UNIVERSITY OF CALIFORNIA

Los Angeles

New Results in Wireless Communications

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering

by

Wei Shi

2000

The dissertation of Wei Shi is approved.

Gregory J. Pottie

Babak Daneshrad

Mario Gerla

Richard D. Wesel, Committee Chair

University of California, Los Angeles

2000

DEDICATION

To my parents, Tom and Frank

Contents

List of Figures

ACKNOWLEDGEMENTS

I would like to give my most sincere thanks to my academic advisor, Professor Rick Wesel, for his wonderful advice, everlasting encouragement and patience. I am truly fortunate to have been accepted as his student. He is the best advisor I can imagine. His brilliant guidance and understanding provided much needed help for me in time of difficulties.

I like to thank Professor Kung Yao for advising me in course study when I came to UCLA. Professor Lieven Vandenberghe gave me many interesting comments and suggestion in my datafusion problem.

I also like to express my gratitude and appreciation to Professor Greg Pottie, Professor Babak Daneshrad and Professor Mario Gerla for serving on my dissertation committee and providing constructive comments.

Many thanks to my group members, Christos, Christine, Xueting, Mark, Adina and Cenk, for being friendly and supportive. Special thanks to Christos, Christine and Xueting for their useful discussions and help in my research. Thanks to Xueting for overwhelmingly joyful chatting.

I can never forget the sweet surprise that Professor Wesel, Celina, Christos, Christine, Xueting and Richard gave to me by wonderful gifts when I had my baby at the hospital.

Very importantly, I like to thank my parents and my sister, Qianwen, for their constant love, support and encouragement through my lifetime. Thanks to Tom Sun, my husband, for his love, understanding and support. He is always there when I need help. Finally I like to thank Frank Sun, my baby, for being my constant source of joy.

VITA

1984	B.S. in Electrical Engineering Huazhong University of Science & Technology Wuhan, P.R.China
1984-1994	Lecturer Nanjing Mechanical & Electrical College Nanjing, P.R.China
1996	M.S. in Electrical Engineering University of Massachusetts, Amherst Amherst, MA

PUBLICATIONS

- W. Shi, T. W. Sun and R. D. Wesel, "Quasiconvexity and optimal binary fusion for distributed detection with identical sensors in generalized Gaussian noise," accepted as a correspondence in *IEEE Transactions on Information Theory.*

- R. D. Wesel,X. Liu and W. Shi, "Trellis codes for periodic erasures," *IEEE Transactions on Communications, Vol.48, No.6,* June 2000.

- W. Shi, T. W. Sun and R. D. Wesel, "Optimal binary distributed detection," *Proc. of 2000 IEEE International Symposium on Information Theory,* June, 2000.

- W. Shi, T. W. Sun and R. D. Wesel, "Optimal binary thresholds for distributed detection in Gaussian noise," *Proc. of the 33th Asilomar Conference on Signals, System & Computers,* Oct., 1999.

- W. Shi and R. D. Wesel, "When the best decision-feedback equalizer is a linear equalizer," *Proc. of the 36th Annual Allerton Conference on Communications, Control and Computing,* Sept., 1998.

- W. Shi and R. D. Wesel, "The effect of mismatch on decision-feedback equalization and Tomlinson-Harashima precoding," *Proc. of the 32th Asilomar Conference on Signals, System & Computers,* Nov., 1998.

- C. G. Cassandras and W. Shi, "Perturbation analysis of multiclass multi-objective queueing systems with 'quality of service' guarantees," *Proc. of the 35th IEEE Conference on Decision and Control,* 1996.

ABSTRACT OF THE DISSERTATION

New Results in Wireless Communications

by

Wei Shi

Doctor of Philosophy in Electrical Engineering
University of California, Los Angeles, 2000
Professor Richard D. Wesel, Chair

A number of new results for wireless communications are presented. First, this thesis shows that finite-impulse-response (FIR) linear equalizer (LE) is, for some channels and input SNRs, the optimal solution among all the possible FIR DFEs with the same number of feedforward taps. When the FIR-DFE does not outperform the FIR-LE, we show that a corresponding reverse-time FIR-DFE will. The second result investigates the behavior of Decision-Feedback Equalization (DFE) and Tomlinson-Harashima Precoding (THP) with channel estimation error in wireless communications. The third result shows the robustness of space-time turbo codes in flat fading channels and frequency selective channels with 2 transmit antennas and 2 receive antennas. The last result presents a technique to find the optimal threshold τ and admissible fusion rule for the binary hypothesis detection problem with n identical and independent sensors in wireless sensor network. The property of quasiconvexity holds for additive generalized Gaussian

noises and some non-Gaussian noise distributions.

Chapter 1

Introduction

In communication channels with intersymbol interference (ISI), a decision-feedback equalizer (DFE) can be used at the receiver side. When infinite-impulse-response (IIR) feedforward and feedback filters are employed, the best decision-feedback equalizer (DFE) includes a nonzero feedback term unless the channel is a simple impulse. However, this is not true in the practically important case where the feedforward filter is constrained to have a finite number of taps. Usually the number of feedforward taps is fixed while the number of feedback taps can be varied in some range. Chapter 2 shows that finite-impulse-response (FIR) linear equalizer (LE) is, for some channels and input SNRs, the optimal solution among all the possible FIR DFEs with the same number of feedforward taps. When the FIR-DFE does not outperform the FIR-LE, we show that a corresponding reverse-time FIR-DFE will.

However the DFE structure has error propagation, so we consider Tomlinson-Harashima Precoding (THP) in Chapter 3 as well as DFE for use in an asymmetric manner on indoor wireless LANs by Sesay and Gibbard [1, 2, 3]. In the asymmetric system, THP is used for the link from the base station to the portable while DFE is used in the reverse link. Thus, in each case the portable does as little processing as possible. One concern with such an approach is the possible channel mismatch of the precoding employed by the THP transmitter and the possible channel mismatch of the DFE receiver with respect to the actual multipath at the time of transmission.

In Chapter 3, we investigate the behavior of DFE and THP with channel estimation error. Results include computation of the output SNR of a DFE with channel mismatch at the receiver and a lower bound on the achievable rates for THP with channel mismatch at both transmitter and receiver and with channel mismatch only at the transmitter. In each case we also present high SNR asymptotic results.

The scenario of one transmit antenna and one receive antenna can be extended to multiple transmit and receive antennas in wireless communications. High data rate wireless communications have motivated the use of multiple transmit and receive antennas. In Chapter 4, we consider the performance of turbo codes with 2 transmit antennas and 2 receive antennas for flat fading channels. We consider robustness within and among families of channels with the same singular values (SV's). Iterative decoding is the same as for standard parallel concatenated convolutional codes, except that the extrinsic information about encoder outputs should also be exchanged.

Chapter 5 shows the robustness of turbo codes in frequency selective channels with 2 transmit antennas and 2 receive antennas. The robustness is achieved in terms of proximity of channel capacity. G.Bauch [4] proposed a MAP equalizer/decoder for space-time coded signals over frequency selective channels. This chapter extends this MAP equalizer/decoder to space-time turbo codes and shows that the decoding results is close to channel capacity.

Chapter 6 is an application of wireless communications to wireless sensor networks. In wireless sensor networks, efforts have been made in designing the thresholds of individual sensors and the optimal fusion rule in the fusion center to get optimal performance. In Chapter 6, we present a technique to find the optimal threshold τ for the binary hypothesis detection problem with n identical and independent sensors in wireless sensor network. The sensors all use an identical and single threshold τ to make local decisions, and the fusion center makes a global decision based on the n local binary decisions. For generalized Gaussian noises and some non-Gaussian noise distributions, we show that for any admissible fusion rule, the probability of error is a quasiconvex function of threshold τ. Hence, the problem decomposes into a series of n quasiconvex optimization problems that may be solved using well known techniques.

Assuming equal *a-priori* probabilities, we give a sufficient condition of the non-Gaussian noise distribution $g(x)$ for the probability of error to be quasiconvex. Furthermore, this technique is extended to Bayes risks and Neyman-Pearson criteria.

The main contribution, presented in this thesis, are:

- The channel conditions required for the FIR LE to be the best FIR DFE (Chapter 2).
- Analysis of channel mismatch effect on DFE and THP (Chapter 3).
- An iterative decoding algorithm for space-time turbo-codes (Chapter 4).
- A technique of time-varying symbol-wise linear transformation (Chapter 4).

- An application of MAP equalizer/decoder to space-time turbo codes (Chapter 5).

- A sufficient condition of the non-Gaussian noise distribution $g(x)$ for the probability of error to be quasiconvex (Chapter 6).

- A technique of designing optimal systems for binary distributed detection (Chapter 6).

Chapter 2

When the Best Decision-Feedback Equalizer is a Linear Equalizer

Most DFE analysis studies the DFE with IIR filters [5, 6, 7, 8]. The IIR DFE strictly outperforms IIR LE except for the trivial channel with just one tap. The FIR DFE was introduced in [7, 9, 10]. Al-Dhahir and Cioffi established finite-length analogs of the well-known IIR MMSE-DFE results and examined the similarities and differences between FIR MMSE-DFE and IIR MMSE-DFE in [10]. This chapter shows that the restriction to finite length filters [10, 9] allows FIR LE to outperform any FIR DFE with nonzero feedback for some channels and for a practical range of input SNR. A comparison of quadratic forms reveals whether or not an LE is optimal in a particular DFE configuration. This chapter demonstrates that, for those channels where the best FIR DFE is a FIR LE, it is beneficial to process blocks backwards in time using a FIR DFE. Finally

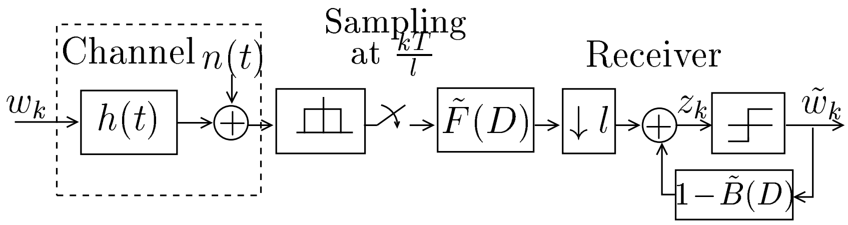

Figure 2.1: FIR DFE system

this chapter also examines fractionally spaced ($\frac{T}{2}$-spaced) FIR DFEs by means of output SNR and compares them with T-spaced FIR DFEs, over the same time span and with the same number of taps. Based on [11, 12], one might expect that these $\frac{T}{2}$-spaced DFEs to behave like IIR DFEs and strictly outperform FIR LEs. However, for any n-tap channel in the T-spaced case, the $\frac{T}{2}$-spaced FIR DFEs do not have IIR DFE behavior since the subchannels always have common zeros. Thus, even in $\frac{T}{2}$-spaced case, FIR DFE does not strictly outperform FIR LE.

2.1 T-spaced FIR DFE

Figure 2.1 shows a fractionally spaced FIR DFE, where $h(t)$ is the cascade of a channel impulse response and a transmitter pulse-shaping filter, and $n(t)$ is additive white Gaussian noise with variance $N_0/2$. In this chapter, we consider T-spaced (i.e. symbol-spaced) FIR DFE with the oversampling factor $l = 1$ and $\frac{T}{2}$-spaced FIR DFE with $l = 2$. An anti-alias lowpass filter with gain $\sqrt{T}$ over the frequency range $-\frac{l\pi}{T} \leq \omega \leq \frac{l\pi}{T}$ is used before the sampler in FIR DFE systems [9], where T is the symbol interval. The variance of the noise samples is $\frac{N_0}{2}l$. For simplicity, $H(D)$ refers to the discrete time impulse response of the cascade of transmitter filter, channel, anti-aliasing lowpass filter and sampler.

In Figure 2.1, let $\tilde{F}(D)$ and $\tilde{B}(D)$ be feedforward and feedback filters with finite length constraints. Assume $\tilde{F}(D)$ has N_f taps and $\tilde{B}(D)$ has N_b taps (excluding the monic term). Despite the finite length restriction, $\tilde{B}(D)$ still must be monic and causal. For the feedforward filter $\tilde{F}(D)$, the choice of which N_f consecutive taps to make nonzero is an important consideration. The integer delay $\triangle$ specifies the position of these taps. Specifically, the nonzero taps are the coefficients of $D^{-\triangle}, \cdots, D^{N_f-\triangle-1}$.

The maximum number N_b of useful feedback taps depends on the parameter $\triangle$. For a given N_f and a channel impulse response $H(D) = h_0 + h_1 D + \cdots + h_v D^v$, the delay parameter $\triangle$ can vary from 0 to $N_f + v - 1$. Correspondingly, the useful number of feedback taps N_b must satisfy

$$N_b = N_f + v - 1 - \triangle \tag{2.1}$$

Note that, when $N_b = 0$, the FIR DFE is actually an FIR LE.

Consider the four-tap maximum phase channel $H(D) = 0.05 + 0.5D + 2D^2 + 4D^3$ under the restriction of $N_f = 3$. Figure 2.2 shows output SNR vs. input SNR for the possible values of $\triangle$.Output SNR is the SNR at the input of the slicer. Given $N_f = 3$, N_b is chosen to have the maximum number of useful feedback taps with $\triangle$ in the range of [0,5]. The DFE system has the best performance when $N_f = 3, N_b = 0$ and $\triangle = 5$ for input SNR in the range of [5dB, 50dB]. We have the surprising result that the optimal FIR DFE is simply a FIR LE since no feedback filter $\tilde{B}(D)$ exists. For input $SNR < 20dB$, the FIR LE performs close to the optimal IIR DFE.

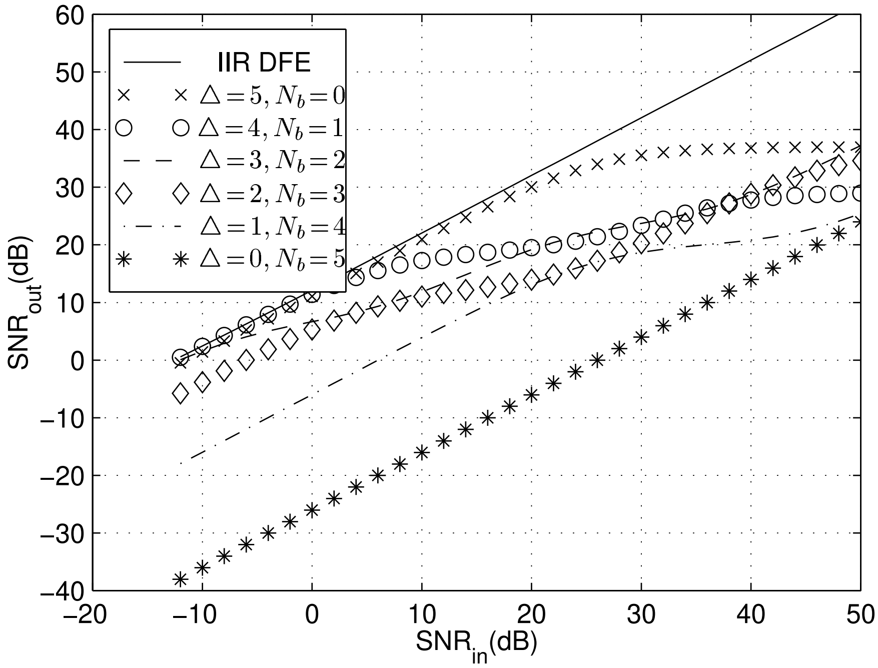

Figure 2.2: SNR_{out} vs. SNR_{in} for the possible choice of $\triangle$, $N_f = 3$ and $H(D) = 0.05 + 0.5D + 2D^2 + 4D^3$

2.2 Channels where FIR LE is optimal

This section identifies what kind of channels can perform better with FIR LE ($N_b = 0$) than all the possible FIR DFEs ($0 < N_b \leq N_f + v - 1$) for some given N_f and a specific input SNR.

