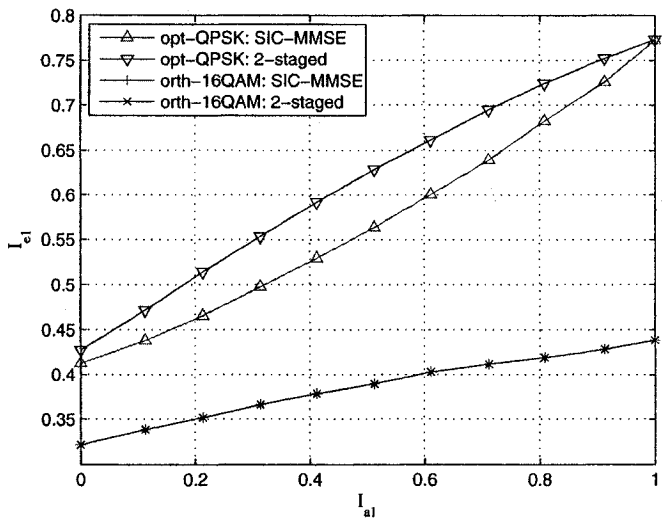
Figure 6.5: EXIT chart for rate tradeoffs between coding rate R_c and symbol rate L/T.

6.2.3 Modulation Symbol Rate L/T vs. Constellation Size 2^Q

The rate tradeoff between modulation symbol rate L/T versus constellation size 2^Q has been studied in Subsection 3.4.4. It is shown that, for a given inner ST modulation rate $R_m = Q \cdot L/T$, one seeks to choose a constellation size as small as possible till the modulation rate in symbol reaches N_t. Here, EXIT functions of the inner MUD are provided to demonstrate this point further. In this simulation, EXIT functions of Scheme 1 and Alamouti scheme are compared at $P/\sigma_z^2 = 3dB$. The corresponding EXIT functions are presented in Fig. 6.6 for $R_m = 3$ and $R_m = 4$ bits per channel use, respectively. As can been seen from these figures, the optimal Scheme 1 outperforms the Alamouti scheme significantly. This suggests that, for given modulation rate R_m, a minimal integer Q should be selected to satisfy $Q \geq R_m/N_t$.



(a). $R_m = 3$



(b). $R_m = 4$

Figure 6.6: EXIT functions of the inner MUD for $R_m = 3$ and $R_m = 4$.
Note: "opt" indicates the optimal design, i.e., Scheme 1; "orth" indicates Alamouti scheme.

6.3 Conclusions

In this chapter, the effects on overall error performance at the turbo receiver are discussed for each constituent component of the turbo transmitter. The tradeoffs among the three rate-related components are studied. It is shown that, in general, the maximum symbol rate $L/T(\leq N_t)$ shall be chosen. In addition, the coding rate R_c and constellation size 2^Q should be chosen such that $I_{e1}(0) \approx R_c$.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

In this thesis, we proposed an efficient turbo MIMO transceiver framework that is capacity-achieving and yet with reasonable complexity. First, the design of the turbo MIMO transmitter was studied and design criteria were developed for the inner ST modulator. Secondly, a new efficient stand-alone MUD has been proposed. Based on this, a generalized SIC-MMSE soft MUD was developed for the turbo MIMO receiver. Finally, rate tradeoff issues were discussed and design guidelines were provided.

7.1.1 Design of Turbo MIMO Transmitter

The turbo MIMO transmitter is a serial concatenation of conventional outer encoder(s) and an inner ST modulator. The abundance of conventional codes for the outer encoder allows us to focus on the design of the inner ST modulation, which is also referred to as CSTM.

In Chapter 3, by extending the EXIT chart technique to turbo MIMO systems, the convergence behavior of the turbo MIMO receiver was studied. Based on this study, two new design criteria were developed, i.e., *Channel Capacity and Error-*

Performance Criteria. It was shown that the inner space-time modulator shall (a) maximize the average mutual information between a bit and the received signal and (b) minimize the pair-wise error performance of the codeword pairs that differ at only one symbol within a modulation block. These two design criteria together reflect a joint optimization of data rate and error rate. To guarantee the convergence of the iterative receiver, the channel capacity of the inner ST modulations must be maximized. Once convergence is achieved, the error performance is optimized by maximizing the rank and determinant of the dispersion matrix of each individual symbol. The latter *Error-Performance Criterion* is much easier to apply than the well-known Tarokh's *Rank and Determinant Criteria.*

With the new design criteria, an optimal LD scheme was proposed. Simulation results were provided to verify the design criteria and to demonstrate the merits of the proposed inner modulator.

