CHANNEL ESTIMATION IN CODED MODULATION SYSTEMS

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

by

Zohreh Andalibi

Saskatoon, Saskatchewan, Canada

Permission to Use

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

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

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

Acknowledgments

It is a pleasure to thank many people during my studies who made this dissertation possible. A few words mentioned here cannot adequately express my appreciation.

I would like to express my deepest gratitude to my supervisor, Professor Ha H. Nguyen, whose expertise, understanding and patience throughout my research program added considerably to my graduate experience. I appreciate his vast knowledge and skills and it has been my honor to work under his supervision. It is difficult to overstate my gratitude to my other Ph.D. supervisor, Prof. Eric Salt. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make Ph.D. fun for me. Throughout my studies period, he provided encouragement, sound advice, good teaching, good company, and lots of good ideas.

I would also like to thank the other members of my committee, Professors Aryan Saadat Mehr, FangXiang Wu, Rajesh Karki, David M. Klymyshyn for the assistance they provided at all levels of the research project.

I also had a really meaningful time in Saskatoon with great pleasure and fun. This is because of my lab-mates: Tran, Son, Nam, Ha, Duy, Tung, Simin, Quang, Brian and Eric in Communications Theories Research Group (CTRG) and Telecommunication Research Labs (TRLabs).

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my husband and best friend, Atabak, without whose love, encouragement and editing assistance, I would not have finished this thesis.

Finally, I gratefully acknowledge the TRLabs and the Department of Electrical and Computer Engineering at the University of Saskatchewan for their financial supports of my studies.

Abstract

With the outstanding performance of coded modulation techniques in fading channels, much research efforts have been carried out on the design of communication systems able to operate at low signal-to-noise ratios (SNRs). From this perspective, the so-called iterative decoding principle has been applied to many signal processing tasks at the receiver: demodulation, detection, decoding, synchronization and channel estimation. Nevertheless, at low SNRs, conventional channel estimators do not perform satisfactorily. This thesis is mainly concerned with channel estimation issues in coded modulation systems where different diversity techniques are exploited to combat fading in single or multiple antenna systems.

First, for single antenna systems in fast time-varying fading channels, the thesis focuses on designing a training sequence by exploiting signal space diversity (SSD). Motivated by the power/bandwidth efficiency of the SSD technique, the proposed training sequence inserts pilot bits into the coded bits prior to constellation mapping and signal rotation. This scheme spreads the training sequence during a transmitted codeword and helps the estimator to track fast variation of the channel gains. A comprehensive comparison between the proposed training scheme and the conventional training scheme is then carried out, which reveals several interesting conclusions with respect to both error performance of the system and mean square error of the channel estimator.

For multiple antenna systems, different schemes are examined in this thesis for the estimation of block-fading channels. For typical coded modulation systems with multiple antennas, the first scheme makes a distinction between the iteration in the channel estimation and the iteration in the decoding. Then, the estimator begins iteration when the soft output of the decoder at the decoding iteration meets some specified reliability conditions. This scheme guarantees the convergence of the iterative receiver with iterative channel estimator. To accelerate the convergence process and decrease the complexity of successive iterations, in the second scheme, the channel

estimator estimates channel state information (CSI) at each iteration with a combination of the training sequence and soft information. For coded modulation systems with precoding technique, in which a precoder is used after the modulator, the training sequence and data symbols are combined using a linear precoder to decrease the required training overhead. The power allocation and the placement of the training sequence to be precoded are obtained based on a lower bound on the mean square error of the channel estimation. It is demonstrated that considerable performance improvement is possible when the training symbols are embedded within data symbols with an equi-spaced pattern. In the last scheme, a joint precoder and training sequence is developed to maximize the achievable coding gain and diversity order under imperfect CSI. In particular, both the asymptotic performance behavior of the system with the precoded training scheme under imperfect CSI and the mean square error of the channel estimation are derived to obtain achievable diversity order and coding gain. Simulation results demonstrate that the joint optimized scheme outperforms the existing training schemes for systems with given precoders in terms of error rate and the amount of training overhead.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APP	A Posteriori Probability
AWGN	Additive White Gaussian Noise
BER	Bit-Error-Rate
BICM	Bit-Interleaved Coded Modulation
BICM-ID	Bit-Interleaved Coded Modulation systems with Iterative Decoding
BICM-MIMO	Bit-Interleaved Coded Modulation in Multiple-Input Multiple-Output
CICE	Combined Iterative Channel Estimator
CRB	Cramer-Rao Bound
CSI	Channel State Information
DNA	Dispersive Nucleo Algebraic
DS-CDMA	Direct-Sequence Code Division Multiple Access
EXIT	EXtrinsic Information Transfer
FER	Frame-Error-Rate
FIM	Fisher Information Matrix
GPS	Global Positioning System
ICSI	Imperfect Channel State Information
ID	Iterative Decoding
LLR	Log-Likelihood Ratio
LS	Least Square

MAP	Maximum A Posteriori probability
MAP	Maximum A Posteriori
MI	Mutual Information
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
MRC	Maximum-Ratio Combining
MSE	Mean-Square Error
O-PPSAM	Optimal Precoded Pilot Symbol Assisted Modulation
OFDM	Orthogonal Frequency Division Multiplexing
PEP	Pairwise Error Probability
PPSAM	Precoded Pilot Symbol Assisted Modulation
PSAM	Pilot Symbol Assisted Modulation
QAM	Quadrature Amplitude Modulation
QPSK	quadrature phase-shift keying
SAICE	Switch-Augmented conventional Iterative Channel Estimator
SIED	Sparsely Interleaved Estimation and Decoding
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SNRs	Signal-to-Noise Ratios
SO-PPSAM	Sub-Optimal Precoded Pilot Symbol Assisted Modulation

SOVA	soft-output Viterbi algorithm
SSD	Signal Space Diversity
TDM	Time Division Multiplexed

1. Introduction

1.1 Introduction

In the last three decades, the explosive growth of mobile and wireless communications has radically changed human life. Nowadays, wireless communications cover a wide variety of applications, from cellular telephony to global positioning system (GPS).

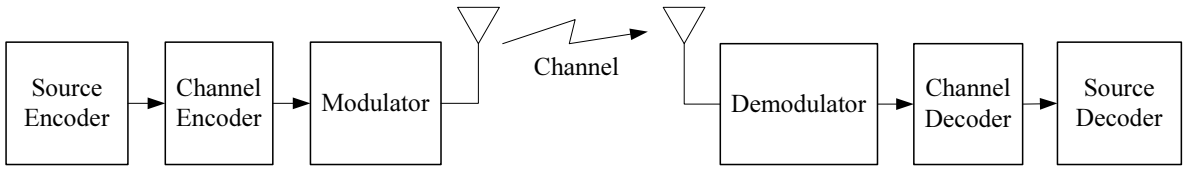

Figure 1.1 Block diagram of a digital wireless communication system.

A wireless communication system deals with information or data transmission from one point to another via the atmosphere or free space. Fig. 1.1 illustrates the basic elements of a digital wireless communication system. The source generates either analog signals such as speech, audio, image, and video, or digital data such as text or multimedia. The source encoder prepares the signal from the source in a compact digital form, typically in binary format. The generated binary data is then processed by a channel encoder so that the binary data sequences can be reliably reproduced at the receiver. The channel-encoded data stream is then modulated to generate waveforms for transmission over a physical channel such as a high-frequency radio link. Due to the open nature of the wireless channel, unfortunately, the channel is subject to various noise sources. At the receiver, the operations at the transmitter

are reversed so as to restore the original source information. The reversed operations include demodulator, channel decoder and source decoder.

An important requirement for providing reliable transmission technology for wireless applications is to have an accurate description of the wireless channel. In general the channel models strongly depend on the radio architecture [1].

Typically, wireless channels contain objects that scatter, reflect or diffract the transmitted signal. The effect is such that several delayed copies of the original transmitted signal arrive at the receiver. These effects may reduce the power and change the phase of the transmitted signal in different ways. There are two general aspects of such a power reduction that require separate treatments. One aspect is the large-scale effect which corresponds to the characterization of the signal power over large distances or the time-average behaviors of the signal. This is called attenuation or path loss and sometimes large-scale fading. The large-scale fading is more relevant to designing link budget. The other aspect is the rapid fluctuation in the amplitude and power of the signal and this is called small-scale fading, or just fading. It relates to the characteristics of the signal over short distances or short time intervals. In fact it is the rapid fluctuation of the amplitude of a radio signal over a short period of time that limits the transmission reliability over a wireless channel and motivates extensive research activities in the last two decades.

In short-scale, where the transmitted signal travels different paths, channels are called multipath channels. In this case, the channel can be modeled by a linear time-varying impulse response, $h(t, \tau)$, which may be approximated by a number of delta functions. Mathematically,

$$h(t, \tau) = \sum_{j=1}^{J} \alpha_j(t)\delta(\tau - \tau_j(t)), \tag{1.1}$$

where $\alpha_j(t)$ and $\tau_j(t)$ are, respectively, the overall attenuation and propagation delay at time t from the transmitter to the receiver on path j. The nature of the multipath channel is such that the amplitudes of these delta functions, $\alpha_j(t)$, are random processes. Further, each component fades independently, i.e., $\alpha_j(t)$ are independent. The

randomness of the channel impulse response mainly originates from the multipath and random locations of objects in the environment. Therefore fading is usually described in statistical terms that show the variations of the channel distortions over time and frequency. Indeed, different statistical models are needed to describe the variation behavior of the amplitude and phase of the received signal in different propagation environments.

How rapidly the channel fades is affected by how fast the receiver and/or transmitter are moving. Motion causes a Doppler shift in the received signal components. Based on this parameter, the channel can be considered fast or slow fading, which characterizes how rapidly the channel filter taps change over a symbol duration[1]. Therefore, whether the channel can be categorized as fast or slow fading depends on both the channel and the application. Regarding the channel model in (1.1), another important parameter is how many channel filter taps are necessary to adequately represent the wireless channel. Based on the underlying application, two typical cases of fading channels are:

- *Flat fading*: Here the channel is represented by a single tap. This applies when most of the paths arrive within one symbol time. This model is suitable for narrowband transmission.

- *Frequency-selective fading*: This applies to wideband channels, where the transmitted signal arrives over multiple symbol times. Equivalently, the channel needs to be represented by multiple taps.

The channel model in (1.1) is the analog model. An useful discrete-time baseband model of (1.1) can be obtained by applying the sampling theorem. Assume the transmitted signal is bandlimited. From (1.1), a discrete-time baseband model of the

[1]Symbol duration is the time duration to transmit one information symbol, or one constellation point in the signal constellation. For example, if the transmission rate is 2-Mbps, and if a rate-1/2 convolutional code and 16-quadrature amplitude modulation (16-QAM) constellation is used, then the symbol rate is 1-Msymbol/sec, and consequently the symbol duration is 1 μsec.

system in terms of channel filter taps is given by

$$y_k = \sum_l h_k[l]x_{k-l} + \eta_k, \tag{1.2}$$

where $h_k[l]$ is the lth channel filter tap and η_k is the additive white Gaussian noise (AWGN) sample at time k. The notations x_k and y_k represent the transmitted and received signals, respectively.

The most popular probabilistic model for the channel filter taps is Rayleigh fading, in which $h_k[l]$ is modeled as a zero-mean circularly symmetric complex Gaussian random variable with variance σ_l^2. Basically, this fading model is applicable when there is no line-of-sight path from the transmitter to the receiver. Rayleigh fading is viewed as a reasonable model for tropospheric and ionospheric signal propagation as well as the effect of heavily built-up urban environments on radio signals.

At the receiver, detection of the originally transmitted signals x_k's is performed based on y_k's. There are two common approaches for detection. In the first approach the receiver does not have any knowledge about $h_k[l]$'s. This type of receiver is called a non-coherent receiver. It has been shown that the error probability of such a non-coherent detection in fading channel is very poor, or one needs to use a large amount of power for reliable communications [2].

The second approach requires that the channel gains are tracked at the receiver so that they are known at the receiver (but still random). In practice, this is done either by sending a known sequence (called a pilot or training sequence) or by using data symbols detected earlier in a decision directed manner. Knowing the channel gains, coherent detection can then be performed. However, it has also been shown that this type of receiver has only about 3dB improvement in signal-to-noise ratio (SNR) compared to non-coherent detection [2] and its performance is still very far away from the performance achieved over an AWGN (i.e., no fading) channel. The reason is that there is a significant probability that the channel is in a deep fade. Deep fade refers to the event when the overall received signals from different paths add destructively. Fig. 1.2 illustrates this phenomenon.

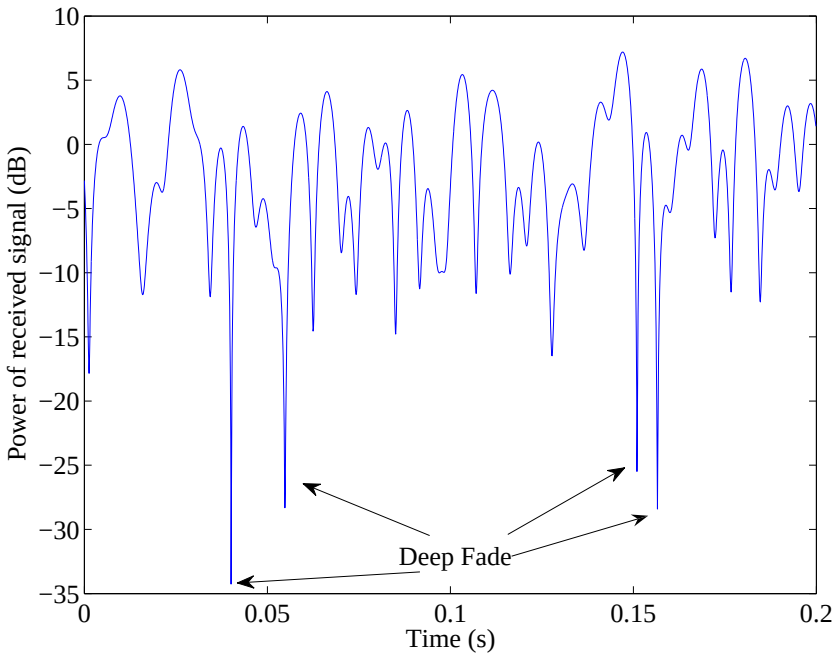

Figure 1.2 Deep fade phenomenon in Rayleigh fading with 100 Hz Doppler shift.

To combat fading and improve the system performance, various *diversity* techniques have been extensively studied in recent years. The main idea behind any diversity technique is to provide different replicas of the transmitted signal at the receiver in such a way that these different replicas are transmitted through independently faded paths. Since the different replicas fade independently, it is less probable to have all copies of the transmitted signal in deep fade simultaneously. Therefore, the receiver can reliably detect the transmitted signal using these received signals.

Diversity techniques can be implemented in different ways. Some of the most common techniques are discussed next.

Time diversity

Interleaving of coded symbols over time is the key operation in time diversity. In effect, this operation sends adjacent signal samples that would experience the same fading over independent fading channels. An alternative way to achieve this diversity for coded systems is to use bit-wise interleaving [3] instead of symbol interleaving. Bit-interleaved coded modulation (BICM) has been shown to be very effective in

exploiting time-diversity.

A drawback of bit-interleaving is that it causes an inherent "random modulation", which significantly reduces the free Euclidean distance when used over AWGN channels. Li and Ritcey [4] proposed a scheme, called BICM with iterative decoding (BICM-ID), to combat this random modulation effect. Their scheme utilizes iterative decoding in combination with hard-decision [4] or soft-decision [5] feedback from the decoder to the demodulator. They have shown that these techniques can provide performance improvement in both AWGN and fading environments.

Frequency diversity

The signal is transmitted using several frequency channels or spread over a wide spectrum that is affected by frequency-selective fading. Practical implementations of frequency diversity include orthogonal frequency division multiplexing (OFDM) in combination with subcarrier interleaving and forward error correction [6], and spread spectrum (for example frequency hopping or direct-sequence code division multiple access (DS-CDMA) [2]). This diversity technique is not considered in the current thesis.

Space diversity

Multiple transmit and/or receive antennas can be used to obtain spatial diversity. Antenna diversity is an effective and practical means of mitigating the effect of multipath fading. The pioneering work on multiple-input multiple-output (MIMO) systems [7] shows that a MIMO system has a multiplexing gain and accordingly higher spectral efficiency over a single-input single-output (SISO) system. Specifically, at high SNRs, the spectral efficiency of a MIMO system increases by m bits/s/Hz for every 3dB increase in SNR, where m is the minimum number of transmit and receive antennas. In contrast, the spectral efficiency of a SISO system increases only 1 bit/s/Hz per 3dB increase in SNR.

Signal space diversity (or modulation diversity)

The key operation in modulation diversity is to apply a certain rotation to a classical signal constellation in such a way that any two signal points have the maximum number of distinct components [8].

Signal space diversity (SSD) was introduced in [9] and [10] as a power and bandwidth efficient technique for communication over fast fading channels. In SSD, an N-dimensional modulation scheme is created by partitioning the data into blocks of N symbols and performing a rotation on each group of N-successive complex (two-dimensional) symbols. With such partitioning and rotation, the diversity order is maximized by increasing the minimum number of distinct components between any two N-dimensional constellation points.

For indoor wireless channels, where both time and frequency diversities may be poor, the channels are typically modeled as block-fading channels. In a block-fading channel model, the channel does not change for a duration of multiple symbols (called block) and then changes independently for the next block. For these channels, the diversity gain can be improved by using linear precoding in combination with the MIMO technique [11]. Using linear precoding increases the diversity by mixing the symbols of different time periods and antennas together.

1.2 Research Motivation

Unlike wireless communication systems without diversity, for systems that exploit diversity the performance difference between coherent and non-coherent detection techniques is significant. It has been shown that as the number of diversity paths increases, the performance of coherent receivers over a fading channel approaches that over an AWGN channel. In contrast, the performance of non-coherent detection in some diversity schemes first improves when the number of diversity paths increases, but then degrades as the number of diversity paths increases further [2].

The significant performance difference between coherent and non-coherent detec-

tions when the number of diversity paths is large amplifies the importance of channel knowledge. Such performance difference is observed under the assumption of perfect channel knowledge for coherent detection. However in practice, the channel taps have to be estimated and tracked. It is therefore important to understand the impact of channel measurement errors on the performance of the coherent detector.

Furthermore, it has been shown that in some cases, having too many diversity paths can have an adverse effect on the system performance due to channel estimation error [12]. Therefore, increasing the diversity gain by combining BICM together with other diversity techniques makes the task of the channel estimator more challenging. Moreover, since using diversity techniques generally adds interference to the transmitted signals, the detection process becomes more complicated. In consequence, extracting the channel information from the noise and interference-corrupted signals takes more hardware resources. On the other hand, the study in [13] shows that the effect of channel estimation error can be mitigated by designing channel estimators that exploit diversity techniques. This suggests that, if properly designed, channel estimation may benefit from diversity techniques.

Motivated by the above discussions, how to effectively utilize the diversity resources in order to improve the channel estimation quality and the system error performance of wireless data transmission is the main object in this thesis. The detailed motivations and contributions of the research shall be given in each chapter.

1.3 Organization of the Thesis

This thesis is organized in a manuscript-based style. The results obtained are included as the main content of the thesis in the form of published or submitted manuscripts. In each chapter, a brief introduction precedes each manuscript in order to connect the manuscript to the main context of the thesis. The remainder of the thesis is organized as follows.

Chapter 2 presents some fundamental knowledge about channel estimation as it

is the main subject of this thesis. First a classical channel estimator is introduced for coded modulation systems. Then an iterative channel estimator is discussed for BICM-ID systems. Such discussion also motivates the need of efficient channel estimation schemes in order to exploit the advantages of coded modulation systems.

Two main types of channel characteristics are considered in this research which are time-varying and time-invariant block fading channels. The first manuscript in Chapter 3 is concerned with channel estimation in a BICM-ID system under a time-varying fading channel. To improve the performance of the system, a new training scheme is introduced by exploiting the SSD technique. The proposed training scheme helps the channel estimator to track the variation of the channel even at low SNR. To facilitate this improvement, a soft iterative channel estimator is also developed to work with the proposed training scheme.

The second part of the thesis, which includes Chapters 4, 5, and 6, contributes to channel estimation under block-fading environment. In Chapter 4, the performance of BICM systems is analyzed using an iterative channel estimator under space diversity (i.e., MIMO channels). In this chapter, two different schemes for estimating the MIMO channels are suggested. In both schemes, convergence behavior of the system is studied. In the first scheme, in order to guarantee performance improvement through successive iterations, conditions are put on the output of the decoder. A less computationally intensive approach is then developed for higher code rates in the second scheme. In this scheme, both the training and data segments of the observation are used in each iteration.

Further results for BICM-MIMO systems are presented in Chapter 5. Motivated by the fact that combining training and data in channel estimation improves the performance of the system as shown in Chapter 4, the training sequence is embedded in data symbols using a space-time precoder at the transmitter. In the conventional scheme, the training overhead used for obtaining channel knowledge is proportional to a power of two of the number of transmit antennas. This overhead is reduced significantly by embedding pilot symbols within data symbols before precoding.

To further extend the result obtained in Chapter 5 and to exploit the advantages of using the space-time precoder, in Chapter 6, the precoder is designed taking into account the channel estimation error. In particular, the precoder is devised to mitigate the effect of channel estimation error on the error performance of the precoded BICM-MIMO system. To design a precoder in the presence of channel estimation error, the coding gain and diversity order of the precoded BICM-MIMO system are evaluated in a block-fading channel under imperfect CSI.

Finally, Chapter 7 summarizes the contributions of this thesis and suggests potential research problems for future works.

References

[1] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. New York: Cambridge University Press, 2005.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2004.

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

[4] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 715–724, Apr. 1999.

[5] Y. Li, X. G. Xia, and G. Wang, "Simple iterative methods to exploit the signal-space diversity," *IEEE Trans. Commun.*, vol. 53, pp. 32–38, Jan. 2005.

[6] J.A.C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[7] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, Jul. 2003.

[8] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1732–1742, May 2007.

[9] K. Boulle and J. C. Belfiore, "Modulation schemes designed for the Rayleigh channel," *Proc. Conf. on Inform. Sciences and Systems*, pp. 288–293, Mar. 1992.

[10] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1453–1467, Jul. 1998.

[11] N. Gresset, L. Brunel, and J. J. Boutros, "Space-time coding techniques with bit-interleaved coded modulations for MIMO block-fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2156–2178, May 2008.

[12] L. L. Chong, L. B. Milstein, "The effects of channel estimation errors on a space-time spreading CDMA system with dual transmit and dual receive diversity," *IEEE Trans. Commun.*, vol. 52, pp. 1145–1151, July 2004.

[13] H. Niu and J. Ritcey, "Pilot-symbol-assisted LDPC coded BICM over correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 849–859, Sep. 2003.

2. Background

As discussed in Chapter 1, channel estimation error limits the performance of a coded modulation system and this has been the subject of various studies in recent years [1–4]. Channel estimation can be performed using blind, semi-blind or data-aided methods. The blind channel estimators are mostly based on differential detection or second and higher order statistics of the received signals [5]. However, the performance of blind channel estimators is not comparable to that of data-based channel estimation. The semi-blind and data-aided estimators, in which a training sequence (or pilot) is inserted into the data sequence, are able to provide a more accurate estimate than the blind methods.

Two major types of training sequences are time division multiplexed (TDM) training and superimposed training. These two types of training for channel estimation have been discussed in [6–10] for fading channels. In TDM, a training sequence is multiplexed with data symbols periodically in time with a known pattern, which gives the position, size and power of the training sequence with respect to data information [7]. The position of a training sequence depends on the rate of channel variation such that the channel estimator can track the variation of the channel using the training sequence. In general, for a time-varying fading channel, the period of the training sequence should be less than $1/2(f_D T_s)$, where f_D is the Doppler shift and T_s is the symbol duration in second. This period provides better tracking of the time-varying channel since the Nyquist sampling theorem for the pilot symbol insertion is satisfied. On the other hand, for superimposed training, the training sequence and data symbols are added arithmetically and transmitted together [8, 11]. The effects

of using these training schemes have been evaluated by measuring the mean square error (MSE) of the channel estimator or the asymptotic error rate of the system [7]. While TDM, or also called pilot symbol assisted modulation (PSAM), has a better error performance at high SNR for slowly time-varying channels, the superimposed training outperforms PSAM in fast time-varying fading channels [7].

At very low SNRs, classical channel estimators (i.e., training-based estimators) can no longer be successfully applied [12, 13]. However, since the invention of turbo codes [14], researchers have been working on communication systems able to operate at very low SNRs. Consequently the need for channel estimators delivering accurate estimates at very low SNRs arises. Inspired by the turbo principle, several authors have proposed using information of data symbols, delivered by the decoder, to improve the estimation quality. This is referred to as turbo (or iterative) channel estimation. Indeed, the poor performance of conventional estimators at low SNRs is attributed to the lack of information about the data symbols.

This chapter discusses channel estimation in coded modulation systems. More specifically, it distinguishes classical channel estimation from iterative channel estimation. In the first section, a single-antenna system model with time-varying fading channel is considered and the channel estimation is developed with PSAM. Then in the second section, the iterative channel estimation will be explained for BICM. The result can be easily extended to other scenarios (with different channel characteristics).

2.1 Classical Channel Estimation

Classical channel estimation is usually characterized by type of the training scheme and the criterion used for estimation. To illustrate the operation of classical channel estimation in coded modulation systems, the block diagram of a typical coded modulation system is illustrated in Fig. 2.1. Without loss of generality, a simple channel is considered to explain the operation of the channel estimator. The following assumptions and settings are considered in this section:

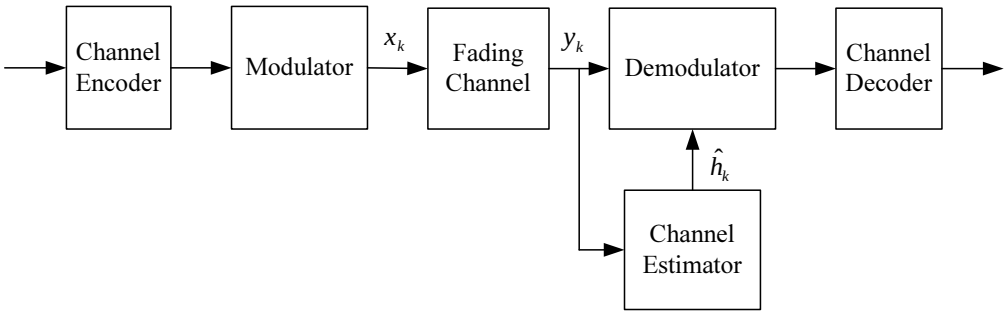

Figure 2.1 Block diagram of a coded modulation system with a classical channel estimator.

- The channel is assumed to be time-varying flat fading. In this case the channel can be modeled by a single tap, h_k, at each time index k. The sequence $\{h_k\}$ contains complex Gaussian random variables with zero mean. The autocorrelation function $R(n) = E\{h_k^* h_{k+n}\} = \sigma_h^2 J_0(2\pi f_D T_s n)$ is given by the Jakes' model [15], where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and σ_h^2 is the variance of channel coefficient h_k.

- A pilot symbol, denoted by $x^{(\mathrm{p})}$ with a period of $L \geq 1/2(f_D T_s)$ is inserted in the data stream as shown in Fig. 2.2.

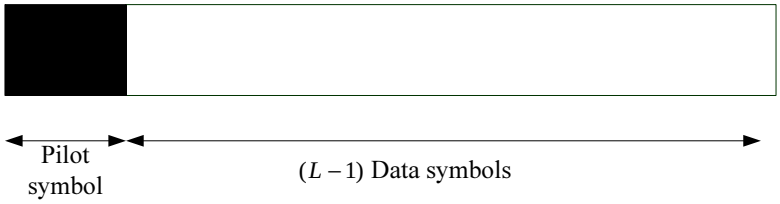

Figure 2.2 Representation of one period of PSAM.

- The input-output model of the channel is given by

$$y_k = h_k x_k + \eta_k, \tag{2.1}$$

where x_k and y_k are the transmitted and the received signals, respectively. The noise component, η_k, is zero-mean complex Gaussian random variable with variance N_0. The linear model in (2.1) is also valid for the training phase. For

Table 2.1 Objective functions for different channel estimators

Maximum likelihood (ML)	$\max_{h_k} p(y_k\|h_k, x_k)$
Least square (LS) error	$\min_{h_k} \|y_k - h_k x_k\|^2$
Maximum a posteriori (MAP) probability	$\max_{h_k} p(h_k\|y_k)$
Minimum mean square error (MMSE)	$\min_{\hat{h}_k} E\{\|h_k - \hat{h}_k\|^2\}$

convenience, x_k is set to $x^{(\mathrm{p})}$ and y_k is denoted by $y_k^{(\mathrm{p})}$ during the training phase. Also let E_s and E_p be the energy per data symbol (x_k) and pilot symbol ($x^{(\mathrm{p})}$), respectively.

The problem is to estimate $\{h_k\}$ using two sources of information: the observation at the channel output in the training phase (i.e., $\{y_k^{(\mathrm{p})}\}$) and the known pilot symbol $x^{(\mathrm{p})}$. Several channel estimators have been proposed using the following criteria: the minimum mean square error (MMSE), the maximum a posteriori (MAP) probability, the least square (LS) error and the maximum likelihood (ML) function. The objective functions which these channel estimators minimize or maximize are given in Table. 2.1, where $p(x|y)$ and $E\{x\}$ refer to probability of x given y and expected value of x, respectively.

The solutions for these estimators based on the linear model given in (2.1) during the training phase are given as follows:

- Since $y_k^{(\mathrm{p})} = h_k x^{(\mathrm{p})} + \eta_k$, and η_k has a Gaussian distribution, it can be shown that maximizing the likelihood function and minimizing the LS error lead to the same estimator, given by,

$$\hat{h}_k = \frac{1}{|x^{(\mathrm{p})}|^2}(x^{(\mathrm{p})})^* y_k^{(\mathrm{p})}. \tag{2.2}$$

- If h_k and y_k are jointly Gaussian, which is the case here, then the MAP estimator has the same expression as the MMSE estimator. The estimator is given by,

$$\hat{h}_k = \frac{1}{N_0 + \sigma_h^2 E_p}\sigma_h^2 (x^{(\mathrm{p})})^* y_k^{(\mathrm{p})}. \tag{2.3}$$

It should be mentioned that the channel taps are only estimated at training time indexes. These estimates are then interpolated and used for other data time indexes. The sinc filter and Gaussian filter have been used for PSAM but the Wiener filter is the optimum interpolator in terms of minimizing the variance of the estimation error [10]. In this case, the channel estimator prepares an estimate of the channel gain h_k using the M nearest pilot symbols by solving the Wiener-Hopf equations [16]. If the M nearest received pilot symbols to the kth position of the channel gain is collected in vector $\mathbf{y}_k^{(\mathrm{p})}$, then the coefficients of the Wiener filter that minimizes the variance of the estimation error (i.e., the MMSE estimator) are obtained by,

$$\mathbf{a}_k \;=\; E\{\mathbf{y}_k^{(\mathrm{p})}(\mathbf{y}_k^{(\mathrm{p})})^H\}^{-1}E\{h_k(\mathbf{y}_k^{(\mathrm{p})})^H\}, \tag{2.4}$$

and the channel estimate is,

$$\hat{h}_k \;=\; \mathbf{a}_k^T\mathbf{y}_k^{(\mathrm{p})}. \tag{2.5}$$

Based on the Jakes' model of the correlated time-varying channel, one has

$$E\{\mathbf{y}_k^{(\mathrm{p})}(\mathbf{y}_k^{(\mathrm{p})})^H\}_{i,j} = \begin{cases} E_p\sigma_h^2 + N_0, & i = j \\ E_p\sigma_h^2 J_0(2\pi f_D T_s|i-j|L), & i \neq j \end{cases}, \tag{2.6}$$

and

$$E\{h_k(\mathbf{y}_k^{(\mathrm{p})})^H\}_i = \sigma_h^2 J_0(2\pi f_D T_s|iL-k|)(x^{(\mathrm{p})})^*. \tag{2.7}$$

Then a closed-form expression for the channel estimate can be found using (2.4) and (2.5).

When the statistic of the noise is known, the MMSE estimator has the minimum error variance and it also has the desirable property that the resulting estimate is uncorrelated with the estimation error [4], and hence it is selected for this research. However, in general the estimator can be selected to have the smallest degradation on system error performance or to have less complexity. The effect of channel estimation error on system performance depends on the type of receiver and characteristics of the channel and therefore needs further investigation for each case. For example

under Rayleigh fading channel and Gaussian noise in MIMO, the ML and MMSE estimators have the same impacts on the asymptotic error rate performance of the ML receiver [18].

For evaluating the impact of channel estimation error on system performance, it is still required to obtain the variance of the estimation error (or MSE). However, sometimes the calculation of MSE is cumbersome. Hence, it is useful to establish lower bounds on the attainable MSE.

Lower bounds on the variance of channel estimation error

There are several lower bounds on the MSE attained by a channel estimator that show the efficiency of the channel estimator. They can be used as benchmarks for assessing the performance of a channel estimator and show the relation between the performance of the channel estimator and parameters of a system. The tightest lower bound is the Barankin bound [19] which is very difficult to compute. The Bhattacharyya bound [20, 21] is simpler to evaluate but it is still more complicated than one of its special version: the Cramer-Rao bound (CRB).

The CRB is widely used as a performance benchmark for a channel estimator. The bound states that the MSE matrix of any unbiased estimator $\hat{\mathbf{h}}$ is lower bounded as [16]:

$$\mathrm{MSE}(\hat{\mathbf{h}}) \equiv E\{[\hat{\mathbf{h}} - \mathbf{h}][\hat{\mathbf{h}} - \mathbf{h}]^H\} \geq \mathbf{J}(\mathbf{h})^{-1},$$

where $\mathbf{J}(\mathbf{h})$ is the complex Fisher information matrix, defined by,

$$\mathbf{J}(\mathbf{h}) \;=\; E\left\{\left[\frac{\partial \ln p(\mathbf{y},\mathbf{h})}{\partial \mathbf{h}^*}\right]\left[\frac{\partial \ln p(\mathbf{y},\mathbf{h})}{\partial \mathbf{h}^*}\right]^H\right\}. \tag{2.8}$$

The vectors $\mathbf{y}$ and $\mathbf{h}$ are vectors of observations and corresponding channel samples, respectively. The expectation is taken over $\mathbf{y}$ and $\mathbf{h}$. The CRB has been calculated in [28] for a time-varying Rayleigh fading channel.

After calculating the CRB for a training-based channel estimation, it can be shown that the CRB is a function of training sequence. Thus in most cases, the CRB is used

as a criterion to design training pattern [29]. However, to design other parameters of a system, it is required to examine the impact of channel estimation error on other benchmarks used for performance evaluation. This issue will be studied in this research. In the following an example of BER performance degradation due to the imperfection of channel estimation is discussed.

Effect of channel estimation error on the performance of coded modulation systems

As an example, Fig. 2.3 compares BER performance of a coded modulation system utilizing training-based MMSE estimator and that of a system with perfect channel state information (CSI). The coded modulation system employs a rate-1/2 convolutional code with constraint length 5 and generator polynomial $[1 + D + D^3 + D^4; 1 + D + D^4]$. The modulation is quadrature phase-shift keying (QPSK). The normalized Doppler shift $f_D T_s$ is assumed to be 0.02. The curves are plotted versus E_b/N_0, where E_b/N_0 is the energy per information bit over the noise variance. For the simulated system (i.e., rate-1/2 coding and QPSK modulation) $E_s/N_0 = E_b/N_0$.

As can be seen in Fig. 2.3, the training-based channel estimator leaves a big performance gap compared to the case of perfect CSI. The main reason is that the training-based MMSE estimator processes only the observation samples that depend on the pilot symbols. The observations corresponding to the data symbols are not taken into account in the estimator. Another drawback of the training-based channel estimation is that the transmission of pilot symbols required by the estimator reduces the spectral and power efficiency. Indeed, the pilot symbols – unlike the data symbols – do not carry any information.

2.2 Iterative Channel Estimation in BICM Systems

The drawback of using only a training sequence in training-based channel estimation can be mitigated by utilizing the soft information of data symbols, fed back from the decoder. To exploit the fed back symbols efficiently (i.e., soft information),

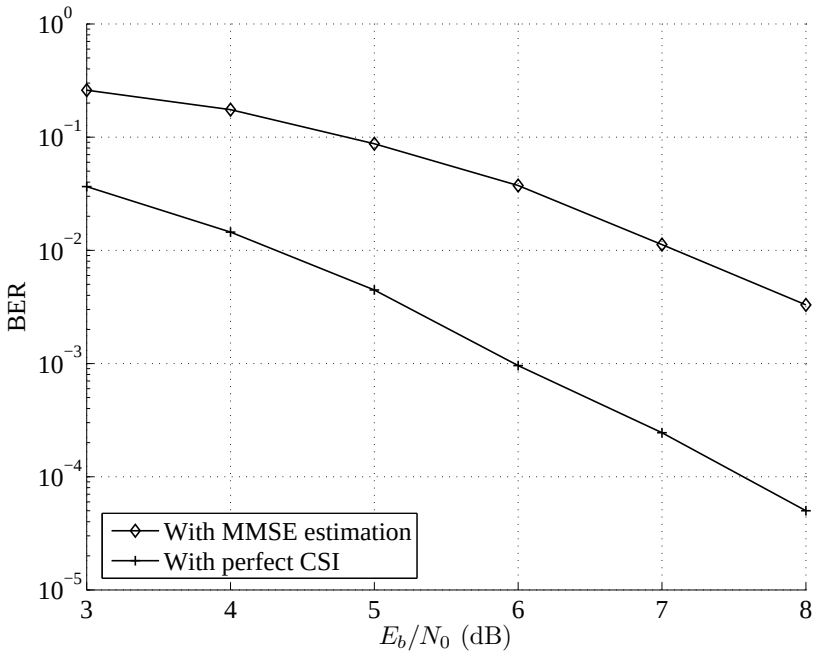

Figure 2.3 Comparison of BER achieved with training-based MMSE estimator and with perfect CSI when $f_D T_s = 0.02$.

a special coding structure is required. The key in such a coding structure is that they can be decoded iteratively. This coding structure was introduced first in 1993 by Claude Berrou's research group and it is termed a turbo code [14]. The performance of turbo codes closely approaches the Shannons limit while possessing a reasonable decoding complexity.