Let the channel impulse response $H(D) = h_0 + h_1 D + \cdots + h_v D^v$. Define two matrices H and $\tilde{H}$ as

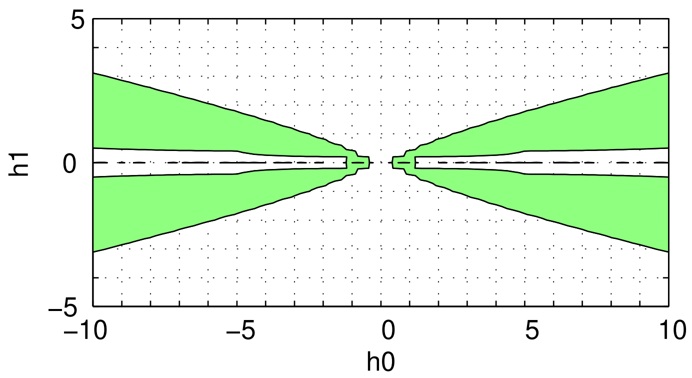

Figure 2.3: Two-tap channels $H(D) = h_0 + h_1 D$ with optimal LE for $N_f = 4$ and input SNR 30dB

$$H = \begin{bmatrix} h_0 & h_1 & \cdots & h_v & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_v & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_v \end{bmatrix} \tag{2.2}$$

and

$$\tilde{H} = \begin{bmatrix} & H & \\ 0_{N_b \times (\triangle+1)} & & I_{N_b} \end{bmatrix} \tag{2.3}$$

where $H \in R^{N_f \times (N_f+v)}$, $\tilde{H} \in R^{(N_f+N_b)\times(N_f+v)}$, I_n refers to the $n \times n$ identity matrix and $0_{m\times n}$ refers to the $m \times n$ matrix of all zeros.

In order to have the FIR minimum-mean-square-error (MMSE) LE as the optimal solution for any given N_f, the MMSE with $N_b = 0$ (i.e. $\triangle = N_f + v - 1$) must be less than the MMSE with $0 < N_b \leq N_f + v - 1$. This condition turns out to be satisfied if and only if the following inequalities hold for each possible combination of $\triangle$ and N_b under the restriction of $\triangle + N_b = N_f + v - 1$ and $0 < N_b \leq N_f + v - 1$.

$$\tilde{H}^T(\triangle+1)\left[\tilde{H}\tilde{H}^T+\frac{1}{SNR}\begin{bmatrix} I_{N_f} & 0_{N_f\times N_b} \\ 0_{N_b\times N_f} & 0_{N_b\times N_b}\end{bmatrix}\right]^{-1}\tilde{H}(\triangle+1)$$
$$\leq H^T(N_f+v)\,[HH^T+\frac{1}{SNR}I_{N_f}]^{-1}H(N_f+v) \tag{2.4}$$

where $\tilde{H}(n)$ and $H(m)$ refer to the nth column of $\tilde{H}$ and the mth column of H respectively. Here $SNR = \frac{\bar{\varepsilon}_w}{N_0/2}$ is the input SNR.

In Figure 2.3, the shaded area shows the set of two-tap channels $H(D) = h_0 + h_1 D$ having LE as the best solution for $N_f = 4$ and $SNR_{in} = 30dB$. This set of channels is a subset of maximum phase two-tap channels.

2.3 Reverse-Time Channels

For those channels where the best DFE is an LE, it is beneficial to process blocks backward in time using a DFE.

Figure 2.7 shows the original LE system with delay parameter $\triangle$, where $H(D) = \sum_{i=0}^{v} h_i D^i$ and $\tilde{F}(D) = \sum_{i=0}^{f} h_i D^{-\triangle+i}$, and Figure 2.8 shows the reverse-time LE system, where $H_r(D) = \sum_{i=0}^{v} h_{v-i} D^i$ and $\tilde{F}_r(D) = \sum_{i=0}^{v} f_{N_f-1-i} D^{-(N_f+v-1-\triangle)+i}$.

It is easily shown that

$$H_r(D) = D^v H(D^{-1}). \tag{2.5}$$

$$\tilde{F}_r(D) = D^{-v}\tilde{F}(D^{-1}). \tag{2.6}$$

2.3.1 Equivalent Unbiased Output SNRs

From (2.5) and (2.6), $H_r(D)\tilde{F}_r(D) = H(D^{-1}\tilde{F}(D^{-1}$. Given the unbiased original LE, then the reverse time LE is also unbiased. So the residual ISI in both systems as shown in Figures 2.7 and 2.8 have the same second moments and the output noise has the same variance from 2.6. Therefore, the system in 2.7 has the same output SNR as the system in 2.8.

2.3.2 Optimal DFE in Reverse-Time Channel $H_r(D)$

For channels $H(D)$ satisfying (2.4), the LE $\tilde{F}(D)$ with $\triangle = N_f - 1 - v$ is the best among all the possible LEs with $0 \leq \triangle < N_f - 1 - v$. As shown in the preceding subsection 2.3.1, for the reverse-time channels $H_r(D)$, $\tilde{F}_r(D)$ with $\triangle = 0$ is the best among all the possible LEs with $0 < \triangle \leq N_f - 1 - v$. However the optimal LE $\tilde{F}_r(D)$ with $\triangle = 0$ introduces the post-cursor residual ISI. Thus the DFE is optimal since the feedback can be used to eliminate all the possible post-cursor residual ISI.

Figure 2.9 shows the performance of the reverse-time channel $H_r(D) = 4 + 2D + 0.5D^2 + 0.05D^3 = D^3H(D^{-1})$, where the original channel $H(D)$ is the four-tap channel as in Figure 2.2. The FIR DFEs with $N_b \geq 3$ have negligible performance loss compared to IIR DFE. However the output SNR of FIR DFEs with $N_b \leq 2$ is bounded and is less than 0dB due to the residual ISI. The optimal FIR LE with $N_f = 3, \triangle = 0, N_b = 0$ for the reverse-time channel $H_r(D)$ has the same performance as the FIR LE with $N_f = 3, \triangle = 5, N_b = 0$ for the original channel $H(D)$. The FIR DFEs outperform the FIR LE because the feedback can eliminate the post-cursor ISI when $N_b \geq 3$.

2.4 $\frac{T}{2}$-spaced FIR DFE

The feedforward filter converts the system to minimum phase in zero-forcing(ZF) IIR DFE and to near minimum phase in MMSE IIR DFE, Mai and Sayed demonstrated in [12] that, for FIR channels, fractionally spaced FIR LE can be a perfect zero-forcing equalizer under length and zero conditions and therefore can perform as well as IIR LE for noise free FIR channels. We expect the fractionally spaced FIR DFEs have performance similar to the IIR DFE. We examine the $\frac{T}{2}$-spaced FIR DFE performance by an example in this section.

The $\frac{T}{2}$-spaced FIR DFE consists of the $\frac{T}{2}$-spaced feedforward filter $\tilde{F}(D)$ and a symbol-spaced (or T-spaced) feedback filter $\tilde{B}(D)$ as shown in Figure 2.1 when $l = 2$. Thus, the useful number of feedback taps $N_b = N_f/2 + v - 1 - \triangle$ and $0 \leq \triangle \leq N_f/2 + v - 1$, where the delay parameter $\triangle$ is a multiple of symbol interval T as defined in the T-spaced case.

We consider the output SNR for IIR DFE, T-spaced FIR DFE and $\frac{T}{2}$-spaced FIR DFE. In the IIR DFE system, a matched filter is used before the sampler instead of the anti-aliasing lowpass filter. Assume the continuous time channel (the transmitter pulse shape is not included) is $\tilde{h}(t) = 0.3\delta(t) + \delta(t-T)$. The transmitter pulse shape is a sinc function with bandwidth $W = \frac{1}{2T}$ for T-spaced FIR DFE while the transmitter pulse shape is $f_{rc}(t)$, a raised cosine with rolloff factor $\alpha = 1$, for $\frac{T}{2}$-spaced FIR DFE.

In Figures 2.4 and 2.5, the output SNR is compared over the same time span and, thus, N_f is twice as large for $\frac{T}{2}$-spaced FIR DFE as for T-spaced FIR DFE. For the $\frac{T}{2}$-spaced case, FIR LE is optimal only for input SNR in the range of [42dB, 50dB] while, for the T-spaced case, FIR LE is optimal for input SNR in the range [30dB, 50dB]. FIR DFE with $N_b = 1$ in the $\frac{T}{2}$-spaced case performs much better than FIR DFE with $N_b = 1$ in the T-spaced case.

Figures 2.5 and 2.6 compare T-spaced and $\frac{T}{2}$-spaced DFEs with the same number of feedforward taps. These plots show that the $\frac{T}{2}$-spaced FIR LE with $N_f = 4$ is optimal only for input SNR in [10dB, 42dB]. The result also shows that, with same number of feedback taps N_b, T-spaced FIR DFE in twice the time span does much better than the $\frac{T}{2}$-spaced FIR DFE for this channel.

Thus,the $\frac{T}{2}$-spaced FIR DFEs do not have the expected IIR DFE behavior. The $\frac{T}{2}$-spaced FIR DFE with $N_b = 1$ does not strictly outperform $\frac{T}{2}$-spaced FIR LE for all input SNR. Second, the $\frac{T}{2}$-spaced FIR LE in Figure 2.6 has residual ISI since its output SNR is bounded as input SNR increases without bound.

By the two subchannel view of a $\frac{T}{2}$-spaced linear equalizer [12], the $\frac{T}{2}$-spaced FIR DFE is equivalent to two T-spaced subchannels followed by the slicer and the feedback filter. The two subchannels have common zeros for our continuous channel model $\tilde{h}(t)$, which is a two-tap channel in the T-spaced case. So the conditions under which fractionally spaced FIR equalizers have IIR equalizer behavior are violated.

When the transmitter pulse shape is $f_{rc}(t)$ and $\tilde{h}(t) = \sum_{k=0}^{k=n} a_k \delta(t-kT)$, which is an n-tap channel for the T-spaced case, the two discrete subchannels can be represented respectively by vectors $C_e = [a_0 \; a_1 \; \cdots \; a_n]$ and $C_o = [a_0/2 \; (a_0 + a_1)/2 \; \cdots \; (a_{n-1} + a_n)/2 \; a_n/2]$. It is easily shown that vector C_o is just the convolution of vectors C_e and [1/2 1/2]. By polynomial representation of vectors, let $C_e(z) = \sum_{k=0}^{k=n} a_k z^k$ and then $C_o(z)$ can be expressed as $C_o(z) = \frac{1}{2}(z+1)C_e(z)$. So the subchannels always have common zeros. Specificly, for our example of a two-tap channel, $n = 1, \; a_0 = 0.3$ and $a_1 = 1$.

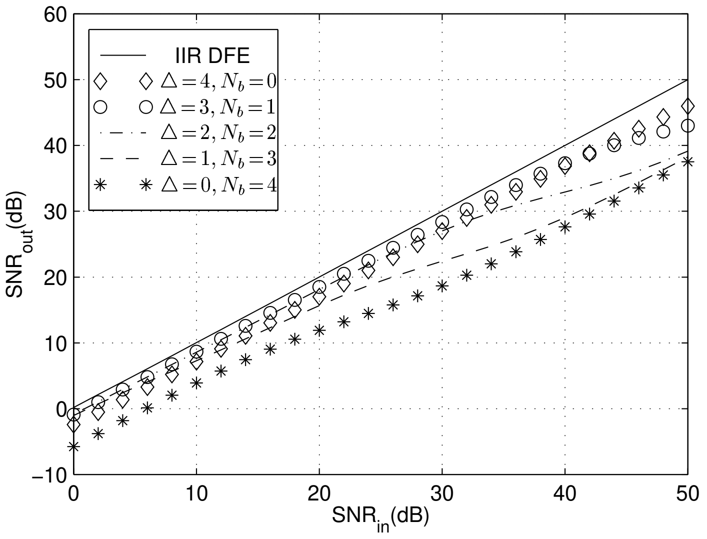

Figure 2.4: SNR_{out} vs. SNR_{in}, $N_f = 8$ in $\frac{T}{2}$-spaced MMSE FIR DFE

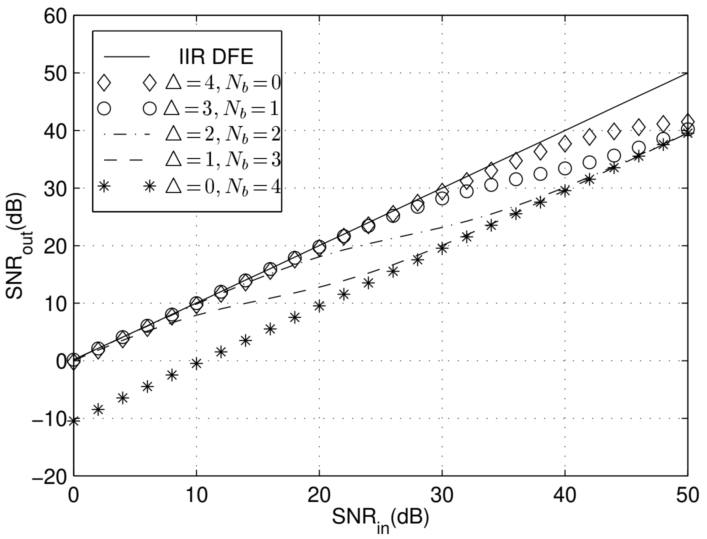

Figure 2.5: SNR_{out} vs. SNR_{in}, $N_f = 4$ in T-spaced MMSE FIR DFE

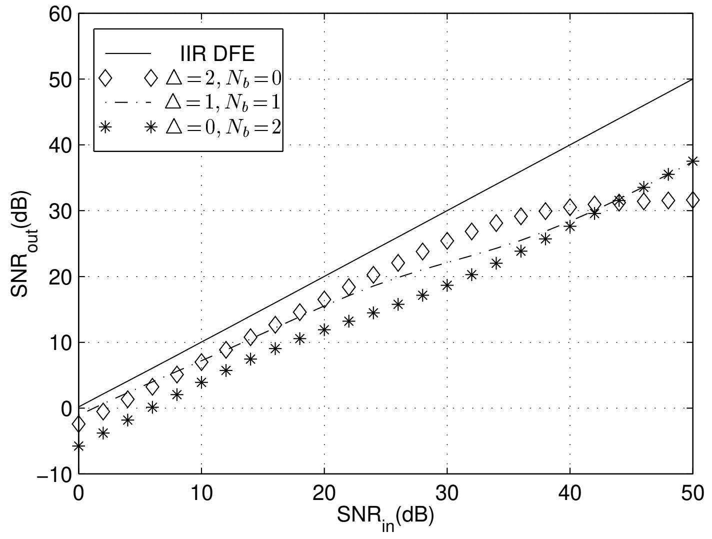

Figure 2.6: SNR_{out} vs. SNR_{in}, $N_f = 4$ in $\frac{T}{2}$-spaced MMSE FIR DFE

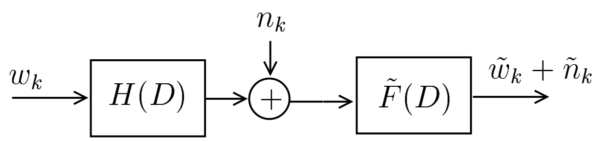

Figure 2.7: The original FIR LE system

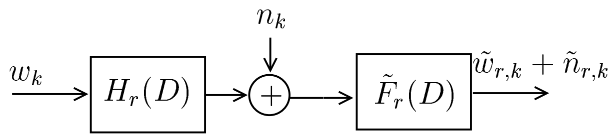

Figure 2.8: The reverse-time FIR LE system

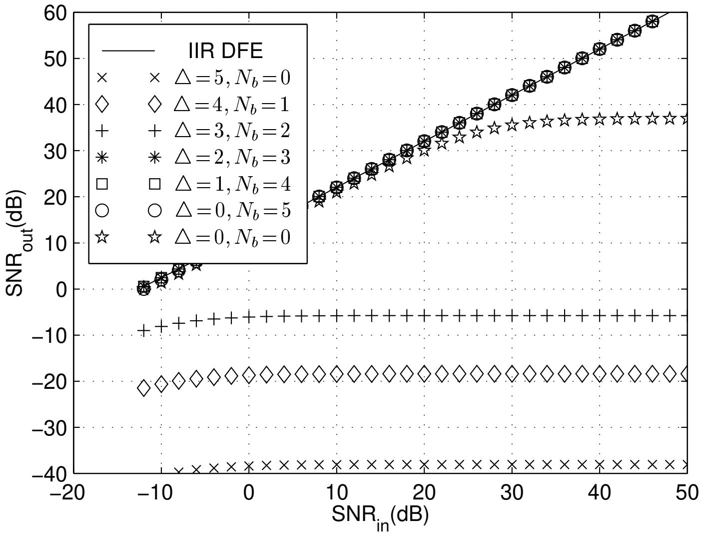

Figure 2.9: SNR_{out} vs. SNR_{in} for the possible choice of $\triangle$, $N_f = 3$ and $H_r(D) = 4 + 2D + 0.5D^2 + 0.05D^3$

Chapter 3

The Effect of Mismatch on Decision-Feedback Equalization and Tomlinson-Harashima Precoding

In high speed wireless communication systems, channels are modeled as multipath channels. Using equalization and channel decoding is one way, with low complexity at the receiver, to achieve the near capacity performance. Another way to achieve the near capacity performance is using MAP equalizer/decoder at the receiver with high complexity for long delay channels. The second method will be considered in Chapter 5. This chapter considers the effect of channel mismatch in wireless communications on a Decision-Feedback Equalizer (DFE)[5] and Tomlinson-Harashima Precoding (THP) [13, 14, 15].