7.1.2 Stand-Alone PIC MUD

Suboptimal MUD schemes based on interference cancellation are preferable in both multi-user communications and single-user MIMO communications due to their low complexity and good error performance.

In Chapter 4, a new MUD scheme was proposed, which can be regarded as a partial PIC scheme. The crucial component in the proposed scheme is a nonlinear MMSE estimator employed as soft decision device. By exploiting the knowledge of symbol alphabet, the nonlinear estimation was developed in sense of MMSE. The proposed nonlinear estimator generates more accurate estimates of interfering symbols and thus leads to better linear estimates at next PIC stage. Since the nonlinear estimation involves little computation, the overall computational complexity is close to that of the conventional PIC schemes. Simulations have been carried out and the results have demonstrated that the proposed scheme significantly outperforms the

linear MMSE detector and other recent partial PIC schemes.

7.1.3 Turbo MIMO Receiver

In Chapter 5, a generalized SIC-MMSE detection has been proposed to be used as the inner soft-output detector. The new detector exploits the useful information provided by channel observation and *a priori* information fed back from the outer decoder as well. A nonlinear MMSE estimator has been developed for the proposed generalized SIC-MMSE detection. The nonlinear estimator is the key component that combines *a priori* information and a so-called preliminary estimate gleaned from the channel observation together in estimating interfering symbols. Depending on the methods to obtain preliminary estimates from the channel observation, we proposed two generalized SIC-MMSE detectors. They are the 2-staged SIC-MMSE and the recursive SIC-MMSE detectors. Analytical study including an EXIT chart analysis indicated that both of the proposed detectors significantly outperform the conventional SIC-MMSE detector with reduced complexity. Motivated by the different behaviors of the two proposed detectors, a hybrid scheme that uses the 2-staged detector at early iterations and the recursive detector at later iterations can be implemented to yield the best error performance with least iterations. Simulation results have also been provided to demonstrate the merits of the proposed schemes.

7.1.4 Design Tradeoffs in Turbo MIMO Transmitter

In reality, the choice of the overall data rate will be set according to the channel capacity for a given SNR region. Given overall data rate, rate allocation among constituent components at the transmitter can affect the overall error performance at the receiver.

In Chapter 6, the effects of the transmitter components on overall error performance were presented. For given overall data rate, the rate tradeoffs among three

rate-related parameters were studied. Design guidelines have been provided.

7.2 Future Works

Based on our study in the filed of turbo MIMO transceiver. Two future research topics can be identified.

- Our results are developed on the assumptions that 1) channel is spatially uncorrelated and 2) CSI is perfectly known at the receiver but unknown to the transmitter. However, in practice, the above assumptions can not be satisfied all the time. Future research works could include 1) extend this study to spatially correlated channel and 2) other CSI scenarios, particularly when the CSI is partially known at the receiver.

- The effects of constellation mapping has been discussed in this study. However, it is well-known that the mapping affects the overall error performance. A future study could be conducted to find the optimal constellation mapping pattern.

Bibliography

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol 10, pp. 585-595, Nov. 1999.

[2] G. J. Foschini, M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311-335, 1998.

[3] D. Gesbert, M. Shafi, D. S. Shiu, P. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, Special Issue on MIMO Systems, pt. I, vol. 21, pp. 281-302, Apr. 2003.

[4] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, Mar. 1998.

[5] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

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

[7] Y. Liu and M. Fitz, "space-time turbo codes," *13th Annual Allerton Conf. on Commun. Control and Computing*, Sep. 1999.

[8] D. Cui and A. M. Haimovich, "Design and performance of turbo space-time coded modulation," *IEEE GLOBECOM'00*, vol. 3, pp1627-1631, Nov. 2000.