A turbo code consists of two simple convolutional codes, concatenated in parallel and separated by an interleaver. At the receiver, each convolutional code is separately decoded but in order to help each other, both decoders exchange some soft extrinsic information about the transmitted bits in an iterative fashion. In contrast to the hard information, the soft information keeps track of the level of reliability (measured by probability) on the latest decisions made on the transmitted bits. The exchanged information is extrinsic information meaning that the information sent by one decoder is only the extra information brought by another decoder. In other words, the extrinsic information received by one decoder at a given iteration does not contain the output by that decoder in the previous iteration.

Because of its outstanding performance, this iterative processing at the receiver, called turbo principle, has further been applied to other processing tasks in the receiver: demodulation [22], equalization [23], multi-user detection [24], MIMO detection [25], synchronization [26] and channel estimation [27]. When involved in an iterative process, all these tasks are referred to as "turbo" followed by their regular names, e.g., turbo demodulation, turbo equalization, turbo multi-user detection.

The association of a channel encoder and the discrete-time equivalent channel, separated by an interleaver in a BICM system, may be regarded as the serial concatenation of two codes. Therefore, the principle of turbo decoding can also be applied to channel estimation and decoding leading to turbo (or iterative) channel estimation. This technique was investigated for the first time in [27]. Research studies suggest that soft iterative channel estimators can lead to a significant gain in the performance of the receiver [6]. Iterative channel estimation is a popular semi-blind channel estimation that uses soft information of data fed back from the decoder.

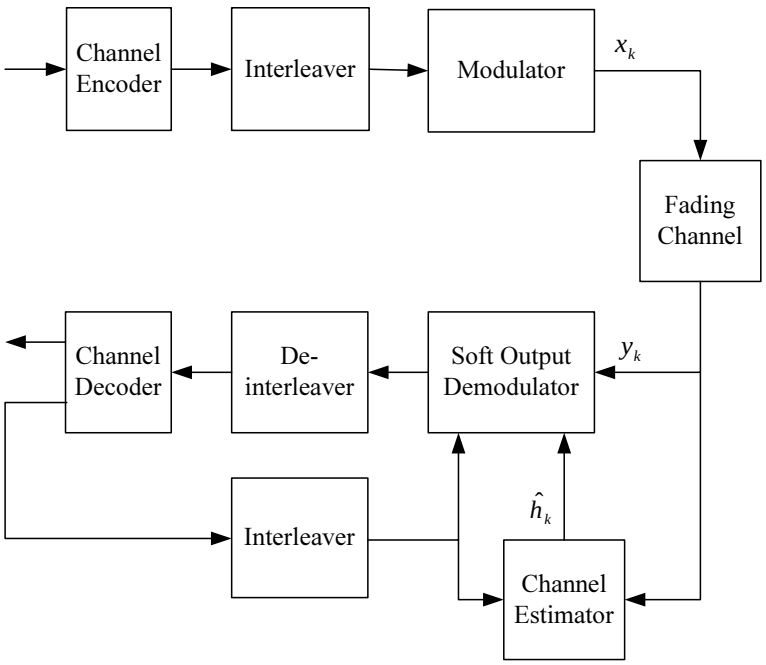

Figure 2.4 Block diagram of a BICM-ID system with iterative channel estimation.

To explain how an iterative channel estimator works, the block diagram of a BICM system with an iterative receiver is shown in Fig. 2.4. The model is quite general and

can apply to any specific system considered in subsequent chapters.

2.2.1 Transmitter of a BICM System

The transmitter is built from the following fundamental blocks. First a channel encoder with a rate-R_c error-correcting code converts the vector of information bits $\mathbf{b}$ into a codeword $\mathbf{c}$. Next, the coded bits are interleaved by a bit interleaver. The bit interleaver breaks the fading correlation by which the coded bits are corrupted. Then the interleaved coded bits are mapped into signal points of a modulation constellation. We now describe the role of each fundamental block in more detail.

Channel Encoder

The encoder applies the bijection between the input information vector $\mathbf{b}$ and the codeword $\mathbf{c}$. The length of $\mathbf{c}$ is $1/R_c$ times higher than the length of $\mathbf{b}$. The error correcting code can be, for example, one of the following:

- Linear block codes (like Reed-Solomon codes).
- Trellis codes (like convolutional codes). Traditionally, convolutional codes are considered for BICM systems. Indeed, they have the double advantage of being simple in encoding and decoded with soft-input soft-output decoders [31]. Convolutional codes are used in this thesis.
- Concatenated codes. These codes can be constructed by concatenating several component codes (block or convolutional codes) [31] at the price of increased decoding complexity.

Interleaver

The interleaver scrambles the coded bits. This is the main component of a BICM system. It is crucial when performing iterative joint detection and decoding because it enhances the independence between *extrinsic* and *a priori* probabilities in both the soft-input soft-output detector and decoder. It is also very important for ML

decoding (if such a decoding is tractable) because it limits the interference in the same time period between two erroneous bits of an error event. The interleaver can be a pseudo-random or a semi-deterministic interleaver with some deterministic constraints.

Modulator

The interleaved coded bits are de-multiplexed into blocks of m bits, which are fed to the mapper that converts them into a constellation symbol. The bijection between the bit vectors and constellation symbols is called mapping or labeling. The number of points in the constellation is equal to $M = 2^m$. At each channel use, the mapper reads m coded bits and generates one modulation symbol. In general, the mapping is not unique and will be chosen depending on specific channel and application. The Gray mapping is a very well-known mapping since it minimizes the number of different bits between two neighbors in the constellation, which essentially minimizes the bit error rate of an uncoded system. However, in many cases, better performance can be achieved by using other mapping techniques (in fact the Gray mapping is shown to be the worst mapping for BICM with ideal interleaving [32]). This thesis considers quadrature amplitude modulation (QAM) as it achieves a good compromise between spectral efficiency (in bits/s/Hz or bits/dim) and performance.

Training sequence for the purpose of channel estimation is embedded in this stage. The method of embedding depends on the application. Typically in fast time-varying channels, superimposed training is added to the data stream, whereas in slow fading, time-multiplexed training is preferred [7].

2.2.2 Iterative Receiver

An ideal BICM receiver would perform a maximum likelihood decoding, which requires joint demodulation, convolutional decoding and perfect knowledge of CSI. Due to the presence of the random bit interleaver, the implementation of this receiver is intractable in practice. Instead, the suboptimal receiver of the system includes three

separate blocks: the soft-input soft-output demodulator, the soft-input soft-output decoder, and the channel estimator.

Soft-Input Soft-Output Decoder

Decoding of an error correcting code has always been a topic of interest. Obviously, the decoder depends on the code employed. For the convolutional codes used in this thesis, the soft-input soft-output decoder could be based on the soft-output Viterbi algorithm (SOVA) or the maximum a posteriori probability (MAP) algorithm [31].

Soft-Input Soft-Output Demodulator

A soft-input soft-output detector can achieve near maximum likelihood performance through iteration with the decoder. Any technique that a soft-input soft-output demodulator uses, converts the received vector of each time period into extrinsic probabilities on the coded bits thanks to the *a priori* probabilities on the coded bits. Some practical soft-input soft-output demodulators for BICM are:

- Maximum a posteriori probability (MAP) demodulators [33].
- Approximate MMSE-APP demodulators [34].
- MMSE soft-output demodulators [35].

Algorithms that can be used for a soft-input soft-output demodulator depend on diversity techniques and parameters of the modulator. The algorithms applied in this thesis are explained in detail in the following chapters for each considered system.

Interactions between Demodulator and Decoder

In any given iteration, the interaction between the demodulator and the channel decoder can be explained as follows. After the channel estimation is performed (as explained in the next section), a soft-input soft-output demodulator demodulates the data. The soft-output demodulator computes the extrinsic information for the

interleaved bits, from the received symbols. To obtain the extrinsic information, the demodulator also exploits the *a priori* information of the coded bits coming from the decoder and the channel estimates. To calculate the extrinsic information, the following *a posteriori* probability (APP) expression is available for any coded bit c_l of any time period. The received signal (i.e., observation) during the considered time period is y_k. The APP probability of a coded bit c_l is defined and computed by Bayes' theorem as:

$$\begin{aligned} \mathrm{APP}(c_l) &= p(c_l|y_k) \\ &= \frac{p(y_k|c_l)p(c_l)}{p(y_k)} \end{aligned} \tag{2.9}$$

In the above expression, $\mathrm{APP}(c_l)$ is expressed as a function of three probability quantities:

- At each iteration, the probabilities given by the output of the decoder are independent from the received signal y_k. They are called a priori probabilities on the coded bits c_l and denoted by $p(c_l)$. At the first iteration, no *a priori* information is available at the demodulator input, so it equally considers all the constellation points.

- The probability $p(y_k)$ depends on the transmitted coded bits, the a priori probability and the AWGN, and is not computable. Fortunately, it can be shown that this quantity is not necessary for the iterative processing.

- The conditioned observation $p(y_k|c_l)$ can be decomposed into more explicit probabilities. By marginalizing over the set of labelings having the lth bit equal to c_l, i.e., $\mathbf{c} = \{c_1, \ldots, c_l, \ldots, c_m\}$, we have,

$$p(y_k|c_l) = \sum_c p(y_k, \mathbf{c}|c_l) = \sum_c p(y_k|\mathbf{c}, c_l)p(\mathbf{c}) \tag{2.10}$$

  where the condition over $\mathbf{c}, c_l$ is equivalent to a condition over the corresponding modulation symbol, x. Using the AWGN distribution, one has

$$p(y_k|\mathbf{c}, c_l) = p(y_k|x) = \frac{1}{\sqrt{2\pi N_0}} \exp \frac{-|y_k - xh_k|^2}{2N_0}. \tag{2.11}$$

The coded bits transmitted during the same time period are supposed to be independent. Thus, in (2.10) $p(\mathbf{c}) = \prod_{i \neq l} p(c_i)$.

In the above, the APP(c_l), computed in the demodulator, can also be expressed as Extr(c_l)$p(c_l)$, with Extr(c_l) and $p(c_l)$ are two independent variables, and Extr(c_l) is called the extrinsic probability. Moreover, using the normalization, the extrinsic probability that the lth coded bit equals 1 is expressed as

$$\text{Extr}(c_l) \;=\; \frac{p(y_k|c_l=1)}{p(y_k|c_l=1)+p(y_k|c_l=0)}. \tag{2.12}$$

Therefore,

$$\text{Extr}(c_l) \;=\; \frac{\sum_{x\in\Omega_l}\left\{\exp\left(-\frac{|y_k-xh_k|^2}{2N_0}\right)\prod_{i\neq l}p(c_i)\right\}}{\sum_{x\in\Omega}\left\{\exp\left(-\frac{|y_k-xh_k|^2}{2N_0}\right)\prod_{i\neq l}p(c_i)\right\}} \tag{2.13}$$

where Ω is the set of all symbols x generated by the QAM mapper, i.e., $|\Omega| = 2^m$ and the subset Ω_l for $l = 1, \ldots, m$, is restricted to include symbols x in which the lth coded bit is equal to 1. One way to express the reliability information or soft information is by the log-likelihood ratio (LLR) , which is defined by $\text{LLR}(c_l) = \ln\frac{p(c_l=1)}{p(c_l=0)}$ or equivalently $\ln\frac{\text{Extr}(c_l=1)}{\text{Extr}(c_l=0)}$.

Then the *extrinsic* information are de-interleaved and become the *a priori* soft information to be used in the channel decoder shown in Fig. 2.4. The channel decoder uses the MAP algorithm to compute the extrinsic information for all coded bits, which are used again in the next iteration in the demodulator. This process is repeated until the iterative receiver converges.

It should be noted that in (2.11) and (2.13), it is assumed that h_k is perfectly known. However, in practice it should be estimated and thus the performance of the demodulator is affected by the accuracy of the estimation.

Iterative Channel Estimator

The interaction between the soft-iterative channel estimator and the channel decoder and the demodulator can be explained as follows. At the first iteration, channel

estimator uses only the training sequence to estimate CSI. It uses the same processing as training-based channel estimators explained in section 2.1.

In subsequent iterations, soft information under the form of extrinsic LLRs on coded bits output by the decoder are fed back to the estimator in order to improve its performance. The channel estimator uses the soft information from the channel decoder to compute new estimates of the channel coefficients using expected values of the data symbols. Therefore the interleaved soft information from the decoder is fed back to the estimator to calculate the expected values and variances of the data symbols, i.e., $E\{x_k\}$ and $\sigma^2_{x_k}$. Calculating $E\{x_k\}$ and $\sigma^2_{x_k}$ from LLRs depends on the mapping rule in constellation set. For example Table 2.2 shows the mapping rules and the expressions for mean values of BPSK and QPSK modulation schemes in terms of LLRs.

The first example is explained as follows. Consider the mapping from one bit to a BPSK symbol as $b_1 \in \{0, 1\}$. Moreover assume that the labeling is such that $x_k = -1$ when the transmitted bit is 0 and $x_k = +1$ when it is 1. Then for each symbol there is one LLR fed-back from the decoder. This LLR(b_1) acts as *a priori* information of symbol x_k. Thus $E\{x_k\} = p(x = +1) - p(x = -1) = p(b_1 = 1) - p(b_1 = 0)$. Dividing by $p(b_1 = 1) + p(b_1 = 0) = 1$ and using the definition of LLR, one has $E\{x_k\} = \frac{\exp(\text{LLR}(b_1))-1}{\exp(\text{LLR}(b_1))+1} = \tanh(\text{LLR}(b_1)/2)$. The variance is $\sigma^2_{x_k} = 1 - |E\{x_k\}|^2$.

The refined estimates of the channel gains are then sent to the demodulator.

It should be pointed out that the iterations related to the feedback of soft extrinsic information from the decoder to the estimator (referred to as "estimation iterations") may differ from those related to the feedback of the same information from the decoder to the demodulator (referred to as "decoding iterations"). For instance, two decoding iterations may be performed per estimation iteration.

Since iterative channel estimation has interactions with the decoder and the demodulator through LLRs, its performance may be affected by the design parameters of the channel encoder and modulator. Moreover, since for the first iteration, the

Table 2.2 Conversion of soft information to the mean value of symbols

Modulation scheme	Bits	Symbol	Mean Value
BPSK	b_1	x_k	$\tanh \frac{\mathrm{LLR}(b_1)}{2}$
	1	$+1$	
	0	-1	
QPSK	$b_1 b_2$	x_k	$\tanh \frac{\mathrm{LLR}(b_1)}{2} + j \tanh \frac{\mathrm{LLR}(b_2)}{2}$
	00	$1+j$	
	01	$1-j$	
	10	$-1+j$	
	11	$-1-j$	

channel is estimated using the training sequence, the estimation performance also depends on the training sequence.

To analyze the effect of the channel estimation error on the performance of a BICM system with an iterative receiver, different quantities and tools have been used in [36] and [6], which include the pairwise error probability (PEP), asymptotic BER and extrinsic information transfer (EXIT) chart. These quantities were used to develop channel estimators for different channel scenarios in order to improve the performance. These quantities may also be used to analyze the effect of BICM system parameters on channel estimation. In parts of this research, PEP and EXIT chart shall be considered.

2.3 Summary

Two types of channel estimations for coded modulation systems have been briefly discussed in this chapter. For training-based channel estimation, with certain assumptions on channel properties, different criteria are examined. Among those criteria, MMSE estimator is selected as it has a promising performance. By evaluating the coded modulation system with training-based channel estimation, it was shown that there are two drawbacks. To compensate for the first one, which is the lack of

information, the use of an iterative channel estimation has been discussed. In Chapter 3, this estimator is developed for BICM systems that exploit SSD. Since it has been shown that one of the parameters that the performance of the channel estimator depends on is the training sequence, Chapter 3 investigates the training design for a time-varying system that exploits SSD technique. In Chapter 4, an algorithm for the iteration process in the iterative channel estimator is applied. This algorithm is considered for MIMO block-fading channels. The second drawback is the signalling overhead of the training sequence. This issue is investigated in Chapters 5 and 6 for MIMO block-fading channels under the CRB and PEP performance criteria.

References

[1] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data- aided channel estimation and perfect interleaving," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 2358–2368, Dec. 2001.

[2] A. Maaref and S. Aissa, "Optimized rate-adaptive PSAM for MIMO MRC systems with transmit and receive CSI imperfections," *IEEE Trans. Commun.*, vol. 57, pp. 821–830, Mar. 2009.

[3] J. Wang, O. Y. Wen, and S. Li, "Soft-output MMSE MIMO detector under ML channel estimation and channel correlation," *IEEE Signal Processing Letters*, vol. 16, pp. 667–670, Aug. 2009.

[4] M. Coldrey and P. Bohlin, "Training-based MIMO systems, Part I: Performance comparison," *IEEE Trans. Signal Processing*, vol. 55, pp. 5464–5476, Nov. 2007.

[5] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems", *IEEE Trans. Veh. Technol.*, vol. 56, pp. 670–685, Mar. 2007.

[6] Y. Huang and J. A. Ritcey, "Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2549–2558, Sep. 2005.

[7] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1403–1418, May 2004.

[8] X. Meng, J. K. Tugnait, and S. He, "Iterative joint channel estimation and data detection using superimposed training: Algorithm and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1873–1880, Jul. 2007.

[9] H. Niu and J. Ritcey, "Pilot-symbol-assisted LDPC coded BICM over correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 849–859, Sep. 2003.

[10] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[11] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, pp. 240–254, Jan. 2004.

[12] X. Wautelet, A. Dejonghe, and L. Vandendorpe, "MMSE-based fractional turbo receiver for space-time BICM over frequency-selective MIMO fading channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1804–1809, Jun. 2004.

[13] R. Otnes and M. Tuchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1918–1923, Nov. 2004.

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

[15] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley and Sons Inc., 1974.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall PTR, 1993.

[17] M. Nicoli, S. Ferrara, and U. Spagnolini, "Soft-iterative channel estimation: Methods and performance analysis," *IEEE Trans. Signal Processing*, vol. 55, pp. 2993–3006, June 2007.

[18] G. Taricco and E. Biglieri, "Spacetime decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1874–1888, July 2005.

[19] E. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 20, pp. 477–501, 1949.

[20] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, Inc., 2001.

[21] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Shankya*, vol. 8, pp. 1–14, 1946.

[22] S. ten Brink, J. Speidel, and J.C. Yan, "Iterative demapping and decoding for multilevel modulation," *Proc. IEEE Globecomm Conf.*, pp. 579–584, Nov. 1998.

[23] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. on Telecommun.*, Vol. 6, pp. 507–511, Sep. 1995.

[24] X. Wang and V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, Vol. 47, pp. 1046–1061, Jul. 1999.

[25] A. Tonello, "Space-time bit-interleaved coded modulation over frequency selective fading channels with iterative decoding," *Proc. IEEE Globecomm Conf.*, pp. 1616–1620, Nov. 2000.

[26] V. Lottici and M. Luise, "Carrier phase recovery for turbo-coded linear modulations," *Proc. IEEE Inter. Conf. Commun.*, Vol. 3, pp. 1541–1545, Aug. 2002.

[27] M. Sandell, C. Luschi, P. Strauch, and R. Yan, "Iterative channel estimation using soft decision feedback," *Proc. IEEE Globecomm Conf.*, pp. 3728–3733, Nov. 1998.

[28] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 3055–3069, Dec. 2002.

[29] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using the affine precoding framework and the CRB," *IEEE Trans. Signal Processing*, vol. 54, pp. 940–954, Mar. 2006.

[30] K.-P. Yar and W. E. Stark, "Performance of Reed-Solomon coded M-ary modulation systems with overlapped symbols," *Proc. IEEE Military Commun. Conf.*, vol. 1, pp. 364–370, Oct. 2005.

[31] R. H Morelos-Zaragoza, *The Art of Error Correcting Coding*. New Jersey: John Wiley, 2006.

[32] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 944–957, May 2001.

[33] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.

[34] L. Ping, "Approximate MMSE-APP estimation for linear systems with binary inputs," *IEEE Commun. Letters*, vol. 9, pp. 172–174, Feb. 2005.

[35] G. M. Kraidy and P. Rossi, "Full-diversity iterative MMSE receivers with space-time precoders over block-fading MIMO channels," in *Proc. IEEE Int. Conf. Wireless Commun and Signal Processing*, (Suzhou), pp. 1–5, Oct. 2010.

[36] I.-W. Lai, S. Godtmann, T.-D. Chiueh, G. Ascheid, and H. Meyr, "Asymptotic BER analysis for MIMO-BICM with zero-forcing detectors assuming imperfect CSI," *Proc. IEEE Int. Conf. Commun.*, pp. 1238–1242, May 2008.

3. Channel Estimation in Bit Interleaved Coded Modulation with Iterative Decoding

Published as:

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt, "Channel Estimation in Bit Interleaved Coded Modulation with Iterative Decoding", *IET Communications*, vol. 4, pp. 2095-2103, Nov. 2010.

In the previous chapter, classical and iterative channel estimators for time-varying fading channels have been discussed. It was shown that the performance of a classical channel estimator depends on the training sequence while for the iterative channel estimator, the performance depends also on the modulator design (i.e., mapping and the constellation set). On the other hand, for a time-varying fading channel it was shown in [3] that the performance of a BICM system is improved by introducing extra diversity. This extra diversity is provided by applying certain rotation after the modulator, a technique called signal space diversity (SSD). Indeed applying SSD can be considered as a part of the modulator.

The manuscript included in this chapter studies iterative channel estimation in a BICM system with the SSD technique. In addition, different from the conventional training sequence for a simple BICM system as in [12], a training sequence is inserted before the modulator. The iterative channel estimator is designed to work effectively with the SSD training sequence.

Channel Estimation in Bit Interleaved Coded Modulation with Iterative Decoding

Zohreh Andalibi, Ha H. Nguyen, J. Eric Salt

Abstract

This paper improves the fading channel estimation in bit-interleaved coded modulation systems with iterative decoding (BICM-ID) and signal space diversity (SSD) by embedding a training sequence. Existing training schemes work well at high signal-to-noise ratio (SNR) or slowly time-varying channels while the applications of BICM-ID are beneficial at low SNR and fast time-varying fading channels. Motivated by the power/bandwidth efficiency of the SSD technique and the fact that superimposed training outperforms pilot symbol assisted modulation (PSAM) training over relatively fast time-varying channels, a new superimposed training sequence is explored. The proposed training sequence inserts pilot bits into the coded bits prior constellation mapping and signal rotation. This becomes a superimposed training sequence in the rotated symbols and helps the estimator to track fast variation of the channel gains. A soft iterative channel estimator is developed to work with the superimposed training sequence. Performance of the proposed scheme, namely SSD-pilot, is shown to be superior to PSAM scheme. To gauge the performance improvement achieved with the proposed channel estimation, an analytical bound on the asymptotic bit error probability for BICM-ID using SSD over correlated fading channels is provided. The Cramer-Rao bound on the mean-square error of the channel estimator is also derived to evaluate the performance of the iterative channel estimator.

Index terms

Manuscript received June 16, 2009; revised June 8, 2010.

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt are with the Department of Electrical & Computer Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon, SK, Canada S7N 5A9, and TRLabs of Saskatchewan. Emails: z.andalibi@usask.ca, ha.nguyen@usask.ca, eric.salt@usask.ca.

Signal space diversity, bit-interleaved coded modulation, iterative decoding, error performance, channel estimation, training sequences, pilot-symbol assisted modulation, Cramer-Rao bound.

3.1 Introduction

Signal space diversity (SSD) was introduced in [1] and [2] as a power and bandwidth efficient technique for communication over fading channels. In SSD, an N-dimensional modulation scheme is created by partitioning the data into blocks of N symbols and performing a rotation on each group of N-successive complex (two-dimensional) symbols. With such partitioning and rotation, the diversity order is maximized by increasing the minimum number of distinct components between any two N-dimensional constellation points. Among the various applications of SSD, bit-interleaved coded modulation (BICM) with iterative decoding (ID) over fast fading channels has been studied in [3]. In [3], the transmitter and receiver are designed under the assumption that the channel state information (CSI) is known at the receiver. However, in practical applications, this is not the case and the CSI must be estimated. Imperfect CSI due to channel estimation degrades the performance, especially at low signal-to-noise ratio (SNR) regime. This drawback can be mitigated by using a soft-iterative channel estimator [4].

Soft iterative channel estimation has been intensively studied in turbo-coded systems, see e.g., [5], [6] and [7] and references therein. Soft iterative channel estimation uses soft information from the soft-input soft-output (SISO) decoder in a semi-blind fashion to improve channel estimation performance, and consequently the system's bit-error-rate (BER) , over fading channels. The concept of semi-blind estimation is rooted in functions of both known and unknown signals. In the context of wireless communications, known signals can be either time-multiplexed or superimposed training sequence. In a time-multiplexed training sequence, pilot symbols are inserted regularly in the data stream. Such schemes, which are commonly called pilot-symbol assisted modulation (PSAM) schemes perform well over slowly time varying channels [8, 9]. To work in a fast fading channel the pilot symbols in a PSAM scheme

have to be inserted more often, which causes bandwidth expansion. Alternatively a superimposed training scheme, which in effect superimposes a training signal on each transmitted symbol, can be used. Superimposed training schemes work well in fast fading channels provided that the SNR is sufficiently high. Such schemes are also effective in multi-path channels [10].

Although there has been no study on the impact of channel estimation on the design of the transmitter and receiver for a BICM-ID-SSD system, the use of PSAM or superimposed schemes was widely considered for coded and un-coded systems, see, e.g., [11–17]. In particular, using pilot symbols for channel estimation in turbo coded systems has been examined in [11]. The study assumed that the channel is flat fading with a gain that depends on time but is constant for the duration of a symbol. Such channels are referred to as correlated fading channels. The study in [11] used an iteratively filtered PSAM (IF-PSAM) algorithm in the receiver. Reference [12] proposes sparsely interleaved estimation and decoding (SIED) as an alternative algorithm for iterative channel estimation in BICM-ID systems. In these papers, pilot symbols are used to initialize the channel estimates and then these estimates are refined in subsequent iterations by an interpolator that uses information fed back from the decoder. In [12] it was shown that in order to improve the performance, the refining phase of the channel estimation is not needed for each iteration of decoding stage. In [13], the performances of PSAM and superimposed training schemes are compared using the worst-case un-coded BER as the performance measure. It was shown for uncoded systems at high SNR that PSAM performs better for slowly time-varying fading channels and the superimposed training scheme performs better in fast time-varying fading channels. This observation also applies to coded systems. For BICM systems, performance of PSAM has been studied in [14] for a range of SNR. In particular, the relation between fade rates and required power of pilots has been investigated using extrinsic information transfer (EXIT) chart. Similar analysis

Hereafter the abbreviation BICM-ID-SSD refers to a system that implements bit-interleaved coded modulation, iterative decoding and signal space diversity technique.

has been investigated for coded systems with superimposed schemes in [15] using BER simulation. However, at low SNR in a fast time-varying fading environment, conventional training schemes fail. On the other hand, since BICM works well at low SNR and over a fast time-varying channel, it is of interest to develop a new training scheme for this system.

This paper proposes and examines alternative training sequences for a BICM-ID-SSD system operating over a correlated fading channel. Specifically, the paper investigates injecting pilots at the bit level instead of the symbol level. The intent is to exploit SSD technique to improve the performance of the iterative receiver. The proposed method inserts pilot bits into the coded bits before constellation mapping and signal rotation. As a result, the rotated superimposed symbols, which shall be referred to as SSD-pilots, can be considered as a superimposed training sequence that is inherent to the rotated transmitted symbols. A soft iterative receiver is developed for the bit level embedded training sequence. In particular, the estimator in the proposed receiver uses the training sequence in addition to the soft information from the decoder to produce estimates of the time-varying channel gains for the demodulator. Moreover, the demodulator uses the training part in the received signal to provide reliable *extrinsic* information for the decoder. It shall be demonstrated that performance of the proposed receiver is significantly better than that of the conventional PSAM scheme.

The paper is organized as follows. The system model of BICM-ID-SSD is presented in Section 3.2. In this section, the structure of the training sequence is described and the transmitter and receiver are developed for the proposed method of training design. Section 3.3 analyzes the performance of the BICM-ID-SSD system over a correlated fading channel. In particular a lower bound of the asymptotic BER is first obtained under the perfect CSI. Then the Cramer-Rao bound (CRB) for the proposed channel estimation method is derived. Section 4.5 provides numerical results and performance comparisons. Section 6.6 offers conclusion.

Notation: Upper and lower boldface letters denote matrices and column vectors,

respectively. Superscripts $(\cdot)^H$ and $(\cdot)^T$ indicate Hermitian and transpose, respectively. For matrix $\mathbf{A}$, $\mathbf{A}_{i,j}$ denotes its (i,j)th entry while $\mathbf{A}_j$ indicates the jth row. The log-likelihood ratio of bit b is defined as $\Lambda^{(b)} = \log\left[\frac{P(b=1)}{P(b=-1)}\right]$.

3.2 System Model

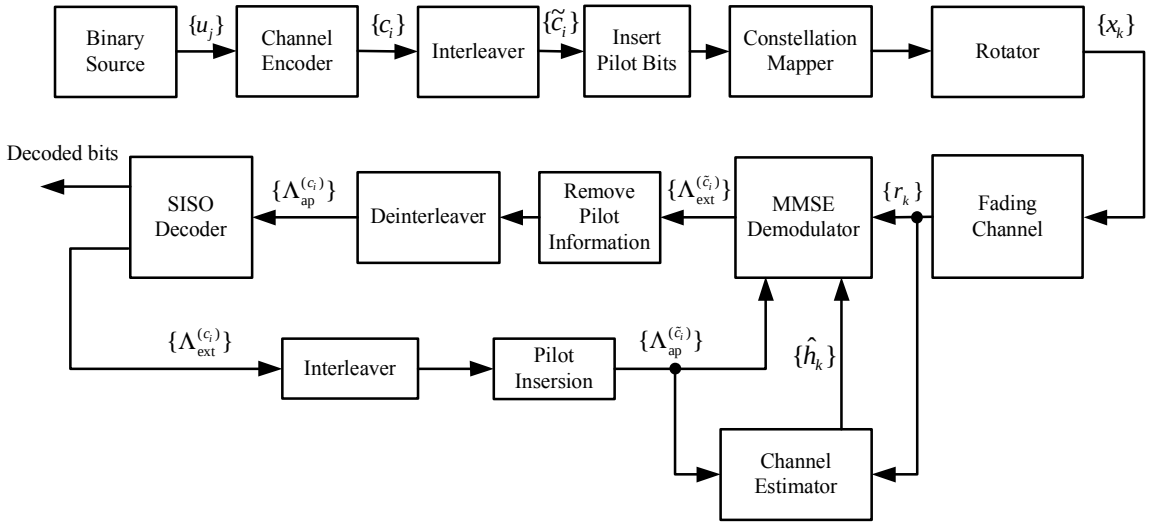

Figure 3.1 Block diagram of BICM-ID-SSD with the proposed pilot insertion and iterative channel estimation.

The block diagram of the transmitter and receiver in discrete-time equivalent baseband is shown in Fig. 4.2.1.

3.2.1 Transmitter

A sequence $\{u_j\}$, $1 \leq j \leq L$, of information bits is first encoded by a rate-r convolutional encoder. The encoded bits $\{c_i\}$, $1 \leq i \leq L/r$ are then passed to a bit-interleaver with length of L/r. Next, the interleaved sequence $\{\tilde{c}_i\}$ is segmented into groups of $(N-N_p)\times m$ bits, where N is the size of the rotation matrix, N_p is the number of pilot symbols in N rotated symbols and m is the number of bits carried by one symbol of a QAM constellation whose size is $|\Omega| = 2^m$. Next, known pilot bits $\{p_n\}$, $n = 1, 2, \ldots, mN_p$, is inserted in the segmented group of $(N - N_p) \times m$ bits. The pilot bits can be placed at the beginning, at the end, or somewhere in between as long as every m pilot bits are inserted as a block in the segmented group

and mapped to one QAM constellation point. This is necessary so that the channel estimator can work with known pilot QAM symbols in the initialization phase. Each expanded group of size Nm is then mapped to one complex N-dimensional "super" symbol, $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$.

There are many possible mappings between Nm bits and a super symbol. The sigma mapping scheme in [3] facilitates a low-complexity soft-output minimum mean-square error (MMSE) demodulator and for that reason is used here. In general, the sigma mapping relates a binary vector $\mathbf{z}$ to the super symbol $\mathbf{s}$ with the equation,

$$\mathbf{s} \quad = \quad \mathbf{V}(2\mathbf{z} - 1) = \mathbf{V}\mathbf{b}, \tag{3.1}$$

where $\mathbf{V}$, $\mathbf{z}$ and $\mathbf{b}$ are defined as follows:

- $\mathbf{V} = \text{diag}\underbrace{(\mathbf{v}, \mathbf{v}, \ldots, \mathbf{v})}_{N\mathbf{v}'s}$ is an $N \times Nm$ block diagonal matrix and $\mathbf{v} = [v_1, v_2, \ldots, v_m]$ is the basis vector whose elements are unit-norm complex numbers.
- $\mathbf{z}$ is an $Nm \times 1$ binary vector of interleaved coded bits and pilot bits whose entries are either 1 or 0.
- $\mathbf{b}$ is a binary vector whose entries are ± 1 and represents both pilot and coded bits.

With the proper design of $\mathbf{v}$, each component s_i is guaranteed to belong to a standard two-dimensional QAM constellation Ω [3]. As an example, Fig. 3.2 shows the sigma mapping of 4-QAM and 16-QAM.

Rotation is accomplished with a matrix, $\mathbf{G}$, which could take on one of several forms. One form which can be used when N is a power of 2 [3], is

$$\mathbf{G} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \alpha_1 & \ldots & \alpha_1^{N-1} \\ 1 & \alpha_2 & \ldots & \alpha_2^{N-1} \\ \vdots & \vdots & & \vdots \\ 1 & \alpha_N & \ldots & \alpha_N^{N-1} \end{pmatrix}, \tag{3.2}$$

where $\alpha_i = \exp\left(\jmath\frac{2\pi}{4N}\right)\exp\left(\jmath\frac{2\pi(i-1)}{N}\right)$ and $\jmath = \sqrt{-1}$. The matrix rotates a super symbol, $\mathbf{s}$ into a rotated super symbol, denoted by $\mathbf{x} = \mathbf{Gs}$.

Let $\mathbf{\Gamma} = \mathbf{GV}$ and introduce subscripts as the time index to the super symbols and rotated super symbols. Then the lth rotated super symbol can be represented by,

$$\begin{aligned}\mathbf{x}_l &= [x_{(l-1)N+1}, \ldots, x_{(l-1)N+N}]^T \\ &= \mathbf{\Gamma}[b_{(l-1)Nm+1}, \ldots, b_{(l-1)Nm+Nm}]^T \\ &= \mathbf{\Gamma}\mathbf{b}_l, \quad 1 \le l \le \frac{1}{N-N_p}\frac{L}{rm}. \end{aligned} \tag{3.3}$$

Thus by showing the time index in form of $k = (l-1)N + i$, $1 \le i \le N$, the kth transmitted complex symbol is given by,

$$x_k = \mathbf{\Gamma}_i \mathbf{b}_l = x_k^{(d)} + x_k^{(p)}, \quad k = 1, 2, \ldots, \frac{N}{N-N_p}\frac{L}{rm}, \tag{3.4}$$

where $x_k^{(d)}$ and $x_k^{(p)}$ are the data and pilot portions of the transmitted symbols, respectively, and $l = \lfloor \frac{k}{N} \rfloor$ and $i = k-(l-1)N$. For the ease of exposition, it is assumed that all the pilot bits take on the same value of 1.

It is of interest to measure the effective energy of the system. The energy efficiency of SSD-pilot scheme can be expressed as

$$\varrho_{\text{SSD-pilot}} = \frac{(N-N_p)E_s}{(N-N_p)E_s + N_p E_p} = \frac{N-N_p}{N-N_p+\eta N_p}, \tag{3.5}$$

when $E_p = \eta E_s$, and E_p and E_s are the energy per pilot symbol and the energy per data symbol before rotation, respectively. For the purpose of comparison to PSAM, it is pointed out that the energy efficiency of PSAM is given by $\frac{M-1}{M-1+\eta}$, where M is the pilot symbol spacing [12].

3.2.2 Iterative Receiver

The sequence of received symbols, denoted $\{r_k\}$, is given by,

$$r_k = h_k x_k + w_k, \quad k = 1, 2, \ldots, \frac{N}{N-N_p}\frac{L}{rm}, \tag{3.6}$$

where w_k is a noise sample and h_k is a channel gain at time k. The sequence of $\{w_k\}$ is a set of i.i.d. zero-mean circularly symmetric Gaussian random variables

with variance N_0. The sequence $\{h_k\}$ contains complex Gaussian random variables with zero mean and autocorrelation $R(k) = \sigma_h^2 J_0(2\pi f_d T_s k)$, which is given by the Jakes's model [18]. The function $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, f_d is the Doppler spread (assumed to be known at the receiver) and T_s is the symbol duration.

