

DESIGN AND ANALYSIS OF ITERATIVELY DECODABLE CODES
FOR ISI CHANNELS

A Dissertation

by

DŨNG NGOC DOAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Electrical Engineering

DESIGN AND ANALYSIS OF ITERATIVELY DECODABLE CODES
FOR ISI CHANNELS

A Dissertation

by

DŨNG NGOC DOAN

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Krishna R. Narayanan
(Chair of Committee)

Scott L. Miller
(Member)

Narasimha Reddy
(Member)

Andreas Klappenecker
(Member)

Chanan Singh
(Head of Department)

August 2004

Major Subject: Electrical Engineering

ABSTRACT

Design and Analysis of Iteratively Decodable Codes

for ISI Channels. (August 2004)

Dũng Ngoc Doan, B.E., Hanoi University of Technology;

M.E., Asian Institute of Technology

Chair of Advisory Committee: Dr. Krishna R. Narayanan

Recent advancements in iterative processing have allowed communication systems to perform close to capacity limits with manageable complexity. For many channels such as the AWGN and flat fading channels, codes that perform only a fraction of a dB from the capacity have been designed in the literature. In this dissertation, we will focus on the design and analysis of near-capacity achieving codes for another important class of channels, namely inter-symbol interference (ISI) channels. We propose various coding schemes such as low-density parity-check (LDPC) codes, parallel and serial concatenations for ISI channels when there is no spectral shaping used at the transmitter. The design and analysis techniques use the idea of extrinsic information transfer (EXIT) function matching and provide insights into the performance of different codes and receiver structures. We then present a coding scheme which is the concatenation of an LDPC code with a spectral shaping block code designed to be matched to the channel's spectrum. We will discuss how to design the shaping code and the outer LDPC code. We will show that spectral shaping matched codes can be used for the parallel concatenation to achieve near capacity performance. We will also discuss the capacity of multiple antenna ISI channels. We study the effects of transmitter and receiver diversities and noisy channel state information on channel capacity.

To my parents and my family

ACKNOWLEDGMENTS

I would like to thank my friends from the Wireless Communications Laboratory for fruitful discussions and help during my stay in College Station.

Special thanks go to professors in the Wireless Communications group for their excellent teaching. I have learned much more from both course work and the process of doing research. I would like to thank Professor C. Georghiades, Director of the WCL, for his insightful advice. I also would like to thank Mrs. Tammy and Mrs. Sonny for their wonderful administrative services.

I would like to thank my committee members, Dr. Scott Miller, Dr. Narasimha Reddy and Dr. Andreas Klappenecker, for their constructive comments and advice on my research.

My master's program adviser at AIT, Dr. Nandan Rajatheva, deserves lots of credit. He introduced me to the coding theory, especially turbo codes.

I would like to sincerely thank my Ph.D. adviser, Dr. Krishna Narayanan. He has patiently helped me in many ways, from forming an idea to making plans for my future. I admire his kindness and intelligence.

Many thanks go to my financial supporters: NSF and FTP, from which my Ph.D. adviser has generously provided me support for the entire time of my study at TAMU.

Heartfelt thanks go to Aude Esperbe for her constant love, encouragement and support. She has made my life more durable and fulfilled.

My parents deserve the most credit for their nurturing, love and patience. It would be impossible to follow this long term of study without the support of my family.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION AND ORGANIZATION OF THE DISSERTATION	1
	A. Introduction	1
	B. Organization of the Dissertation	2
II	BACKGROUND	4
	A. ISI Channel and Capacity	4
	1. ISI Channel	4
	2. Channel Capacity	4
	a. Unconstrained Inputs	5
	b. Constrained Inputs	6
	c. i.i.d. Capacity	7
	d. Tight Lower Bound on Channel Capacity	8
	B. Channel Coding	9
	1. Turbo Codes	9
	2. Serial Concatenation	9
	3. LDPC Codes	10
	C. Soft-Input Soft-Output Decoding	11
	D. Transfer of Soft Information in Iterative Equalization and Decoding	13
	E. EXIT Functions	14
	1. Properties of EXIT Functions	15
	2. Using EXIT Functions for Code Design	16
III	CONVERGENCE OF TURBO EQUALIZATION FOR ISI CHANNELS	17
	A. Introduction	17
	B. Precoded ISI Channels	19
	C. Measures of Performance	22
	D. Channel Mixture	26
	1. Code Design Based on Expected Extrinsic LLR Transfer Diagram	26
	a. Optimizing the Mixture	27

CHAPTER	Page
2. Code Design Based on Extrinsic Information Transfer (EXIT) Functions	29
E. Design of LDPC Codes	33
1. Using EXIT Functions for LDPC Code Design	33
a. Use of LDPC Codes Optimum for the AWGN Channel	34
b. Rate Loss due to Non-iterative Equalization	35
c. Rate Loss due to Sub-optimal Equalizers	36
F. Conclusion	37
IV CODE DESIGN AND ANALYSIS WITH SPECTRAL SHAPING	38
A. Introduction	38
B. Serial Concatenation	41
1. Block Inner Codes for Spectral Shaping	42
2. Threshold	49
3. Trellis Complexity	50
C. Parallel Concatenation	51
1. Simulation Results	56
D. Conclusion	57
V ITERATIVE PACKET COMBINING SCHEMES FOR INTER-SYMBOL INTERFERENCE CHANNELS	59
A. Introduction	59
B. System Model	62
C. Combining Techniques	63
1. ML Combining (MLC)	64
2. Iterative Combining (IC)	65
D. Distance-Spectrum Based Properties of Precoded ISI Channels	66
1. Recursive Property	66
2. Frame Error Rate Property	68
3. Interleaving Gain Property for Non-fading Precoded ISI Channels	69
4. Interleaving Gain Property for Fading Precoded ISI Channels	73
E. Results	75
F. Conclusion	80

CHAPTER	Page
VI	
CAPACITY OF MULTIPLE ANTENNA ISI CHANNELS AND SUMMARY	82
A. Introduction	82
B. Channel Model	84
C. Perfect CSI at Transmitter and Receiver	86
1. MISO Discrete-Time ISI System	87
2. SIMO Discrete-Time ISI System	93
3. MIMO Discrete-Time ISI System	96
D. Information Rate for Nearest Neighbor Decoding	96
1. Estimated CSI	97
2. Decoding Metric	97
3. Generalized Mutual Information	99
a. Without Deterministic Channel Estimation Error	103
b. Without Random Channel Estimation Error . . .	107
E. Conclusion	108
REFERENCES	109
APPENDIX A	117
APPENDIX B	126
VITA	128

LIST OF TABLES

TABLE	Page
I State transition for the combined trellis code and the $1-D$ channel with good and bad input mappings	55

LIST OF FIGURES

FIGURE	Page
1	A rate 1/3 turbo encoder. 9
2	A serial concatenation. 10
3	The bipartite graph of an LDPC encoder. 11
4	System model of a serial concatenation with iterative equalization. . . 13
5	System model for the serial concatenation of a outer code and an ISI channel 19
6	Concatenation of a rate-1 recursive inner coder (precoder) and an ISI channel. 20
7	The equivalent precoded ISI channel, where $M = \max(L, J)$ 20
8	Loss of mutual information at the first iteration: Plot of $I_{E,o}^{(1)}$ as a function of E_b/\mathcal{N}_0 for the 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel. 25
9	Density evolution of the 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and the $(1 + D + D^2, 1 + D^2)$ outer convolutional code; the precoding polynomial $h(D) = 1 \oplus D$ 29
10	Channel mixture at 1.2 dB: 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and $(1 + D + D^2, 1 + D^2)$ outer convolutional code, precoding poly- nomial $h(D) = 1 \oplus D$ 31
11	Channel mixture at 1.3 dB: 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and $(1 + D + D^2, 1 + D^2)$ outer convolutional code, precoding poly- nomial $h(D) = 1 \oplus D$ 32
12	EXIT chart for an ISI channel and an LDPC code designed based on EXIT function matching. 35
13	EXIT functions for different sub-optimal equalizers. 36

FIGURE	Page
14	System model for the serially concatenated scheme. 41
15	Capacity for the dicode channel. 45
16	EXIT functions for the concatenation of biphasic and dicode for different values of E_b/\mathcal{N}_0 46
17	Capacity loss due to the use of sub-optimal input PSD $P_X(e^{jw})$ as opposed to the optimal water-filling PSD $P_{X,\text{opt}}(e^{jw})$. The first 2 codewords are 6 and 9 (in decimal representation); the third codeword is numbered near each curve; and the fourth codeword is the complement of the third codeword. 48
18	Capacity for the $0.38 - 0.6D - 0.6D^2 + 0.38D^3$ ISI channel. 49
19	The parallel concatenation for the ISI channel. 51
20	EXIT functions of the rate half Pusch and Weinrichter's code used in the parallel concatenated scheme of rate 1/4 over the $0.8+0.6D$ ISI channel. 53
21	Different input mappings for $f_0 + f_1D$ channels when $\text{sign}(f_0f_1) = -1$ (dashed lines: input bit is equal to 1; solid lines: input bit is equal to 0) (a) Good mapping; (b) Bad mapping. 54
22	EXIT functions of the parallel concatenation for the $1 - D$ channel. 56
23	Bit error rate of the new parallel concatenation for the $1 - D$ channel. 57
24	Retransmissions of the same data packet over the channel can be viewed as a parallel concatenation of K trellis codes. 62
25	Frame error rate for the (0.89443, 0.44721) ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$ 76
26	FER bounds for the (0.89443, 0.44721) ISI channel; 'uniform' interleaver of sizes $N = 500$ and $N = 2000$; precoding polynomial $h(D) = 1 \oplus D$; numbers of transmissions $K = 2$ and $K = 3$ 77

FIGURE	Page
27	Frame error rate for the (0.5, 0.7071, 0.5) ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$ 78
28	Frame error rate for the (0.67082, 0.5, 0.3873, 0.31623, 0.22361) ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$ 79
29	Coding gains for the (0.5, 0.7071, 0.5) ISI channel; interleaver sizes $N = 500$ and $N = 2000$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$ 80
30	Frame error rate for the (0.5, 0.7071, 0.5) ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; mismatched channel at receiver $\hat{\mathbf{f}} = (0.42814, 0.79556, 0.42814)$; number of transmissions $K = 2$ 81
31	Trellis of 2-tap (f_0, f_1) channels; for the precoded channel, the precoder polynomial $h(D) = 1 \oplus D$; $a = f_0 + f_1$; $b = f_0 - f_1$ 117
32	Plot of $R_{1 0}$ and $R_{0 1}$ as functions of E_b/\mathcal{N}_0 for various 2-tap ISI channels. 125

CHAPTER I

INTRODUCTION AND ORGANIZATION OF THE DISSERTATION

A. Introduction

A typical communication system consists of a transmitter, a transmission medium (or channel), and a receiver. Different designs for the transmitter and the receiver are necessary, depending on channel characteristics, system complexity, and available resources. For the last 50 years, ever since Shannon presented the landmark paper on communication theory, many practical systems have been realized. The discovery of turbo codes by Berrou *et al.* [1] and the rediscovery of low-density parity-check (LDPC) codes [2] have significant impacts on the way a modern communication system is built.

Recent advancements in iterative (turbo) processing have allowed communication systems to perform close to capacity limits with manageable complexity. For many channels such as the AWGN and flat fading channels, codes that perform only a fraction of a dB from the capacity have been designed in the literature. In this dissertation, we will focus on the design and analysis of near-capacity achieving codes for another important class of channels, namely inter-symbol interference (ISI) channels.

ISI results from non-Nyquist Tx/Rx filters, transmission of wide band signals over a band limited channel, over sampling at higher than Nyquist rate, or due to non-ideal frequency response of the channel. ISI can be understood as the received signal at the channel output being dependent on signals at past, current, and possibly future time instants. ISI causes distortion to the transmitted signals, which is undesirable in most applications. ISI channels are often encountered in many practical systems

The journal model is *IEEE Transactions on Automatic Control*.

such as cellular networks, code division multiple access (CDMA) systems with large spreading gain, wireless LAN with high data rate, and multi-track magnetic recording systems.

For reliable communication through ISI channels, different methods can be used to combat the detrimental effect of ISI. Among them, spectrum shaping, channel equalization, and channel coding are the most popular methods used. In this dissertation, we focus on coding for ISI channels. We extensively use iterative principles for the design and analysis of iteratively decodable codes. We will give motivations of our research in each individual chapter.

This chapter is intentionally made brief to give a general picture about problems discussed in this dissertation. More detailed introductions can be found in subsequent chapters.

B. Organization of the Dissertation

The dissertation is organized as follows.

In Chapter II, we provide some relevant background, which is helpful for the exposition of subsequent chapters.

In Chapter III, we study code design and analysis for ISI channels without spectral shaping, and in the context of serial concatenation. We study convergence of iterative decoding/equalization using extrinsic information transfer (EXIT) functions. We present new results pertaining to the convergence of precoded ISI channels, and use these to design convolutional codes for ISI channels. We also design LDPC codes for these channels using EXIT functions. Our technique will be shown to provide much insight into code design and analysis.

In Chapter IV, we propose low complexity coding schemes for ISI channels which

perform close to the channel capacity by using spectral shaping techniques. We use codes that are spectrally matched to the channel for serial and parallel concatenations. Analysis using EXIT functions and BER simulations shows that both schemes surpass the independent identically distributed (i.i.d.) channel capacity and outperform other existing schemes.

In Chapter V, we extend results in Chapter III to automatic retransmission request (ARQ) schemes. We propose a novel combining scheme, namely iterative combining, which is shown to significantly improve system performance in terms of both frame error rate and convergence.

In Chapter VI, we consider the problem of finding the capacity for multiple antenna ISI channels with and without perfect channel state information. We study the effects on channel capacity of channel estimation errors. We will discuss possible capacity-achieving coding schemes for multiple antenna ISI channels.

CHAPTER II

BACKGROUND

In this chapter, we present some background on ISI channels, channel capacity, various types of coding and decoding schemes, and techniques to analyze them.

A. ISI Channel and Capacity

1. ISI Channel

The output of a discrete-time ISI channel, denoted by y_n , can be expressed as

$$y_n = \sum_{\ell=0}^L f_{\ell} \cdot x_{n-\ell} + z_n = r_n + z_n, \quad 0 \leq n < N, \quad (2.1)$$

where $x_0^{N-1} = (x_0, x_1, \dots, x_{N-1})$ is the channel input sequence of length N , r_n is the noiseless channel output, $f_0^L = (f_0, f_1, \dots, f_L)$ is the channel state information (CSI), and $z_0^{N-1} = (z_0, z_1, \dots, z_{N-1})$ is a sequence of additive white Gaussian noise (AWGN) samples with zero mean and variance σ_z^2 . The channel output at time n , y_n , depends not only on the useful input signal at time n , x_n , but also on other interfering signals x_{n-L}, \dots, x_{n-1} as well as the AWGN z_n . The parameter L is called channel memory, which is often used as an indicator of how severe the effect of ISI is. It also dictates receiver complexity. The ISI channel can be represented by either a trellis with M^L states for M -ary modulation or a finite state machine with L delay elements.

2. Channel Capacity

In this section, we will discuss the capacity of ISI channels for both constrained (limited to a finite alphabet) and unconstrained channel inputs. We will use the results in this section to design capacity achieving codes in Chapter IV.

a. Unconstrained Inputs

Channel capacity, denoted by C , is defined as

$$C = \triangleq \lim_{N \rightarrow \infty} \sup_{p_{X_0^{N-1}}} \frac{1}{N} I(X_0^{N-1}; Y_0^{N-1}) \quad (2.2)$$

where $I(X_0^{N-1}; Y_0^{N-1})$ is the mutual information between input $\{X_n\}$ and output $\{Y_n\}$ and $p_{X_0^{N-1}}$ is the probability distribution function (pdf) of the input. We use upper case letters to represent random variables.

It has been shown in [3] that the capacity of the discrete-time ISI channel for unconstrained inputs is obtained as

$$C_G = \frac{1}{2\pi} \int_0^\pi \log \left(1 + \frac{P_{X,\text{opt}}(e^{jw}) |F(e^{jw})|^2}{\sigma_z^2} \right) dw, \quad (2.3)$$

where $F(e^{jw})$ is the discrete Fourier transform of f_0^L , $P_{X,\text{opt}}(e^{jw})$ is the optimum water-filling power spectrum density (PSD) found according to

$$P_{X,\text{opt}}(e^{jw}) = \begin{cases} \lambda - \frac{\sigma_z^2}{|F(e^{jw})|^2}, & \text{if } \lambda > \frac{\sigma_z^2}{|F(e^{jw})|^2}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.4)$$

where λ is the water-level and is chosen such that

$$\frac{1}{\pi} \int_0^\pi P_{X,\text{opt}}(e^{jw}) dw = \mathbb{E}[|X_n|^2] = E_s.$$

The capacity-achieving inputs X_n are correlated Gaussian random variables with zero mean and covariances

$$r_X(k) = \mathbb{E}[X_{n+k} X_n] = \frac{1}{\pi} \int_0^\pi P_{X,\text{opt}}(e^{jw}) \cos(kw) dw.$$

We are interested in computing the loss in capacity of an ISI channel when the input PSD is restricted to a fixed $P_X(e^{jw})$. Define the capacity loss as the difference

between the capacity C_G and the maximal achievable mutual information through the channel when the input has PSD $P_X(e^{jw})$:

$$\Delta C(P_X) \triangleq C_G - \mathcal{I}(X; Y | P_X), \quad (2.5)$$

where

$$\mathcal{I}(X; Y | P_X) = \lim_{N \rightarrow \infty} \sup_{\substack{p_{X_0^{N-1}} \\ P_X(e^{jw}) \text{ fixed}}} \frac{1}{N} I(X_0^{N-1}; Y_0^{N-1}). \quad (2.6)$$

It can be easily shown from [3] that

$$\mathcal{I}(X; Y | P_X) = \frac{1}{2\pi} \int_0^\pi \log \left(1 + \frac{P_X(e^{jw}) |F(e^{jw})|^2}{\sigma_z^2} \right) dw \quad (2.7)$$

$$\Delta C(P_X) = \frac{1}{2\pi} \int_0^\pi \log \left(\frac{\sigma_z^2 + P_{X, \text{opt}}(e^{jw}) |F(e^{jw})|^2}{\sigma_z^2 + P_X(e^{jw}) |F(e^{jw})|^2} \right) dw. \quad (2.8)$$

Note that the supremum in (2.6) is satisfied by a Gaussian input.

b. Constrained Inputs

Channel capacity for constrained inputs, denoted by $C_{\mathcal{A}}$, is defined as

$$C_{\mathcal{A}} \triangleq \lim_{N \rightarrow \infty} \sup_{\substack{p_{X_0^{N-1}} \\ X_n \in \mathcal{A}}} \frac{1}{N} I(X_0^{N-1}; Y_0^{N-1}) \quad (2.9)$$

where \mathcal{A} is a finite alphabet that X_n are constrained to.

It should be noted that the result in subsection a is only applicable to unconstrained inputs. Finding the true capacity of an ISI channel when channel inputs are constrained to a finite alphabet is still an open problem although tight lower and upper bounds have been found recently [4], [5].

For constrained inputs, there has been no close-form expression found for the channel capacity; and it is not clear what input PSD $P_X(e^{jw})$ achieves capacity in the

constrained case. In the next subsection, we discuss a special case when the channel input is identical independently distributed (i.i.d.). In this case, channel capacity can be computed.

c. i.i.d. Capacity

When the input to the channel is i.i.d., the input-output mutual information in this case (often called i.i.d. capacity) is defined as

$$C_{\text{i.i.d.}} \triangleq \lim_{N \rightarrow \infty} \sup_{P_{X_0^{N-1}} = \prod_{n=1}^N P_{X_n}} \frac{1}{N} I(X_0^{N-1}; Y_0^{N-1}). \quad (2.10)$$

For unconstrained inputs, the i.i.d. capacity is easily computed by replacing $P_{X, \text{opt}}(e^{jw})$ in (2.7) with $P_X(e^{jw}) \equiv E_s$, resulting in

$$C_{\text{i.i.d.,G}} = \frac{1}{2\pi} \int_0^\pi \log \left(1 + \frac{E_s |F(e^{jw})|^2}{\sigma_z^2} \right) dw.$$

For constrained inputs, we rewrite the mutual information $I(X_0^{N-1}; Y_0^{N-1})$ in (2.10) as

$$I(X_0^{N-1}; Y_0^{N-1}) = h(Y_0^{N-1}) - h(Y_0^{N-1} | X_0^{N-1}), \quad (2.11)$$

where $h(\cdot)$ and $h(\cdot|\cdot)$ represent entropy and conditional entropy, respectively. The second term in the right hand side of (2.11) is actually the entropy of the AWGN term, which is easily computed (namely it is equal to $\frac{N}{2} \log(2\pi e \sigma_z^2)$). The first term can be computed by using the Asymptotic Equipartition Property (AEP) for random processes

$$h(Y_0^{N-1}) = \lim_{N \rightarrow \infty} \log \Pr(Y_0^{N-1}) \quad \text{almost surely,}$$

where $\Pr(Y_0^{N-1})$ is the probability of receiving sequence Y_0^{N-1} , which can be easily computed by using the BCJR forward recursion. We will discuss BCJR algorithm shortly.

Since there is no close-formed expression for constrained capacity, i.i.d. capacity is often used as a lower bound on the true capacity of the channel. We are interested in i.i.d. capacity because in several coded systems, the channel inputs (codewords) behave like i.i.d. sequences when the codeword length is large and an interleaver is used in front of the channel.

d. Tight Lower Bound on Channel Capacity

In the previous subsection, we have discussed channel capacity when the channel input is i.i.d.. The i.i.d. capacity can be used as a lower bound on the channel capacity. However, it can be intuitively clear that the i.i.d. input may result in a large capacity loss for constrained alphabets. To find a tighter lower bound, we need to maximize the quantity $h(Y_0^{N-1})$. Obviously, any input sequence $\{x_0^{N-1}\}$ can be used to find a lower bound on capacity. A tighter lower bound can be found by not constraining the input to be i.i.d.. In general, it is hard to find a sequence $\{x_0^{N-1}\}$ of finite alphabet that maximizes the entropy $h(Y_0^{N-1})$. However, for the low rate region (or equivalently, low signal to noise ratio region), the input $\{x_0^{N-1}\}$ that maximizes the noiseless channel output energy $\frac{1}{N} \sum_{n=0}^{N-1} |r_n|^2$ often results in higher entropy $h(Y_0^{N-1})$, and consequently higher mutual information. This input sequence can be found by using an algorithm which is similar to the Viterbi algorithm to find the maximum path metric.

B. Channel Coding

1. Turbo Codes

A conventional turbo code, or parallel concatenated convolutional codes (PCCC), consists of two identical component recursive systematic convolutional (RSC) codes and a random interleaver, as shown in Fig. 1. The role of the interleaver is to spread out error patterns that cause small output (Hamming) weight at a component code so that the output weight at the other component code is large. This does not necessarily mean that the overall minimum output weight of the turbo code is large; however, small Hamming weight outputs appear less frequently.

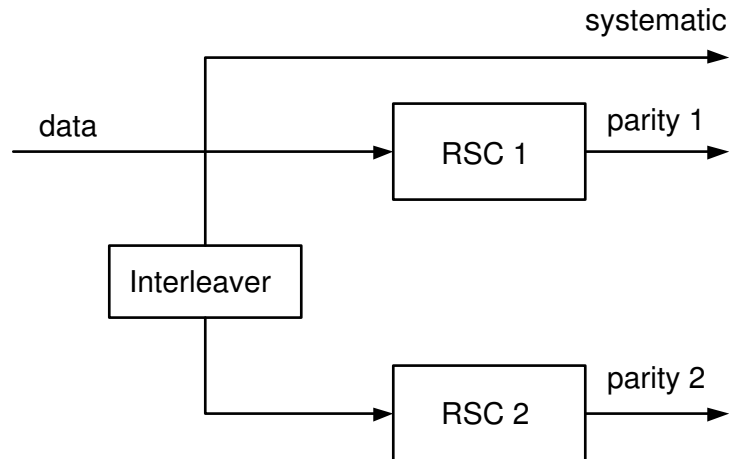


Fig. 1. A rate 1/3 turbo encoder.

2. Serial Concatenation

Similar to the turbo code, a serially concatenated code is formed by an outer code, a random interleaver, and an inner code as shown in Fig. 2. The inner and outer codes are normally chosen to be convolutional codes; and in this case it is often known

as serially concatenated convolutional code (SCCC). It has been shown [6] that by choosing the inner code to be recursive, good distance spectrum of the overall code can be obtained. Since an SCCC normally has a higher minimum Hamming distance compared to that of a PCCC, SCCC is often preferred in some application which requires a lower error floor.

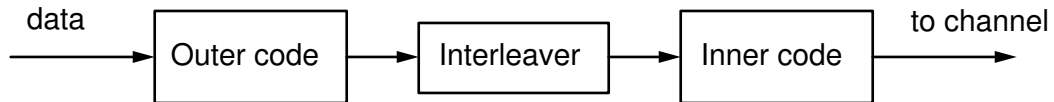


Fig. 2. A serial concatenation.

3. LDPC Codes

Another important class of codes we will use widely in the dissertation is LDPC codes. An LDPC code is a block code whose parity check matrix H contains mostly zeros in its entries. The LDPC encoder can be represented by a bipartite graph with two types of nodes, namely bit nodes and check nodes. The graph is drawn from the parity check matrix H whose rows represent parity check nodes and columns represent coded bit nodes. Entries with value one in the H matrix represent the connections between the corresponding bit and check nodes in the graph. A simple example depicting this is given in Fig. 3.

As we can see from the figure, there is the irregularity in the graph of the code. In other words, the number of connections stemming from coded bit nodes (or parity check nodes) is not uniform. The irregularity of an LDPC code is often represented by

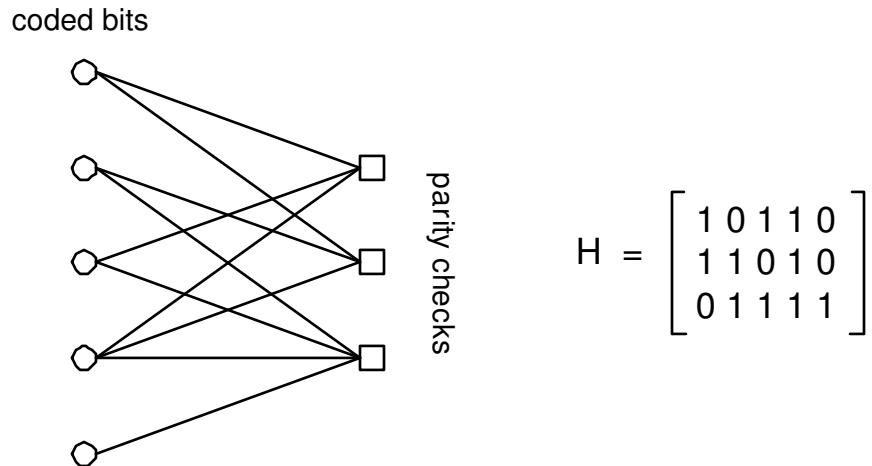


Fig. 3. The bipartite graph of an LDPC encoder.

using degree profiles [7]. Degree profiles (edge perspective) are a pair of polynomials

$$\lambda(x) = \sum_{i=1}^{d_l} \lambda_i x^{i-1} \quad \text{and} \quad \rho(x) = \sum_{i=1}^{d_r} \rho_i x^{i-1}$$

where λ_i represents the percentage of edges from bit nodes (left nodes) which have i connections (edges) to check nodes (right nodes) and ρ_i represents the percentage of edges from check nodes which have i connections to bit nodes. The example in Fig. 3 has the following degree profiles:

$$\lambda(x) = \frac{1}{10} + \frac{6}{10}x + \frac{3}{10}x^2 \quad \text{and} \quad \rho(x) = \frac{6}{10}x^2 + \frac{4}{10}x^3.$$

C. Soft-Input Soft-Output Decoding

We now discuss soft-input soft-output (SISO) decoding for a trellis-based code. Since an ISI channel can be represented by a trellis, SISO decoding algorithms for trellis-based channel codes can be directly applied to the decoding of ISI channels. (Typically, this is referred to as channel equalization.) In the second part of the dissertation,

we will use extensively the a posteriori probability (APP) decoding with the BCJR algorithm [8]. Basically, BCJR is a recursive algorithm whose inputs are channel soft values (either in terms of probabilities or log likelihood ratios), and where forward and backward recursions are used to compute soft outputs. The BCJR algorithm outputs joint probabilities of the state sequence $\{S_n\}$ given the observation sequence Y_0^{N-1} as follows:

$$\sigma_n(m', m) = \Pr(S_{n-1} = m'; S_n = m; Y_0^{N-1}).$$

Let us define the following probability functions:

- Forward recursion: $\alpha_n(m) = \Pr(S_n = m; Y_0^{n-1})$
- Backward recursion: $\beta_n(m) = \Pr(Y_n^{N-1} | S_n = m)$
- Transition probability: $\gamma_n(m', m) = \Pr(S_n = m; Y_n | S_{n-1} = m')$.

Using Markov properties, it can be shown that:

$$\begin{aligned} \sigma_n(m', m) &= \alpha_n(m') \gamma_n(m', m) \beta_{n+1}(m) \\ \alpha_n(m) &= \sum_{m'} \alpha_{n-1}(m') \gamma_n(m', m) \\ \beta_n(m) &= \sum_{m'} \beta_{n+1}(m') \gamma_{n+1}(m, m'). \end{aligned}$$

For actual implementation, the normalized version of the BCJR algorithm is often used. This helps prevent numerical overflows in the forward and backward recursions.

The normalized forward and backward recursions are implemented as follows:

$$\begin{aligned} \alpha_n(m) &= \frac{\sum_{m'} \alpha_{n-1}(m') \gamma_n(m', m)}{\sum_{m', m} \alpha_{n-1}(m') \gamma_n(m', m)} \\ \beta_n(m) &= \frac{\sum_{m'} \beta_{n+1}(m') \gamma_{n+1}(m, m')}{\sum_{m', m} \beta_{n+1}(m') \gamma_{n+1}(m, m')}. \end{aligned}$$

D. Transfer of Soft Information in Iterative Equalization and Decoding

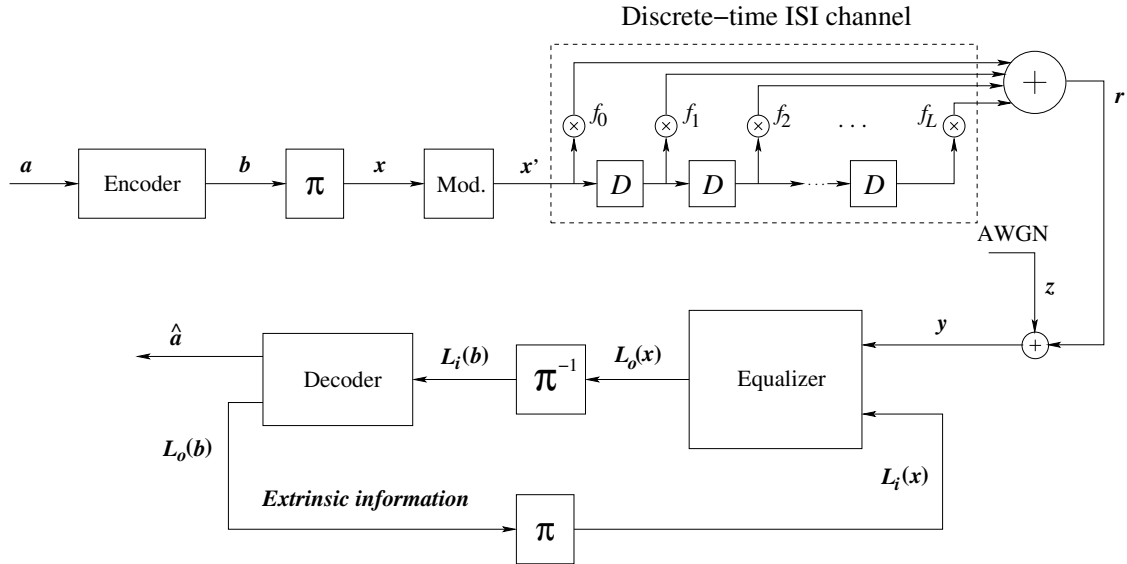


Fig. 4. System model of a serial concatenation with iterative equalization.

We first introduce the concept of iterative equalization (aka turbo equalization). Fig. 4 depicts the system model when iterative equalization takes place. The iterative equalization is an iterative procedure where soft information for coded bits is exchanged between the soft-input soft-output (SISO) equalizer and the SISO decoder. Each round of exchanging the information between the equalizer and the decoder is called one iteration. Many iterations may be necessary before final hard decision.

We now discuss the concept of effective channels for the system given in Fig. 4. During the m th iteration, the inner decoder (inner SISO, or Equalizer) provides the outer decoder (outer SISO) extrinsic LLR information $L_{\text{ext,o}}^{(m)}(x_n)$. We define the expected value (or mean) of the extrinsic log likelihood ratio (LLR) as

$$\phi_o^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}} [X_n' L_{\text{ext}}^{(m)}(x_n)], \quad (2.12)$$

where the notation $\mathbb{E}_{\mathbf{z}; \mathbf{x}}$ denotes the expectation over noise \mathbf{z} and all channel inputs

\mathbf{x} ,

$$L_{\text{ext}}^{(m)}(x_n) = \log \frac{P^{(m)}(X_n = 1)}{P^{(m)}(X_n = 0)},$$

with $P^{(m)}(x_n)$ being the extrinsic probabilities given by the soft-input soft-output (SISO) equalizer (inner SISO), $x'_n = (2x_n - 1)$ is the BPSK modulated signal of x_n , and m denotes the iteration number. Thus, $\phi_o^{(m)}$ represents a measure of how reliable the hypothetical channel as seen by the outer SISO is. The higher $\phi_o^{(m)}$ is, the more reliable the effective channel as seen by the outer SISO during the m th iteration is. Similarly, $\phi_i^{(m)}$ is the mean LLR of the equivalent channel as seen by the inner SISO during the $(m + 1)$ th iteration. The turbo equalization process can be visualized by thinking of each SISO block as a dynamic transformer which operates on the mean LLR's ($\phi_i^{(m)}$ and $\phi_o^{(m)}$). At medium to high E_b/\mathcal{N}_0 regions, as the iterations proceed, $\phi_i^{(m)}$ and $\phi_o^{(m)}$ increase with high probability. This process is repeated until $\phi_i^{(m)}$ and $\phi_o^{(m)}$ reach infinity, which signifies convergence of the decoding algorithm. However, at low E_b/\mathcal{N}_0 , $\phi_i^{(m)}$ and $\phi_o^{(m)}$ do not increase rapidly or, sometimes, do not increase beyond some value at all.

However, decoding for LDPC codes is often done based on the message passing algorithm [9].

E. EXIT Functions

In the literature, there have been many other measures of performance proposed. We will use in this section the extrinsic mutual information transfer (EXIT) functions [10]. Other measures of performance will be discussed further in Chapter III. Let $I_A^{(m)}$ represent the mutual information between the channel input X and the a priori LLR of X at iteration m , denoted by $L_{\text{ap}}^{(m)}$, where $L_{\text{ap}}^{(m)}$ has a Gaussian distribution;

and $I_E^{(m)}$ represent the mutual information between the channel input X and the extrinsic LLR of X , denoted by $L_{\text{ext}}^{(m)}$. That is,

$$I_A^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}}[I(X; L_{\text{ap}}^{(m)})], \quad L_{\text{ap}}^{(m)} \sim \mathcal{N}(x' \phi_{\text{ap}}^{(m)}, 2\phi_{\text{ap}}^{(m)})$$

$$I_E^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}}[I(X; L_{\text{ext}}^{(m)})]$$

where $\phi_{\text{ap}}^{(m)}$ denotes the mean of the a priori LLR. The EXIT curve is a function between a priori mutual information $I_A^{(m)}$ and extrinsic mutual information $I_E^{(m)}$.

It can be seen that $I_A^{(m)}$ and $I_E^{(m)}$ are in the range $[0, 1]$ and directly related to the mean LLR $\phi^{(m)}$. The mutual information reaches value one as $\phi^{(m)}$ approaches infinity. This ensures the convergence of the iterative equalization/decoding process. The mutual information can be used to study the convergence of turbo equalization for ISI channels. We now discuss some key properties of EXIT functions for decoders and equalizers.