[9] D. Tujkovic, "Recursive space-time trellis codes for turbo coded modulation", *Proc. of GlobeCom 2000*, San Francisco.

[10] B. M. Hochwald and S. ten Brink "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Comm.*, vol. 51, pp. 389-399, Mar 2003.

[11] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multiple antennas," *Bell labs. Tech. J.*,vol. 1, no. 2, pp. 41-59, 1996.

[12] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, pp. 14-16, Jan. 1999.

[13] H. El Gamal and A. R. Hammons Jr., "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2321-2334. Sep. 2001.

[14] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804-1824, July 2002.

[15] R. W. Heath and A. Paulraj, "Linear Dispersion Codes for MIMO Systems Based on Frame Theory," *IEEE Trans. on Signal Processing*, vol. 50, No. 10, pp. 2429-2441, October 2002.

[16] X. Ma and G. B. Giannakis, "Full-Diversity Full-Rate Complex-Field Space-Time Coding," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2917-2930, July 2003.

[17] L. Zhang and D. Tse, "Diversity and mutiplexing: A fundamental tradeoff in multiple antenna channels" *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073-96, May 2003.

[18] G. D. Forney Jr., "Concatednated Codes", *M. I. T. Press*, Cambridge, Massachusetts, 1966.

[19] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, pp. 1064-1070, Geneva, Switzerland, May 1993.

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

[21] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," in *Proc. Vehicle Technology Conf.*, pp. 473-478, Sept. 2000.

[22] A. van Zelst, R. van Nee, and G. A. Awater, "Turbo-BLAST and its performance," in *Proc. Vehicle Technology Conf.*, vol. 2, pp. 1282-1286, May 2001.

[23] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Proc.*, vol. 50, no. 10, pp. 2538-2546, Oct. 2002.

[24] S. Haykin, M. Sellathurai, Y. Jong and T. Willink, "Turbo-MIMO for wireless communications," *IEEE Comm. Magazine*, pp. 48-53, Oct. 2004.

[25] P. Robertson, E. Villebrun and P. Hoecher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. Int. Conf. Commun.*, pp.1009-1013, June 1995.

[26] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol 20, Issue: 2, pp.284-287, Mar. 1974.

[27] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Comm.*, vol. 47, pp. 1046-1061, July 1999.

[28] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean square error equalization using *a priori* information," *IEEE Trans. Signal Processing*, vol. 50, pp. 673-683, Mar. 2002.

[29] M. Tüchler, R. Koetter, and A. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Comm.*, vol. 50, No. 5. pp. 754-767, May. 2002.

[30] T. Rapaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2001

[31] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill.

[32] D. Tse, P. Viswanath and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1859-1874, Sep. 2004.

[33] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," in *Math. Comput.*, vol. 44, pp. 463-471, Apr. 1985

[34] Y. L. C. de Jong and T. J Willink, "Iterative tree search detection for MIMO

wireless systems," *Proc. VTC 2002*, Vancouver, BC, Canada, vol. 2, 24-28, pp. 1041-45, Sept. 2002.

[35] J. Hagenauer, P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," *GLOBECOM 89., IEEE*, vol. 3, pp. 47.1.1-47.1.7, Nov. 1989.

[36] B. Vucetic and J. Yuan, *Turbo Codes: Principles and Applications*, Kluwer, 2000

[37] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, no. 13, pp. 1117-1118, Jun. 1999.

[38] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 40, pp. 1727-1737, Oct. 2001.

[39] S. ten Brink, "Exploiting the chain rule of mutual information for the design of iterative decoding schemes," in *Proc. 39th Allerton Conf. on Communications, Control and Computing*, Monticello, Urbana-Champaign, Oct. 2001.

[40] Z. Wu and X. F. Wang, "Design of coded space-time modulation," *IEEE International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 2, pp. 1059-1064, Jun. 13-16, 2005.

[41] Z. Wu and X. F. Wang, "A new concatenated encoding and space-time modulation scheme for MIMO wireless communications," Accepted by *IEE Processings Communications*, Dec. 2006.