An iterative receiver operates on the received signals in an iterative manner. In the first iteration, the received signals are filtered to produce initial estimates of the channel gains, denoted $\{\hat{h}_k\}$. In essence, only the known information, i.e., pilot portions of the transmitted signals, $\{x_k^{(p)}\}$, is used. The extrinsic information for the de-mapped bits, $\{\Lambda_{\text{ext}}^{(\tilde{c}_i)}\}$, is obtained in the same iteration using the soft-output MMSE demodulator proposed in [19]. The mismatched demodulator uses $\{\hat{h}_k\}$, the *a priori* information of the coded bits, $\{\Lambda_{\text{ap}}^{(c_i)}\}$, which is "0", and the extrinsic information of the pilot bits, $\{\Lambda_{\text{ap}}^{(p_n)}\}$ to obtain $\{\Lambda_{\text{ext}}^{(\tilde{c}_i)}\}$. The demodulator can be simplified without sacrificing significant accuracy by using a smaller number for $\Lambda_{\text{ap}}^{(p_n)}$, say ± 100, which has a theoretical value of $\pm\infty$. After demodulation, the extrinsic information of the de-mapped pilot bits, $\{\Lambda_{\text{ext}}^{(p_n)}\}$, is removed and the sequence is de-interleaved. Next a SISO decoder takes the de-interleaved version of $\{\Lambda_{\text{ext}}^{(\tilde{c}_i)}\}$ and treats it as a priori information $\{\Lambda_{\text{ap}}^{(c_i)}\}$ to compute $\{\Lambda_{\text{ext}}^{(c_i)}\}$ of the coded bits. Here, the SISO decoder uses the maximum a posteriori probability (MAP) algorithm [20].

In subsequent iterations, soft information from the decoder is used to improve the performance of the channel estimator. The channel estimator uses such information to compute new estimates of the channel coefficients using expected values of the data symbols. Therefore the interleaved $\{\Lambda_{\text{ext}}^{(c_i)}\}$ from the decoder are fed back to the estimator to calculate the expected values and variances of the data symbols, i.e., $E\{x_k\}$ and $\sigma_{x_k}^2$. Using (3.4), $E\{x_k\}$ and $\sigma_{x_k}^2$ are computed as,

$$E\{x_k\} = \mathbf{\Gamma}_i E\{\mathbf{b}_l\} \tag{3.7}$$

$$\sigma_{x_k}^2 = \sum_{j=1}^{Nm} |\mathbf{\Gamma}_{i,j}|^2 \sigma_{b_{Nm(l-1)+j}}^2 = \frac{1}{N} \sum_{j=1}^{Nm} \sigma_{b_{Nm(l-1)+j}}^2, \tag{3.8}$$

where the $\{\Lambda_{\text{ap}}^{(\tilde{c}_i)}\}$ should be used to calculate $E\{b_i\} = \tanh\left[\frac{\Lambda^{(b_i)}}{2}\right]$ and $\sigma_{b_i}^2 = 1 - |E\{b_i\}|^2$. For the pilot bits, since $\{p_n\}$ are known at the receiver, one has $E\{p_n\} = p_n$.

The demodulator uses the $\{\Lambda_{\text{ap}}^{(\tilde{c}_i)}\}$ and $\Lambda_{\text{ap}}^{(p_n)}$ directly to calculate $\{\Lambda_{\text{ext}}^{(\tilde{c}_i)}\}$ for the rest of the iterations, as in the first iteration. The detailed operation of the channel estimator is discussed next.

3.2.3 Channel Estimator

In this section, an iterative MMSE channel estimator is developed for the SSD-pilot training sequence. It is shown in Section 4.5 that inserting pilot bits with the coded bits lowers error probability in each successive iteration. If the transmitted sequence was known at the receiver, then the best linear MMSE estimates of the channel gains would be found by using the Wiener interpolator. However, using the pilot portions of the transmitted signals for the first iteration and the soft information of the transmitted signals for the successive iterations produces a good approximation of the channel gains.

The kth channel gain is estimated by a filter of length $(2K_0 + 1)$, where K_0 is a constant. The tap weights of this filter are updated in each iteration. To get a compact expression for $\hat{h}_k$ the following definitions are used:

- The observation vector $\breve{\mathbf{r}}_k = [r_{k-K_0}, \ldots, r_k, \ldots, r_{k+K_0}]^T$ is the collection of the $(2K_0 + 1)$ received symbols nearest to r_k.
- The pilot portion of the transmitted symbols is $\breve{\mathbf{x}}_k^{(p)} = [x_{k-K_0}^{(p)}, \ldots, x_k^{(p)}, \ldots, x_{k+K_0}^{(p)}]^T$, which is a known vector at the receiver.
- The vector of the transmitted symbols, $\breve{\mathbf{x}}_k = [x_{k-K_0}, \ldots, x_k, \ldots, x_{k+K_0}]^T$ is the collection of the $(2K_0 + 1)$ symbols that surround x_k.

In general the choice of K_0 is governed by the trade-off between complexity and performance. A detailed discussion on how to choose K_0 and its effect on the performance can be found in [12] and [11].

- The covariance matrix of the observation vector is obtained with $E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}$.
- The covariance vector, which has length $(2K_0+1)$ is obtained by $E\{\breve{\mathbf{r}}_k h_k^*\}$.

With these definitions, the linear MMSE estimate of the channel gain is

$$\hat{h}_k \;=\; \left(E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}^{-1}E\{\breve{\mathbf{r}}_k h_k^*\}\right)^H \breve{\mathbf{r}}_k. \tag{3.9}$$

Expressions for the covariance matrix and covariance vector must be obtained to calculate $\hat{h}_k$ with (3.9) for the first iteration, where only the pilot portions of the received signals are known. For the pilot portion $E\left\{x_i^{(p)}(x_j^{(p)})^*\right\} = x_i^{(p)}(x_j^{(p)})^*$, $\forall i,j$ and $E\{x_i^{(p)}\} = x_i^{(p)}$. For the data portion of the transmitted signal, $E\left\{x_i^{(d)}(x_j^{(d)})^*\right\} = \left(\frac{N-Np}{N}\right)E_s$ for $i=j$, $E\left\{x_i^{(d)}(x_j^{(d)})^*\right\} = 0$ for $i \neq j$ and $E\{x_i^{(d)}\} = 0$ for the first iteration. Assembling this information leads to

$$E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}_{i,j} = \begin{cases} \left(|\breve{x}_{k,i}^{(p)}|^2 + \frac{N-Np}{N}E_s\right)\sigma_h^2 + N_0, & i=j \\ \breve{x}_{k,i}^{(p)}(\breve{x}_{k,j}^{(p)})^*\sigma_h^2 J_0(2\pi f_d|i-j|T_s), & i\neq j \end{cases}, \tag{3.10}$$

where $E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}_{i,j}$ is the (i,j)th element of the covariance matrix and $\breve{x}_{k,i}^{(p)} = x_{k+i-(K_0+1)}^{(p)}$ is the ith element of vector $\breve{\mathbf{x}}_k^{(p)}$. In addition, $E\{\breve{\mathbf{r}}_k h_k^*\}_i = \sigma_h^2 J_0(2\pi f_d|i|T_s)\breve{x}_{k,i}^{(p)}$. Note that the above calculations imply that the matrix $E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}$ and vector $E\{\breve{\mathbf{r}}_k h_k^*\}$ can be calculated off-line for the first iteration.

In the next iterations, for calculating the covariance matrix and covariance vector, the soft information of data is taken into account. Using (3.7), $E\{x_i(x_j)^*\} = |E\{x_i\}|^2 + \sigma_{x_i}^2$ if $i=j$ and $E\{x_i(x_j)^*\} = E\{x_i\}E\{x_j\}^*$ for $i\neq j$. Then the elements of the covariance matrix are,

$$E\{\breve{\mathbf{r}}_k\breve{\mathbf{r}}_k^H\}_{i,j} = \begin{cases} \left(|E\{\breve{x}_{k,i}\}|^2 + \sigma_{\breve{x}_{k,i}}^2\right)\sigma_h^2 + N_0, & i=j \\ \left(E\{\breve{x}_{k,i}\}E\{\breve{x}_{k,j}\}^*\right)\sigma_h^2 J_0(2\pi f_d|i-j|T_s), & i\neq j \end{cases}, \tag{3.11}$$

and the entries of the covariance vector are obtained as $E\{\breve{\mathbf{r}}_k h_k^*\}_i = \sigma_h^2 J_0(2\pi f_d|i|T_s)\times E\{\breve{x}_{k,i}\}$. In the above last two expressions $E\{\breve{x}_{k,i}\}$ and $\sigma_{\breve{x}_{k,i}}^2$ are computed using (3.7) and (8), respectively, and by associating $\breve{x}_{k,i}$ to the ith element of vector $\breve{\mathbf{x}}_k$, i.e., $\breve{x}_{k,i} = x_{k+i-(K_0+1)}$.

3.3 Bound for Asymptotic Error Performance and Cramer-Rao Bound for Channel Estimation

3.3.1 Analytical Bound of Asymptotic BER with Perfect CSI

The union bound on the BER for BICM-ID-SSD is helpful for evaluating the system performance in a correlated fading channel. The bound for a rate-r convolutional code can be written as,

$$P_b \;\leq\; \frac{1}{r}\sum_{d=d_H}^{\infty} c_d f(d,\Psi,\zeta), \tag{3.12}$$

where parameters d, d_H, c_d, Ψ, and ζ as well as function $f(\cdot)$ are defined as: d is the Hamming distance, d_H is the free Hamming distance of the code, c_d is the total input weight of error events at Hamming distance d, Ψ is the N-dimensional complex constellation set, and ζ is the mapping rule. The function $f(d,\Psi,\zeta)$ denotes the average pairwise error probability (PEP) , which depends on the Hamming distance d, the signal constellation set Ψ and the mapping rule ζ. In general, $f(d,\Psi,\zeta)$ can be approximated as in [3] by,

$$f(d,\Psi,\zeta) \;\approx\; \frac{1}{2}\left(\delta(\mathbf{G},\Omega,\zeta)\right)^d, \tag{3.13}$$

where, for an uncorrelated Rayleigh fading channel,

$$\delta(\mathbf{G},\Omega,\zeta) = \frac{1}{N}\frac{1}{m2^m}\sum_{u=1}^{N}\left[\sum_{k:s_k\in\Omega}\sum_{j=1}^{m}\prod_{i=1}^{N}\left(1+\frac{|\mathbf{G}_{i,u}(s_k-s_{k\sim j})|^2}{4N_0}\right)^{-1}\right]. \tag{3.14}$$

In the above s_k is a constellation point in Ω and notation $k\sim j$ is used to refer to a signal whose mapping is the same as that of s_k, except that the jth bit is toggled (i.e. complemented). As an example for 16-QAM, if s_k has the mapping 1111 and $j=2$ then $s_{k\sim j}$ would be the signal that has the mapping $11\tilde{1}1 = 1101$. As an illustration, the signals s_{15} and $s_{15\sim 2}$ are shown in Fig. 3.2 for the sigma mapping scheme.

The union bound over a correlated fading channel is now derived using the procedure of [3]. To account for the correlation, the joint probability density function of the channel is expressed as

$$p(\mathbf{h}) \;=\; \frac{1}{\pi^N \det\mathbf{C}}\exp\left(-\mathbf{h}^H\mathbf{C}^{-1}\mathbf{h}\right), \tag{3.15}$$

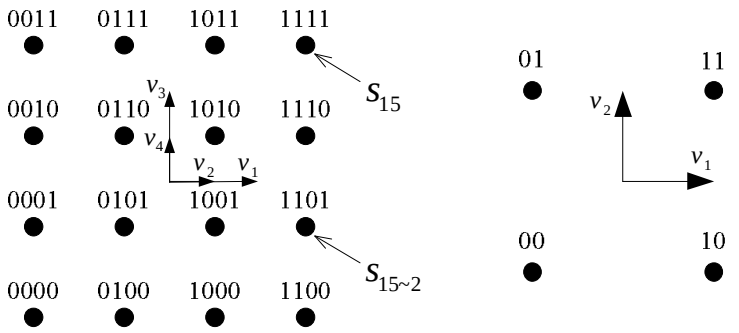

Figure 3.2 Sigma mappings for 16-QAM and 4-QAM. For 16-QAM the basis vectors are $v_1 = 2v_2 = \sqrt{1.6}\exp(\jmath 0)$, $v_3 = 2v_4 = \sqrt{1.6}\exp(\jmath\frac{\pi}{2})$. For 4-QAM, $v_1 = \exp(\jmath 0)$ and $v_2 = \exp(\jmath\frac{\pi}{2})$.

where $\mathbf{h} = [h_1, \ldots, h_N]^T$ is a correlated fading vector and $\mathbf{C} = E\{\mathbf{h}\mathbf{h}^H\}$ is an $N \times N$ covariance matrix, with

$$\mathbf{C}_{i,j} = \begin{cases} \sigma_h^2, & i = j \\ \sigma_h^2 J_0(2\pi f_d(i-j)T_s), & i \neq j \end{cases}. \tag{3.16}$$

After some manipulation and then diagonalizing the data vector, i.e., $\mathbf{D}_u = \text{diag}(|\mathbf{G}_{1,u}(s_k - s_{k\sim j})|^2, \ldots, |\mathbf{G}_{N,u}(s_k - s_{k\sim j})|^2)$, and using the method in [21], one recognizes that $f(d, \Psi, \zeta)$ has the same form as (3.13), but with $\delta(\mathbf{G}, \Omega, \zeta)$ given as

$$\delta(\mathbf{G}, \Omega, \zeta) = \frac{1}{N}\frac{1}{m2^m}\sum_{u=1}^{N}\left[\sum_{k:s_k\in\Omega}\sum_{j=1}^{m}\prod_{i=1}^{N}\left(1 + \frac{\lambda_{i,u}}{4N_0}\right)^{-1}\right], \tag{3.17}$$

where $\lambda_{i,u}$ are the eigenvalues of $\mathbf{D}_u\mathbf{C}$. It is clear that there are at most N nonzero eigenvalues. It is also noted that the number of eigenvalues equals the average number of distinct channel states in one rotated super symbol. Therefore, it follows from (3.13) that the diversity order of the system must be less than or equal to Nd_H. If the rank of $\mathbf{D}_u\mathbf{C}$ is less than N, then the number of nonzero eigenvalues and therefore the diversity order is less than or equal to the rank of $\mathbf{C}$ times d_H. For example, if $f_dT_s = 0.02$, the rank of $\mathbf{C}$ is 8 when $N = 32$. This implies that the maximum diversity order is $8d_H$.

3.3.2 Cramer-Rao Bound for Mean-Square Error of the Channel Estimation

Cramer-Rao bound (CRB) is widely used as a benchmark for the performance of any channel estimator. The bound states that the mean-square error (MSE) matrix of any unbiased estimator $\hat{\mathbf{h}}$ is lower bounded by [22],

$$\text{MSE}(\hat{\mathbf{h}}) \equiv E\{[\hat{\mathbf{h}} - \mathbf{h}][\hat{\mathbf{h}} - \mathbf{h}]^H\} \geq \mathbf{J}(\mathbf{h})^{-1},$$

where $\mathbf{J}(\mathbf{h})$ is the complex Fisher information matrix defined by,

$$\mathbf{J}(\mathbf{h}) = E\left\{\left[\frac{\partial \ln p(\mathbf{r},\mathbf{h})}{\partial \mathbf{h}^*}\right]\left[\frac{\partial \ln p(\mathbf{r},\mathbf{h})}{\partial \mathbf{h}^*}\right]^H\right\}. \tag{3.18}$$

A closed form expression for $\mathbf{J}(\mathbf{h})$ is found by taking the expectation of (5.6) over $\mathbf{r}$ and $\mathbf{h}$. To this end, it is convenient to express $p(\mathbf{r},\mathbf{h})$ as $p(\mathbf{r}|\mathbf{h}) \cdot p(\mathbf{h})$, where $p(\mathbf{h})$ is defined in (3.15), $p(\mathbf{r}|\mathbf{h}) = \frac{1}{(\pi N_0)^{\frac{N}{2}}} \exp\left(-\frac{1}{N_0}\|\mathbf{r} - \mathbf{X}_D\mathbf{h}\|^2\right)$, $\mathbf{r} = [r_1, \ldots, r_N]^T = \mathbf{X}_D\mathbf{h} + \mathbf{w}$ and $\mathbf{X}_D = \text{diag}(x_1, \ldots, x_N)$. After some manipulations, the Fisher information matrix is given in terms of $\mathbf{X}_D$ as follows:

$$\mathbf{J}(\mathbf{h}) = \frac{1}{N_0} E\{\mathbf{X}_D\mathbf{X}_D^H\} + \mathbf{C}^{-1}. \tag{3.19}$$

Note that $E\{\mathbf{X}_D\mathbf{X}_D^H\}$ is a diagonal matrix which has $E\{|x_i|^2\} = |E\{x_i\}|^2 + \sigma_{x_i}^2$ on its diagonal, $i = 1, \ldots, N$. The terms $E\{x_i\}$ and $\sigma_{x_i}^2$ can be numerically calculated based on (3.7) and (3.8). The vast simulations are arranged even at low SNR to obtain an accurate CRB.

3.4 Simulation Results

The performance of the proposed method is investigated using a computer simulation to calculate the MSE of the channel estimator and the resulting BER. The same convolutional code, data frame, bit-interleaver and modulation format is used in all simulations. The channel code is a rate-1/2 convolutional code with constraint length 5 and generator polynomial $[1 + D + D^3 + D^4; 1 + D + D^4]$. The data frame consists of $11,996$ data bits and 4 tail bits. The bit-interleaver has length $L = 2 \times 12,000$. The

modulation is quadrature phase-shift keying (QPSK) with sigma mapping, where the basis vector is $\mathbf{v} = [1, \jmath]$.

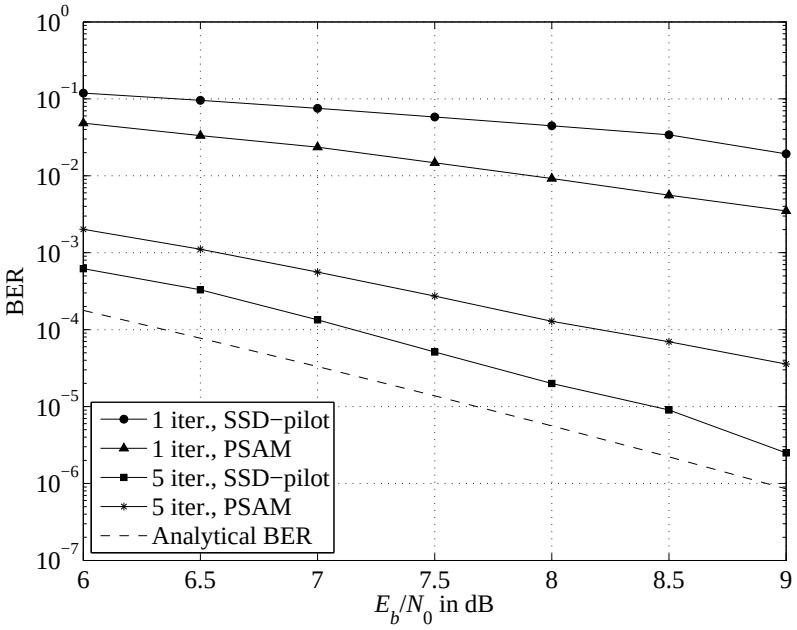

Figure 3.3 Comparison of BER obtained from SSD-pilot and PSAM with the iterative channel estimator when $N = 16$, $f_d T_s = 0.02$ and $N_p = 1$, and for 1 and 5 iterations.

Fig. 4.9 plots BER versus E_b/N_0, where E_b is the energy per information bit, and the normalized fade rate is taken to be $f_d T_s = 0.02$. The size of the rotation matrix is $N = 16$ while $N_p = 1$ and $\eta = 1$. From (3.5) this means that the energy efficiency is $\varrho = 0.9375$, since $E_p = E_s = 2$. The length of the Wiener filter is set to $2K_0 + 1$ with $K_0 = 40$. This has the Wiener filter estimating each channel tap with 81 channel realizations. For each value of E_b/N_0 considered, 10^5 independent frames were run.

The performance of the SSD-pilot scheme is compared to the performance of the PSAM training scheme as well as the analytical bound in Fig. 4.9. Note that for a fair comparison, the channel estimator used with PSAM is also based on the iterative approach with rotated data symbols. For the PSAM scheme, the pilot symbol is $1 + \jmath$ and the pilot spacing is $M = 16$, which yields the same energy efficiency as that of SSD-pilot scheme. The other simulation parameters are the same as in SSD-pilot scheme.

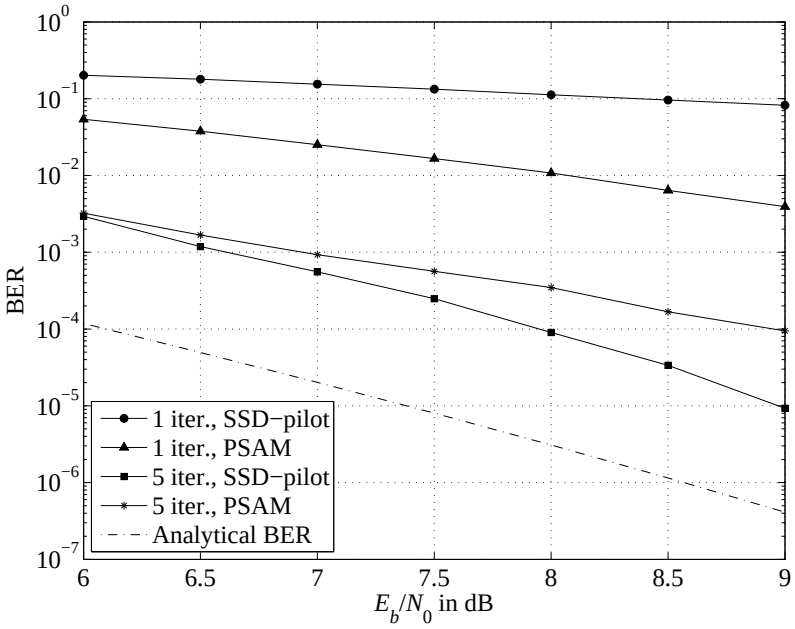

Figure 3.4 Comparison of BER obtained from SSD-pilot and PSAM with the iterative channel estimator when $N = 8$, $f_d T_s = 0.05$ and $N_p = 1$, and for 1 and 5 iterations.

There are five curves in Fig. 4.9. The top two provide the performance after the first iteration for the SSD-pilot and the PSAM schemes. The middle two curves show the performance after 5 iterations. The bottom curve, shown as a dashed line, is a plot of the bound. The curves for the first iteration show that the performance of the proposed scheme is about 2 dB worse than the conventional PSAM scheme. This is expected because only the pilot portions of the transmitted symbols are used for channel estimation in the first iteration in SSD-pilot scheme. However, the results are quite different after 5 iterations. The SSD-pilot scheme is from 1 to 1.5 dB better than the PSAM scheme, depending on E_b/N_0. The reason for this is that the pilot information is embedded in the rotated symbols for the SSD-pilot scheme and not for the PSAM scheme. Since the SSD-pilots are embedded in the SSD received symbols, the demodulator can make use of this information for its training. In contrast, the pilots in PSAM, which are not embedded in the rotated symbols cannot be used by the demodulator. In the end the SSD-pilot information used in the demodulator iteratively improves the overall performance of the receiver. The improvement is such

that after 5 iterations the performance closely approaches the analytical bound. Note that there is a gap between the union bounds and the BER curves, which is because the union bound is calculated assuming perfect CSI.

Fig. 3.4 has the same curves as Fig. 4.9 for a fade rate of $f_dT_s = 0.05$. For this larger fade rate, parameters N, N_p and K_0 are changed to accommodate better channel tracking. The following values were used: $N = 8$, $N_p = 1$, and $K_0 = 45$. For the PSAM scheme, M was changed from 16 to 8 to make the energy efficiency the same. It is observed that the SSD-pilot method also outperforms conventional PSAM under this faster fade rate. SSD-pilot gain is as high as 1.25 dB at lower BERs. From these two figures it can be seen that the tracking ability of SSD-pilot is better than PSAM scheme.

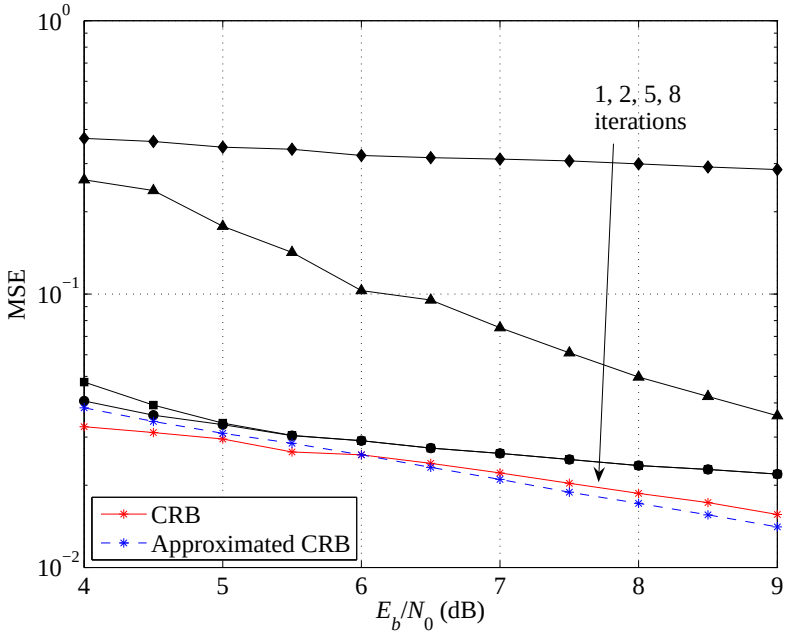

Figure 3.5 MSE obtained from SSD-pilot with the iterative channel estimator when $N = 16$, $f_dT_s = 0.02$ and $N_p = 1$, and for 1, 2, 5 and 8 iterations.

The MSE performance of the SSD-pilot estimator is shown in Fig. 4.10 and Fig. 3.6 after 1, 2, 5 and 8 iterations for $f_dT_s = 0.02$ and $f_dT_s = 0.05$, respectively. The CRBs of the channel estimator are also plotted in Fig. 4.10 and Fig. 3.6 (dashed lines) for

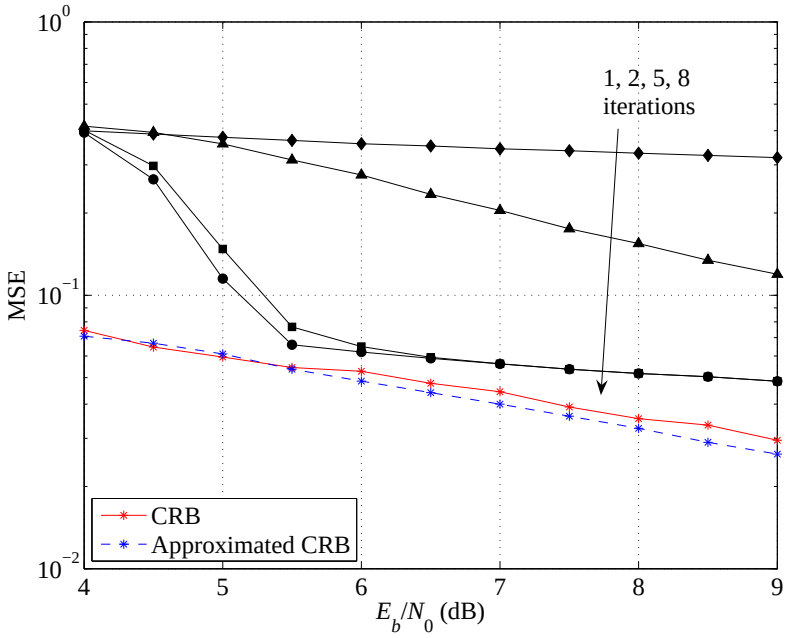

Figure 3.6 MSE obtained from SSD-pilot with the iterative channel estimator when $N = 8$, $f_d T_s = 0.05$ and $N_p = 1$, and for 1, 2, 5 and 8 iterations.

comparison. The simulation results after 5 iterations are very close to the CRBs in both figures. It is observed that MSE curves for $E_b/N_0 \geq 5$ dB exhibit flat slopes with higher numbers of iterations. This is explained as follows. In high E_b/N_0 region, the genie condition (i.e., error-free feedback from the decoder) is basically achieved after a sufficiently large number of iterations (say 5 iterations). This condition implies that, $E\{|x_i|^2\} = E_s$ in (3.19). Therefore the slope of the CRB can be approximated by $(\frac{E_s}{N_0}\mathbf{I}_N + \mathbf{C}^{-1})^{-1}$ where $\mathbf{I}_N$ is an identity matrix with size N and $E_p = E_s$. This approximation is also plotted in Figs. 4.10 and 3.6 and the approximated curves are seen to tightly approach the CRB curves.

3.5 Conclusions

In this paper, a superimposed training sequence for a BICM-ID-SSD system over a correlated fading channel has been proposed. Information of the pilot bits, which are inserted into the coded bits before modulation and signal rotation, has been used

by the channel estimator and the demodulator. The soft iterative channel estimator was developed to exploit the pilot information and the soft information of the rotated symbols. In order to analyze the effectiveness of using SSD-pilot symbols, the analytical bound of the BER over a correlated channel with perfect CSI was obtained. Simulation results verified that, compared to PSAM, the BER performance is improved by about 1 dB at the BER level of 10^{-4} under both fade rates of 0.02 and 0.05 and at the same energy efficiency. Finally, a Cramer-Rao bound for the MSE of the estimated channel coefficients was derived. Simulation results showed that the MSE of the iterative channel estimator closely approaches the Cramer-Rao bound after 5 iterations.

References

[1] K. Boulle and J. C. Belfiore, "Modulation schemes designed for the Rayleigh channel," *Proc. CISS Conf.*, pp. 288–293, Mar. 1992.

[2] J. Boutros and E. Viterbo, "Signal space diversity: A power- and bandwidth-efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1453–1467, Jul. 1998.

[3] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1732–1742, May 2007.

[4] K. Choi, "Pilot overhead reduction in turbo coded OFDM systems employing an iterative channel estimation under low signal-to-noise ratio environments," *IET Communications*, vol. 4, no. 3, pp. 312–321, 2010.

[5] J.-W. Kim and W.-S. Yoon, "Turbo code with iterative channel estimator using soft-output of turbo decoder," *Electronics Letters*, vol. 36, no. 18, pp. 1560–1562, 2000.

[6] M. Nicoli, S. Ferrara, and U. Spagnolini, "Soft-iterative channel estimation:

Methods and performance analysis," *IEEE Trans. Signal Processing.*, vol. 55, pp. 2993–3006, Jun. 2007.

[7] M. A. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3165–3173, Nov. 2006.

[8] C. Tellambura and V. K. Bhargava, "Convolutionally coded binary PSAM for Rayleigh fading channels," *Electronics Letters*, vol. 28, no. 16, pp. 1503–1505, 1992.

[9] S. Godtmann, L. I-Wei, G. Ascheid, C. Tzi-Dar, and H. Meyr, "Tight approximation of the bit error rate for BICM(-ID) assuming imperfect CSI," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 11, pp. 4468–4473, 2008.

[10] Y. Gong, Z. Ding, T. Ratnarajah, and C. F. N. Cowan, "Turbo channel estimation and equalisation for a superposition-based cooperative system," *IET Communications*, vol. 3, no. 11, pp. 1790–1799, 2009.

[11] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 1697–1705, Sep. 2001.

[12] Y. Huang and J. A. Ritcey, "Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2549–2558, Sep. 2005.

[13] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Processing.*, vol. 52, pp. 1403–1418, May 2004.

[14] H. Niu and J. Ritcey, "Pilot-symbol-assisted LDPC coded BICM over correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2076–2082, Sep. 2005.

[15] H. M. Karkhanechi and B. C. Levy , "An Efficient Adaptive Channel Estimation Algorithm for MIMO OFDM Systems–Study of Doppler Spread Tolerance," *Journal of Signal Processing Systems*, vol. 56, no. 2–3, pp. 261–271, Sep. 2009.

[16] V. Ramon, X. Wautelet, L. Vandendorpe, and L. V. D. Perre, "Low-complexity iterative estimators of quasi static frequency selective channels," *Proc. IEEE Global Telecommun. Conf.*, (Washington, DC), pp. 3002-3007, Nov. 2007.

[17] M. Teimouri, N. Rezaee, and A. Hedayat, "Concatenated coded modulation techniques and orthogonal space-time block codes in the presence of fading channel estimation errors," *IET Communications*, vol. 4, no. 2, pp. 135–143, 2010.

[18] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley and Sons Inc., 1974.

[19] L. Ping, "Approximate MMSE-APP estimation for linear systems with binary inputs," *IEEE Commun. Letters*, vol. 9, pp. 172–174, Feb. 2005.

[20] S. Benedetto, D. Divsalar, G. Montorsi, and F.Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan. 1997.

[21] K. Leeuwin, J. C. Belfore, and G. K. Kaleh, "Chernoff bound of trellis-coded modulation over correlated Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 2506–2511, Aug. 1994.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. new Jersey: Prentice-Hall PTR, 1993.

4. Analyzing BICM-MIMO Systems with Channel Estimation Error

Published as:

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt, "Analyzing BICM-MIMO Systems with Channel Estimation Error", to appear in *IET Communications*.

As mentioned in Chapter 2, the two main channels considered in this thesis are fast time-varying and slow (or block-fading) fading channels. The manuscript in Chapter 3 concentrated on BICM systems in fast-time varying channels. In fact, the diversity techniques that are employed in these systems are time diversity in the form of SSD. However, using space diversity is attractive in slow time-varying flat fading channels, since other diversity resources are poor in these environments. Using MIMO technology in conjunction with BICM is considered as one of the attractive space-time coding techniques.

The manuscript included in this chapter studies the operation of channel estimators in BICM-MIMO systems, especially an iterative channel estimator. As mentioned in Chapter 2, the iteration of the channel estimator differs from the iteration of the decoder and demodulator. Built on the fact that the iterative channel estimator uses the output of the decoder at each iteration, the performance of the channel estimator depends on the reliability of the output of the decoder. A threshold is therefore used to indicate in which decoding iteration the iterative channel estimator should be activated. In addition, it is shown that combining training sequence with soft information at each iteration of the decoder makes the estimation iteration to follow

the iteration of the decoder. Simulation results are provided to confirm that the two proposed estimation schemes can significantly improve the BER performance of the BICM-MIMO when compared to conventional schemes.

Analyzing BICM-MIMO Systems with Channel Estimation Error

Zohreh Andalibi, Ha H. Nguyen, J. Eric Salt

Abstract

The performance of bit-interleaved coded modulation in multiple-input multiple-output (BICM-MIMO) systems using an iterative channel estimator is analyzed. In a conventional iterative channel estimator, after initialization with the training phase, the channel estimator switches to the data phase. However, such a conventional iterative channel estimator does not always improve the performance of the receiver. In order to guarantee the performance improvement, a condition on when the output of the decoder should be used by the estimator is determined. Such a condition is related to the reliability of the soft information utilized by the channel estimator. The key in establishing this relationship is to use the mutual information (MI) that the observation vector has about the channel gains given the output of the decoder at each iteration. In this switch-augmented conventional iterative channel estimator, referred to as SAICE, the condition is theoretically found and indicates the needed reliability of the soft information for the channel estimator at the switching time. The switch-augmented scheme guarantees performance improvement of the iterative receiver with each iteration, however, it might need many iterations to converge for moderate to low signal-to-noise ratios (SNRs) . A less computationally intensive approach is to use both the training and data segments of the observation. This approach produces a combined iterative channel estimator (CICE) for BICM-MIMO systems. The performance behavior of the BICM-MIMO system is illustrated through the extrinsic information transfer (EXIT) chart with imperfect channel state information (CSI). Analytical results are verified with computer simulations.

Manuscript received September 27, 2010; revised February 1, 2012.

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt are with the Department of Electrical & Computer Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon, SK, Canada S7N 5A9, and TRLabs of Saskatchewan. Emails: z.andalibi@usask.ca, ha.nguyen@usask.ca, eric.salt@usask.ca.

Index terms

Bit-interleaved coded modulation, MIMO, iterative receiver, channel estimation, EXIT chart.

4.1 Introduction

Implementing bit-interleaved coded modulation (BICM) in a multiple-input multiple-output (MIMO) system can significantly improve the system's spectral efficiency, even at a low signal-to-noise ratio (SNR) [1, 2]. This achievement has been demonstrated largely with coherent detection and an assumption of *perfect* knowledge of the channel state information (CSI) at the receiver [3–5]. However, only imperfect channel state information (ICSI) can be realized in practice by a data-aided or semi-blind channel estimator in the receiver. The channel estimation error limits the performance of a BICM system and this has been the subject of various studies in recent years.

By investigating the effects of ICSI, solutions have been proposed to improve the system performance of coherent receivers in a BICM system [6–8]. Depending on how the system performance is defined, previous investigations can be classified in to two main categories: 1) information theoretic approaches, and 2) signal processing approaches. The information theoretic approaches have been used in [6] to calculate the achievable data rate for a MIMO system in the presence of channel estimation error. In particular, the effect of channel estimation error was represented as an SNR degradation, which was used to determine the number of pilot symbols needed in the pilot-symbol-assisted modulation (PSAM) based channel estimator. In [9], power and time allocation schemes were designed for single-input single-output systems, under both time-multiplexed and superimposed pilots.

The signal-processing approaches were used in [5,7,8,10,11] to investigate the effect of channel estimation error on the bit-error-rate (BER). In particular, this approach was used in [10] to relate the channel estimation error to both the BER and frame-error-rate (FER) . A similar approach was carried out in [7] and the authors suggest that an adaptive-rate PSAM can be used for a MIMO system with a maximum-ratio

combining (MRC) receiver. In [11] a BICM-MIMO system, which is also the system of interest in this paper, was analyzed to determine the effect of ICSI. This effect was represented as an SNR degradation for the case of a zero-forcing receiver and PSAM. A soft metric was derived in [5] for the MIMO receiver by taking the variance of the estimation error into account. The overhead required for the training sequence was established based on computer simulation.