DFE and THP have been considered for use in indoor wireless LANs on an asymmetric manner proposed by Sesay and Gibbard [1, 2, 3]. In the asymmetric system, THP is used for the link from the basestation to the portable while DFE is used in the reverse link. Thus, in each case the portable does as little processing as possible. One concern with such an approach is the possible channel mismatch of the precoding employed by the THP transmitter and the possible channel mismatch of the DFE receiver with respect to the actual multipath at the time of transmission. Another concern is the possible mismatch introduced by using finite-impulse-response (FIR) feedback and feedforward filters.

In this chapter, for the THP, we quantitatively describe the information rates achievable over the entire range of SNR under the MMSE or ZF criterion with mismatch and quantify loss from information rates without mismatch. We consider two mismatch conditions. In the first scenario, the transmitter and receiver share the same incorrect channel estimate. In the second scenario, the receiver has perfect knowledge of the channel, but the transmitter still has an incorrect estimate. This second mismatch condition models the case where mismatch is primarily due to a delay in communicating channel estimates from the receiver to the transmitter.

For the DFE, we quantitatively describe the output SNR over the entire range of input SNR under MMSE criterion with mismatch and quantify loss from output SNR without mismatch. Unlike THP, we only consider one mismatch, which is the channel mismatch at the receiver.

The loss from the information rates achievable in THP without mismatch at a particular SNR depends on the specific channel impulse response and the severity of the mismatch. Several impulse response and mismatch conditions are studied as examples. At high SNR, the asymptotic limitations on the information rates are a direct result of the residual ISI caused by the channel estimation mismatch at both transmitter and receiver sides. With only the transmitter suffering channel estimation error, a correctly designed feedforward filter allows the information rate to increase without bound as SNR increases. However transmitter mismatch can still cause non-negligible loss.

Similar to THP with channel mismatch at the receiver, the loss from the output SNR in a DFE without mismatch at a particular input SNR depends on the specific channel impulse response and the severity of the mismatch. The output SNR is bounded even with input SNR unbounded.

3.1 Mismatched DFE

3.1.1 Mismatch Model

We first briefly review ideal DFE analysis and introduce mismatch. Figure 3.1 shows the mismatched DFE system. $\hat{B}(D)$ and $\hat{F}(D)$ correspond to the feedback and feedforward filters of the DFE designed for $\hat{H}(D) \neq H(D)$. $\hat{B}(D)$ must be causal and monic. $\hat{F}(D)$ needs to be linear and time-invariant. Figure 3.2 is equivalent to Figure 3.1 assuming no error propagation. The noise $\hat{n}_k$ is noise n_k filtered by $\hat{F}(D)$. Even the ZF-DFE has nonzero ISI due to the mismatch. Under mismatch, $H(D)\hat{F}(D) - \hat{B}(D)$ has a term at time 0 (the zero-lag tap) and $H(D)\hat{F}(D)$ is not monic. Let g_0 represent the zero-lag tap of $H(D)\hat{F}(D)$. Thus, Figure 3.3 is equivalent to Figure 3.2.

3.1.2 Output SNR in DFE system

Both MMSE-DFE and ZF-DFE are considered in the analysis of the output SNR for the mismatched DFE. In Figure 3.3, the ISI term y_2 contains no current term of w. We can write the output SNR as:

$$SNR_1 = \frac{g_0^2 \sigma_w^2}{\sigma_{y_2}^2 + \sigma_{\hat{n}}^2} \tag{3.1}$$

An alternative SNR is:

$$SNR_2 = \frac{\sigma_w^2}{\sigma_{y_1}^2 + \sigma_{\hat{n}}^2} \tag{3.2}$$

The SNR_1 is unbiased while SNR_2 is not. SNR_1 derived in (3.1) and SNR_2 in (3.2) are equivalent when $g_0 = 1$.

Figure 3.4 shows how channel estimation mismatch reduces the output SNR on example 2 tap channels. The residual ISI limits the output SNR to be bounded even as input SNR increases without bound.

This is true in general. In the limit of high SNR, MMSE-DFE and ZF-DFE are equivalent. Thus, $SNR_1 \to \frac{g_{0,ZF}^2 \sigma_w^2}{\sigma_{y_2,ZF}^2}$ as $\sigma_n^2 \to 0$ where $g_{0,ZF}$ and $\sigma_{y_2,ZF}^2$ are the zero forcing values. Similarly, $SNR_2 \to \frac{\sigma_w^2}{\sigma_{y_1,ZF}^2}$ where $y_{1,ZF}$ is the value y_1 in Figure 3.2, produced by the mismatched ZF-DFE.

3.2 Mismatched THP

Figure 3.5 shows a general mismatched THP in which both transmitter and receiver incorrectly estimate the channel to be $\hat{H}(D)$. Thus, $\hat{B}(D)$ and $\hat{F}(D)$ correspond to the feedback and feedforward filters of the THP designed for $\hat{H}(D) \neq H(D)$. As in THP without mismatch, $\hat{B}(D)$ must be causal and monic

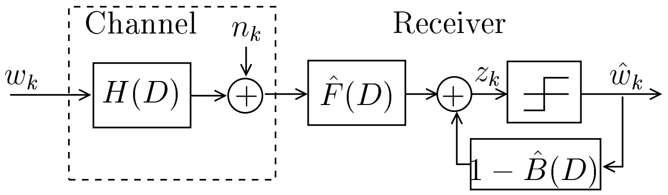

Figure 3.1: DFE system with mismatch.

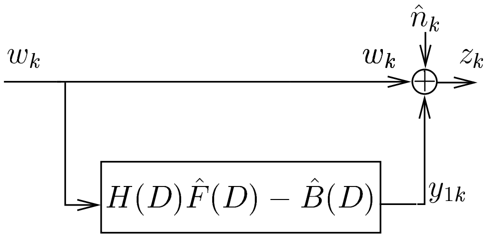

Figure 3.2: Equivalent system to Figure 3.1.

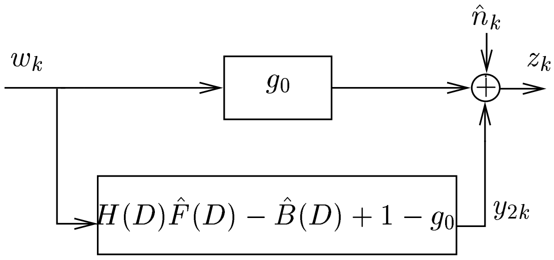

Figure 3.3: Equivalent system to Figure 3.2.

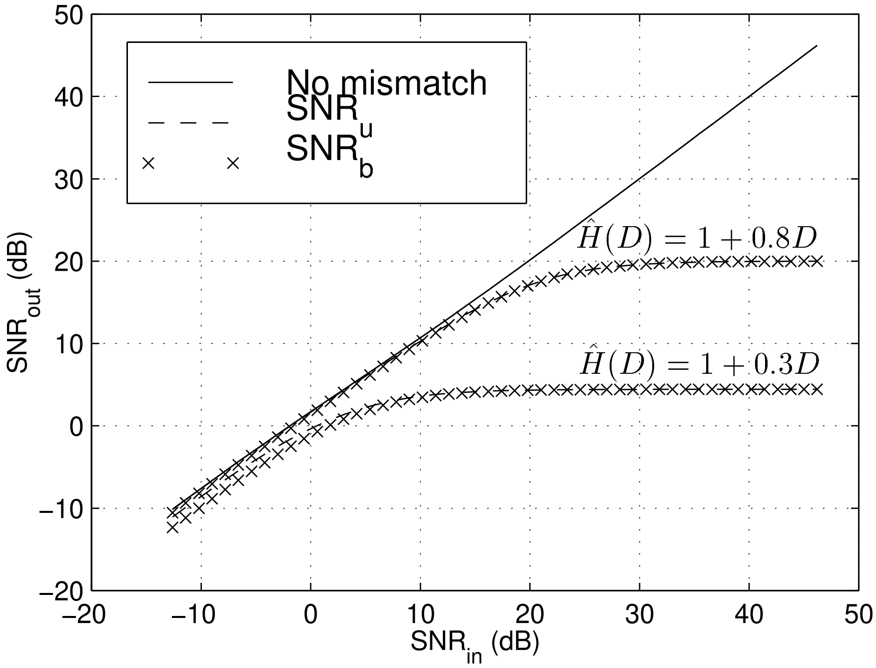

Figure 3.4: Output SNR vs. input SNR for MMSE-DFE with impulse response $H(D) = 1 + 0.9D$

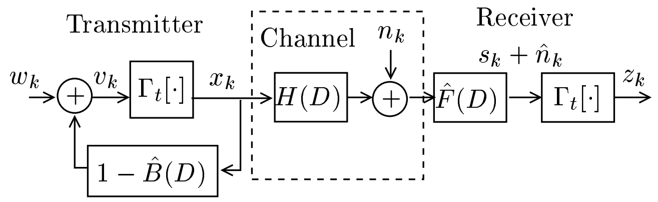

Figure 3.5: THP communication system with mismatch.

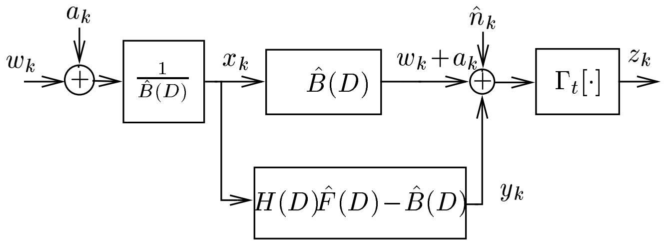

Figure 3.6: THP system with mismatch equivalent to Figure 3.5.

while $\hat{F}(D)$ needs to be linear and time-invariant. The modulo operation Γ_t is a mapping from the real numbers $\Re$ to $(-t/2, t/2]$. Specifically, $\Gamma_t[v_k] = v_k + a_k$ where a_k is the integer multiple of t for which $\Gamma_t[v_k] \in (-t/2, t/2]$. If w_k is an *i.i.d.* sequence with a uniform distribution on $(-t/2, t/2]$, then x_k will also be an *i.i.d.* sequence with a uniform distribution on $(-t/2, t/2]$ regardless of the choice of $\hat{B}(D)$. n_k is a sequence of white Gaussian noise and $\hat{n}_k$ is noise n_k filtered by $\hat{F}(D)$.

For large alphabet PAM [16], the THP transmitter output power is roughly $t^2/12$. Thus we restrict our attention to that value of t for which $t^2/12 = P$, the power constraint. Smaller values ($t < \sqrt{12P}$) give a lower achievable information rate. Larger values of t cause the power constraint to be violated by an i.i.d. PAM w_k sequence. Thus, $t > \sqrt{12P}$ would require w_k to be carefully constructed to satisfy $E[x^2] \leq P$.

Figure 3.6 presents a structure equivalent to Figure 3.5 with a_k as defined in the first paragraph of this section. y_k is the ISI term, which is the output of the filter $L(D) = H(D)\hat{F}(D) - \hat{B}(D)$ with input x_k. The filter $L(D)$ is useful in the calculation of the variance of y_k, which is required for the lower bound of information rates as shown in section 3.2.1.

ZF-THP with mismatch has residual ISI due to the mismatch. Thus, the ISI term y_k in Figure 3.6 is nonzero. Unlike the ZF-THP without mismatch, the ZF-THP with mismatch can not transform an AWGN channel with ISI to an AWGN channel without ISI. The residual ISI becomes more severe as the mismatch increases. Spectral factorization techniques [9] make the $\hat{F}(D)$ an all-pass filter. Thus, the filtered noise $\hat{n}_k$ is white as with the ZF-THP without mismatch.

MMSE-THP, whether having mismatch or not, has residual ISI. Of course, the residual ISI is larger in MMSE-THP with mismatch than without mismatch. As in ZF-THP with mismatch, the severity of ISI depends mainly on the mismatch severity. $\hat{F}(D)$ is not an all-pass filter by spectral factorization techniques [9] so the filtered noise $\hat{n}_k$ is not white for the MMSE-THP.

3.2.1 Channel Mismatch at both Transmitter and Receiver

We now analyze the achievable information rates for the scenario described above, where both transmitter and receiver incorrectly estimate the channel to be $\hat{H}(D) \neq$

$H(D)$. Because of the nonzero ISI for both MMSE-THP and ZF-THP with mismatch, ideal interleaving [17] is assumed. Thus, the channel looks like a discrete memoryless channel(DMC) with input w and output $\Gamma_t[w+y+\hat{n}]$. This assumption introduces a loss, but it is consistent with how THP is used in practice.

The analysis of the information rates is identical for both MMSE-THP and ZF-THP with mismatch, resulting in a lower bound on the achievable rates for either case.

In Figure 3.6, the output can be written as:

$$\begin{aligned} z &= \Gamma_t[w + a + y + \hat{n}] \\ &= \Gamma_t[w + y + \hat{n}]. \end{aligned} \tag{3.3}$$

In the following, the index k is omitted as above when no confusion results. The mutual information of this channel is given as:

$$\begin{aligned} I(w; z) &= h(z) - h(z|w) \\ &= h(\Gamma_t[w + y + \hat{n}]) - h(\Gamma_t[y + \hat{n}]|w). \end{aligned} \tag{3.4}$$

A lower bound on C_M can be found by assigning w an i.i.d. uniform distribution on $(-t/2, t/2]$, where C_M is the capacity of the channel produced by the mismatched MMSE-THP or ZF-THP with ideal interleaving. For such a w, x is i.i.d. uniform on $(-t/2, t/2]$. Let $G(D) = H(D)\hat{F}(D)$. In general, $G(D)$ can be expressed as

$$G(D) = \sum_{k=-\infty}^{\infty} g_k D^k \tag{3.5}$$

In Figure 3.5, the input to the modulo operator $\Gamma_t[\cdot]$ at the receiver side can be expressed as

$$S(D) + \hat{N}(D) = X(D)G(D) + \hat{N}(D) \tag{3.6}$$

$$= g_0 X(D) + R(D). \tag{3.7}$$

where

$$R(D) = \sum_{k \neq 0} g_k D^k X(D) + \hat{N}(D). \tag{3.8}$$

Using the discrete time index, we have $s_k = g_0 x_k + r_k$. Thus, s_k contains the scaled current value of x_k and the remaining term r_k, which consists of past and future scaled values of x_k as well as the noise term $\hat{n}_k$. If x_k is an i.i.d. sequence with uniform distribution on $(-t/2, t/2]$, then r_k is independent of $g_0 x_k$. Typically $|g_0| \leq 1$. Thus a lower bound on $h(z) = h(\Gamma_t[g_0 x + r])$ can be obtained by conditioning r

$$h(z) \geq h(z|r) \tag{3.9}$$

$$= log_2(|g_0|t). \tag{3.10}$$

Removing conditioning always increases entropy, so

$$h(z|w) = h(\Gamma_t[w + a + y + \hat{n}]|w) \tag{3.11}$$

$$= h(\Gamma_t[y + \hat{n}]|w) \tag{3.12}$$

$$\leq h(\Gamma_t[y + \hat{n}]). \tag{3.13}$$

An upper bound on $h(\Gamma_t[y + \hat{n}])$ is the maximum differential entropy for a distribution with region of support $(-t/2, t/2]$ and variance equal to $\text{VAR}(y + \hat{n})$. A truncated Gaussian distribution [18, 17] achieves this maximum differential entropy. Let $\mathcal{T}(\sigma^2, t)$ be a zero mean Gaussian truncated to $(-t/2, t/2]$ with variance (after truncation) of σ^2. Then,

$$h(\Gamma_t[y + \hat{n}]) \leq h(\mathcal{T}(\sigma^2, t)). \tag{3.14}$$

Combining (3.10) and (3.14) produces the desired lower bound on C_M,

$$C_M \geq \log_2(|g_0|t) - h(\mathcal{T}(\sigma^2, t)). \tag{3.15}$$

[17] shows how to calculate $h(\mathcal{T}(\sigma^2, t))$.