1. Properties of EXIT Functions

Let \mathcal{T}_{eq} and \mathcal{T}_{dec} be EXIT functions (output extrinsic mutual information as a function of input a priori mutual information) for the equalizer and the decoder, respectively. Then the following two important properties can be shown [11].

- Convergence Property: For the turbo equalization to converge,

$$\mathcal{T}_{\text{dec}}^{-1}(x) < \mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0), \quad 0 \leq x < 1.$$

- Area Property: If the extrinsic information is from an erasure channel, under optimal decoding, the code rate R is related to the area under the EXIT function

of the decoder as follows

$$\int_0^1 \mathcal{T}_{\text{dec}}^{-1}(x) dx = R$$

or

$$\int_0^1 \mathcal{T}_{\text{dec}}(x) dx = 1 - R.$$

By simulation, it has been also observed that even the extrinsic information is from a Gaussian channel, the area under the EXIT function approximates very well the code rate.

These two properties will be used for code design. We mention briefly the idea in the following subsection; and further discussion on this will be revisited in later chapters when we design specific codes.

2. Using EXIT Functions for Code Design

Basing on the convergence and area properties of EXIT functions, it can be seen that the optimal code is the one whose EXIT function satisfies

$$\mathcal{T}_{\text{dec}}^{-1}(x) = \mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0), \quad 0 < x \leq 1.$$

In other words, the area under the EXIT function of the equalizer is the achievable information rate (or, loosely speaking, i.i.d. capacity). Thus, the problem of designing optimal codes becomes a curve fitting problem. This significantly reduces the design complexity since the EXIT functions of the decoder and the equalizer can be computed separately. By not combining EXIT functions of the equalizer and the decoder, more insights into code design, achievable information rate, and rate loss can be seen easily. We will discuss more about this in later chapters.

CHAPTER III

CONVERGENCE OF TURBO EQUALIZATION FOR ISI CHANNELS

In this chapter, we present new results pertaining to the design and analysis of codes for ISI channels when there is no spectral shaping, and in the context of iterative equalization. We first study the design and analysis of convolutional codes with binary precoding. An analytical proof is presented to show that precoding results in a loss in fidelity during the first iteration for all 2-tap ISI channels. Based on this and the convergence properties of precoded and non-precoded channels, it is shown that by using a mixture of precoded and non-precoded parts, better performance can be achieved. Due to the recursiveness introduced from precoding, the codes required to achieve good performance are very simple codes and, hence, the resulting schemes provide good bit error rate performance at low receiver complexity. We then propose an LDPC code design technique based on EXIT functions. This technique will be shown very accurate and results in good codes. By not combining EXIT functions of the equalizer and the decoder as done in the literature, more insights into code design, achievable information rate, and rate loss can be seen easily.

A. Introduction

Most of the work on turbo equalization has considered only finite impulse response ISI channels which are non-recursive [12], [13]. When viewed as a serial concatenation scheme with an outer code, there is no interleaving gain due to the ISI channel. Consequently, either powerful codes such as turbo codes and LDPC codes should be used as the outer code; or when convolutional codes are used as in [12], the resulting gain is rather small.

A precoding technique that makes the ISI channel appear recursive without in-

creasing the equalization complexity has been used in the area of magnetic recording with very high code rates [14], [15]. Other than in this application, there has been limited work in considering such precoding for general ISI channels and lower rate codes except for [16], [17]. The union bound on the bit error rate (BER) on precoded ISI channels with an outer code was presented in [17]. Such a distance spectrum based analysis is useful only for the high signal to noise ratio (SNR) and is practically useless for lower values of SNR.

In this chapter, we analyze the performance of the iterative equalization using a technique similar to that of density evolution [9] or the extrinsic information transfer (EXIT) chart [18, 10]. We design convolutional codes and LDPC codes for ISI channels. When convolutional codes are used as outer codes, precoding is also utilized.

In [15], a similar study was used to describe the effect of precoding on the convergence of turbo equalization. There a measure of performance called fidelity of the equivalent channels as seen by the decoder and the equalizer at every iteration was defined. It was shown that the initial fidelity at the output of the equalizer at the first iteration is important to the convergence of iterative decoding [15]. It was also shown through simulations that precoding reduces the initial fidelity; however, this was not proved in [15].

Two new results are presented in this chapter. An analytical proof is given to show that the initial fidelity is worse for precoded channels than non-precoded channels for all 2-tap ISI channels and all values of SNR. Using this fact, it is shown that by using a mixture of precoded and non-precoded ISI channels, improved performance can be achieved. Analysis via density evolution shows that the resulting schemes provide close to capacity performance (within a few tenths of a dB) with very simple outer codes and, hence, with fairly low decoding complexity.

The remainder of the chapter is organized as follows. In Section B, we present the

system model. In Section C, we study some measures of performance, which are useful to study density evolution. We prove some key properties of precoding that affect the convergence of turbo equalization. In Section D, we consider the convergence analysis of coded schemes over ISI channels. We propose a system that uses precoding only for a fraction of the coded bits and show how to optimize this fraction. In Section E, we design LDPC codes for ISI channel using the EXIT function matching technique. Finally, we conclude the chapter in Section F.

B. Precoded ISI Channels

We will discuss coding techniques for ISI channels in the context of the serial concatenation with iterative equalization as shown in Fig. 5. When the encoder is chosen

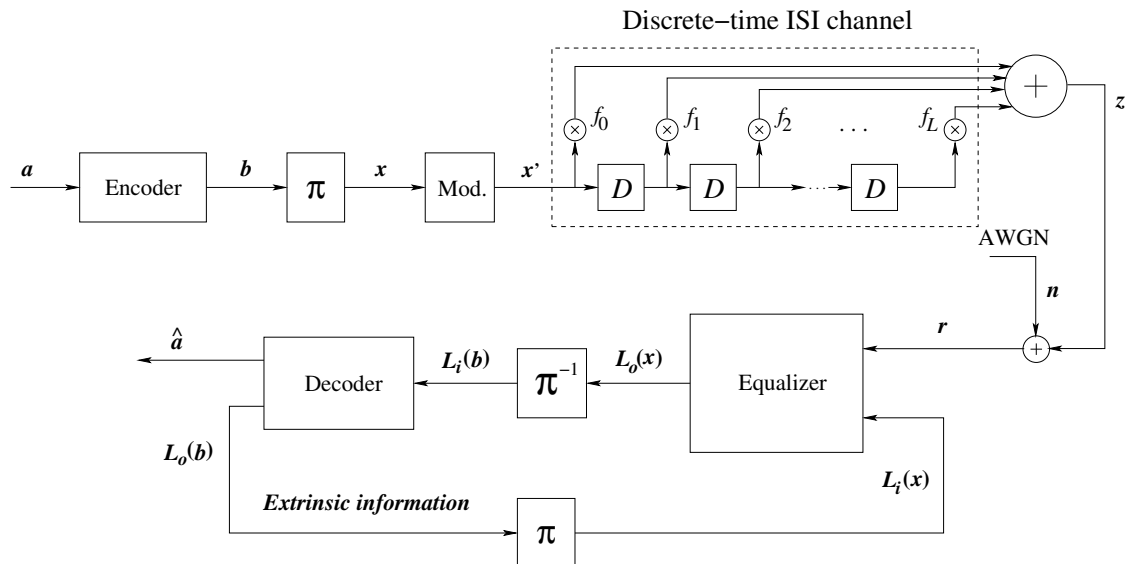


Fig. 5. System model for the serial concatenation of an outer code and an ISI channel

to be a convolutional code, we use precoding before channel transmission. However, when LDPC codes are used, precoding before channel transmission is not necessary. The reason behind this is explained as following.

Since the ISI channel is typically a feed-forward (non-recursive) code, combination of the outer convolutional code and the channel corresponds to a serial concatenation with non-recursive inner code. It is well-known from [6] that when the inner code of a serial concatenated code is not recursive, an interleaving gain does not result and, hence, performance of such a scheme is not good. Nevertheless almost all previous work in the turbo equalization area uses such a system.

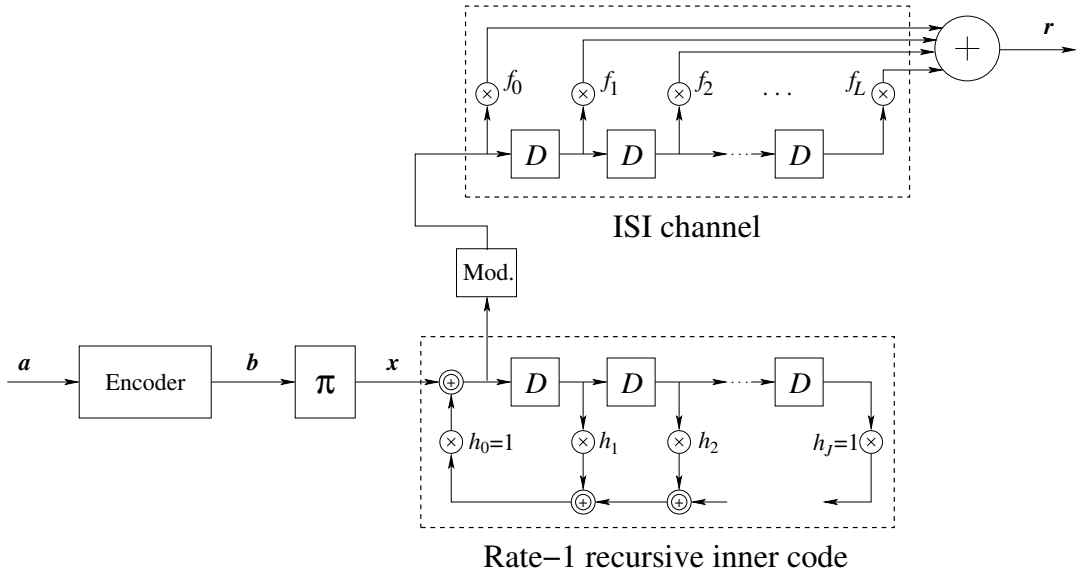


Fig. 6. Concatenation of a rate-1 recursive inner coder (precoder) and an ISI channel.

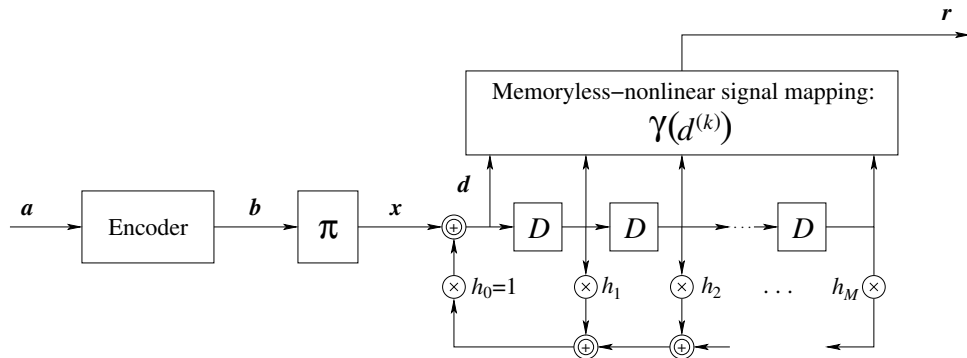


Fig. 7. The equivalent precoded ISI channel, where $M = \max(L, J)$.

We now discuss a technique that makes the non-recursive channel appear recursive and, hence, results in an interleaving gain. The idea is to encode the coded data coming from the outer code using a rate-one recursive convolutional encoder (binary precoder) prior to transmission over the channel as shown in Fig. 6.

The binary precoder is a convolutional code with memory J and is specified by the polynomial

$$h(D) = \bigoplus_{i=1}^J h_i D^i, \quad h_i \in \{0, 1\}, \quad h_0 = h_J = 1, \quad (3.1)$$

where \oplus denotes binary modulo addition. Figure 6 shows the concatenation of the binary precoder and an equivalent discrete-time ISI channel. We can represent this concatenation by combining the channel memory and the precoder memory into a canonic form as illustrated in Fig. 7. When $J \leq L$, the number of states in the trellis of the precoded ISI channel is the same as that of the non-precoded ISI channel, thus decoding complexity remains the same. The mapping function in Fig. 7 is defined as

$$\gamma(\mathbf{d}^{(n)}) \triangleq \sum_{i=0}^L f_i \mu(d_{n-i}), \quad (3.2)$$

where $\mathbf{d}^{(n)} = (d_n, d_{n-1}, \dots, d_{n-L})$, and $\mu(d_n)$ is the modulation function depending on which modulation format is used. For BPSK modulation, we have $\mu(d_n) = 2d_n - 1$. Note that the input $\mathbf{d} = (d_0, d_1, \dots, d_{N-1})$ to the mapper is now an encoded version of a codeword \mathbf{x} of the outer code and satisfies the following relation:

$$d_n = \bigoplus_{i=1}^J h_i d_{n-i} \oplus x_n. \quad (3.3)$$

We can rewrite (3.3) as

$$x_n = \bigoplus_{i=0}^J h_i d_{n-i}. \quad (3.4)$$

In this chapter, we focus on the convergence of precoded ISI channels. In Chapter V, we will show that when precoding is used for both static and fading ISI channels, recursiveness is obtained. Thus, by employing precoding, interleaving gains are possible for all ISI channels.

It must be emphasized here that the purpose of the proposed precoding is quite different from conventional precoding schemes such as Tomlinson-Harashima (TH) precoding, which are aimed at canceling the ISI at the transmitter. The precoding technique considered here simply attempts to make the ISI channel appear recursive and does not attempt to cancel ISI. As a result, unlike TH precoding, channel knowledge is not required at the transmitter.

C. Measures of Performance

In Chapter II, we have discussed the concept an equivalent channel in the turbo equalization. We have used the mean of extrinsic LLR and extrinsic mutual information as measures of performance of the equivalent channel. It is possible to use other measures of performance other than the mean LLR and mutual information for the reliability of the equivalent channels. We will consider the following closely related quantities in this chapter.

- *Mean of the extrinsic LLR:*

$$\phi^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}} [X'_n L_{\text{ext}}^{(m)}(x_n)],$$

where $X'_n = (2X_n - 1)$ is the BPSK modulated signal of X_n ; (We use upper-case letters to denotes random variables, and lower-case letters to denote realizations.)

- *Fidelity [15]:*

$$\pi^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}} \left[X'_n \tilde{X}_n^{(m)} \right],$$

where $\tilde{X}_n^{(m)} = 1 \times P^{(m)}(\hat{X}_n = 1) + (-1) \times P^{(m)}(\hat{X}_n = 0) = \tanh \left(L_{\text{ext}}^{(m)}(x_n)/2 \right)$ is the soft estimate of X'_n at the m th iteration;

- *Extrinsic mutual information [10]:*

$$\begin{aligned} I_{\text{E}}^{(m)} &\triangleq \mathbb{E}_{\mathbf{z}; \mathbf{x}} [I(X_n; L_{\text{ext}}^{(m)}(x_n))] \\ &= \frac{1}{2} \sum_{x'_n \in \{\pm 1\}} \int_{-\infty}^{+\infty} p_L(\psi|X = x'_n) \log_2 \frac{2p_L(\psi|X = x'_n)}{p_L(\psi|X = -1) + p_L(\psi|X = +1)} d\psi, \end{aligned} \quad (3.5)$$

where $p_L(\cdot|\cdot)$ denotes the conditional probability density function (pdf) of the extrinsic LLR $L_{\text{ext}}^{(m)}(x_n)$;

- *Expected ratio of the probability of correct decision to the probability of erroneous decision:*

$$\overline{\text{CER}}^{(m)} \triangleq \mathbb{E}_{\mathbf{z}; x_0^{n-1}, x_{n+1}^{N-1}} \left[\frac{P^{(m)}(\hat{X}_n = 1|X_n = 1)}{P^{(m)}(\hat{X}_n = 0|X_n = 1)} + \frac{P^{(m)}(\hat{X}_n = 0|X_n = 0)}{P^{(m)}(\hat{X}_n = 1|X_n = 0)} \right],$$

where $P^{(m)}(\cdot|\cdot)$ denotes the conditional extrinsic *a posteriori* probability at the m th iteration.

It can be seen that the fidelity $\pi^{(m)}$ and the mutual information $I_{\text{E}}^{(m)}$ are in the range $[0, 1]$ and directly related to the mean LLR $\phi^{(m)}$. The fidelity and mutual information reach value one as $\phi^{(m)}$ approaches infinity. Similarly, when $\phi^{(m)}$ reaches infinity, so does $\overline{\text{CER}}^{(m)}$. This ensures the convergence of the iterative equalization/decoding process. Any of the above quantities can be used in order to study the convergence of the turbo equalization for ISI channels. The mean LLR results

in a diagram that is easy to visualize. However, analytical results will be presented for other performance measures. When the mutual information is used instead of the mean LLR, the resulting diagram is the EXIT chart [10].

Now, we study the effect of precoding on these performance measures. We first prove the following property.

Property 1: For every 2-tap ISI channel, at a given E_b/\mathcal{N}_0 , $\pi_{\text{o,np}}^{(1)} \geq \pi_{\text{o,p}}^{(1)}$, where $\pi_{\text{o,np}}^{(1)}$ and $\pi_{\text{o,p}}^{(1)}$ are the initial reliabilities for non-precoded and precoded 2-tap ISI channels, respectively.

The proof is provided in Appendix A.

We now investigate closely the relationship between $\pi_{\text{o}}^{(1)}$ and $\overline{CER}^{(1)}$. We prove that under certain conditions, it is possible to prove directly that $\overline{CER}^{(1)}$ for the 2-tap non-precoded channel is higher than that for the precoded one. First, we note that $P_{\text{np}}(\hat{X}_n = 1|X_n = 0)$, $P_{\text{np}}(\hat{X}_n = 0|X_n = 1)$, $P_{\text{p}}(\hat{X}_n = 1|X_n = 0)$, and $P_{\text{p}}(\hat{X}_n = 0|X_n = 1)$, where the subscripts p and np denote precoded and non-precoded channels, respectively, are random variables with respect to the additive noise and transmitted sequences \mathbf{x} . We assume that pdfs of those random variables exist and are continuous. Let $p_{\text{np};1|0}$, $p_{\text{np};0|1}$, $p_{\text{p};1|0}$, and $p_{\text{p};0|1}$ be pdfs of $P_{\text{np}}(\hat{X}_n = 1|X_n = 0)$, $P_{\text{np}}(\hat{X}_n = 0|X_n = 1)$, $P_{\text{p}}(\hat{X}_n = 1|X_n = 0)$, and $P_{\text{p}}(\hat{X}_n = 0|X_n = 1)$, respectively. Then we have the following property.

Property 2: If $p_{\text{np};1|0}$ and $p_{\text{p};1|0}$ cross over each other at only one point whose abscissa is in $(0, 1)$ and if $p_{\text{np};0|1}$ and $p_{\text{p};0|1}$ cross over each other at only one point whose abscissa is in $(0, 1)$, then for every 2-tap ISI channel, at a given E_b/\mathcal{N}_0 , $\overline{CER}_{\text{np}}^{(1)} \geq \overline{CER}_{\text{p}}^{(1)}$, where the subscripts p and np represent for the precoded and non-precoded channels, respectively.

The proof is provided in Appendix B.

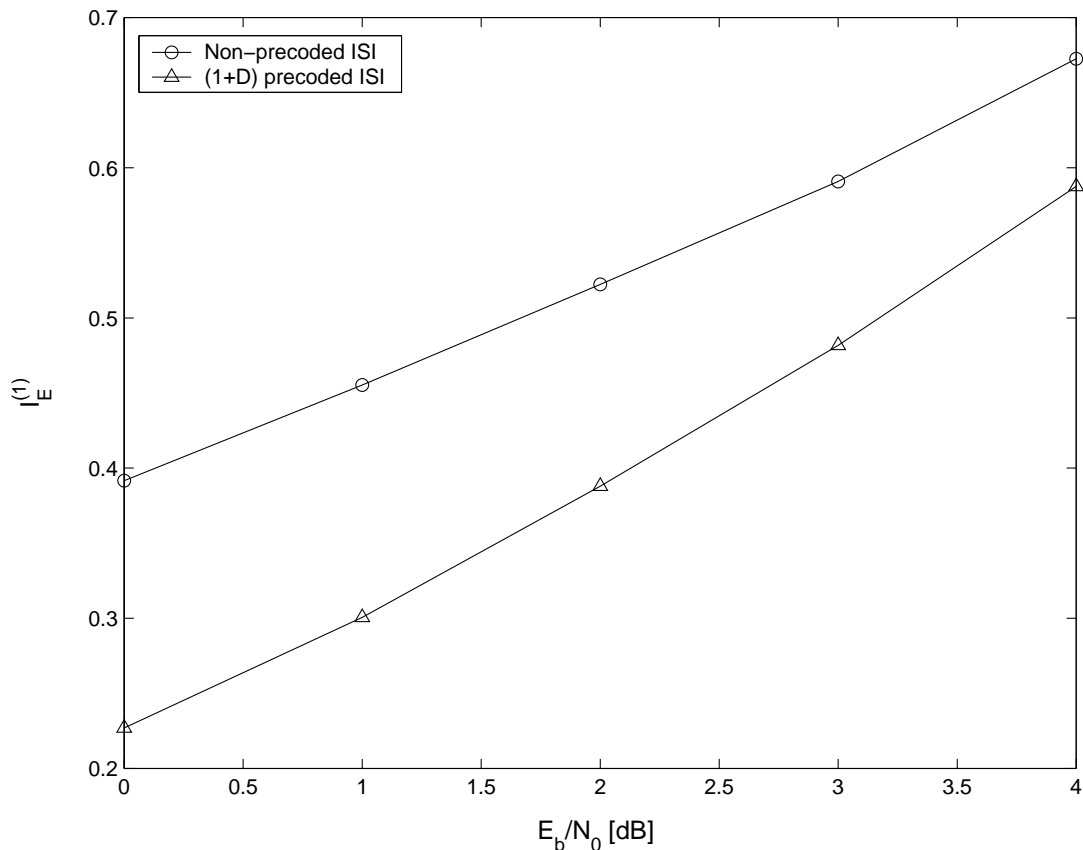


Fig. 8. Loss of mutual information at the first iteration: Plot of $I_{E,o}^{(1)}$ as a function of E_b/N_0 for the 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel.

We are unable to prove that precoding causes a degradation of soft information quality as measured by the mean LLR $\phi_o^{(1)}$ and the mutual information $I_E^{(1)}$. Thus, we resort to the Monte-Carlo simulation. Figure 8 compares loss in $I_E^{(1)}$ for the 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel. It can be seen that the loss in $I_E^{(1)}$ due to precoding can be as much as 2 dB in E_b/N_0 compared to no precoding. For ISI channels with more than 2 taps, we also observed the loss in $I_E^{(1)}$.

Following is another crucial property that affects the convergence.

Property 3 [19]: Let $\mathcal{F}_{i,p}$ and $\mathcal{F}_{i,np}$ be the transfer functions (the output mean LLR as functions of the input mean LLR) of the precoded and non-precoded inner

SISO (equalizer), respectively. Asymptotically, $\mathcal{F}_{i,np}(\phi_i^{(m)}) \rightarrow 2/\sigma_z^2$ as $\phi_i^{(m)} \rightarrow \infty$ while $\mathcal{F}_{i,p}(\phi_i^{(m)}) \rightarrow \infty$ as $\phi_i^{(m)} \rightarrow \infty$.

D. Channel Mixture

In this section, we will demonstrate that by using only a fraction of precoded bits to the channel, better performance can be achieved. In the following first subsection, we will use the expected extrinsic LLR transfer diagram to design codes because it is easier to visually see the convergence of turbo equalization. For the threshold computation, it has been pointed out by ten Brink [10] and Narayanan [15] that mutual information transfer diagram (EXIT chart) and fidelity transfer diagram should be used, respectively. Thus, we will also use EXIT chart here to study our system and point out why using mutual information is more robust than using mean LLR in code design.

1. Code Design Based on Expected Extrinsic LLR Transfer Diagram

From Properties 1 and 2 and the discussion in the previous section, we can see that $\phi_{o,p}^{(m)} \leq \phi_{o,np}^{(m)}$ for small m . Thus, with a fixed outer code, the increase of mean LLR (or, fidelity; or, mutual information) for non-precoded channels is higher than that for precoded channels during the first few iterations. (When m is small.) In contrast, from Property 3, we see that the convergence behavior of the precoded system is faster as the number of iterations increases. This is because the asymptotic slope of $\mathcal{F}_{i,p}$ for the precoded channel is higher than that of $\mathcal{F}_{i,np}$ for the non-precoded channel. This implies that the non-precoded channel is better during the first few iterations whereas the precoded one is better during the later iterations. In order to combine the advantages of the precoded and non-precoded bits, we propose to

combine the precoded and non-precoded ISI channels by partitioning the interleaved codewords of the outer code into two blocks. A precoder is used with one of the block whereas it is not used with the other. Hence, the effective inner channel is now a mixture of precoded and non-precoded channels. This technique is similar to doping proposed by ten Brink in [20]. However, the fraction of precoded and non-precoded bits is usually not very small like in doping. Let a fraction λ of the bits of the outer code word (after interleaving) be precoded and the fraction $(1 - \lambda)$ be non-precoded. Then, if the transfer characteristics of the precoded and non-precoded channels are $\mathcal{F}_{i,p}$ and $\mathcal{F}_{i,np}$, the transfer function of the mixture is $\mathcal{F}_{i,mix} = \lambda\mathcal{F}_{i,p} + (1 - \lambda)\mathcal{F}_{i,np}$. The optimum value of λ depends on the actual channel, precoder, and the outer code. We now propose a technique to determine this value λ .

a. Optimizing the Mixture

Let \mathcal{F}_o be the transfer function of the outer code. At a given value of E_b/\mathcal{N}_0 , let a be the abscissa of the intersection of the curves $\mathcal{F}_{i,p}^{-1}$ and \mathcal{F}_o , b be the abscissa of the intersection of $\mathcal{F}_{i,np}^{-1}$ and \mathcal{F}_o , and finally c be the abscissa of the intersection of $\mathcal{F}_{i,np}^{-1}$ and $\mathcal{F}_{i,p}^{-1}$. If any of the above pair of curves intersects at more than one point, then let a be the smallest value and b be the largest value.

Condition 1: In order for the mixture to provide better results than either the precoded ISI channel or the non-precoded ISI channel, the following condition must be satisfied: $a \leq c \leq b$.

Proof: This can be seen since for the values of E_b/\mathcal{N}_0 for which the crossover between $\mathcal{F}_{i,np}^{-1}$ and $\mathcal{F}_{i,p}^{-1}$ is above the curve \mathcal{F}_o , both the curves $\mathcal{F}_{i,p}^{-1}$ and $\mathcal{F}_{i,np}^{-1}$ will intersect \mathcal{F}_o and, hence, for any linear combination of them, $\mathcal{F}_{i,mix}^{-1}$ will also intersect \mathcal{F}_o and, hence, there will be a fixed point.

This property makes it much easier for us to find a near optimum value of λ . We

first note that a lower bound on the threshold $(E_b/\mathcal{N}_0)_{\text{Th}}$ with the optimum λ occurs for that value of E_b/\mathcal{N}_0 for which $a = b = c$. The following approach is then used to find the optimum λ .

We first compute the value of E_b/\mathcal{N}_0 (say $(E_b/\mathcal{N}_0)_{\text{Th}}$) for which the crossover between the curves $\mathcal{F}_{\text{i,p}}^{-1}$ and $\mathcal{F}_{\text{i,np}}^{-1}$ is on the curve \mathcal{F}_o . This implies that for all values of $E_b/\mathcal{N}_0 < (E_b/\mathcal{N}_0)_{\text{Th}}$ the crossover point will lie above the curve \mathcal{F}_o and, hence, no further improvement is possible. Once $(E_b/\mathcal{N}_0)_{\text{Th}}$ is determined then the value of λ for which there are no fixed points between \mathcal{F}_o and $\mathcal{F}_{\text{i,mix}}^{-1}$ is the optimum λ . It should be noted that this procedure is much simpler than evaluating all possible λ and their associated thresholds, since we directly find the optimum threshold first and then find the optimum λ .

The above procedure was verified to provide improved results over purely precoded or non-precoded channels. As an example, the advantage of using mixed ISI channels is demonstrated through Fig. 9. The outer code is the rate-half, 4-state convolutional code with polynomial: $(1 + D + D^2, 1 + D^2)$. The channel has 2 taps: $f_0 = \sqrt{0.5}, f_1 = -\sqrt{0.5}$. For this channel, the binary input i.i.d. capacity is 0.8 dB. We found that $\lambda = 0.55$ and the precoder is $1 \oplus D$. The threshold for the precoded system is 1.45 dB. With the optimized mixture, the threshold is 1.0 dB, which is only 0.2 dB away from the binary input i.i.d capacity of this ISI channel. In this case, we obtain the coding gain of 0.45 dB. It should be noted that the complexity of the proposed scheme is significantly smaller than using a turbo outer code with turbo equalization, since the outer code is only a 4-state convolutional code. Also note that either of the proposed techniques (precoding or mixture precoding) does not increase the receiver complexity beyond that of turbo equalization with the non-precoded channel.

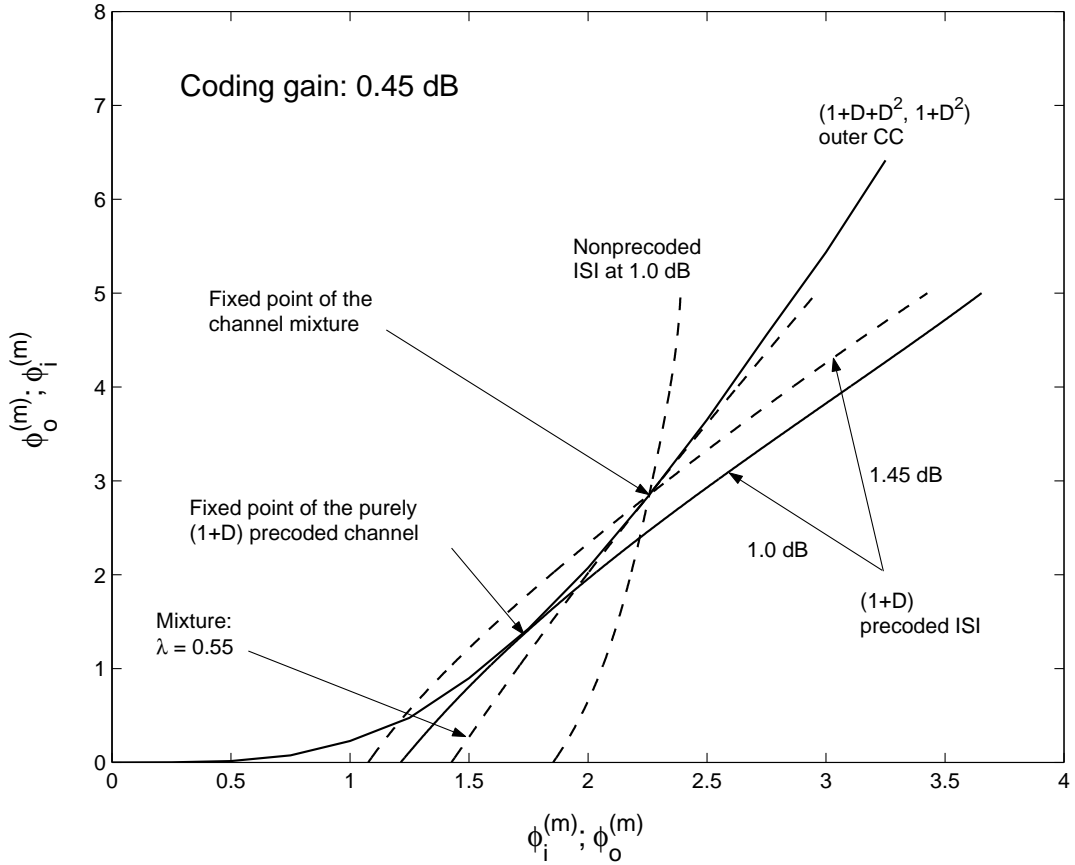


Fig. 9. Density evolution of the 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and the $(1 + D + D^2, 1 + D^2)$ outer convolutional code; the precoding polynomial $h(D) = 1 \oplus D$.

2. Code Design Based on Extrinsic Information Transfer (EXIT) Functions

In the previous subsection, we have used the expected extrinsic LLR transfer diagram \mathcal{F} along with the assumption that the pdfs of the LLRs are symmetric Gaussian for code design. This approach has been used in [21] and [22]. However, this approach can lead to inaccuracies as described below. First note that if, $p_{L,p}$, $p_{L,np}$, and $p_{L,mix}$ are the pdfs of extrinsic LLRs at the output of the inner SISO for precoded, non-

precoded, and mixed channels, respectively. Then,

$$p_{L,\text{mix}} = \lambda p_{L,\text{p}} + (1 - \lambda)p_{L,\text{np}}. \quad (3.6)$$

When the mean LLR input to the inner SISO approaches infinity, the mean LLR at the output of the inner SISO also approaches infinity; however, the pdf cannot be assumed to be Gaussian, since it is linear combination of two Gaussian pdfs according to (3.6). This leads to inaccuracies particularly since in this case, even though the output LLR from the inner SISO approaches infinity, it does not mean that the equivalent channel as seen by the outer code is fully reliable. Because the mean LLR contributed by the non-precoded part is upper bounded by $2/\sigma_z^2$, and the pdf of the extrinsic LLR is Gaussian with variance twice the mean, the equivalent channel as seen by the outer code for the coded bits transmitted using the non-precoded part still has some amount of uncertainty. Thus, we need to use a more robust tool to design code. Here, we choose EXIT chart [10] since it has been shown by many authors that this technique gives highly accurate results.

Let \mathcal{T}_i be the mutual information transfer function (we call ‘transfer function’ for short, but this should not be confused with the mean LLR transfer functions, as defined in the previous section) of the inner SISO, i.e., the output mutual information $I_E^{(m)}$ (extrinsic part) as a function of input mutual information $I_A^{(m)}$ (a priori part) at a given value of E_b/\mathcal{N}_0 . That is, $I_E^{(m)} = \mathcal{T}_i(I_A^{(m)}, E_b/\mathcal{N}_0)$. Also let $\mathcal{T}_{i,\text{p}}$, $\mathcal{T}_{i,\text{np}}$, and $\mathcal{T}_{i,\text{mix}}$ be the transfer functions for the precoded, non-precoded, and mixed channels, respectively. It is difficult to get closed-form expressions for $\mathcal{T}_{i,\text{p}}$ and $\mathcal{T}_{i,\text{np}}$ for general ISI channels. Thus, we use Monte-Carlo simulations to find $\mathcal{T}_{i,\text{p}}$ and $\mathcal{T}_{i,\text{np}}$. In general, there is no linear relation between $\mathcal{T}_{i,\text{mix}}$ and the pair $\{\mathcal{T}_{i,\text{p}}, \mathcal{T}_{i,\text{np}}\}$, except for some special cases, such as when pdfs of extrinsic LLRs at the outputs of the inner SISO blocks for both non-precoded and precoded channels are symmetric and consistent

[9]. This can be easily observed from the definition of mutual information as defined in (3.5). To find $\mathcal{T}_{i,\text{mix}}$, the pdf of the LLR for the channel mixture can be computed by first computing $p_{L,p}$ and $p_{L,np}$ through simulations and then using (3.6). Once $p_{L,\text{mix}}$ is obtained, it is straightforward to compute $\mathcal{T}_{i,\text{mix}}$ by replacing p_L in (3.5) with $p_{L,\text{mix}}$.