In fact the information theoretic and signal processing approaches are connected. The SNR degradation is a key parameter in modeling the estimation error in both approaches (e.g., in [6] and [11]). The SNR degradation is a function of the mean squared channel estimation error (MSE). The function depends on the structure of the receiver and the operation of the channel estimator. The MSE, on the other hand, depends on the statistics of the channel, the estimation criterion and more importantly the design of the training in a training-based channel estimator.

Despite the abundance of literature on training-based channel estimators, other research works suggest that soft iterative channel estimators can improve the performance of the receiver. In [12], a comprehensive analysis of the soft channel estimation has been done for multi-user systems. Specifically, the MSEs of different soft iterative channel estimators were derived and compared. Yet, the interactions among the detector, decoder and estimator were not clearly considered in [12]. Understanding these interactions could be useful in designing effective channel coding in coded modulation systems in the presence of ICSI. The interactions among different blocks in an iterative receiver are best understood by the extrinsic information transfer (EXIT) chart technique.

Recently, EXIT chart analysis in the presence of channel estimation error has gained some interests for BICM-OFDM [13], BICM [14] and BICM-MIMO [15] systems. Reference [13] showed that, for BICM systems with different types of channel estimation algorithm employed by the soft-iterative estimator, the system performance improves with iterations. The authors illustrated the effect of the error using BER curves and transfer charts of the joint detector and estimator. In [15], EXIT curves

for a BICM-MIMO detector were analyzed when only the training-based channel estimator was used. In these two references, the channel estimator and the detector are considered jointly as one block in the iterative receiver.

EXIT curves were used mainly to investigate the property of the detector under channel estimation error. To our knowledge, only the works in [14] and [16] use the EXIT chart to take into account the effect of the channel estimator on the property of the detector. In [14], the effect of channel estimation error on the EXIT chart was analyzed using the so-called correlation parameter with a PSAM-based estimator. It is shown that lower-rate codes can recover the performance loss due to imperfect CSI and they are more robust to channel estimation error. In [16], the authors showed that switching from PSAM to a soft-iterative estimator, referred to as an iterative filtering algorithm, does not always improve the performance of the receiver. Therefore, the authors put conditions on when the output of the decoder should be used by the estimator. They called this algorithm "sparsely interleaved estimation and decoding" (SIED). However, in [16] those conditions are neither based on performance of the channel estimator nor theoretical derivations. Instead, the conditions in [16] test the convergence of the decoder using computer simulation.

In this paper, the switching time is related to the reliability of soft information based on the performance of the channel estimator. The key in establishing this relationship is using the mutual information (MI) that the observation vector has about the channel gains. This switch-augmented conventional iterative channel estimator will be referred to as SAICE. In contrast to the analysis in [16], only one condition is placed on the mutual information at the input of the channel estimator. This condition is theoretically found and indicates the needed reliability of soft information for the channel estimator at the switching time.

The switch-augmented scheme guarantees performance improvement of the iterative receiver with each iteration, however, it might need many iterations to converge for moderate to low SNRs. A less computationally intensive approach is to use both the training and data segments of the observation. This approach produces a channel

estimator for BICM-MIMO systems that will be referred to as a combined iterative channel estimator (CICE), since it combines both the training and data segments. Since this new channel estimator makes use of training sequence and soft information from the output of the decoder in every iteration, its performance also improves with every iteration. This helps the iterative receiver converge at higher code rates for low SNRs. At moderate to high SNRs, the complexity of the receiver is reduced since the receiver converges with a lower number of iterations.

Notation: Upper and lower boldface letters denote matrices and column vectors, respectively. Superscripts $(\cdot)^H$ and $(\cdot)^T$ indicate Hermitian and transpose operations, respectively. For matrix $\mathbf{A}$, $\mathbf{A}_{i,j}$ denotes its (i,j)th entry. The log-likelihood ratio of bit b is defined as $\Lambda^{(b)} = \log\left[\frac{P(b=1)}{P(b=-1)}\right]$. For convenience, the log-likelihood ratio shall be referred to as the Λ-value in this paper. In addition, $E\{\cdot\}$ denotes the expectation of a random variable and $\mathbf{X} \sim \mathcal{CN}(\boldsymbol{\Upsilon}, \boldsymbol{\Theta})$ signifies the random matrix $\mathbf{X}$ has complex normal distribution with mean $\boldsymbol{\Upsilon}$ and covariance matric $\boldsymbol{\Theta}$.

4.2 BICM-MIMO System with Iterative Receiver

4.2.1 Transmitter

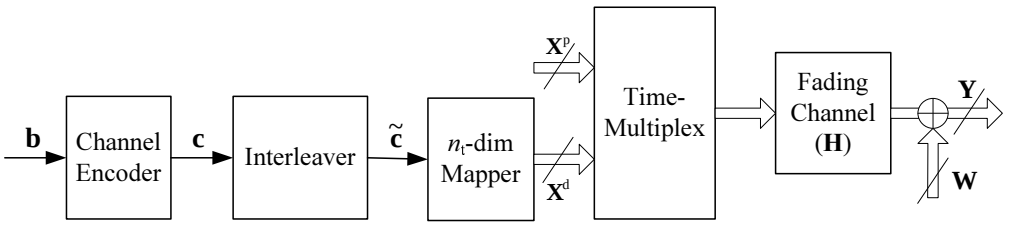

Figure 4.1 Block diagram of a BICM-MIMO transmitter.

The block diagram of a BICM-MIMO system with an iterative receiver is shown in Fig. 4.1. A channel encoder with a rate-R_c error-correcting code converts the vector of information bits $\mathbf{b}$ into a codeword $\mathbf{c}$. Next, the coded bits are interleaved by a pseudo-random interleaver with length L_i to produce the interleaved codeword $\tilde{\mathbf{c}}$. Then the coded bits of $\tilde{\mathbf{c}}$ are mapped to a complex-valued super symbol of a multi-dimensional

signalling set, $\mathbf{\Psi} = \{\underbrace{\mathbf{\Omega} \times \mathbf{\Omega} \times \cdots \times \mathbf{\Omega}}_{n_\mathrm{t}}\}$, where $\mathbf{\Omega}$ is 2^m-QAM constellation, and n_t is the number of transmit antennas.

The next block is the time-multiplexer. It has several inputs. Input $\mathbf{X}^\mathrm{d}$ is the data matrix whose rows contain super data symbols, i.e., $\mathbf{X}^\mathrm{d} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_L^T]^T$ and $\mathbf{x}_k$ is a data super symbol coming from the n_t-dimensional mapper as $\mathbf{x}_k = [x_{(k-1)n_\mathrm{t}+1}, x_{(k-1)n_\mathrm{t}+2}, \ldots, x_{kn_\mathrm{t}}]$. Here, k is the time index and x_i's are complex data symbols belonging to the 2^m-QAM constellation $\mathbf{\Omega}$. Input $\mathbf{X}^\mathrm{p}$ is the training signal. The length of $\mathbf{X}^\mathrm{p}$ is n_p super symbols and the length of $\mathbf{X}^\mathrm{d}$ is L. The time-multiplexer concatenates $\mathbf{X}^\mathrm{d}$ to $\mathbf{X}^\mathrm{p}$ to form blocks of length $L + n_\mathrm{p}$ super symbols. To exploit the time diversity gain associated with the use of an interleaver, L_i is selected such that $L_\mathrm{i}/(n_\mathrm{t}m)$ (i.e., the number of super symbols contained in one codeword) be much greater than $L+n_\mathrm{p}$. The output of the time-multiplexer is transmitted on n_t antennas for spatial diversity.

The channel is assumed to be frequency non-selective with block fading. Let n_r be the number of receive antennas. The $n_\mathrm{t} \times n_\mathrm{r}$ channel matrix $\mathbf{H}$ remains constant for each block duration (i.e., over $L+n_\mathrm{p}$) super symbols, but changes independently from block to block. It has a zero-mean complex Gaussian distribution with covariance matrix of $\mathbf{C_H}$, i.e., $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_H})$. The additive noise is indicated as matrix $\mathbf{W}$ in Fig. 4.1. The output of the channel is shown as matrix $\mathbf{Y}$. It is given by

$$\mathbf{Y} = \mathbf{Y}^\mathrm{p} = \mathbf{X}^\mathrm{p}\mathbf{H} + \mathbf{W}^\mathrm{p} \tag{4.1}$$

during training, and

$$\mathbf{Y} = \mathbf{Y}^\mathrm{d} = \mathbf{X}^\mathrm{d}\mathbf{H} + \mathbf{W}^\mathrm{d} \tag{4.2}$$

during data transmission. In general, the noise matrices $\mathbf{W}^\mathrm{p}$ and $\mathbf{W}^\mathrm{d}$ are zero-mean Gaussian with covariance matrices $\mathbf{C}_{\mathbf{W}^\mathrm{p}}$ and $\mathbf{C}_{\mathbf{W}^\mathrm{d}}$, respectively. The dimensions of $\mathbf{Y}$ are different for training and data transmission phases. During training, the dimensions are $n_\mathrm{p} \times n_\mathrm{r}$, whereas the dimensions are $L \times n_\mathrm{r}$ during data transmission.

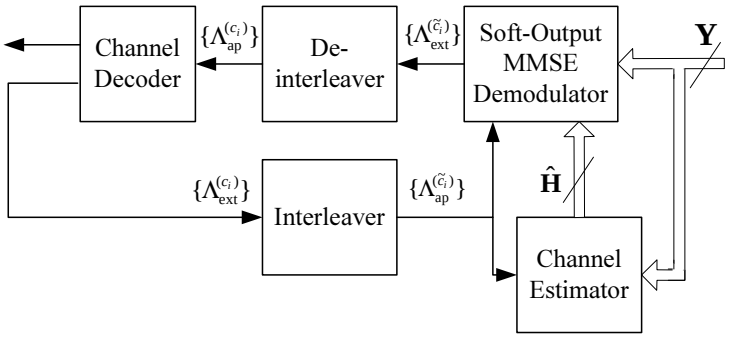

Figure 4.2 Block diagram of the iterative receiver with channel estimation.

4.2.2 Iterative Receiver

The iterative receiver is shown in Fig. 4.2 as a system of 5 blocks. The channel estimator is the block of main interest in this paper and shall be discussed in detail in Section 4.3. It produces a quick estimate of the channel using $\mathbf{Y}^{\text{p}}$ on the first iteration and refines the estimates on subsequent iterations.

After the channel estimation is performed using the training signal, the soft-input soft-output demodulator uses the MMSE criterion to demodulate the data. There is a number of practical demodulation approaches for MIMO systems but the soft-output MMSE demodulator is used here, since it has low complexity and good performance. The soft-output MMSE demodulator can only be used if the coded bits are mapped on each transmit antenna independently [17], which is the case considered in this paper. The soft-output MMSE demodulator computes the extrinsic information for the interleaved bits, $\{\Lambda_{\text{ext}}^{(\tilde{c}_i)}\}_{i=1}^{L_{\text{i}}}$, from the received symbols. To obtain Λ-values, the demodulator exploits the *a priori* information of the coded bits coming from the decoder, $\{\Lambda_{\text{ap}}^{(\tilde{c}_i)}\}$, and the channel estimate $\hat{\mathbf{H}}$. In the first iteration, the demodulator assumes that the *a priori* Λ-values are zero.

The de-interleaved outputs, i.e., $\{\Lambda_{\text{ap}}^{(c_i)}\}$, become the *a priori* Λ-values used in the

If the coded bits are not mapped independently, other demodulators such as a sphere decoding demodulator can be used.

channel decoder shown in Fig. 4.2. The channel decoder uses the log-MAP algorithm to compute the extrinsic Λ-values $\{\Lambda_{\text{ext}}^{(c_i)}\}$ for all coded bits, which are used again in the next iteration in the demodulator.

It should be pointed out that to be consistent with the concept of an EXIT chart, mutual information (MI) at the inputs of the demodulator and the decoder are defined as $I_{\text{ap}}^{(\tilde{c})}$ and $I_{\text{ap}}^{(c)}$, respectively. The MI at the outputs of the demodulator and the decoder are defined as $I_{\text{ext}}^{(\tilde{c})}$ and $I_{\text{ext}}^{(c)}$, respectively.

The structure of Fig. 4.2 is well known. However, the performance of the receiver can vary greatly depending on the operation of the iterative channel estimator. In this paper, an operation is suggested to improve the performance of the receiver at each iteration and to lessen the complexity of the receiver. The operation includes the iterative algorithm and the criterion used by the iterative channel estimator to refine the channel estimates. A theoretical derivation for the channel estimator is given in the next section.

4.3 Channel Estimator

The iterative channel estimator has three inputs: (i) the noisy channel observation $\mathbf{Y}$, which can be $\mathbf{Y}^{\text{d}}$ or $\mathbf{Y}^{\text{p}}$, (ii) the training signal $\mathbf{X}^{\text{p}}$, and (iii) the *a priori* information $\{\Lambda_{\text{ap}}^{(\tilde{c}_i)}\}$ of the interleaved bits. The estimator has one output which is an estimate of the channel coefficient matrix, $\hat{\mathbf{H}}$. A simple block diagram that shows inputs and output is given in Fig. 4.3.

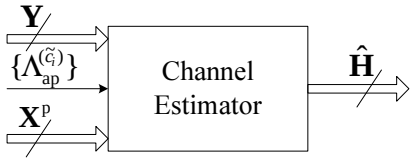

Figure 4.3 Inputs and output of the channel estimator.

Although the log-MAP algorithm can be applied to both block codes and convolutional codes, it is most efficient and practical for decoding convolutional codes based on their well-defined trellis diagrams.

To study the behavior of the iterative channel estimator as an independent block in the iterative receiver, it is necessary to determine how much information the inputs of the channel estimator have about the channel and how to use this information to estimate the channel gains. The key in measuring this information is the MI at the input of the channel estimator.

In a conventional iterative channel estimator, after initialization with the training phase, the channel estimator switches to the data phase. During the training phase, the information that the channel observations (i.e., $\mathbf{Y}^{\mathrm{p}}$) contain about the channel gains at the input of the channel estimator can be expressed by

$$I(\mathbf{H};\mathbf{Y}^{\mathrm{p}}) = h(\mathbf{H}) - h(\mathbf{H}|\mathbf{Y}^{\mathrm{p}}) \tag{4.3}$$

where $h(\Xi) \equiv -\int p(\Xi=\xi)\cdot\log p(\Xi=\xi)\mathrm{d}\xi$ is the entropy of random variable Ξ. To find a closed-form expression for this MI, it is assumed that $\mathbf{H} \sim \mathcal{CN}(\mathbf{0},\mathbf{C}_{\mathbf{H}})$ which implies that $h(\mathbf{H}) = \log \pi e|\mathbf{C}_{\mathbf{H}}|$. The expression of $h(\mathbf{H}|\mathbf{Y}^{\mathrm{p}})$ as stated in [18], is bounded by $\log \pi e|\mathbf{C}_{\mathbf{H}|\mathbf{Y}^{\mathrm{p}}}|$, where $\mathbf{C}_{\mathbf{H}|\mathbf{Y}^{\mathrm{p}}} = \mathbf{C}_{\mathbf{H}}^{-1} + (\mathbf{X}^{\mathrm{p}})^{H}\mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1}\mathbf{X}^{\mathrm{p}}$ is the conditioned covariance of $\mathbf{H}$. Therefore,

$$I(\mathbf{H};\mathbf{Y}^{\mathrm{p}}) = \log|\mathbf{I}_{n_{\mathrm{t}}} + \mathbf{C}_{\mathbf{H}}(\mathbf{X}^{\mathrm{p}})^{H}\mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1}\mathbf{X}^{\mathrm{p}}|. \tag{4.4}$$

During the data phase, at each iteration $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$ and $\mathbf{Y}^{\mathrm{d}}$ are the inputs. The Λ-values are known random variables which are independent of the channel gains. Therefore $I(\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\};\mathbf{H}) = 0$. However, to use the soft information, define the mutual information $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\})$, which indicates how much information $\mathbf{Y}^{\mathrm{d}}$ has about $\mathbf{H}$ knowing Λ-values.

Since Λ-values are known in subsequent iterations, (4.2) can be rewritten as

$$\begin{aligned}\mathbf{Y}^{\mathrm{d}} &= E\{\mathbf{X}^{\mathrm{d}}\}\mathbf{H} + (\mathbf{X}^{\mathrm{d}} - E\{\mathbf{X}^{\mathrm{d}}\})\mathbf{H} + \mathbf{W}^{\mathrm{d}} \\ &= E\{\mathbf{X}^{\mathrm{d}}\}\mathbf{H} + \mathbf{V},\end{aligned} \tag{4.5}$$

where $\mathbf{V}$ is approximately a zero-mean Gaussian matrix with covariance matrix $\mathbf{C}_{\mathbf{V}}$, which represents the total effective noise matrix in subsequent iterations. The entries

of $E\{\mathbf{X}^{\mathrm{d}}\}$ are calculated using $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$ at each iteration by $E\{\mathbf{X}_{i,j}^{\mathrm{d}}\} = \sum_{x\in\Omega} x \cdot p(\mathbf{X}_{i,j}^{\mathrm{d}} = x)$. The detailed derivations of the probability $p(\mathbf{X}_{i,j}^{\mathrm{d}} = x)$ from Λ-values are given in [19] (note that the calculation depends on the mapping rule in $\boldsymbol{\Omega}$).

Moreover, the covariance matrix $\mathbf{C_V}$ is given by,

$$\mathbf{C_V} = \mathbf{C}_{\mathbf{W}^{\mathrm{d}}} + \left(E\{\mathbf{X}^{\mathrm{d}}\mathbf{C_H}(\mathbf{X}^{\mathrm{d}})^H\} - E\{\mathbf{X}^{\mathrm{d}}\}\mathbf{C_H}(E\{\mathbf{X}^{\mathrm{d}}\})^H\right)$$

It follows that the entries of $\mathbf{C_V}$ are

$$(\mathbf{C_V})_{i,j} = \begin{cases} (\mathbf{C}_{\mathbf{W}^{\mathrm{d}}})_{i,i} + \sum_{k=1}^{n_{\mathrm{t}}} \sigma^2_{\mathbf{X}_{i,k}^{\mathrm{d}}} (\mathbf{C_H})_{k,k}\,, & i = j \\ (\mathbf{C}_{\mathbf{W}^{\mathrm{d}}})_{i,j}\,, & i \neq j. \end{cases} \tag{4.6}$$

where $\sigma^2_{\mathbf{X}_{i,k}^{\mathrm{d}}} = \sum_{x\in\Omega} |x|^2 \cdot p(\mathbf{X}_{i,k}^{\mathrm{d}} = x) - |E\{\mathbf{X}_{i,k}^{\mathrm{d}}\}|^2$ is the variance of $\mathbf{X}_{i,k}^{\mathrm{d}}$.

Finally, the MI at the input of the channel estimator during the data phase can be obtained by

$$\begin{aligned} I(\mathbf{H}; \mathbf{Y}^{\mathrm{d}}, \{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}) &= \log |\mathbf{C_H}{\mathbf{C_{H|Y}}}^{-1}| \\ &= \log |\mathbf{I}_{n_{\mathrm{t}}} + \mathbf{C_H}(E\{\mathbf{X}^{\mathrm{d}}\})^H \mathbf{C_V}^{-1} E\{\mathbf{X}^{\mathrm{d}}\}| \end{aligned}$$

Now, the question is how to use this information to estimate the channel. Different channel estimators can be used to exploit this information. To quantify the amount of information exploited by the channel estimator, a new parameter can be defined as $I(\hat{\mathbf{H}}; \mathbf{H})$, which measures the information at the output of the channel estimator. Among different channel estimators, it can be shown that the MMSE channel estimator fully exploits the mutual information at the input of the channel estimator. Specifically, with the MMSE channel estimator $I(\mathbf{H}; \hat{\mathbf{H}}^{\mathrm{p}}) = I(\mathbf{H}; \mathbf{Y}^{\mathrm{p}})$ and $I(\mathbf{H}; \hat{\mathbf{H}}^{\mathrm{d}}) = I(\mathbf{H}; \mathbf{Y}^{\mathrm{d}}, \{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\})$, where $\hat{\mathbf{H}}^{\mathrm{p}}$ and $\hat{\mathbf{H}}^{\mathrm{d}}$ are the channel estimations at the training and data phases, respectively. For a linear channel estimator $\hat{\mathbf{H}}^{\mathrm{p}} = \mathbf{A}^{\mathrm{p}}\mathbf{Y}^{\mathrm{p}}$, where $\mathbf{A}^{\mathrm{p}}$ is the matrix of the linear channel estimator in the training phase and $\hat{\mathbf{H}}^{\mathrm{d}} = \mathbf{A}^{\mathrm{d}}\mathbf{Y}^{\mathrm{d}}$, where $\mathbf{A}^{\mathrm{d}}$ is the matrix of the linear channel estimator in the data phase. Expressions for the linear MMSE channel estimators at the training and data

phases are given by

$$\begin{aligned}\mathbf{A}^{\mathrm{p}} &= E\{\mathbf{H}(\mathbf{Y}^{\mathrm{p}})^H\}(E\{\mathbf{Y}^{\mathrm{p}}(\mathbf{Y}^{\mathrm{p}})^H\})^{-1} \\ &= \left(\mathbf{C}_{\mathbf{H}}^{-1} + (\mathbf{X}^{\mathrm{p}})^H\mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1}\mathbf{X}^{\mathrm{p}}\right)^{-1}(\mathbf{X}^{\mathrm{p}})^H\mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1} \qquad (4.7)\end{aligned}$$

and

$$\mathbf{A}^{\mathrm{d}} = \left(\mathbf{C}_{\mathbf{H}}^{-1} + (E\{\mathbf{X}^{\mathrm{d}}\})^H\mathbf{C}_{\mathbf{V}}^{-1}E\{\mathbf{X}^{\mathrm{d}}\}\right)^{-1}(E\{\mathbf{X}^{\mathrm{d}}\})^H\mathbf{C}_{\mathbf{V}}^{-1}, \qquad (4.8)$$

respectively.

As mentioned, the conventional iterative channel estimator is not guaranteed to converge. For the estimator to converge, the information at the input of the channel estimator during the data phase should be more than the information during the training. This means that $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}) > I(\mathbf{H};\mathbf{Y}^{\mathrm{p}})$ must hold for each iteration. Since the soft information from the decoder is not as reliable as training signal for the first few iterations, this condition is not guaranteed. Thus, in general, although increasing the number of iterations increases the reliability of the soft output from the decoder and increases $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\})$ iteratively, this does not guarantee that $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}) > I(\mathbf{H};\mathbf{Y}^{\mathrm{p}})$ at each iteration.

To solve the problem, a constraint is placed on the input of the estimator. The mutual information $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\})$ can be tested at each iteration and the channel estimator is enabled in the data phase on iteration where $I(\mathbf{H};\mathbf{Y}^{\mathrm{d}},\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}) > I(\mathbf{H};\mathbf{Y}^{\mathrm{p}})$. This translates to placing a constraint on the soft information at the output of the decoder, which depends on channel length and system's parameters. The number of iterations with which the condition is met depends on coding design. This new scheme is referred to as a switch-augmented conventional iterative channel estimator (SAICE) .

Another way to guarantee increasing mutual information at the input of the channel estimator is to use information from the previous iterations. Exploiting the inherent latency of bit-interleaver and the property of a block-fading channel, the observation vector can be constructed from both the training and data segments of the

observations in this approach. The mutual information at the input of the channel estimator then can be expressed as $I(\mathbf{H}; [\mathbf{Y}^{\mathrm{p}}, \mathbf{Y}^{\mathrm{d}}], \{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\})$ and it is given by,

$$I(\mathbf{H}; [\mathbf{Y}^{\mathrm{p}}, \mathbf{Y}^{\mathrm{d}}], \{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}) = \log |\mathbf{I}_{n_{\mathrm{t}}} + \mathbf{C}_{\mathbf{H}}(\mathbf{X}^{\mathrm{p}})^H \mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1} \mathbf{X}^{\mathrm{p}} + \mathbf{C}_{\mathbf{H}}(E\{\mathbf{X}^{\mathrm{d}}\})^H \mathbf{C}_{\mathbf{V}}^{-1} E\{\mathbf{X}^{\mathrm{d}}\}| \quad (4.9)$$

A new MMSE channel estimator can be devised such that it uses both training and data segments of the observation matrices at each iteration. This estimator is given by $\hat{\mathbf{H}}^{\mathrm{c}} = [\mathbf{A}^{\mathrm{p_c}} \mathbf{A}^{\mathrm{d_c}}][\mathbf{Y}^{\mathrm{p}} \mathbf{Y}^{\mathrm{d}}]^T$ and shall be referred to as a combined iterative channel estimator (CICE). $\mathbf{A}^{\mathrm{p_c}}$ and $\mathbf{A}^{\mathrm{d_c}}$ indicate the training and data phases of the combined estimator, respectively. They can be obtained by,

$$[\mathbf{A}^{\mathrm{p_c}} \mathbf{A}^{\mathrm{d_c}}] = E\{\mathbf{H}[\mathbf{Y}^{\mathrm{p}} \mathbf{Y}^{\mathrm{d}}]^H\} (E\{[\mathbf{Y}^{\mathrm{p}} \mathbf{Y}^{\mathrm{d}}][\mathbf{Y}^{\mathrm{p}} \mathbf{Y}^{\mathrm{d}}]^H\})^{-1}.$$

It follows that

$$\mathbf{A}^{\mathrm{p_c}} = \left(\mathbf{C}_{\mathbf{H}}^{-1} + (E\{\mathbf{X}^{\mathrm{d}}\})^H \mathbf{C}_{\mathbf{V}}^{-1} E\{\mathbf{X}^{\mathrm{d}}\} + (\mathbf{X}^{\mathrm{p}})^H \mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1} \mathbf{X}^{\mathrm{p}}\right)^{-1} (\mathbf{X}^{\mathrm{p}})^H \mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1}, \quad (4.10)$$

and

$$\mathbf{A}^{\mathrm{d_c}} = \left(\mathbf{C}_{\mathbf{H}}^{-1} + (E\{\mathbf{X}^{\mathrm{d}}\})^H \mathbf{C}_{\mathbf{V}}^{-1} E\{\mathbf{X}^{\mathrm{d}}\} + (\mathbf{X}^{\mathrm{p}})^H \mathbf{C}_{\mathbf{W}^{\mathrm{p}}}^{-1} \mathbf{X}^{\mathrm{p}}\right)^{-1} (E\{\mathbf{X}^{\mathrm{d}}\})^H \mathbf{C}_{\mathbf{V}}^{-1} \quad (4.11)$$

In contrast to the mean-square error or correlation parameter analysis, the above MI analysis can be used to determine when the information at the input of the channel estimator is beneficial. In addition, it can be applied to any channel estimator, since it is measured at the input of the channel estimator. Besides, for a correlated channel environment it can be used to design the training sequence in order to maximize the information at the input of the channel estimator.

4.4 Good Approximations of the Proposed Channel Estimators

The statistical property of the channel and the system's parameters affect the performance of the channel estimator. To reduce the complexity of the analysis that determines these effects, the following simplifying assumptions are made.

- The channel is assumed to be Rayleigh fading. This means the channel coefficients are i.i.d. complex Gaussian, i.e., $\mathbf{C_H} = E\{\mathbf{H}\mathbf{H}^H\} = n_\mathrm{r}\sigma_h^2\mathbf{I}_{n_\mathrm{t}}$, where σ_h^2 is the variance of each entry.

- The noise corrupting the channel is i.i.d. AWGN. This means the entries of noise matrices in (4.1) and (4.2) are i.i.d., i.e., $\mathbf{C}_{\mathbf{W}^\mathrm{p}} = E\{\mathbf{W}^\mathrm{p}(\mathbf{W}^\mathrm{p})^H\} = n_\mathrm{r}N_0\mathbf{I}_{n_\mathrm{p}}$ and $\mathbf{C}_{\mathbf{W}^\mathrm{d}} = E\{\mathbf{W}^\mathrm{d}(\mathbf{W}^\mathrm{d})^H\} = n_\mathrm{r}N_0\mathbf{I}_L$, where N_0 is the one-sided spectral density of each noise source in the channel.

- The training signal is chosen such that $(\mathbf{X}^\mathrm{p})^H\mathbf{X}^\mathrm{p} = n_\mathrm{p}E_\mathrm{p}I_{n_\mathrm{t}}$, where $E_\mathrm{p} = \frac{1}{n_\mathrm{t}n_\mathrm{p}}\mathrm{trace}\{(\mathbf{X}^\mathrm{p})^H\mathbf{X}^\mathrm{p}\}$. For i.i.d. Rayleigh channel coefficients, this training maximizes $I(\mathbf{H}, \hat{\mathbf{H}})$.

In addition, two signal to noise ratios are introduced to simplify the notation. They are $\mathrm{SNR_d} = \frac{\sigma_h^2 E_\mathrm{s}}{N_0}$ and $\mathrm{SNR_p} = \frac{\sigma_h^2 E_\mathrm{p}}{N_0}$, where E_s is the energy per data symbol. Moreover, $I_\mathbf{H}$ is used to indicate the MI at the input of the channel estimator in general and $I_\mathbf{H}^\mathrm{p}$, $I_\mathbf{H}^\mathrm{d}$ and $I_\mathbf{H}^\mathrm{c}$ are simplified notations used for $I(\mathbf{H};\mathbf{Y}^\mathrm{p})$, $I(\mathbf{H};\mathbf{Y}^\mathrm{d},\{\Lambda_\mathrm{ap}^{(\tilde{c}_i)}\})$ and $I(\mathbf{H};[\mathbf{Y}^\mathrm{p},\mathbf{Y}^\mathrm{d}],\{\Lambda_\mathrm{ap}^{(\tilde{c}_i)}\})$, respectively.

The channel estimator in (4.10) has about the same complexity as the conventional estimators of (4.7) and (4.8). However the complexity of the estimators given by (4.8) and (4.10) can be significantly reduced with two simplifying approximations. To simplify the term $(E\{\mathbf{X}^\mathrm{d}\})^H\mathbf{C}_\mathbf{V}^{-1}E\{\mathbf{X}^\mathrm{d}\}$ in (4.10) and (4.8), the entries of diagonal matrix $\mathbf{C_V}$, which are

$$(\mathbf{C_V})_{i,i} \;=\; n_\mathrm{r}N_0 + n_\mathrm{r}\sigma_h^2\sum_{j=1}^{n_\mathrm{t}}\sigma^2_{\mathbf{X}^\mathrm{d}_{i,j}}, \tag{4.12}$$

are approximated by

$$(\mathbf{C_V})_{i,i} \;\approx\; n_\mathrm{r}N_0 + n_\mathrm{r}\sigma_h^2\tilde{n}_\mathrm{t}E_s, \tag{4.13}$$

The Rayleigh fading and AWGN assumptions assures $\mathbf{C_V}$ is diagonal.

where $\tilde{n}_{\rm t}$ represents the effect of residual interference and is defined by $\tilde{n}_{\rm t} \equiv \frac{1}{LE_s}\sum_{i=1}^{L}\sum_{j=1}^{n_{\rm t}}\sigma^2_{\mathbf{X}^{\rm d}_{i,j}}$. From the approximation in (4.13) it follows that

$$(E\{\mathbf{X}^{\rm d}\})^H\mathbf{C}_{\mathbf{V}}^{-1}E\{\mathbf{X}^{\rm d}\} \approx \frac{(E\{\mathbf{X}^{\rm d}\})^H E\{\mathbf{X}^{\rm d}\}}{n_{\rm r}N_0 + n_{\rm r}\sigma_h^2\tilde{n}_{\rm t}E_s}. \tag{4.14}$$

Yet one more approximation is made. Assuming a long interleaver and i.i.d. Λ-values of the data symbols, the off-diagonal elements of matrix $(E\{\mathbf{X}^{\rm d}\})^H E\{\mathbf{X}^{\rm d}\}$ are negligible. Therefore this matrix can be approximated by

$$(E\{\mathbf{X}^{\rm d}\})^H E\{\mathbf{X}^{\rm d}\} \;\approx\; \tilde{L}E_s\mathbf{I}_{n_{\rm t}}, \tag{4.15}$$

where $\tilde{L} \equiv \frac{1}{n_{\rm t}E_s}\sum_{i=1}^{L}\sum_{j=1}^{n_{\rm t}}|E\{\mathbf{X}^{\rm d}_{i,j}\}|^2$.

The simplifying approximations lead to compact expressions for the estimators of (4.8) and (4.10). The estimator of (4.8) simplifies to

$$\mathbf{A}^{\rm d} \approx \frac{\sigma_h^2/N_0}{1+(\tilde{n}_{\rm t}+\tilde{L})\mathrm{SNR}_{\rm d}}(E\{\mathbf{X}^{\rm d}\})^H, \tag{4.16}$$

while the estimator of (4.10) simplifies to

$$\begin{aligned}\mathbf{A}^{\rm p_c} &\approx \frac{[(1+\tilde{n}_{\rm t}\mathrm{SNR}_{\rm d})\sigma_h^2/N_0]\,(\mathbf{X}^{\rm p})^H}{1+n_{\rm p}\mathrm{SNR}_{\rm p}+((1+n_{\rm p}\mathrm{SNR}_{\rm p})\tilde{n}_{\rm t}+\tilde{L})\mathrm{SNR}_{\rm d}}\\ \mathbf{A}^{\rm d_c} &\approx \frac{[\sigma_h^2/N_0]\,(E\{\mathbf{X}^{\rm d}\})^H}{1+n_{\rm p}\mathrm{SNR}_{\rm p}+((1+n_{\rm p}\mathrm{SNR}_{\rm p})\tilde{n}_{\rm t}+\tilde{L})\mathrm{SNR}_{\rm d}}.\end{aligned} \tag{4.17}$$

It is useful to examine the performance of the channel estimators at the following extreme (corner) cases:

- The performance of the iterative channel estimator for the first few iterations with low SNR.
- The performance of the iterative channel estimator after a large number of iterations (sufficient for convergence) with high SNR.

For the first few iterations and at low SNR the soft information at the output of the channel decoder is not reliable, i.e., $E\{\mathbf{X}^{\rm d}\} \simeq \mathbf{0}$. This leads to $\mathbf{A}^{\rm p_c} \simeq \mathbf{A}^{\rm p}$ and

$\mathbf{A}^{\mathrm{d_c}} \simeq \mathbf{A}^{\mathrm{d}} \simeq \mathbf{0}$. Therefore, in this case it is reasonable to use only the training signal to reduce complexity. In using only the training signal, the MI at the input of both channel estimators is $\log\left(1 + n_{\mathrm{p}}\mathrm{SNR}_{\mathrm{p}}\right)$.

For high SNR and after a large number of iterations, the so-called genie condition can be approached and $E\{\mathbf{X}^{\mathrm{d}}\}$ can be approximated by $\mathbf{X}^{\mathrm{d}}$. In this case, the approximations made earlier in this section are more accurate. Moreover, $\tilde{n}_{\mathrm{t}} \to 0$ and $\tilde{L} \to L$ which implies $\mathbf{A}^{\mathrm{d}} \approx \frac{\sigma_h^2/N_0}{1+L\mathrm{SNR}_{\mathrm{d}}}(E\{\mathbf{X}^{\mathrm{d}}\})^H$, $\mathbf{A}^{\mathrm{p_c}} \approx \frac{\sigma_h^2/N_0}{1+n_{\mathrm{p}}\mathrm{SNR}_{\mathrm{p}}+L\mathrm{SNR}_{\mathrm{d}}}(\mathbf{X}^{\mathrm{p}})^H$ and $\mathbf{A}^{\mathrm{d_c}} \approx \frac{\sigma_h^2/N_0}{1+n_{\mathrm{p}}\mathrm{SNR}_{p}+L\mathrm{SNR}_{\mathrm{d}}}(E\{\mathbf{X}^{\mathrm{d}}\})^H$. Treating these approximations as equalities leads to a tight upper bound for $I_{\mathbf{H}}^{\mathrm{c}}$ which is $\log\left(1 + n_{\mathrm{p}}\mathrm{SNR}_{\mathrm{p}} + L\mathrm{SNR}_{\mathrm{d}}\right)$. Furthermore, for the SAICE and for the conventional iterative channel estimator, the upper bound simplifies to $\log\left(1 + L\mathrm{SNR}_{\mathrm{d}}\right)$.

Under the approximations made in the two corner cases, $I_{\mathbf{H}}$ for both the conventional and proposed channel estimators does not depend on the modulation order nor the number of antennas. To be more precise, the upper bound of the CICE depends on the length of the channel, the power of the data symbols, and the power and time duration of the training signal. For the channel estimator with the switch-augmented scheme, the upper bound only depends on the length of the channel and power of the data symbols.

In other more typical cases, the conclusions for the corner cases do not apply. The reason for this is that $\tilde{n}_{\mathrm{t}}$ does not contribute to $I_{\mathbf{H}}$ in either corner case but does most of the time. Except for the corner cases, the number of transmit antennas affects the performance of the iterative channel estimators, since $\tilde{n}_{\mathrm{t}}$ depends on the number of transmit antennas. This is illustrated with simulation results in section 4.5.

The insight gain from the analysis is now used to develop a new iterative algorithm for both the conventional and proposed channel estimators. The conventional soft-iterative channel estimator uses (4.7) for initialization and switches to (4.8) to refine the estimated channel gains. However, the analysis shows that for the first few iterations $I_{\mathbf{H}}^{\mathrm{d}}$ is less than $I_{\mathbf{H}}^{\mathrm{p}}$, which means that the soft-iterative channel estimator

(which is responsible for $I_{\mathbf{H}}^{\mathrm{d}}$) should be turned off for the first few iterations. To optimize performance, the soft-iterative channel estimator should be turned on when $I_{\mathbf{H}}^{\mathrm{d}}$ exceeds the value of $I_{\mathbf{H}}^{\mathrm{p}}$.