[42] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Info. Theory*, vol. 32, pp. 85-96, Jan. 1986.

[43] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[44] R. Lupas, S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. inform. Theory*, vol. 35, pp. 123-136, Jan. 1989.

[45] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*,, vol. 8, pp. 641-649, May 1990.

[46] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Select. Areas Commun.*, vol. 12, no. 5, pp. 796-807, June 1994,

[47] G. J. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841-1852, Nov. 1999.

[48] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509-519, Apr. 1990.

[49] D. Divsalar, M. K. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 258-268, Feb. 1998.

[50] R. M. Buehrer and S. P. Nicoloso, "Comments on partial parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 658-661, May 1999.

[51] W. H. Chin, A. G. Constantinides, and D. B. Ward, "Parallel multistage detection for multiple antenna wireless systems," *Electron. Lett.*, vol. 38, no. 12, pp. 597-599, Jun. 2002.

[52] S. Tseng and H. Lee, "An adaptive partial parallel multistage detection for MIMO systems," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 587-591, Apr. 2005.

[53] L. K. Rasmussen and I. J. Oppermann,"Ping-pong effects in linear parallel interference cancellation for CDMA," *IEEE Trans. Wireless Commn.*, vol. 2, no. 2, pp. 357-363, Mar. 2003.

[54] D. R. Brown III, "Multistage parallel interference cancellation: convergence behavior and improved performance through limit cycle mitigation," *IEEE Trans. Signal Processing*, vol. 53, pp. 283-294, Jan. 2005.

[55] Z. Wu and X. F. Wang, "Parallel Interference Cancellation Aided By Nonlinear MMSE Estimation," *The International Association of Science and Technology for Development (IASTED) The 4th International Conference on Communication, Internet and Information Technology (CIIT 2006)*, Nov. 2006.

[56] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[57] Z. Wu and X. F. Wang, "A new symbol-alphabet-aware multi-user detection scheme," *IEEE Comm. Lett.*, vol. 10, pp. 582 - 584, Issue 8, Aug. 2006.

[58] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. inform. Theory*, vol. 43, no. 3, pp. 858-871, May 1997.

[59] A. Amari, *Differential-Geometrical Methods in Statistics*, Springer-Verlag, New York, 1985.

[60] A. Lampe and J. Huber, "On improved multiuser detection with iterated soft decision interference cancellation", In *Proc. of CTMC at Int. Conf. Comm. (ICC'99)*, pp. 172-176, June 1999, Vancouver, Canada.

[61] C. Sgraja, W. G. Teich, A. Engelhart and J. Lindner, "Multiuser/multisubchannel detection based on recurrent neural network structures for linear modulation schemes with general complex-valued symbol alphabet", *Proc. COST 262 Workshop on Multiuser Detection in Spread Spectrum Communications*, pp. 45-52, Jan. 2001, Schloss Reisensburg, Germany.

[62] X. F. Wang, W.-S. Lu, and A. Antoniou, "Constrained minimum-BER multiuser detection," *IEEE Trans. Signal Processing*, vol. 48, pp. 2903-2909, Oct. 2000.

[63] Y. Zhang, "Reduced complexity iterative multiuser detection for DS/CDMA with FEC," in *Int. Conf. on Univ. Personal Comm.*, no. 12, (San Diego, U.S.A), pp. 1014, Oct. 1997.

[64] P. D. Alexander, A. J. Grant, and M. C. Reed, "Performance analysis of an iterative decoder for code-division mulitple-access," *European Trans. Telecomm.*, vol. 9, pp. 419426, Sept./Oct. 1998.

[65] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linkping Univ., Sweden, 1996.

[66] Z. Wu and X. F. Wang, "A New Turbo Multi-User Receiver with Recursive SIC-MMSE Detection," Accepted by *IEEE Wireless Communications and Networking Conference 2007*, Dec. 2006.

[67] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," in *Proc. Int. Conf. Communications*, pp. 858-862, June 1999.

[68] A W. Marshall and I. Olkin, *Ineqalities: Theory of Majorization and Its Application*, Academic Press, Inc. (London) Ltd., 1979.