The ISI y and the noise $\hat{n}$ are independent. Thus $\text{VAR}(y+\hat{n})$ equals $\text{VAR}(y)+\text{VAR}(\hat{n})$. From Figure 3.6, y is i.i.d.(white sequence) x filtered by $L(D)$ (as defined in section 3.1.1) and $\hat{n}$ is white noise n filtered by $\hat{F}(D)$. Both of the inputs are white so the calculation of the variance of $\hat{n}$ is similar to that of y. Therefore, only the derivation of $\text{VAR}(y)$ is considered here.

The input to the filter $L(D)$ is a white sequence with uniform distribution on $(-t/2, t/2]$ with variance $\frac{t^2}{12}$. Then the variance of y can be obtained by:

$$\begin{aligned} \text{VAR}(y) &= \frac{t^2}{12}\frac{1}{2\pi}\int_{-\pi}^{\pi} |L(e^{-jw})|^2 dw && (3.16) \\ &= \frac{t^2}{12} L_0 && (3.17) \end{aligned}$$

where the term L_0 is the zero lag tap of $L(D)L^*(D^{-*})$. If $L(D)$ is IIR, the zero lag tap L_0 can be found by partial fraction expansion.

In formula (3.15), note that g_0, $\hat{n}$ and σ^2 depend on $\hat{F}(D)$ and $\hat{B}(D)$ and thus will be different for different mismatch settings. Also g_0, $\hat{n}$ and σ^2 are not the same for ZF-THP with mismatch as for MMSE-THP with mismatch.

3.2.2 Perfect Channel Knowledge at Receiver

When the receiver has perfect knowledge of the channel and knows that the transmitter has some error in its channel estimate, $\hat{H}(D)$, the receiver can design the feedforward filter $\hat{F}(D)$ to meet ZF or MMSE criteria based on the *mismatched* feedback filter $\hat{B}(D)$ and the *actual* channel impulse response $H(D)$.

Selection of $\hat{F}(D)$ for ZF-THP with Mismatch

In Figure 3.6, the $\hat{F}(D)$ should be selected to make the ISI term, y, zero. Since the receiver knows both the *mismatched* feedback filter $\hat{B}(D)$ and the *actual* channel impulse response $H(D)$, The receiver can let $\hat{F}(D) = \frac{\hat{B}(D)}{H(D)}$ for zero-forcing. Thus, the THP system has no ISI.

Selection of $\hat{F}(D)$ for MMSE-THP with Mismatch

For minimizing mean squared error $E[(y+\hat{n})^2]$, the selection of $\hat{F}(D)$ is equivalent to the selection of $\hat{F}(D)$ in the MMSE-DFE system given $H(D)$ and the feedback filter $\hat{B}(D)$. Following [5] and [9], we can get $\hat{F}(D)$ based on $\hat{B}(D)$ and $H(D)$,

$$\hat{F}(D) = \frac{\hat{B}(D)H^*(D^{-*})}{\parallel \rho \parallel^2 (Q(D) + \frac{1}{SNR_{\text{MFB}}})}. \tag{3.18}$$

where

$$Q(D) = H(D)H^*(D^{-*})/ \parallel \rho \parallel^2 \tag{3.19}$$

$$\parallel \rho \parallel^2 = \sum_{k=-\infty}^{\infty} h_k^2 \tag{3.20}$$

$$H(D) = \sum_{k=-\infty}^{\infty} h_k D^k \tag{3.21}$$

$$SNR_{\text{MFB}} = \frac{t^2 \parallel \rho \parallel^2}{12\sigma_n^2} \tag{3.22}$$

The SNR_{MFB} is the matched-filter bound [9]. The average input energy equals $\frac{t^2}{12}$, and σ_n^2 is the variance of AWGN noise. The feedforward filter $\hat{F}(D)$ in (3.18) should be scaled to make $H(D)\hat{F}(D)$ monic as required in an unbiased MMSE-DFE system. When $\hat{H}(D)$ equals $H(D)$, $\hat{B}(D)$ and $\hat{F}(D)$ are the same as $B(D)$ and $F(D)$ respectively in MMSE-THP without mismatch.

The formula derived in (3.15) for THP with channel mismatch at both transmitter and receiver still applies to THP with perfect channel knowledge at the receiver where $g_0 = 1$ since the THP with perfect channel knowledge at the receiver is unbiased.

3.2.3 Tightness of the lower bound

In this section, the tightness of the lower bound in (3.15) is discussed. In the limit of high SNR, MMSE-THP with mismatch is identical to ZF-THP with mismatch. Asymptotically, the achievable rates are identical for both MMSE-THP and ZF-THP with mismatch. Since the tightness of the lower bound is considered in the range of high SNR, the ZF-THP is used for simplicity.

Two-Tap Channel

Let $H(D) = 1 + aD$ and $\hat{H}(D) = 1 + \hat{a}D$. $\hat{H}(D)$ is assumed to be a minimum phase channel and $d = |a - \hat{a}|$. For simplicity, d is assumed to be less than 1. For this mismatched ZF-THP system, the feedforward filter $\hat{F}(D) = 1$ and the feedback filter $\hat{B}(D) = \hat{H}(D)$. From Figure 3.6, the ISI term y_k has the D-transform $Y(D) = (a - \hat{a})D \cdot X(D)$. x_k is an i.i.d. sequence with uniform distribution in $(-t/2, t/2]$ if w_k is an i.i.d. sequence uniformly distributed in $(-t/2, t/2]$. Thus, y_k is an i.i.d. sequence uniformly distributed in $(-td/2, td/2]$. With $d < 1$, $\Gamma_t(y) = y$ and

$$
\begin{aligned}
h(\Gamma_t[y]) &= h(y) && (3.23)\\
&= \log_2(dt). && (3.24)
\end{aligned}
$$

$$h(\Gamma_t[y+\hat{n}]|w) = h(\Gamma_t[y+\hat{n}]). \tag{3.25}$$

w is independent of y and $\hat{n}$ in the mismatched ZF-THP for two-tap channel since $y_i = x_{i-1}$, x_{i-1} is independent of w_i and n_i is independent of w_i.

In the limit of high SNR, noise term $\hat{n}$ can be ignored. Thus, (3.25) converges to

$$h(\Gamma_t[y]) = \log_2(dt). \tag{3.26}$$

as derived in (3.24).

Combining (3.4), (3.10), (3.25) and (3.24), the tighter lower bound should converge to

$$\log_2(t) - \log_2(dt) = \log_2 \frac{1}{d}. \tag{3.27}$$

$g_0 = 1$ in this two-tap channel case. (3.27) shows the limit of the lower bound that considers the ISI term y as uniformly distributed in $(-dt/2, dt/2]$.

In the limit of high SNR, the entropy of the truncated Gaussian is identical to the entropy of Gaussian with the variance $\Theta^2 = \frac{(dt)^2}{12}$. Thus, the lower bound derived in (3.15) converges to

$$\log_2(t) - \frac{1}{2}\log_2(2\pi e\Theta^2) = \log_2 \frac{1}{d} - \frac{1}{2}\log_2 \frac{\pi e}{6}. \tag{3.28}$$

Subtracting (3.28) from (3.27) yields

$$\frac{1}{2}\log_2 \frac{\pi e}{6} = 0.255\ bits \tag{3.29}$$

This shows that the lower bound derived in (3.15) is actually looser by 0.255 bits due to the maximum truncated Gaussian entropy.

General Channel

Without the restriction of two-tap channel, the distribution of the ISI term y is not uniformly distributed any more and (3.24) does not hold. Generally, the number of taps in $L(D)$ increases as the number of taps in $H(D)$ increases. Theoretically, the distribution of the ISI term y can be determined. x_i is i.i.d. uniformly distributed in $(-t/2, t/2]$. The distribution of y is closer to Gaussian distribution as the number of taps in $L(D)$ increases. Thus, the lower bound derived in (3.15) is tighter as $H(D)$ has more taps.

Figure 3.7 shows that the achievable rate in mismatched THP behaves similarly to the output SNR in mismatched DFE for mismatch at both transmitter and receiver. This is the direct result of residual ISI in the mismatched DFE and THP.

3.3 Achievable rates of THP with PAM constellation

In this section, the mutual information with different PAM constellation size is demonstrated. All the PAM constellation points are constrained in the interval $(-t/2, t/2]$, and symmetric about the origin for even constellation size. The output z of the modulo at the receiver side is demodulated by hard decoding. Let $\hat{w}$ represent the output of the demodulator.

The analytic expression $I(w; \hat{w})$ in THP with mismatch in the limit of high SNR is obtained and simulation results are conducted from low to high range of SNR.

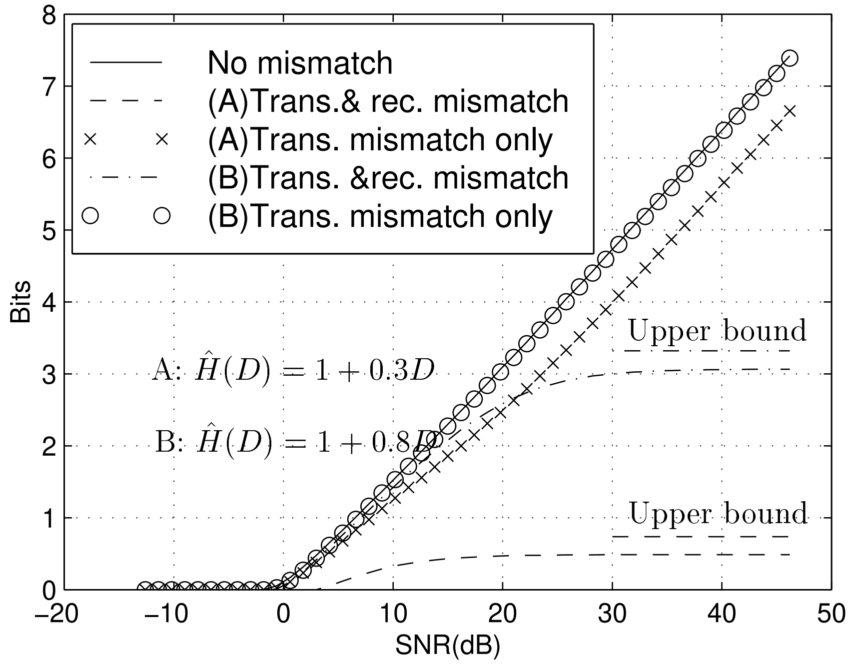

Figure 3.7: Achievable rate vs. SNR for MMSE-THP with impulse response $H(D) = 1 + 0.9D$. All curves are lower bounds except the two upper bounds identified in the plot

3.3.1 THP without mismatch

For THP without mismatch, the lower bound derived in (3.15) is tight in the limit of high SNR. This lower bound increases without bound as SNR increases. However, the mutual information with PAM M converges to $\log_2 M$ in the limit of high SNR.

The results show that the information rate shown by lower bound in (3.15) is greater than the mutual information with PAM4, PAM8, PAM16 and PAM32 as the input. The tight lower bound is achieved by assigning w i.i.d. uniformly distributed in $(-t/2, t/2]$, which is the extreme case of PAM constellation size.

3.3.2 THP with mismatch

Two-Tap Channel

Let $H(D) = 1 + aD$ and $\hat{H}(D) = 1 + \hat{a}D$. $\hat{H}(D)$ is assumed to be a minimum phase channel and $d = |a - \hat{a}|$. For simplicity, d is assumed to be less than 1. For this mismatched ZF-THP system, the feedforward filter $\hat{F}(D) = 1$ and the feedback filter $\hat{B}(D) = \hat{H}(D)$ in the limit of high SNR.

For i.i.d. PAM w_k with constellation size M, w_k is the data symbol as a M-valued random variable in the interval $(-t/2, t/2]$, symmetric about the origin for even constellation size. The M values can be represented by index $i = 0, 1, 2, \cdots, M-1$ from left to right in the interval. w_k takes the M values independently and with equal probability.

From Figure 3.5, $x_k = \Gamma_t[w_k - \hat{a}x_{k-1}]$. For a fairly large index k, x_k is assumed to take countable infinity rational values in $(-t/2, t/2]$ with equal probability. Thus, ISI y is assumed to take countable infinity rational values in $(-dt/2, dt/2]$

with equal probability. The assumptions are verified by the simulation results.

In the limit of high SNR, noise $\hat{n}$ is negligible. Therefore, the error occurs only due to ISI y.

Let's consider the mutual information $I(w;\hat{w}) = H(\hat{w}) - H(\hat{w}|w)$. $H(\hat{w}) = \log_2 M$ by symmetry in the input w. To determine $H(\hat{w}|w)$, consider the conditional probability $P(\hat{w} = j|w = i), i, j = 0, 1, \cdots, M-1$ first.

The demodulator decides $\hat{w} = j$ if the output of the modulo at the receiver side falls into the decision region (with length $\frac{t}{M}$) of PAM constellation size with index j.

If $dt \leq \frac{t}{M}$, that is $d \leq \frac{1}{M}$. Then, the decision made at the demodulator is correct. This means $P(\hat{w} = j|w = i) = \delta_{ij}$. Thus,

$$I(w;\hat{w}) = \log_2 M \tag{3.30}$$

$$\leq -\log_2 d \tag{3.31}$$

If $dt \geq \frac{t}{M}$, that is $d \geq \frac{1}{M}$. Then, the decision made at the demodulator has error. This means $P(\hat{w} = j|w = i) \neq 0$, for $i = 0, 1, \cdots, M-1$ and $j = i, i \pm 1, \cdots, i \pm \frac{s}{2} - 1, i \pm \frac{s}{2}$. Where s is the number of symbols the demodulator decides incorrectly when the real input is $w = i$. If j does not belong to $0, 1, \cdots, M-1$, then j should be added by M if $j < 0$ and be subtracted by M if $j > M-1$. For simplicity, the addition and subtraction by M is not be considered in the following analysis.

The number of symbols the demodulator decides incorrectly, s, depends on the parameters d(the scaling factor of the ISI y) and M (the PAM constellation size).

$$s = 2 \cdot \lceil \frac{d - \frac{1}{M}}{2} \cdot \frac{1}{\frac{1}{M}} \rceil \tag{3.32}$$

$$= 2 \cdot \lceil \frac{dM - 1}{2} \rceil. \tag{3.33}$$

The transition probability can be expressed as follows:

$$P(\hat{w} = i | w = i) = \frac{1}{dM} \tag{3.34}$$

$$P(\hat{w} = j | w = i) = \frac{1}{dM}, j = i, i \pm 1, \cdots, i \pm \frac{s}{2} - 1 \tag{3.35}$$

$$P(\hat{w} = i \pm \frac{s}{2} | w = i) = \frac{1}{2}(1 - \frac{s-1}{dM}). \tag{3.36}$$

The mutual information $I(w; \hat{w})$ can be obtained through (3.36),

$$I(w; \hat{w}) = (1 - \frac{s-1}{dM})(\log_2(dM - (s-1)) - 1) - \log_2 d \tag{3.37}$$

The following relationship holds,

$$I(w; \hat{w}) \leq -\log_2 d \tag{3.38}$$

This means that $I(w; \hat{w})$ has an upper bound $-\log_2 d$ for every finite PAM constellation size M. The proof is given as follows.