As with the conventional estimator, the performance of the switch-augmented estimator is a sole function of $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$. For this to be true the switching threshold must be a sole function of $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$. This threshold is established by $I_{\mathbf{H}}^{\mathrm{d}} > I_{\mathbf{H}}^{\mathrm{p}}$ which is equivalent to

$$\tilde{L} \;>\; n_{\mathrm{p}}\mathrm{SNR}_{\mathrm{p}}\left(\frac{1}{\mathrm{SNR}_{\mathrm{d}}} + \tilde{n}_{\mathrm{t}}\right), \tag{4.18}$$

when the approximations in this section hold. Since $\tilde{L}$ and $\tilde{n}_{\mathrm{t}}$ are sole functions of $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$, the threshold is a sole function of $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$.

It should be pointed out that even with the switching scheme, in some cases at low SNR, the receiver fails to converge, while the switching scheme mitigates the convergence problem it does not eliminate it. The proposed estimator, which has no switch, does not have a convergence problem.

The guarantee of convergence is the main advantage of the proposed estimator over the SAICE. Convergence is guaranteed since the performance of the proposed channel estimator improves after each iteration. In addition, the proposed channel estimator reduces the convergence time of the iterative receiver. This translates to a lower-complexity receiver since the decoder and the demodulator need less iterations to converge.

4.5 Evaluation Results

The performance of a BICM-MIMO system is evaluated using the EXIT charts and BER curves for the switch-augmented and the proposed channel estimators. For a fair comparison, $L = 30$ (the length that the channel stays the same) is fixed for all the simulations, while the channel coefficients have a Rayleigh distribution with a variance of 1, i.e., $\sigma_h^2 = 1$. A rate-1/2 convolutional code and a bit-interleaver with length $L_{\mathrm{i}} = 4,800$ are used. In addition, E_p is set to E_s, n_{p} is set to n_{t} and thus

$\mathrm{SNR} = \mathrm{SNR_d} = \mathrm{SNR_p}$ for all the figures.

4.5.1 Mutual Information at the Input of the Channel Estimator

A Monte-Carlo simulation is used to calculate the MI at the input of the channel estimator. However, closed-form expressions and approximations found in Section 4.4 are tested and found to be valid for the all simulations.

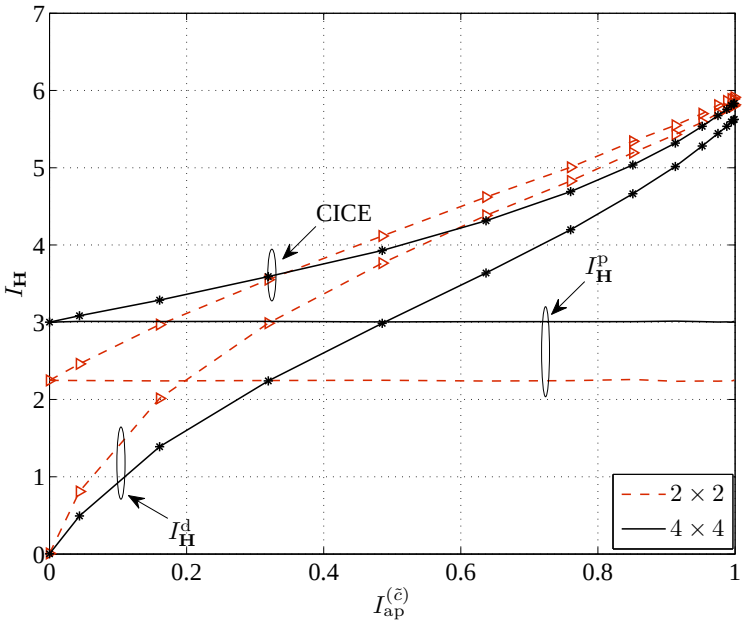

Figure 4.4 The MI at the input of the conventional channel estimator with training and data phases and the CICE, with QPSK modulation, $n_{\mathrm{t}} = n_{\mathrm{r}} = n_{\mathrm{p}} = 2$ and $n_{\mathrm{t}} = n_{\mathrm{r}} = n_{\mathrm{p}} = 4$, where $\mathrm{SNR_d} = 3\mathrm{dB}$.

Fig. 4.4 plots $I_{\mathbf{H}}$ for a range of $I_{\mathrm{ap}}^{(\tilde{c})}$, where $\mathrm{SNR_d} = 3\mathrm{dB}$ for 2×2 and 4×4 MIMO configurations. The input to the channel estimator, $\{\Lambda_{\mathrm{ap}}^{(\tilde{c}_i)}\}$ is generated for a given $I_{\mathrm{ap}}^{(\tilde{c})}$. The performance of the conventional channel estimator is shown with two curves; one with $I_{\mathbf{H}}^{\mathrm{p}}$ at initialization and with the $I_{\mathbf{H}}^{\mathrm{d}}$ for the iterations in the data phase. The performance of the CICE is also plotted and compared with the conventional estimator for QPSK with sigma mapping scheme proposed in [20]. It is observed that $I_{\mathbf{H}}^{\mathrm{c}}$ starts from $I_{\mathbf{H}}^{\mathrm{p}}$ and converges to 5.93 and 5.98 which are the upper bound for the two MIMO configurations, respectively. At the beginning $I_{\mathbf{H}}^{\mathrm{c}}$ is

larger than $I_{\mathbf{H}}^{\mathrm{d}}$, which is essential for the convergence of the iterative receiver in BICM systems. It should be pointed out that since $n_{\mathrm{p}} = n_{\mathrm{t}}$ in Fig. 4.4, $I_{\mathbf{H}}^{\mathrm{p}}$ is larger for $n_{\mathrm{t}} = 4$ than for $n_{\mathrm{t}} = 2$. Moreover, at the so-called genie condition, i.e., $I_{\mathrm{ap}}^{(\tilde{c})} = 1$, both MIMO configurations approach the upper bounds for both conventional estimator and the proposed CICE.

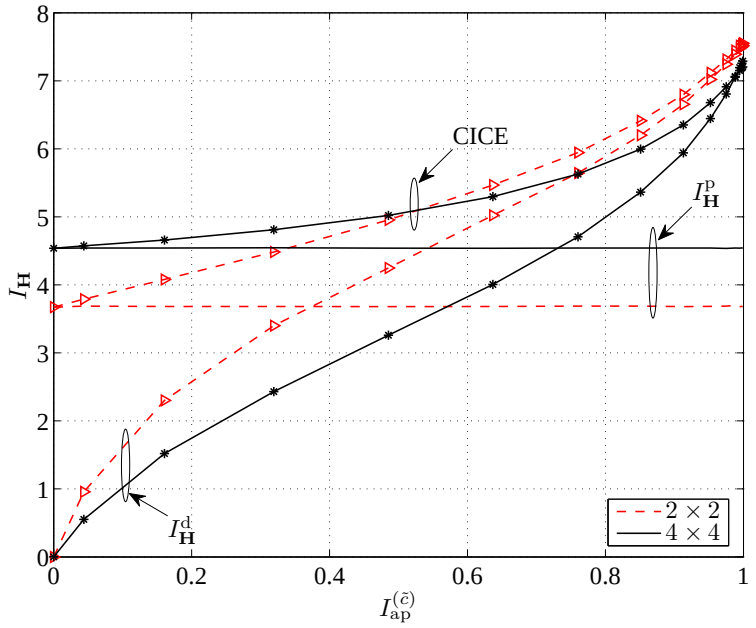

Figure 4.5 The MI at the input of the conventional channel estimator with training and data phases and the CICE, with 16QAM modulation, $n_{\mathrm{t}} = n_{\mathrm{r}} = n_{\mathrm{p}} = 2$ and $n_{\mathrm{t}} = n_{\mathrm{r}} = n_{\mathrm{p}} = 4$, where $\mathrm{SNR}_{\mathrm{d}} = 8\mathrm{dB}$.

The curves for the MI at the input of the channel estimators are also plotted in Fig. 4.5 for 16QAM, where $\mathrm{SNR}_{\mathrm{d}} = 8\mathrm{dB}$. From Fig. 4.5 it is observed that for 16QAM, $I_{\mathbf{H}}^{\mathrm{d}}$ for switch-augmented scheme crosses $I_{\mathbf{H}}^{\mathrm{p}}$ at a larger value of $I_{\mathrm{ap}}^{(\tilde{c})}$ than for QPSK. The reason for this is the $\mathrm{SNR}_{\mathrm{p}}$ in Fig. 4.5 is larger than that used in Fig. 4.4.

To examine the sensitivity of the CICE estimator with respect to $\mathrm{SNR}_{\mathrm{d}}$, $I_{\mathbf{H}}^{\mathrm{c}}$ versus $\mathrm{SNR}_{\mathrm{d}}$ is plotted in Fig. 4.6 for QPSK and 16QAM. Three values 0, 0.5 and 1 for $I_{\mathrm{ap}}^{(\tilde{c})}$ are considered in simulations. From Fig. 4.6 it is clear the MI at the input of the channel estimator is a linear function of SNR and does not change significantly as the SNR increases from 2 to 10. It should be pointed out that $I_{\mathbf{H}}^{\mathrm{c}}$ for $I_{\mathrm{ap}}^{(\tilde{c})} = 0$ corresponds

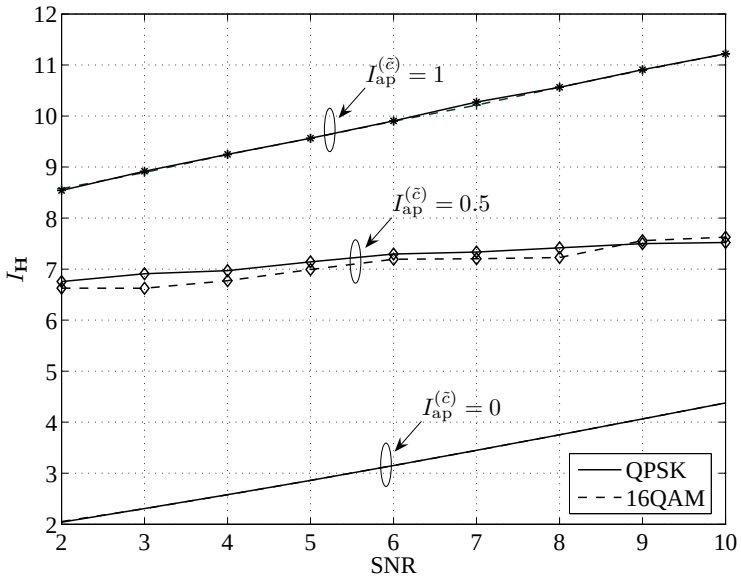

Figure 4.6 The MI of the CICE versus SNR for QPSK and 16QAM for a range of $I_{\mathrm{ap}}^{(\tilde{c})}$, where $n_{\mathrm{t}} = n_{\mathrm{r}} = n_{\mathrm{p}} = 2$.

to the training phase and SNR corresponds to $\mathrm{SNR_p}$ for that case. From the result, it is clear that the MI is basically the same for a given SNR for both QPSK and 16QAM.

From Figs. 4.4, 4.5 and 4.6, it appears that the operation of the channel estimator, if the channel estimator is considered independently, does not depend on the order of modulation. Moreover, the system's parameters that most affect the operation of the channel estimators are SNR and n_{t} in addition to channel length.

4.5.2 EXIT Chart with ICSI

In this section the EXIT chart is examined to understand the convergence behavior of an iterative system with imperfect CSI. In the previous section, channel estimator was treated as an independent block. Here, the interactions among the decoder, detector and channel estimator are determined through the EXIT chart. To do this the trajectory curves, which are found iteratively, are obtained.

In Fig. 4.7, the channel code is a rate-1/2 convolutional code with constraint length 5 and generator polynomial $\mathbf{G} = [33; 31]$. The parameter $\mathrm{SNR_d}$ is set to 3dB

and parameters n_{t}, n_{r} and n_{p} are all set to 2. The modulation is QPSK, performed independently on each antenna.

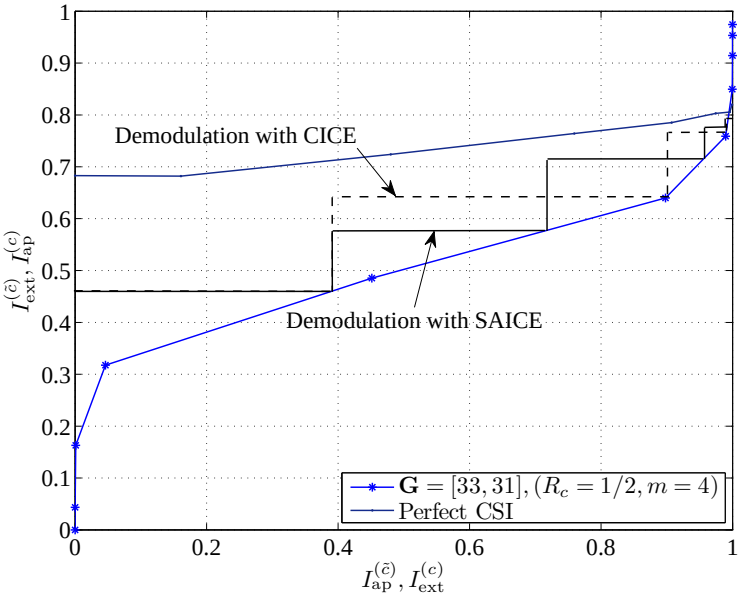

Figure 4.7 EXIT chart of the BICM-MIMO receiver with two methods of channel estimation for QPSK, where SNR = 3dB.

The interaction among the three blocks begins with the initialization of the channel estimator. Therefore, for the first iteration, $I_{\mathbf{H}}$ is 2.2487 with $I_{\mathrm{ap}}^{(\tilde{c})} = 0$. Using $I_{\mathbf{H}} = 2.2487$ and $I_{\mathrm{ap}}^{(\tilde{c})} = 0$ then it follows that $I_{\mathrm{ext}}^{(\tilde{c})} = 0.32$. From Fig. 4.7, $I_{\mathrm{ap}}^{(c)} = I_{\mathrm{ext}}^{(\tilde{c})} = 0.3210$ which leads to $I_{\mathrm{ext}}^{(c)} = 0.75$. This value is the input for the two channel estimators and the demodulator for the second iteration (i.e., $(I_{\mathbf{H}}, I_{\mathrm{ap}}^{(\tilde{c})}) \rightarrow I_{\mathrm{ext}}^{(\tilde{c})}(I_{\mathrm{ap}}^{(c)}) \rightarrow I_{\mathrm{ext}}^{(c)}(I_{\mathrm{ap}}^{(\tilde{c})}) \rightarrow (I_{\mathbf{H}}, I_{\mathrm{ap}}^{(\tilde{c})}) \rightarrow \cdots$). This information exchange process continues until the trajectory curve reaches a point on the coding curve (called optimum point). The outputs of switch-augmented estimator and CICE are used by the demodulator to produce the curves in Fig. 4.7 (shown as solid and dash lines, respectively). The interactions, which illustrated by these curves, can be seen clearer by examining the MI values at each iteration shown in Table 4.1. In the second and the last two columns of Table 4.1, two values are inserted in each column. The first value corresponds to the system with the SAICE and the second one corresponds to the system with the CICE. It should be pointed out that for the first iteration only one value is represented which

Table 4.1 Interaction of the MI for three iterations.

Number of iterations	$I_{\text{ap}}^{(\tilde{c})}$	$I_{\mathbf{H}}^{\text{p/d}}$	I_{H}^{c}	$I_{\text{ext}}^{(\tilde{c})} = I_{\text{ap}}^{(c)}$	$I_{\text{ext}}^{(c)}$
1	0	2.2412	2.2412	0.46	0.39
2	0.39	3.4235	4.0070	0.58/0.64	0.72/0.91
3	0.72/0.91	4.7389	5.6351	0.72/0.78	0.96/0.98
4	0.96/0.98	5.5952	5.8588	0.78/0.80	0.99/1

is obtained from the system using only the training sequence.

For comparison, the transfer function of the demodulator under perfect CSI assumption is also plotted in Fig. 4.7. It is observed that the demodulators with both estimators of SAICE and CICE converge after 4 iterations. However the demodulator with the SAICE has more residual error than the CICE. In fact the performance of the demodulator with the CICE is the same as the demodulator with perfect CSI after 4 iterations.

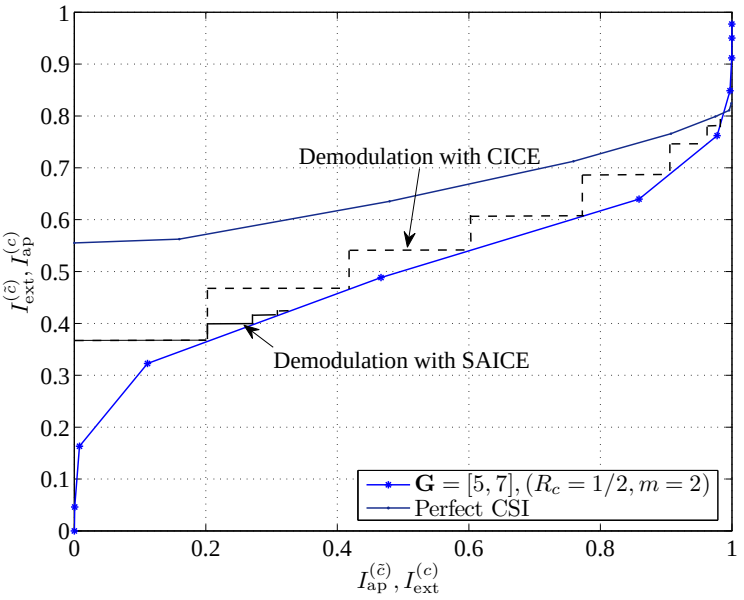

Figure 4.8 EXIT chart of the BICM-MIMO receiver with two methods of channel estimation for 16QAM, where SNR = 8dB.

The result of the same simulation performed for 16QAM is shown in Fig. 4.8. Since the soft information of the 16QAM demodulator is not as reliable as that of

the QPSK demodulator at the first few iterations, the channel code with constraint length 3 and generator polynomial $[7; 5]$ is used for 16QAM and SNR_d is set to 8dB. The other simulation parameters are the same as that for Fig. 4.7. From Fig. 4.8, the demodulator with the SAICE fails to converge after 4 iterations. However the demodulator with the CICE converges after 7 iterations.

4.5.3 BER Performance

The BER performance of the proposed algorithm for the receivers in Fig. 4.7 and Fig. 4.8 are compared for different channel estimators in Fig. 4.9 and Fig. 4.10, respectively.

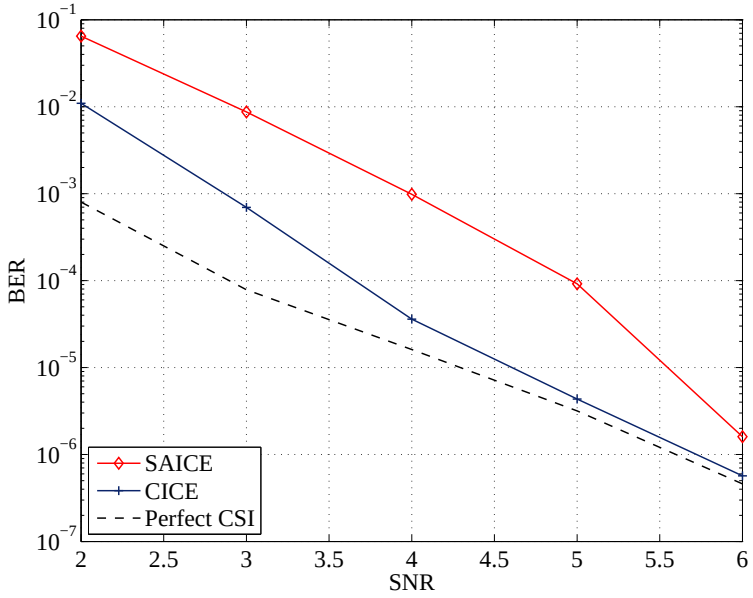

Figure 4.9 BER of the BICM-MIMO receiver with QPSK modulation after 5 iterations for SAICE and CICE.

Fig. 4.9 plots BER versus SNR, where for each value of SNR considered, 10^5 independent frames were run to obtain a BER value. Compared in the figure are performances of two receivers, one with the CICE and the other with the SAICE. The BER curve of the receiver with perfect CSI is also plotted to serve as a benchmark. The curves in Fig. 4.9 show that the performance of the CICE is about 1.5 dB better than the performance with the SAICE after 5 iterations at BER $= 10^{-5}$. Since the

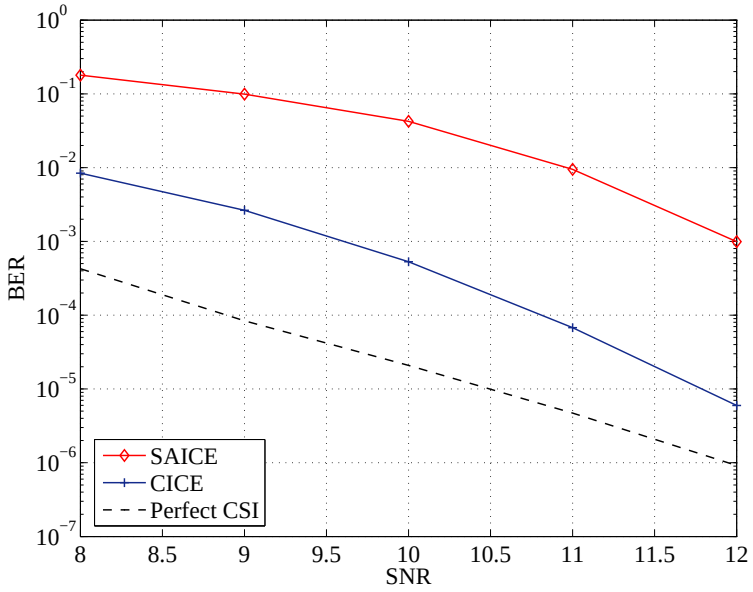

Figure 4.10 BER of the BICM-MIMO receiver with 16QAM modulation after 5 iterations for SAICE and CICE.

CICE uses the information of both pilot and data to estimate the channel coefficients, it has the better performance at the beginning of the iterative process. Thus it facilitates convergence and therefore, iteratively improves the overall performance of the receiver. It is also noted that the BER performance with the CICE approaches that of the receiver that has perfect CSI.

The performance curves for 16QAM in Fig. 4.10 indicate an improvement of about 2dB at BER $= 10^{-3}$. The reason is the soft information at the output of the 16QAM demodulator is less reliable than the soft information at the output of the QPSK demodulator at the first few iterations. Thus for 16QAM the improvement in reliability of the soft outputs from the channel decoder is more helpful to the estimators and thus demodulator in the beginning. In other words, for higher-order modulation the performance of the receiver is more sensitive to the performance of the channel estimator.

4.6 Conclusions

The problem of channel estimation for BICM-MIMO systems has been investi-

gated in this paper. First, the conventional iterative channel estimator was improved using a switching scheme (SAICE). The improvement involves switching from the training phase to the data phase based on the performance of the iterative channel estimator. The demodulator with the SAICE, however, does not converge quickly at low SNR and with a higher-order modulation. To overcome this problem, a new estimator, referred to as CICE, that guarantees convergence is devised utilizing both the training and data segments on every iteration. For the development of both SAICE and CICE, a closed-form expression for the mutual information at the input of the channel estimator was derived. Carefully crafted approximations for the MI in the corner cases show that for the CICE, the MI is a function of the system's parameters, channel length and power allocation to the training sequence. They also show that the MI at the input of the SAICE is only a function of the system's parameters and the channel length after the first iteration. EXIT chart results show that for QPSK, the demodulator with both estimators converges after 4 iterations for an SNR of 3dB. Moreover, the BER results show that the performance of the receiver with the more robust CICE approaches the performance of the receiver that has perfect CSI and a moderate to high SNR. Changing the modulation to 16QAM increases the number of iterations required for convergence.

References

[1] S. H. Muller-Weinfurtner, "Coding approaches from multiple antenna transmission in fast fading and OFDM," *IEEE Trans. Signal Processing.*, vol. 50, pp. 2442–2450, Oct. 2002.

[2] T. Yang and J. Yuan, "Performance of MIMO-BICM with parallel interference canceller on slow fading channels," *Electronics Letters*, vol. 42, pp. 1292–1293, 2006.

[3] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Signal Processing.*, vol. 44, pp. 927–945, May 1998.

[4] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in rayleigh fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 944–957, May 2001.

[5] G. Taricco, , and E. Biglieri, "SpaceTime decoding with imperfect channel estimation," *IEEE Trans. on Wireless Commun.*, vol. 4, pp. 1874–1888, Jul. 2005.

[6] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data- aided channel estimation and perfect interleaving," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 2358–2368, Dec. 2001.

[7] A. Maaref and S. Aissa, "Optimized rate-adaptive PSAM for MIMO MRC systems with transmit and receive CSI imperfections," *IEEE Trans. Commun.*, vol. 57, pp. 821–830, Mar. 2009.

[8] J. Wang, O. Y. Wen, and S. Li, "Soft-output MMSE MIMO detector under ML channel estimation and channel correlation," *IEEE Signal Processing Letters*, vol. 16, pp. 667–670, Aug. 2009.

[9] M. Coldrey and P. Bohlin, "Training-based MIMO systemspart I: Performance comparison," *IEEE Trans. Signal Processing.*, vol. 55, pp. 5464–5476, Nov. 2007.

[10] M. Teimouri, N. Rezaee, and A. Hedayat, "Concatenated coded modulation techniques and orthogonal space-time block codes in the presence of fading channel estimation errors," *IET Commun.*, vol. 4, pp. 135–143, 2010.

[11] I.-W. Lai, S. Godtmann, T.-D. Chiueh, G. Ascheid, and H. Meyr, "Asymptotic BER analysis for MIMO-BICM with zero-forcing detectors assuming imperfect CSI," *Proc. IEEE Int. Conf. Commun.*, (Beijing), pp. 1238–1242,, May 2008.

[12] M. Nicoli, S.Ferrara, and U. Spagnolini, "Soft-iterative channel estimation: Methods and performance analysis," *IEEE Trans. Signal Processing.*, vol. 55, pp. 2993–3006, Jun. 2007.

[13] Y. nan Lee, A. Ashikhmin, and J.-T. Chen, "Impact of soft channel construction on iterative channel estimation and data decoding for multicarrier systems," *IEEE Trans. Commun.*, vol. 7, pp. 2762–2770, Jul. 2008.

[14] Y. Huang and J. A. Ritcey, "EXIT chart analysis of BICM-ID with imperfect channel state information," *IEEE Commun. Letters*, vol. 7, pp. 434–436, Sep. 2003.

[15] C. Novak, G. Lechner, and G. Matz, "MIMO-BICM with imperfect channel state information: EXIT chart analysis and LDPC code optimization," *Proc. Asilomar Conf. Signals, Systems and Computers*, (Pacific Grove, CA), pp. 443–447, Oct. 2008.

[16] Y. Huang and J. A. Ritcey, "Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2549–2558, Sep. 2005.

[17] N. Gresset, J. J. Boutros, and L. Brunel, "Optimal linear precoding for BICM over MIMO channels," in *ISIT*, (Chicago, IL), p. 66, Jul. 2004.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, Inc., 2006.

[19] M. A. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3165–3173, Nov. 2006.

[20] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1732–1742, May 2007.

5. Training Design for Precoded BICM-MIMO Systems in Block-Fading Channels

Published as:

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt, "Training design for precoded BICM-MIMO systems in block-fading channels", *EURASIP Journal on Wireless Communications and Networking 2012*, 2012:80.

As shown in the previous chapter, one way to implement space-time coding techniques is to concatenate multiple antennas with BICM. Further improvements in diversity and coding gain are provided by applying a linear precoder in BICM-MIMO systems. On the other hand, it was shown that using the information of training sequence in subsequent iterations of an iterative channel estimator results in significant performance improvement. However, by increasing the number of transmit antennas, the training overhead required for the initialization of the channel estimator increases.

Given the importance of training overhead (i.e., bandwidth efficiency) and motivated by the performance improvement using training information in subsequent iteration, the manuscript in this chapter studies a new training design for BICM-MIMO system with a linear precoder. In particular, the training sequence is inserted before the precoder and within data symbols. The manuscript designs training symbols and their positions (i.e., training pattern) by minimizing the CRB. It is shown that depending on system parameters, the training overhead decreases at least by the power of two while the error performance of the system improves significantly.

Training Design for Precoded BICM-MIMO Systems in Block-Fading Channels

Zohreh Andalibi, Ha H. Nguyen, J. Eric Salt

Abstract

In order to improve bandwidth efficiency and error performance, a new training scheme is proposed for bit-interleaved-coded modulation in multiple-input multiple-output (BICM-MIMO) systems. Typically, in a block-fading channel, the training overhead used for obtaining channel knowledge is proportional to a power of 2 of the number of transmit antennas. However, this overhead can be reduced by embedding pilot symbols within data symbols before precoding. The values, positions, and the number of pilot symbols are found by minimizing the Cramer-Rao bound on the channel estimation error. Computer simulations are presented to demonstrate the advantage of the proposed scheme over other training methods, in terms of both the mean-square-error of the channel estimation and the system's frame-error-rate.

Index terms

BICM-MIMO, block fading, channel estimation, training design, pilot symbols, Cramer-Rao bound, iterative receiver.

5.1 Introduction

The pioneering work on multiple-input multiple-output (MIMO) systems [1] shows that a MIMO system can provide a multiplexing gain and accordingly high spectral

Manuscript received June 16, 2011; revised December 8, 2011, and accepted March 4, 2012.

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt are with the Department of Electrical & Computer Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon, SK, Canada S7N 5A9, and TRLabs of Saskatchewan. Emails: z.andalibi@usask.ca, ha.nguyen@usask.ca, eric.salt@usask.ca.

efficiency over slow fading channels. On the other hand, to achieve a high diversity order, space–time transmission techniques can be implemented at the transmitter [2,3]. To achieve both high diversity order and coding gain in coded modulation systems, the concept of space–time transmission has also been applied [4,5]. In such systems, space–time transmission is typically implemented using a linear space–time matrix, or equivalently a linear precoder, so that a single modulation symbol is efficiently transmitted across multiple transmit antennas. Among many research works on precoder design for coded modulation systems with multiple antennas, the design that considers all the relevant components of the transmitter, namely precoding, modulation, and interleaver, can be found in [5–7]. Specifically, a *full-rate* precoder with any size and for any number of transmit antennas is designed in [6] to maximize the achievable diversity order and coding gain in MIMO block-fading channels.

It is shown in [6] that the maximum achievable diversity order can be realized by an iterative receiver that employs a soft-input soft-output detector [5] and under the assumption of having the *perfect* channel state information (CSI) at the receiver. In practice, however, CSI has to be estimated using a channel estimator and it is never perfect. Two types of channel estimators have been used for MIMO block-fading channels in coded modulation systems, i.e., training-based and semi-blind channel estimators [8, 9]. In both types of channel estimators, known signals are used to estimate the CSI at the first iteration of the iterative receiver.

Conventionally, for block-fading channels, known signals or the training sequence is included at the beginning of each data block, which is called time-multiplexed training or pilot symbol-assisted modulation (PSAM) scheme [10]. This scheme however reduces bandwidth efficiency of MIMO systems, since the amount of training overhead needed is at least a power of 2 of the number of transmit antennas [11] to ensure the identifiability of the MIMO channel. A straightforward application of the PSAM scheme to a BICM-MIMO system would be time-multiplex data information with the training information after the precoder.

As an alternative to the above conventional PSAM scheme, a potential benefit can

be sought by time-multiplexing data information with the training information *before* the precoder in the transmitter. This new approach shall reduce the required training overhead compared to the conventional PSAM, since the transmitted training symbols are spread over more time periods; thanks to the precoder. This approach shall be referred to as precoded PSAM (PPSAM) . Investigating power and time allocations of the training symbols in PPSAM scheme is the main objective of this article.

Moreover, by multiplexing the training sequence before precoder, training symbols can be exploited in both the initialization and iteration phases of the iterative channel estimation process. This is different from a conventional iterative channel estimator using PSAM scheme, in which training sequence is only used at the initialization phase. A natural question is whether the optimal training design for the initialization phase using PPSAM scheme is still optimal for subsequent iterations of an iterative channel estimator. On the one hand, the channel estimation error at the initialization phase translates to an SNR shift in the BER performance [8]. On the other hand, the channel estimation error from the last iteration of the iterative estimator has a strong impact on the error floor of the BER performance [12]. Therefore, optimal training sequence should be designed carefully that considers both initialization and iteration phases.

One of different criteria that have been used to design training sequences is the minimization of the Cramer-Rao bound (CRB) of the channel estimation error [10]. This criterion shall be used in this article due to two main reasons. First, it is directly related to the channel estimation error. Second, since the CRB is a lower bound on the mean-squared-error (MSE) of any unbiased estimator, designing training sequences using this criterion would be applicable to many estimation algorithms. Other design criteria, such as maximizing the channel capacity [8] and minimizing the outage probability [13], are based on some specific channel estimation algorithms.

The article is organized as follows. The system model of BICM-MIMO is presented in Section 5.2. In Section 5.3 a lower bound on the MSE of the channel estimator is obtained and the training sequence is designed by minimizing this bound. Section 5.4

provides numerical results and comparisons. Section 5.5 concludes the article.

5.2 System Model

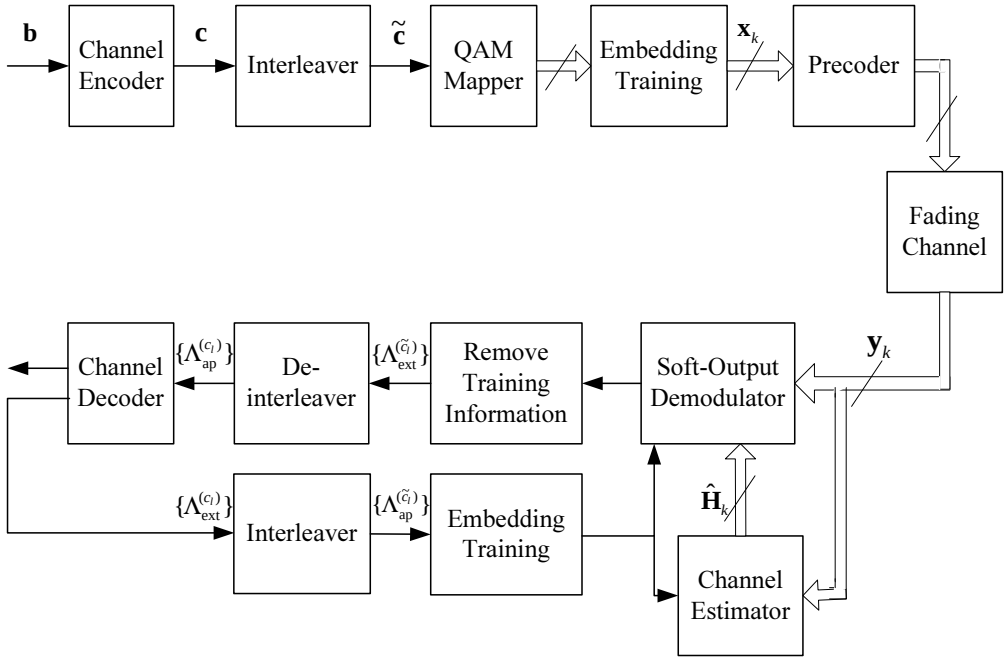

Figure 5.1 Block diagram of a BICM-MIMO system with a linear precoder and proposed training insertion.

Fig. 5.1 shows the block diagram of a BICM-MIMO system under consideration. At the transmitter, a channel encoder with a rate-r error-correcting code converts the vector of information bits $\mathbf{b}$ into a codeword $\mathbf{c}$. The coded bits are then interleaved by a random interleaver as described in [6] to produce the interleaved codeword $\tilde{\mathbf{c}}$. The interleaved codewords are segmented into groups of $(Nn_{\text{t}} - N_{\text{p}}) \times m$ bits, where N is the spreading factor of the precoder, n_{t} is the number of transmit antennas, N_{p} is the number of pilot symbols in Nn_{t} precoded symbols and m is the number of bits carried by one symbol of a QAM constellation whose size is $|\Omega| = 2^m$. Next, the coded bits are mapped to $(Nn_{\text{t}} - N_{\text{p}})$ QAM constellation points. In this step, N_{p} known pilot symbols are inserted in every segmented group of $(Nn_{\text{t}} - N_{\text{p}})$ data symbols to produce N super-symbols. Here, each super-symbol refers to a group of n_{t} consecutive symbols. Investigating the positions and the number of pilot symbols

(i.e., $N_{\rm p}$) to be used in each $Nn_{\rm t}$ symbols is the main objective of this article.

Every group of N super-symbols is then spread over N time periods using a linear precoder $\mathbf{G}$. The $Nn_{\rm t} \times Nn_{\rm t}$ matrix $\mathbf{G}$ multiplies a vector of $Nn_{\rm t}$ QAM symbols at the precoder input, and generates $Nn_{\rm t}$ symbols to be transmitted over $n_{\rm t}$ antennas, over N time periods. This is illustrated in Fig. 5.2. Let $\mathbf{x}_k = [x_{(k-1)Nn_{\rm t}+1}, x_{(k-1)Nn_{\rm t}+2}, \ldots, x_{(k-1)Nn_{\rm t}+Nn_{\rm t}}]$ be the kth vector to be precoded. Then, $\mathbf{x}_k\mathbf{G}$ gives the precoded symbols. Here, x_i's are complex data or pilot symbols belonging to the 2^m-QAM constellation Ω. It is assumed that the data symbols x_i's are i.i.d with variance σ_x^2. After precoding, precoded symbols are transmitted through $n_{\rm t}$ transmit antennas over a block-fading channel.