[69] H. Boche and E. A. Jorswieck, "On Schur-convexity of expectation of weighted

sum of random variables with applications", *Journal of Inequalities in Pure Applied Mathematics*, vol. 5, Issue 2, Article 46, 2004.

[70] M. E. Bock, P. Diaconis, F. W. Huffer and M. D. Perlman, "Inequalities for linear combinations of Gamma random variables", *Canada J. Statistics*, vol. 15, pp. 387-395, 1987.

[71] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymtotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1772-1793, Jul. 2002.

[72] S. Dolinar and D. Divsalar, "Weight distribution for turbo codes using random and nonrandom permutations," *JPL Progress Report* 42-122, Aug. 15, 1995.

[73] M. Tüchler, S. ten Brink and J. Hagenauer, "Measures for tracing convergence of iterative decoding algorithms," in *Proc. 4th Int. ITG Conf. Source and Channel Coding*, Berlin, Germany, Jan. 2002.

[74] B. Scanavino, G. Montorsi and S. Benedetto, "Convergence properties of iterative decoders working at bit and symbol level," in *Proc. 2001 IEEE Global Telecommunications Conf. (GLOBECOM 01)*, vol. 2, pp. 10371041, 2001.

[75] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[76] B. Vucetic and J. Yuan, *Space-Time Coding*, John Wiley & Sons Ltd, England, 2003.

[77] S. Lin, D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice Hall, New Jersey, 1983.

[78] T. Richardson and R. Urbanke, "Capacity of low density parity check codes

under message passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no.2, pp. 599-618, Feb. 2001.

[79] H. Dai and H. V. Poor, "Iterative space-time processing for multiuser detection in multipath CDMA channels," *IEEE Trans. Signal. Proc.*, vol. 50, no. 9, pp. 2216-2127, Sept. 2002.

[80] K. Li and X. Wang, "EXIT chart analysis of turbo multiuser detection," *IEEE Trans. Wireless Comm.*, vol. 4, no. 1, pp. 300-311, Jan. 2005.

Appendix A

Proof of Theorem 1

The maximization of $\tilde{C}$ in equation (3.14) is considered in terms of modulation matrix $\mathbf{G}$ and L (number of symbols per modulation block).

Denote $\mathbf{D}_t$ as the matrix consisting of rows and columns $(t-1)N_t+1$ to tN_t of $\mathbf{G}\mathbf{G}^H$, i.e., $\mathbf{D}_t$ is the t-th diagonal block of size $N_t \times N_t$ of $\mathbf{G}\mathbf{G}^H$, then we have

$$\tilde{C} \leq \frac{1}{T} \log \left[\prod_{t=1}^{T} \det \left(\mathbf{I}_{N_r} + \frac{T}{L} \frac{P}{N_t \sigma_z^2} \mathbf{H}\mathbf{D}_t\mathbf{H}^H \right) \right] \tag{A.1}$$

The equality in (A.1) holds on when $\mathbf{G}\mathbf{G}^H = \mathrm{diag}[\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_T]$, i.e., $\mathbf{G}\mathbf{G}^H$ is block diagonal. In such a case, we can further write

$$\tilde{C} \leq \frac{1}{T} \log \left[\prod_{t=1}^{T} \det \left(\mathbf{I}_{N_r} + \frac{T}{L} \frac{P}{N_t \sigma_z^2} C_t^2 \mathbf{H}\mathbf{H}^H \right) \right] \tag{A.2}$$

where $C_t^2 = \mathrm{tr}(\mathbf{D}_t)/N_t$. Note that to ensure that P is the average energy per channel use at each receive antenna, one needs to have $\mathrm{tr}(\mathbf{G}\mathbf{G}^H) = L$. Hence, $\sum_{t=1}^{T} C_t^2 = \sum_{t=1}^{T} \mathrm{tr}(\mathbf{D}_t)/N_t = L/N_t$. From (A.2), it can be checked that $\tilde{C}$ is maximized only when $C_t^2 = \frac{L}{N_t T}$ for any $t = 1, 2, \ldots, T$. Apparently, this is possible only if equation (3.15) is satisfied.

Appendix B

Proof of Theorem 2

To solve the problem, we will first introduce the following two definitions [68][69].