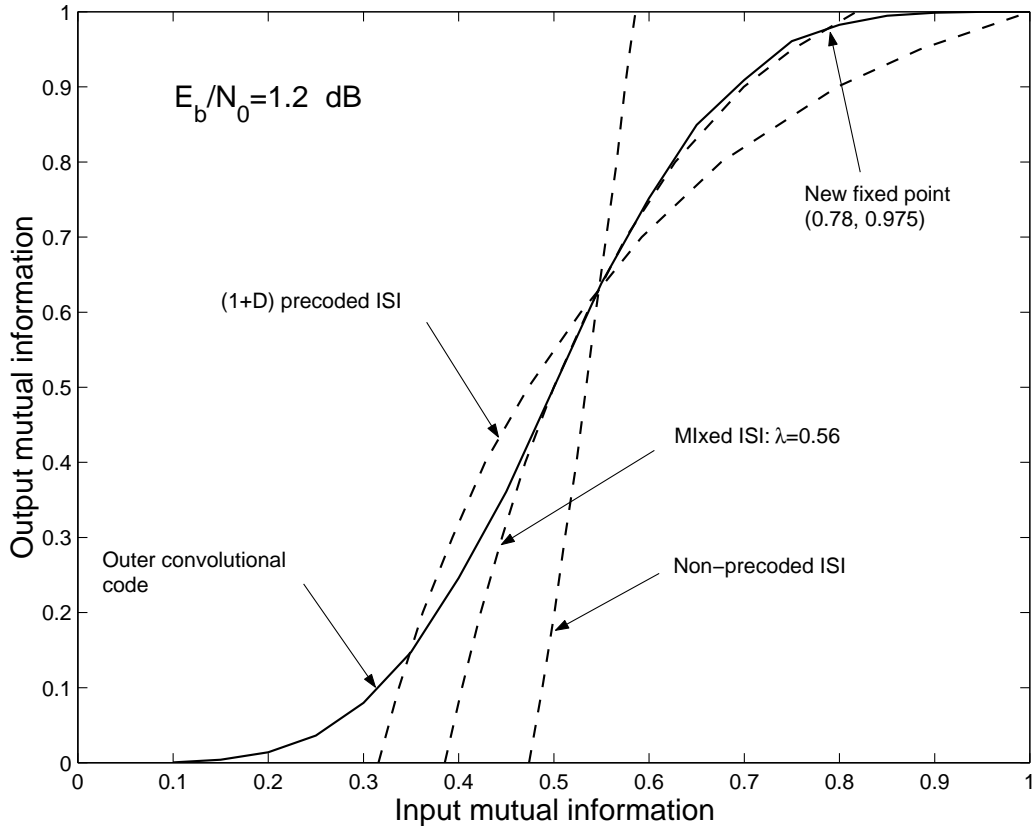


Fig. 10. Channel mixture at 1.2 dB: 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and $(1 + D + D^2, 1 + D^2)$ outer convolutional code, precoding polynomial $h(D) = 1 \oplus D$.

We are now interested in value $\mathcal{T}_{i,\text{mix}}(1, E_b/N_0)$, i.e., the output extrinsic mutual information $I_E^{(m)}$ when the input a priori mutual information $I_A^{(m)}$ is equal 1. That is when the input a priori LLR is equal infinity. From the proof of Property 3, we know that both $p_{L,p}^\infty$ and $p_{L,np}^\infty$ are Gaussian with means ∞ and $2/\sigma_z^2$, respectively, where

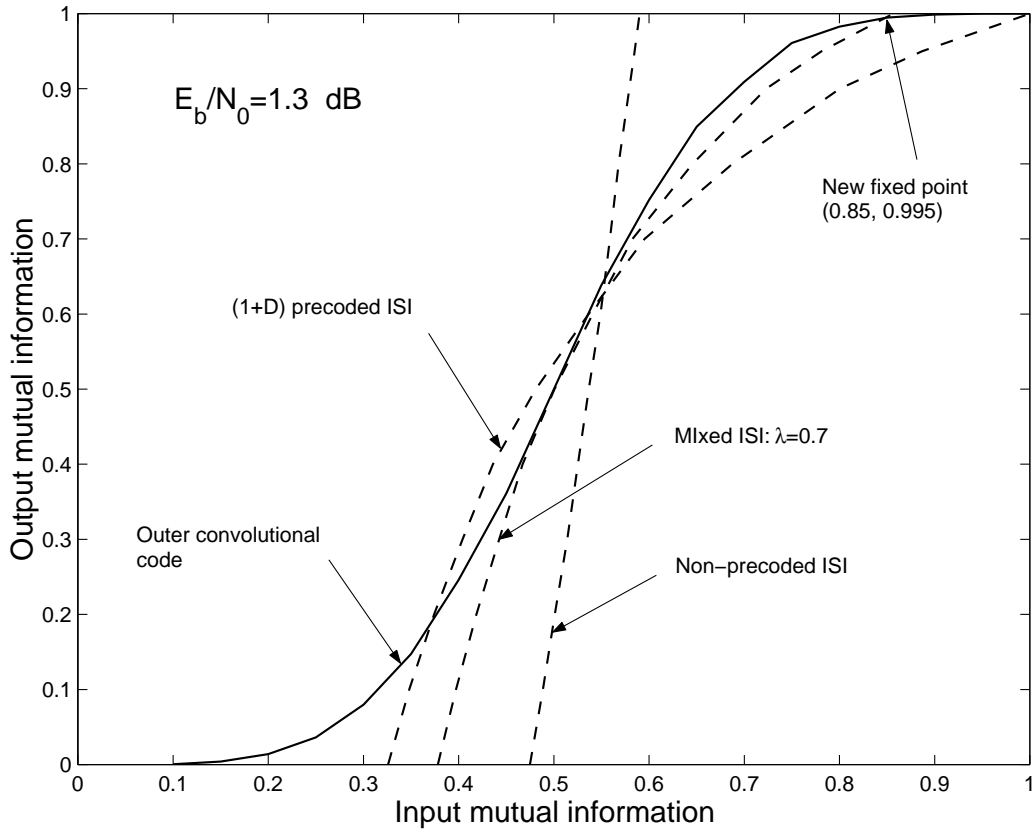


Fig. 11. Channel mixture at 1.3 dB: 2-tap $(\sqrt{0.5}, -\sqrt{0.5})$ ISI channel and $(1 + D + D^2, 1 + D^2)$ outer convolutional code, precoding polynomial $h(D) = 1 \oplus D$.

$p_{L,p}^\infty, p_{L,np}^\infty$ denote pdfs for precoded and non-precoded channels when the input a priori LLR is infinite. Consequently, $\mathcal{T}_{i,\text{mix}}(1, E_b/\mathcal{N}_0) \leq 1$, where the equality holds when $E_b/\mathcal{N}_0 = \infty$, or when $\lambda = 1$ (purely precoded channel). Let \mathcal{T}_o be the transfer function of the outer SISO. If $\mathcal{T}_o(\mathcal{T}_{i,\text{mix}}(1, E_b/\mathcal{N}_0)) < 1$, then there exists a fixed point before point $(1, 1)$. Denote $(I_A^{(\infty)}, I_E^{(\infty)})$ as the fixed point of the turbo equalization/decoding process. When both the a priori and extrinsic LLRs are assumed to be Gaussian distributed, there is a one-to-one correspondence between the fixed point $(I_A^{(\infty)}, I_E^{(\infty)})$ and the value of BER [10]. When either $I_E^{(\infty)} \rightarrow 1$ or $I_A^{(\infty)} \rightarrow 1$, $\text{BER} \rightarrow 0$. For simple outer codes such as convolutional codes, both $I_A^{(\infty)} < 1$ and $I_E^{(\infty)} < 1$ and therefore

the BER can never be made zero. Consequently, an error floor always exists when a channel mixture is used. However, in practice, we can restrict our attention to a finite but low BER and the channel mixture is still useful.

We use the same channel and outer code as in the previous subsection and a similar procedure to optimize the fraction λ ; however, we only require that there are no fixed points in the range (I_A, I_E) , which can be determined given a BER of interest [10]. The results of the channel mixture are given in Figures 10 and 11. The threshold for the precoded channel is 1.6 dB, which is 0.15 dB different from the one computed by the mean LLR transfer diagram in the previous subsection. If the BER is required to be of 10^{-3} , the optimum λ is 0.56 and the threshold for the channel mixture is 1.2 dB (see Fig. 10), and when $BER = 10^{-4}$, the optimum λ is 0.7 and the resulting threshold is 1.3 dB (see Fig. 11).

E. Design of LDPC Codes

We have presented some key properties of EXIT functions that are useful to our code design purpose in Chapter II). In this chapter, we present the LDPC code design idea in the context of iterative equalization.

1. Using EXIT Functions for LDPC Code Design

Basing on the convergence and area properties of EXIT functions, we can state the following about design of LDPC codes:

1. An LDPC code is optimal for a given ISI channel, E_b/\mathcal{N}_0 and equalizer if the EXIT function of the LDPC code for a given (λ, ρ) satisfies $\mathcal{T}_{\text{dec}}^{-1}(x) = \mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0)$. Therefore, the LDPC code design problem can be stated as one of finding degree profiles (λ, ρ) such that the EXIT transfer function for the

LDPC has the same shape as that of $\mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0)$.

2. The achievable information rate for a given equalizer under iterative equalization is the area of the curve under the transfer function $\mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0)$. In other words, the capacity loss due to the use of sub-optimum code with EXIT function $\mathcal{T}_{\text{dec}}^{-1}(x)$ is the area of the gap between $\mathcal{T}_{\text{dec}}^{-1}(x)$ and $\mathcal{T}_{\text{eq}}(x, E_b/\mathcal{N}_0)$.

We note here that precoding does not increase information rate of ISI channels. Thus, when LDPC outer codes are chosen, precoding is unnecessary. However, precoding is crucial in designing convolutional outer codes for ISI channels because it helps improve the convergence in the last stage of iterations.

In [19], we use a Gaussian approximation together with EXIT functions to design LDPC codes. This technique can be simplified to a simple linear optimization problem, in contrast to previous work in the literature where complex nonlinear optimization was used to track the pdf of information messages. We refer the reader to [19] for more detailed of the design technique.

By not combining EXIT functions of the equalizer and the decoder (joint EXIT function), more insight into code design, achievable information rate, and rate loss can be seen easily. We state in the following some important results (See Fig. 12).

a. Use of LDPC Codes Optimum for the AWGN Channel

It can be shown that an LDPC code that is optimal for the AWGN channel has a flat EXIT function. Since the EXIT function for an ISI channel has a slope, thus, using LDPC codes optimal for the AWGN channel results in a loss in rate. This rate loss is the area in between the two EXIT functions.

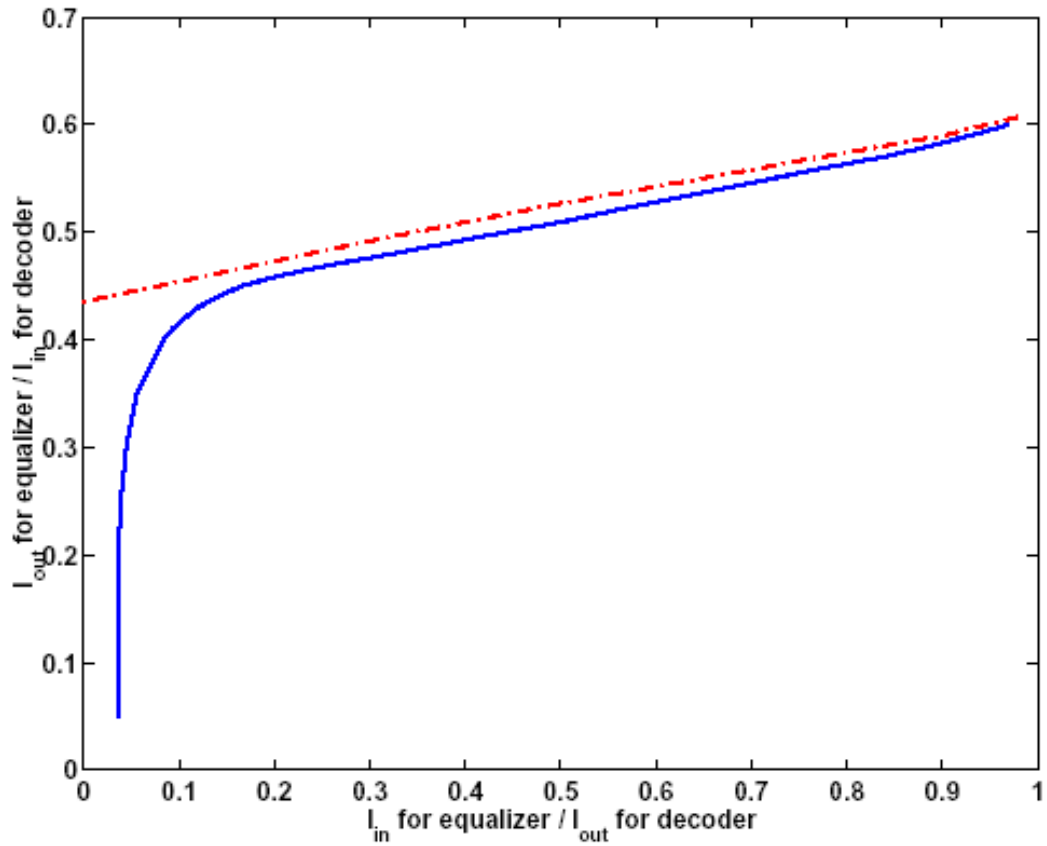


Fig. 12. EXIT chart for an ISI channel and an LDPC code designed based on EXIT function matching.

b. Rate Loss due to Non-iterative Equalization

In some applications, iterative equalization is not viable because of its high complexity. In this case, the maximum achievable rate is the area of a square dictated by the first point in the EXIT function of the channel. Thus, this loss is the same as that when LDPC codes optimal for the AWGN channel are used for ISI channels.

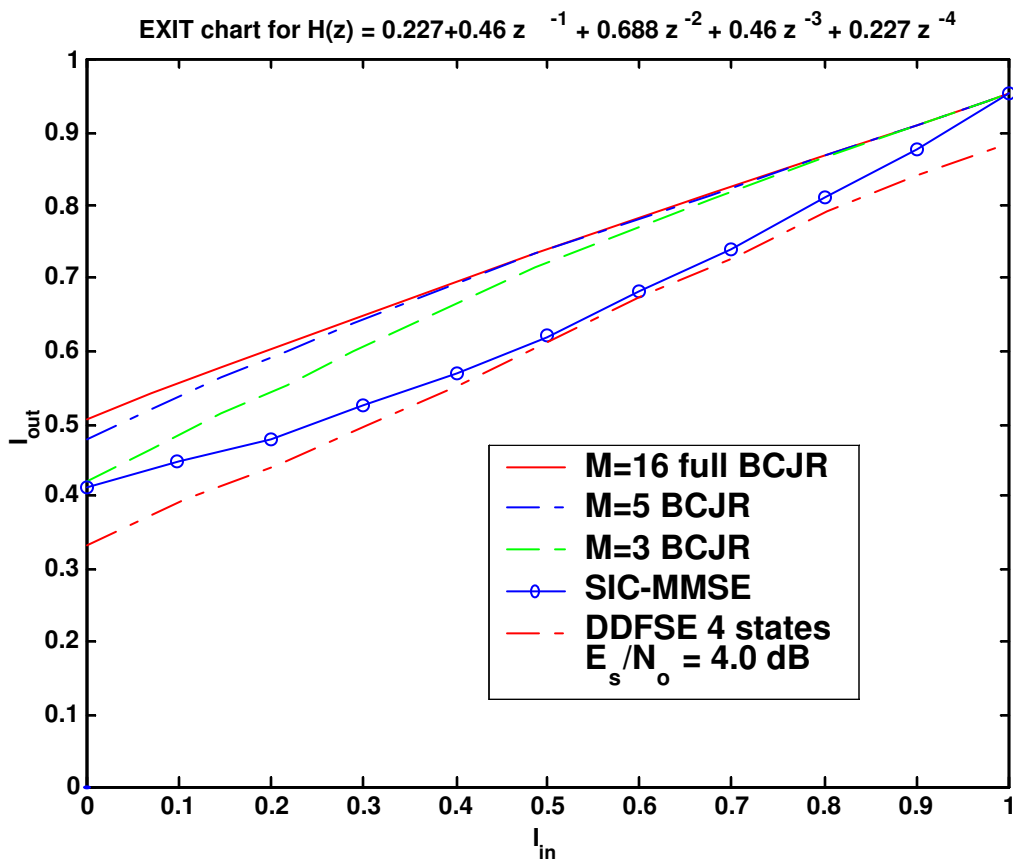


Fig. 13. EXIT functions for different sub-optimal equalizers.

c. Rate Loss due to Sub-optimal Equalizers

It is obvious that the EXIT function for the optimal equalizer (BCJR algorithm) has the highest rate (or, area under the EXIT function) compared to all the other sub-optimal equalizers. Thus, the rate loss due to the sub-optimal equalization is the area of the gap between the optimal and sub-optimal EXIT functions. Fig. 13 shown EXIT functions for the optimal and sub-optimal equalizers. We refer the reader to [19] for more details on the descriptions of different types of sub-optimal equalizers. Here, it can be interpreted that the first points of the EXIT functions for different sub-optimal equalizers are lower than that of the optimal equalizer. This indicates

that when using LDPC codes optimal for the AWGN channel, or when there is no iteration in the equalization/decoding process, there is a significant rate loss.

However, the last points of the EXIT functions for most sub-optimal equalizers are the same as that of the optimal equalizer, except for the delay decision feed-back sequence estimation BCJR (DDFSE-BCJR). This means that if an outer code is a convolutional code, there is almost no rate loss when iterative equalization is present. This is because the last point of the EXIT function dictates the performance.

F. Conclusion

We have demonstrated that precoding is important for ISI channels to achieve coding gains. We have shown via density evolution that precoding provides excellent performance and does not increase complexity or decrease coding rate. We showed analytically that precoding always results in a loss in fidelity during the first stage of iteration for 2-tap channels. For channels with longer memory Monte-Carlo simulations show the same. Using this fact and by properly designing the outer code and precoder, low complexity systems can be realized. It is also interesting that designing rules based on distance spectrum are not necessarily true for iterative decoding in the low E_b/\mathcal{N}_0 regions. We have shown that by using a mixture of precoded and non-precoded channels, better performance can be achieved. Recently, Tuchler and Hagenauer [23] proposed a serially concatenated scheme using mixed outer codes, instead of the channel mixture proposed in this chapter. It seems interesting to combine both the channel mixture and mixed outer codes to get close the capacity performance. This problem is under our current investigation. Finally, we presented design of LDPC codes using EXIT functions. This technique has been proved to be less complex; and resulted in good codes.

CHAPTER IV

CODE DESIGN AND ANALYSIS WITH SPECTRAL SHAPING

In this chapter, we propose two low complexity coding schemes for inter-symbol interference (ISI) channels which perform close to the channel capacity. The first scheme is a serial concatenation of an outer code and a spectral shaping inner code. The second scheme is a parallel concatenation of two component trellis codes which are designed to be spectrally matched to the channel. Analysis using EXIT functions and BER simulations shows that both schemes surpass the independent identically distributed (i.i.d.) channel capacity and outperform other existing schemes.

A. Introduction

For unconstrained channel inputs (not restricted to finite constellations), the optimal channel input is governed by the water-filling spectrum [3] (also shown in Chapter II). In the low rate region, from empirical observations, unconstrained and constrained channel capacity curves lie very close to each other. Furthermore, in this case the channel output (interference plus noise) is mainly contributed by the additive Gaussian component. Therefore, coding techniques for unconstrained input signals can be used as a guideline to design codes for the constrained case. Consequently, codes with flat spectra suffer from a rate loss when used for ISI channels. The capacity loss can be quantified exactly for the unconstrained inputs. Unfortunately, for the finite alphabet inputs, there is still no strong result on the relationship between the information rate and the code spectrum. For more details on this, the reader is referred to [24]. Thus, it is necessary to use some form of spectral shaping for ISI channels. Designing good codes based on spectral shaping is the main idea and the contribution of this chapter.

Matched spectral-null (MSN) codes [25] guarantee that the nulls of codewords are matched to the nulls of the channel frequency response. The concatenation of an MSN inner code and an outer convolutional code was presented in [26]. This scheme was shown to perform quite well for the dicode channel (with impulse response $1 - D$). This suggests that by using a spectral shaping inner code concatenated with a powerful outer code, such as an LDPC code or a turbo code, significant coding gains can be obtained. There is not much work available on this except for [27]. However, the technique in [27] is quite complex and becomes unmanageable when the channel has a large memory. The spectral shaping proposed in this chapter is different from matching the spectral nulls and is based on matching the water-filling spectrum. In the low rate region, the shaping code focuses the signal power near the peaks of the channel response in contrast to MSN codes, where channel input sequences are ensured to have nulls at frequencies where the channel frequency response has nulls. In [19] and [28], good LDPC codes were designed for partial response channels without spectral shaping. In this chapter, we use the technique proposed in [19] to design good LDPC codes for general ISI channels with spectral shaping.

Wolf and Ungerboeck [29] used the concept of set partitioning to design trellis codes for partial response channels. The code was chosen such that the minimum Euclidean distance between permitted noiseless (excluding AWGN) channel outputs is increased. This can be done by considering the set of noiseless channel outputs as an expanded set of signals being transmitted through an AWGN channel. Haeb [30] extended this idea to the multidimensional case. A coding gain of up to 7 dB compared to the uncoded case was achieved by a simple 8-state trellis code.

Designing turbo codes matched to ISI channels poses a challenging problem due to tight restrictions of the component codes in the parallel concatenation. Pusch, Weinrichter, and Taferner [31] proposed a turbo-coding scheme for the $1 - D^2$ channel.

By using the fact that the $1 - D^2$ channel can be considered as two interleaved independent $1 - D$ channels [29], the component codes were designed to have power spectrum density (PSD) matched to the $1 - D$ channel. The coding gain obtained is quite modest (around 1.5 dB away from the i.i.d. capacity at rate 1/4 for the information block length of 1000 bits). For other channels, to our best knowledge, there is still no result on the design of turbo codes matched to the channel.

In this chapter, we apply concepts of spectral shaping and set partitioning to design iteratively decodable codes matched to ISI channels. The main tools used to find good codes are extrinsic information transfer (EXIT) functions [10]. We propose two simple coding schemes which perform better than the binary i.i.d. capacity and outperform other existing schemes.

The first scheme is a serial concatenation of an outer code and a spectral shaping (ISI-matched) inner code. This scheme is based on the observation that the binary i.i.d. capacity of the concatenation between the inner code and the ISI channel is higher than that of the channel alone in the low rate region. For some special cases, such as when the inner code is the biphasic code and the channel is the dicode channel, coding for the AWGN channel can be applied.

The second scheme is a parallel concatenation of two component trellis codes designed such that the minimum Euclidean distance between permitted noiseless channel outputs is as large as possible. Moreover, the trellis transitions are taken from codewords of a block codebook whose PSD is matched to that of the channel. Component trellis codes can be designed by using existing techniques, such as those proposed in [29] and [30]. However, some critical modifications are required, which will be discussed in more details in later sections.

B. Serial Concatenation

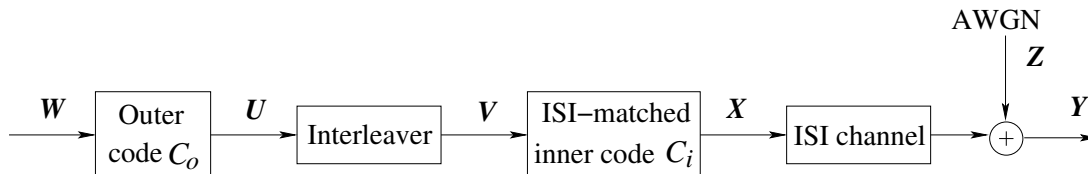


Fig. 14. System model for the serially concatenated scheme.

The generic system model (without the receiver) is shown in Fig. 14. An outer code C_o is serially concatenated with a spectral shaping (ISI-matched) inner code C_i . The inner code is spectrally matched to the ISI channel. In general, the choice of an outer code can be an LDPC code, turbo code, repeat accumulate (RA) code, or even a simple repetition code depending on the actual channel realization and the ISI-matched inner code. We use EXIT functions to find good codes instead of tracking the evolution of pdfs of messages [28], which is computationally complex. For the case when the outer code is an LDPC code, a simple technique proposed in [19] is used.

In [19], an LDPC code is designed such that the LDPC decoder EXIT function matched to the equalizer EXIT function. This is done so that the LDPC code rate is as high as possible. It should be noted that the code rate is approximately equal to the area under the decoder EXIT function [11], which is upper-bounded by the equalizer EXIT function.

The overall receiver is an iterative receiver which performs turbo equalization by passing extrinsic messages on the coded bits between the decoder and the equalizer. The decoder implements the message passing algorithm for the outer code. The equalizer works on the combined trellis of the inner code and the ISI channel.

1. Block Inner Codes for Spectral Shaping

We use an inner code before the ISI channel so that the combined inner code and ISI channel has a higher i.i.d. capacity than that of the channel alone for the desired operating rate region. In [27], good trellis codes that nearly optimize the i.i.d capacity for the combined channel are proposed and shown to provide good performance. However, this technique is quite complicated, even for channels with short memories. Here, we propose the design of an inner code based on a simple spectral shaping technique, rather than on maximizing the mutual information as done in [27]. Both block and trellis codes can be used for shaping; however, we consider only block codes in this chapter. The basic idea is to design a block code such that the PSD of codewords is matched closely to the water-filling spectrum. For a wide range of E_b/\mathcal{N}_0 , for most channels, the channel spectrum (square of the magnitude of the channel frequency response) is a good approximation to the water-filling spectrum. We focus on block codes of length $(L + 1)$, where L is the channel memory. To make the exposition simple, we consider only the BPSK modulation. Extension to other code lengths and modulation schemes is straightforward.

Let \mathcal{X} be the codebook (inner shaping code alone) consisting of $M = 2^{(L+1)R}$ codewords $x^{(0)}, \dots, x^{(M-1)}$, where $x^{(m)} = (x_0^{(m)}, \dots, x_L^{(m)})$, $x_i^{(m)} \in \{\pm 1\}$, and R is the code rate in bits. The DFT of codeword $x^{(m)}$ is defined as

$$X^{(m)}(e^{jw}) \triangleq \sum_{l=0}^L x_l^{(m)} e^{-jlw}. \quad (4.1)$$

Further, we define the partial PSD of $x^{(m)}$ as

$$P_X^{(m)}(e^{jw}) \triangleq |X^{(m)}(e^{jw})|^2. \quad (4.2)$$

Define the bit-reversed (or, mirror) and complement sequences of $x = (x_0, \dots, x_L)$,

respectively, as

$$x_{\text{rev}} = (x_L, \dots, x_0) \quad x_{\text{comp}} = (-x_0, \dots, -x_L). \quad (4.3)$$

Therefore, the DFTs of x_{rev} and x_{comp} are, respectively, $X_{\text{rev}}(e^{jw}) = e^{jLw} X^*(e^{jw})$ and $X_{\text{comp}}(e^{jw}) = -X(e^{jw})$. Consequently, the partial PSDs of x , x_{rev} , and x_{comp} are the same. This helps reduce the code searching complexity, as being shown shortly.

The PSD of a coded sequence obtained by serially concatenating codewords from \mathcal{X} is [32]

$$P_X(e^{jw}) = \frac{1}{M(L+1)} \sum_{m=0}^{M-1} P_X^{(m)}(e^{jw}). \quad (4.4)$$

In order to design a good code, $M/2$ codewords (due to the symmetry we can include their complements) must be chosen jointly out of 2^{L+1} length- $(L+1)$ sequences so that the capacity loss resulting from using $P_X(e^{jw})$ is minimal or, equivalently, $\mathcal{I}(X; Y|P_X)$ is maximized. Since we are interested in low rates (or, equivalently, low SNR), we can further simplify the search for M sequences which result in the highest $\mathcal{I}(X; Y|P_X)$ by applying the approximation $\log(1+x) \approx x$, for small x , to (2.7). With this we obtain the cost function:

$$\begin{aligned} \mathcal{I}(X; Y|P_X) &\approx \frac{1}{2\pi} \int_0^\pi \frac{P_X(e^{jw}) |F(e^{jw})|^2}{\sigma_z^2} dw \\ &= \frac{2}{M(L+1)} \sum_{m=0}^{M/2} \tilde{\mathcal{I}}(X; Y|P_X^{(m)}), \end{aligned} \quad (4.5)$$

where

$$\tilde{\mathcal{I}}(X; Y|P_X^{(m)}) = \frac{1}{2\pi} \int_0^\pi \frac{P_X^{(m)}(e^{jw}) |F(e^{jw})|^2}{\sigma_z^2} dw \quad (4.6)$$

are partial cost functions.

It can be seen from (4.5) that the partial PSD of each sequence linearly con-

tributes to the channel capacity. We can use (4.5) as a guideline for designing the block code. The block code will consist of $M/2$ distinct sequences and $M/2$ complement sequences of them. We want to choose $M/2$ distinct codewords, which are not complements of each other, whose $\tilde{\mathcal{I}}(X; Y|P_X^{(m)})$ is as large as possible. However, since the overall transmitted codebook is obtained by concatenating a small codebook, the chosen codebook should also possess good distance properties in order to provide good BER performance. Let $r^{(m)}$ denote the noiseless channel output when the channel input is $x^{(m)}$. We want the minimum Euclidean distance between any pair of outputs in the set $\{r^{(0)}, \dots, r^{(M-1)}\}$ to be high. Thus any choice of $\{x^{(0)}, \dots, x^{(M-1)}\}$ that simultaneously provides large $\mathcal{I}(X; Y|P_X)$ and results in a high minimum Euclidean distance between the corresponding outputs is a good choice. Since we are usually interested in small values of M , this can be easily done. For the ISI channels investigated in this chapter, codes designed to maximize $\mathcal{I}(X; Y|P_X)$ also result in the highest free Euclidean distance between corresponding channel outputs (among all possible candidates $\{x^{(0)}, \dots, x^{(M-1)}\}$). If this is not the case, we must compute the binary i.i.d. capacities for different codebooks combined with the ISI channel and pick the one which has the highest binary i.i.d. capacity. It should be noted here that the complexity of the combined trellis of the block code and the ISI channel is lower compared to that of the ISI channel alone for the desired working rate region (see Subsection 3). Thus, the search for the best sequences is not computationally intensive. Furthermore, we pick the first pair of codewords to be $\{\pm \text{sign}(\mathbf{f})\}$. This guarantees good spectral matching. The above design procedure can be used to design good block shaping codes; however, no optimality can be claimed. The efficiency of the proposed designed algorithm will be illustrated through the following examples.

Example 1: Block spectral shaping for the $1 \pm D^K$ channel

It is well-known that the $1 \pm D^K$ channel can be considered as K time-interleaved

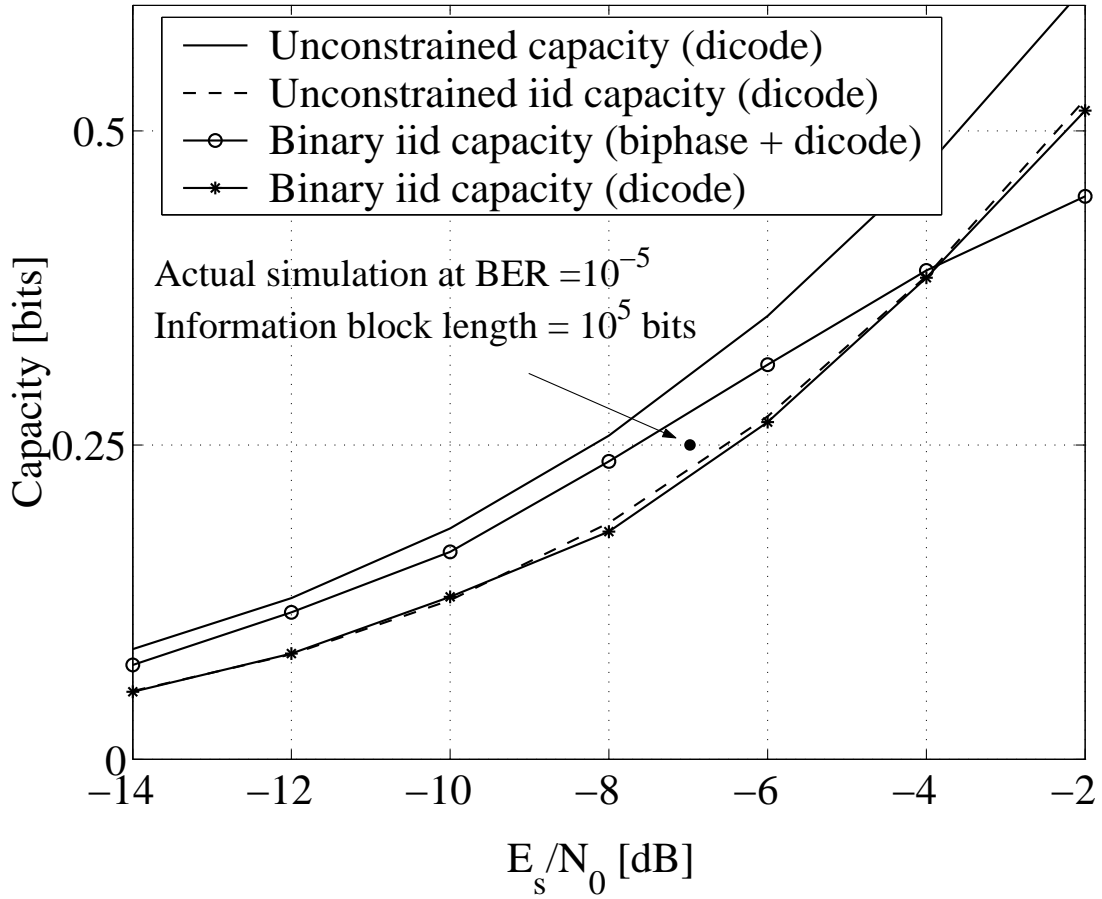


Fig. 15. Capacity for the dicode channel.

$1 \pm D$ channels [29] with independent coding and decoding. Coding for the $1 - D$ channel can be applied to the $1 + D$ channel by complementing the even coded bits. Therefore, we need only to consider the dicode channel. The biphasic code is the simplest rate half block code of length two which maps -1 to $x^{(0)} = (-1, 1)$, and 1 to $x^{(1)} = (1, -1)$. This corresponds to $M = 2$, $R = 1/2$. It is easily seen that the spectrum of the biphasic code is exactly the same as that of the dicode channel [25], which is a good approximation to the water-filling spectrum at low SNR. The i.i.d. capacity and EXIT functions of the combined biphasic code and dicode channel are shown in Fig. 15 and Fig. 16, respectively. We also plot the capacity for the dicode

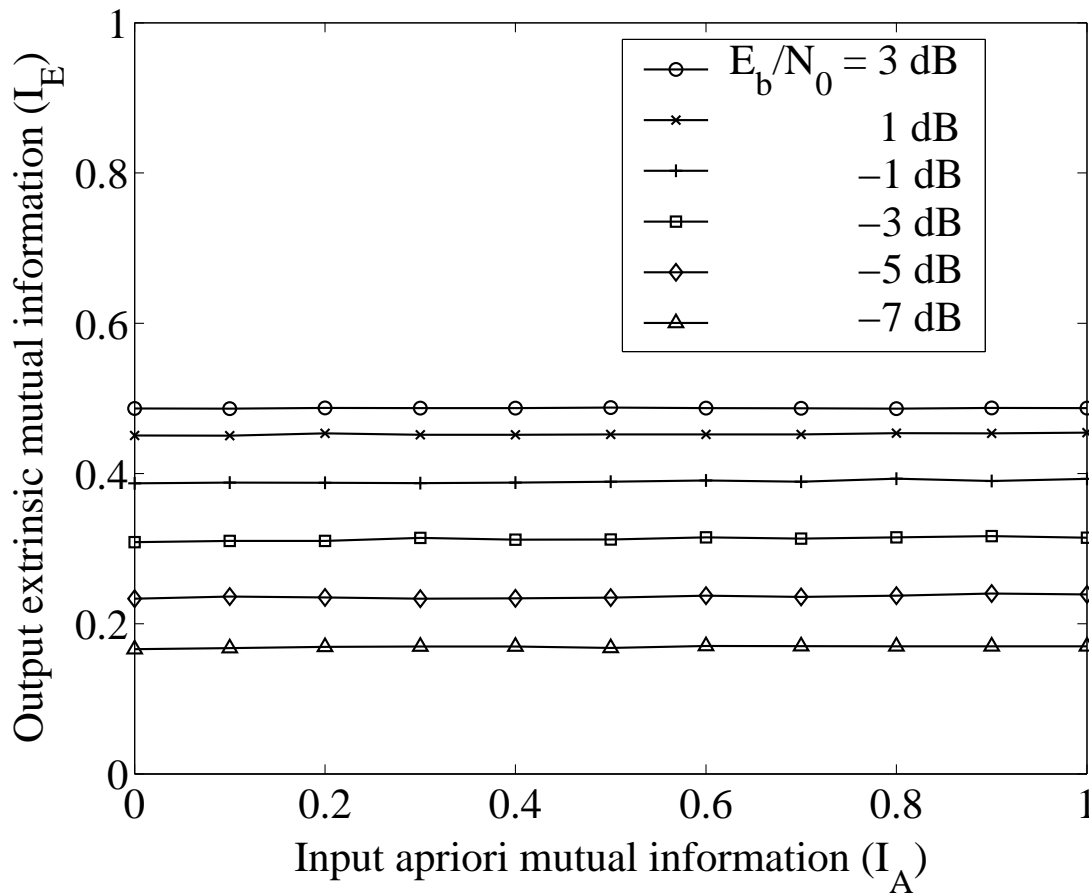


Fig. 16. EXIT functions for the concatenation of biphasic and dicode for different values of E_b/N_0 .

channel for comparison. Two key observations are made from Figs. 15 and 16. First, from the capacity curves, we can see that the combined biphasic and dicode scheme has better i.i.d. capacity than the dicode channel alone when the rate is less than approximately 0.35 bits per channel use. Second, from the EXIT diagram, we can see that the combined biphasic and dicode scheme has almost flat EXIT functions for all values of E_b/N_0 simulated. This means that the combined biphasic and dicode scheme acts as an equivalent AWGN channel. (Note that for the AWGN channel, the EXIT functions are flat.) Thus, we can use any codes that are optimal or near optimal for

the AWGN channel here. We choose to use LDPC codes since they perform very well over the AWGN channel. Others candidates can be irregular repeat accumulate (IRA) codes and turbo codes.