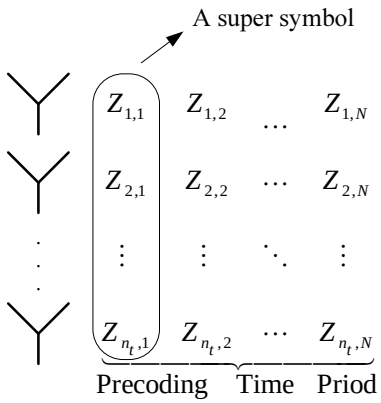

Figure 5.2 Spreading a precoded symbol over n_t antennas and N time periods - denoted by $\mathbf{Z}_k = [Z_{1,k}, Z_{2,k}, \ldots, Z_{n_{\rm t},k}]^T$.

With $n_{\rm t}$ transmit antennas and $n_{\rm r}$ receive antennas, the channel is modeled by an $n_{\rm t} \times n_{\rm r}$ matrix. For frequency-flat Rayleigh fading, coefficients of the channel matrix are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with variance σ_h^2. The channel is assumed to be block fading with n_c different channel realizations during each codeword. For the kth symbol to be precoded, $\mathbf{x}_k$, the $Nn_{\rm t} \times Nn_{\rm r}$ extended channel matrix, $\mathbf{H}_k$, can be written as

$$\mathbf{H}_k = \text{diag}\left\{\underbrace{\mathbf{H}_k^{[1]}, \ldots, \mathbf{H}_k^{[1]}}_{N/n_s}, \mathbf{H}_k^{[2]}, \ldots, \mathbf{H}_k^{[2]}, \ldots, \mathbf{H}_k^{[n_s]}, \ldots, \mathbf{H}_k^{[n_s]}\right\}, \qquad (5.1)$$

where n_s is the number of distinct channel realizations during N time periods of each codeword. To simplify the notation it is also assumed that n_s divides N. For example, if the length of a codeword is 64 and $n_c = 32$, then choosing $N = 2$ would make $n_s = 1$, whereas choosing $N = 4$ gives $n_s = 2$. Notation $\mathbf{H}_k^{[t]}$ refers to the $n_t \times n_r$ complex matrix that defines the tth channel realization included in n_s channel realizations. The extended channel input/output relationship is expressed by

$$\mathbf{y}_k = \mathbf{x}_k \mathbf{G} \mathbf{H}_k + \mathbf{w}_k \tag{5.2}$$

where $\mathbf{y}_k = [y_{(k-1)Nn_r+1}, y_{(k-1)Nn_r+2}, \ldots, y_{(k-1)Nn_r+Nn_r}]$ is the received vector at the kth precoding time period and $\mathbf{w}_k$ is the noise vector with size $1 \times Nn_r$ whose components are i.i.d zero-mean circularly symmetric Gaussian random variables with variance N_0.

It is noted from (5.2) that although both data and pilot symbols are precoded, the part of the precoder that multiplies the pilot symbols depends on the positions of the pilot symbols in $\mathbf{x}_k$. Equivalently, the design of the pilot symbols is governed by the properties of the precoder used. Since this study adopts the transmission framework and precoder design in [6], it is useful to review the properties of the precoder proposed in [6].

In general, the properties of the precoder in [6] are established by the maximum-likelihood decoding analysis and an assumption of ideal channel interleaving. Specifically, this linear precoder which achieves full diversity order and maximum coding gain satisfies the following two conditions:

- A genie condition, which guarantees orthogonal and equal norm sub-rows in the linear precoding matrix. Each sub-row has size n_t in a precoding matrix with size $Nn_t \times Nn_t$.

- Dispersive nucleo algebraic (DNA) condition, which is based on Proposition 2

In practice, since n_s is typically an approximated value over some range and since N can be selected, such an assumption can be fulfilled.

in [6], forces null and orthogonal nucleotides with size $s' = N/n_s$. Nucleotides refer to subparts of sub-rows with size s'.

A linear precoder that satisfies the above two sets of conditions is called DNA-cyclo precoder and has the best performance in terms of achieving diversity and coding gains with low complexity receiver when $N \leq n_{\mathrm{t}}$. It is suggested in [6] that to generate one class of such a precoder, a $Ns' \times Ns'$ cyclotomic rotator, denoted by $\mathbf{\Phi}$, that satisfies the genie condition is first selected. Then the orthogonal nucleotides are placed inside an $Nn_{\mathrm{t}} \times Nn_{\mathrm{t}}$ matrix and they are separated with null nucleotides. Therefore, the DNA-cyclo precoder matrix can be expressed by subparts of a cyclotomic rotator as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[1][1]} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[N][1]} \\ \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[1][2]} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[N][2]} \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[1][Ns']} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/s'} \otimes \mathbf{\Phi}^{[N][Ns']} \end{bmatrix} \tag{5.3}$$

where $\mathbf{\Phi}^{[i][j]}$ is the ith sub-row of the jth row of $\mathbf{\Phi}$ with size $1 \times s'$, $\mathbf{I}_n$ is an identity matrix with size $n \times n$ and $\otimes$ denotes the Kronecker product.

The properties that shall be useful for the problem considered in this article, which are implied directly from the genie and DNA conditions, are $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}_{Ns'}$ and $\mathbf{\Phi}^{[i][t]}(\mathbf{\Phi}^{[j][t]})^H = \frac{1}{N}\delta(i-j)$. It is also useful to point out that each component of $\mathbf{\Phi}$ has an exponential form with a scaling factor of $\frac{1}{\sqrt{Ns'}}$.

The iterative receiver is also shown in Fig. 5.1. The channel estimator produces an estimate of the channel using the minimum MSE (MMSE) criterion based on the training sequence. Details about channel estimation with the proposed method of inserting training sequence shall be given in Section 5.3. After channel estimation is performed using the training signal, the soft-input soft-output demodulator uses the MMSE criterion to demodulate the data. The soft-output MMSE demodulator computes the extrinsic information for the interleaved bits, $\{\Lambda_{\mathrm{ext}}^{(\tilde{c}_l)}\}$, from the received symbols. To obtain Λ-values, the demodulator exploits the *a priori* information of

the coded bits coming from the decoder, $\{\Lambda_{\text{ap}}^{(\tilde{c}_l)}\}$, and the channel estimate $\hat{\mathbf{H}}_k$. In the first iteration, the demodulator assumes that the *a priori* Λ-values are zero, except for the pilot symbols. For the corresponding bits of the pilot symbols, the demodulator uses a large number, say ± 100 as their *a priori* Λ-values. The de-interleaved outputs, i.e., $\{\Lambda_{\text{ap}}^{(c_l)}\}$, become the *a priori* Λ-values used in the channel decoder shown in Fig. 5.1 after removing the information of pilot symbols. The channel decoder uses the maximum *a posteriori* probability (MAP) algorithm to compute the extrinsic Λ-values $\{\Lambda_{\text{ext}}^{(c_l)}\}$ for all coded bits, which are used again in the next iteration in the demodulator. In subsequent iterations, soft information from the decoder is used to improve the performance of the channel estimator. The detailed operation of the iterative channel estimator is discussed in the following sections.

5.3 Training Design and Channel Estimator

As discussed before, the criterion used for training design in this article is the CRB on the channel estimation error. The bound states that the MSE of any unbiased estimator is lower bounded by the trace of inverse of complex Fisher information matrix (FIM) [14]. To derive FIM, the relation between the channel input and channel output during one block-length, i.e., N/n_s time periods, whose corresponding channel matrix is $\mathbf{H}_k^{[t]}$, is of interest. In the following, index k is omitted, since it suffices to consider the transmission of a single precoded symbol for the purpose of channel estimation. With the previously described structure of the precoder, the channel output during one super-symbol time is given by

$$\mathbf{y}^{[i,t]} = \left(\mathbf{I}_{n_{\text{r}}} \otimes \left(\sum_{\tau=1}^{Ns'} \mathbf{x}^{[\tau]} \otimes \mathbf{\Phi}^{[i,t][\tau]}\right)\right) \mathbf{h}^{[t]} + \mathbf{w}^{[i,t]}; \quad t = 1, \ldots, n_s,\ i = 1, \ldots, s' \quad (5.4)$$

where $\mathbf{y}^{[i,t]} = \mathbf{y}^{[(t-1)s'+i]}$ represents the $((t-1)s'+i)$th received symbol during N time periods, with size $n_{\text{r}} \times 1$. Moreover, $\mathbf{h}^{[t]}$ is the column vector formed by vertically stacking the columns of an $n_{\text{t}} \times n_{\text{r}}$ channel realization matrix $\mathbf{H}^{[t]}$ and $\mathbf{x}^{[\tau]}$'s are constructed by splitting $\mathbf{x}$ in Ns' sub-vectors with size $1 \times n_{\text{t}}/s'$. In the following, we call these sub-vectors $\mathbf{x}^{[\tau]}$'s *nucleo* symbols.

It is quite obvious from (5.4) that, to have all the received super-symbols, $\mathbf{y}^{[i,t]}$, contain training information, there should be at least one pilot nucleo (i.e., n_{t}/s' pilot symbols) in each group of Ns' nucleos to be precoded.

With the above structure of the proposed training sequence, the number of pilot symbols in Nn_{t} transmitted symbols would be $N_{\mathrm{p}} = n_{\mathrm{p}} \times n_{\mathrm{t}}/s'$, where n_{p} nucleo symbols in a symbol to be precoded are assigned to training sequence. Therefore, (5.4) can be rewritten as

$$\mathbf{y}^{[i,t]} = \left(\mathbf{I}_{n_{\mathrm{r}}} \otimes \left(\sum_{\tau \in \mathcal{I}_{\mathrm{d}}} \mathbf{x}_{\mathrm{d}}^{[\tau]} \otimes \mathbf{\Phi}_{\mathrm{d}}^{[i,t][\tau]} + \sum_{\tau \in \mathcal{I}_{\mathrm{p}}} \mathbf{x}_{\mathrm{p}}^{[\tau]} \otimes \mathbf{\Phi}_{\mathrm{p}}^{[i,t][\tau]} \right) \right) \mathbf{h}^{[t]} + \mathbf{w}^{[i,t]}, \tag{5.5}$$

where $\mathcal{I}_{\mathrm{d}}$ and $\mathcal{I}_{\mathrm{p}}$ are sets of indexes from $\{1, \ldots, Ns'\}$, that are assigned to data and pilot nucleos, respectively, and $|\mathcal{I}_{\mathrm{d}}| + |\mathcal{I}_{\mathrm{p}}| = (Ns' - n_{\mathrm{p}}) + n_{\mathrm{p}} = Ns'$. Note that the subscripts "d" and "p" are used to differentiate between data and pilot nucleos. For convenience, the notations $\mathbf{\Phi}_{\mathrm{p}}^{[i,t][\tau]}$ and $\mathbf{\Phi}_{\mathrm{d}}^{[i,t][\tau]}$ are used to refer to sub-rows of $\mathbf{\Phi}$ that are multiplied by pilot and data nucloes, i.e., $\mathbf{x}_{\mathrm{p}}^{[\tau]}$ and $\mathbf{x}_{\mathrm{d}}^{[\tau]}$, respectively. Furthermore, in the following the notation $\mathbf{T}^{[i,t]}$ is used for $\mathbf{I}_{n_{\mathrm{r}}} \otimes \left(\sum_{\tau \in \mathcal{I}_{\mathrm{p}}} \mathbf{x}_{\mathrm{p}}^{[\tau]} \otimes \mathbf{\Phi}_{\mathrm{p}}^{[i,t][\tau]} \right)$.

The derivation of FIM is given in the next section. Pilot symbols are exploited at the initialization phase and in subsequent iterations considering the special structure of the training sequence. In general, training design can be investigated for these two phases separately. However, for the precoder adopted in this article, the optimal training design obtained for the initialization phase turns out to also be optimal for the iteration phase. Nevertheless, the optimal numbers of pilot nucleos in these two phases of channel estimation are not the same.

5.3.1 Fisher Information Matrix

The key steps in deriving the FIM in the *initialization phase* are now given. Without loss of generality we drop superscript t in (5.5) and perform all the derivations for the first block period (i.e., $t = 1$). Collecting all the observations during the first block period of length s' in a vector $\boldsymbol{\varphi}$, the FIM for the channel estimation problem

at the initialization phase is defined and computed as

$$\begin{aligned}\mathrm{FIM}^{\mathrm{init}}(n_{\mathrm{p}},\mathbf{x}_{\mathrm{p}},\mathcal{I}_{\mathrm{p}}) &= E_{\boldsymbol{\varphi},\mathbf{h}}\left\{\left[\frac{\partial \ln p(\boldsymbol{\varphi},\mathbf{h})}{\partial \mathbf{h}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\varphi},\mathbf{h})}{\partial \mathbf{h}^*}\right]^H\right\} \\ &= E_{\mathbf{h}}\left\{E_{\boldsymbol{\varphi}}\left\{\left[\frac{\partial \ln p(\boldsymbol{\varphi}|\mathbf{h})}{\partial \mathbf{h}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\varphi}|\mathbf{h})}{\partial \mathbf{h}^*}\right]^H\Big|\mathbf{h}\right\}\right\} \\ &+ E_{\mathbf{h}}\left\{\left[\frac{\partial \ln p(\mathbf{h})}{\partial \mathbf{h}^*}\right]\left[\frac{\partial \ln p(\mathbf{h})}{\partial \mathbf{h}^*}\right]^H\right\} \qquad (5.6)\end{aligned}$$

where $\mathrm{FIM}^{\mathrm{init}}(n_{\mathrm{p}},\mathbf{x}_{\mathrm{p}},\mathcal{I}_{\mathrm{p}})$ shows the dependency of FIM on those parameters of interest. Using the i.i.d. assumption on noise and data, $p(\boldsymbol{\varphi}|\mathbf{h})$ can be approximated as a complex normal distribution with mean $\mu=[\mu_1^T,\ldots,\mu_{s'}^T]^T$ and covariance $\mathbf{R}_{\boldsymbol{\varphi}}=\mathrm{diag}[\mathbf{R}_1,\ldots,\mathbf{R}_{s'}]$. Moreover, it follows from (5.5) that $\mu_i=E_{\boldsymbol{\varphi}}\{\mathbf{y}^{[i]}|\mathbf{h}\}=\mathbf{T}^{[i]}\mathbf{h}$ and

$$\mathbf{R}_i = \mathcal{H}\left(\sigma_x^2 I_{n_t/s'}\otimes((\boldsymbol{\Phi}_d^{[i]})^T(\boldsymbol{\Phi}_d^{[i]})^*)\right)\mathcal{H}^H+N_0\mathbf{I}_{n_{\mathrm{r}}} \qquad (5.7)$$

where $\mathcal{H}=(\mathbf{H}^{[1]})^T$ and $\boldsymbol{\Phi}_d^{[i]}$ is the ith sub-matrix of $\boldsymbol{\Phi}$ with size $(Ns'-n_{\mathrm{p}})\times s'$ that is assigned to data symbols.

The i.i.d. assumptions on noise and data make the FIM additive. Specifically, $\mathrm{FIM}^{\mathrm{init}}(n_{\mathrm{p}},\mathbf{x}_{\mathrm{p}},I_{\mathrm{p}})=\sum_{i=1}^{s'}\mathrm{FIM}_i^{\mathrm{init}}$. The quantity $\mathrm{FIM}_i^{\mathrm{init}}$ is obtained as follows:

$$\mathrm{FIM}_i^{\mathrm{init}}=E_{\mathbf{h}}\left\{E_{\mathbf{y}}\left\{\frac{\partial \ln p(\mathbf{y}|\mathbf{h})}{\partial \mathbf{h}^*}\left(\frac{\partial \ln p(\mathbf{y}|\mathbf{h})}{\partial \mathbf{h}^*}\right)^H\Big|\mathbf{h}\right\}\right\}+\sigma_h^{-2}\mathbf{I}_{n_tn_{\mathrm{r}}}.$$

We know that

$$\ln p(\mathbf{y}|\mathbf{h})=\text{Constant}-\ln|\mathbf{R}_i|-(\mathbf{y}-\mu_i)^H\mathbf{R}_i^{-1}(\mathbf{y}-\mu_i). \qquad (5.8)$$

and $\frac{\partial \ln|\mathbf{R}_i|}{\partial h_l^*}=\text{trace}\left(\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial h_l^*}\right)$. Therefore,

$$\frac{\partial \mathbf{R}_i}{\partial h_l^*} = \mathcal{H}\left(\sigma_x^2 I_{n_t/s'}\otimes((\boldsymbol{\Phi}_d^{[i]})^T(\boldsymbol{\Phi}_d^{[i]})^*)\right)\boldsymbol{\Sigma}_l^T \qquad (5.9)$$

where $\boldsymbol{\Sigma}_l$ is an $n_{\mathrm{r}}\times n_t$ null matrix with only a single element of 1 at position $(\lfloor\frac{l-1}{n_t}\rfloor+1,(l-1 \bmod n_t)+1)$. The derivative of the third term in (5.8) is

$$\frac{\partial(\mathbf{y}-\mu_i)^H\mathbf{R}_i^{-1}(\mathbf{y}-\mu_i)}{\partial h_l^*} = -\frac{\partial \mu_i^H}{\partial h_l^*}\mathbf{R}_i^{-1}(\mathbf{y}-\mu_i)+(\mathbf{y}-\mu_i)^H\frac{\partial \mathbf{R}_i^{-1}}{\partial h_l^*}(\mathbf{y}-\mu_i)$$

where $\frac{\partial \mathbf{R}_i^{-1}}{\partial h_l^*} = -\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial h_l^*}\mathbf{R}_i^{-1}$ and $\frac{\partial \mathbf{R}_i}{\partial h_l^*}$ is given by (5.9). In addition,

$$\frac{\partial \mu_i^H}{\partial h_l^*} = \frac{\partial \mathbf{h}^H}{\partial h_l^*}(\mathbf{T}^{[i]})^H = \mathbf{e}_l^T(\mathbf{T}^{[i]})^H$$

where $\mathbf{e}_l$ is an $n_t n_{\rm r} \times 1$ null vector with a single element 1 at position l.

Using all the above equations and after some manipulations, one has

$$\begin{aligned}(\mathrm{FIM}_i^{\rm init})_{l,j} &= E_{\mathbf{h}}\{\mathbf{e}_l^T(\mathbf{T}^{[i]})^H\mathbf{R}_i^{-1}\mathbf{T}^{[i]}\mathbf{e}_j \\ &+ \mathrm{tr}(\mathbf{R}_i^{-1}\mathcal{H}\mathbf{A}^{[i]}\mathbf{\Sigma}_l^T\mathbf{R}_i^{-1}\mathbf{\Sigma}_j(\mathbf{A}^{[i]})^H\mathcal{H}^H)\} + \sigma_h^{-2}\delta(l-j),\end{aligned}$$

where $\mathbf{A}^{[i]} \equiv \left(\sigma_x^2 I_{n_t/s'} \otimes ((\mathbf{\Phi}_d^{[i]})^T(\mathbf{\Phi}_d^{[i]})^*)\right)$.

Using the fact that $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB})$ and summing over s' quantities $\mathrm{FIM}_i^{\rm init}$, the total FIM is given by,

$$\mathrm{FIM}^{\rm init}(n_{\rm p}, \mathbf{x}_{\rm p}, \mathcal{I}_{\rm p}) = E_{\mathbf{h}}\left\{\sum_{i=1}^{s'}\mathbf{R}_i^{-1} \otimes \left((\mathbf{X}_{\rm p}^{[i]})^H\mathbf{X}_{\rm p}^{[i]}\right) + \mathbf{R}_i^{-1} \otimes \mathbf{Q}_i\right\} + s'\sigma_h^{-2}\mathbf{I}_{n_t n_{\rm r}} \quad (5.10)$$

where $\mathbf{X}_{\rm p}^{[i]} = \sum_{\tau\in\mathcal{I}_{\rm p}} \mathbf{x}_{\rm p}^{[\tau]} \otimes \mathbf{\Phi}_{\rm p}^{[i][\tau]}$, and $\mathbf{Q}_i = (\mathbf{A}^{[i]})^T\mathcal{H}^T\mathbf{R}_i^{-1}\mathcal{H}^*(\mathbf{A}^{[i]})^*$.

For designing training sequence, (5.10) can be simplified further using numerical calculation. Using numerical calculation, it is observed that for a Rayleigh-distributed channel, the matrix $E_{\mathbf{h}}\{\mathbf{R}_i^{-1}\}$ in (5.10) is approximately a diagonal matrix, $\alpha\mathbf{I}_{n_{\rm r}}$. This observation means that $E_{\mathbf{h}}\{\mathbf{Q}_i\}$ can be approximated by $n_r\alpha\sigma_h^2(\mathbf{A}^{[i]})^T(\mathbf{A}^{[i]})^*$. Then, by performing the expectation operation and using the factorization property of the Kronecker product, (5.10) can be represented as

$$E_{\mathbf{h}}\left\{\sum_{i=1}^{s'}\alpha\mathbf{I}_{n_{\rm r}} \otimes \left((\mathbf{X}_{\rm p}^{[i]})^H\mathbf{X}_{\rm p}^{[i]}\right) + \alpha\mathbf{I}_{n_{\rm r}} \otimes n_r\alpha\sigma_h^2(\mathbf{A}^{[i]})^T(\mathbf{A}^{[i]})^*\right\} + s'\sigma_h^{-2}\mathbf{I}_{n_{\rm r}} \otimes \mathbf{I}_{n_{\rm t}} =$$

$$\mathbf{I}_{n_{\rm r}} \otimes \left(\sum_{i=1}^{s'}\alpha\left((\mathbf{X}_{\rm p}^{[i]})^H\mathbf{X}_{\rm p}^{[i]}\right) + n_r\alpha^2\sigma_h^2(\mathbf{A}^{[i]})^T(\mathbf{A}^{[i]})^* + s'\sigma_h^{-2}\mathbf{I}_{n_{\rm t}}\right)$$

Moreover, using the property of the Kronecker product $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$, it follows that $(\mathbf{X}_{\rm p}^{[i]})^H\mathbf{X}_{\rm p}^{[i]} = \sum_{\tau\in\mathcal{I}_{\rm p}}\sum_{\tau'\in\mathcal{I}_{\rm p}}((\mathbf{x}_{\rm p}^{[\tau]})^H\mathbf{x}_{\rm p}^{[\tau']}) \otimes ((\mathbf{\Phi}_{\rm p}^{[i][\tau]})^H\mathbf{\Phi}_{\rm p}^{[i][\tau']})$.

Using the matrix inversion lemma, one has $\mathbf{R}_i^{-1} = \left(\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H + N_0\mathbf{I}_{n_{\rm r}}\right)^{-1} = N_0^{-1}\mathbf{I}_{n_r} + N_0^{-2}\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H\left(I_{n_r} + N_0^{-1}\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H\right)^{-1}$. Therefore, for high SNR, $E\{\mathbf{R}_i^{-1}\}$ can be approximated by $N_0^{-1}\mathbf{I}_{n_r}$.

Therefore (5.10) can be further simplified to

$$\mathrm{FIM}^{\mathrm{init}}(n_{\mathrm{p}}, \mathbf{x}_{\mathrm{p}}, \mathcal{I}_{\mathrm{p}}) = \mathbf{I}_{n_{\mathrm{r}}} \otimes \left(\alpha \sum_{\tau \in \mathcal{I}_{\mathrm{p}}} \sum_{\tau' \in \mathcal{I}_{\mathrm{p}}} (\mathbf{x}_{\mathrm{p}}^{[\tau]})^H \mathbf{x}_{\mathrm{p}}^{[\tau']} \otimes \sum_{i=1}^{s'} (\mathbf{\Phi}_{\mathrm{p}}^{[i][\tau]})^H \mathbf{\Phi}_{\mathrm{p}}^{[i][\tau']} + n_r \alpha^2 \sigma_h^2 \sum_{i=1}^{s'} (\mathbf{A}^{[i]})^T (\mathbf{A}^{[i]})^* + s' \sigma_h^{-2} \mathbf{I}_{n_{\mathrm{t}}} \right) \tag{5.11}$$

In general, the second term in (5.11) depends on $\mathcal{I}_{\mathrm{p}}$, but not on the training symbols, whereas the first term depends on both $\mathbf{x}_{\mathrm{p}}$ and $\mathcal{I}_{\mathrm{p}}$. Although both terms depend on n_{p}, how $\mathrm{FIM}^{\mathrm{init}}$ depends on n_{p} is determined by $\mathcal{I}_{\mathrm{p}}$. Therefore, in the following $\mathcal{I}_{\mathrm{p}}$ and $\mathbf{x}_{\mathrm{p}}$ are first optimized. Then n_{p} is determined for the optimized $\mathcal{I}_{\mathrm{p}}$.

For the *iteration phase*, specifically the last iteration, estimation and detection are implemented using information about the data symbols as well as the pilot symbols. Thus, the parameter of interest in deriving FIM is $\boldsymbol{\theta} = [\mathbf{h}^T \mathbf{x}_{\mathrm{d}}]^T$. Moreover, $\mu_i = E_{\boldsymbol{\varphi}}\{\mathbf{y}^{[i]}|\boldsymbol{\theta}\} = \left(\mathbf{I}_{n_{\mathrm{r}}} \otimes \left(\sum_\tau \mathbf{x}^{[\tau]} \otimes \mathbf{\Phi}^{[i][\tau]}\right)\right) \mathbf{h}$ and $\mathbf{R}_i = N_0 \mathbf{I}_{n_{\mathrm{r}}}$. By replacing $\boldsymbol{\theta}$ in (5.6) for $\mathbf{h}$ and after some manipulations, the FIM for channel estimation in the iteration phase is given by

$$\mathrm{FIM}^{\mathrm{iter}}(n_{\mathrm{p}}, \mathbf{x}_{\mathrm{p}}, \mathcal{I}_{\mathrm{p}}) = N_0^{-1} \mathbf{I}_{n_{\mathrm{r}}} \otimes \left(\frac{N s' - n_{\mathrm{p}}}{N} \sigma_x^2 \mathbf{I}_{n_{\mathrm{t}}} + \sum_{\tau \in \mathcal{I}_{\mathrm{p}}} \sum_{\tau' \in \mathcal{I}_{\mathrm{p}}} (\mathbf{x}_{\mathrm{p}}^{[\tau]})^H \mathbf{x}_{\mathrm{p}}^{[\tau']} \otimes \sum_{i=1}^{s'} (\mathbf{\Phi}_{\mathrm{p}}^{[i][\tau]})^H \mathbf{\Phi}_{\mathrm{p}}^{[i][\tau']} \right) + s' \sigma_h^{-2} \mathbf{I}_{n_{\mathrm{t}} n_{\mathrm{r}}}. \tag{5.12}$$

5.3.2 Optimization of Training Symbols and Their Positions

This section is first concerned with minimizing the CRB expression for the initialization phase. The minimization is under a constraint on the power budget for the training sequence. Such a constraint is expressed as

$$\sum_{\tau \in \mathcal{I}_{\mathrm{p}}} \left(\mathbf{x}_{\mathrm{p}}^{[\tau]} \otimes \mathbf{\Phi}_{\mathrm{p}}^{[i][\tau]}\right) \left((\mathbf{x}_{\mathrm{p}}^{[\tau]})^H \otimes (\mathbf{\Phi}_{\mathrm{p}}^{[i][\tau]})^H\right) \;\; \leq \;\; P_t. \tag{5.13}$$

Using the properties of the precoder employed in this study, the above constraint can be simplified to $\frac{s'}{N} \sum_{\tau' \in \mathcal{I}_{\mathrm{p}}} \mathbf{x}_{\mathrm{p}}^{[\tau]} (\mathbf{x}_{\mathrm{p}}^{[\tau]})^H \leq P_t$. The other obvious constraint is that the training symbols should be selected from QAM constellation Ω. Then, the training

symbols, $\mathbf{x}_\mathrm{p}$'s and their positions, specified by $\mathcal{I}_\mathrm{p}$, are obtained by solving the following constrained optimization problem:

$$\begin{aligned}\min_{\mathbf{x}_\mathrm{p},\mathcal{I}_\mathrm{p}} \mathrm{CRB}^{\mathrm{init}}(n_\mathrm{p},\mathbf{x}_\mathrm{p},\mathcal{I}_\mathrm{p}) &= \min_{\mathbf{x}_\mathrm{p},\mathcal{I}_\mathrm{p}} \mathrm{tr}(\mathrm{FIM}^{\mathrm{init}}(n_\mathrm{p},\mathbf{x}_\mathrm{p},\mathcal{I}_\mathrm{p})^{-1}) \\ \text{s.t. } &\begin{cases} \frac{s'}{N}\sum_{\tau\in\mathcal{I}_\mathrm{p}} \mathbf{x}_\mathrm{p}^{[\tau]}(\mathbf{x}_\mathrm{p}^{[\tau]})^H \le P_t \\ (\mathbf{x}_\mathrm{p}^{[\tau]})_j \in \Omega,\ j=1,\ldots,n_\mathrm{t}/s',\ \tau\in\mathcal{I}_\mathrm{p} \end{cases}\end{aligned} \tag{5.14}$$

where $(\mathbf{x}_\mathrm{p}^{[\tau]})_j$ is the jth pilot symbol in the τth pilot nucleo and the FIM is given in (5.11).

To proceed, lets consider two separate cases for problem (5.14): $n_\mathrm{p}=1$ and $n_\mathrm{p}\ge 2$.

Case 1 ($n_\mathrm{p}=1$): In this case the FIM is simplified to

$$\begin{aligned}\mathbf{I}_{n_\mathrm{r}} \otimes \Bigg(&\alpha\left((\mathbf{x}_\mathrm{p}^{[\tau]})^H\mathbf{x}_\mathrm{p}^{[\tau]}\right)\otimes\sum_{i=1}^{s'}\left((\mathbf{\Phi}_\mathrm{p}^{[i][\tau]})^H\mathbf{\Phi}_\mathrm{p}^{[i][\tau]}\right) \\ &+n_r\alpha^2\sigma_h^2\sum_{i=1}^{s'}(\mathbf{A}^{[i]})^T(\mathbf{A}^{[i]})^* + s'\sigma_h^{-2}\mathbf{I}_{n_\mathrm{t}}\Bigg),\end{aligned} \tag{5.15}$$

Because of the shift-invariant property of (5.15) with respect to τ, τ can be any value in the set $\{1,2,\ldots,Ns'\}$. For simplicity, set $\tau=1$ and the superscript τ is omitted. Using the fact that if $\mathbf{X}>0$ then $\mathrm{tr}(\mathbf{X}^{-1})\ge\sum_i 1/(\mathbf{X})_{i,i}$, the original optimization problem is simplified by minimizing the lower bound of the objective function.

On the other hand, $\sum_{i=1}^{s'}\left(\left(\mathbf{\Phi}_\mathrm{p}^{[i][\tau]}\right)^H\mathbf{\Phi}_\mathrm{p}^{[i][\tau]}\right)=\frac{1}{N}\mathbf{I}_{s'}$, $\sum_{i=1}^{s'}(\mathbf{A}^{[i]})^T(\mathbf{A}^{[i]})^*=\frac{\sigma_x^4}{s'}\left((\frac{Ns'-1}{N})^2+(\frac{1}{N})^2\right)\mathbf{I}$ and the constraint is $\frac{s'}{N}\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}^H=\frac{s'}{N}\sum_{j=1}^{n_\mathrm{t}/s'}|(\mathbf{x}_\mathrm{p})_j|^2$. Therefore, it is not hard to see that the solution of the simplified optimization problem is $|(\mathbf{x}_\mathrm{p})_1|^2=|(\mathbf{x}_\mathrm{p})_2|^2=\cdots=|(\mathbf{x}_\mathrm{p})_{n_\mathrm{t}/s'}|^2=\frac{NP_t}{n_\mathrm{t}}$. It means that all pilot symbols should have the same power. For example, one can select corner points of the QAM constellations for the training symbols.

Case 2 ($n_\mathrm{p}\ge 2$): In this case there are two options for the placements of pilot nucleos. The first option is to group all pilot nucleos in one single cluster and the second option is to spread pilot nucleos. It can be shown that the CRB is invariant

with respect to a shift of the placements of pilot nucleos in both options. Therefore, it suffices to select one cluster or one spread placement. However, the precoder has been designed such that the soft-output demodulator works with uncorrelated inputs and putting pilot nucleos between data nucleos may violate this condition. That condition is satisfied when $\mathbf{A}^{[i]}$ has a diagonal form. The implication of this property is to place pilot nucloes equi-spaced in $\mathbf{x}_k$ and $\mathcal{I}_{\rm p} = \{i_0 + kn;\ \ k = 0, \ldots, n_{\rm p} - 1\}$, where $n = Ns'/n_{\rm p}$ and $i_0 \in \{1, \ldots, n\}$, which leads to $\mathbf{A}^{[i]} = \sigma_x^2 \frac{Ns' - n_{\rm p}}{Ns'} \mathbf{I}_{s'}$. In this selection it is supposed that $n_{\rm p}$ is divisible by Ns'.

Then the FIM in (5.11) can be represented by

$$\mathbf{I}_{n_{\rm r}} \otimes \left(\frac{1}{N} \alpha \sum_{\tau \in \mathcal{I}_{\rm p}} (\mathbf{x}_{\rm p}^{[\tau]})^H \mathbf{x}_{\rm p}^{[\tau]} \otimes \mathbf{I}_{s'} + n_r \alpha^2 \sigma_x^4 \sigma_h^2 \frac{1}{s'} (\frac{Ns' - n_{\rm p}}{N})^2 \mathbf{I}_{n_{\rm t}} + s' \sigma_h^{-2} \mathbf{I}_{n_{\rm t}} \right) \quad (5.16)$$

To obtain the above expression of the objective function, the following property has been used:

$$\left(\sum_{i=1}^{s'} (\mathbf{\Phi}_{\rm p}^{[i][\tau]})^H \mathbf{\Phi}_{\rm p}^{[i][\tau']} \right)_{n,l} = \begin{cases} \frac{1}{N}, & \tau = \tau'; n = l \\ 0, & \text{otherwise} \end{cases} \quad (5.17)$$

Moreover, the only term that depends on the training symbols is $\sum_{\tau \in \mathcal{I}_{\rm p}} (\mathbf{x}_{\rm p}^{[\tau]})^H \mathbf{x}_{\rm p}^{[\tau]}$ in (5.16).

Finally, using the constraint on training power, which can be written as

$$\frac{s'}{N} \sum_{\tau \in \mathcal{I}_{\rm p}} \sum_{j=1}^{nt/s'} |(\mathbf{x}_{\rm p}^{[\tau]})_j|^2 \leq P_t, \quad (5.18)$$

the solution is given by $\sum_\tau |(\mathbf{x}_{\rm p}^{[\tau]})_j|^2 = \frac{NP_t}{n_{\rm t}};\ j = 1, \ldots, n_{\rm t}/s'$.

Now consider the training design for the iteration phase. Observe that all the terms in (5.12) have diagonal forms with equal diagonal elements, except $\sum_{\tau \in \mathcal{I}_{\rm p}} \sum_{\tau' \in \mathcal{I}_{\rm p}} (\mathbf{x}_{\rm p}^{[\tau]})^H \mathbf{x}_{\rm p}^{[\tau']} \otimes \sum_{i=1}^{s'} (\mathbf{\Phi}_{\rm p}^{[i][\tau]})^H \mathbf{\Phi}_{\rm p}^{[i][\tau']}$. This means that the solution of problem (5.14), but with $\text{FIM}^{\text{init}}(n_{\rm p}, \mathbf{x}_{\rm p}, \mathcal{I}_{\rm p})$ replaced by $\text{FIM}^{\text{iter}}(n_{\rm p}, \mathbf{x}_{\rm p}, \mathcal{I}_{\rm p})$, is to choose equal diagonal elements for this term. Therefore, the training sequence designed for the initialization is also optimal for the iteration phase.

Table 5.1 Optimum n_{p} for several sets of parameters $\{n_{\mathrm{t}}, n_{\mathrm{r}}, N\}$

n_{t}	n_{r}	N	n_{p}
2	2	2	1
4	2	2	2
4	2	4	4
4	4	2	1
4	4	4	1

In summary, by selecting pilot nucleos such that the sum of the powers of their corresponding pilot symbols with the same indexes are equal, the bound on CRB is minimized. The above condition can give different selections for pilot symbols from a two-dimensional constellation. It should be pointed out, however, that not all selections guarantee that pilot symbols belong to standard QAM constellations.

5.3.3 Determination of the Number of the Training Symbols

For block-fading channels, the number of pilot nucleos, i.e., n_{p}, should be as small as possible that meets the power constraint. Using a larger value for n_{p} wastes bandwidth and does not change the system performance.

The optimum numbers of the training symbols in the initialization phase and iteration phase are not the same. This is explained as follows. At the initialization, by looking at (5.7), it is observed that the first term in (5.11) is an increasing function of n_{p}. However, the second term is a decreasing function of n_{p} that is multiplied by n_r. Therefore, n_{p} that minimizes the CRB are determined by the summation of these two terms, which is also determined by the value of n_{r}. Table 5.1 gives several examples of optimal n_{p} for different sets of n_{t}, n_{r} and N. For the iteration phase, the expression in (5.12) means that the CRB in the iteration phase always increases by increasing n_{p}. Since it is assumed that there is perfect information about the data symbols in the iteration phase, which is not the case in reality, it is most appropriate

to select n_{p} considering only the initialization phase.

To demonstrate the optimal training design, Fig. 5.3 shows a graphical structure for a simple example, where $P_t = 4\sigma_x^2$, $n_{\mathrm{p}} = 2$, $N = 2$, $n_{\mathrm{t}} = 4$ and $n_{\mathrm{r}} = 2$. In this example, $n_{\mathrm{s}} = 1$. Then the size of pilot nucleos should be $n_{\mathrm{t}}/s' = 2$, where $s' = N/n_s = 2$.