Proof: From inequality $M > \frac{1}{d}$, $dM - 1 > 0$ holds.

Let $0 \leq \theta \leq 1$. From (3.33) s can be expressed as

$$s = 2(\frac{dM - 1}{2} + \theta) \tag{3.39}$$

$$= dM - 1 + 2\theta. \tag{3.40}$$

So that $s + 1 = dM + 2\theta > dM$, which is equivalent to

$$dM - (s - 1) < 2. \tag{3.41}$$

From (3.40), $s-1 = dM+2\theta-2 = dM-2(1-\theta) < dM$, which is equivalent to

$$1 - \frac{s-1}{dM} > 0. \tag{3.42}$$

From (3.41), (3.42) and (3.37), $I(w;\hat{w}) \leq -\log_2 d$, which completes the proof.

For the two-tap channel, $-\log_2 d$ is the upper bound of the mutual information of the THP with mismatch when the input w is i.i.d. PAM and uniformly distributed in interval $(-t/2, t/2]$.

General Channel

Let $H(D) = 1 + \sum_{i=1}^{n} h_i D^i$ and $\hat{H}(D) = 1 + \sum_{i=1}^{n} \hat{h}_i D^i$. Furthermore, $\hat{H}(D)$ is assumed to be minimum phase. The feedforward filter $\hat{F}(D) = 1$ and the feedback filter $\hat{B}(D) = \hat{H}(D)$ in the limit of high SNR.

Define $\Delta H(D) = H(D) - \hat{H}(D) = \sum_{i=1}^{n} \Delta h_i D^i$. where $\Delta h_i = h_i - \hat{h}_i$.

Let

$$\Delta h_{min} = \min_{1\leq i\leq n} |\Delta h_i|. \tag{3.43}$$

Furthermore, $\Delta h_{min} < 1$ is assumed.

From Figure 3.6, the ISI y can be expressed as

$$y_k = \sum_{i=0}^{n} \Delta h_i \cdot x_{k-i}. \tag{3.44}$$

From (3.43), the conditional entropy of z on w can be lower bounded as

$$h(z|w) = h(\Gamma_t(y)|w) \tag{3.45}$$

$$\geq h(\Gamma_t(\Delta h_{min} \cdot x) \tag{3.46}$$

$$= \log_2(\Delta h_{min} \cdot t). \tag{3.47}$$

Thus $I(w; z)$ can be upper bounded as

$$I(w; z) \leq \log_2 t - \log_2(\Delta h_{min} \cdot t) \tag{3.48}$$

$$= -\log_2(\Delta h_{min}). \tag{3.49}$$

So $-\log_2(\Delta h_{min})$ is the upper bound of the mutual information of the THP with mismatch if the input w is i.i.d. uniformly distributed in interval $(-t/2, t/2]$.

If w is i.i.d. PAM M, it can be proved that

$$I(w; \hat{w}) \leq -\log_2(\Delta h_{min}). \tag{3.50}$$

Proof: Let $\hat{w}_1$ be the output of the two-tap channel as in subsection 3.3.2 with $d = \Delta h_{min}$.

Since the ISI y in the two-tap channel as in 3.3.2 with $d = \Delta h_{min}$ is less than the ISI y in the actual general channel, thus,

$$I(w; \hat{w}) \leq I(w; \hat{w}_1) \leq -\log_2(\Delta h_{min}). \tag{3.51}$$

which completes the proof.

For the general channel, $-\log_2(\Delta h_{min})$ is the upper bound of the mutual information of the THP with mismatch when the input w is i.i.d. PAM and uniformly distributed in interval $(-t/2, t/2]$.

3.4 Examples

Figures 3.8-3.11 show simple examples of how channel estimation mismatch reduces achievable rate. For mismatch at both transmitter and receiver, residual ISI limits maximum achievable rate to be bounded even as SNR increases without bound. This effect is analogous to the "irreducible error rate" effect seen in unequalized systems faced with intersymbol interference. With severe mismatch, the bounded maximum achievable rate is very low.

As in the case of THP without mismatch, the overall performance of MMSE-THP with mismatch is generally better than the performance of ZF-THP with the same mismatch. At lower and moderate SNR, the information rate in the MMSE-THP with mismatch is higher than in ZF-THP with the same mismatch. At high SNR both ZF-THP and MMSE-THP with mismatch converge to the same limit since the feedback filter $\hat{B}(D)$ and the feedforward filter $\hat{F}(D)$ are identical.

When only the transmitter suffers channel estimation error, a correctly designed feedforward filter allows the information rate to increase without bound as SNR increases. Figures 3.8-3.9 show that for small amount of mismatch at the transmitter, properly designed receiver filters can closely approach the rate achievable without mismatch. In the case of small mismatch, this conclusion by analyzing the channel capacity is similar to that by Pottie and Eyuboğlu [19] by analyzing the SNR. In their paper, their explanation is that small mismatch results no new deep nulls in the channel and hence a linear equalizer is almost as good as a DFE; the properly designed receiver feedforward filters does this function.

However, Figures 3.10-3.11 show that as transmitter mismatch becomes severe, non-negligible loss is incurred despite perfect channel knowledge at the receiver and therefore a gap between the information rates with perfect knowledge at the receiver side and without mismatch exists. This loss is analogous to the loss of a system with only a linear equalizer as compared to a DFE or THP structure, since the receiver can only mitigate mismatch with a feedforward filter.

At small SNR, ZF-THP with perfect knowledge at the receiver side may have lower information rate than ZF-THP with mismatch at both transmitter and receiver side as a result of noise enhancement introduced by the linear equalizer.

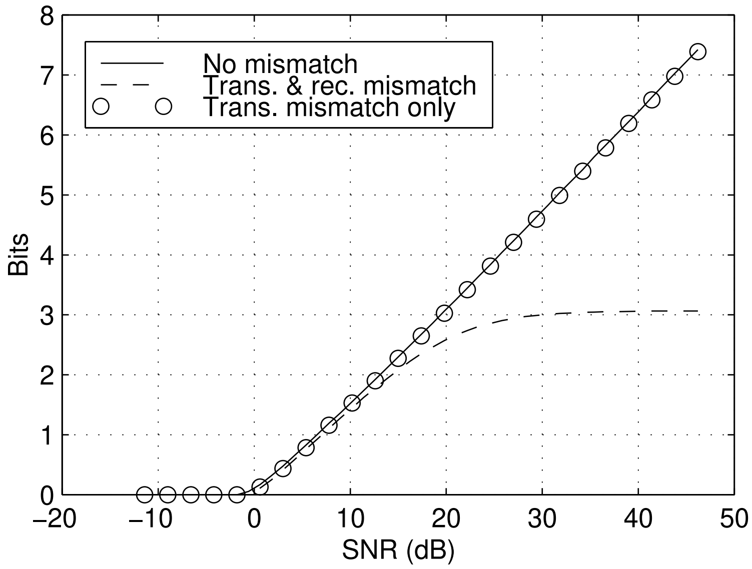

Figure 3.8: MMSE-THP with impulse response $H(D) = 1 + 0.9D$ and stale impulse response $\hat{H}(D) = 1 + 0.8D$

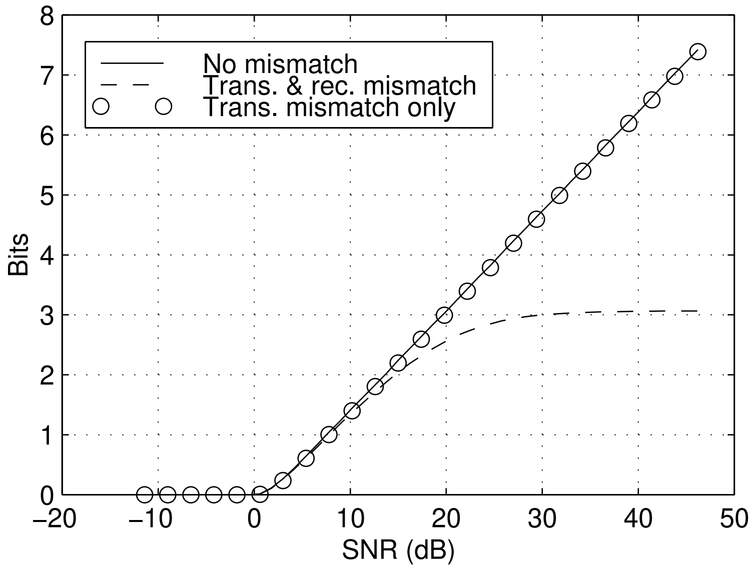

Figure 3.9: ZF-THP with impulse response $H(D) = 1 + 0.9D$ and stale impulse response $\hat{H}(D) = 1 + 0.8D$

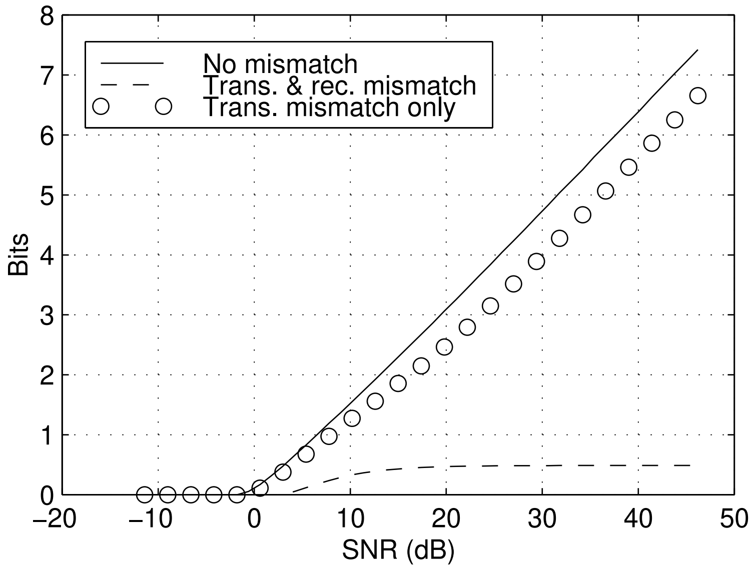

Figure 3.10: MMSE-THP with impulse response $H(D) = 1 + 0.9D$ and stale impulse response $\hat{H}(D) = 1 + 0.3D$

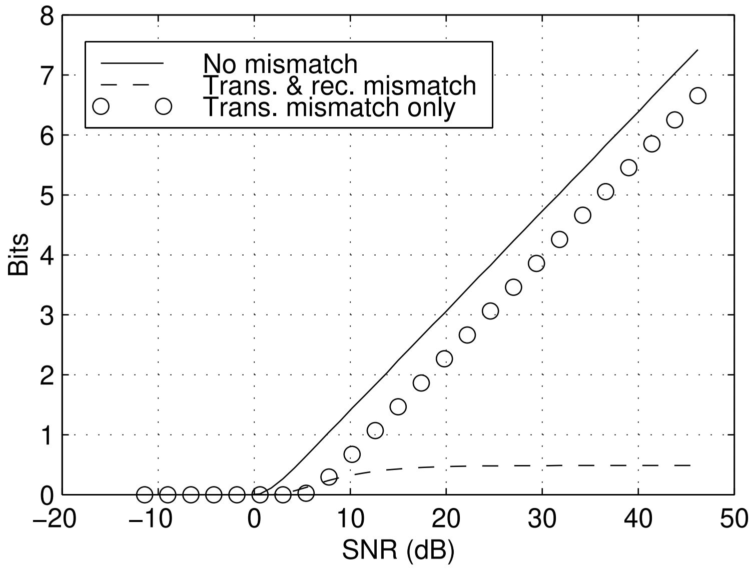

Figure 3.11: ZF-THP with impulse response $H(D) = 1 + 0.9D$ and stale impulse response $\hat{H}(D) = 1 + 0.3D$

Chapter 4

Robustness of Space-Time Turbo Codes for flat fading

Recently, the goal of providing high data rate wireless communications has motivated the use of multiple transmit and receive antennas. Foschini and Gans [20] showed the extraordinarily large capacity promised by using multiple transceiver antennas for the flat Rayleigh fading. This opened the new area of space-time codes, which take advantage of spatial diversity with multiple transceiver antennas and temporal diversity provided by channel coding and interleaving.

Tarokh *et al.* [21, 22] proposed the design criterion for space-time codes for flat Rayleigh fading in a variety of scenarios. To achieve a high level of diversity, turbo codes [23] can be used to get excellent performance in fading channels with multiple transceiver antennas. In [24], Liu *et al.* proposed a rank criterion of full space diversity QPSK space-time turbo codes. In [25, 26], Stefanov and Duman showed that a simple QPSK turbo code outperforms the space-time block and trellis codes [21, 22] significantly.

Foschini [27] proposed the diagonally-layered structure which can achieve a tight lower bound on the capacity. Tarokh *et al.* [28] extended this layered structure by a group interference suppression method and employed the combination of array processing and channel coding.

Unlike [28], this chapter uses the reliability of the signals from the other transmit antenna to perform soft cancellation within the forward-backward decoder.

This chapter shows the robustness of turbo codes in flat fading with 2 transmit antennas and 2 receive antennas. The robustness is measured in terms of consistent bit error rate (BER) for a specified proximity to channel capacity. For flat fading, we consider robustness for families of channels with the same set of singular values (SV's). We have found that the robustness is determined by the difference of the 2 SV's. When the SV's are similar, a straightforward turbo decoding approach provides robust performance. When the SV's differ significantly, robustness may still be obtained, but requires a time-varying linear transformation (TVLT) in both the transmitter and the receiver.

However TVLT technique is powerful in broadcasting communications, when the transmitter has no knowledge about the channels but the receiver has the knowledge of channel. For all the channels with the same set of singular values, the gain by using TVLT for the worst channels is much larger than the slight loss by using TVLT for the best channels. So the overall performance for all the channels with or without the same set of singular values is significantly improved.

4.1 Channel and system models

4.1.1 Space-time channels and theoretical limits

As in [20], we assume that the fading is fixed for a transmission block, but varies between blocks. Let the 2×2 complex matrix H represent the channel matrix for a given transmission block, where h_{ij} is the path gain from transmit antenna j to receive antenna i, $i, j \in \{1, 2\}$. The input-output equation for the multiple-input multiple-output (MIMO) channel is

$$\mathbf{y} = H \cdot \mathbf{x} + \mathbf{n},$$

or explicitly

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \tag{4.1}$$

where x_j is the transmitted signal at the transmit antenna j, which is a two-dimensional (2-D) constellation point with energy $\bar{P}_x$ per dimension,$\bar{P}_x = P_x/2$. y_i is the received signal at receive antenna i. Noise n_i is i.i.d. additive white Gaussian noise (AWGN) with variance $N_0/2$ per dimension.

When the transmitter knows nothing about the channel and hence transmits x_1 and x_2 with the same energy, the maximum mutual information (MI) is

$$I(\mathbf{x}; \mathbf{y}) = \log_2 \left(1 + \frac{\lambda_1^2 P_x}{N_0}\right) + \log_2 \left(1 + \frac{\lambda_2^2 P_x}{N_0}\right), \tag{4.2}$$

where λ_1 and λ_2 are singular values of channel matrix H. As defined in [20], the signal-to-noise ratio (SNR) is the average SNR at each receiver branch. In terms of the singular values, $SNR = \frac{(\lambda_1^2 + \lambda_2^2) P_x}{2N_0}$.