Example 2: Block spectral shaping for the $f_0 + f_1D$ channel with multilevel PAM modulation

The frequency response of this channel is quite similar to that of the $1 - D$ channel when f_0 and f_1 are of different signs; and similar to that of the $1 + D$ channel when f_0 and f_1 are of the same sign. Thus we can use the biphasic coding idea here with a small modification. Length-2 codewords of the rate-half block code $B_{\text{ISI}2}$ are obtained as follows. The first coded bit of a codeword is the same as the input symbol. The second coded bit is the same as the input symbol if f_0 and f_1 have the same sign, otherwise, the negative of the input symbol. Note that $B_{\text{ISI}2}$ has the same codewords as the biphasic code when f_0 and f_1 have opposite signs, and the BPSK modulation is used. Moreover, we need only one bit for the channel state information at the transmitter (to know whether f_0 and f_1 are of the same sign or not). For the binary input case, the i.i.d. capacity and EXIT functions of the combined block code $B_{\text{ISI}2}$ and $f_0 + f_1D$ channel also have similar behaviors as those of the biphasic code over the dicode channel for a large range of channel coefficients f_0 and f_1 . This suggests that we can use codes optimized for the AWGN channel as outer codes to achieve a very good performance. This block code will be used again in the next section (see Example 5).

Example 3: Block spectral shaping for the 4-tap $\mathbf{f} = (0.38, -0.6, -0.6, 0.38)$ ISI channel

Here $L = 3$. We fix the code rate $R = 1/2$ and, therefore, $M = 4$. The four codewords that maximize $\mathcal{I}(X; Y | P_X)$ are $x^{(0)} = (1, -1, -1, 1)$ and $x^{(1)} = -x^{(0)}$, $x^{(2)} = (-1, -1, 1, 1)$ and $x^{(3)} = -x^{(2)}$ (or, $\{9, 6, 3, 12\}$ in decimal representation, re-

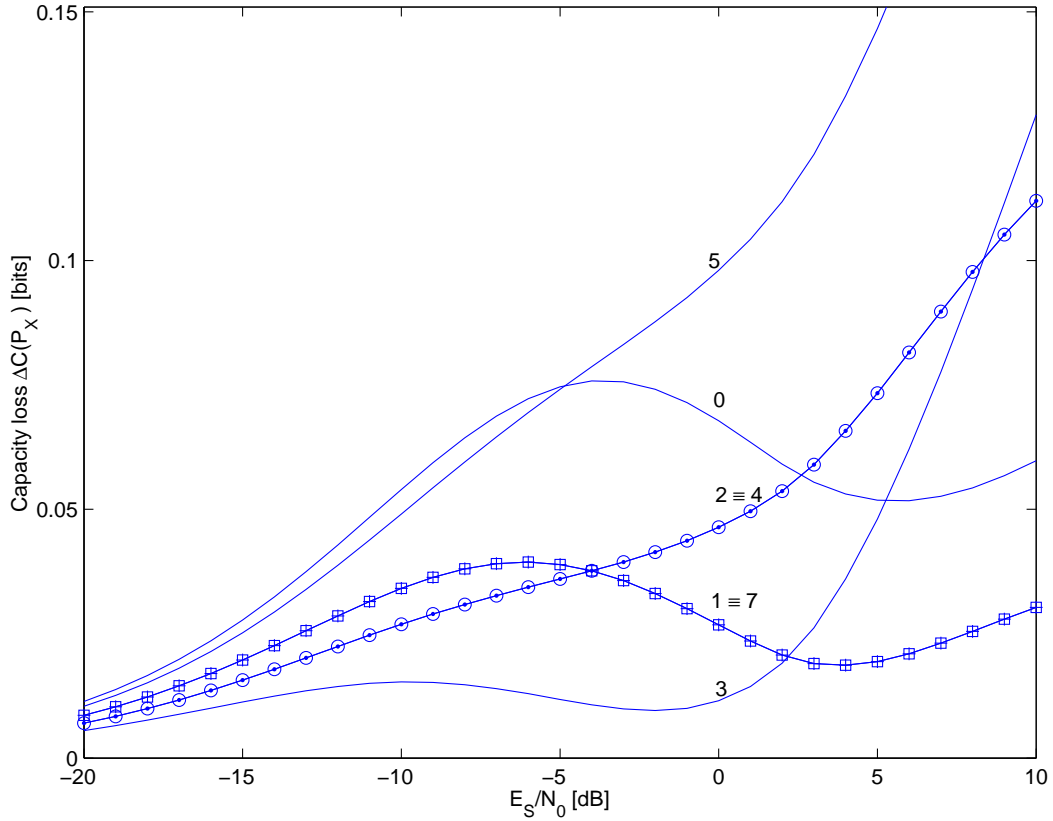


Fig. 17. Capacity loss due to the use of sub-optimal input PSD $P_X(e^{jw})$ as opposed to the optimal water-filling PSD $P_{X,\text{opt}}(e^{jw})$. The first 2 codewords are 6 and 9 (in decimal representation); the third codeword is numbered near each curve; and the fourth codeword is the complement of the third codeword.

spectively). We plot the capacity loss due to the use of non-optimal input PSD $P_X(e^{jw})$ as opposed to the optimal water-filling PSD $P_{X,\text{opt}}(e^{jw})$. In Fig. 17, we plot the i.i.d capacity of the combined block code and channel by fixing two codewords of the block code to be $(1, -1, -1, 1) = \text{sign}(\mathbf{f})$ and $(-1, 1, 1, -1) = -\text{sign}(\mathbf{f})$, and by varying the other two. The third codeword (decimal representation) is numbered near each curve; and the fourth codeword is the complement of the third codeword. Clearly, in the low rate region, the designed codebook $\mathcal{X} = \{9, 6, 3, 12\}$ results in the least loss in rate. This code also results in the highest minimum (squared) Euclidean

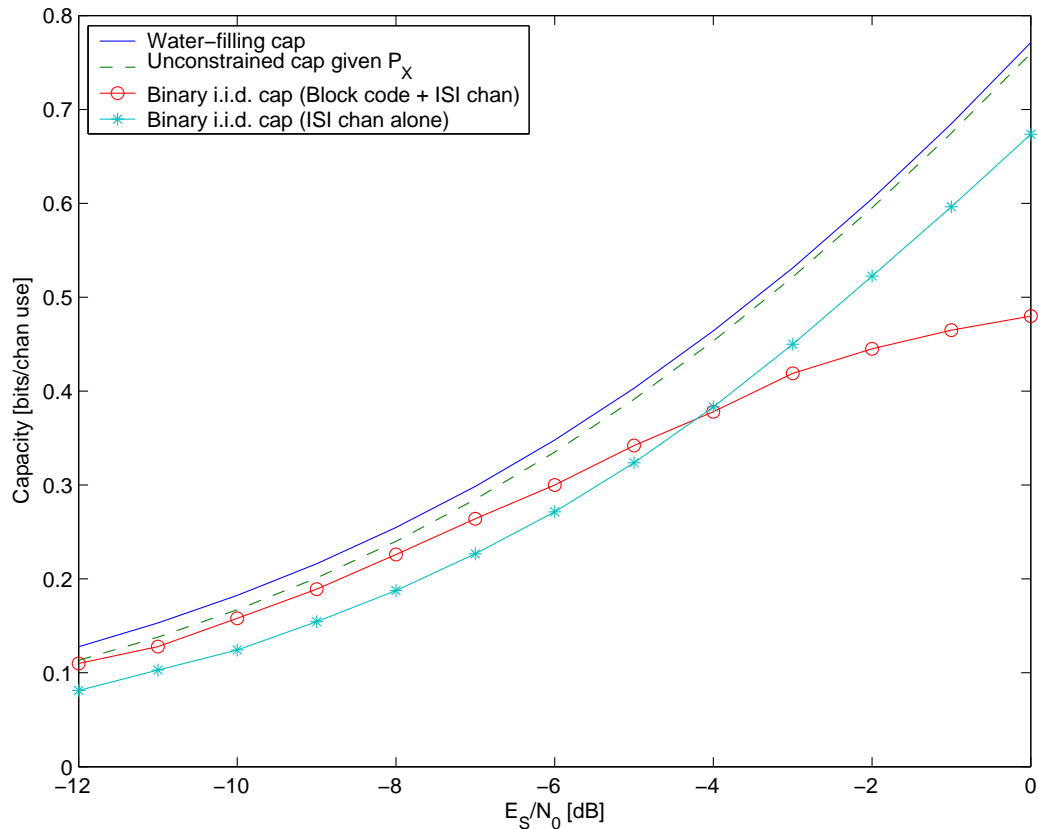


Fig. 18. Capacity for the $0.38 - 0.6D - 0.6D^2 + 0.38D^3$ ISI channel.

distance between corresponding outputs, which is equal to 8.4544. To show that the proposed procedure results in good codes, in Fig. 18, the i.i.d capacity with the block code is shown in comparison to the i.i.d capacity of the ISI channel without the code. It can be seen that for code rates below 0.35, a significant coding gain is obtained. At low rates, the constrained i.i.d capacity is close to the unconstrained capacity with the same input PSD $P_X(e^{jw})$.

2. Threshold

To show that the proposed scheme performs well, we compute the threshold for the LDPC code with the spectral shaping inner code. As an example, we consider the 2-

tap channel with the impulse response $0.8 + 0.6D$. The inner code is the binary block code B_{ISI2} as mentioned in Example 2. Since the combination of this spectral shaping inner code and the 2-tap $0.8 + 0.6D$ ISI channel behaves as an equivalent AWGN (with flat EXIT functions), coding techniques used for the AWGN channel can be directly applied here. The outer code is an irregular LDPC code of rate $1/2$ optimized for AWGN channel with left and right degree profiles given by $\lambda(x) = 0.23403x + 0.21242x^2 + 0.1469x^5 + 0.10284x^6 + 0.30381x^{19}$ and $\rho(x) = 0.71875x^7 + 0.28125x^8$, taken from [33]. Similar degree profiles can also be obtained by a technique proposed in [19]. The overall rate is $1/4$. The i.i.d. capacity in E_b/\mathcal{N}_0 at rate $1/4$ is -0.5 dB before shaping and -1.5 dB after shaping. Thus, the upper limit on the achievable improvement in E_b/\mathcal{N}_0 for the serially concatenated scheme is 1 dB. By using density evolution, the threshold of the proposed scheme was found to be -1.2 dB. We also use Monte-Carlo simulation to verify the performance of the proposed scheme. The simulated E_b/\mathcal{N}_0 required to get the BER of 10^{-5} is -1.0 dB for an information block length of 10^5 bits. It can be seen that the proposed scheme outperforms the i.i.d. capacity by 0.5 dB and is only 0.5 dB away from the achievable limit.

3. Trellis Complexity

One important property of the spectral shaping block code is that when combined with an $(L + 1)$ -tap ISI channel, the total number of states in the joint trellis is equal to the codebook size $M = 2^{(L+1)R}$ as opposed to 2^L as the number of states of the ISI channel alone. The rate region such that the decoding complexity of the combined block code and ISI channel is lower than that of the channel without block coding satisfies the following equation

$$2^L \cdot 2 \geq 2^{(L+1)R} \cdot 2 \cdot \frac{1}{R} \Leftrightarrow L \geq (L + 1)R - \log(R).$$

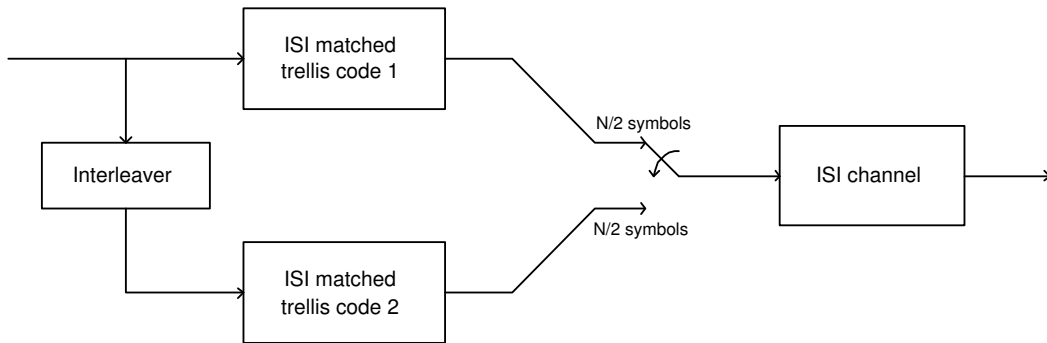


Fig. 19. The parallel concatenation for the ISI channel.

C. Parallel Concatenation

In the previous section, we proposed the serially concatenated scheme for use in ISI channels. In practice, the parallel concatenated scheme is also widely used. In this section, we focus on the parallel scheme. The proposed scheme is shown in Fig. 19. It is a parallel concatenation of two ISI-matched trellis component codes which are designed such that their PSDs are matched to the water-filling PSD of the channel, and that the output minimum Euclidean distance is as large as possible. The component encoders are in general non-systematic. The transmission of the coded signals to the channel needs to be slightly modified compared to the conventional way done in turbo codes. The coded signals coming out of the first (upper) ISI-matched trellis component code are first transmitted block-wise. The second (lower) ISI-matched trellis component code starts to transmit only after the upper code has finished its transmission. That is, bit-wise multiplexing [1] is not used here. With this schedule, the approach can be applied to any ISI channel, and is not restricted only to the $1 - D^2$ channel as in [31]. A similar transmission schedule was proposed in [34], [35], but there, only turbo decoding was considered. The overall receiver is an iterative decoder which performs iterative decoding similar to that in turbo codes.

Here, each component decoder works on the combined trellis of the ISI-matched trellis code and the ISI channel.

The code trellis is built to satisfy the following conditions:

Condition 1: The outputs along the branch transitions are chosen from the set of block codewords to satisfy the spectrum property (discussed in the previous section).

Condition 2: The minimum Euclidean distance at the noiseless channel outputs is as large as possible [29], [30].

Condition 3: The trellis represents a recursive encoder.

It should be noted here that for the case of ISI channels, recursiveness means that when a channel input error event is of weight 1, the corresponding noiseless channel output error event is of infinite weight with probability 1 (with respect to the length of the channel input error event). For further discussion on this we refer the reader to [34].

Condition 1 results in a shaping gain and Condition 2 results in a coding gain. Note that we will suffer from a rate loss for Conditions 1 and 2 to be satisfied. If the rate loss is smaller than the coding/shaping gain, then the combined trellis code and ISI channel scheme will have a higher i.i.d. capacity than that of the original ISI channel.

It can be shown that [30] Conditions 1 and 2 can be satisfied by the serial concatenation of a trellis code designed by set partitioning and a spectral shaping block code. A similar result can be found in [36]. We will not discuss set partitioning in the chapter. The reader is referred to [29] and [30] for more details on this. Condition 3 is satisfied by choosing an appropriate mapping of the input bits to the trellis branches. However, to obtain good codes for use in the parallel concatenation, it is important to optimize this mapping. We will use EXIT functions to find suitable input mappings.

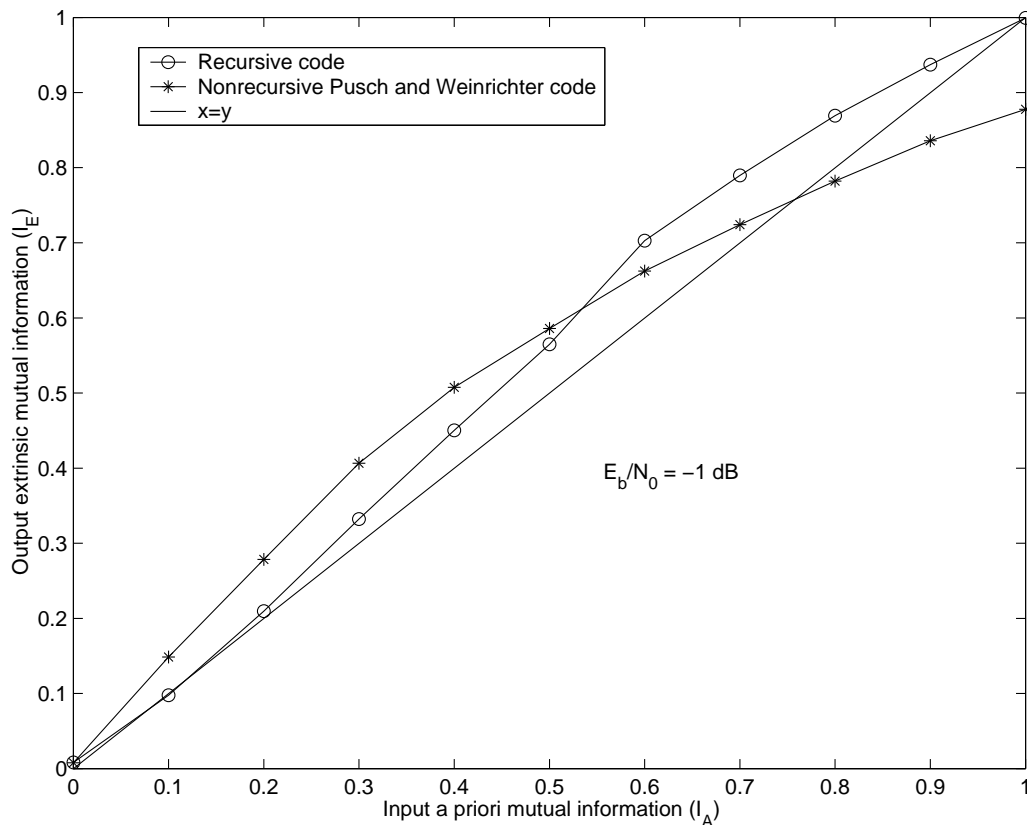


Fig. 20. EXIT functions of the rate half Pusch and Weinrichter’s code used in the parallel concatenated scheme of rate 1/4 over the $0.8 + 0.6D$ ISI channel.

To demonstrate the idea, we present two simple examples.

Example 4:

The ISI-matched trellis component codes are taken from [36] with a slight modification to make the trellis recursive. It was also pointed out in [31] that the non-recursive codes introduced in [36] would have poor performance when used in a parallel concatenated scheme. The trellis code consists of 4 states. From the EXIT functions (see Fig. 20), we see that this code should not be used in a parallel concatenated scheme. It is because when there is no input a priori information, the output extrinsic information is very low. Thus, the iteration will get stuck from the beginning.

Example 5:

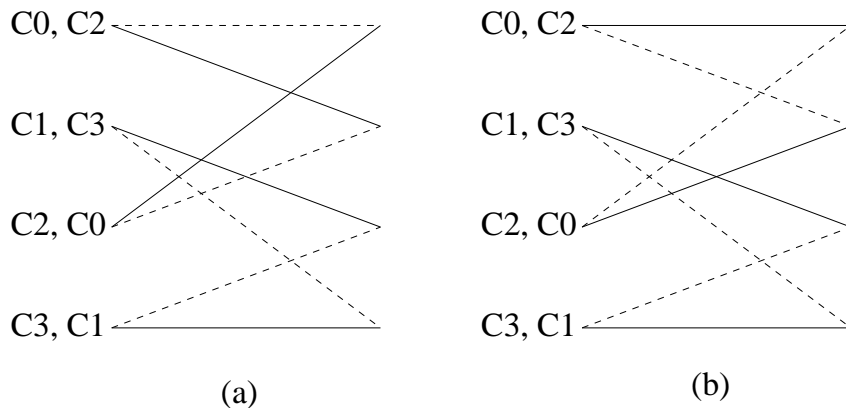


Fig. 21. Different input mappings for $f_0 + f_1 D$ channels when $\text{sign}(f_0 f_1) = -1$ (dashed lines: input bit is equal to 1; solid lines: input bit is equal to 0) (a) Good mapping; (b) Bad mapping.

The example is adopted from [30] with a modification that the trellis encoder is made recursive. However, it should be noted here that direct application of the techniques in [29] and [30] generally will not result in good component codes for use in the parallel concatenation. The main problem we have to solve here is to make the joint trellis of the ISI-matched trellis component code and the ISI channel recursive so that the interleaving gain can be achieved. It should be noted here that the optimal input mapping (with respect to the decoding threshold) of the trellis encoder depends on the channel state information. Fig. 21 shows two different mappings of information bits. The trellises of the combined 4-state 4-PAM recursive trellis code, 4-PAM block code $B_{\text{ISI}2}$ (as in Example 2), and $f_0 + f_1 D$ channel (or RTBC for short) with two different mappings are given in Table I.

Table I. State transition for the combined trellis code and the $1-D$ channel with good and bad input mappings

Start state	Input k -tuple (Good)	Input k -tuple (Bad)	Output n -tuple	End state	Noiseless channel output
0	0	1	-3 3	3	-2 6
0	1	0	1 -1	0	2 -2
1	0	1	-3 3	3	-6 6
1	1	0	1 -1	0	-2 -2
2	0	0	3 -3	4	4 -6
2	1	1	-1 1	7	0 2
3	0	0	-1 1	4	0 -6
3	1	1	3 -3	7	-4 2
4	0	1	-3 3	1	0 6
4	1	0	1 -1	2	4 -2
5	0	1	-3 3	1	-4 6
5	1	0	1 -1	2	0 -2
6	0	0	3 -3	6	6 -6
6	1	1	-1 1	5	2 2
7	0	0	3 -3	6	2 -6
7	1	1	-1 1	5	-2 2

1. Simulation Results

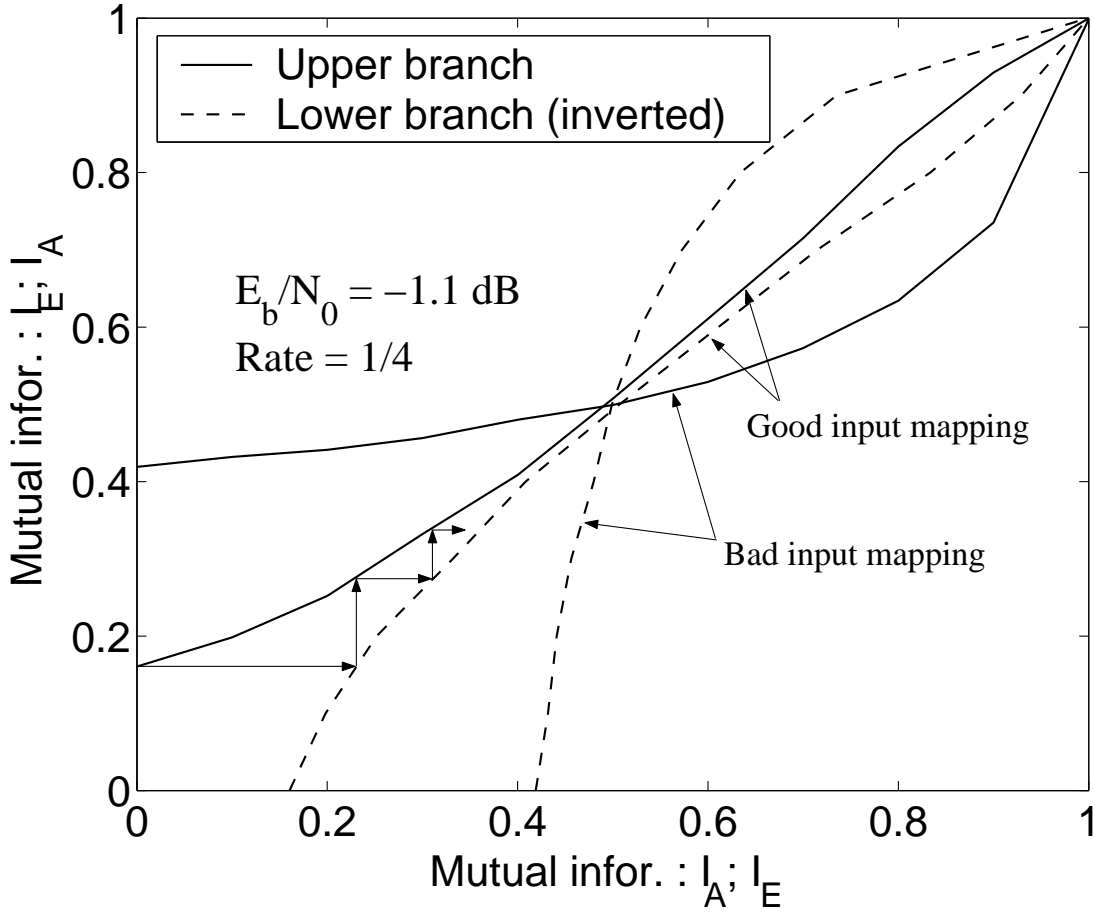


Fig. 22. EXIT functions of the parallel concatenation for the $1 - D$ channel.

An interesting feature about the EXIT function of the RTBC scheme (see Fig. 22) is that the output extrinsic mutual information reaches value 1 when the input a priori mutual information is 1. Note that this property is required for a parallel concatenated scheme to have a finite threshold; and hence this scheme is suitable for the parallel concatenation. As seen, the threshold for this scheme is about $E_b/N_0 = -1.15$ dB, which offers 0.75 dB improvement compared to the case without spectral shaping. The i.i.d. capacity of the dicode channel without spectral shaping is $E_b/N_0 = -0.4$

dB at rate 1/4. Monte-Carlo BER simulation for short block length also verifies the usefulness of the scheme (see Fig. 23). At BER of 10^{-5} and interleaver length of 1000 bits, the performance of the new scheme is approximately 0.3 dB better than the scheme proposed by Pusch et al. [31]. The decoder for each branch (upper and lower RTBCs) has only 8 states compared to 24 states of the scheme in [31].

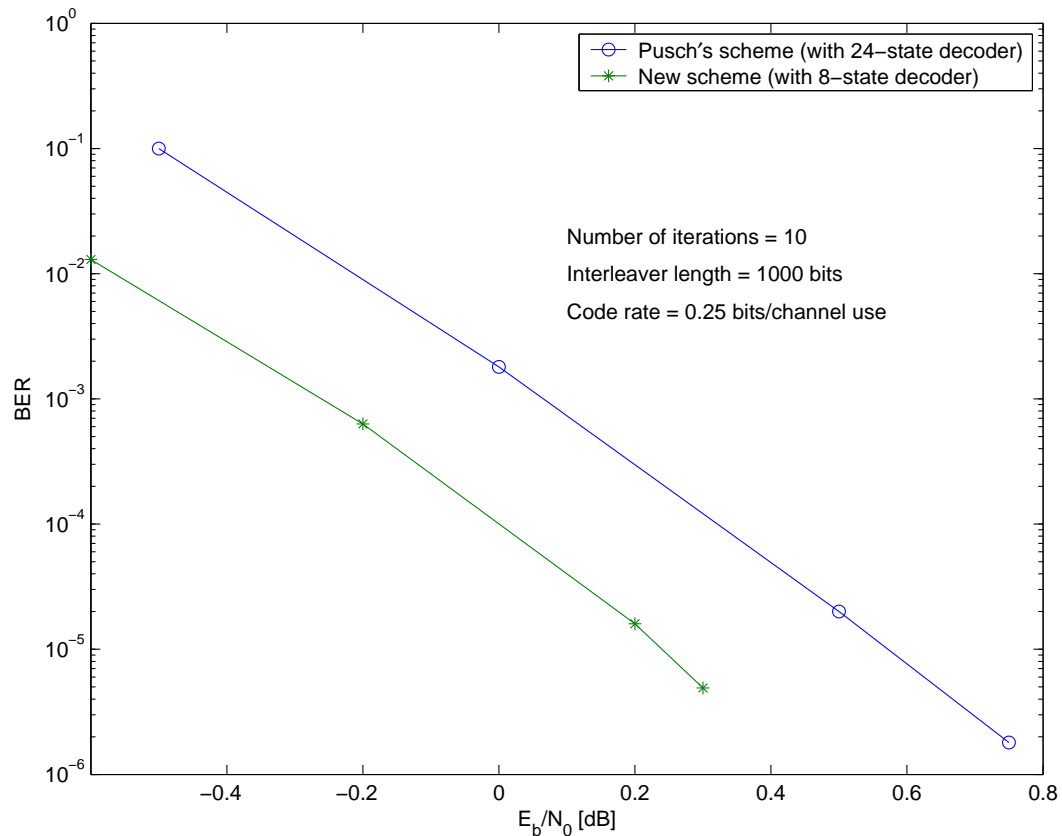


Fig. 23. Bit error rate of the new parallel concatenation for the $1 - D$ channel.

D. Conclusion

We have presented different simple coding techniques for ISI channels which perform well in the medium to low rate region. The basic idea is to use a spectral shaping block code in order to shape the spectrum of the transmitted signals to the desired

water-filling spectrum. We showed that the spectral shaping can be combined directly with the trellis of the component encoders in order to build parallel concatenated schemes. The proposed schemes were shown to provide performance better than the i.i.d. capacity and to perform only with a few tenths of a dB away from the capacity limit. Significant performance gains were shown for short block lengths.

CHAPTER V

ITERATIVE PACKET COMBINING SCHEMES FOR INTER-SYMBOL
INTERFERENCE CHANNELS

In this chapter, we extend results on precoding in Chapter III to automatic repeat request (ARQ) schemes. We study packet combining techniques for retransmission schemes over ISI channels. Two types of combining schemes are investigated, namely, maximum likelihood combining (MLC) and iterative combining (IC). By first employing a precoding technique and then by interpreting the ISI channel as a trellis code, the transmissions of the same data packet at different times through the channel can be treated as the parallel concatenation of recursive trellis codes. If interleavers are used in between retransmissions, ‘turbo’ coding gains can be achieved by iterative equalization. It is shown that IC provides excellent performance and outperforms other forms of combining in terms of frame error rate (FER) performance both analytically and through simulations.

A. Introduction

In communication systems, there are two fundamental techniques used for error control: Forward error correction (FEC) and automatic repeat request (ARQ). In FEC schemes, the receivers have the ability to correct erroneous data, while in pure ARQ schemes, the receivers can only detect if there are errors in the transmission. Automatic repeat request schemes are widely used in data communication where transmission delay is acceptable. In Type I ARQ schemes, uncoded data is transmitted in packets and when an erroneous packet is detected, the receiver can request the transmitter to re-send the same packet of data until no error is detected. Hybrid ARQ schemes (Type II or Type III ARQ schemes) which combine FEC and ARQ are

also commonly used [37].

It is well known that the throughput of an ARQ scheme can be improved by using packet combining. In packet combining, all received copies of the data packet are combined to form a more reliable estimate of the transmitted data. In Chase's combining scheme [38], a codeword from a code of rate R is repeated K times to form a codeword from a lower rate code of rate R/K . Each received copy is weighted by its corresponding reliability, which depends on the instantaneous gain of the channel. Cyclic redundancy check (CRC) bits usually form an error detection code and are used by the receiver to determine if the packet has been decoded correctly. In [39], a version of packet combining was proposed in which K copies of a data packet are combined into a single packet of the same length as that of the original transmitted data packet by averaging the soft decision values from the constituent copies. It was shown that for AWGN channels, this scheme is equivalent to *Maximum Likelihood* (ML) diversity combining.

Since packet data systems can tolerate the latency due to a large interleaver size, turbo codes are ideally suited for packet data transmission. Although application of the turbo principle to coded systems over ISI channels has recently been investigated by some authors [12], [13], its performance in ARQ schemes has not been studied in detail. Balachandran and Anderson [40] proposed a turbo-like scheme where component encoders of the original turbo code were replaced by two equivalent discrete-time ISI channels. The decoder is an iterative decoder that iteratively decodes (or, equalizes) the received signals to form an estimate of the data. The performance of this parallel scheme is improved only marginally compared to using a Viterbi decoder to jointly decode both the transmissions, when perfect channel state information (CSI) is available at the receiver. In [41], a low complexity linear iterative turbo-equalization scheme for dual ISI channels was investigated. Similar to this chapter, [41] also ad-

dresses turbo equalization of ISI channels alone (without an outer code) and, hence, the results in [41] are applicable to the special case of one retransmission. However, in contrast to this chapter, [41] does not consider any form of precoding. As shown in this chapter, this precoding is essential to obtain an interleaving gain.

The focus of this chapter is in designing packet combining schemes for combining multiple retransmissions through ISI channels in order to reduce the FER (or, equivalently, the number of retransmissions required to successfully decode a packet). We first show that the poor performance of the parallel concatenation scheme in [40] is because it is a parallel concatenation of two non-recursive codes (ISI channels). We use a technique recently proposed in [15] in the context of serial concatenation to make the ISI channel appear recursive. With this technique, multiple transmissions through the ISI channel become parallel concatenation of *recursive* codes. Hence, multiple transmissions of uncoded (or, coded) data through ISI channels can be iteratively combined to result in significantly lower FER than other forms of packet combining schemes. It is shown both analytically and through simulations that IC when used with precoding results in excellent FER and outperforms the scheme in [40], without increasing the complexity or decreasing the data rate. Moreover, the IC schemes require very little modification at the transmitter, and are a promising solution to systems that enjoy low cost and easy-to-implement terminals.

The remainder of the chapter is organized as follows. In Section B, we present the system model. In Section C, we study ML combining and IC schemes. In Section D, we present some important properties of precoded ISI channels and derive upper performance bounds based on the uniform interleaver concept for the IC schemes. We provide some results and discussion in Section E. Finally, we conclude the chapter in Section F.

B. System Model

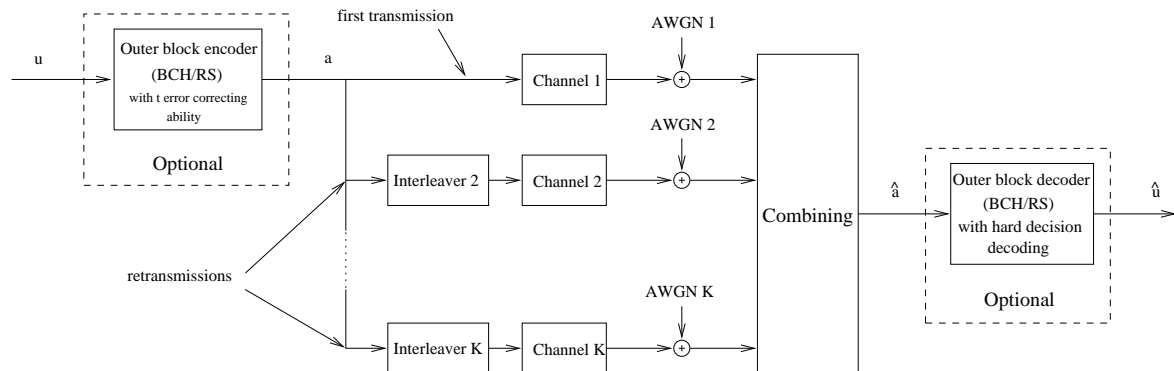


Fig. 24. Retransmissions of the same data packet over the channel can be viewed as a parallel concatenation of K trellis codes.