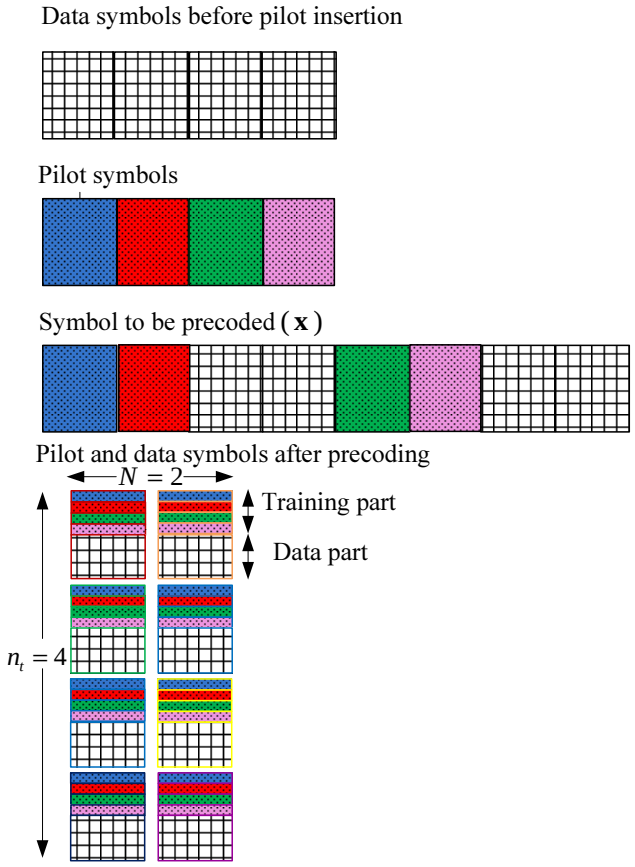

Figure 5.3 Structure of the proposed scheme for the training sequence - when $N = 2$, $n_{\mathrm{t}} = 4$, $n_{\mathrm{r}} = 2$ and $n_{\mathrm{p}} = 2$.

5.3.4 Channel Estimation

For the channel estimation task, one can view the received vector during one block length as $\boldsymbol{\varphi}^{[t]} = [(\mathbf{y}^{[1,t]})^T, (\mathbf{y}^{[2,t]})^T, \ldots, (\mathbf{y}^{[s',t]})^T]^T$.

At the initialization, the mean and covariance matrix of this vector are given in Section 5.3.1. By treating the data symbols as nuisance parameters, the MMSE

channel estimate can be found as [14]

$$\hat{\mathbf{h}}^{[t]} = \sigma_h^2 \mathbf{T}^H \left(\sigma_h^2 \mathbf{T}\mathbf{T}^H + \mathbf{R}_{\boldsymbol{\varphi}^{[t]}} \right) \boldsymbol{\varphi}^{[t]} \tag{5.19}$$

where $\mathbf{T} = [(\mathbf{T}^{[1]})^T, \ldots, (\mathbf{T}^{[s']})^T]^T$.

In the subsequent iterations, soft information from the decoder is used to improve the performance of the channel estimator. The channel estimator uses such information to compute new estimates of the channel coefficients using expected values of the data symbols. Therefore, the interleaved $\{\Lambda_{\text{ext}}^{(c_i)}\}$ from the decoder are fed back to the estimator to calculate the expected values of the data symbols, i.e., $E\{\mathbf{x}_\text{d}\}$. The entries of $E\{\mathbf{x}_\text{d}\}$ are calculated using $\{\Lambda_{\text{ap}}^{(\tilde{c}_i)}\}$ at each iteration by $E\{(\mathbf{x}_\text{d})_i\} = \sum_{x \in \Omega} x \cdot p((\mathbf{x}_\text{d})_i = x)$. The detailed derivations of the probability $p((\mathbf{x}_\text{d})_i = x)$ from Λ-values are given in [15] (note that the calculation depends on the mapping rule in $\boldsymbol{\Omega}$).

To verify the results obtained in this section, Section 5.4 compares numerically the MSE performance of the above channel estimator obtained with the optimal and suboptimal training sequences.

5.4 Illustrative Results

In this section, the frame-error-rate (FER) and MSE performances of BICM-MIMO systems using a MMSE iterative channel estimator are presented. The space–time precoder is the DNA-cyclo precoder that satisfies the properties outlined in Section 5.2. We consider quadrature phase-shift keying (QPSK) modulation with Gray mapping.

The MSE performance of a BICM-MIMO for a codeword length of 4×1024 bits is shown for a 4×2 block-fading MIMO channel in Fig. 5.4, when $n_c = 2$. In this figure, E_b is the energy per information bit. The code used is the 16-state convolutional code with generator polynomials $(23, 35)$ in octal form. In Fig. 5.4, the MSE curves are obtained after 1 and 5 iterations of the iterative channel estima-

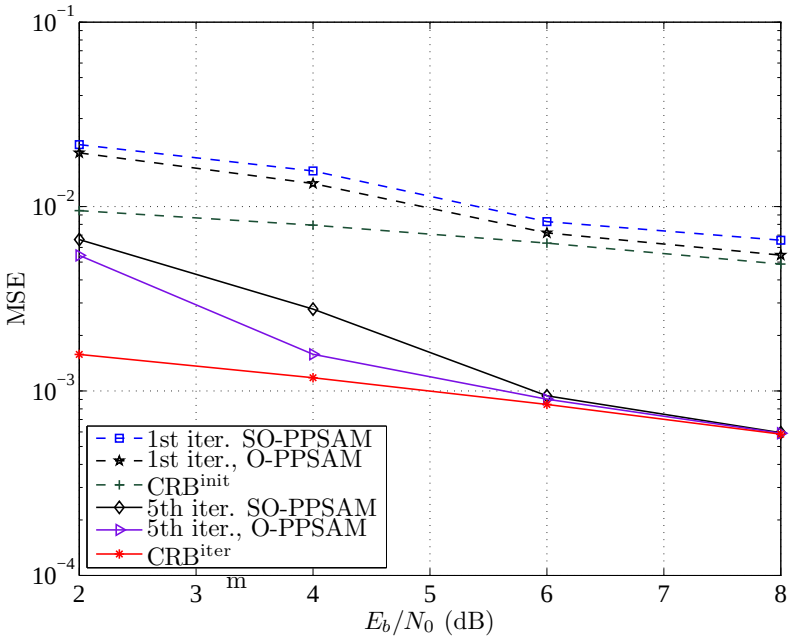

Figure 5.4 Comparison of MSE performance obtained with the optimal PPSAM and the sub-optimal PPSAM - over a 4×2 block-fading channel with $n_c = 2$, when $N = 2$ and $n_{\text{p}} = 2$ after 1 and 5 iterations of iterative channel-estimation/demodulation/decoding.

tion/demodulation/decoding, with the following cyclotomic rotator [16]:

$$\mathbf{\Phi} = \frac{1}{2}\begin{bmatrix} 1 & 1 & e^{j6\pi/15} & -e^{j6\pi/15} \\ e^{j2\pi/15} & je^{j2\pi/15} & -e^{j8\pi/15} & je^{j8\pi/15} \\ e^{j4\pi/15} & -e^{j4\pi/15} & e^{j10\pi/15} & e^{j10\pi/15} \\ e^{j6\pi/15} & -je^{j6\pi/15} & -e^{j12\pi/15} & -je^{j12\pi/15} \end{bmatrix}$$

and when the setting for N, n_s, n_p and P_t in Fig. 5.3 are used. The channel is generated randomly and is assumed to be Rayleigh distributed. For the purpose of comparison, the results for MSE performances of the optimal PPSAM, denoted by O-PPSAM and the suboptimal PPSAM, denoted by SO-PPSAM as well as the CRB are shown in Fig. 5.4.

For SO-PPSAM, two pilot nucleos are inserted as one cluster in front of data nucleos in a symbol to be precoded. In contrast, in the case of O-PPSAM, the optimized training sequence embeds the pilot nucleos at the first and third positions

of $Ns' = 4$ positions for nucleos. The MSE curves show that the performance of the optimal scheme is better than the sub-optimum scheme for the first iteration (i.e., initialization). In fact the MSE performance of the proposed scheme closely approaches the CRB at high E_b/N_0 after 5 iterations.

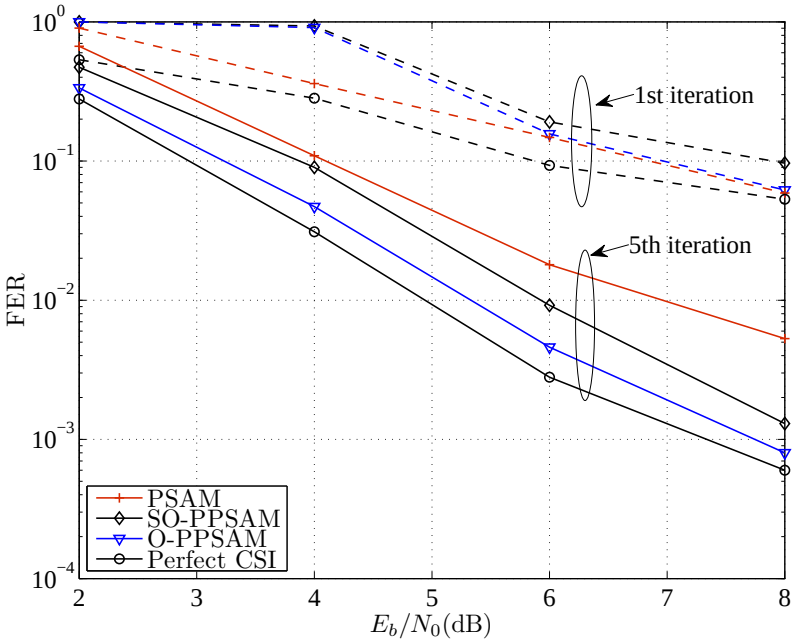

Figure 5.5 Comparison of FER performance obtained with the optimal PPSAM, sub-optimal PPSAM and PSAM scheme - over a 4×2 block-fading channel with $n_c = 2$, when $N = 2$ and $n_\mathrm{p} = 2$ after 1 and 5 iterations of iterative channel-estimation/demodulation/decoding.

In Fig. 5.5, the FER performance of the system with the PPSAM schemes is compared with the conventional PSAM training scheme for the same system parameters as in Fig. 5.4. The top curve is the FER performance of the system with the conventional PSAM training scheme. Note that for a fair comparison, the training scheme in PSAM also meets the training power constraint as $\mathrm{trace}(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^H) = P_t$, where $\mathbf{X}_\mathrm{p}$ is the training matrix placed at the beginning of each block of the precoded symbols. The optimal option for PSAM scheme in terms of minimizing the FER as proposed in [11] is to select $\mathbf{X}_\mathrm{p}$ to have orthogonal columns. The simplest option is $\sqrt{2 \times \sigma_x^2/n_\mathrm{t}}\mathbf{I}_{n_\mathrm{t}} = \sqrt{\sigma_x^2}\mathbf{I}_{n_\mathrm{t}}$, which results in the same power budget as that of the

proposed scheme. As can be seen from Fig. 5.5, the O-PPSAM scheme offers 0.5 dB performance gain as compared to the SO-PPSAM scheme at FER= 10^{-2}. In comparison with PSAM, the performance of the PSAM scheme is about 0.5–1.5 dB worse than the proposed scheme depending on E_b/N_0 after 5 iterations. This is expected because the pilot information is embedded in the precoded symbols for the proposed scheme and not for the PSAM scheme. In this way, the demodulator can also make use of this information. Note, however, that for the first iteration, since there is no information about data, PSAM works the best. More importantly, while the proposed scheme uses a little bandwidth for training information (for the system considered in this figure the training overhead is $n_{\mathrm{p}} \times n_{\mathrm{t}}/s' = 4$), the training overhead of PSAM scheme is $n_{\mathrm{t}} \times n_{\mathrm{t}} = 16$, which is quadruple.

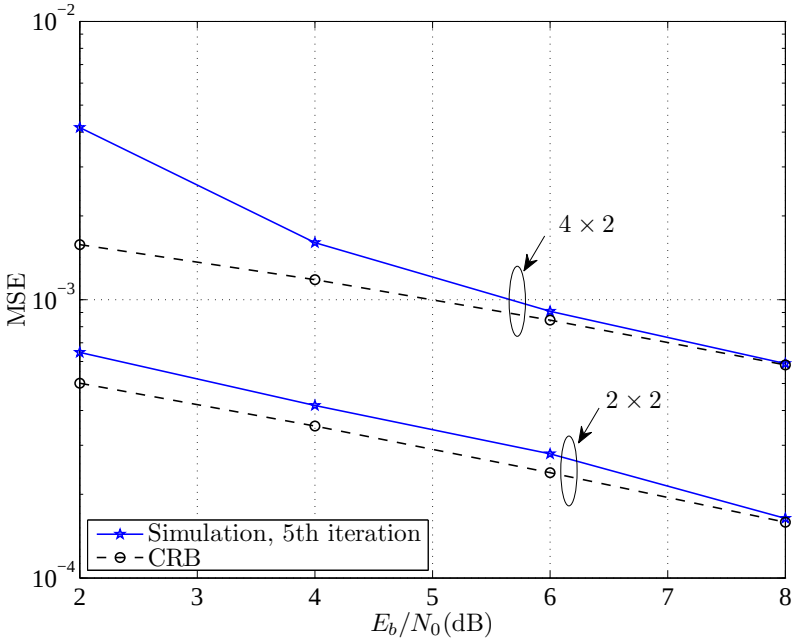

Figure 5.6 Comparison of MSE performance obtained with the optimal PP-SAM for 2×2 and 4×2 block-fading channel with $n_c = 2$, when $N = 2$ and $n_{\mathrm{p}} = 2$ after 5 iterations of iterative channel-estimation/demodulation/decoding.

To investigate the effect of the number of transmit antennas, two different systems, one with 2×2 channel and one with 4×2 MIMO channel, are compared in Figs. 5.6

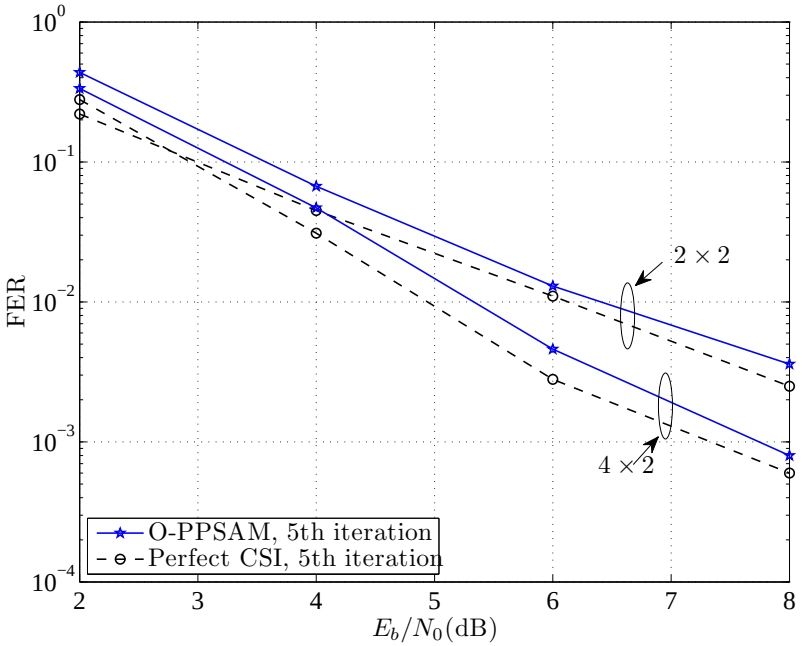

Figure 5.7 Comparison of FER performance obtained with the optimal PP-SAM for 2×2 and 4×2 block-fading channel with $n_c = 2$, when $N = 2$ and $n_{\mathrm{p}} = 2$ after 5 iterations of iterative channel-estimation/demodulation/decoding.

and 5.7 in terms of MSE and FER, respectively. For both channels, $n_p = 2$ and the optimum scheme are used when $N = 2$, while other system parameters are the same as those used for Fig. 5.4. As can be seen from Fig. 5.6, the MSE of the channel estimation increases when increasing the number of transmit antennas. This is expected because there are more channels to be estimated for the same amount of training information and power as done in the comparison. Nevertheless, the gain in diversity by using more antennas can still improve the overall FER performance as seen in Fig. 5.7.

5.5 Conclusion

In this article, a new training design for a BICM-MIMO system over a block-fading channel has been proposed. The design inserts pilot symbols into the data symbols before precoding. The new training sequence improves bandwidth efficiency as com-

pared to the conventional PSAM scheme and can also be used by the demodulator in the receiver. In order to design the optimal training symbols and their positions, the CRB on the channel estimations at the initialization and at the iteration phases are minimized. Compared to PSAM, performance improvement achieved with the proposed training is about 1.5 dB at a FER level of 10^{-2}.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, July 2003.

[2] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," vol. 16, pp. 1451–1458, Oct. 1998.

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

[4] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1453–1467, July 1998.

[5] J. Boutros, N. Gresset, and L. Brunel, "Turbo coding and decoding for multiple antenna channels," in *International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 1–8, 2003.

[6] N. Gresset, L. Brunel, and J. Boutros, "Spacetime coding techniques with bit-interleaved coded modulations for mimo block-fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2156–2178, May 2008.

[7] N. Gresset, J. J. Boutros, and L. Brunel, "Optimal linear precoding for BICM over MIMO channels," in *ISIT*, (Chicago, IL), p. 66, Sept. 2004.

[8] M. Coldrey and P. Bohlin, "Training-based MIMO systems, Part I: Performance comparison," *IEEE Trans. Signal Processing*, vol. 55, pp. 5464–5476, Nov. 2007.

[9] M. Nicoli, S.Ferrara, and U. Spagnolini, "Soft-iterative channel estimation: Methods and performance analysis," *IEEE Trans. Signal Processing*, vol. 55, pp. 2993–3006, June 2007.

[10] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1403–1418, May 2004.

[11] G. Taricco and E. Biglieri, "Spacetime decoding with imperfect channel estimation," *IEEE Trans. on Wireless Commun.*, vol. 4, pp. 1874–1888, July 2005.

[12] Y. Huang and J. A. Ritcey, "Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2549–2558, Sept. 2005.

[13] P. Piantanida and S. M. Sadough, "On the outage capacity of a practical decoder accounting for channel estimation inaccuracies," *IEEE Trans. Commun.*, vol. 57, pp. 1341–1350, May 2009.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall PTR, 1993.

[15] M. A. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3165–3173, Nov. 2006.

[16] G. M. Kraidy and P. Rossi, "Full-diversity iterative MMSE receivers with space-time precoders over block-fading MIMO channels," in *IEEE Int. Conf. Wireless Commun and Signal Processing*, (Suzhou), pp. 1–5, Oct. 2010.

6. Precoder Design for BICM-MIMO Systems under Channel Estimation Error

Submitted as:

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt, "Precoder Design for BICM-MIMO System under Channel Estimation Errors" (submitted to *Springer Wireless Personal Communications*).

In Chapter 5, it was shown that putting training sequence before the precoder with an equi-spaced and equi-power pattern improves the error performance and decreases the required training overhead in precoded BICM-MIMO systems. Part of this improvement is due to sending less training which results in sending more data symbols in one codeword and consequently exploiting more diversity and coding gain. In this chapter the effect of the training sequence on the diversity order and coding gain is studied for an arbitrary precoder. To compensate better for the signalling overhead used for the training, the part of the precoder used for sending training symbols and the training symbols are designed jointly to maximize the achievable diversity and coding gain. To investigate these effects, the coding gain and diversity order are obtained by analyzing the PEP under the imperfect CSI and for a precoded training sequence. It is shown that the proposed precoder and training sequence lead to a less training overhead and a further improvement in error performance compared to the approach presented in Chapter 5.

Precoder Design for BICM-MIMO Systems under Channel Estimation Error

Zohreh Andalibi, Ha H. Nguyen, J. Eric Salt

Abstract

If properly designed, the use of a linear precoder can achieve the maximum coding gain and diversity order in bit-interleaved coded modulation with multiple-input multiple-output (BICM-MIMO) systems. However, such maximum coding gain and diversity order are compromised under the practical scenario of having imperfect channel state information (CSI) at the receiver. To alleviate the impact of imperfect CSI on the coding gain and diversity order, joint linear precoder and training pattern are designed in this paper. The design is carried out by considering both the pairwise error probability (PEP) and the mean square error (MSE) of the channel estimator. The effectiveness of the proposed design is illustrated by comparing its performance with the performance obtained when only the training pattern is designed for an precoder optimized under the perfect CSI. In particular, simulation results show that a 1.5 dB gain is achieved by the proposed design.

Index terms

BICM-MIMO, pairwise error probability, diversity order, coding gain, linear precoder, pilot symbols, mean-square error, iterative receiver.

6.1 Introduction

To enjoy the maximum coding gain in a BICM-MIMO system with iterative demodulation/decoding, CSI needs to be perfectly known at the iterative receiver. Conventionally, CSI is estimated using a known training sequence [1,2] that is embedded

Zohreh Andalibi, Ha H. Nguyen and J. Eric Salt are with the Department of Electrical & Computer Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon, SK, Canada S7N 5A9, and TRLabs of Saskatchewan. Emails: z.andalibi@usask.ca, ha.nguyen@usask.ca, eric.salt@usask.ca.

within data symbols at the transmitter. When a space-time precoder is used, it has been shown that inserting training sequence between data symbols before the precoder leads to a significant performance improvement in [3]. Inserting training before precoding also saves time and space for sending data symbols as compared to the conventional training scheme (i.e., pilot-symbol assisted modulation (PSAM) scheme).

In particular, reference [3] designs training sequences to work with a full-rate linear precoder. This full-rate precoder, proposed in [4], is called *dispersive nucleo algebraic* (DNA) precoder. The DNA precoder was shown in [4] to maximize the coding gain and diversity order of a BICM-MIMO system in a block-fading environment. It should be pointed out, however, that the DNA precoder is devised under the assumption of perfect CSI and maximum-likelihood (ML) detection. This means that the DNA precoder might not effectively minimize the impact of channel estimation error. This is because by adding a training sequence before the precoder, a part of the precoder is used for the training sequence and the main purpose of using the DNA precoder for data detection is compromised.

In this paper, the impact of imperfect CSI is analyzed by obtaining the asymptotic pairwise error probability (PEP). The derivation is based on the classical analysis approach of ML detector in [4] and MMSE detector in [5]. Then the coding gain and diversity order are quantified from the obtained PEP expression. It will be shown how the channel estimation error is characterized by the structure of the precoder. The precoder and the training sequence are then jointly designed by maximizing the coding gain and diversity order and minimizing MSE of the channel estimation.

The paper is organized as follows. The system model of BICM-MIMO is presented in Section 6.2. In Section 6.3, the asymptotic PEP performance is analyzed for the MMSE detector under a block-fading channel and imperfect CSI. Then, the channel estimation error of an iterative channel estimator with the training sequence is obtained. The precoder and the training sequence are designed in Section 6.4. Section 6.5 provides numerical results and performance comparison. Section 6.6 concludes

the paper.

6.2 System Model

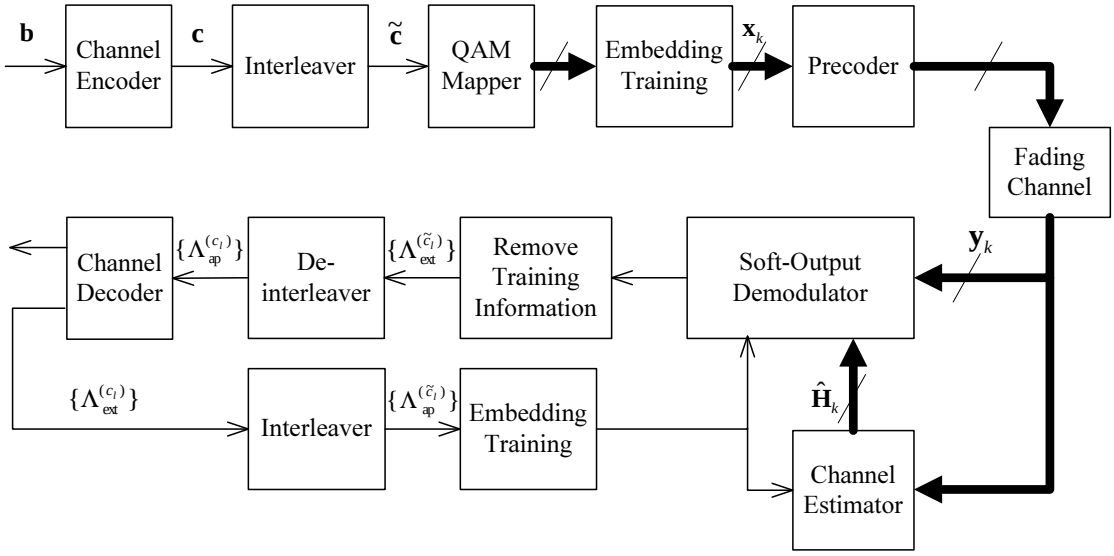

Figure 6.1 Block-diagram of a BICM-MIMO system with linear precoder and iterative receiver.

The BICM-MIMO system model considered in this paper is presented in Fig. 6.1. The vector of information bits $\mathbf{b}$ is first encoded with a rate-r binary convolutional code into a codeword $\mathbf{c}$. The coded bits are then interleaved and mapped to QAM constellation points. For the purpose of channel estimation, pilot symbols are embedded in data symbols at this stage. Let $\mathbf{x}_k = [x_{(k-1)Nn_\text{t}+1}, x_{(k-1)Nn_\text{t}+2}, \ldots, x_{(k-1)Nn_\text{t}+Nn_\text{t}}]$ be the kth symbol vector to be precoded and L be the time duration in which the CSI is nearly the same. Here, N is the spreading factor of the precoder and n_t is the number of transmit antennas. In one block length of L, the first P symbol vectors to be precoded contain pilot symbols and for the rest of $(L/N) - P$ symbol vectors, $\mathbf{x}_k$ has only data symbols. The data symbols are segmented into groups of Nn_t symbols, while for those symbol vectors that contain both data and pilots, N_p pilot symbols are inserted in every segmented group of $(Nn_\text{t} - N_\text{p})$ data symbols to produce $\mathbf{x}_k$. The selection of P and N_p for the training sequence shall be discussed in Section 6.4. The resulting Nn_t symbols in $\mathbf{x}_k$, are then precoded by matrix $\mathbf{G}_k$. As an exam-

ple, Fig. 6.2 shows one block length of precoded symbols with the training insertion scheme where $n_{\mathrm{t}} = 2$, $N = 2$, $P = 1$, $N_p = 2$, and $L = 10$. The design and analysis in this paper only focuses on the first part of each block that contains both data and pilot symbols.

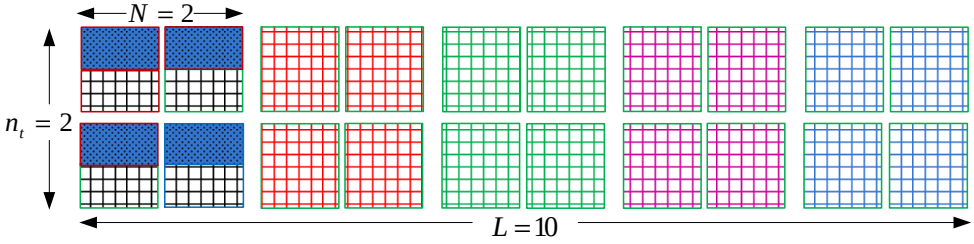

Figure 6.2 Example of one block of precoded symbols with training sequence.

After precoding, precoded symbols are transmitted through n_{t} transmit antennas over a block-fading channel with n_c different channel realizations during each codeword. Each of these n_c channel realizations, denoted by $\mathbf{H}_t$, is the same for whole block length L. It has size $n_{\mathrm{t}} \times n_{\mathrm{r}}$ and changes independently in the next block. For frequency-flat Rayleigh fading, coefficients of the channel matrix are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with variance σ_h^2. For simplicity, it is assumed that $\sigma_h^2 = 1$ and N divides L. Then, corresponding to the kth vector to be precoded, $\mathbf{x}_k$, the $Nn_{\mathrm{t}} \times Nn_{\mathrm{r}}$ extended channel matrix, $\mathbf{H}_k$, can be written as $\mathbf{H}_k = \mathrm{diag}\left\{ \underbrace{\mathbf{H}_t, \ldots, \mathbf{H}_t}_{N}, \right\}$. By definition, $t = \lfloor \frac{k-1}{(L/N)} \rfloor + 1$ and $1 \leq t \leq n_c$. The extended channel input/output relationship is expressed by

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

where $\mathbf{y}_k = [y_{(k-1)Nn_{\mathrm{r}}+1}, y_{(k-1)Nn_{\mathrm{r}}+2}, \ldots, y_{(k-1)Nn_{\mathrm{r}}+Nn_{\mathrm{r}}}]$ is the received vector at the kth precoding time period and $\mathbf{w}_k$ is the noise vector with size $1 \times Nn_{\mathrm{r}}$, whose components are i.i.d zero-mean circularly symmetric Gaussian random variables with variance N_0. $\mathbf{G}_k$ is the precoder matrix that can be partitioned into two parts. The first part contains only the rows of $\mathbf{G}_k$ that are multiplied by data symbols. This part does not change with the time index and is denoted by $\mathbf{G}^{(\mathrm{d})}$. The second part contains the rows that are multiplied by the pilot symbols in $\mathbf{x}_k$ and is denoted by

$\mathbf{G}_k^{(\mathrm{p})}$. In essence, the purpose of $\mathbf{G}^{(\mathrm{d})}$ is to implement space-time coding on data symbols, while $\mathbf{G}_k^{(\mathrm{p})}$ is used to assist channel estimation.

At the receiver, data symbols are detected using a soft-input soft-output MMSE detector. Among practical demodulation approaches for MIMO systems, the soft-output MMSE demodulator is used as it has low complexity and good performance. Under the assumption of perfect CSI, the MMSE detector calculates the lth component of the kth symbol vector at each iteration as [6],

$$\begin{aligned} \hat{x}_{l,k} &= (\mathbf{y}_k - \mathbf{z}_l\tilde{\mathbf{H}}_k)(\tilde{\mathbf{H}}_k^H\mathbf{\Gamma}_l\tilde{\mathbf{H}}_k + N_0\mathbf{I}_{Nn_\mathrm{r}})^{-1}\tilde{\mathbf{h}}_{l,k}^H, \\ & \quad 1 \le l \le Nn_\mathrm{t}, \end{aligned} \tag{6.2}$$

where $\tilde{\mathbf{H}}_k = \mathbf{G}_k\mathbf{H}_k$ is an equivalent correlated MIMO channel, whose correlation is due to $\mathbf{G}_k$. Furthermore, $\mathbf{z}_l = [\bar{x}_1, \bar{x}_2, \cdots, 0, \bar{x}_{l+1}, \cdots, \bar{x}_{Nn_\mathrm{t}}]$ and $\mathbf{\Gamma}_l = \mathrm{diag}(\bar{\sigma}_{x_1}^2, \cdots, \sigma_x^2, \bar{\sigma}_{x_{l+1}}^2, \cdots, \bar{\sigma}_{x_{Nn_\mathrm{t}}}^2)$, where $\bar{x}_l$ and $\bar{\sigma}_{x_l}^2$ are the *a priori* mean and variance of the lth coded symbol, x_l, calculated from the soft information delivered by the decoder (i.e., the log-likelihood-ratios, $\Lambda_{\mathrm{ap}}^{(\tilde{c}_l)}$) at each iteration, respectively. Lastly, $\tilde{\mathbf{h}}_{l,k} = \mathbf{G}_{l,k}\mathbf{H}_k$ is the lth row of $\tilde{\mathbf{H}}_k$ generated from the lth row of $\mathbf{G}_k$ (written as $\mathbf{G}_{l,k}$).

The asymptotic performance analysis is based on the assumption of perfect soft information from the decoder and therefore the error-free feedback analysis of (6.2) is of interest. With such an assumption, $\mathbf{z}_l = \mathbf{0}$ and $\mathbf{\Gamma}_l = \sigma_x^2\mathbf{I}_{Nn_\mathrm{t}}$. It is not hard to see that the lth detected symbol under error-free feedback from the decoder can be modeled as

$$\hat{x}_{l,k} = \mu_{l,k}x_{l,k} + \eta_{l,k}, \tag{6.3}$$

where $\mu_{l,k}$ is the mean, given by $\tilde{\mathbf{h}}_{l,k}(\sigma_x^2\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k + N_0\mathbf{I})^{-1}\tilde{\mathbf{h}}_{l,k}^H$, and $\eta_{l,k}$ is the noise component with distribution of $\mathcal{N}(0, \sigma_{\eta_{l,k}}^2)$, where $\sigma_{\eta_{l,k}}^2 = \mu_{l,k}(1-\mu_{l,k})$. It can be seen that as long as $\mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}$, $\mu_{l,k}$ and $\sigma_{l,k}^2$ for detection of data symbols can be replaced by $\mu_{l,t}$ and $\sigma_{l,t}^2$ for all indexes k corresponding to $t = \lfloor\frac{k-1}{(L/N)}\rfloor + 1$.

To consider the effect of imperfect CSI, the above MMSE detector can be adapted by modeling the channel estimation error as an additional SNR loss. This applies well for linear MMSE channel estimation since the interference introduced by

channel estimation error is independent of the data symbols [7]. Specifically, (6.2) and (6.3) are modified by replacing $\mathbf{H}_k \rightarrow \hat{\mathbf{H}}_k$ and $N_0 \rightarrow N_0 + n_{\mathrm{t}}\sigma_x^2\sigma_e^2$, where $\hat{\mathbf{H}}_k = \mathrm{diag}\Big\{\underbrace{\hat{\mathbf{H}}_t, \ldots, \hat{\mathbf{H}}_t}_{N},\Big\}$. Here, $\hat{\mathbf{H}}_t$ and σ_e^2 are the estimated channel and the variance of estimation error, respectively [8]. The estimation error is defined as $\mathbf{E}_t = \mathbf{H}_t - \hat{\mathbf{H}}_t$, where $\hat{\mathbf{H}}_t$ is obtained by an iterative channel estimator. The channel estimation process will be explained in detail in the next section.

6.3 Performance Analysis

6.3.1 PEP under Imperfect CSI

As explained earlier, the first P symbols in each L/N precoded symbols contain pilot symbols. As a result, the data rate for the first part is $(Nn_{\mathrm{t}} - N_{\mathrm{p}})/Nn_{\mathrm{t}}$ and that for the second part is one. In this situation, it is not straightforward to evaluate the total error performance since it is hard to find an interleaver to uniformly spread the error events over all $L \times n_c$ time periods (i.e., one codeword). Furthermore, finding a closed-form expression for the total error performance is not useful in terms of investigating the effect of the channel estimation error. Therefore, the PEP for two parts of the block, i.e., the hybrid part and the data part, should be assessed separately. The PEP for a full-rate linear precoder (i.e. the second part of the block length) has been completely analyzed in [9]. As such, the following analysis focuses on the PEP calculation corresponding to the first part.

The average PEP corresponding to the error event with Hamming weight d is denoted by $f(d, \Omega, \zeta)$, which depends on the constellation Ω and the mapping rule ζ. The function $f(d, \Omega, \zeta)$ is obtained by averaging the conditional PEP over the channel realization. Let $\mathbf{c}$ and $\breve{\mathbf{c}}$ be the transmitted and decoded codewords, respectively. Also let the corresponding transmitted and decoded symbol vectors be $\boldsymbol{\chi}$ and $\breve{\boldsymbol{\chi}}$, which differ in d positions. The PEP, conditioned on the channel realization $\tilde{\mathcal{H}}$ and the Hamming weight d can be expressed as $P(\mathbf{c} \rightarrow \breve{\mathbf{c}}|\tilde{\mathcal{H}}, d) = P(\boldsymbol{\chi} \rightarrow \breve{\boldsymbol{\chi}}|\tilde{\mathcal{H}}, d)$.

Knowing that a correct decision is taken when the LLR is positive, then, for a

given set of channel realizations $\tilde{\mathcal{H}}$, $P(\boldsymbol{\chi} \to \breve{\boldsymbol{\chi}}|\tilde{\mathcal{H}}, d) = P(\sum_{\iota=1}^{d} \mathrm{LLR}_\iota < 0)$. Thus, $P(\mathbf{c} \to \breve{\mathbf{c}}|d) = E_{\tilde{\mathcal{H}}}\left\{P(\sum_{\iota=1}^{d} \mathrm{LLR}_\iota < 0)\right\} = \int_{-\infty}^{0} p_{E_{\tilde{\mathcal{H}}}\{\sum_{\iota=1}^{d} \mathrm{LLR}_\iota\}}(x)dx$, where $E_{\tilde{\mathcal{H}}}\{\cdot\}$ indicates the expectation over $\tilde{\mathcal{H}}$.

The set $\tilde{\mathcal{H}}$ contains d channel realization vectors. Each of them with size Nn_{r} is used to transmit one symbol. Furthermore, for simplicity the error correcting code is selected such that $d \geq (Nn_{\mathrm{t}} - N_{\mathrm{p}})n_c$. By this assumption, the maximum number of independent channel realization vectors is $(Nn_{\mathrm{t}} - N_{\mathrm{p}}) \times n_c$ for an $(Nn_{\mathrm{t}} - N_{\mathrm{p}}) \times Nn_{\mathrm{r}}$ equivalent MIMO n_c-block-fading channel (i.e., n_c number of the equivalent correlated channel $\mathbf{G}^{(\mathrm{d})}\mathbf{H}_k$ used for sending only data symbols).