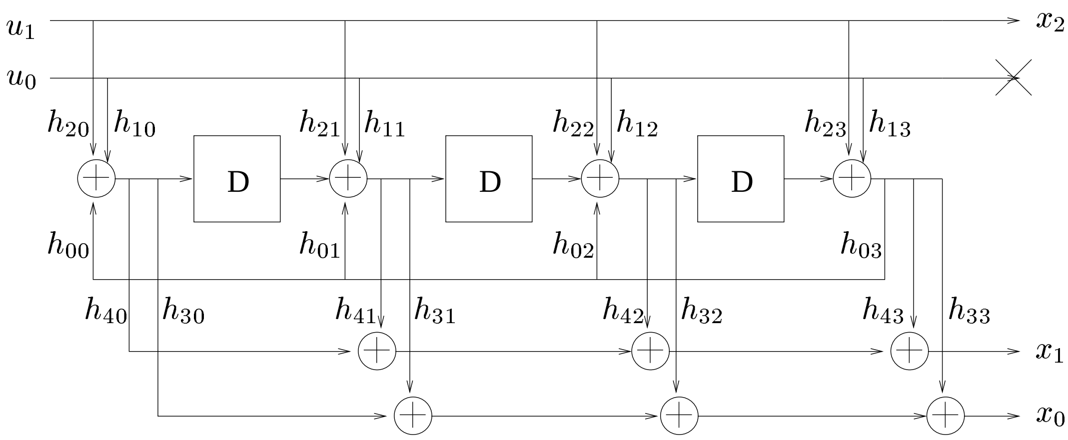

Figure 4.1: Encoder structure with 3 memory elements

Suppose the transmitter performs a linear transformation on the transmitted signal x_1 and x_2 by an 2×2 unitary matrix Q_t and the receiver performs linear transformation on the received signal y_1 and y_2 by an 2×2 unitary matrix Q_r. Then we have the following

$$Q_r \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = Q_r \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} Q_t \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + Q_r \begin{bmatrix} n_1 \\ n_2 \end{bmatrix},$$

where the effective channel $\tilde{H} = Q_r H Q_t$ has the same singular values as H, and the same MI by (4.2). Let the channel matrix H have the singular value decomposition (SVD) as $H = USV^*$ with the singular values λ_1 and λ_2, then $\tilde{H}$ has an SVD as $\tilde{H} = \tilde{U}S\tilde{V}^*$ with the same singular values λ_1 and λ_2, since both $\tilde{U} = Q_r U$ and $\tilde{V} = Q_r^* V$ are unitary matrices. The noise $\begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix} = Q_r \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$ is still i.i.d. AWGN with variance $N_0/2$ per dimension. So if the original channel is in the family of channels with singular values λ_1 and λ_2, the channel $\tilde{H} = Q_r H Q_t$ is still in the family.

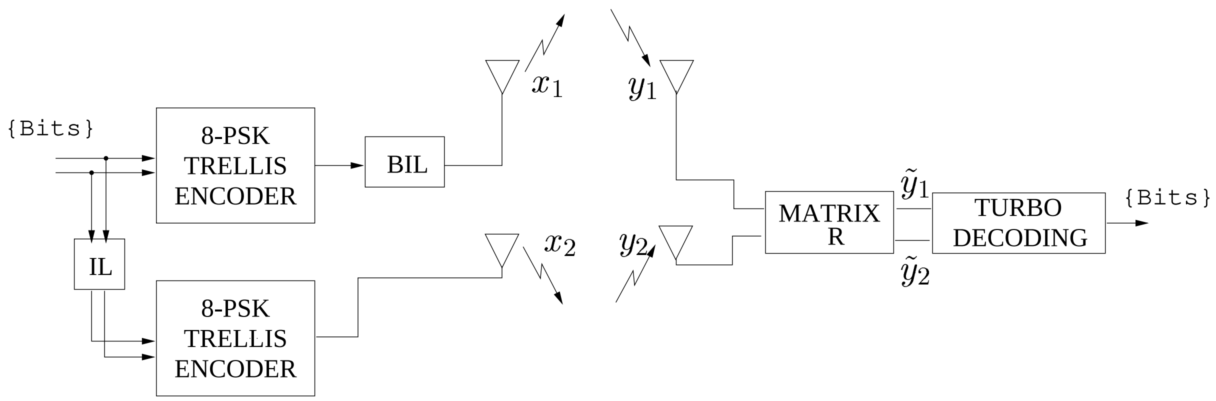

Figure 4.2: Space-time model with turbo-coding.

4.1.2 Space-time turbo codes

The constituent encoders we use in the symbol-interleaved turbo-trellis code of Figure 4.2 are depicted in Figure 4.1. They are characterized by the octal polynomial $(h_0, h_1, h_2, h_3, h_4) = (13, 1, 17, 13, 17)$, the meaning of which is described in Figure 4.1. These constituent encoders are optimal by the criteria proposed in [29, 30]. For the upper encoder the input bit u_0 is punctured. For the lower encoder the input bit u_1 is punctured.

The interleaver of Figure 4.2 is a 4098-symbol randomly generated interleaver with 2 input bits per symbol. The input bits are flipped in each interleaved symbol, as shown in Figure 4.2, which avoids repetition of the systematic bit. A symbolwise block-interleaver is used at the output of the upper encoder. x_1 is the transmitted signal in the upper transmit antenna which is the block-interleaved version of the output of the upper encoder. x_2 is the transmitted signal in the lower transmit antenna, which is the output of the lower encoder.

At the receiver, y_1 and y_2 are the linear combinations of x_1 and x_2 distorted by

AWGN noise received by the two receive antennas. After a linear transformation R as shown in Figure 4.2 at the receiver, which will be discussed in the next section, the received signals y_1 and y_2 become $\tilde{y}_1$ and $\tilde{y}_2$. So the inputs to the turbo decoder are $\tilde{y}_1$ and $\tilde{y}_2$.

Figure 4.3 shows the decoding diagram. The only channel input to the upper decoder is the block-deinterleaved version of $\tilde{y}_1$. Similarly $\tilde{y}_2$ is the only channel input to the lower decoder. The channel input $\tilde{y}_i$ not considered by one decoder is accounted for through the extrinsic information on x_i provided by the other decoder.

Both upper and lower decoders use the forward-backward algorithm to get the reliabilities of input and the outputs of corresponding encoders. They exchange the reliabilities for the input in the same way as in standard parallel concatenated convolutional codes. Each decoder also sends the reliabilities about the outputs of the corresponding encoders to the other decoder. The upper and lower decoders are connected by interleavers, which is standard in soft-input soft-output (SISO) networks [31].

The crucial part of the forward-backward algorithm for the lower forward-backward iterative decoding is

$$
\begin{aligned}
&\gamma_k(m', m) = Pr(\tilde{y}_{2k}, S_k = m | S_{k-1} = m') \\
&= Pr(S_k = m | S_{k-1} = m') \cdot Pr(\tilde{y}_{2k} | S_{k-1} = m', S_k = m) \\
&= Pr(u_k; I) \cdot Pr(\tilde{n}_{2k} = \tilde{y}_{2k} - \tilde{h}_{21}\, x_{1k} - \tilde{h}_{22}\, x_{2k} \,| x_{2k}) \\
&= Pr(u_k; I) \cdot \sum_{x_{1k}} Pr(\tilde{n}_{2k} = \tilde{y}_{2k} - \tilde{h}_{21}\, x_{1k} - \tilde{h}_{22}\, x_{2k} \,|\, x_{1k},\, x_{2k}) \cdot Pr(x_{1k}; I).
\end{aligned}
$$

We can get a similar expression for the upper decoder. With the extrinsic information about x_1 or x_2, the upper and lower decoders can update its $\gamma_k(m', m)$ and calculate the reliabilities about the input and the output of the corresponding encoder.

Assume that the upper decoder is doing the decoding first in each iteration. In the first iteration, the extrinsic information $P(u; I)$ about the input will be zero (all u's are equally likely), when the logarithm is taken. The soft information $P(x_2; I)$ is needed by the upper encoder. It is possible to overcome the need for soft information $P(x_2; I)$ by the upper decoder by using an upper/lower-triangular decomposition (ULTD).

Let $U = \begin{bmatrix} 1 & u_{12} \\ 0 & 1 \end{bmatrix}$ be the upper triangular matrix and $L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$ be the lower triangular matrix such that $\tilde{H} = UL$. U and L can be uniquely determined for a given $\tilde{H}$ and it can be shown that $l_{21} = \tilde{h}_{21}, l_{22} = \tilde{h}_{22}$. The inverse of matrix U is simply $U^{-1} = \begin{bmatrix} 1 & -u_{12} \\ 0 & 1 \end{bmatrix}$. Doing linear transformation on $\tilde{y}_1$ and $\tilde{y}_2$ by matrix U^{-1}, we get $\hat{y}_1$ and $\hat{y}_2$ as

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 \\ \tilde{h}_{21} & \tilde{h}_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \hat{n}_1 \\ \hat{n}_2 \end{bmatrix} \tag{4.3}$$

where $\begin{bmatrix} \hat{n}_1 \\ \hat{n}_2 \end{bmatrix} = U^{-1} \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix}$. Clearly $\hat{y}_1 = l_{11}x_1 + \hat{n}_1$, so there is no dependency in x_2. $\hat{y}_2 = \tilde{y}_2$ by (4.3), so the channel input to the lower decoder is $\tilde{y}_2$ for all the iterations even with the ULTD performed in the first iteration.

4.2 Signal processing techniques

4.2.1 Linear transformation at the receiver side

From an MI perspective (4.2), all channels with the same SV's are equivalent. However diagonal channels perform best in our turbo decoding simulations because there is no interference from the other transmit antenna. Normally, the channel matrix is not diagonal in the space-time scenario; there is interference between the transmitted signals.

When the receiver has the knowledge of channel H, the receiver can perform linear operations on y_1 and y_2 to separate the transmitted signals. In this way, the receiver tries to create an effective channel $\tilde{H}$ that has the same mutual information but less interference. The linear transformation can be expressed as

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \tag{4.4}$$

Assuming $H = USV^*$ and $R = VU^*$, the effective channel matrix $\tilde{H} = RH = VSV^*$ is symmetric and positive-definite. If $\lambda_1 = \lambda_2$, $\tilde{H}$ is diagonal. This technique improves performance by mitigating variation in U in all cases and V when $\lambda_1 = \lambda_2$. For example,

Let

$$V_1 = \begin{bmatrix} \cos\theta e^{-j\theta_1} & \sin\theta \\ \sin\theta & -\cos\theta e^{j\theta_1} \end{bmatrix}, \tag{4.5}$$

then

$$V_1SV_1^* = \begin{bmatrix} \lambda_1\cos^2\theta + \lambda_2\sin^2\theta & (\lambda_1-\lambda_2)\cos\theta\sin\theta e^{-j\theta_1} \\ (\lambda_1-\lambda_2)\cos\theta\sin\theta e^{j\theta_1} & \lambda_1\sin^2\theta + \lambda_2\cos^2\theta \end{bmatrix}.$$

Another unitary matrix V_2 can be represented as

$$V_2 = \begin{bmatrix} \cos\theta e^{-j\theta_1} & -\sin\theta \\ \sin\theta & \cos\theta e^{j\theta_1} \end{bmatrix}, \tag{4.6}$$

and $V_2SV_2^* = V_1SV_1^*$.

We can see that the larger the difference $|\lambda_1 - \lambda_2|$, the larger the off-diagonal elements in $\tilde{H} = V_1SV_1^*$ will be. Likewise the channel matrix $\tilde{H} = V_1SV_1^*$ has small off-diagonal elements if λ_1 is very close to λ_2. $\tilde{H} = V_1SV_1^* = \lambda_1 I$ if $\lambda_1 = \lambda_2$, where matrix I is identity. So all the channels with the identical singular values have the exactly same effective channel $\tilde{H}$ and exactly the same performance.

4.2.2 Time-varying linear transformations

When the λ_1 and λ_2 are far apart, the channels with interference perform noticeably worse in our simulations than the diagonal channels. This can be observed from expression (4.5). In this case, the values in the off-diagonal elements depend on the specific value θ or the unitary matrix V. For different θ, the performance of channels can be noticeably different.

To achieve robustness (i.e. similar performance for a family of channels with the same λ_1 and λ_2), a time-varying linear transformation at both the transmitter and the receiver can be introduced to make the performance on the family of channels consistent. This technique in a sense transforms the unitary matrix V in

(4.5) or (4.6) to a time-varying unitary matrix $T_k V$, where T_k is a unitary matrix at the transmitter and is periodically varying for every channel transmission. Let

$$T_k = \begin{bmatrix} \cos\psi_i e^{-j\gamma_j} & \sin\psi_i \\ \sin\psi_i & -\cos\psi_i e^{j\gamma_j} \end{bmatrix}, \tag{4.7}$$

ψ_i can be some value from 0 to π, i.e. $\psi_i = \frac{\pi}{p_1} i, i = 1 \cdots, p_1$ and $\gamma_j = \frac{\pi}{p_2} j, j = 0 \cdots, p_2 - 1$, where p_1 and p_2 can be any positive integer numbers. The index k is related with i and j as $k = (j-1)p_1 + i$.

To the extent that T_k cycles through a set of matrices that finely samples the space of possible V's, performance will be consistent throughout the class. We have found that the worst V's benefit substantially from this time variation. The best V's suffer a performance loss, but not as much as the gain for the worst V's.

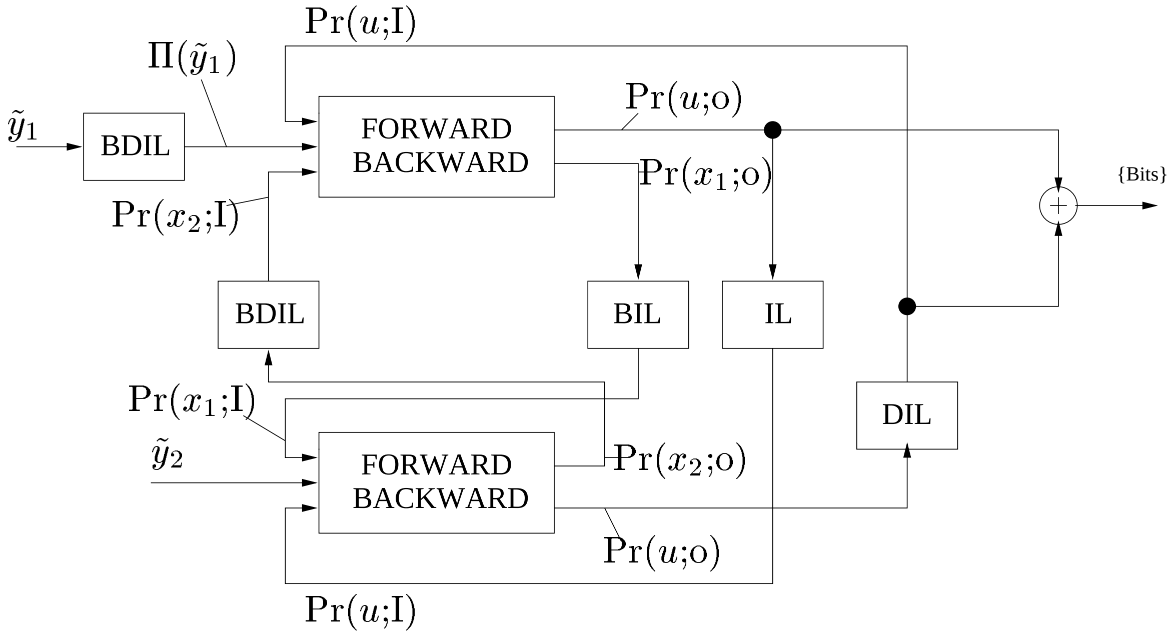

Figure 4.3: Space-time iterative decoding.