The system model is shown in Fig. 24. A block of N bits of data is encoded by an optional outer code and then modulated and transmitted K times over a channel corrupted by ISI and AWGN. For ease of exposition, we assume that the modulation is binary phase shift keying (BPSK), although the proposed techniques are applicable to any memoryless modulation. The receive filter is a whitened matched filter corresponding to the channel impulse response $h(t)$. The output samples of the equivalent discrete-time ISI channel can be expressed as

$$y_n = r_n + z_n, \quad 0 \leq n < N \quad (5.1)$$

where $r_n = \sum_{l=0}^L (2x_{n-l} - 1)f_l$ (for BPSK), $x_t = 1/2$ for $t < 0$, is the noiseless channel output, $\mathbf{f} = (f_0, f_1, \dots, f_L)$ is the equivalent discrete-time channel coefficient vector, L is the channel memory, and $\mathbf{z} = (z_0, z_1, \dots, z_{N-1})$ is a sequence of independent identically distributed (i.i.d.) Gaussian noise samples with zero mean and variance $K\mathcal{N}_0/(2E_b)$, where E_b is the average energy per bit. The input to the channel $\mathbf{x} = (x_0, x_1, \dots, x_{N-1})$ is a sequence of i.i.d. input data with $x_n \in \{0, 1\}$ in the

absence of an outer code, or \mathbf{x} is a codeword from an outer code. The outer code can be a convolutional code which admits easy soft-decision decoding or a block code such as a Bose Chaudhuri Hocquenghem (BCH) or a Reed-Solomon (RS) code, which is typically decoded using a hard decision algebraic decoder. In the most general case, the channel can be time-variant, with the static (time-invariant) channel being a special case. At the receiver, an equalizer is implemented by using the Viterbi algorithm (VA) for MLC schemes while the outer code is decoded using a hard decision decoder. When considering IC, the Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm [8] is used for equalization. The equalizer output is then fed to the outer decoder (if the outer code is used).

C. Combining Techniques

During the first transmission, the received packet is equalized and decoded using the outer code and then checked for errors using an error detection code. A CRC or the outer code itself may serve as the error detection code. When the packet is determined to be in error, a retransmission is requested. What is retransmitted during the retransmissions and how the retransmissions are combined depend on the exact combining technique used at the receiver. Here, we consider two types of combining schemes which will be discussed in the following Subsections. It should be emphasized that the focus of this chapter is on techniques to combine the retransmissions for the purpose of equalization only, that is, the packet combining techniques themselves do not take advantage of the outer code, if it is present. The decoding is done after the packets are combined. Therefore, the presence or absence of the outer code does not have a significant impact on the performance of the combining schemes. Extension to the case of combining using the outer code is straightforward, though usually it is

computationally complex.

1. ML Combining (MLC)

Here, the retransmissions are identical to the original transmission and the receiver uses the VA to combine the retransmissions using an ML rule. MLC uses the VA to search the trellis path that has minimum metric. For ISI channels, the path metric for a single received copy of the transmitted packet \mathbf{x} is given below

$$M_{\text{single}} = \sum_{n=0}^{N-1} (y_n - \hat{r}_n)^2 \quad (5.2)$$

$$\equiv \sum_{n=0}^{N-1} (\hat{r}_n^2 - 2\hat{r}_n y_n) \quad (5.3)$$

where $\mathbf{y} = (y_0, y_1, \dots, y_{N-1})$ is the received signal and $\hat{r}_n = \sum_{l=0}^L (2\hat{x}_{n-l} - 1)f_l$, where \hat{x}_n is the tentatively estimated version of the transmitted data x_n . In (5.3) we have dropped the term $\sum_{n=0}^{N-1} y_n^2$ since it is the same for all possible transmitted packets. Suppose that K copies of the transmitted packet \mathbf{x} have been received, say $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K$. It can be easily shown that when the ISI channel is time-invariant, the branch metric at time k is

$$\lambda_n = \sum_{r=1}^K (\hat{r}_n^2 - 2\hat{r}_n y_{r,n}) \quad (5.4)$$

where $y_{r,n}$ is the received signal at time k of the r -th copy. Correspondingly, the ML path metric for combining K received copies is

$$M_{\text{MLC}} = \sum_{n=0}^{N-1} (\hat{r}_n^2 - 2\hat{r}_n \bar{y}_n) \quad (5.5)$$

where

$$\bar{y}_n = \frac{1}{K} \sum_{r=1}^K y_{r,n}.$$

From (5.3) and (5.5), it is clear that for ISI channels, the ML path metric for multiple received copy combining has a form similar to that for single received copy combining.

2. Iterative Combining (IC)

In the case of IC, during a retransmission, an interleaver is placed at the input to the ISI channel as shown in Fig. 24. Different interleavers are used for different retransmissions. This is the only modification that is required at the transmitter. We begin by noting that, the equivalent discrete-time ISI channel can be treated as a rate-one *feed-forward* convolutional encoder with constraint length $(L + 1)$. The ISI channel is characterized by the the polynomial

$$f(D) = \sum_{l=0}^L f_l D^l \quad (5.6)$$

and is normalized so that $\sum_{l=0}^L f_l^2 = 1$ if the channel is time-invariant. Multiple retransmissions through the ISI channel can be considered as a concatenation of convolutional codes. When an interleaver is used in between the transmissions, it represents a parallel concatenated turbo code where the ISI channels are the component codes. Multiple transmissions are combined using a turbo decoder that iteratively forms estimates for the transmitted data packet \mathbf{x} based on the received copies $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K$. The receiver consists of K component soft-input soft-output (SISO) decoders. In order to fully make use of available soft information from the component SISO decoders, when $K > 2$, the parallel structure [42] is used in iterative decoding. During the completion of the $(m + 1)$ th iteration, the soft information (extrinsic information) produced by the i -th SISO is given by

$$L_{\text{ext},i}^{(m+1)}(x_n) = \mathcal{G}\left(\mathbf{y}_i, L_{\text{ext},j}^{(m)}(x_l); 1 \leq j \leq K, j \neq i; 0 \leq l < N\right),$$

$$1 \leq i \leq K, \quad 0 \leq n < N$$

where \mathcal{G} is the MAP decoding function which can be implemented using either the BCJR algorithm or the SOVA or any variation of these algorithms. In this chapter, we consider only the Max-Log-MAP implementation [43].

D. Distance-Spectrum Based Properties of Precoded ISI Channels

It is more convenient to use the D -transform notation for the sequences. We define the following (See Chapter III for more details on precoding.)

- Channel input: $x(D) = x_0 \oplus x_1 D \oplus x_2 D^2 \oplus \dots$

- Precoder output: $d(D) = d_0 \oplus d_1 D \oplus d_2 D^2 \oplus \dots$

- Noiseless channel output:

$$r(D) = r_0 + r_1 D + r_2 D^2 + \dots = \gamma(d^{(0)}) + \gamma(d^{(1)})D + \gamma(d^{(2)})D^2 + \dots$$

- Modulated signal: $\mu(d(D)) = \mu(d_0) + \mu(d_1)D + \mu(d_2)D^2 + \dots$

We now discuss some important properties of precoded ISI channels in the light of distance-spectrum based analysis.

1. Recursive Property

Property 1: When the input error sequence to channel is of weight one, the output channel error sequence is of infinite weight.

Proof: It is clear from (3.4) that $x(D) = d(D)h(D)$, or $d(D) = x(D)/h(D)$. We now consider two sequences at the input to the precoder $x_1(D)$ and $x_2(D)$. Correspondingly, we have the precoded output sequences $d_1(D) = x_1(D)/h(D)$ and $d_2(D) = x_2(D)/h(D)$. Without including AWGN, the received sequences corresponding to $d_1(D)$ and $d_2(D)$ are $r_1(D)$ and $r_2(D)$, respectively. The squared Euclidean

distance between these two received sequences is

$$\begin{aligned}
\delta^2(r_1(D), r_2(D)) &= \|\varepsilon_r(D)\|^2 = \sum_{n=0}^{N-1} |(r_{2,k} - r_{1,k})|^2 \\
&= \sum_{n=0}^{N-1} \left| \gamma(d_2^{(n)}) - \gamma(d_1^{(n)}) \right|^2 \\
&= \|f(D)[\mu(d_2(D)) - \mu(d_1(D))]\|^2 \\
&= \|f(D)\varepsilon_\mu(d(D))\|^2
\end{aligned}$$

where $\varepsilon_r(D) = r_2(D) - r_1(D)$ and $\varepsilon_\mu(d(D)) = \mu(d_2(D)) - \mu(d_1(D))$. Let us consider an input sequence $x_1(D)$ and another input sequence $x_2(D) = x_1(D) \oplus D^n$, for some $0 \leq n < N$. This corresponds to an input error event $\varepsilon_x(D) = x_1(D) \oplus x_2(D) = D^n$. The error sequence at the output of the precoder $\varepsilon_d(D) = d_1(D) \oplus d_2(D) = D^n/h(D)$ contains an infinite number of non-zero coefficients when $N \rightarrow \infty$. In other words, $d_1(D)$ and $d_2(D)$ differ in infinitely many positions and consequently $\varepsilon_\mu(d(D))$ is an infinite weight sequence. Though $\varepsilon_\mu(d(D))$ is of infinite weight, there are some cases that the sequence $f(D)\varepsilon_\mu(d(D))$ is of finite weight. Equivalently, there is no guarantee that the squared Euclidean distance $\delta^2(r_1(D), r_2(D))$ is of infinite weight for a *particular* sequence $\varepsilon_\mu(d(D))$. We can easily verify this claim by considering channels which have nulls, such as a 2-tap channel with coefficients $f_0 = -f_1$. In this case, when the error sequence $\varepsilon_\mu(d(D)) = 2 + 2D + 2D^2 + \dots$ or $\varepsilon_\mu(d(D)) = -2 - 2D - 2D^2 - \dots$ (for BPSK), we have $\delta^2(r_1(D), r_2(D)) = 4f_0^2 + 4f_1^2 = 4$. Fortunately, the probability of appearance of an infinite weight sequence $\varepsilon_\mu(d(D))$ such that $\delta^2(r_1(D), r_2(D)) = \|f(D)\varepsilon_\mu(d(D))\|^2$ is of finite weight is asymptotically (in length N) zero. Therefore, when we consider the average of $\delta^2(r_1(D), r_2(D))$ over all possible infinite weight error sequences $\varepsilon_\mu(d(D))$, $\mathbb{E}_{\varepsilon_\mu}[\delta^2(r_1(D), r_2(D))]$ is of infinite weight. \square

We use this recursive property of precoded ISI channels in our IC ARQ scheme.

The retransmission schemes and packet combining techniques for precoded ISI channels are similar to those discussed in Section C, except that at the receiver, the precoded channel is treated as the actual ISI channel.

2. Frame Error Rate Property

Property 2: When the input to the channel $x(D)$ is a sequence of i.i.d. random variables, the probability of a frame error is the same for both precoded and non-precoded ISI channels when ML decoding is used. Therefore, for Type I or Type II ARQ schemes, the probability of requesting a retransmission after the first transmission is unaffected due to precoding.

Proof: Consider an error event ε . For the non-precoded ISI channel, corresponding to the error event ε , let $x(D)$, $\hat{x}(D)$, $r_{\text{np}}(D)$, and $\hat{r}_{\text{np}}(D)$ be the input sequence, estimated input sequence, noiseless channel output sequence, and estimated noiseless channel output sequence, respectively. We have

$$\varepsilon_x(D) = x(D) \oplus \hat{x}(D) \quad (5.7)$$

as the input error sequence associated with ε and

$$\varepsilon_{y,\text{np}}(D) = \hat{r}_{\text{np}}(D) - r_{\text{np}}(D) \quad (5.8)$$

as the channel output error sequence associated with ε . We have

$$\varepsilon_{y,\text{np}}(D) = \varepsilon_\mu(x(D))f(D) \quad (5.9)$$

where $\varepsilon_\mu(x(D)) = \mu(\hat{x}(D)) - \mu(x(D))$. The distance spectrum of the non-precoded ISI channel is defined as the set Δ_{np} that contains all possible values of squared

Euclidean error distances $\Delta_{\text{np}}^2(\varepsilon)$ defined as

$$\Delta_{\text{np}}^2(\varepsilon) = \|\varepsilon_{y,\text{np}}(D)\|^2 = \|\varepsilon_\mu(x(D))f(D)\|^2, \quad \forall x(D) \neq \hat{x}(D).$$

For the precoded ISI channel, the input error sequence $\varepsilon_x(D)h(D)$ to the precoded channel produces the channel output error sequence

$$\varepsilon_{y,\text{p}}(D) = \varepsilon_\mu(x(D)h(D)/h(D))f(D) = \varepsilon_{y,\text{np}}(D). \quad (5.10)$$

Since the input sequence $x(D)$ has the i.i.d. property, $\varepsilon_x(D)$ is i.i.d. and, hence, $\varepsilon_x(D)h(D)$ is also i.i.d.. This means that the distance spectrum of the precoded ISI channel, denoted by Δ_{p} , is the same as that of the non-precoded ISI channel, Δ_{np} .¹ Thus the FER of the precoded ISI channel is the same as that of the non-precoded ISI channel. \square

3. Interleaving Gain Property for Non-fading Precoded ISI Channels

One of the important parameters for ARQ schemes is the expected number of transmissions that must be made before a packet is accepted, denoted by T_r . It is well known that for packet combining schemes, T_r is upper and lower bounded by [39]

$$1 + \sum_{n=1}^{\infty} \prod_{i=1}^K P(E_i) \leq T_r \leq 1 + \sum_{n=1}^{\infty} P(E_n) \quad (5.11)$$

where $P(E_n)$ is the probability of a retransmission request is made while decoding the packet formed by combining K received copies. We can interpret $P(E_n)$ as the probability of an error event after combining K received copies of the data packet given that the combining of the previous $(K - 1)$ copies was in error. In this Section, we focus on the calculation of $P(E_n)$ for IC schemes by assuming ML decoding. We

¹When the length of the packet N is large, we can ignore the trellis terminating effect.

show the following property.

Property 3: For large E_b/N_o , the union bound on $P(E_n)$ denoted by $P_U(E_n)$ satisfies

$$P_U(E_n) \propto \begin{cases} N, & \text{for non-precoded ISI channels} \\ N^{2-K}, & \text{for precoded ISI channels.} \end{cases}$$

Proof: We begin by noting that the mapping function in (3.2) is non-linear. Thus in the evaluation of the error performance for an ISI channel, we cannot assume that the transmitted (correct) sequence is the all-zero sequence. The squared Euclidean error distance caused by an error event ε depends on the transmitted sequence $x(D)$. Note that when $x(D)$ is i.i.d. the input error sequence associated with an error event ε , $\varepsilon_x(D) = x(D) \oplus \hat{x}(D)$ is also i.i.d. and is one of the possible input sequences.

Define $A^{C_{\text{ISI}}}(w, \delta^2)$ as the average number of error events of the ISI channel (or, the precoded ISI channel) that have squared Euclidean error distance δ^2 and w data bit errors. Assuming that $(K-1)$ uniform interleavers [44] are used during the retransmissions over the ISI channel, the distance spectrum of the parallel concatenation of K ISI channels is approximated as

$$A^{C_P}(w, \delta^2) \approx \binom{N}{w}^{1-K} \underbrace{\sum_{\Delta_1^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \sum_{\Delta_2^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \cdots \sum_{\Delta_n^2=\Delta_{\min}^2}^{\Delta_{\max}^2}}_{\Delta_1^2+\Delta_2^2+\dots+\Delta_n^2=\delta^2} \cdot \prod_{i=1}^K A^{C_{\text{ISI}}}(w, \Delta_i^2). \quad (5.12)$$

We can interpret $A^{C_P}(w, \delta^2)$ as the average number of error events of the parallel concatenation that have squared Euclidean error distance δ^2 and w data bit errors. The multiplicity $\binom{N}{w}^{1-K}$ in (5.12) is due to the contribution of $(K-1)$ uniform

interleavers. Applying the union bounding technique, $P(E_n)$ is upper bounded by

$$\begin{aligned} P(E_n) &\leq P_U(E_n) \\ &= \sum_{\delta^2 \in \Delta} \sum_{w=w_{\min}}^N A^{C_P}(w, \delta^2) Q\left(\sqrt{\frac{\delta^2 E_b}{2K N_0}}\right) \end{aligned} \quad (5.13)$$

where Δ is the set of all possible squared Euclidean error distances of the parallel concatenation.

To see the effect of the precoder on the performance of the IC schemes, we follow the approach in [44] and [45]. For large interleaver size N , the average number of error events of the ISI channel $A^{C_{\text{ISI}}}(w, \delta^2)$ is approximated as [45]

$$A^{C_{\text{ISI}}}(w, \delta^2) \approx \sum_{n=1}^{n_{\max}} \binom{N}{n} A^{C_{\text{ISI}}}(w, \delta^2, n) \quad (5.14)$$

where $A^{C_{\text{ISI}}}(w, \delta^2, n)$ is the average number of error events that have squared Euclidean error distance δ^2 , and w data bit errors, which are formed by concatenating n sub-error events, and n_{\max} is the largest number of sub-error events generated by weight w input error sequences $\varepsilon_x(D)$. Substituting (5.14) into (5.12), we have

$$\begin{aligned} A^{C_P}(w, \delta^2) &\approx \binom{N}{w}^{1-K} \underbrace{\sum_{\Delta_1^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \sum_{\Delta_2^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \cdots \sum_{\Delta_n^2=\Delta_{\min}^2}^{\Delta_{\max}^2}}_{\Delta_1^2+\Delta_2^2+\dots+\Delta_n^2=\delta^2} \\ &\quad \cdot \prod_{i=1}^K \sum_{n_i=1}^{n_{\max}} \binom{N}{n_i} A^{C_{\text{ISI}}}(w, \Delta_i^2, n_i). \end{aligned} \quad (5.15)$$

By using the approximation $\binom{N}{n} \approx \frac{N^n}{n!}$ for large N , we obtain

$$\begin{aligned}
A^{CP}(w, \delta^2) &\approx \underbrace{\sum_{\Delta_1^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \sum_{\Delta_2^2=\Delta_{\min}^2}^{\Delta_{\max}^2} \cdots \sum_{\Delta_n^2=\Delta_{\min}^2}^{\Delta_{\max}^2}}_{\Delta_1^2+\Delta_2^2+\dots+\Delta_n^2=\delta^2} \sum_{n_1=1}^{n_{\max}} \sum_{n_2=1}^{n_{\max}} \cdots \\
&\cdot \sum_{n_n=1}^{n_{\max}} N^{n_1+n_2+\dots+n_n-w(K-1)} \\
&\cdot \frac{(w!)^{n-1}}{n_1!n_2!\dots n_n!} \prod_{i=1}^K A^{C_{\text{ISI}}}(w, \Delta_i^2, n_i). \tag{5.16}
\end{aligned}$$

We now focus on terms that mostly contribute to the right hand side of (5.13). That is, we find the largest possible power of N , denoted by L_N . From (5.16) we have $L_N = Kn_{\max} - w(K-1)$. For the case of non-precoded ISI channels, the minimum value of input error sequence weight w that can produce finite weight squared Euclidean error distance $\Delta_{\text{np}}^2(\varepsilon)$ is $w_{\min} = 1$. The maximum number of sub-error events in this case is $n_{\max} = w$. Thus, $L_{N,\text{np}} = K \cdot 1 - 1 \cdot (K-1) = 1$. For precoded ISI channels, the minimum value of input error sequence weight w that can produce finite weight squared Euclidean error distance $\Delta_{\text{p}}^2(\varepsilon)$ is $w_{\min} = 2$ and correspondingly $n_{\max} = \lfloor \frac{w}{2} \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer part of x . Thus, for the lowest weight ($w = 2$) input error sequences, we have $L_{N,\text{p}} = K \cdot 1 - 2(K-1) = 2 - K$. Property 2 follows immediately. \square

It should be noted here that the above analysis shows the dependence of the truncated union bound on the probability of retransmission $P_U(E_n)$ on the length N for precoded and non-precoded systems. For the range of E_b/\mathcal{N}_0 and N for which the truncated union bound is tight, the above result applies to the probability of retransmission $P(E_n)$ as well. However, for the range of E_b/\mathcal{N}_0 for which the truncated union bound becomes loose such as when N becomes very large for a fixed E_b/\mathcal{N}_0 , the result is not necessarily valid for $P(E_n)$ and, hence, is weak. In the following, when

making conclusions about $P(E_n)$ based on the above analysis, it is to be assumed that the E_b/\mathcal{N}_0 is large enough that the union bound is tight.

Property 2 states that for non-precoded ISI channels, the union bound on $P(E_n)$ increases when the interleaver size N increases (this is of course undesirable) while for precoded ISI channels, the union bound on $P(E_n)$ decreases when N increases for the case $K \geq 3$. It is interesting to note that for non-precoded ISI channels, $P_U(E_n)$ increases linearly with N for both MLC and IC, hence, iterative combining offers no substantial improvement in performance. This explains the results in [40]. However, for precoded ISI channels, IC can significantly improve the performance over MLC.

It is interesting to note from Property 2 that when the number of transmissions $K = 2$, $P_U(E_n)$ is independent of interleaver size N for precoded ISI channels. In other words, there is no coding gain for precoded ISI channels for the case $K = 2$. This result agrees with that for repeat-accumulate codes [46] with repetition codes of length two. For the case of non-precoded ISI channels, $P_U(E_n)$ increases as N increases. Thus, precoded ISI channels still outperform non-precoded ISI channels in terms of $P_U(E_n)$, even when $K = 2$.

4. Interleaving Gain Property for Fading Precoded ISI Channels

So far, we have only considered time-invariant ISI channels. We next show that an interleaving gain results for time-variant ISI channels as well.

Property 4: An interleaving gain is present for time-variant (multi-path fading) ISI channels if a precoder is used. Hence, Property 2 is also valid for multi-path fading channels.

Proof: We first show that for time-variant ISI channels, when precoding is used, the recursiveness is also obtained. That is, we have to show that a weight one input error sequence $\varepsilon_x(D)$ produces a squared Euclidean error distance of infinite weight.

At time k , we denote the realization of the channel coefficient vector as $\mathbf{f}(n) = (f_0(n), f_1(n), \dots, f_L(n))$. The channel polynomial is now $f(D, n) = \sum_{l=0}^L f_l(n)D^l$. Taking the average of the squared Euclidean error distance of Rayleigh fading channels over all possible realizations of the channel coefficient vector $\mathbf{f}(n)$, we have

$$\overline{\delta^2}(\varepsilon) = \mathbb{E}_{\mathbf{f}}[\delta^2(\varepsilon)] = \sum_{n=0}^{N-1} \overline{\Delta_n^2}(\varepsilon) \quad (5.17)$$

where

$$\begin{aligned} \overline{\Delta_n^2}(\varepsilon) &= \mathbb{E}_{\mathbf{f}}[\Delta_n^2(\varepsilon)], & 0 \leq n < N \\ &= \mathbb{E}_{\mathbf{f}} \left[\left| \sum_{l=0}^L f_l(n) \varepsilon_{\mu}(d_{n-l}) \right|^2 \right] \\ &= \sum_{l=0}^L \sigma_{f_l}^2 \varepsilon_{\mu}^2(d_{n-l}) \end{aligned} \quad (5.18)$$

where $\sigma_{f_l}^2$ is the variance of the l th fading tap and $\varepsilon_{\mu}(d_n)$ is the n th coefficient of the error polynomial $\varepsilon_{\mu}(d(D))$. In (5.18), we have assumed independent Rayleigh fading on each of the taps. We can see that the average squared Euclidean error distance of the Rayleigh fading channels depends on their power profiles. Again, if $\varepsilon_x(D) = D^t$ for $t \geq 0$, we will have $\varepsilon_{\mu}(d(D))$ of infinite weight and, hence, $\overline{\delta^2}(\varepsilon)$ is of infinite weight.

It is possible to show that the pairwise error probability (PEP) caused by the error event ε , denoted by $P(\varepsilon)$, is a function of the averaged squared Euclidean error distances $\left\{ \overline{\Delta_n^2}(\varepsilon); 0 \leq n < N \right\}$ and E_b/\mathcal{N}_0 . That is $P(\varepsilon) = g\left(E_b/\mathcal{N}_0, \overline{\Delta_n^2}(\varepsilon); 0 \leq n < N\right)$, where $g(\cdot)$ is some deterministic function. Note that for the case of time-invariant ISI channels, $g(\cdot)$ becomes the Gaussian tail function $Q(\cdot)$. We now proceed in the same light as for the case of time-invariant channels. It is easy to show that the probability of a retransmission request being made while decoding the packet

formed by combining K received copies averaged over all realizations of the channel coefficient vector is upper bounded as

$$\begin{aligned} \overline{P}(E_n) \leq \sum_{\overline{\Delta}_n^2(\varepsilon)} \sum_{w=w_{\min}}^N \overline{A^{CP}}\left(w, \overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right) \\ \cdot g\left(E_b/\mathcal{N}_0, \overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right) \end{aligned} \quad (5.19)$$

where $\overline{A^{CP}}\left(w, \overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right)$ is the average number of error events of the parallel concatenation that have average squared Euclidean error distances $\left\{\overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right\}$ so that at fixed E_b/\mathcal{N}_0 the function $g\left(E_b/\mathcal{N}_0, \overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right)$ has a fixed value, and w data bit errors. Since recursiveness is also obtained when precoding is used, we can show that $\overline{A^{CP}}\left(w, \overline{\Delta}_n^2(\varepsilon); 0 \leq n < N\right)$ depends on the value of N , which exactly follows the same behavior as in Property 2. \square

It has also been shown that precoding for ISI channels significantly improves convergence of IC schemes. The convergence-based analysis for precoded ISI channels is discussed more detailed in Chapter III. Further discussion on this can be also found in [34].

E. Results

The performance of combining schemes has been investigated for three different channels. The first channel has one interfering symbol (2-tap channel), with impulse response (IR) $f_2(D) = 0.89443 + 0.44721D$. The second channel has 3 taps with IR $f_3(D) = 0.5 + 0.7071D + 0.5D^2$ and is taken from [40] for comparison purposes. The other channel has 5 taps with IR $f_5(D) = 0.67082 + 0.5D + 0.3873D^2 + 0.31623D^3 + 0.22361D^4$. As mentioned in [15], every precoder of the form $h(D) = 1 \oplus D^J$ results in an interleaving gain. In our simulations, we choose the precoder of the simplest form, i.e.: $h(D) = 1 \oplus D$. We see that this precoder, though simple, performs well

for all tested channels. To save simulation time, but still be able to demonstrate the performance of the MLC and IC schemes, we fix the number of transmissions to $K = 2$, and record the FER at the output of the equalizer. For the IC schemes, S -random interleavers of sizes 500 and 2000 bits ($S = 12$ and 17, respectively) and six iterations were used in all cases.

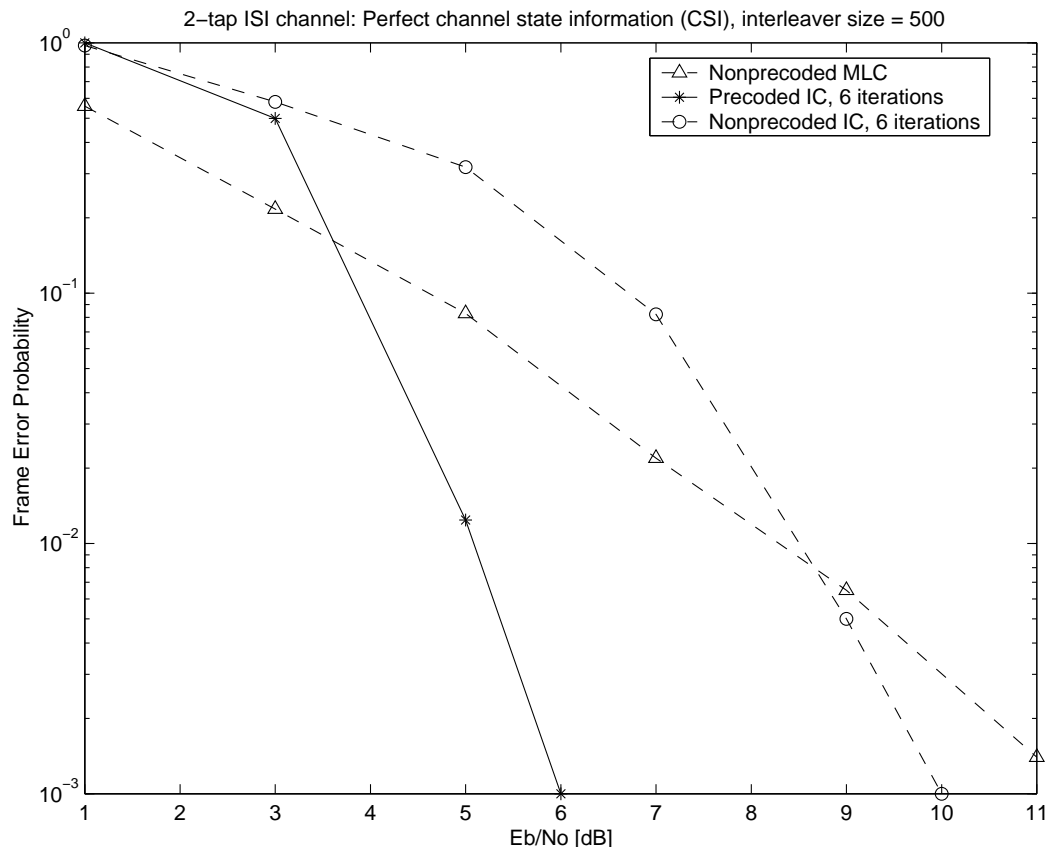


Fig. 25. Frame error rate for the $(0.89443, 0.44721)$ ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$.

Figure 25 shows the performance of the MLC and IC schemes over the channel $f_2(D)$ with one interfering symbol. At the second transmission, and at FER of 10^{-3} , precoded IC offers a coding gain of approximately 4 dB for interleaver size $N = 500$ over non-precoded IC. Comparing MLC schemes, precoded IC offers a coding gain

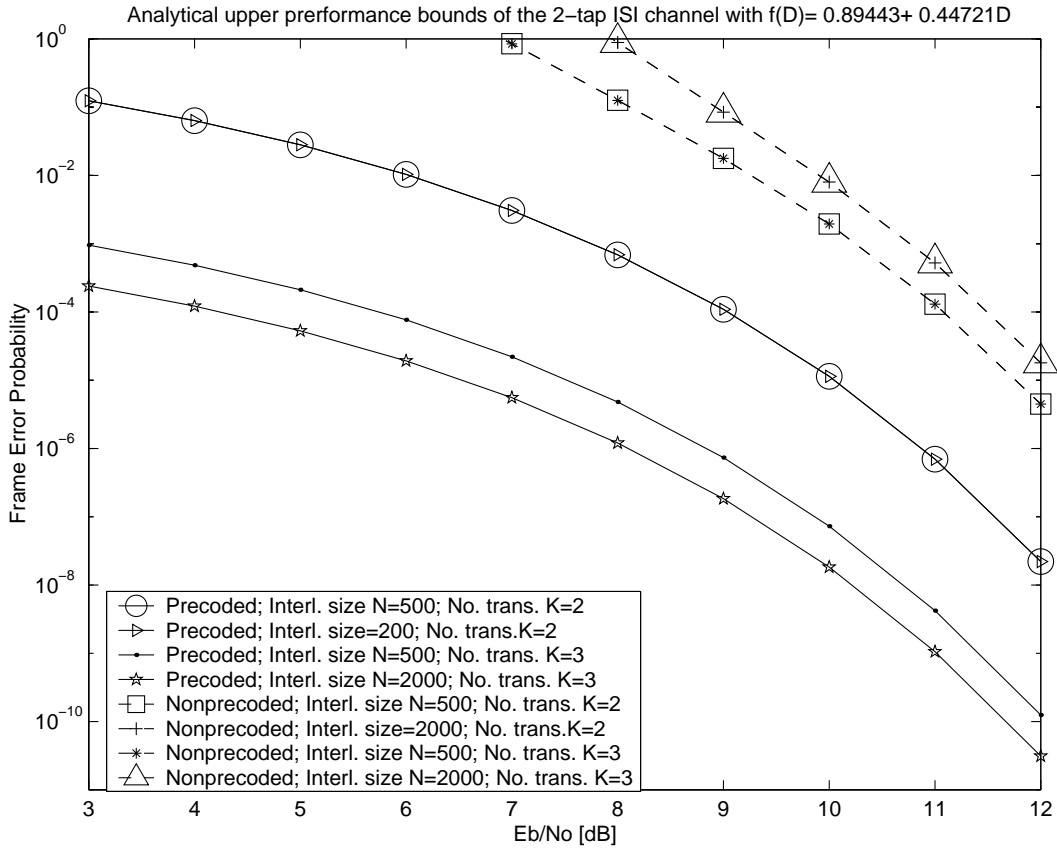


Fig. 26. FER bounds for the $(0.89443, 0.44721)$ ISI channel; ‘uniform’ interleaver of sizes $N = 500$ and $N = 2000$; precoding polynomial $h(D) = 1 \oplus D$; numbers of transmissions $K = 2$ and $K = 3$.

of roughly 5.5 dB, also at $FER=10^{-3}$. We also plot upper bound curves for this ISI channel (see Fig. 26) for different values of interleaver sizes as well as numbers of transmissions. In the evaluation of the bound, only significant terms in the distance spectrum of the parallel concatenation of the ISI channels are included. For the ISI channel with 2 taps, we can apply quasi-regularity properties [47], [48] to compute the distance spectrum. The bound is useful when used to predict the error floor of the system at high E_b/\mathcal{N}_0 where simulation is burdensome or impossible. We can see that when the number of transmissions $K = 3$ the precoded ISI channel can achieve interleaving gains in contrast to the non-precoded ISI channel.

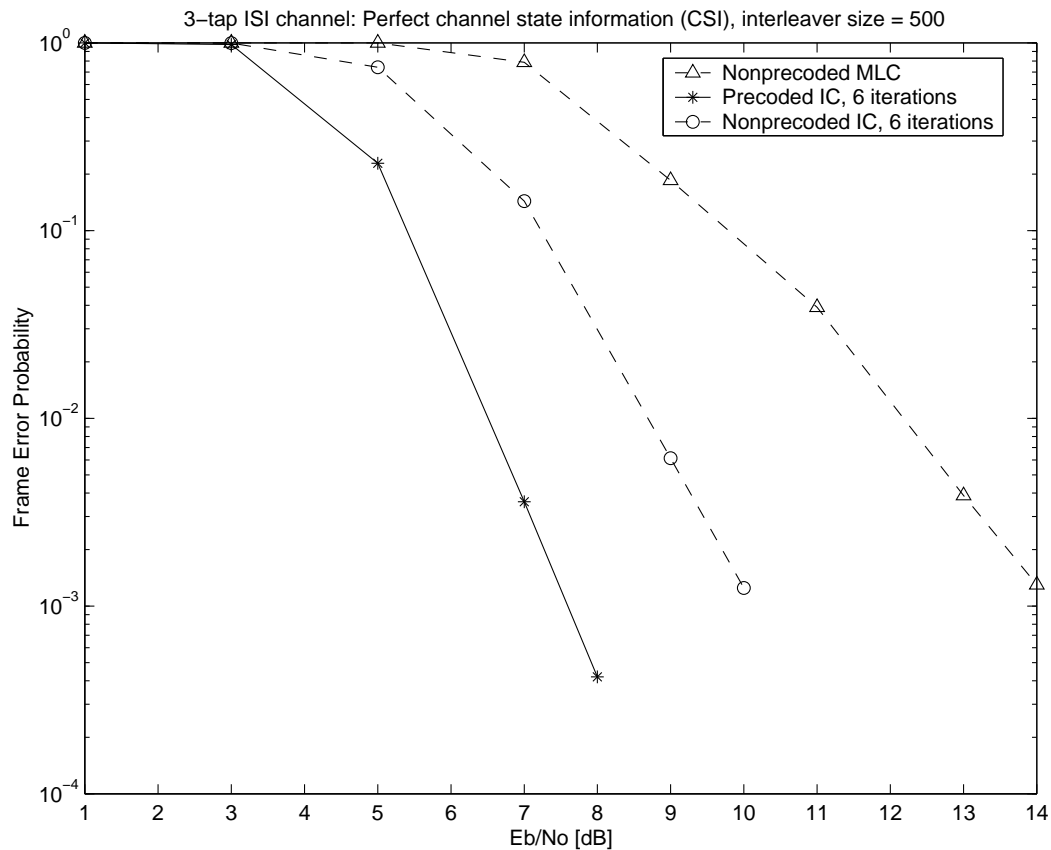


Fig. 27. Frame error rate for the $(0.5, 0.7071, 0.5)$ ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$.