The d random variables, LLR_ι, can be grouped into $(Nn_{\mathrm{t}} - N_{\mathrm{p}})n_c$ independent groups. Specifically, let $\mathrm{LLR}_{i,l,t}$ be the log-likelihood ratio of the ith bit in the constellation symbol transmitted on the lth antenna of the tth block, $t = 1, \ldots, n_c$, $l = 1, \ldots, Nn_{\mathrm{t}} - N_{\mathrm{p}}$ and $i = 1, \ldots, \kappa_{l,t}$, where $\kappa_{l,t}$ is the number of bits transmitted on the lth antenna of the tth block. Then by definition $\sum_t \sum_l \kappa_{l,t} = d$ and it follows that $\mathrm{LLR} = \sum_t \sum_l \sum_i \mathrm{LLR}_{i,l,t}$ is the sum of $(Nn_{\mathrm{t}} - N_{\mathrm{p}})n_c$ independent random variables.

Let $\Delta_{i,l,t}$ denote the Euclidean distance between the two signal points $\chi_{i,l,t}$ and $\breve{\chi}_{i,l,t}$ that determine the calculation of $\mathrm{LLR}_{i,l,t}$. Then using (6.3) for the MMSE detector, the $\mathrm{LLR}_{i,l,t}$, conditioned on $\Delta_{i,l,t}^2$, is given by

$$\begin{aligned} \mathrm{LLR}_{i,l,t} &= \frac{|\hat{\chi}_{i,l,t} - \breve{\chi}_{i,l,t}|^2 - |\hat{\chi}_{i,l,t} - \chi_{i,l,t}|^2}{\sigma_{\eta_{l,t}}^2} \\ &= \frac{\mu_{l,t}^2 \Delta_{i,l,t}^2 + 2\mathrm{Re}(\Delta_{i,l,t}^2 \mu_{l,t} \eta_{l,t})}{\sigma_{\eta_{l,t}}^2} \end{aligned} \tag{6.4}$$

It should be pointed out that the distance spectrum $\{\Delta_{i,l,t}^2\}$ depends on the modulation type, its size and the mapping rule. It can be seen that, all $\mathrm{LLR}_{i,l,t}$ for a given t and l share the same equivalent channel vector $\tilde{\mathbf{h}}_{l,t}$ through parameters $\mu_{l,t}$ and $\sigma_{\eta_{l,t}}^2$. Defining $\Delta_{l,t}^2 = \sum_i \Delta_{i,l,t}^2$ and using the relation between $\mu_{l,t}$ and $\sigma_{\eta_{l,t}}^2$, one has $\mathrm{LLR}_{l,t}|\hat{\mathbf{H}}_t \sim \mathcal{N}\left(\frac{\mu_{l,t}}{1-\mu_{l,t}}\Delta_{l,t}^2, 2\frac{\mu_{l,t}}{1-\mu_{l,t}}\Delta_{l,t}^2\right)$ and $\mathrm{LLR}_t|\hat{\mathbf{H}}_t \sim \mathcal{N}\left(\sum_l \frac{\mu_{l,t}}{1-\mu_{l,t}}\Delta_{l,t}^2, \sum_l 2\frac{\mu_{l,t}}{1-\mu_{l,t}}\Delta_{l,t}^2\right)$. Therefore, the expectation over $\hat{\mathbf{H}}_t$ can be done instead by the expectation over the

random variable $\xi_t = \sum_l \frac{\mu_{l,t}}{1-\mu_{l,t}}\Delta_{l,t}^2$. The pdf of ξ_t can be obtained for the MMSE detector. For this detector, the variable ξ_t is equivalent to $\sum_l \Delta_{l,t}^2 \mathbf{G}_l^{(\mathrm{d})}\hat{\mathbf{H}}_k(\hat{\mathbf{H}}_k^H\hat{\mathbf{H}}_k + \tilde{N}_0/\sigma_x^2\mathbf{I})^{-1}\hat{\mathbf{H}}_k^H(\mathbf{G}_l^{(\mathrm{d})})^H$, where $\mathbf{G}_l^{(\mathrm{d})}$ is the lth row of $\mathbf{G}^{(\mathrm{d})}$. From the definition of $\hat{\mathbf{H}}_k$ it can be shown that

$$\begin{aligned}\xi_t =& \sum_l \Delta_{l,t}^2 \mathrm{trace}\left(\tilde{\mathbf{G}}_l^{(\mathrm{d})}\hat{\mathbf{H}}_t\left(\hat{\mathbf{H}}_t^H\hat{\mathbf{H}}_t + \frac{\tilde{N}_0}{\sigma_x^2}\mathbf{I}\right)^{-1}\hat{\mathbf{H}}_t^H(\tilde{\mathbf{G}}_l^{(\mathrm{d})})^H\right)\\ =& \mathrm{trace}\left(\sum_l \Delta_{l,t}^2\tilde{\mathbf{G}}_l^{(\mathrm{d})}\mathbf{U}\mathbf{\Lambda}_t\mathbf{U}^H(\tilde{\mathbf{G}}_l^{(\mathrm{d})})^H\right)\end{aligned} \tag{6.5}$$

In the above $\tilde{\mathbf{G}}_l^{(\mathrm{d})}$ indicates an $N \times n_{\mathrm{t}}$ matrix that is constructed from $\mathbf{G}_l^{(\mathrm{d})}$ as $\tilde{\mathbf{G}}_l^{(\mathrm{d})} = [\mathbf{G}_{(l,1:n_{\mathrm{t}})}^{(\mathrm{d})}; \ldots; \mathbf{G}_{(l,(N-1)n_{\mathrm{t}}+1:Nn_{\mathrm{t}})}^{(\mathrm{d})}]$. Note that $\hat{\mathbf{H}}_t\left(\hat{\mathbf{H}}_t^H\hat{\mathbf{H}}_t + \frac{\tilde{N}_0}{\sigma_x^2}\mathbf{I}\right)^{-1}\hat{\mathbf{H}}_t^H$ is interpreted as signal-to-interference-plus-noise ratio (SINR) at the output of MMSE detector in an un-precoded MIMO system [10]. Using eigenvalue decomposition, the eigenvalues of $\mathbf{\Lambda}_t$ can be approximated by a Gamma distribution with parameters (α, β), where [10]

$$\alpha = \frac{(n_{\mathrm{t}}-1)\rho^2}{\sigma^2} \tag{6.6}$$

$$\beta = \frac{1}{n_{\mathrm{r}}}\frac{(\mathrm{SNR}^e)^2\sigma^2}{\rho}. \tag{6.7}$$

In the above expressions, ρ, σ^2 and SNR^e are defined as [5, 10]

- $\rho = \frac{\kappa - \mathrm{SNR}^e(1-\gamma)+1}{2\mathrm{SNR}^e\gamma}$, where
  $\kappa = \left((\mathrm{SNR}^e)^2(1-\gamma)^2 + 2\mathrm{SNR}^e(1+\gamma)+1\right)^{1/2}$,
  and $\gamma \sim \frac{n_{\mathrm{t}}}{n_{\mathrm{r}}}$,
- $\sigma^2 = \rho - \frac{1+\mathrm{SNR}^e(1+\gamma)-\kappa}{2\mathrm{SNR}^e\gamma\kappa}$, and
- $\mathrm{SNR}^e = \frac{n_{\mathrm{r}}\sigma_x^2(\sigma_h^2-\sigma_e^2)}{N_0+n_{\mathrm{t}}\sigma_e^2\sigma_x^2}$ is the received SNR per transmit antenna for imperfect CSI.

Now let define $\mathbf{M}_t$ as an $n_{\mathrm{t}} \times n_{\mathrm{t}}$ Hermitian square-root matrix of $\Xi_t = \sum_{l=1}^{Nn_{\mathrm{t}}-N_{\mathrm{p}}} \Delta_{l,t}^2(\tilde{\mathbf{G}}_l^{(\mathrm{d})})^H\tilde{\mathbf{G}}_l^{(\mathrm{d})}$. Then $\xi_t = \mathrm{trace}\left(\mathbf{M}_t\mathbf{\Lambda}_t\mathbf{M}_t\right)$ and $\mathbf{M}_t\mathbf{\Lambda}_t\mathbf{M}_t$ is a random matrix that has a Wishart distribution with α degrees of freedom and parameter matrix Ξ_t [11]. By taking the expectation over ξ_t, the characteristic function of LLR_t

is given by

$$\begin{aligned}\psi_t(j\nu) &= E_{\xi_t}\left\{e^{\nu(j-\nu)\left(\mathrm{trace}\left(\mathbf{M}_t\mathbf{\Lambda}_t\mathbf{M}_t\right)\right)}\right\} \\ &= \prod_{u=1}^{n_t}\left(1-\nu(j-\nu)\beta\upsilon_{t,u}\right)^{-\alpha}, \end{aligned} \tag{6.8}$$

where $\upsilon_{t,u}$ is the uth eigenvalue of Ξ_t. Moreover, there are n_c independent $\psi_t(j\nu)$ and therefore,

$$\psi(j\nu) = \prod_{t=1}^{n_c}\prod_{u=1}^{n_t}\left(1-\nu(j-\nu)\beta\upsilon_{t,u}\right)^{-\alpha}. \tag{6.9}$$

Based on this characteristic function and using the partial fraction expansion and after some manipulations, the asymptotic expression of $f(d,\Omega,\zeta)$ can be approximated by

$$f(d,\Omega,\zeta)_{N_0\to 0} \approx \begin{pmatrix} 2\lfloor\alpha N_\delta\rfloor - 1 \\ \lfloor\alpha N_\delta\rfloor \end{pmatrix} \prod_{\vartheta=1}^{n_\delta}\left(\frac{1}{\beta\delta_\vartheta^2}\right)^{\lambda_\vartheta\alpha} \tag{6.10}$$

where $\{\delta_\vartheta^2\}$ is the set of n_δ distinct nonzero eigenvalues extracted from the sequence defined by $\upsilon_{t,u}$ values. Each δ_ϑ^2 is repeated λ_ϑ times and $N_\delta = \sum_{\vartheta=1}^{n_\delta}\lambda_\vartheta$ is the total number of eigenvalues. It should be pointed out that to use partial fraction expansion theorem, the integer value of the exponent $\sum_{\vartheta=1}^{n_\delta}\alpha\lambda_\vartheta = \alpha N_\delta$ was used in (6.10). Also the result obtained in (6.10) agrees with Eq. (24) of [4] in the case of perfect CSI (i.e., when $\sigma_e^2 = 0$), except that $\alpha = n_\mathrm{r}$ and $\beta = 1/N_0$ for ML detector.

The diversity associated with the considered Hamming weight d from (6.10) is $\sum_{\vartheta=1}^{n_\delta}\alpha\lambda_\vartheta = \alpha N_\delta$. The coding gain defined as the coefficient dividing N_0, which is given by $\left(\prod_{\vartheta=1}^{n_\delta}\left(N_0\beta\delta_\vartheta^2\right)^{\lambda_\vartheta}\right)^{\frac{1}{N_\delta}}$.

From the above equations, it is obvious that coding gain and diversity order depend on $\mathbf{G}^{(\mathrm{d})}$ through δ_ϑ^2 and N_δ, respectively. The coding gain and the diversity order are also functions of σ_e^2 through β and α, respectively. In addition, the diversity order depends on the number of pilot symbols by N_δ. In contrary to BICM-MIMO systems without precoding, σ_e^2 should also be itself a function of the precoder design in precoded BICM-MIMO systems. The reason is that the information of both *precoded* pilot and data symbols is used in each iteration of the iterative channel estimator. Therefore, for designing the precoder this relationship shall be explained in the next section.

6.3.2 MSE of the Channel Estimator

Without loss of generality, the following channel estimation process is explained for the first block (i.e., when $t = 1$). For the iterative linear MMSE channel estimation, the estimate of the channel, $\hat{\mathbf{H}}$, is obtained using a linear filter as $\mathbf{A}\breve{\mathbf{Y}}$, where $\mathbf{A} = E\left\{\hat{\mathbf{H}}(\breve{\mathbf{Y}})^H\right\} E\left\{\breve{\mathbf{Y}}(\breve{\mathbf{Y}})^H\right\}^{-1}$, $\breve{\mathbf{Y}}$ is the $NP \times n_{\mathrm{r}}$ observation matrix during P precoding time periods that is constructed from the vector $\mathbf{y}_1, \ldots, \mathbf{y}_P$ as $\breve{\mathbf{Y}} = [(\mathbf{y}_1^{[1]})^T, \ldots, (\mathbf{y}_1^{[N]})^T, \ldots, (\mathbf{y}_P^{[1]})^T, \ldots, (\mathbf{y}_P^{[N]})^T]^T$, where $\mathbf{y}_k^{[i]} = [y_{(k-1)Nn_{\mathrm{r}}+(i-1)n_{\mathrm{r}}+1}, \ldots, y_{(k-1)Nn_{\mathrm{r}}+in_{\mathrm{r}}}]$.

From (6.1), $\mathbf{y}_k^{[i]}$ can be expressed in terms of data and training symbols as

$$\mathbf{y}_k^{[i]} = \mathbf{x}_k \mathbf{G}_k^{[i]} \mathbf{H} + \mathbf{w}_k^{[i]} = \mathbf{x}_k^{\mathrm{p}} \mathbf{G}_k^{[i,\mathrm{p}]} \mathbf{H} + \mathbf{x}_k^{\mathrm{d}} \mathbf{G}^{[i,\mathrm{d}]} \mathbf{H} + \mathbf{w}_k^{[i]},$$

where for convenience, the notations $\mathbf{G}_k^{[i]}$ refers to the ith sub-rows of $\mathbf{G}_k$ with size $Nn_t \times n_t$, $i = 1 : N$, and $\mathbf{G}^{[i,\mathrm{d}]}$ and $\mathbf{G}_k^{[i,\mathrm{p}]}$ refer to the rows of $\mathbf{G}_k^{[i]}$ multiplied by data and pilot symbols, respectively. Here, the data and pilot symbols are denoted by $\mathbf{x}_k^{(\mathrm{d})}$ and $\mathbf{x}_k^{(\mathrm{p})}$, respectively. Then, when the data symbols are decoded correctly, the covariance matrix of the channel estimation error, denoted by $\mathbf{\Sigma}$, after some manipulations is obtained by

$$\mathbf{\Sigma} = \left(\sigma_h^{-2} \mathbf{I}_{n_{\mathrm{t}}} + \frac{P\sigma_x^2}{N_0} \sum_{i=1}^{N} (\mathbf{G}^{[i,\mathrm{d}]})^T (\mathbf{G}^{[i,\mathrm{d}]})^* + \frac{1}{N_0} \sum_{k=1}^{P} \sum_{i=1}^{N} (\mathbf{G}_k^{[i,\mathrm{p}]})^T (\mathbf{x}_k^{\mathrm{p}})^T (\mathbf{x}_k^{\mathrm{p}})^* (\mathbf{G}_k^{[i,\mathrm{p}]})^* \right)^{-1} \otimes \mathbf{I}_{n_{\mathrm{r}}} \tag{6.11}$$

Now σ_e^2 is the trace of $\mathbf{\Sigma}$ which is a function of the precoder matrix and pilot symbols.

6.4 Designing the Precoder

To design the precoder, one can try to maximize achievable diversity order and coding gain. However, it can be shown that they are maximized when all the eigenvalues δ_ϑ^2 are as equal as possible (i.e., $N_\delta = n_c n_{\mathrm{t}}$ and $\delta_\vartheta^2 = \delta^2$). On the other hand coding gain and diversity order are decreasing functions of the variance of estimation error, σ_e^2, which as well depends on precoder structure. Therefore, the designing problem

can be converted to jointly optimize $\mathbf{G}^{(\mathrm{d})}$ to make eigenvalues equal and $\mathbf{G}^{(\mathrm{d})}$, $\mathbf{G}_k^{(\mathrm{p})}$ and $\mathbf{x}_k^{(\mathrm{p})}$ to minimize σ_e^2. Moreover, all constraints can be summarized as,

$$\begin{cases} \frac{1}{P}\mathrm{trace}\sum_{k=1}^{P}\sum_{i=1}^{N}(\mathbf{G}_k^{[i,\mathrm{p}]})^T(\mathbf{x}_k^{(\mathrm{p})})^T(\mathbf{x}_k^{(\mathrm{p})})^*(\mathbf{G}_k^{[i,\mathrm{p}]})^* \leq P_t \\ {\mathbf{x}^{(\mathrm{p})}}_{j,k} \in \Omega; 1 \leq j \leq N_{\mathrm{p}} \\ \mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}_{Nn_{\mathrm{t}}} \end{cases}$$

In general two objective functions in the optimization problem may not be optimized with the same $\mathbf{G}^{(\mathrm{d})}$. However, in our model, it can be shown that if $\mathbf{G}^{(\mathrm{d})}$ is extracted from the DNA precoder proposed in [4], it would be optimum for both functions with some conditions explained in the following.

First, let's review the structure of the DNA precoder as

$$\mathbf{G}^{\mathrm{DNA}} = \begin{bmatrix} \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[1][1]} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[N][1]} \\ \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[1][2]} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[N][2]} \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[1][N^2]} & \cdots & \mathbf{I}_{n_{\mathrm{t}}/N} \otimes \mathbf{\Phi}^{[N][N^2]} \end{bmatrix} \tag{6.12}$$

where $\mathbf{\Phi}$ is an $N^2 \times N^2$ rotator matrix with orthogonal sub-rows (i.e., $\mathbf{\Phi}^{[i][j]} \perp \mathbf{\Phi}^{[i'][j']}; \forall i \neq i' \ \& \ j = j'$). $\mathbf{\Phi}^{[i][j]}$ is the ith sub-row of the jth row of $\mathbf{\Phi}$ with size N and $\otimes$ denotes the Kronecker product.

Now let's explain the conditions. It can be seen from (6.11) that σ_e^2 has its minimum when the second and the third terms are scaled identity matrices, i.e., $\sum_{i=1}^{N}(\mathbf{G}^{[i,\mathrm{d}]})^T(\mathbf{G}^{[i,\mathrm{d}]})^* = \varrho\mathbf{I}_{n_{\mathrm{t}}}$ and $\sum_{k=1}^{P}\sum_{i=1}^{N}(\mathbf{G}_k^{[i,\mathrm{p}]})^T(\mathbf{x}_k^{(\mathrm{p})})^T(\mathbf{x}_k^{(\mathrm{p})})^*(\mathbf{G}_k^{[i,\mathrm{p}]})^* = \nabla\mathbf{I}_{n_{\mathrm{t}}}$. Interestingly, selecting $Nn_{\mathrm{t}} - N_{\mathrm{p}}$ rows of the $Nn_{\mathrm{t}} \times Nn_{\mathrm{t}}$ DNA precoder matrix, satisfies the first constraint as $\sum_{i=1}^{N}(\mathbf{G}^{[i,\mathrm{d}]})^T(\mathbf{G}^{[i,\mathrm{d}]})^* = (N - (N_{\mathrm{p}}/n_{\mathrm{t}}))\mathbf{I}_{n_{\mathrm{t}}}$, if N_{p} divided by n_{t}/N. Then to maximize the diversity order and coding gain N_{p} should have its minimum value, n_{t}/N. This choice also minimizes the training overhead.

The remaining design parameters are $\mathbf{G}_k^{(\mathrm{p})}, \mathbf{x}_k^{(\mathrm{p})}$ and P such that they minimize σ_e^2 subject to the remaining constraints. One selection for $\mathbf{x}_k^{(\mathrm{p})}$ would be repeating

the same symbol $N_{\rm p} = n_{\rm t}/N$ times and then utilizing it for all P precoded symbols. Then, $(\mathbf{x}^{(\rm p)})^T(\mathbf{x}^{(\rm p)})^* = |x^{(\rm p)}|^2\mathbf{1}_{n_{\rm t}/N}$, where $\mathbf{1}_{n_{\rm t}/N}$ is an $n_{\rm t}/N \times n_{\rm t}/N$ matrix in which all components have unit values and $|x^{(\rm p)}|^2$ is the power of selected pilot symbol. By this selection for $\mathbf{x}^{(\rm p)}$ and considering the structure of matrix $\mathbf{G}^{(\rm d)}$, it is suggested that the block diagonal form for $\mathbf{G}_k^{[i,\rm p]}$ is

$$\mathbf{G}_k^{[i,\rm p]} = \text{diag}\left[\mathbf{\Theta}_k^{[i][1]}, \mathbf{\Theta}_k^{[i][2]}, \ldots, \mathbf{\Theta}_k^{[i][n_{\rm t}/N]}\right] \tag{6.13}$$

where $\mathbf{\Theta}$ is an $P \times Nn_{\rm t}$ matrix that should be designed and $\mathbf{\Theta}_k$ is the kth row of $\mathbf{\Theta}$. Here, $\mathbf{\Theta}_k^{[i][j]}$ is the jth sub-row with size N of $\mathbf{\Theta}_k^{[i]}$, where $\mathbf{\Theta}_k^{[i]}$ is the ith sub-row of $\mathbf{\Theta}_k$ with size $n_{\rm t}$. Then,

$$\begin{aligned}\sum_{k=1}^{P}\sum_{i=1}^{N}(\mathbf{G}_k^{[i,\rm p]})^T(\mathbf{x}_k^{(\rm p)})^T(\mathbf{x}_k^{(\rm p)})^*(\mathbf{G}_k^{[i,\rm p]})^* \quad &= \\ |x_p|^2\sum_{k=1}^{P}\sum_{i=1}^{N}(\mathbf{\Theta}_k^{[i]})^T(\mathbf{\Theta}_k^{[i]})^* \quad &\end{aligned} \tag{6.14}$$

It can be seen that for the above equation to be a scaled identity matrix, P should be selected such that $P \geq n_{\rm t}/N$. This result matches with the fact that for obtaining the CSI at the receiver at the first iteration, spanning all antennas with training is necessary, since there is no information about data at the first iteration. The expression in (6.14) is an identity matrix if $\mathbf{\Theta}$ is row-orthogonal and $\mathbf{\Theta}_k \neq \mathbf{\Theta}_{k'}; \forall k, k'$. The other constraint which is $\mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}$ satisfies if

$$\mathbf{\Theta}_k^{[i][j]} \quad \perp \quad \mathbf{\Phi}^{[i'][j']}; \forall k \;\&\; i \neq i' \;\&\; j = j' \tag{6.15}$$

The above conditions are meet when $\mathbf{\Theta}_k^{[i][j]} = \exp\left(2\pi\sqrt{-1}(j-1)(k-1)(N/n_t)\right)\mathbf{\Phi}^{[i][n]}; i = 1:N, j = 1:n_t/N, k = 1:n_t/N$, where n is the index of the row of $\mathbf{\Phi}$ which is not used in $\mathbf{G}^{(\rm d)}$.

In summary the parameters are designed as

- $N_{\rm p} = n_{\rm t}/N$.
- $P = n_{\rm t}/N$.

Although the proposed design attains the minimum on σ_e^2, the optimal design may not be unique.

- $\mathbf{G}^{(\mathrm{d})} = \mathbf{G}^{\mathrm{DNA}}_{\frac{n_t}{N}+1:Nn_t}$.
- $\mathbf{G}_k^{(\mathrm{p})}$: (6.15) and (6.13) are used to provide $\mathbf{G}_k^{(\mathrm{p})}$. It is noted that for $N = n_t$, $\mathbf{G}_k = \mathbf{G}^{\mathrm{DNA}}$.

6.5 Illustrative Results

In this section, the FER and MSE performances of precoded BICM-MIMO systems using the proposed precoder and the training sequence are presented. We consider quadrature phase-shift keying (QPSK) modulation with Gray mapping. Note that, to determine the FER the errors in the whole codeword (including hybrid and pure symbols) are counted.

The FER and MSE performance of a BICM-MIMO for a codeword length of 4×1024 bits are shown for a 4×2 block-fading MIMO channels in Fig. 6.3 and Fig. 6.4, respectively. In these figures, $n_c = 2$ and E_b is the energy per information bit. The code used is the 16-state convolutional code with generator polynomials $(23, 35)$ in octal form. In Fig. 6.3, the FER curves are obtained after 5 iterations of the iterative channel estimation/demodulation/decoding and when $N = 2$. To verify the results obtained in Section 6.3, the FER curves of the following four systems are compared. For a fair comparison, in the first three cases, $P = 2$.

1. The system with the proposed precoder and training sequence, $N_{\mathrm{p}} = n_{\mathrm{t}}/N = 2$.
2. The system using the full DNA precoder with the training sequence proposed in [3]. This system has its optimum performance when $N_{\mathrm{p}} = 4$ pilot symbols are inserted equi-spaced in each group of 12 data symbols.
3. The system using the full DNA precoder and with the same power budget for the training symbols as in the first system, i.e., with $N_{\mathrm{p}} = 2$.
4. The system with perfect CSI.

From the results presented in Fig. 6.3, it is clear that the proposed precoder

outperforms the DNA precoder. More importantly, while the proposed scheme uses a little bandwidth for training information (for the system considered in this figure the training overhead is $PN_{\mathrm{p}} = 4$), the training overhead with the DNA precoder is $2 \times 4 = 8$, which is double. The improvement is even more impressive if considering the fact that the training power for using the proposed precoder is half of that using DNA precoder. The improvement in FER is about 0.25 dB at the FER of 10^{-3}. Furthermore, by comparing the FER curves of the first and the third systems, one can see that for the same training overhead of $N_{\mathrm{p}} = 2$, for both the proposed precoder and the DNA precoder, the improvement is about 1.5 dB at FER of 10^{-2}.

Similar results are also observed in Fig. 6.4 for the MSE curves. The system parameters are the same as those used for Fig. 6.3. It can be seen that for the same amount of training information and power ($N_p = 2$ for both the proposed precoder and the DNA precoder), the performance improvement by our proposed design is about 2 dB at the MSE of 10^{-3}.

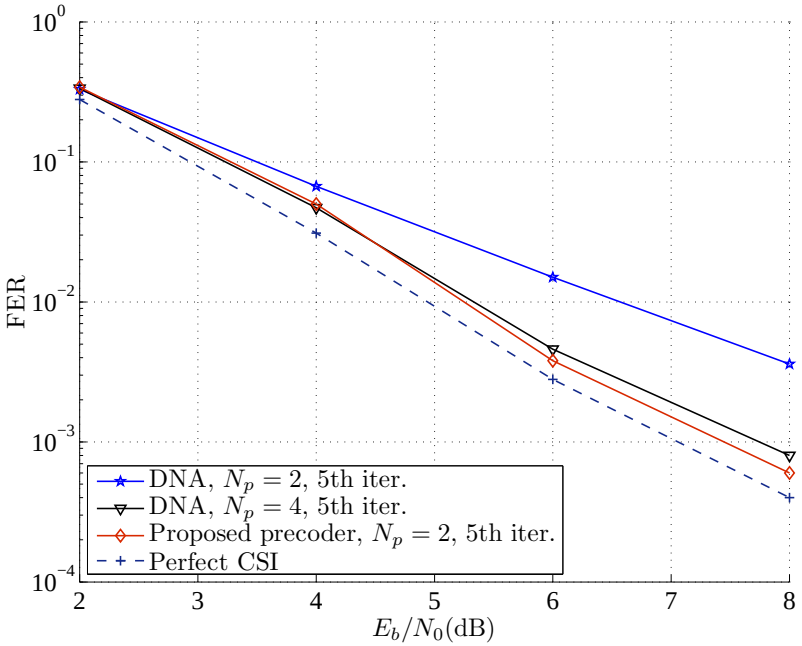

Figure 6.3 FER performances obtained with the proposed precoder and DNA precoder over a 4×2 block-fading channel, when $N = 2$.

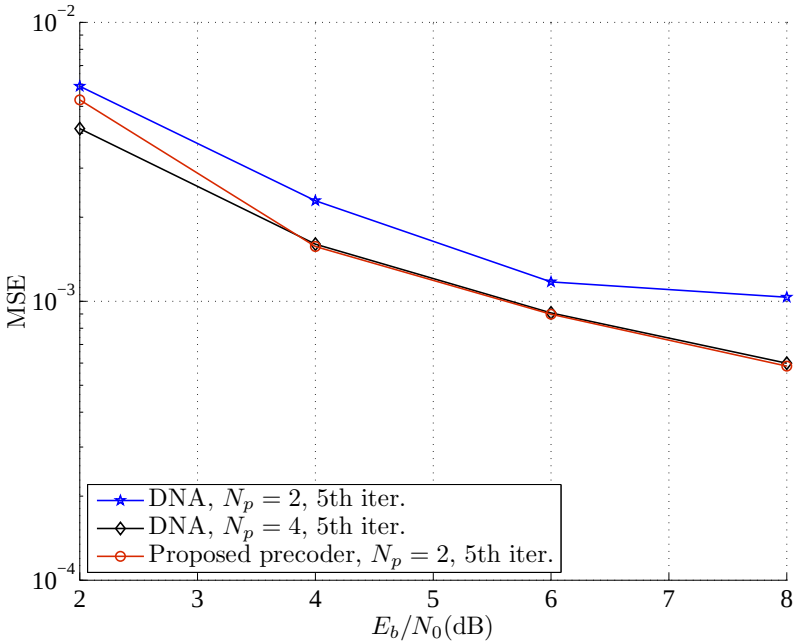

Figure 6.4 MSE performances obtained with the proposed precoder and DNA precoder over a 4×2 block-fading channel, when $N = 2$.

6.6 Conclusion

In this paper, a linear precoder for a BICM-MIMO system is devised that takes into account channel estimation error. The design uses the same structure of DNA precoders for precoding data symbols. However, for the pilot symbols, the subparts of the precoder assigned to precode pilot symbols change over time. The pilot symbols are also designed in order to effectively assist the proposed precoder. In order to design the precoder and the training sequence, the coding gain and diversity order under imperfect CSI are maximized. Compared to the DNA precoder, performance improvement achieved with the proposed precoder is about 1.5 dB at a FER level of 10^{-2} for the same power budget on training. Moreover, for a 4×2 MIMO channel, the amount of signalling overhead that can be eliminated for transmitting data symbols is 4 symbols, while maintaining the same FER performance.

References

[1] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1403–1418, May 2004.

[2] G. Taricco, "Optimum receiver design and performance analysis of arbitrarily correlated Rician fading MIMO channels with imperfect channel state information," *IEEE Trans. Inform. Theory*, vol. 56, pp. 1114–1134, Mar. 2010.

[3] Z. Andalibi, H. H. Nguyen, and J. E. Salt, "Training design for precoded BICM-MIMO systems in block-fading channels," *EURASIP J. Wireless Commun. and Networking*, vol. 80, Mar. 2012.

[4] N. Gresset, L. Brunel, and J. Boutros, "Spacetime coding techniques with bit-interleaved coded modulations for MIMO block-fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2156–2178, May 2008.

[5] I.-W. Lai, G. Ascheid, H. Meyr, and T.-D. Chiueh, "Asymptotic BER analysis for MIMO-BICM with MMSE detection and channel estimation," *Proc. IEEE Int. Conf. Commun.*, pp. 1–5, June 2011.

[6] N. Gresset, *New Space-Time Coding Techniques With Bit-Interleaved Coded Modulations*. PhD thesis, ENST, Paris, France, Dec 2004.

[7] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.

[8] M. Coldrey and P. Bohlin, "Training-based MIMO systems, Part I: Performance comparison," *IEEE Trans. Signal Processing*, vol. 55, pp. 5464–5476, Nov. 2007.

[9] Z. Andalibi, H. H. Nguyen, and J. E. Salt, "Asymptotic performance analysis of precoded BICM-MIMO under channel estimation errors," *Proc. IEEE Int. Conf. Commun.*, June 2012.

[10] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inform. Theory*, vol. 52, pp. 271–286, Jan. 2006.

[11] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proc. Radar, Sonar and Navig.*, vol. 147, pp. 162–168, Aug. 2000.

7. Summary and Suggestions for Further Study

7.1 Summary

The main contribution of this thesis was to develop iterative channel estimation for BICM systems incorporating different diversity techniques and under two distinct types of channel environments, namely slow fading and fast fading. The thesis also focused on exploiting different diversity techniques to mitigate the effect of imperfect CSI. Specifically, the main contributions of this thesis are summarized as follows:

- A novel iterative channel estimation technique has been developed for BICM-ID systems. Signal space diversity has been exploited to construct a new training sequence in Chapter 3. Using a rotator which implements signal space diversity before the transmission spreads the training sequence over the time duration of a rotated symbol. This facilitates the channel estimator to track the variation of the fading channel over time. An upper-bound of the BER expression has also been derived. Simulation results with different training schemes revealed that performance improvement by SSD-pilot is realized through successive iterations.

- In the second contribution, space diversity has been considered in block-fading channels to examine the tradeoff between exploiting diversity techniques and the effect of channel estimation error on BICM systems. By increasing the diversity order (i.e., by increasing the number of transmit antennas), interference from other antennas on the received signal increases and the channel estimation error makes the convergence process of an iterative receiver more challenging. The convergence problem due to channel estimation error has been solved in

Chapter 4 by working out conditions on the estimation iteration versus decoding iteration. Better performance improvement has been obtained by combining training and data information at each estimation iteration.

- For the third contribution, a space-time coding technique has been implemented using a DNA linear precoder in Chapter 5. By spreading data symbols over time and space using a linear precoder, training symbols have been suggested to be used with the precoder. The training pattern in this new scheme has been designed by minimizing the CRB. The proposed training pattern uses less time and space for sending data. Furthermore, the channel estimator has been developed to work with this new training scheme in an iterative manner.

- In the last contribution, a further improvement in the system error rate of precoded BICM-MIMO, as well as bandwidth and power allocation of the training sequence have been provided by designing joint precoder and training scheme. In Chapter 6, the PEP of the system and the MSE of the channel estimator have been derived and used as the criteria to design joint training sequence and precoder. Performance comparison revealed that employing joint optimized precoder and training sequence leads to a considerable performance improvement of BICM-MIMO systems with a small power and signalling overhead allocated to the training sequence.

7.2 Suggestions for Further Studies

The work done in this thesis might constitute a significant step toward the design of channel estimators and training sequences for systems exploiting various diversity techniques under different channel environments. While conducting our research works, several issues arose that should be interesting for further studies. These issues are elaborated next.

- In Chapters 3, 5 and 6, precoders for improving diversity and coding gain have been adopted. Then training sequence has been designed by exploiting pre-

coders such that pilot symbols are precoded with data symbols. By inserting pilot symbols before the precoder, the demodulator uses the information of pilot symbols to facilitate the convergence and to increase the reliability of the extrinsic soft information. Further improvement might be realized by inserting pilot bits before the channel encoder. By this structure, the decoder can also exploit the information of pilot bits in the same way of the demodulator. However, the main purpose of using pilot bits is to estimate the channel. Therefore, channel encoder and bit-interleaver should be designed carefully such that pilot bits after encoding and interleaving can still be distinguishable by the channel estimator and the demodulator.

- The pilot symbols in Chapters 5 and 6 have been designed for estimating block-fading channels in BICM-MIMO systems. The pilot symbols are inserted at the beginning of each block independently. However, in fast time-varying fading channels, to track the variation of CSI, new design for training sequence with intended correlation among pilot symbols is useful.

- As presented in this thesis, our work concentrated on channel estimation and the use of diversity techniques in frequency-flat fading channels. In frequency-selective fading channels, one can combine frequency diversity with coded modulation systems to combat fading and improve the system performance. To estimate the channel efficiently with less training overhead and reasonable performance, implementation of this diversity technique can be modeled by a precoder (i.e., a frequency precoder) in single-antenna systems and with space-frequency precoder in MIMO systems. With this model, one can follow the same procedure of Chapter 3 for constructing an iterative channel estimation and designing training sequence to track the CSI in frequency and time domains.

- Currently, much research efforts have focused on relaying techniques and cooperative networks as new implementations of diversity techniques. Optimal training design and placement in these implementations are of interest and worthwhile to study.

A. Copyright Permission

This appendix documents IET's permission to reproduce the manuscripts in Chapters 3 and 4 of this thesis. The manuscript in Chapter 5 is published by an open-access journal and the copyright rests with the authors.

From: Bond,Anna [ABond@theiet.org]
Sent: 23 July 2012 06:53
To: Andalibi, Zohreh
Subject: RE: Re:COM-2010-0858.R1 and COM-2010-0047.R1 permission request

Dear Zohreh Andalibi,

Permission to reproduce as requested is given, provided that the source of the material including the author, title, date, and publisher is acknowledged.

A reproduction fee is not due to the IET on this occasion because you are the author and it is for educational purposes.

Thank you for your enquiry. Please do not hesitate to contact me should you require any further information.

Kind regards,

Anna Bond
Editorial Assistant (Letters)
The IET
www.theiet.org

T: +44 (0)1438 767269

F: +44 (0)1438 767317

The Institution of Engineering and Technology, Michael Faraday House, Six Hills Way, Stevenage, SG1 2AY, United Kingdom

——Original Message——

From: Rowley,Paul

Sent: Monday, July 23, 2012 1:45 PM

To: Bond,Anna

Subject: FW: Re:COM-2010-0858.R1 and COM-2010-0047.R1 permission request

Anna,

Please see the below permission request for 2 papers in IET Communications.

The first paper is due to be published in issue 11, this will probably go to press around mid-August.

Thanks

Paul

——Original Message——

From: z.andalibi@usask.ca

Sent: Friday, July 20, 2012 2:21 PM

To: Rowley,Paul

Cc: Ha.Nguyen@usask.ca

Subject: Re:COM-2010-0858.R1 and COM-2010-0047.R1

Dear Dr. Paul Rowley

COM-2010-0858.R1 and COM-2010-0047.R1

I am Zohreh Andalibi, the first author of the following two papers, "Analyzing BICM-

MIMO Systems with Channel Estimation Error", to appear in IET Communications.
and
"Channel Estimation in Bit Interleaved Coded Modulation with Iterative Decoding", IET Communications, vol. 4, pp. 2095-2103, Nov. 2010.

Currently, I am writing my thesis. Would you please give me the permission to include these two papers in my thesis.

Thank you in advance for your time and consideration,

Best regards,

Zohreh Andalibi