Definition 1: For two nonnegative real vectors $\mathbf{x}, \mathbf{y} \in \mathcal{R}_+^n$ with descending ordered elements, i.e. $x_1 \geq x_2 \geq \ldots \geq x_n$ and $y_1 \geq y_2 \geq \ldots \geq y_n$, if $\sum_{k=1}^{m} x_k \geq \sum_{k=1}^{m} y_k, m = 1, 2, \ldots, n-1$ and $\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k$, It is said that $\mathbf{x}$ majorizes $\mathbf{y}$ and written as $\mathbf{x} \succ \mathbf{y}$.

Definition 2: A real-valued function Ψ defined on $\mathcal{R}_+^n$ is said to be Schur-convex if

$$\mathbf{x} \succ \mathbf{y} \Longrightarrow \Psi(\mathbf{x}) \geq \Psi(\mathbf{y}). \tag{B.1}$$

If the *a priori* information is perfect, we can obtain from (3.6)

$$I_{e1}^{(i)}(1) = \frac{1}{Q} \log \left(1 + \frac{P}{N_t \sigma_z^2} \tilde{\mathbf{h}}_i^H \tilde{\mathbf{h}}_i \right) \tag{B.2}$$

As can be seen from (B.2), maximizing $P_r(I_{e1}^{(i)}(Q) \geq \rho)$ is equivalent to maximizing $P_r(\|\tilde{\mathbf{h}}_i\|^2 \geq \xi)$.

To maximize the outage probability $P_r(\|\tilde{\mathbf{h}}_i\|^2 \geq \xi)$, the following lemma is introduced [70]

Lemma 1: Consider Y_i, $, \forall 1 \leq i \leq n$, *i.i.d.* Gamma random variables with density $p(y) = e^{-\beta y} y^{\alpha-1} \beta^{\alpha} / \Gamma(\alpha)$. Let $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$ with θ_i for all i and $W =$

$\sum_{i=1}^{n} \theta_i Y_i$. $P_r(W \geq t)$ is Schur-convex in θ for

$$t \geq \frac{(n\alpha + 1)(\theta_1 + \theta_2 + \ldots + \theta_n)}{\beta} \tag{B.3}$$

Noting that $\tilde{\mathbf{h}}_i = \overline{\mathbf{H}}\mathbf{m}_i$, we have

$$\|\tilde{\mathbf{h}}_i\|^2 = \sum_{m=1}^{N_r} \mathbf{h}_m^H \mathbf{M}_i \mathbf{M}_i^H \mathbf{h}_m \tag{B.4}$$

where $\mathbf{h}_m$ is the m-th column vector of $\mathbf{H}^H$. The following decomposition is assumed

$$\mathbf{M}_i \mathbf{M}_i^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{B.5}$$

where $\mathbf{\Lambda} = \text{diag}[\lambda_n], \forall n = 1, 2, ..., N_t$. By Lemma 5 in [1], since $\mathbf{U}^H \mathbf{h}_m$ has the same distribution as $\mathbf{h}_m$, we need only consider $\mathbf{M}_i \mathbf{M}_i^H = \mathbf{\Lambda}$. Then we have

$$\|\tilde{\mathbf{h}}_i\|^2 = \sum_{n=1}^{N_t} \lambda_n \sum_{m=1}^{N_r} |h_{mn}|^2 \tag{B.6}$$

For the power constraint, we have $\sum_{n=1}^{N_t} \lambda_n = 1$. Without loss of generality, we assume $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N_t} \geq 0$. .

By *Lemma 2*, since the random variable $\sum_{m=1}^{N_r} |h_{mn}|^2$ in (B.6) is identically chi-square distributed with $2N_r$ degrees of freedom, the outage probability $P_r(\|\tilde{\mathbf{h}}_i\|^2 \geq \xi)$ is maximized if all the eigenvalues of $\mathbf{M}_i \mathbf{M}_i^H$ are equal. With $\text{tr}(\mathbf{M}_i \mathbf{M}_i^H) = 1$, it implies that $\lambda_i = 1/N_t, \forall 1 \leq i \leq N_t$.