Figures 27 and 28 show the performance results of the MLC and IC schemes over the 3-tap $f_3(D)$ and 5-tap $f_5(D)$ ISI channels respectively. The IC schemes result in significant improvements in terms of FER over MLC schemes. Precoded IC has the best performance and offers a coding gain of more than 6 dB over MLC for interleaver size 500. Precoded IC is still superior to non-precoded IC. By using precoding, improvements of about 2.5 dB for channel $f_3(D)$ and 3.0 dB for channel $f_5(D)$ with interleaver size $N = 500$ are obtained.

Figure 29 shows the results of MLC and IC schemes over the 3-tap channel $f_3(D)$ for interleavers of sizes 500 and 2000. It can be seen that for non-precoded IC, at

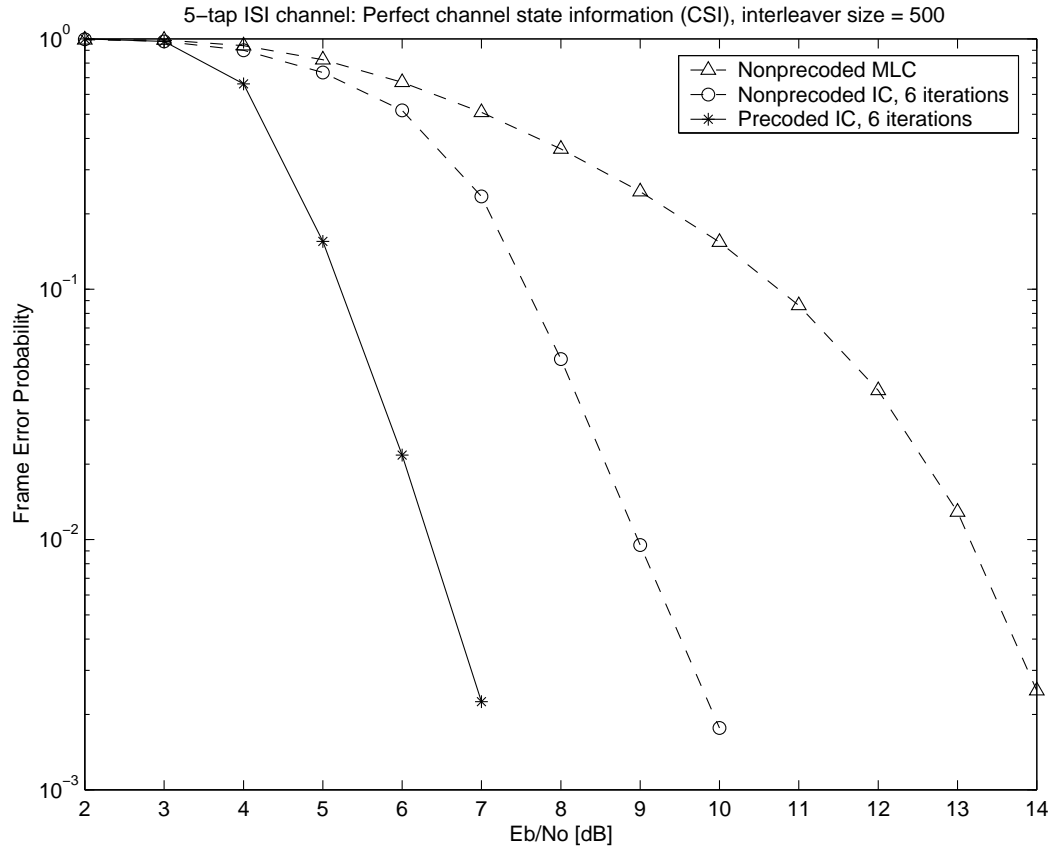


Fig. 28. Frame error rate for the $(0.67082, 0.5, 0.3873, 0.31623, 0.22361)$ ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$.

a fixed value of E_b/\mathcal{N}_0 , FER increases by a factor of roughly four times when the interleaver size increases from 500 to 2000 bits. However, FER curves for precoded IC remain almost unchanged when the interleaver size varies. Note that in our case $K = 2$ and these results perfectly reflect the analytical explanations in Section D.

Figure 30 shows the performance results of MLC and IC schemes for the 3-tap channel $f_3(D)$ when the decoder is mismatched to the channel. Similar to [40], the channel estimates used in the decoder are: $\hat{f}_0 = 0.42814$, $\hat{f}_1 = 0.79556$, and $\hat{f}_2 = 0.42814$. It can be seen that the IC schemes offer a potential robustness to mismatched decoding. From Fig. 27 and Fig. 30, it can be seen that the performance

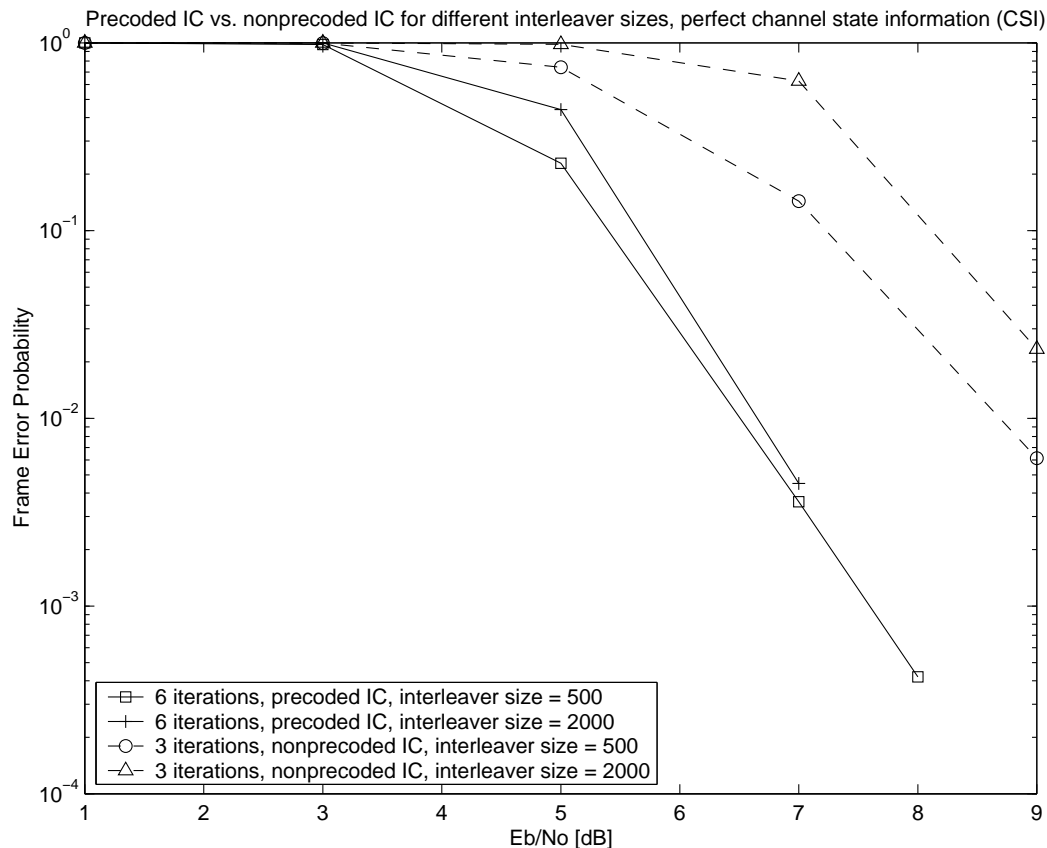


Fig. 29. Coding gains for the $(0.5, 0.7071, 0.5)$ ISI channel; interleaver sizes $N = 500$ and $N = 2000$; precoding polynomial $h(D) = 1 \oplus D$; number of transmissions $K = 2$.

with mismatched decoding is about 0.75 dB of the performance with perfect channel knowledge for the three tap ISI channel at a FER of 10^{-2} . In terms of FER, the precoded IC scheme offers coding gains of 2.5 dB and 7.5 dB over the non-precoded IC and MLC schemes respectively when the interleaver is of size $N = 500$.

F. Conclusion

The iterative principle applied to ARQ systems has been investigated for various ISI channels. By using binary precoding before the channel, the parallel concatenation of ISI channels follows a similar behavior as that of a multiple turbo code. Interleaving

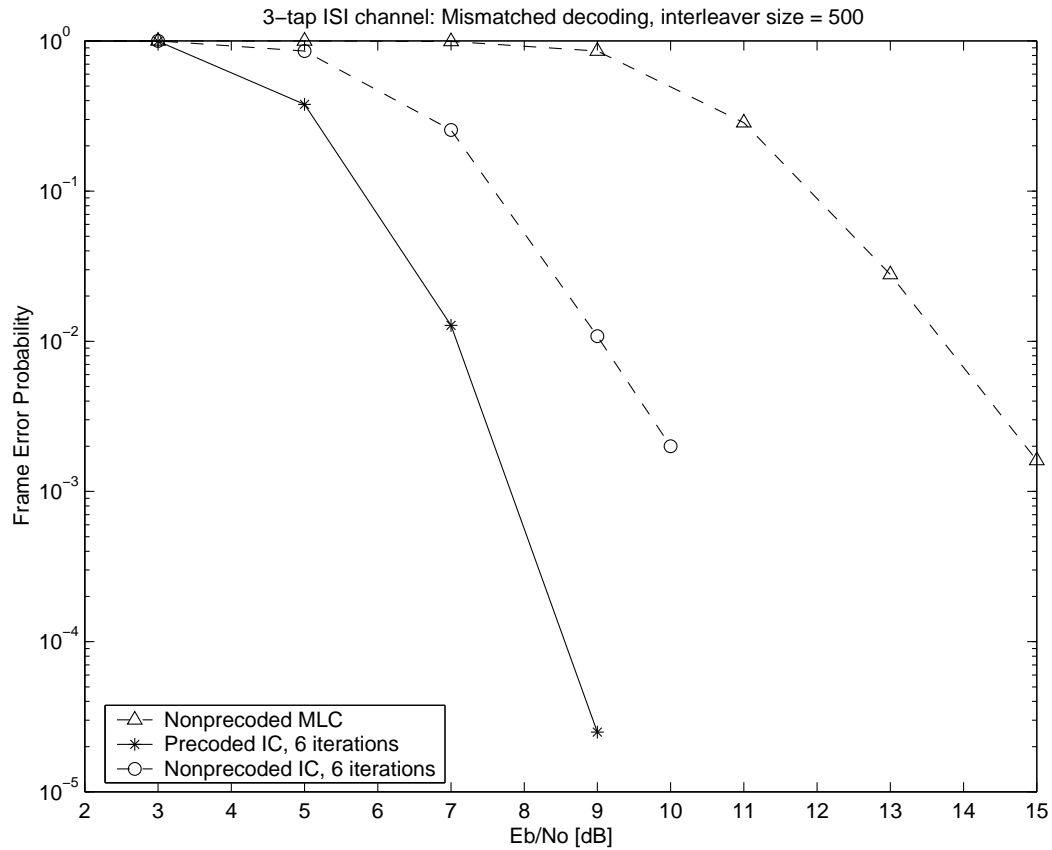


Fig. 30. Frame error rate for the $(0.5, 0.7071, 0.5)$ ISI channel; interleaver size $N = 500$; precoding polynomial $h(D) = 1 \oplus D$; mismatched channel at receiver $\hat{\mathbf{f}} = (0.42814, 0.79556, 0.42814)$; number of transmissions $K = 2$.

gains can be achieved in contrast to the scheme in [40]. In all cases investigated, precoded IC outperforms non-precoded IC and MLC with significant coding gains both when perfect channel state information is available and when the receiver uses mismatched channel estimates.

CHAPTER VI

CAPACITY OF MULTIPLE ANTENNA ISI CHANNELS AND SUMMARY

We consider the problem of finding the capacity for multiple antenna discrete-time ISI systems. We first investigate the capacity when both the transmitter and the receiver have perfect channel state information (CSI). It is shown that the capacity of the multiple-input single-output (MISO) ISI system is the same as that of the single-input multiple-output (SIMO) ISI system, which can be computed as the capacity of a single-input single-output ISI system with an equivalent composite frequency response. When the transmitter, however, has no knowledge of the CSI and the receiver utilizes a noisy (estimated) CSI, mismatched capacity is found. The effects on channel capacity of random and deterministic channel estimation errors are quantified.

A. Introduction

The capacity of an single-input single-output discrete-time ISI channel can be computed using the so called water-filling solution [3], analogously to the water-filling solution for the continuous-time ISI channel [49]. Finding the channel capacity becomes challenging for a more general case when there are N_T transmit antennas and N_R receive antennas in the system, or a multiple-input multiple-output (MIMO) system. Each channel, the channel from a certain transmit antenna to a receive antenna, is modeled as a discrete-time ISI channel with L interfering symbols (memory L). This kind of system is often encountered in, for example, cellular networks, wireless LAN with high data rate, and multi-track magnetic recording systems.

There has been a large quantity of work dedicated to MIMO systems because of the promising enhancement in capacity without power and bandwidth penalties. However, there has been no closed-form expression for the capacity of these system

when the ISI is present, even for a simpler case when there is one antenna at the transmitter or one antenna at the receiver. The authors in [50] claimed that the extension of the results for single antenna ISI channels to MIMO ISI channels is possible; however, no explicit capacity expression was given [50]. The problem of finding capacity for MIMO ISI channels was discussed by Diggavi in [51]; and it was pointed out that finding optimum power allocation is a hard problem. So the author contented with uncorrelated inputs (flat spectrum) in both space and time.

Recently, the authors in [52] computed the capacity for MISO ISI channels. It is based on the OFDM approach by splitting the channel spectra into different small bands of flat frequency response. The optimum power allocation was found. However, the approach is quite complicated and a compact expression for the capacity was not given. The evaluation of the capacity becomes challenging when the block length N becomes large because it involves matrix multiplications and determination of determinants of matrices of size N -by- N . Some related results can be deduced from the literature in the area of multiple-access and broadcast channels [53], [54], however, the problem setups there are quite different.

Computing capacity for MIMO systems when only noisy CSI is available at either the transmitter or the receiver, or both attracts much attention lately [55], [56], [57], [58]. However, these results are only available for either ISI-free channels or optimal receivers (using optimal decoding metric at the receiver). When receivers use mismatched decoding metric, some channel capacity results can be found in [59] and [60] for single antenna channels; but no such a result is available for MISO or SIMO ISI channels.

In this chapter, we extend the results in [3], [50], [59], and [60] to the MISO ISI and SIMO ISI systems. We provide closed-form expressions for the capacity of these systems. We first assume that the CSI is known at both the transmitter and the

receiver. Our approach is quite conventional (similar to [3]), and is simpler compared to [52]. The results provide a lot of insight and lead us to many useful interpretations.

We then investigate a more practical problem when the transmitter has no CSI, and the receiver has only noisy estimates of the CSI. This reflects the real life situation in which the receiver has to estimate the CSI through the received signals or through some backup channels (for example, the control channel in the down-link of a cellular network). In this case, we compute the mismatched capacity.

Although the results for the case when the transmitter and the receiver have perfect CSI are interesting; and to the best of our knowledge they have not been mentioned before, we believe the major contribution of this chapter is in computing the mismatched capacity.

B. Channel Model

The outputs of an MIMO discrete-time ISI system are expressed as

$$y_{r,n} = \sum_{t=1}^{N_T} \sum_{l=0}^L f_{r,t,l} \cdot x_{t,n-l} + z_{r,n}, \quad 1 \leq r \leq N_R, \quad 0 \leq n < N \quad (6.1)$$

where N_T is the number of transmit antennas, N_R is the number of receive antennas, N is the block length, $\{y_{r,n}\}$ is the output sequence at the r -th receive antenna, $\{x_{t,n}\}$ is the input sequence to the t -th transmit antenna, $f_{r,t,l}$ is the l -th tap in the channel with memory L between the t -th transmit antenna and the r -th receive antenna, and $\{z_{r,n}\}$ is a sequence of complex white Gaussian noise added to the r -th receive antenna with zero mean and autocorrelation $\mathbb{E}[z_{r,n_1} z_{r,n_2}^*] = \mathcal{N}_0 \delta_{n_1 - n_2}$, where $\delta_0 = 1$ and $\delta_n = 0$ for $n \neq 0$.¹ We assume that each of $N_T N_R$ channels has finite energy and

¹If the noise is colored, we can use a whitening filter for the noise before decoding. This operation is information lossless. However, care should be made when designing such a filter so that it is physically realizable.

remains unchanged during the duration of N symbols, and changes independently between blocks (block fading channels). We also assume that the channel input has finite energy. That is,

$$\sum_{t=1}^{N_T} \mathbb{E}[|x_{t,n}|^2] \leq E_s, \quad 0 \leq n < N$$

where $\mathbb{E}[\cdot]$ stands for expectation operation.

As done in [3] and [50], we use the complex N -circular Gaussian channel (NCGC) as

$$y_r[n] = \sum_{t=1}^{N_T} \sum_{l=0}^{N-1} f_{r,t}[n] x_t[n-l] + z_r[n], \quad 1 \leq r \leq N_R, \quad 0 \leq n < N \quad (6.2)$$

where $f_{r,t}[0, N-1]$ is obtained by padding $N-L-1$, assuming $L < N$, zeros to the end of $f_{r,t,l}$, i.e.,

$$f_{r,t}[n] = \begin{cases} f_{r,t,n}, & \text{if } 0 \leq n \leq L \\ 0, & \text{if } L < n < N, \end{cases}$$

and $u[n] = u_{n \bmod N}$, where $\{u_n\}$ is a generic sequence. We use the notation: $u[n_1, n_2] = \{u_{n_1}, \dots, u_{n_2}\}$, for $n_1 \leq n_2$. By the argument similar to [3], it can be shown that asymptotically with respect to the block length N , the channels defined in (6.1) and (6.2) are equivalent in terms of capacity. However, if cyclic prefixes are added to input sequences $\{x_{t,n}\}$, both channel models are equivalent for all values of N . Thus, we need to consider only the NCGC.

We define the discrete Fourier transform (DFT) and inverse discrete Fourier

transform (IDFT) pair as follows:

$$\mathbf{DFT:} \quad U[k] \triangleq \sum_{n=0}^{N-1} u[n] \Omega_N^{-kn}, \quad 0 \leq k < N, \quad (6.3)$$

$$\mathbf{IDFT:} \quad u[n] \triangleq \frac{1}{N} \sum_{k=0}^{N-1} U[k] \Omega_N^{kn}, \quad 0 \leq n < N \quad (6.4)$$

where $\Omega_N = e^{j2\pi/N}$, $j = \sqrt{-1}$, and $u[0, N-1]$ is an arbitrary sequence (either real or complex valued). This will be used in the subsequent sections.

C. Perfect CSI at Transmitter and Receiver

In this section, we investigate the capacity for MISO and SIMO discrete-time ISI channels assuming that perfect CSI is available at both the transmitter and the receiver. We start with MISO ISI channels. Similar results are then applied to the SIMO ISI channels. The technique computing the capacity here is quite conventional and is adopted from [3] and [50]. Specifically, we convert the time-domain channels to the frequency-domain ones by the help of the DFT of N -circular sequences. Then we convert them back to the desired time-domain representation by using the IDFT. This technique has been used widely recently to cope with ISI channels [53], [54]. We show that the resulting capacity of both channels are equivalent, as expected, even though this simple result has not been mentioned for ISI systems before. (This is true for ISI-free systems [61].)

1. MISO Discrete-Time ISI System

By omitting the index representing the receive antenna, the output of an MISO complex NCGC is given by

$$y[n] = \sum_{t=1}^{N_T} \sum_{l=0}^{N-1} f_t[l] x_t[n-l] + z[n], \quad 0 \leq n < N \quad (6.5)$$

with power constraints:

$$\sum_{t=1}^{N_T} \mathbb{E}[|x_{t,n}|^2] \leq E_s, \quad 0 \leq n < N. \quad (6.6)$$

The following result is the generalization of [3] and [50] for the case of multiple inputs. Note that in [3], the real-valued channel model is assumed.

Theorem 1: The capacity of the MISO complex NCGC (6.5) with respect to the block length N is

$$C_N^{\text{MISO}} = \frac{1}{N} \sum_{k=0}^{N-1} \log \left(1 + \frac{P_{\mathbf{X},\text{opt}}[k] \|\mathbf{F}[k]\|^2}{\mathcal{N}_0} \right)$$

where $\|\mathbf{F}[k]\|^2 = \sum_{t=1}^{N_T} |F_t[k]|^2$, $F_t[k]$ is the DFT of $f_t[0, N-1]$, and

$$P_{\mathbf{X},\text{opt}}[k] = \max \left(\lambda - \frac{\mathcal{N}_0}{\|\mathbf{F}[k]\|^2}, 0 \right), \quad 0 \leq k < N \quad (6.7)$$

where the water level λ is found through the power constraint

$$\frac{1}{N} \sum_{k=0}^{N-1} P_{\mathbf{X},\text{opt}}[k] = E_s.$$

Moreover, the capacity achieving input sequence is a circularly symmetric complex Gaussian sequence with zero mean and its correlation sequence $r_{x_{t,i}}[0, N-1]$ is the

IDFT of $F_t^*[k]F_i[k]\frac{P_{\mathbf{X},\text{opt}}[k]}{\|\mathbf{F}[k]\|^2}$, i.e.,

$$\begin{aligned} r_{x_t,i}[m] &\triangleq \mathbb{E}[x_t[n+m]x_i^*[n]] \\ &= \frac{1}{N} \sum_{k=0}^{N-1} F_t^*[k]F_i[k]\frac{P_{\mathbf{X},\text{opt}}[k]}{\|\mathbf{F}[k]\|^2}\Omega^{mk}, \quad 1 \leq t, i \leq N_T, \quad 0 \leq m < N. \end{aligned} \quad (6.8)$$

Proof: By definition, the capacity of the channel (6.5) with respect to block length N can be expressed as

$$\begin{aligned} C_N^{\text{MISO}} &= \sup_{p_x} I_N^{\text{MISO}} \\ &\triangleq \sup_{p_x} \frac{1}{N} I(x_1[0, N-1], \dots, x_{N_T}[0, N-1]; y[0, N-1]) \end{aligned} \quad (6.9)$$

where the supremum is taken over all pdf's p_x of the joint sequence $\{x_1[0, N-1], \dots, x_{N_T}[0, N-1]\}$ satisfying the per-block power constraint:

$$\sum_{t=1}^{N_T} \sum_{n=0}^{N-1} \mathbb{E}[|x_t[n]|^2] \leq N E_s. \quad (6.10)$$

It will be clear that the capacity C_N^{MISO} defined in (6.9) is still the same even when stronger per-symbol power constraints

$$\sum_{t=1}^{N_T} \mathbb{E}[|x_t[n]|^2] \leq E_s, \quad 0 \leq n < N \quad (6.11)$$

are imposed.

Taking the DFT of (6.5) and using the linearity property of DFT, we get a set of N parallel complex Gaussian channels as follows:

$$Y[k] = \sum_{t=1}^{N_T} F_t[k]X_t[k] + Z[k], \quad 0 \leq k < N \quad (6.12)$$

where $X_t[0, N-1]$, $Y[0, N-1]$, $Z[0, N-1]$, and $F_t[0, N-1]$ are, respectively, the DFTs of $x_t[0, N-1]$, $y[0, N-1]$, $z[0, N-1]$, and $f_t[0, N-1]$. Using Parseval's

theorem, the power constraint (6.10) becomes

$$\sum_{t=1}^{N_T} \sum_{k=0}^{N-1} \mathbb{E}[|X_t[k]|^2] \leq N^2 E_s.$$

Defining $\mathbf{X}[k] = [X_1[k] \cdots X_{N_T}[k]]^T$ and $\mathbf{F}[k] = [F_1[k] \cdots F_{N_T}[k]]^T$, (6.12) becomes

$$Y[k] = \mathbf{F}^T[k] \mathbf{X}[k] + Z[k], \quad 0 \leq k < N.$$

Since the DFT is invertible, the transform in (6.12) is information lossless. Thus,

$$I_N^{\text{MISO}} = \frac{1}{N} I(\mathbf{X}[0, N-1]; Y[0, N-1]) \leq \frac{1}{N} \sum_{k=0}^{N-1} I(\mathbf{X}[k]; Y[k]) \quad (6.13)$$

with equality iff $Y[k]$ are independent. The terms $I(\mathbf{X}[k]; Y[k])$ can be computed by using the technique proposed in [61] by performing the single value decomposition (SVD) of $\mathbf{F}^T[k]$. However, this would lead to a complicated expression. Here, we are interested in a closed-form expression for the capacity. Thus we choose to apply a different approach. We write

$$\begin{aligned} I(\mathbf{X}[k]; Y[k]) &= \mathfrak{h}(Y[k]) - \mathfrak{h}(Y[k]|\mathbf{X}[k]), \quad 0 \leq k < N \\ &= \mathfrak{h}(Y[k]) - \mathfrak{h}(Z[k]), \end{aligned} \quad (6.14)$$

where $\mathfrak{h}(\cdot)$ denotes differential entropy, and the second equality is because the input $\mathbf{X}[0, N-1]$ and noise $Z[0, N-1]$ are independent. It is easy to verify that [50], [61] $\mathfrak{h}(Z[k]) = \log(\pi e N \mathcal{N}_0)$. So our next step is to find $\mathfrak{h}(Y[k])$.

Applying the Cauchy-Bunyakovski-Schwartz inequality² twice to the variance of

²The Cauchy-Bunyakovski-Schwartz inequality states: $|\sum_k x_k y_k|^2 \leq \sum_k |x_k|^2 \sum_i |y_i|^2$, with equality iff $y_k = \lambda x_k^*$, $\forall k$, where λ is a constant.

$Y[k]$ we get

$$\begin{aligned}
\mathbb{E}[|Y[k]|^2] &= \sum_{t=1}^{N_T} \sum_{i=1}^{N_T} F_t[k] F_i^*[k] \mathbb{E}[X_t[k] X_i^*[k]] + N\mathcal{N}_0 \\
&\stackrel{(a)}{=} \left| \sum_{t=1}^{N_T} F_t[k] \sum_{i=1}^{N_T} F_i^*[k] \mathbb{E}[X_t[k] X_i^*[k]] \right| + N\mathcal{N}_0 \\
&\stackrel{(b)}{\leq} \left(\sum_{t=1}^{N_T} |F_t[k]|^2 \right)^{1/2} \left(\sum_{t=1}^{N_T} \left| \sum_{i=1}^{N_T} F_i^*[k] \mathbb{E}[X_t[k] X_i^*[k]] \right|^2 \right)^{1/2} + N\mathcal{N}_0 \quad (6.15) \\
&\stackrel{(c)}{\leq} \|\mathbf{F}[k]\| \left[\sum_{t=1}^{N_T} \left(\sum_{m=1}^{N_T} |F_m[k]|^2 \right) \left(\sum_{i=1}^{N_T} |\mathbb{E}[X_t[k] X_i^*[k]]|^2 \right) \right]^{1/2} + N\mathcal{N}_0 \\
&= \|\mathbf{F}[k]\|^2 \left(\sum_{t=1}^{N_T} \sum_{i=1}^{N_T} |\mathbb{E}[X_t[k] X_i^*[k]]|^2 \right)^{1/2} + N\mathcal{N}_0
\end{aligned}$$

where $\|\mathbf{F}[k]\|^2 \triangleq \sum_{t=1}^{N_T} |F_t[k]|^2$. (a) is obvious because the variance is always non-negative. (b) becomes equality iff

$$\sum_{i=1}^{N_T} F_i^*[k] \mathbb{E}[X_t[k] X_i^*[k]] = p[k] F_t^*[k], \quad (6.16)$$

where $p[k]$ is some deterministic function. (c) becomes equality iff

$$\mathbb{E}[X_t[k] X_i^*[k]] = q_t[k] F_i^*[k] \quad (6.17)$$

where $q_t[k]$ is some deterministic function. Substituting (6.17) into (6.16) we get

$$q_t[k] \|\mathbf{F}[k]\|^2 = p[k] F_t^*[k]. \quad (6.18)$$

We can safely assume that $\|\mathbf{F}[k]\|^2 \neq 0$. If this is not the case, then it means $F_t[k] = 0$, $1 \leq t \leq N_T$. In this case, we do not assign any power to the frequency k .

Thus, equivalently, (6.18) becomes

$$q_t[k] = p[k] \frac{F_t^*[k]}{\|\mathbf{F}[k]\|^2}. \quad (6.19)$$

Substituting (6.19) into (6.17) we get

$$\mathbb{E}[X_t[k]X_i^*[k]] = NP_{\mathbf{X}}[k] \frac{F_t^*[k]F_i[k]}{\|\mathbf{F}[k]\|^2} \quad (6.20)$$

where $P_{\mathbf{X}}[k] = p[k]/N$. By equating $t = i$ in the equation above, it can be easily seen that $p[k]$ (or, equivalently, $P_{\mathbf{X}}[k]$) must be non-negative. Summing up (6.20) over $0 \leq k < N$ and $1 \leq t \leq N_T$ for $t = i$ and using the power constraint (6.10) gives

$$\frac{1}{N} \sum_{k=0}^{N-1} P_{\mathbf{X}}[k] \leq E_s. \quad (6.21)$$

Substituting (6.16) into (6.15) and using the fact that $p[k]$ (or $P_{\mathbf{X}}[k] = p[k]/N$) is non-negative, we get

$$\mathbb{E}[|Y[k]|^2] = NP_{\mathbf{X}}[k]\|\mathbf{F}[k]\|^2 + N\mathcal{N}_0.$$

It is well-known that among random variables of the same variance, a circularly symmetric complex Gaussian distributed random variable has the largest entropy.

Thus, (6.14) follows

$$I(\mathbf{X}[k]; Y[k]) \leq \log \left(\frac{\mathbb{E}[|Y[k]|^2]}{N\mathcal{N}_0} \right) = \log \left(1 + \frac{P_{\mathbf{X}}[k]\|\mathbf{F}[k]\|^2}{\mathcal{N}_0} \right) \quad (6.22)$$

with equality iff $\mathbf{X}[k]$ are circularly symmetric complex Gaussian random variables with zero mean and correlation given in (6.20).

Combining (6.22) and (6.13), we get

$$I_N^{\text{MISO}} \leq \frac{1}{N} \sum_{k=0}^{N-1} \log \left(1 + \frac{P_{\mathbf{X}}[k] \|\mathbf{F}[k]\|^2}{\mathcal{N}_0} \right) \quad (6.23)$$

with equality iff $\mathbf{X}[k]$ are independent circularly symmetric complex Gaussian random variables. We can think of (6.23) as the mutual information of an equivalent *scalar* channel with composite frequency response $\|\mathbf{F}(e^{jw})\| = \left(\sum_{t=1}^{N_T} |F_t(e^{jw})|^2 \right)^{1/2}$, where $F_t(e^{jw}) = \sum_{n=0}^{N-1} f_t[n] e^{-jnw} = \sum_{l=0}^L f_{t,l} e^{-jlw}$. The input to this equivalent channel is a sequence of *one-dimension* random variables $X_{\text{eq}}[0, N-1]$ with power constraint (6.21). The solution to this equivalent problem is the well-known water-filling solution [49]. The resulting optimum power allocation of $X_{\text{eq}}[0, N-1]$, denoted by $P_{\mathbf{X},\text{opt}}[k]$, is given in (6.7). Note that the water-filling solution is still the same for the case when we have per-symbol power constraint as given in (6.11). Using [50, Theorem 4], the capacity achieving sequence in time domain $x_t[0, N-1]$, $1 \leq t \leq N_T$ is the circularly symmetric complex Gaussian sequence with zero mean and correlation given in (6.8).

□

When the block length tends to infinity, we have the following result.

Corollary 1: The capacity of the MISO discrete-time ISI channel is

$$C^{\text{MISO}} = \frac{1}{2\pi} \int_0^{2\pi} \log \left(1 + \frac{P_{\mathbf{X},\text{opt}}(e^{jw}) \|\mathbf{F}(e^{jw})\|^2}{\mathcal{N}_0} \right) dw \quad (6.24)$$

where $\|\mathbf{F}(e^{jw})\|^2 = \sum_{t=1}^{N_T} |F_t(e^{jw})|^2$, and $F_t(e^{jw}) = \sum_{l=0}^L f_{t,l} e^{-jlw}$.

The capacity is achieved when the input to the t -th antenna is a circularly symmetric complex Gaussian sequence with zero mean and PSD

$$P_{X_t}(e^{jw}) = |F_t(e^{jw})|^2 \frac{P_{\mathbf{X},\text{opt}}(e^{jw})}{\|\mathbf{F}(e^{jw})\|^2}, \quad 1 \leq t \leq N_T \quad (6.25)$$

where

$$P_{\mathbf{X},\text{opt}}(e^{jw}) = \max \left(\lambda - \frac{\mathcal{N}_0}{\|\mathbf{F}(e^{jw})\|^2}, 0 \right),$$

where the water-level λ is chosen to satisfy the power constraint

$$\frac{1}{2\pi} \int_0^{2\pi} P_{\mathbf{X},\text{opt}}(e^{jw}) dw = E_s. \quad (6.26)$$

Proof: By the same argument in [3], the capacity of the NCGC is the same as that of the actual channel as the block length $N \rightarrow \infty$. Thus,

$$C^{\text{MISO}} = \lim_{N \rightarrow \infty} C_N^{\text{MISO}}.$$

We use the property of Riemann integrals,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} g(\|\mathbf{F}[k]\|^2) = \lim_{N \rightarrow \infty} \frac{1}{2\pi} \sum_{k=0}^{N-1} g(\|\mathbf{F}(w_k)\|^2) \Delta_w = \frac{1}{2\pi} \int_0^{2\pi} g(\|\mathbf{F}(w)\|^2) dw$$

where $g(\cdot)$ is any continuous real-valued function, $\|\mathbf{F}[k]\|^2 = \|\mathbf{F}(w_k)\|^2$, $w_k = k\Delta_w$, $0 \leq k < N$, and $\Delta_w = 2\pi/N$. We get the expressions for the capacity and input PSD as given in (6.24) and (6.25), respectively. \square

2. SIMO Discrete-Time ISI System

The following results are similar to those for MISO ISI channels computed in the previous subsection.

Theorem 2: The capacity of the SIMO complex NCGC (6.2) is

$$C_N^{\text{SIMO}} = \frac{1}{N} \sum_{k=0}^{N-1} \log \left(1 + \frac{P_{X,\text{opt}}[k] \|\mathbf{F}[k]\|^2}{\mathcal{N}_0} \right)$$

where $\|\mathbf{F}[k]\|^2 = \sum_{r=1}^{N_R} |F_r[k]|^2$, $F_r[k]$ is the DFT of $f_r[0, N-1]$, and

$$P_{X,\text{opt}}[k] = \max\left(\lambda - \frac{\mathcal{N}_0}{\|\mathbf{F}[k]\|^2}, 0\right), \quad 0 \leq k < N \quad (6.27)$$

where the water-level λ is found through the power constraint

$$\frac{1}{N} \sum_{k=0}^{N-1} P_{X,\text{opt}}[k] = E_s.$$

Moreover, the capacity achieving input sequence is a circularly symmetric complex Gaussian sequence with zero mean and its correlation sequence $r_x[0, N-1]$ is the IDFT of $P_X[0, N-1]$, i.e.,

$$r_x[m] \triangleq \mathbb{E}[x[n+m]x^*[n]] = \frac{1}{N} \sum_{k=0}^{N-1} P_{X,\text{opt}}[k] \Omega^{mk}, \quad 0 \leq m < N. \quad (6.28)$$

When the block length tends to infinity, we have the following result.