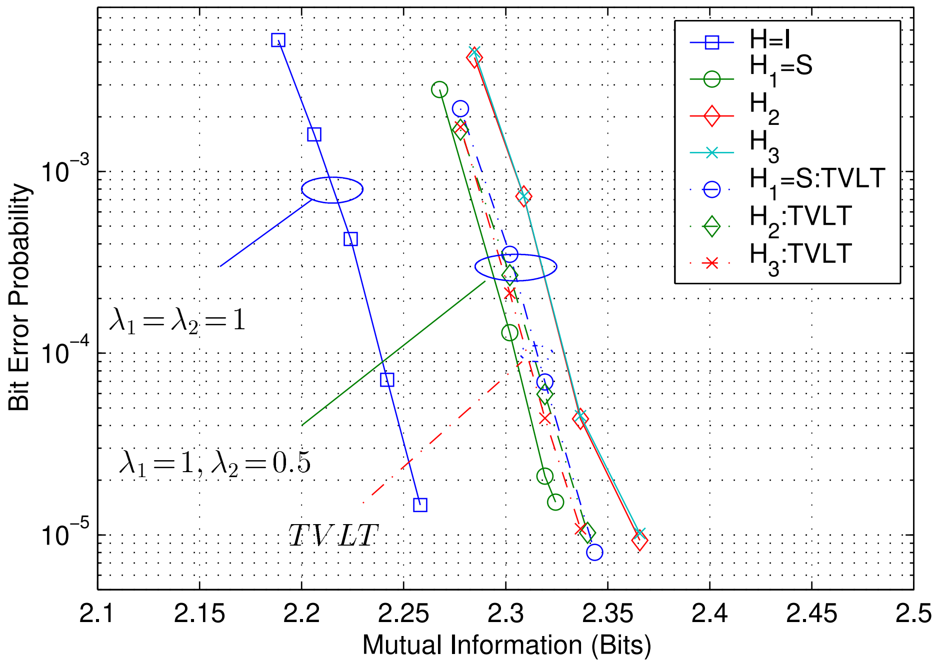

Figure 4.4: Bit error rate vs. mutual information for two families of channels with different sets of singular values. The notation $H_i : TVLT$ represents the time-varying linear transformation for channel H_i used at both the transmitter and receiver.

4.3 Simulation

4.3.1 Proximity to Capacity

To have an idea about how closely the space-time turbo codes perform to the MI expressed in (4.2), we consider the plot of BER versus MI in plots of simulation results using Monte Carlo simulation with 15 iterations. All the simulations are terminated when 80 block errors occur. The standard plots of BER versus SNR make a fair performance comparison between different channels difficult so we use MI (4.2) instead of SNR, as in [32].

Figure 4.4 shows the BER versus MI for two families of channels. The $\tilde{H} = I$ represents all channels with singular values $\lambda_1 = \lambda_2 = 1$ and the other 3 channels all with $\lambda_1 = 1, \lambda_2 = 0.5$. With the assumption that the receiver has perfect knowledge of the channel, all different channels with $\lambda_1 = \lambda_2 = 1$ have the same effective channel $\tilde{H} = I$, which is the identity matrix. At BER=10^{-5}, this class of channels needs 2.26 bits per 2 antennas. This family of channels requires an excess MI of 0.13 bits per antenna. At BER=10^{-5}, all 3 channels with $\lambda_1 = 1, \lambda_2 = 0.5$ need 2.33-2.35 bits per 2 antennas; they require an excess MI of 0.165-0.175 bits per antenna.

By using TVLT at both the transmitter and the receiver, the channels with singular values $\lambda_1 = \lambda_2 = 1$ still have the effective channel $\tilde{H} = I$ and perform the same as without using TVLT at both the transmitter and the receiver. However, for the channels with singular values with $\lambda_1 = 1, \lambda_2 = 0.5$, the TVLT technique makes the worst channels H_2 and H_3 perform much better and the best channel $H_1 = S$ perform a little bit worse. By using TVLT, the 3 channels perform much closer and need approximately 2.335-2.341 bits per 2 antennas for BER=10^{-5};

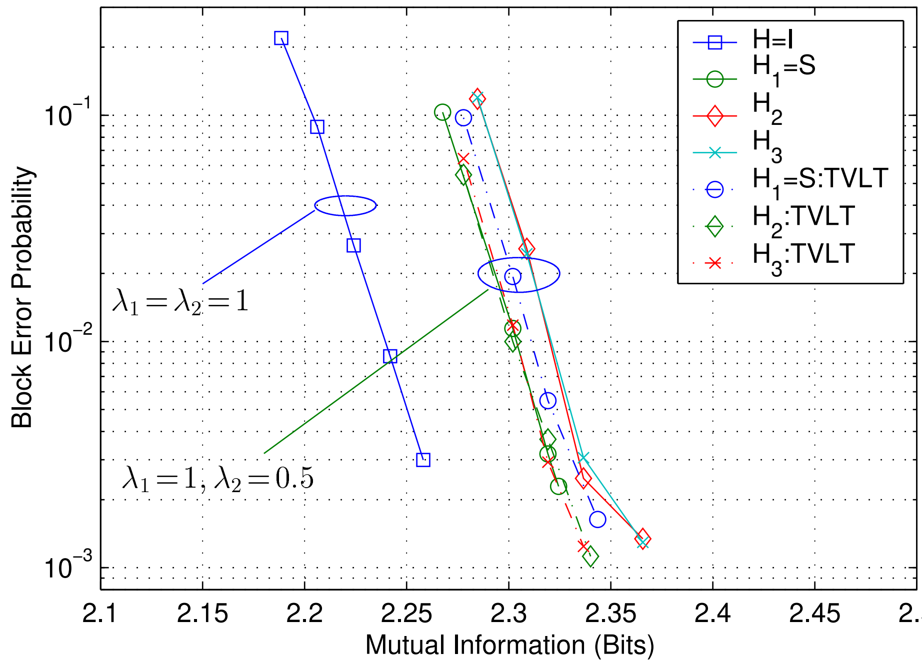

Figure 4.5: Block error rate vs. mutual information for two families of channels with different sets of singular values. The notation $H_i : TVLT$ represents the time-varying linear transformation for channel H_i used at both the transmitter and receiver.

they require an excess MI of 0.167-0.171 bits per antenna.

Furthermore, the equivalent AWGN SNR as described in [32] is used here to get a more familiar term. For $\tilde{H} = I$, the 1.13 bits per antenna corresponds to equivalent AWGN SNR of 0.81dB. For the other 3 channels, the 1.165-1.175 bits per antenna correspond to equivalent AWGN SNR of 0.94dB-1.00dB. Using TVLT, the 1.167-1.171 bits per antenna correspond to equivalent AWGN SNR of 0.96dB-0.98dB. For comparison, the turbo-codes for transmission of 1 bit per channel use in the AWGN channel require an equivalent SNR of 0.7dB-1.0dB (or 0.121-0.176 bits of excess MI) for BER=10^{-5}. There might still be some improvement on the performance of space-time turbo-codes.

4.3.2 Examples

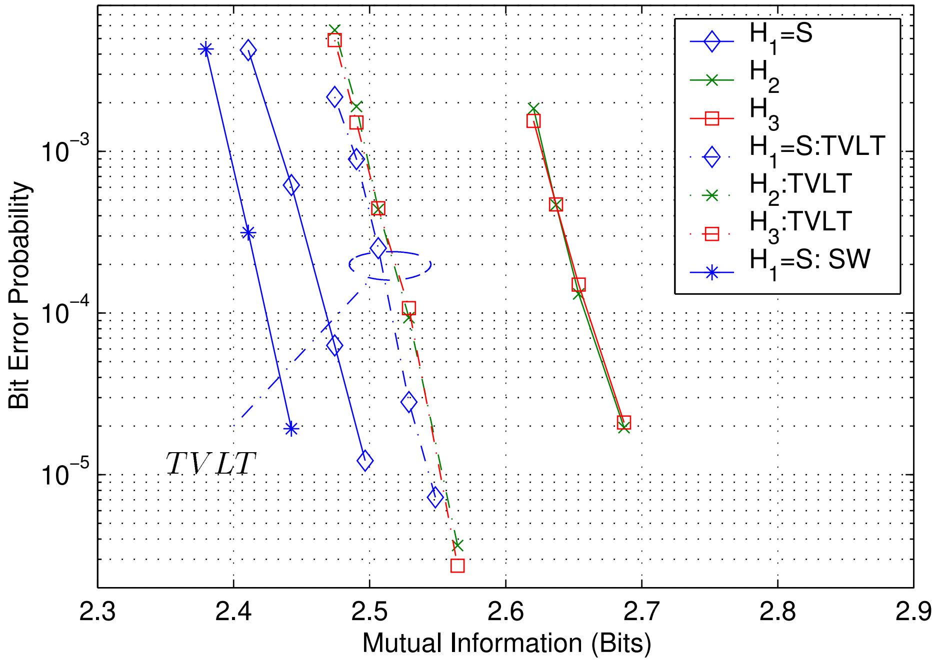

Figure 4.6: Bit error rate vs. mutual information for the family of channels with λ_1 =1, λ_2=0.25. The notation $H_i : TVLT$ represents the time-varying linear transformation for channel H_i used at both the transmitter and receiver. The notation $H_1 = S : SW$ represents the transmitted signals x_1 and x_2 are alternated for each antenna at the transmitter for the diagonal channel $H_1 = S$.

Here, we take a closer look at the 3 channels with singular values $\lambda_1 = 1, \lambda_2 = 0.5$ in Figure 4.4. Let $S = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$. Channel matrix $H_1 = S$ is a diagonal matrix. $H_2 = U_2 S V_2^*$, where $V_2 = \begin{bmatrix} 0.6594 & -0.7518 \\ 0.7518 & 0.6594 \end{bmatrix}$. U_2 can be any unitary matrix in real or complex domain. So the channel matrix H_2 represents a family of matrices which have the same effective channel matrix $\tilde{H}_2 = V_2 S V_2^*$ after the linear transformation at the receiver. Similarly $H_3 = U_3 S V_3^*$, where $V_3 = \begin{bmatrix} 0.7518 & -0.6594 \\ 0.6594 & 0.7518 \end{bmatrix}$. Similar to H_2, channel matrix H_3 represents a family of channel matrices which have the same effective channel matrix $\tilde{H}_3 = V_3 S V_3^*$.

As described in subsection 4.3.1, at BER=10^{-5}, the diagonal channel H_1 requires 2.33 bits of MI per 2 antennas using (4.2) while the other 2 channels H_2 and H_3 require approximately 2.35 bits per 2 antennas; the gap between bits the diagonal channel H_1 requires and the other 2 channels require is 0.02 bits per 2 antennas. The 3 channels require an excess MI of 0.165-0.175 bits per antenna. By using TVLT, the diagonal channel H_1 requires 2.335 bits per 2 antennas and the other 2 channels H_2 and H_3 require approximately 2.341 bits per 2 antennas; the gap between bits the diagonal channel H_1 requires and the other 2 channels require is 0.006 bits per 2 antennas. The 3 channels require an excess MI of 0.167-0.171 bits per antenna by using TVLT.

In Figure 4.6, we consider 3 channels, all with singular values $\lambda_1 = 1, \lambda_2 = 0.25$. Thus $S = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.25 \end{bmatrix}$. Channel matrix $H_1 = S$ is a diagonal matrix. $H_2 = U_2 S V_2^*$ and $H_3 = U_3 S V_3^*$, where V_2 and V_3 as well as U_2 and U_3 are the same as described in Figure 4.4.

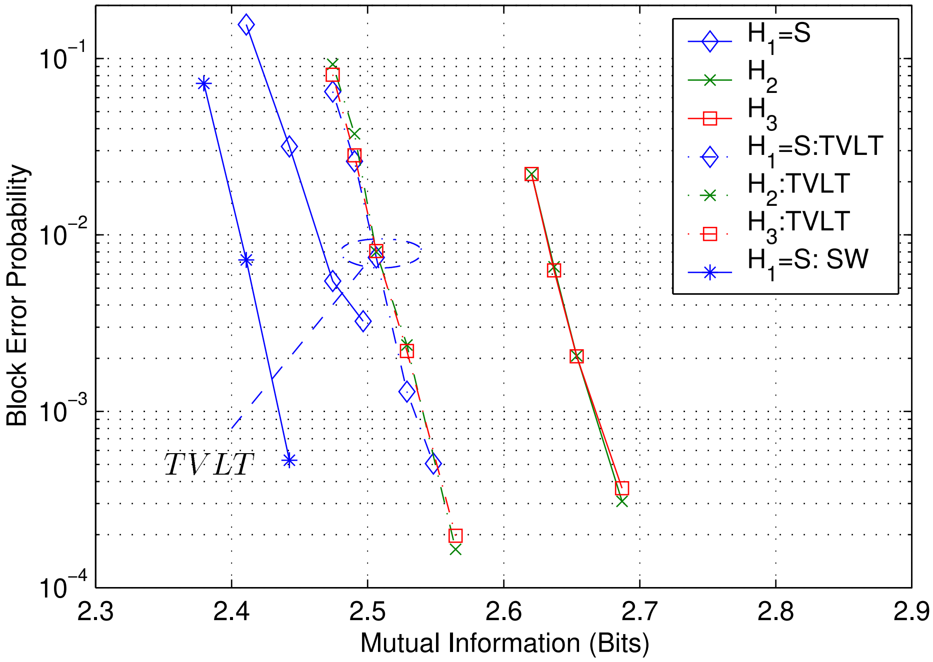

Figure 4.7: Block error rate vs. mutual information for the family of channels with λ_1 =1, λ_2=0.25. The notation $H_i : TVLT$ represents the time-varying linear transformation for channel H_i used at both the transmitter and receiver. The notation $H_1 = S : SW$ represents the transmitted signals x_1 and x_2 alternated for each antenna at the transmitter for the diagonal channel $H_1 = S$.

At BER=10^{-5}, the diagonal matrix H_1 requires approximately 2.5 bits per 2 antennas and the other 2 channels H_2 and H_3 require approximately 2.7 bits per 2 antennas; the gap between bits the diagonal channel H_1 requires and the other 2 channels require is 0.2 bits. Also the 3 channels need an excess MI of 0.25-0.35 bits per antenna.

In this case, the TVLT technique is used with parameters $p_1 = 12, p_2 = 1$. In the simulation, we only consider the linear transformation in the real domain, i.e. $\gamma_0 = 0$ as shown in (4.7). Using this technique, the diagonal channel H_1 performs slightly worse but the other 2 channels H_2 and H_3 perform much better. The performances of the 3 channels are much more consistent by using the TVLT. The performance loss suffered by the diagonal channels H_1 is much less than the performance gain by the 2 channels H_2 and H_3. At BER=10^{-5}, the diagonal matrix H_1 requires almost 2.54 bits per 2 antennas and the other 2 channels require 2.56 bits per 2 antennas; the gap between the bits required by the 3 channels is 0.02 bits. The 3 channels consistently need an excess MI of 0.27-0.28 bits per antenna.

For the diagonal channel $H_1 = S$, the performance is improved by alternating between the transmitted signals x_1 and x_2 for each antenna. At $BER = 10^{-5}$, the diagonal channel $H_1 = S$ requires almost 2.5 bits per 2 antennas. By alternating between the transmitted signals x_1 and x_2 for each antenna, this channel requires only 2.47 bits per 2 antenna.

Figure 4.5 also shows the curve for block error rate vs. MI for all the channels shown in Figure 4.4 with or without the TVLT. Similarly, Figure 4.7 shows the curve for block error rate vs. MI for all the channels shown in Figure 4.6 with or without the TVLT. We can see that the performance in terms of block error rate vs. MI is similar to the performance in terms of BER vs. MI. One interesting observation is that, for the same BER, the block error rate is almost one order less in channels with $\lambda_1 = 1, \lambda_2 = 0.25$ than the channels with $\lambda_1 = 1, \lambda_2 = 0.5$. The largest block error rate occurs when the channels have singular values $\lambda_1 = \lambda_2 = 1$.

4.3.3 Discussions

In the examples, we show that the performance is dictated by the set of singular values. The channel transformation and space-time iterative decoding techniques used in the receiver make the performance robust. When the values of λ_1 and λ_2 are close to each other, the performance in the family of channels is almost the same as the diagonal channel which can achieve turbo-coding performance similar to that in the AWGN channels.