Corollary 2: The capacity of the SIMO discrete-time ISI channel (6.1) is

$$C^{\text{SIMO}} = \frac{1}{2\pi} \int_0^{2\pi} \log\left(1 + \frac{P_{X,\text{opt}}(e^{jw})\|\mathbf{F}(e^{jw})\|^2}{\mathcal{N}_0}\right) dw$$

where $\|\mathbf{F}(e^{jw})\|^2 = \sum_{r=1}^{N_R} |F_r(e^{jw})|^2$, and $F_r(e^{jw}) = \sum_{l=0}^L f_{r,l} e^{-jlw}$. The capacity is achieved when the input is a circularly symmetric complex Gaussian sequence with zero mean and PSD

$$P_{X,\text{opt}}(e^{jw}) = \max\left(\lambda - \frac{\mathcal{N}_0}{\|\mathbf{F}(e^{jw})\|^2}, 0\right)$$

where the water-level λ is chosen to satisfy the power constraint

$$\frac{1}{2\pi} \int_0^{2\pi} P_{X,\text{opt}}(e^{jw}) dw = E_s.$$

The proofs are quite similar to those for the MISO ISI channels.

- *Remark 1:* The capacity of the MISO ISI system is equivalent to that of the SIMO ISI system. This result is known for an ISI free system, but has not been proved for the ISI channel case before.
- *Remark 2:* The optimum transmission scheme for MISO ISI channels is that the inputs to transmit antennas are outputs of a common source whose PSD is the water-filling PSD of the composite channel with response $\|\mathbf{F}(e^{jw})\|$, then being filtered by different filters $F_t^*(e^{jw})/\|\mathbf{F}(e^{jw})\|$. For MISO continuous-time ISI channels, a similar result was presented in [62] but a robust proof was not given.
- *Remark 3:* It was shown in [52] that the channel capacity can be achieved through OFDM with separate coding on each sub-carrier. Our proposed approach is simpler and the results here show that capacity can alternatively be achieved by serial transmission. This result shows that a single Gaussian codebook (whose PSD matches the water-filling spectrum of the composite channel) suffices to achieve capacity unlike in the OFDM case when a codebook of different rates should be used on each sub-carrier. The codewords can then be filtered (beam-forming) by different filters before transmission through the different antennas. Our approach is particularly well suited, when the capacity-achieving frequency band is concentrated within continuous intervals. This is often the case since the equivalent channel $\|\mathbf{F}(e^{jw})\|$ now is a summation of many individual channels. Thus, it is flattened out over the whole range of frequency. The result that serial transmission can be used to achieve capacity for the single antenna ISI channel case is due to Forney and Ungerboeck [63]. The results here show that this is true for the MISO ISI case as well.

3. MIMO Discrete-Time ISI System

The general expression for the capacity of MIMO ISI channels is given in [51]. However, the detailed proof was not included in [51]. Furthermore, the solution for the optimization of input PSD was not found there. Thus, the author used the flat input PSD in both space and time dimensions. For MIMO channels without ISI, in general, the optimization solution also has no closed-form expression except for some asymptotic cases. Thus we leave this open problem of finding capacity for MIMO ISI channels as the future research.

It has been shown that the SIMO and MISO ISI channels behave equivalently and they accept quite similar analysis approaches. Thus, from now on, we only focus on the MISO ISI channels. Similar results can be deduced for the counterpart MISO ISI channels. In the next section, we shall discuss another interesting problem when CSI is not present at the transmitter and the receiver only has a noisy CSI.

D. Information Rate for Nearest Neighbor Decoding

A more practical scenario is when the receiver does not have perfect CSI but only the estimated CSI, or when it is too expensive to have/to implement the system with the precise CSI. The receiver may form the estimated CSI based on the received signals, or through a separate control channel which is independent of the transmitted data. Another similar problem arises when sub-optimum, or reduced-complexity, schemes are used for both the transmitter and the receiver. Instead of using the original channel coefficients, the receiver, or possibly the transmitter, uses the shortened version of them to reduce the encoding and decoding complexity. (In this chapter, we interchangeably use channel coefficients, channel taps, and CSI to mean the same thing.)

1. Estimated CSI

In this subsection, we investigate the case when the estimated channel coefficients $\{\hat{f}_{t,l}\}$ are available only at the receiver. An acceptable representation of $\{\hat{f}_{t,l}\}$ is

$$\hat{f}_{t,l} = \bar{f}_{t,l} + \tilde{f}_{t,l}, \quad 1 \leq t \leq N_T, \quad 0 \leq l \leq L \quad (6.29)$$

where $\bar{f}_{t,l}$ are fixed constants during the transmission block, and $\tilde{f}_{t,l}$ are random variables with $\mathbb{E}[\tilde{f}_{t,l}] = 0$ and $\mathbb{E}[|\tilde{f}_{t,l}|^2] = \sigma_{\tilde{f}}^2$. The values of $\bar{f}_{t,l}$ depend on the channel estimation method used. It is usually assumed that $\bar{f}_{t,l} = \mathbb{E}[f_{t,l}]$ when the channels undergo fading [56]. This represents an unbiased estimator. The results provided here are for one single realization of the channel. When studying outage capacity, averaging over statistics of the channels is necessary. We further assume that $\tilde{f}_{t,l}$ are independent of the channel inputs $x_{t,n}$, and that the processes $\{y_n\}$ and $\{\tilde{f}_{t,l}\}$ are jointly ergodic. By introducing the random terms $\tilde{f}_{t,l}$, we include many practical scenarios, enabling us to investigate the problem more flexibly. The assumption that variances of $\tilde{f}_{t,l}$ are the same is not totally unreasonable because normally the receiver employs only one channel estimation method. It is possible that the channel estimation error becomes smaller when the energy captured by the receiver along some paths is high. This situation can also be dealt with but it will make the expression messier. However, the assumption above is still made since it allows us to have a somewhat cleaner expression, as being seen shortly in the following subsections.

2. Decoding Metric

The optimum decoder requires the knowledge of the pdf of the receive sequence $\{y_n\}$ given the input sequence $\{x_{t,n}\}$ and the estimated CSI $\{\hat{f}_{t,l}\}$. However, this is often hard to obtain, especially when the receiver has to estimate the channel statistics.

Thus, a sub-optimum decoder is used instead. The decoder decodes the received signals by minimizing the following decoding metric

$$D_N(m) = \frac{1}{N} \sum_{n=0}^{N-1} \left| y_n - \sum_{t=1}^{N_T} \sum_{l=1}^L \hat{f}_{t,l} x_{t,n-l}^{(m)} \right|^2 \quad (6.30)$$

where the input sequence to the t -th transmit antenna $\{x_{t,n}^{(m)}\}$ is the m -th codeword taken from a codebook \mathcal{M} of e^{NR} codewords, where R (in nats) is the code rate. Since the transmitter has no knowledge of the CSI, i.i.d. Gaussian codewords are used.

By using the argument similar to that in [3], [59], the MISO NCGC and the true MISO ISI channels are equivalent in terms of capacity when $N \rightarrow \infty$. Thus we only consider the MISO NCGC.

The decoding metric for the MISO NCGC is then

$$\tilde{D}_N(m) = \frac{1}{N} \sum_{n=0}^{N-1} \left| y[n] - \sum_{t=1}^{N_T} \sum_{l=0}^{N-1} \hat{f}_t[n] x_t^{(m)}[n-l] \right|^2 \quad (6.31)$$

where $\hat{f}_t[0, N-1]$ is the sequence obtained from $\hat{f}_{t,l}$ by padding $N-L-1$ zeros at the end, and the indices in the summation in (6.31) are taken modulo N . As $N \rightarrow \infty$, the end effects vanish and, hence, decoding using the metrics (6.30) and (6.31) is equivalent. That is, the decoder chooses message m as

$$m = \arg \min_{\hat{m} \in \mathcal{M}} \tilde{D}_N(\hat{m}). \quad (6.32)$$

By Parseval's theorem, the decoding metric (6.31) becomes

$$\tilde{D}_N(m) = \frac{1}{N^2} \sum_{k=0}^{N-1} \left| Y[k] - \sum_{t=1}^{N_T} \hat{F}_t[k] X_t^{(m)}[k] \right|^2$$

where

$$\hat{F}_t[k] = \bar{F}_t[k] + \tilde{F}_t[k], \quad 1 \leq t \leq N_T, \quad 0 \leq k < N,$$

and $\overline{F}_t[k]$ and $\widetilde{F}_t[k]$ are, respectively, the DFTs of $\overline{f}_t[0, N-1]$ and $\widetilde{f}_t[0, N-1]$. It is easy to verify that $\mathbb{E}[|\widetilde{F}_t[k]|^2] = (L+1)\sigma_f^2$. Moreover, $\widetilde{F}_t[k]$ are independent in space (with index t) and correlated in frequency (with index k). That is, $\mathbb{E}[\widetilde{F}_{t_1}[k_1]\widetilde{F}_{t_2}^*[k_2]] = 0$, $0 \leq k_1, k_2, < N$, $t_1 \neq t_2$; and $\mathbb{E}[\widetilde{F}_t[k_1]\widetilde{F}_t^*[k_2]] \neq 0$, $k_1 \neq k_2$.

3. Generalized Mutual Information

The generalized mutual information (GMI) [60], denoted by $I_{\text{GMI}}^{\text{MISO}}$, of the system assuming the decoding rule (6.32) specifies the highest rate at which the probability of error averaged over all i.i.d. Gaussian input sequences tends to zero as block length N goes to infinity. The following result quantifies the $I_{\text{GMI}}^{\text{MISO}}$, and is the generalization of [59].

Theorem 3: The GMI of the MISO NCGC whose decoder is based on (6.32), and whose channel input is an i.i.d. Gaussian ensemble is

$$I_{\text{GMI}}^{\text{MISO}} = \sup_{\theta < 0} I(\theta) = \sup_{\theta < 0} [\theta \mathcal{T} - \Lambda(\theta)] \quad (6.33)$$

where

$$\begin{aligned} \mathcal{T} &= \frac{1}{2\pi} \int_0^\pi \|\Delta_{\mathbf{F}}(e^{jw})\|^2 E_s / N_T dw \\ &\quad + (L+1)\sigma_f^2 E_s + \mathcal{N}_0, \\ \|\Delta_{\mathbf{F}}(e^{jw})\|^2 &= \sum_{t=1}^{N_T} |F_t(e^{jw}) - \overline{F}_t(e^{jw})|^2, \\ \overline{F}_t(e^{jw}) &= \sum_{n=0}^{N-1} \overline{f}_t[n] e^{-jnw} = \sum_{l=0}^L \overline{f}_{t,l} e^{-jlw}, \\ \|\overline{\mathbf{F}}(e^{jw})\|^2 &= \sum_{t=1}^{N_T} |\overline{F}_t(e^{jw})|^2, \end{aligned} \quad (6.34)$$

and

$$\begin{aligned}\psi(\theta, \mathcal{X}, w) &= \frac{\theta(\|\mathbf{F}(e^{jw})\|^2 E_s/N_T + \mathcal{N}_0)}{1 - \theta E_s \mathcal{X}/N_T} \\ &\quad - \log(1 - \theta E_s \mathcal{X}/N_T), \\ \Lambda(\theta) &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \psi(\theta, \mathcal{X}, w) f_{\mathcal{X}}(\mathcal{X}) d\mathcal{X} dw\end{aligned}\quad (6.35)$$

and $f_{\mathcal{X}}(\mathcal{X})$ is the non-central chi-square pdf with $2N_T$ degrees of freedom, and is given by

$$\begin{aligned}f_{\mathcal{X}}(\mathcal{X}) &= \frac{1}{(L+1)\sigma_{\tilde{f}}^2} \left(\frac{\mathcal{X}}{\|\overline{\mathbf{F}}(e^{jw})\|^2} \right)^{(N_T-1)/2} \\ &\quad \cdot \exp\left(-\frac{\|\overline{\mathbf{F}}(e^{jw})\|^2 + \mathcal{X}}{(L+1)\sigma_{\tilde{f}}^2}\right) \\ &\quad \cdot I_{N_T-1}\left(\frac{\sqrt{\mathcal{X}}\|\overline{\mathbf{F}}(e^{jw})\|}{(L+1)\sigma_{\tilde{f}}^2}\right), \quad \mathcal{X} \geq 0.\end{aligned}\quad (6.36)$$

Proof: Since the codewords are chosen independently according to the Gaussian product distribution, the decoder (6.32) is symmetric and the probability of error averaged over an ensemble of codebooks does not depend on the actual transmitted codeword. Thus, without loss of generality, we assume that the transmitted codeword is $\mathbf{x}^{(1)} = \{x_t^{(1)}[0, N-1], 1 \leq t \leq N_T\}$. Thus, the received signals are expressed as

$$y[n] = \sum_{t=1}^{N_T} \sum_{l=0}^{N-1} f_t[l] x_t^{(1)}[n-l] + z[n].$$

In general, we denote the DFT of $\mathbf{x}^{(m)}$ as $\mathbf{X}^{(m)}$. That is, $\mathbf{X}^{(1)} = [X_1^{(m)}[0, N-1] \cdots X_{N_T}^{(m)}[0, N-1]]^T$, where $X_t^{(m)}[k]$ are i.i.d. complex Gaussian random variables with zero mean and variance E_s/N_T .

Using the property of Riemann integrals [3], the joint ergodicity of processes $y[0, N-1]$ and $\tilde{f}_t[0, N-1]$ (so are their DFTs), and the independence among $\tilde{F}_t[k]$,

$X_t^{(m)}[k]$, and $Z[k]$, the metric accumulated by the correct codeword $\mathbf{X}^{(1)}$ converges almost surely (a.s.) to

$$\begin{aligned}
\mathcal{T} &= \lim_{N \rightarrow \infty} \tilde{D}_N(1) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{k=0}^{N-1} \left| \sum_{t=1}^{N_T} (F_t[k] - \hat{F}_t[k]) X_t^{(1)}[k] + Z[k] \right|^2 \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{N} \left| \sum_{t=1}^{N_T} \{ (F_t[k] - \bar{F}_t[k]) - \tilde{F}_t[k] \} X_t^{(1)}[k] + Z[k] \right|^2 \\
&\stackrel{\text{a.s.}}{=} \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{N} \mathbb{E} \left[\left| \sum_{t=1}^{N_T} \{ [F_t(e^{jw}) - \bar{F}_t(e^{jw})] - \tilde{F}_t(e^{jw}) \} X_t^{(1)}[k] + Z[k] \right|^2 \right] dw, \\
&\stackrel{\text{a.s.}}{=} \frac{1}{2\pi} \int_0^{2\pi} \left\{ \left[\|\Delta_{\mathbf{F}}(e^{jw})\|^2 + N_T(L+1)\sigma_f^2 \right] E_s/N_T + \mathcal{N}_0 \right\} dw,
\end{aligned}$$

where

$$\|\Delta_{\mathbf{F}}(e^{jw})\|^2 \triangleq \sum_{t=1}^{N_T} |F_t(e^{jw}) - \bar{F}_t(e^{jw})|^2.$$

An error occurs when the decoder declares message $m \neq 1$. The probability of error is then equal to $\Pr [\tilde{D}_N(m) < \tilde{D}_N(1)]$. Using the Gartner-Ellis theorem [64], [60], it follows that

$$I_{\text{GMI}}^{\text{MISO}} = - \lim_{N \rightarrow \infty} \frac{1}{N} \log \Pr [\tilde{D}_N(m) < \tilde{D}_N(1)] = \sup_{\theta < 0} [\theta \mathcal{T} - \Lambda(\theta)]$$

where $\Lambda(\theta)$ is the limiting logarithmic moment generating function (MGF)

$$\Lambda(\theta) = \lim_{N \rightarrow \infty} \frac{1}{N} \Lambda_N(N\theta) \tag{6.37}$$

where $\Lambda_N(\theta)$ is the MGF of metric $\tilde{D}_N(m)$, and is defined as

$$\Lambda_N(\theta) \triangleq \mathbb{E}[\exp(\theta \tilde{D}_N(m)) | Y[0, N-1], \bar{F}_t[0, N-1], 1 \leq t \leq N_T].$$

Given $Y[0, N-1]$ and $\overline{F}_t[0, N-1], 1 \leq t \leq N_T$, $\tilde{D}_N(m)$ is simply the summation of N non-central chi-squared random variables with 2 degrees of freedom each. It is easy to verify that

$$\Lambda_N(\theta) = \sum_{k=0}^{N-1} \left[\frac{|Y[k]|^2 \theta / N^2}{1 - \|\widehat{\mathbf{F}}[k]\|^2 \theta E_s / (N N_T)} - \log \left(1 - \|\widehat{\mathbf{F}}[k]\|^2 \theta E_s / (N N_T) \right) \right] \quad (6.38)$$

where

$$\|\widehat{\mathbf{F}}[k]\|^2 = \sum_{t=1}^{N_T} |\overline{F}_t[k] + \tilde{F}_t[k]|^2.$$

Replacing (6.38) into (6.37), using the property of Riemann integrals, and using the ergodicity argument, we obtain

$$\begin{aligned} \Lambda(\theta) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \left[\frac{|Y[k]|^2 \theta / N}{1 - \|\widehat{\mathbf{F}}[k]\|^2 \theta E_s / N_T} - \log \left(1 - \|\widehat{\mathbf{F}}[k]\|^2 \theta E_s / N_T \right) \right] \\ &\stackrel{\text{a.s.}}{=} \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_0^{2\pi} \mathbb{E} \left[\frac{|\sum_{t=1}^{N_T} F_t(e^{jw}) X_t^{(1)}[k] + Z[k]|^2 \theta / N}{1 - \sum_{t=1}^{N_T} |\overline{F}_t(e^{jw}) + \tilde{F}_t(e^{jw})|^2 \theta E_s / N_T} \right. \\ &\quad \left. - \log \left(1 - \sum_{t=1}^{N_T} |\overline{F}_t(e^{jw}) + \tilde{F}_t(e^{jw})|^2 \theta E_s / N_T \right) \right] dw \\ &\stackrel{\text{a.s.}}{=} \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \left[\frac{\theta (\|\mathbf{F}(e^{jw})\|^2 E_s / N_T + \mathcal{N}_0)}{1 - \theta E_s \mathcal{X} / N_T} - \log (1 - \theta E_s \mathcal{X} / N_T) \right] \\ &\quad \cdot f_{\mathcal{X}}(\mathcal{X}) d\mathcal{X} dw, \end{aligned}$$

where the last equality is obtained because $X_t^{(1)}[k]$ and $\tilde{F}_t[k]$ are independent, and the summation $\sum_{t=1}^{N_T} |\overline{F}_t(e^{jw}) + \tilde{F}_t(e^{jw})|^2$ follows the non-central chi-square distribution with $2N_T$ degrees of freedom and is given in (6.36). \square

With the original form of the GMI above, it is hard to gain some intuition behind even though we have used a quite simple set up for the channel estimation error (6.29). So we study some simplified cases.

a. Without Deterministic Channel Estimation Error

We consider the case when the deterministic channel estimation error vanishes, i.e., $\|\Delta_{\mathbf{F}}(e^{jw})\|^2 = 0$, or $\overline{\mathbf{F}}(e^{jw}) = \mathbf{F}(e^{jw})$. This represents an unbiased estimator. When the number of transmit antennas is large, or approaches infinity as in the asymptotic case, we can say more about $\|\mathbf{F}(e^{jw})\|^2/N_T$. For the analysis to be tractable, we assume that the channel coefficients $f_{t,l}$ spatially possess the same statistics (at different transmit antennas) and that they are temporally independent. That is, $f_{t_1,l}$ and $f_{t_2,l}$ are of the same statistics (identical distribution); and f_{t,l_1} and f_{t,l_2} , $l_1 \neq l_2$, are independent. We note here that this does not violate our assumption that the channel coefficients $f_{t,l}$ do not change during the whole transmission block of duration N symbols. What we mean by statistics is that the channels change independently from transmission block to transmission block, and they possess the randomness with the above assumptions. To avoid confusion, we introduce one more time index, denoted by κ , in the definition of channel coefficients to denote different transmission blocks. In this case, we write $f_{t,l}(\kappa)$. Thus, $f_{t,l}(\kappa)$ change with respect to κ , and remain constants for fixed κ . Denote $\mathbb{E}[|f_{t,l}(\kappa)|^2] = \sigma_{f_l}^2$, $0 \leq l \leq L$. Clearly, $f_{t_1,l}(\kappa)$ and $f_{t_2,l}(\kappa)$, $t_1 \neq t_2$, are different realizations of different processes with the same statistics (distribution). Consequently, in the limiting case,

$$\begin{aligned}
\lim_{N_T \rightarrow \infty} \frac{1}{N_T} \|\mathbf{F}(e^{jw})\|^2 &= \lim_{N_T \rightarrow \infty} \frac{1}{N_T} \sum_{t=1}^{N_T} \left| \sum_{l=0}^L f_{t,l}(\kappa) e^{-jlw} \right|^2 \\
&\stackrel{(a)}{=} \mathbb{E} \left[\left| \sum_{l=0}^L f_{t,l}(\kappa) e^{-jlw} \right|^2 \right], \quad \text{a.s.} \\
&\stackrel{(b)}{=} \sum_{l=0}^L \sigma_{f_l}^2 \triangleq \|\sigma_f\|^2
\end{aligned} \tag{6.39}$$

where (a) is true because we assume that $f_{t_1,l}(\kappa)$ are jointly ergodic, and $f_{t_1,l}(\kappa)$ and $f_{t_2,l}(\kappa)$ are of the identical distribution; and (b) is true because we implicitly assume that $\mathbb{E}[f_{t,l}(\kappa)] = 0$. This result implies that the equivalent channel behaves like a flat spectrum slow fading channel.

We are interested in the distribution of the random variable $\phi = \mathcal{X}/N_T$, where \mathcal{X} follows the non-central chi-squared distribution with $2N_T$ degrees of freedom as defined in (6.36). In fact, as $N_T \rightarrow \infty$, ϕ has an impulse distribution at $\sigma_{\bar{f}}^2 + \|\sigma_f\|^2$, i.e.,

$$f_\phi(\mathcal{X}/N_T) = \delta(\mathcal{X}/N_T - \sigma_{\bar{f}}^2 - \|\sigma_f\|^2). \quad (6.40)$$

This can be easily verified by using the fact that $\mathbb{E}[\mathcal{X}] = N_T\sigma_{\bar{f}}^2 + \|\bar{\mathbf{F}}(e^{jw})\|^2$ and $\mathbb{E}[|\mathcal{X}|^2] - \mathbb{E}^2[\mathcal{X}] = N_T\sigma_{\bar{f}}^2 + 2\sigma_{\bar{f}}^2\|\bar{\mathbf{F}}(e^{jw})\|^2$. Thus, $\mathbb{E}[\phi] = \sigma_{\bar{f}}^2 + \|\bar{\mathbf{F}}(e^{jw})\|^2/N_T$ and $\sigma_\phi^2 \triangleq \mathbb{E}[|\phi|^2] - \mathbb{E}^2[\phi] = \sigma_{\bar{f}}^2/N_T + 2\sigma_{\bar{f}}^2\|\bar{\mathbf{F}}(e^{jw})\|^2/N_T^2$. As $N_T \rightarrow \infty$, $\mathbb{E}[\phi] \rightarrow \sigma_{\bar{f}}^2 + \|\sigma_f\|^2$ and $\sigma_\phi^2 \rightarrow 0$, assuming that $\sigma_{\bar{f}}^2$ and $\|\sigma_f\|^2$ are bounded (since $\|\bar{\mathbf{F}}(e^{jw})\|^2 = \|\mathbf{F}(e^{jw})\|^2$).

Substituting (6.39) and (6.40) into (6.35) we obtain

$$\Lambda(\theta) = \frac{\theta(E_s\|\sigma_f\|^2 + \mathcal{N}_0)}{1 - \theta E_s(\sigma_{\bar{f}}^2 + \|\sigma_f\|^2)} - \log(1 - \theta E_s(\sigma_{\bar{f}}^2 + \|\sigma_f\|^2)). \quad (6.41)$$

The metric accumulated by the correct codeword is now

$$\mathcal{T} = (L+1)E_s\sigma_{\bar{f}}^2 + \mathcal{N}_0.$$

By taking the derivative of $I(\theta)$ in (6.33) with respect to θ and equating it to zero,

we obtain

$$0 = a\theta^2 + b\theta + c, \quad (6.42)$$

$$a = E_s(\sigma_{\tilde{f}}^2 + \|\sigma_f\|^2)^2[(L+1)E_s\sigma_{\tilde{f}}^2 + \mathcal{N}_0] > 0,$$

$$b = (\sigma_{\tilde{f}}^2 + \|\sigma_f\|^2)[E_s\|\sigma_f\|^2 - 2\mathcal{N}_0 - (2L+1)E_s\sigma_{\tilde{f}}^2],$$

$$c = L\sigma_{\tilde{f}}^2 - 2\|\sigma_f\|^2.$$

Denote roots of (6.42) by θ_1 and θ_2 . We are interested in a negative real root of (6.42), if it exists.³ Thus, there are two cases we need to consider.

- *When $b \geq 0$:* then $\sigma_{\tilde{f}}^2 \leq \frac{\|\sigma_f\|^2}{2L+1} - \frac{2\mathcal{N}_0}{(2L+1)E_s}$. This automatically gives $c < 0$, and consequently $\Delta = b^2 - 4ac > 0$ since $a > 0$ (always), and both roots are real. Obviously, there must be at least one negative root say $\theta_1 = (-b - \sqrt{\Delta})/(2a) < 0$.
- *When $b < 0$:* then $\sigma_{\tilde{f}}^2 > \frac{\|\sigma_f\|^2}{2L+1} - \frac{2\mathcal{N}_0}{(2L+1)E_s}$. If $\Delta = b^2 - 4ac > 0$, there must be at least one positive root, say $\theta_2 = (-b + \sqrt{\Delta})/(2a) > 0$, since $a > 0$. Since $\theta_1 \theta_2 = c/a$, $\theta_1 < 0$ when $c < 0$. Obviously, $c < 0$ (i.e., $\sigma_{\tilde{f}}^2 < \frac{2\|\sigma_f\|^2}{L}$) implies $\Delta > 0$. Thus, when $\frac{\|\sigma_f\|^2}{2L+1} - \frac{2\mathcal{N}_0}{(2L+1)E_s} < \sigma_{\tilde{f}}^2 < \frac{2\|\sigma_f\|^2}{L}$, there must one positive root and one negative root. When $c \geq 0$ (i.e., $\sigma_{\tilde{f}}^2 \geq \frac{2\|\sigma_f\|^2}{L}$), either $\Delta < 0$ or $\Delta \geq 0$, in which if the latter is true, then both roots must be non-negative.

Thus, the optimum solution is $\theta^* = \theta_1 = (-b - \sqrt{b^2 - 4ac})/(2a) < 0$, when $\sigma_{\tilde{f}}^2 < \frac{2\|\sigma_f\|^2}{L}$. If $\sigma_{\tilde{f}}^2 \geq \frac{2\|\sigma_f\|^2}{L}$, $\theta^* = 0$ and the resulting GMI is equal to zero, regardless of how large the transmit power E_s is. This result specifies the threshold effect of the

³The probability of decoding error is proportional to $e^{\theta E(\tilde{D}_\infty(m))}$, for some positive function $E(\tilde{D}_\infty(m))$. (We are dealing with positive transmission rate.) Thus, the probability of error can approach 0 only when θ is negative.

variance of the random channel estimation error, and is summarized as follows.

Lemma 1: As the number of antennas N_T approaches infinity, the random channel estimation error has the fatal effect on GMI if $\mathbb{E}[f_{t,l}(\kappa)] = 0$ and $\sigma_f^2 \geq \frac{2\|\sigma_f\|^2}{L} = \frac{2}{L} \sum_{l=0}^L \sigma_{f_l}^2$. In this case, regardless of how large the transmit power E_s is, the GMI is equal to zero.

- *Remark 4:* The distribution of channel coefficients in time or space has no influence on GMI because only the total channel gain $\|\sigma_f\|^2$ appears in the expression.
- *Remark 5:* It is interesting to note here that if $\|\Delta_{\mathbf{F}}(e^{jw})\|^2$ vanishes, it can be seen that the metric accumulated by the correct codeword \mathcal{T} depends on $(L + 1)\sigma_f^2$. That is, the effect of the channel estimation error increases by a factor of $(L + 1)$ compared to the frequency non-selective single antenna case. More interestingly, increasing the number of antennas N_T means that the random channel estimation noise sources increase; However, this *does not* worsen the overall effect of random channel estimation noise. This is because the noise sources (variances) are scaled by the signal power transmitted over each antenna (equal to E_s/N_T), which decreases by the same factor as the number of random noise source increases.

When the random channel estimation error diminishes, (i.e., $\sigma_f^2 = 0$), it is easy to verify that the optimization parameter $\theta^* = -1/\mathcal{N}_0$. In this case, the resulting GMI is

$$I_{\text{GMI}} = \log \left(1 + \frac{E_s}{\mathcal{N}_0} \|\sigma_f\|^2 \right)$$

which does not depend on the number of channel coefficients $L + 1$.

b. Without Random Channel Estimation Error

When the random channel estimation noises $\tilde{f}_{t,l}$ disappear, the distribution $f_{\mathcal{X}}(\mathcal{X})$ becomes an impulse at $\|\bar{\mathbf{F}}(e^{jw})\|^2$. That is,

$$f_{\mathcal{X}}(\mathcal{X}) = \delta(\mathcal{X} - \|\bar{\mathbf{F}}(e^{jw})\|^2). \quad (6.43)$$

Replacing (6.43) into (6.35) we get

$$\Lambda(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{\theta(\|\mathbf{F}(e^{jw})\|^2 E_s/N_T + \mathcal{N}_0)}{1 - \theta E_s \|\bar{\mathbf{F}}(e^{jw})\|^2/N_T} - \log(1 - \theta E_s \|\bar{\mathbf{F}}(e^{jw})\|^2/N_T) \right] dw. \quad (6.44)$$

The metric accumulated by the correct codeword \mathcal{T} does not depend on the random channel estimation noise, and is reduced to

$$\mathcal{T} = \frac{1}{2\pi} \int_0^{2\pi} \|\Delta_{\mathbf{F}}(e^{jw})\|^2 E_s/N_T dw + \mathcal{N}_0.$$

The resulting $I_{\text{GMI}}^{\text{MISO}}$ has a similar form as [59] for the single antenna channel. However, the result here is for the MISO case. Thus our result is more general.

In general, the optimum value of θ , denoted by θ^* , is a function of the channel response as well as the channel estimation error. Unfortunately, we are unable to find a closed-form expression for this optimization parameter. Thus, to compute the GMI, we use the numerical method. However, when the constant channel estimation error vanishes, i.e., $\|\Delta_{\mathbf{F}}(e^{jw})\|^2 = 0$, by taking the derivative of $I(\theta) = \theta\mathcal{T} - \Lambda(\theta)$ and solving the quadratic equation, the optimum θ would be: $\theta^* = -1/\mathcal{N}_0$. In this case, the GMI becomes the familiar i.i.d. Gaussian capacity C_G and is given by

$$I_{\text{GMI}}^{\text{MISO}} = C_G = \frac{1}{2\pi} \int_0^{2\pi} \log \left(1 + \frac{E_s \|\mathbf{F}(e^{jw})\|^2}{\mathcal{N}_0} \right) dw. \quad (6.45)$$

E. Conclusion

We have investigated the capacity for MISO and SIMO discrete-time ISI channels in the chapter. First, it was assumed that the CSI is perfectly known at the transmitter and the receiver. In this case, the capacity of the MISO and SIMO discrete-time ISI channels are the same. This result is well-known for the ISI-free channels, but has not been reported before for ISI channels. We then looked into the problem when the receiver has noisy CSI, and the transmitter has no CSI. Also, the receiver employs a sub-optimum decoding metric, instead of the optimum decoder as done in [56]. The code design idea for single antenna ISI channels can be used for multiple antenna ISI channels.

REFERENCES

- [1] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: turbo codes,” in *Proc. IEEE Int. Conf. on Communications*, Geneva, Switzerland, June 1993, pp. 1064–1070.
- [2] R. G. Gallager, “Low-density parity-check codes,” *IRE Trans. Inform. Theory*, vol. 8, pp. 21–28, Jan. 1962.
- [3] W. Hirt and J. L. Massey, “Capacity of the discrete-time Gaussian channel with intersymbol interference,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 380–388, May 1988.
- [4] A. Kavcic, “On the capacity of Markov sources over noisy channels,” in *Proc. IEEE Global Telecommunications Conf.*, Dallas, TX, Nov. 2001, pp. 2997–3001.
- [5] P. O. Vontobel and D. M. Arnold, “An upper bound on the capacity of channels with memory and constraint input,” in *Proc. IEEE Inform. Theory Workshop*, Cairns, Australia, Sept. 2001, pp. 147–149.
- [6] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, “Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 909–926, May 1998.
- [7] M. G. Luby, M. Mitzenmacher, M.A. Shokrollahi, and D. A. Spielman, “Improved low density parity check codes using irregular graphs and belief propagation,” in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 117.

- [8] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.
- [9] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.
- [10] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.
- [11] A. Ashikmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: a model and two properties," in *Proc. Conf. on Inform. Sciences and Systems*, Princeton, NJ, Mar. 2002, pp. 742–747.
- [12] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *Eur. Trans. Commun.*, vol. 6, pp. 507–511, Sep-Oct 1995.
- [13] D. Raphaeli and Y. Zarei, "Combined turbo equalization and turbo decoding," *IEEE Commun. Lett.*, vol. 2, pp. 107–109, Apr. 1998.
- [14] W. Ryan, "Design considerations for concatenating convolutional codes with partial response channels," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sept. 1999, pp. 1324–1333.
- [15] K. R. Narayanan, "Effect of precoding on the convergence of turbo equalization for partial response channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 686–698, Apr. 2001.

- [16] K. R. Narayanan, U. Dasgupta, and B. Lu, “Low complexity turbo equalization with binary precoding,” in *Proc. IEEE Int. Conf. on Communications*, New Orleans, LA, June 2000, pp. 1–5.
- [17] I. Lee, “The effect of a precoder on serially concatenated coding systems with ISI channel,” *IEEE Trans. Commun.*, vol. 49, pp. 1168–1175, July 2001.
- [18] S. ten Brink, “Design of serially concatenated codes based on iterative decoding convergence,” in *Proc. Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sept. 2000, pp. 319–322.
- [19] K. R. Narayanan, D. N. Doan, and R. V. Tamma, “Design and analysis of LDPC codes for turbo equalization with optimal and suboptimal soft output equalizers,” in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2002, pp. 737–746.
- [20] S. ten Brink, “Code doping for triggering iterative decoding convergence,” in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 235.
- [21] D. Divsalar, S. Dolinar, and F. Pollara, “Iterative turbo decoder analysis based on Gaussian density evolution,” in *Proc. IEEE MILCOM*, Los Angeles, CA, Oct. 2000, pp. 202–208.
- [22] D. N. Doan and K. R. Narayanan, “Some new results on the design of codes for intersymbol interference channels based on convergence of turbo equalization,” in *Proc. IEEE Int. Conf. on Communications*, New York, May 2002, pp. 1873–1877.

- [23] M. Tuchler and J. Hagenauer, "EXIT charts of irregular codes," in *Proc. Conf. on Inform. Sciences and Systems*, Princeton, NJ, Mar. 2002, pp. 748–753.
- [24] J. Justesen, "Information rates and power spectra of digital codes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 457–472, May 1982.
- [25] R. Karabed and P. H. Siegel, "Matched spectral-null codes for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 37, pp. 818–855, May 1991.
- [26] R. Karabed and P. H. Siegel, "Matched spectral null codes for partial response channels, parts I and II," in *Proc. IEEE Int. Symp. Information Theory*, Kode, Japan, June 1988, pp. 142–143.
- [27] X. Ma, N. Varnica, and A. Kavcic, "Matched information rate codes for binary ISI channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June 2002, p. 269.
- [28] N. Varnica and A. Kavcic, "Optimized low-density parity-check codes for partial response channels," *IEEE Commun. Lett.*, vol. 7, pp. 168–170, Apr. 2003.
- [29] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 765–77, Aug. 1986.
- [30] R. Haeb, "A modified trellis coding technique for partial response channels," *IEEE Trans. Commun.*, vol. 40, pp. 513–520, Mar. 1992.
- [31] W. Pusch, H. Weinrichter, and M. Taferner, "Turbo-codes matched to the $1-D^2$ partial response channel," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 63.
- [32] K. A. S. Immink, "Spectral nulls codes," *IEEE Trans. Magn.*, vol. 26, pp. 1130–1135, Mar. 1990.

- [33] S.-Y. Chung, Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.
- [34] D. N. Doan and K. R. Narayanan, "Iterative packet combining schemes for inter-symbol interference channels," *IEEE Trans. Commun.*, vol. 50, pp. 560–570, Apr. 2002.
- [35] J. Garcia-Frias and J. D. Villasenor, "Combined turbo detection and decoding for unknown ISI channels," *IEEE Trans. Commun.*, vol. 51, pp. 79–85, Jan. 2003.
- [36] W. Pusch and H. Weinrichter, "Improvement of data transmission on ISI channels by means of channel matched M -ary PAM-trellis codes," in *Proc. Int. Zurich Seminar on Broadband Communications*, Zurich, Switzerland, Feb. 1998, pp. 253–257.
- [37] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamental and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [38] D. Chase, "Code combining—a maximum-likelihood approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. 33, pp. 385–393, May 1985.
- [39] B. A. Harvey and S. B. Wicker, "Packet combining systems based on the Viterbi decoder," *IEEE Trans. Commun.*, vol. 42, pp. 1544–1557, Feb./Mar./Apr. 1994.
- [40] K. Balachandran and J. B. Anderson, "Mismatched decoding of intersymbol interference using a parallel concatenated scheme," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 255–259, Feb. 1998.

- [41] A. Singer, J. Nelson, and R. Koetter, "Linear iterative turbo-equalization (LITE) for dual channels," in *Proc. 33th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 1999, pp. 1670–1674.
- [42] D. Divsalar and F. Pollara, "Multiple turbo codes," in *Proc. IEEE Military Commun. Conf.*, San Diego, CA, Nov. 1995, pp. 279–285.
- [43] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal map decoding algorithms operation in the log domain," in *Proc. IEEE Int. Conf. on Communications*, Seattle, WA, June 1995, pp. 1009–1013.
- [44] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 409–428, Mar. 1996.
- [45] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. Commun.*, vol. 44, pp. 591–600, May 1996.
- [46] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sept. 1998, pp. 201–210.
- [47] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. 33, pp. 196–202, Mar. 1987.
- [48] S. A. Raghavan, J. K. Wolf, and L. B. Milstein, "On the performance evaluation of ISI channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 957–965, May 1993.
- [49] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

- [50] F. D. Neeser and J. L. Massey, “Proper complex random sequences with applications to information theory,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.
- [51] S. N. Diggavi, “On achievable performance of spatial diversity fading channels,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 308–325, Jan. 2001.
- [52] G. Scutari and S. Barbarossa, “Generalized water-filling for multiple transmit antenna systems,” in *Proc. IEEE Int. Conf. on Communications*, Anchorage, Alaska, May 2003, pp. 1668–2672.
- [53] R. S. Cheng and S. Verdu, “Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 773–785, May 1993.
- [54] A. J. Goldsmith and M. Effros, “The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 219–240, Jan. 2001.
- [55] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, “Efficient use of side information in multiple-antenna data transmission over fading channels,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.
- [56] E. Medard, “On the effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 933–946, May 2000.
- [57] E. Visotsky and U. Madhow, “Space-time transmit precoding with imperfect feedback,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

- [58] S. Bhashyam, A. Sabharwal, and B. Aazhang, “Feedback gain in multiple antenna systems,” *IEEE Trans. Commun.*, vol. 50, pp. 785–798, May 2002.
- [59] I. C. Abou-Faycal and A. Lapidoth, “On the capacity of reduced-complexity receivers for intersymbol interference channels,” in *Proc. 21st IEEE Convention of the Electrical and Electronic Engineers in Israel*, Tel-Aviv, Israel, Apr. 2000, pp. 263–266.
- [60] A. Lapidoth and S. Shamai, “Fading channels: how perfect need “perfect side information” be?,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.
- [61] E. Telatar, “Capacity of multi-antenna Gaussian systems,” *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [62] K. C. Zangi and L. G. Kransy, “Capacity-achieving transmitter and receiver pairs for dispersive miso channels,” *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1204–1216, Nov. 2003.
- [63] G. D. Forney and G. Ungerboeck, “Modulation and coding for linear Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2384–2415, Oct. 1998.
- [64] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, New York: Springer-Verlag, 1998.

APPENDIX A

PROOF OF PROPERTY 1, CHAPTER III

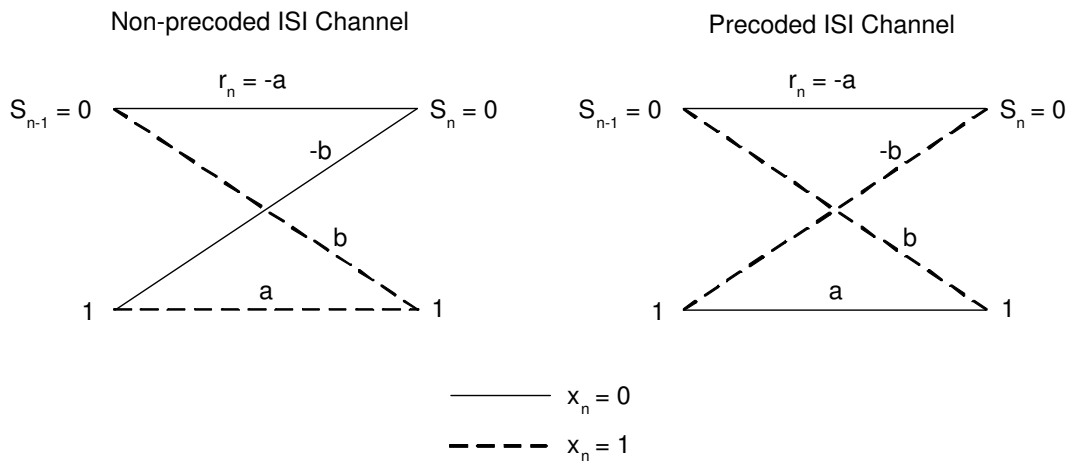


Fig. 31. Trellis of 2-tap (f_0, f_1) channels; for the precoded channel, the precoder polynomial $h(D) = 1 \oplus D$; $a = f_0 + f_1$; $b = f_0 - f_1$.

The trellis diagrams of the precoded and non-precoded ISI channels are presented in Fig. 31. At time instant n , we consider two cases; that is, the transmitted bits $x_n = 1$ and $x_n = 0$. Since there is no a priori knowledge available at the first iteration, the inner SISO assumes that the probability of transmitting x_j , for $j \neq n$, is equally likely. Let $\mathcal{X}_n^1 = \{x_0^{N-1} : x_n = 1\}$ and $\mathcal{X}_n^0 = \{x_0^{N-1} : x_n = 0\}$. We assume that the equalizer uses the BCJR algorithm [8] to compute the a posteriori probabilities. We

obtain the following relations for the forward and backward recursions:

$$\alpha_n(0) = \frac{1}{2}\alpha_{n-1}(0)p_{Z_n}(r_n + a + z_n) + \frac{1}{2}\alpha_{n-1}(1)p_{Z_n}(r_n + b + z_n), \quad (\text{A.1})$$

$$\alpha_n(1) = \frac{1}{2}\alpha_{n-1}(0)p_{Z_n}(r_n - b + z_n) + \frac{1}{2}\alpha_{n-1}(1)p_{Z_n}(r_n - a + z_n), \quad (\text{A.2})$$

$$\beta_n(0) = \frac{1}{2}\beta_{n+1}(0)p_{Z_n}(r_n + a + z_n) + \frac{1}{2}\beta_{n+1}(1)p_{Z_n}(r_n - b + z_n), \quad (\text{A.3})$$

$$\beta_n(1) = \frac{1}{2}\beta_{n+1}(0)p_{Z_n}(r_n + b + z_n) + \frac{1}{2}\beta_{n+1}(1)p_{Z_n}(r_n - a + z_n), \quad (\text{A.4})$$

where $r_n = f_0x'_n + f_1x'_{n-1}$ is the noiseless channel output, f_0 and f_1 are the first and second channel taps, respectively, $a = f_0 + f_1$, $b = f_0 - f_1$, and $p_{Z_n}(z_n) = \frac{1}{\sqrt{2\pi\sigma_z^2}}e^{-z_n^2/(2\sigma_z^2)}$ is the pdf of the n th noise sample. For a fixed transmitted sequence \mathbf{x} , \mathbf{r} is also fixed. Thus, taking the expectation of (A.1)–(A.4) over noise samples results in

$$\mathbb{E}_{z_0^n}[\alpha_n(0)] = C\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)]e^{-(r_n+a)^2/(4\sigma_z^2)} + C\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)]e^{-(r_n+b)^2/(4\sigma_z^2)}, \quad (\text{A.5})$$

$$\mathbb{E}_{z_0^n}[\alpha_n(1)] = C\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)]e^{-(r_n-b)^2/(4\sigma_z^2)} + C\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)]e^{-(r_n-a)^2/(4\sigma_z^2)}, \quad (\text{A.6})$$

$$\mathbb{E}_{z_n^{N-1}}[\beta_n(0)] = C\mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(0)]e^{-(r_n+a)^2/(4\sigma_z^2)} + C\mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(1)]e^{-(r_n-b)^2/(4\sigma_z^2)}, \quad (\text{A.7})$$

$$\mathbb{E}_{z_n^{N-1}}[\beta_n(1)] = C\mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(0)]e^{-(r_n+b)^2/(4\sigma_z^2)} + C\mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(1)]e^{-(r_n-a)^2/(4\sigma_z^2)}. \quad (\text{A.8})$$

In (A.5)–(A.8), we have used $\mathbb{E}_{z_n} \left[\frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-(r_n+m+z_n)^2/(2\sigma_z^2)} \right] = \frac{1}{2\sqrt{\pi\sigma_z^2}} e^{-(r_n+m)^2/(4\sigma_z^2)}$ and $C = \frac{1}{4\sqrt{\pi\sigma_z^2}}$. Since equations (A.5)–(A.8) share the common multiplication constant C , we can drop it without causing any confusion. We are now interested in finding the average values of the left-hand sides of (A.5)–(A.8) over all possible transmitted

channel input sequences \mathbf{x} (or, equivalently over all possible noiseless channel output sequences \mathbf{r}). We first focus on the forward recursion in (A.5). We have the following relations:

$$\begin{aligned} \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=a}[\alpha_n(0)] &\propto \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=a}[\alpha_{n-1}(0)]e^{-a^2/\sigma_z^2} + \\ &+ \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=a}[\alpha_{n-1}(1)]e^{-(a+b)^2/(4\sigma_z^2)}, \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-a}[\alpha_n(0)] &\propto \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-a}[\alpha_{n-1}(0)] + \\ &+ \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-a}[\alpha_{n-1}(1)]e^{-(a-b)^2/(4\sigma_z^2)}, \end{aligned} \quad (\text{A.10})$$

$$\begin{aligned} \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=b}[\alpha_n(0)] &\propto \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=b}[\alpha_{n-1}(0)]e^{-(a+b)^2/(4\sigma_z^2)} + \\ &+ \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=b}[\alpha_{n-1}(1)]e^{-b^2/\sigma_z^2}, \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-b}[\alpha_n(0)] &\propto \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-b}[\alpha_{n-1}(0)]e^{-(a-b)^2/(4\sigma_z^2)} + \\ &+ \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-b}[\alpha_{n-1}(1)]. \end{aligned} \quad (\text{A.12})$$

It should be mentioned here that in (A.9)–(A.12), we actually take the average of the quantity $\mathbb{E}_{z_0^n}[\alpha_n(0)]$ over all realizations of r_0^n . This should not be confused with the restriction that for a particular branch in the trellis, α_n must be in a certain state. In (A.9)–(A.12), we have used the following notations: $\{r_0^{n-1}, r_n = m\}$ is the set of all possible r_0^n of length $n + 1$ such that the n th component $r_n = m$; $\{r_0^{n-1}|r_n = m\}$ is the set of all possible r_0^{n-1} of length n given that $r_n = m$.

From the trellis diagrams in Fig. 31, it is easily seen that $\{r_0^{n-1}|r_n = a\} \equiv \{r_0^{n-1}|r_n = -b\}$ and $\{r_0^{n-1}|r_n = -a\} \equiv \{r_0^{n-1}|r_n = b\}$. Thus,

$$\mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=a}[\alpha_{n-1}(0)] = \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-b}[\alpha_{n-1}(0)] \triangleq \overline{\alpha_{n-1}^{a,-b}}(0), \quad (\text{A.13})$$

$$\mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-a}[\alpha_{n-1}(0)] = \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=b}[\alpha_{n-1}(0)] \triangleq \overline{\alpha_{n-1}^{-a,b}}(0), \quad (\text{A.14})$$

$$\mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=a}[\alpha_{n-1}(1)] = \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-b}[\alpha_{n-1}(1)] \triangleq \overline{\alpha_{n-1}^{a,-b}}(1), \quad (\text{A.15})$$

$$\mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=-a}[\alpha_{n-1}(1)] = \mathbb{E}_{z_0^{n-1}; r_0^{n-1}|r_n=b}[\alpha_{n-1}(1)] \triangleq \overline{\alpha_{n-1}^{-a,b}}(1). \quad (\text{A.16})$$

Since $r_n = -a$ and $r_n = -b$ merge into state $S_n = 0$ and $r_n = a$ and $r_n = b$ merge into state $S_n = 1$, we obtain

$$\overline{\alpha_n^{a,-b}}(0) = \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=a}[\alpha_n(0)] + \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=b}[\alpha_n(0)], \quad (\text{A.17})$$

$$\overline{\alpha_n^{-a,b}}(0) = \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-a}[\alpha_n(0)] + \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-b}[\alpha_n(0)], \quad (\text{A.18})$$

$$\overline{\alpha_n^{a,-b}}(1) = \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=a}[\alpha_n(1)] + \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=b}[\alpha_n(1)], \quad (\text{A.19})$$

$$\overline{\alpha_n^{-a,b}}(1) = \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-a}[\alpha_n(1)] + \mathbb{E}_{z_0^n; r_0^{n-1}, r_n=-b}[\alpha_n(1)]. \quad (\text{A.20})$$

Substituting (A.13)–(A.16) into (A.9)–(A.12) and then summing them according to (A.17)–(A.20), we have relations for the forward recursions at time instant n :

$$\overline{\alpha_n^{a,-b}}(0) \propto e^{-a^2/\sigma_z^2} \overline{\alpha_{n-1}^{a,-b}}(0) + e^{-b^2/\sigma_z^2} \overline{\alpha_{n-1}^{-a,b}}(1) + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha_{n-1}^{-a,b}}(0) + \overline{\alpha_{n-1}^{a,-b}}(1) \right], \quad (\text{A.21})$$

$$\overline{\alpha_n^{-a,b}}(0) \propto \overline{\alpha_{n-1}^{-a,b}}(0) + \overline{\alpha_{n-1}^{a,-b}}(1) + e^{-(a-b)^2/(4\sigma_z^2)} \left[\overline{\alpha_{n-1}^{a,-b}}(0) + \overline{\alpha_{n-1}^{-a,b}}(1) \right], \quad (\text{A.22})$$

$$\overline{\alpha_n^{a,-b}}(1) = \overline{\alpha_n^{-a,b}}(0), \quad (\text{A.23})$$

$$\overline{\alpha_n^{-a,b}}(1) \propto e^{-b^2/\sigma_z^2} \overline{\alpha_{n-1}^{a,-b}}(0) + e^{-a^2/\sigma_z^2} \overline{\alpha_{n-1}^{-a,b}}(1) + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha_{n-1}^{-a,b}}(0) + \overline{\alpha_{n-1}^{a,-b}}(1) \right]. \quad (\text{A.24})$$

Thus, equations (A.21)–(A.24) are the forward recursions on *expected* values directly, which will be used to compute the conditional a posteriori probabilities.

Likewise, by using the fact that $\{r_{n+1}^{N-1} | r_n = a\} \equiv \{r_{n+1}^{N-1} | r_n = b\}$, $\{r_{n+1}^{N-1} | r_n = -a\} \equiv \{r_{n+1}^{N-1} | r_n = -b\}$, and $r_n = -a$ and $r_n = b$ diverge from state $S_{n-1} = 0$, $r_n = a$ and $r_n = -b$ diverge from state $S_{n-1} = 1$, we obtain relations for the backward

recursions on expected values at time instant n as

$$\overline{\beta_n^{a,b}}(0) \propto e^{-a^2/\sigma_z^2} \overline{\beta_{n+1}^{a,b}}(0) + e^{-b^2/\sigma_z^2} \overline{\beta_{n+1}^{-a,-b}}(1) + e^{-(a-b)^2/(4\sigma_z^2)} \left[\overline{\beta_{n+1}^{-a,-b}}(0) + \overline{\beta_{n+1}^{a,b}}(1) \right], \quad (\text{A.25})$$

$$\overline{\beta_n^{-a,-b}}(0) \propto \overline{\beta_{n+1}^{-a,-b}}(0) + \overline{\beta_{n+1}^{a,b}}(1) + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\beta_{n+1}^{a,b}}(0) + \overline{\beta_{n+1}^{-a,-b}}(1) \right], \quad (\text{A.26})$$

$$\overline{\beta_n^{a,b}}(1) = \overline{\beta_n^{-a,-b}}(0), \quad (\text{A.27})$$

$$\overline{\beta_n^{-a,-b}}(1) \propto e^{-b^2/\sigma_z^2} \overline{\beta_{n+1}^{a,b}}(0) + e^{-a^2/\sigma_z^2} \overline{\beta_{n+1}^{-a,-b}}(1) + e^{-(a-b)^2/(4\sigma_z^2)} \left[\overline{\beta_{n+1}^{-a,-b}}(0) + \overline{\beta_{n+1}^{a,b}}(1) \right]. \quad (\text{A.28})$$

The initializations for (A.21)–(A.28) are done as usual, depending on the starting state and whether the trellis is terminated or not. We now calculate the a posteriori probabilities for bit x_n averaged over noise samples and transmitted sequences for the non-precoded and precoded channels. For a fixed transmitted sequence \mathbf{x} , \mathbf{r} is also fixed. Thus, $P(y_0^{n-1}, y_n, y_{n+1}^{N-1} | r_0^{N-1}) = P(y_0^{n-1} | r_0^{n-1}) P(y_n | r_n) P(y_{n+1}^{N-1} | r_{n+1}^{N-1})$. We have

$$\begin{aligned} \mathbb{E}_{\mathbf{z}}[P_{\text{np}}(\hat{X}_n = 1)] &\propto \left[\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)] e^{-(r_n-b)^2/(4\sigma_z^2)} + \right. \\ &\quad \left. + \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)] e^{-(r_n-a)^2/(4\sigma_z^2)} \right] \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(1)], \quad (\text{A.29}) \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{z}}[P_{\text{np}}(\hat{X}_n = 0)] &\propto \left[\mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)] e^{-(r_n+a)^2/(2N+0)} + \right. \\ &\quad \left. + \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)] e^{-(r_n+b)^2/(4\sigma_z^2)} \right] \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(0)], \quad (\text{A.30}) \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{z}}[P_{\text{p}}(\hat{X}_n = 1)] &\propto \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)] e^{-(r_n-b)^2/(4\sigma_z^2)} \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(1)] + \\ &\quad + \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)] e^{-(r_n+b)^2/(4\sigma_z^2)} \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(0)], \quad (\text{A.31}) \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{z}}[P_{\text{p}}(\hat{X}_n = 0)] &\propto \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(0)] e^{-(r_n+a)^2/(4\sigma_z^2)} \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(0)] + \\ &\quad + \mathbb{E}_{z_0^{n-1}}[\alpha_{n-1}(1)] e^{-(r_n-a)^2/(4\sigma_z^2)} \mathbb{E}_{z_{n+1}^{N-1}}[\beta_{n+1}(1)]. \quad (\text{A.32}) \end{aligned}$$

Assuming that $X_n = 1$ was transmitted at time n , the noiseless channel output r_n for the non-precoded channel must be in the set $\{a, b\}$. Thus, the conditional a posteriori probability can be expressed as

$$\begin{aligned} \overline{P}_{\text{np}}(\hat{X}_n = 0|X_n = 1) &= \mathbb{E}_{\mathbf{z}; \mathbf{x} \in \mathcal{X}_n^1} [P_{\text{np}}(\hat{X}_n = 0)] \\ &= \mathbb{E}_{\mathbf{z}; r_0^{n-1}, r_n = a, r_{n+1}^{N-1}} [P_{\text{np}}(\hat{X}_n = 0)] + \\ &\quad + \mathbb{E}_{\mathbf{z}; r_0^{n-1}, r_n = b, r_{n+1}^{N-1}} [P_{\text{np}}(\hat{X}_n = 0)]. \end{aligned} \quad (\text{A.33})$$

The two terms in the right-hand side of (A.33) can be directly calculated by the BCJR algorithm to obtain

$$\mathbb{E}_{\mathbf{z}; r_0^{n-1}, r_n = a, r_{n+1}^{N-1}} [P_{\text{np}}(\hat{X}_n = 0)] \propto \left[e^{-a^2/\sigma_z^2} \overline{\alpha}_{n-1}^{a,-b}(0) + e^{-(a+b)^2/(4\sigma_z^2)} \overline{\alpha}_{n-1}^{a,-b}(1) \right] \overline{\beta}_{n+1}^{a,b}(0), \quad (\text{A.34})$$

$$\mathbb{E}_{\mathbf{z}; r_0^{n-1}, r_n = b, r_{n+1}^{N-1}} [P_{\text{np}}(\hat{X}_n = 0)] \propto \left[e^{-b^2/\sigma_z^2} \overline{\alpha}_{n-1}^{-a,b}(1) + e^{-(a+b)^2/(4\sigma_z^2)} \overline{\alpha}_{n-1}^{-a,b}(0) \right] \overline{\beta}_{n+1}^{a,b}(0). \quad (\text{A.35})$$

Note that in (A.34), the forward and backward recursions, $\overline{\alpha}_{n-1}^{a,-b}$ and $\overline{\beta}_{n+1}^{a,b}$, are connected by the branch with $r_n = a$. Thus, they have a common superscript a . Similarly, in (A.35), $\overline{\alpha}_{n-1}^{-a,b}$ and $\overline{\beta}_{n+1}^{a,b}$ are connected by the branch with $r_n = b$. Thus, they have a common superscript b .

Substituting (A.34) and (A.35) into (A.33), we get

$$\begin{aligned} \overline{P}_{\text{np}}(\hat{X}_n = 0|X_n = 1) &\propto \left\{ e^{-a^2/\sigma_z^2} \overline{\alpha}_{n-1}^{a,-b}(0) + e^{-b^2/\sigma_z^2} \overline{\alpha}_{n-1}^{-a,b}(1) + \right. \\ &\quad \left. + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{-a,b}(0) + \overline{\alpha}_{n-1}^{a,-b}(1) \right] \right\} \overline{\beta}_{n+1}^{a,b}(0). \end{aligned} \quad (\text{A.36})$$

Likewise, we have

$$\begin{aligned} \overline{P}_{\text{np}}(\hat{X}_n = 1|X_n = 0) &= \mathbb{E}_{\mathbf{z}; \mathbf{x} \in \mathcal{X}_n^0} [P_{\text{np}}(\hat{X}_n = 0)] \\ &\propto \left\{ e^{-b^2/\sigma_z^2} \overline{\alpha}_{n-1}^{a,-b}(0) + e^{-a^2/\sigma_z^2} \overline{\alpha}_{n-1}^{-a,b}(1) + \right. \\ &\quad \left. + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{-a,b}(0) + \overline{\alpha}_{n-1}^{a,-b}(1) \right] \right\} \overline{\beta}_{n+1}^{-a,-b}(1), \end{aligned} \quad (\text{A.37})$$

$$\begin{aligned} \overline{P}_{\text{p}}(\hat{X}_n = 0|X_n = 1) &= \mathbb{E}_{\mathbf{z}; \mathbf{x} \in \mathcal{X}_n^1} [P_{\text{p}}(\hat{X}_n = 0)] \\ &\propto e^{-(a-b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{a,-b}(0) \overline{\beta}_{n+1}^{-a,-b}(0) + \overline{\alpha}_{n-1}^{-a,b}(1) \overline{\beta}_{n+1}^{a,b}(1) \right] + \\ &\quad + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{-a,b}(0) \overline{\beta}_{n+1}^{a,b}(0) + \overline{\alpha}_{n-1}^{a,-b}(1) \overline{\beta}_{n+1}^{-a,-b}(1) \right], \end{aligned} \quad (\text{A.38})$$

$$\begin{aligned} \overline{P}_{\text{p}}(\hat{X}_n = 1|X_n = 0) &= \mathbb{E}_{\mathbf{z}; \mathbf{x} \in \mathcal{X}_n^0} [P_{\text{p}}(\hat{X}_n = 0)] \\ &\propto e^{-(a-b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{a,-b}(0) \overline{\beta}_{n+1}^{a,b}(1) + \overline{\alpha}_{n-1}^{-a,b}(1) \overline{\beta}_{n+1}^{-a,-b}(0) \right] + \\ &\quad + e^{-(a+b)^2/(4\sigma_z^2)} \left[\overline{\alpha}_{n-1}^{-a,b}(0) \overline{\beta}_{n+1}^{-a,-b}(1) + \overline{\alpha}_{n-1}^{a,-b}(1) \overline{\beta}_{n+1}^{a,b}(0) \right]. \end{aligned} \quad (\text{A.39})$$

Using the relations $P(\hat{X}_n^{(1)} = 1) = \frac{e^{L_{\text{ext}}^{(1)}(x_n)}}{(1+e^{L_{\text{ext}}^{(1)}(x_n)})}$, $P(\hat{X}_n^{(1)} = 0) = \frac{1}{(1+e^{L_{\text{ext}}^{(1)}(x_n)})}$, and $L_{\text{ext}}^{(1)}(x_n) = \log\left(\frac{P(\hat{X}_n=1)}{P(\hat{X}_n=0)}\right)$,⁴ we get:

$$\begin{aligned} \pi^{(1)} &= \mathbb{E}_{\mathbf{z}; \mathbf{x}} [P(\hat{X}_n = 1|X_n = 1) + P(\hat{X}_n = 0|X_n = 0) - \\ &\quad - P(\hat{X}_n = 0|X_n = 1) - P(\hat{X}_n = 1|X_n = 0)] \\ &= \mathbb{E}_{\mathbf{z}; \mathbf{x}} [2 - 2P(\hat{X}_n = 0|X_n = 1) - 2P(\hat{X}_n = 1|X_n = 0)]. \end{aligned}$$

We have to show:

$$\begin{aligned} &\mathbb{E}_{\mathbf{z}; \mathbf{x}} [P_{\text{p}}(\hat{X}_n = 0|X_n = 1)] - \mathbb{E}_{\mathbf{z}; \mathbf{x}} [P_{\text{np}}(\hat{X}_n = 0|X_n = 1)] + \\ &\mathbb{E}_{\mathbf{z}; \mathbf{x}} [P_{\text{p}}(\hat{X}_n = 1|X_n = 0)] - \mathbb{E}_{\mathbf{z}; \mathbf{x}} [P_{\text{np}}(\hat{X}_n = 1|X_n = 0)] \geq 0. \end{aligned} \quad (\text{A.40})$$

⁴At the first iteration, the a priori information is equal zero.

To prove (A.40), it suffices to show that $R_{1|0} \leq 1$ and $R_{0|1} \leq 1$, where

$$R_{1|0} = \frac{\overline{P}_{\text{np}}(\hat{X}_n = 1|X_n = 0)}{\overline{P}_{\text{p}}(\hat{X}_n = 1|X_n = 0)}, \quad (\text{A.41})$$

$$R_{0|1} = \frac{\overline{P}_{\text{np}}(\hat{X}_n = 0|X_n = 1)}{\overline{P}_{\text{p}}(\hat{X}_n = 0|X_n = 1)}. \quad (\text{A.42})$$

For the 2-tap ($f_0, f_1 = \sqrt{1 - f_0^2}$) channel, it suffices to consider $0.5 \leq f_0^2 \leq 1$ due to the symmetry property of $R_{1|0}$ and $R_{0|1}$ with respect to f_0 and f_1 . The plot of $R_{1|0}$ and $R_{0|1}$ for different 2-tap channels and noise variances is shown in Fig. 32.⁵ It should be noted that $R_{1|0}$ and $R_{0|1}$ are continuous and monotonic functions of f_0 at fixed E_b/\mathcal{N}_0 . It is clear that in all cases, $R_{1|0} \leq 1$ and $R_{0|1} \leq 1$.

It should be pointed out here that computation of $R_{1|0}$ and $R_{0|1}$ does not involve simulating the channel and the equalizer. Furthermore, the numerical problem does not appear in evaluating $R_{1|0}$ and $R_{0|1}$ since they are ratios of probabilities and upper-bounded by 1.

⁵ $R_{1|0}$ and $R_{0|1}$ converge to their stable values very fast (after no more than 20 time indexes used to compute the forward and backward recursions).

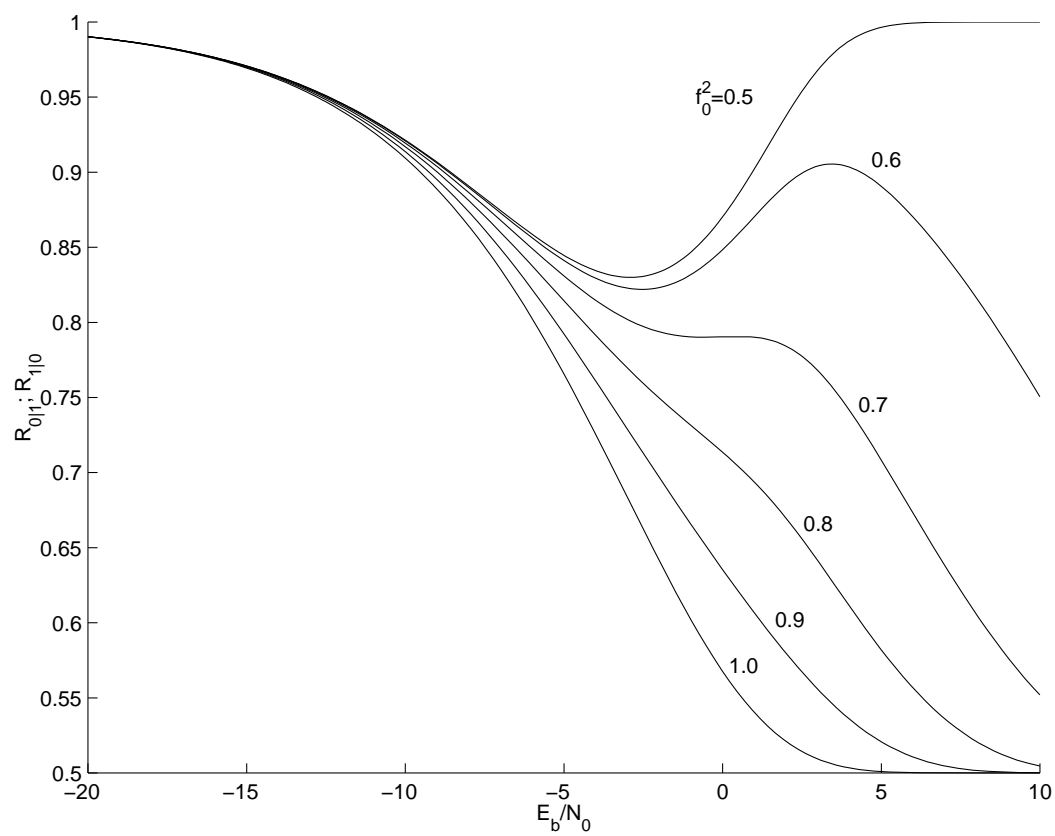


Fig. 32. Plot of $R_{1|0}$ and $R_{0|1}$ as functions of E_b/N_0 for various 2-tap ISI channels.

APPENDIX B

PROOF OF PROPERTY 2, CHAPTER III

To show $\overline{CER}_{\text{np}}^{(1)} \geq \overline{CER}_{\text{p}}^{(1)}$, it is equivalent to show

$$\begin{aligned} \mathbb{E}_{\mathbf{z}; \mathbf{x}} & \left[\frac{P_{\text{np}}^{(1)}(\hat{X}_n = 1 | X_n = 1)}{P_{\text{np}}^{(1)}(\hat{X}_n = 0 | X_n = 1)} - \frac{P_{\text{p}}^{(1)}(\hat{X}_n = 1 | X_n = 1)}{P_{\text{p}}^{(1)}(\hat{X}_n = 0 | X_n = 1)} \right] \\ & + \mathbb{E}_{\mathbf{z}; \mathbf{x}} \left[\frac{P_{\text{np}}^{(1)}(\hat{X}_n = 0 | X_n = 0)}{P_{\text{np}}^{(1)}(\hat{X}_n = 1 | X_n = 0)} - \frac{P_{\text{p}}^{(1)}(\hat{X}_n = 0 | X_n = 0)}{P_{\text{p}}^{(1)}(\hat{X}_n = 1 | X_n = 0)} \right] \geq 0. \quad (\text{B.1}) \end{aligned}$$

Since $P(\hat{X}_n = 1 | X_n = x_n) + P(\hat{X}_n = 0 | X_n = x_n)$ can be normalized to value one, to show (B.1), it suffices to show

$$\mathbb{E}_{\mathbf{z}; \mathbf{x}} \left[\frac{1}{P_{\text{np}}(\hat{X}_n = 0 | X_n = 1)} - \frac{1}{P_{\text{p}}(\hat{X}_n = 0 | X_n = 1)} \right] \geq 0 \quad (\text{B.2})$$

and

$$\mathbb{E}_{\mathbf{z}; \mathbf{x}} \left[\frac{1}{P_{\text{np}}(\hat{X}_n = 1 | X_n = 0)} - \frac{1}{P_{\text{p}}(\hat{X}_n = 1 | X_n = 0)} \right] \geq 0. \quad (\text{B.3})$$

We first prove the following claim:

Claim 1: Let p_U and p_V be pdfs of random variables U and V , respectively, where $U, V \in [0, 1]$. If p_U and p_V satisfy the following conditions: p_U and p_V cross over each other at only one point whose abscissa is in $(0, 1)$ and $\mathbb{E}[U] \geq \mathbb{E}[V]$, then $\mathbb{E}[1/U] \leq \mathbb{E}[1/V]$.

Proof: Let p_U and p_V meet at point A with the abscissa $x_A \in (0, 1)$, and $h = p_U - p_V$. The following must be true about h : $h(x) \leq 0$, for $x \leq x_A$, $h(x) \geq 0$, for

$x > x_A$, and $\int_0^1 h(x)dx = 0$. Let $\Delta = \mathbb{E}[1/U] - \mathbb{E}[1/V]$. Then,

$$\begin{aligned}
 \Delta &= \int_0^1 \frac{h(x)}{x} dx \\
 &= \int_0^{x_A} \frac{h(x)}{x} dx + \int_{x_A}^1 \frac{h(x)}{x} dx \\
 &\leq \frac{1}{x_A} \int_0^{x_A} h(x) dx + \frac{1}{x_A} \int_{x_A}^1 h(x) dx = 0.
 \end{aligned} \tag{B.4}$$

In (B.4), we have used the integral theorem.

By Claim 1 and the fact that $R_{0|1} \leq 1$ and $R_{1|0} \leq 1$, the proof is complete.

VITA

Dũng Ngoc Doan was born in 1976 in Dong Trieu, Quang Ninh, Vietnam. He received his B.E. degree from Hanoi University of Technology, Hanoi, Vietnam in 1997, his M.E. degree from the Asian Institute of Technology (AIT), Bangkok, Thailand in 1999, and his Ph.D. degree from Texas A&M University, College Station, TX, USA in 2004; all in Telecommunications. From January 2000 to October 2003, he was a research assistant in the Wireless Communications Laboratory, Department of Electrical Engineering at Texas A&M University, working toward his Ph.D. degree. From October 2003 to April 2004, he was an intern at the Nokia Research Center, Irving, Texas, working on a CDMA2000 1xEV-DV project. He received best undergraduate and best graduate honors. His research interests are in the area of digital communications with an emphasis on error control coding techniques. His permanent address is: 15 Tran Nhan Tong, Dong Trieu, Quang Ninh, Vietnam. He can be contacted through email: dung_ngoc_doan@yahoo.com.

The typist for this thesis was Dũng Ngoc Doan.