When λ_1 and λ_2 are significantly different, there are two differences from the previous case. First the diagonal channels need more additional mutual information to get the desired BER. Although not shown in this chapter, alternating between the transmitted signals x_1 and x_2 for each antenna can improve performance in this case. Second there is some noticeable gap between the performance of the diagonal channels and the channels with the worst interference. In this case, a time-varying symbol-wise linear transformation technique provides robust performance in the family of channels. We observed that the performance loss suffered by the good channels is much less than the performance gain obtained by the bad channels.

The TVLT technique can be useful for channels with singular values close to each other as shown in Figure 4.4. This result suggests the use of TVLT at both the transmitter and receiver sides for any channels in broadcasting communications, where the transmitter has no knowledge about the channels but the receiver has the perfect knowledge of channel. By using TVLT, the overall performance for all the channels can be improved significantly.

We also observe that, as the singular values get far apart, then the transmission blocks tend to make less block errors but the number of bit errors in each error block tend to be larger. The simulation results show that this phenomenon is more obvious when the TVLT technique is used.

Chapter 5

Robustness of Space-Time Turbo Codes for Frequency Selective Fading

This chapter shows the robustness of turbo codes in frequency selective channels with 2 transmit antennas and 2 receive antennas. The robustness is achieved in terms of proximity of channel capacity. G.Bauch [4] proposed a MAP equalizer/decoder for space-time coded signals over frequency selective channels. This chapter extends this MAP equalizer/decoder to space-time turbo codes and shows that the decoding results are close to channel capacity.

5.1 MAP Equalizer over frequency selective channels

When the channel has multiple paths, an equalizer should be used before the 2 decoders. For the frequency selective channels as shown in Figure 5.1, the block interleaver is used at the outputs of both upper and lower encoders instead of using a block interleaver only at the output of upper encoder as in the flat fading channels in Figure 4.2. At the receiver side as shown in Figure 5.2, the MAP equalizer detects symbols by providing symbol reliability $P(x_1; O)$ to the upper decoder and $P(x_2; O)$ to the lower decoder. The decoders use both the extrinsic information about the input from the other decoder and the output of the corresponding encoder from the MAP equalizer. The outputs of the decoders are the reliability about the input $P(u; O)$ which is fed back to the other decoder and the reliability about the output $P(x_1; O)/P(x_2; O)$ of the corresponding encoder which is fed back to the MAP equalizer. As shown in Figure 5.2, the MAP equalizer and the 2 decoders are connected by interleavers which is standard in SISO networks [31].

The output of the channel is the channel input to MAP equalizer. The iterative decoders for the frequency selective channels do not exchange the extrinsic information about the output of the other encoder as in the flat fading channels. The forward-backward decoders exchange the extrinsic information about the input with each other and exchange the extrinsic information about the output of the corresponding encoder with the MAP equalizer.

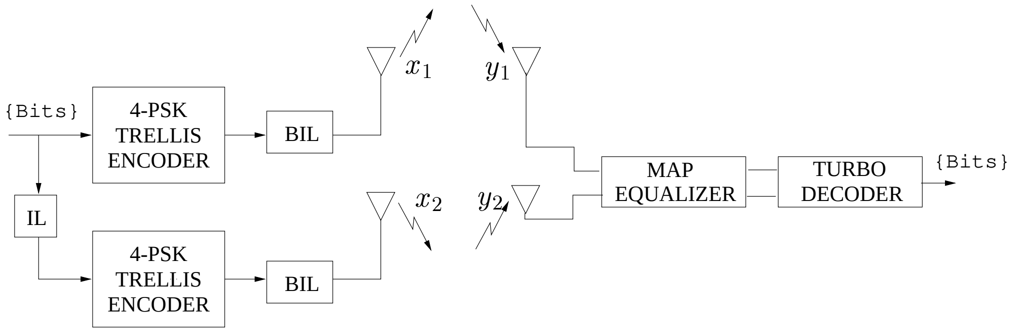

Figure 5.1: Space-time model with turbo codes over frequency selective channel.

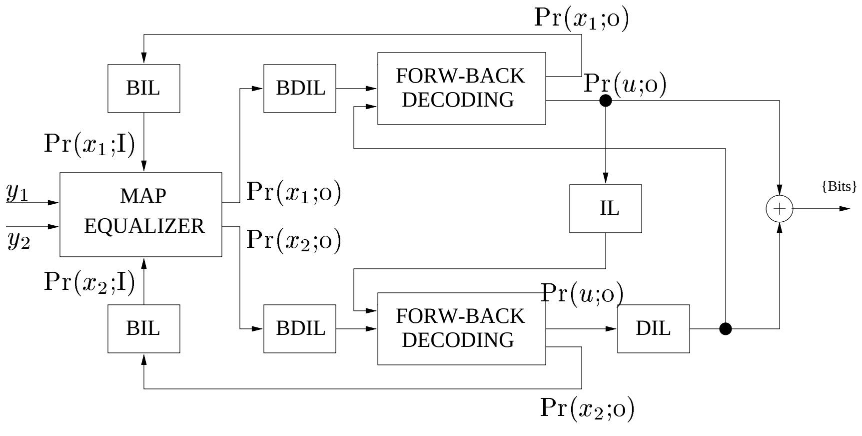

Figure 5.2: Space-time iterative decoding with MAP equalization.

5.1.1 MAP equalizer

The information is transmitted in finite blocks of length N. If $x_1(k)$ and $x_2(k)$ are the transmitted signals at time k with constellation size M, there are M^2 possible code vectors $x(k) = [x_1(k), x_2(k)]$ that can be transmitted at any given time k. With the FIR impulse response of the channel between the transmitter and the receiver, the channel trellis that the MAP equalization uses can be described as follows:

1. State: a state $m(k)$ at time k is defined as

$$m(k) \triangleq (x(k-D), x(k-D+1), \cdots, x(k-2), x(k-1)),$$

where D is the memory depth of the channel.

2. Branch or Transition: a branch or transition $\zeta(k)$ at time k is described as

$$\zeta(k) \triangleq (m(k), m(k+1)).$$

The channel trellis has $|S| = M^{2D}$ states and there are M^2 branches to and from each state.

3. Branch Input: A branch $\zeta(k)$ determines the only one input $x(k)$.

4. Branch Output or Branch Squared Euclidean Distance: a branch determines a noiseless output $y(\zeta(k) = H \cdot [x_1(k), \cdots, x_1(k-D), x_2(k), \cdots, x_2(k-D)]^T$. The squared Euclidean distance of a branch $\zeta(k)$, denoted by $D(k)$, is defined as

$$D(k) \triangleq |r(k) - y(\zeta(k))|^2.$$

The MAP equalizer shown in Figure 5.2 is used to compute the probabilities $Pr(x_1(k);O) = Pr(x_1(k)|\mathbf{r}_0^{N-1})$ and $Pr(x_2(k);O) = Pr(x_2(k)|\mathbf{r}_0^{N-1})$ for $0 \leq k \leq N-1$, given the received signal vectors $\mathbf{r}_0^{N-1} \triangleq [\mathbf{r}(0), \mathbf{r}(1), \cdots, \mathbf{r}(N-1)]$. Besides the received signal vectors $[\mathbf{r}(0), \mathbf{r}(1), \cdots, \mathbf{r}(N-1)]$ the equalizer accepts the extrinsic information $Pr(x_1(k);I)$ and $Pr(x_2(k);I)$ about the code symbols. This extrinsic information is obtained by feedback from the decoder 1 and the decoder 2 respectively.

In the first equalization step (iteration 0) all symbols $x_1(k)$ and $x_2(k)$ are assumed to be equally likely. Let M be the number of constellation points and $c_0, c_1, \cdots, c_{M-1}$ be the constellation points. For each possible symbol $x_i(k) = c_n, n = 0, 1, \cdots, M-1, i = 1, 2$ the MAP equalizer computes the a posteriori information $Pr(x_1(k) = c_n|\mathbf{r}_0^{N-1})$ and $Pr(x_2(k) = c_n|\mathbf{r}_0^{N-1})$ according to

$$Pr(x_i(k) = c_n|\mathbf{r}_0^{N-1}) = \sum_{(m',m):x_i(k)=c_n} Pr(m', m, \mathbf{r}_0^{N-1}) \tag{5.1}$$

where $m' = m(k-1)$ is the state of the trellis at time $k-1$ and $m = m(k)$ is the state of the trellis at time k for the respective transition. The sum is taken over those transitions $\zeta(k) = (m', m)$ which are labeled with the code symbol $x_i(k) = c_n$. The joint probability $Pr(m', m, \mathbf{r}_0^{N-1})$ is the product of three factors, two of which can be calculated by recursive formulae:

$$Pr(m', m, \mathbf{r}_0^{N-1}) = \underbrace{Pr(m', \mathbf{r}_0^{k-1})}_{\alpha_{k-1}(m')} \cdot \underbrace{Pr(m|m') \cdot Pr(\mathbf{r}(k)|m', m)}_{\gamma_k(m',m)} \cdot \underbrace{Pr(\mathbf{r}_{k+1}^{N-1}|m)}_{\beta_k(m)} \tag{5.2}$$

where the $\alpha_k(m)$ is the forward probability of state m at time k and is given by

$$\alpha_k(m) = \sum_{m'} \gamma_k(m', m) \cdot \alpha_{k-1}(m'), \tag{5.3}$$

with initial value $\alpha_0(m) = \frac{1}{|S|}, m = 0, 1, \cdots, |S| - 1$. $\beta_{k-1}(m')$ is the backward probability of state m' at time $k - 1$ and is given by

$$\beta_{k-1}(m') = \sum_{m} \gamma_k(m', m) \cdot \beta_k(m), \tag{5.4}$$

with the initial value at time $N - 1$ as $\beta_{N-1}(m) = \frac{1}{|S|}, m = 0, 1, \cdots, |S| - 1$. and $\gamma(m', m)$ is the branch probability for the existing transitions and can be expressed as the product of an a priori probability and the channel transition probability:

$$\gamma_k(m', m) = Pr(m|m') \cdot Pr(\mathbf{r}(k)|m', m). \tag{5.5}$$

The a priori information is obtained by the feedback of the extrinsic information $Pr(x_1(k); I)$ of decoder 1 and $Pr(x_2(k); I)$ of decoder 2:

$$Pr(m|m') = \prod_{i=1}^{2} Pr(x_i(k); I), \tag{5.6}$$

where the transition $\zeta(k) = (m', m)$ is labeled with the code symbols $c_i, i = 1, 2$. In iteration zero, all transitions are assumed to be equally likely. The channel transition probability is

$$Pr(\mathbf{r}(k)|m', m) = \frac{1}{2\pi\sigma^2} \exp(-\frac{D(k)}{2\sigma^2}). \tag{5.7}$$

The a posteriori information of the equalizer $Pr(x_1(k) = c_n|\mathbf{r}_0^{N-1})$ is the *channel input* to the upper decoder and $Pr(x_2(k) = c_n|\mathbf{r}_0^{N-1})$ is the *channel input* to the lower decoder.

The upper/lower decoders deliver a posteriori probabilities for $i = 1, 2$

$$\tilde{P}r(x_i(k) = c_n|\mathbf{r}_0^{N-1}) = \sum_{(m',m):x_i(k)=c_n} \tilde{P}r(m', m, \mathbf{r}_0^{N-1}) \tag{5.8}$$

$$= \sum_{(m',m):x_i(k)=c_n} \tilde{\alpha}_{k-1}(m')\tilde{\gamma}_k(m', m)\tilde{\beta}_k(m) \tag{5.9}$$

The crucial part of the forward-backward algorithm in the upper decoder is

$$\tilde{\gamma}_k(m', m) = \tilde{P}r(m|m') \tag{5.10}$$

$$= \tilde{P}r(x_1(k); I) \cdot \tilde{P}r(u; I) \tag{5.11}$$

The $\tilde{P}r(x_1(k); I)$ is the block-interleaved $Pr(x_1(k); O)$ which is the output of the MAP equalization.

We know that it is important to feed back only the extrinsic part of the a posteriori information because the correlation between the a priori information used by the equalizer and previous decisions of the equalizer should be minimized. Ideally the a priori information would be an independent estimation. The extrinsic information about a certain symbol is the incremental information about the current symbol obtained through the decoding process from all information which is available for the other symbols in the block. It can be calculated by subtracting the logarithm of the input probabilities of the decoder from the logarithm of the output probabilities.

5.2 Examples

In the simulations of using MAP equalizer/decoder for the frequency selective channels, the parallel concatenated codes we use employ the rate 2/2 QPSK

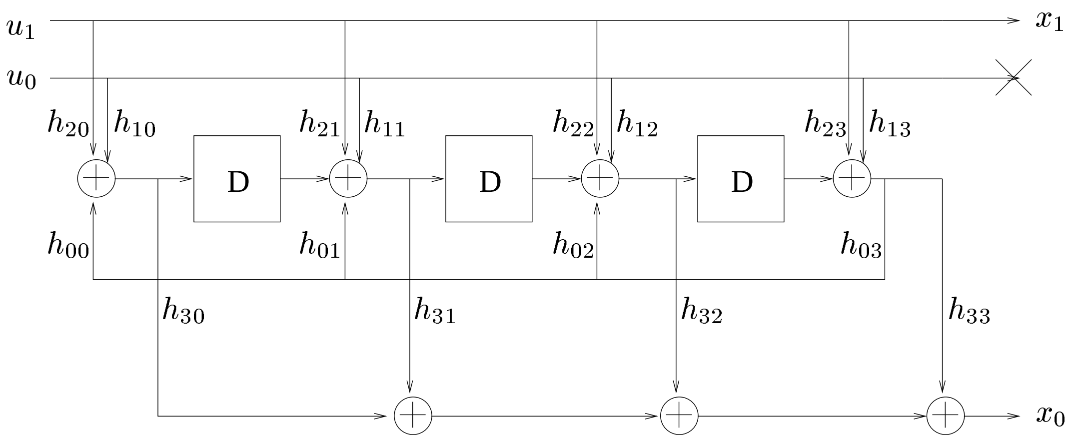

Figure 5.3: Encoder structure with 3 memory elements

with polynomial $(h_0, h_1, h_2, h_3) = (13, 1, 7, 17)$ for each constituent RSC code. The encoder structure for this RSC code can be represented by Figure 5.3. The puncture pattern and the interleaver of the input symbols are the same as in the previous subsection. Here block interleavers are used at both the upper and lower encoder outputs as in Figure 5.1.

Here we consider a 2-tap channel with channel matrix

$$H_1 = \begin{bmatrix} 0.9487 & 0.3162 & 0.9285 & -0.3714 \\ 0.9285 & -0.3714 & 0.9487 & 0.3162 \end{bmatrix}. \tag{5.12}$$

The simulation stops when the number of block errors reaches 50 with 10 iterations. At BER=10^{-5}, the channel needs the capacity of 2.38bits; it needs an excess of 0.38bits as shown in Figure 5.6.

We also consider the structure, where the precoded channel is the inner rate 1 code and a rate 1/2 feedforward convolutional code with polynomial $(G_1(D), G_2(D)) = (1 + D + D^3, 1 + D + D^2 + D^3)$ as the outer code. We did an exhaustive search for the good precoders. With one of the good precoders for the given channel H_1, the bits required for BER in the order of 10^{-3} is 2.45bits. Compared to the structure without the precoder and using the turbo codes, the precoded channels need at least an extra 0.15bits. From the BER point of view, the structure without precoder and using the turbo codes is preferred. Also the structure is preferred in the sense of robustness; one good precoder is optimal only for some MIMO FIR channels and different channels need different precoders.

The results of this chapter also demonstrate the robustness of the space-time turbo-codes in frequency selective fading channels. By using the MAP equalizer and upper and lower decoders, the robustness is achieved in terms of proximity to channel capacity. This robustness in frequency selective fadings is in a sense analogous to the robustness in frequency non-selective fadings as shown in previous chapter.

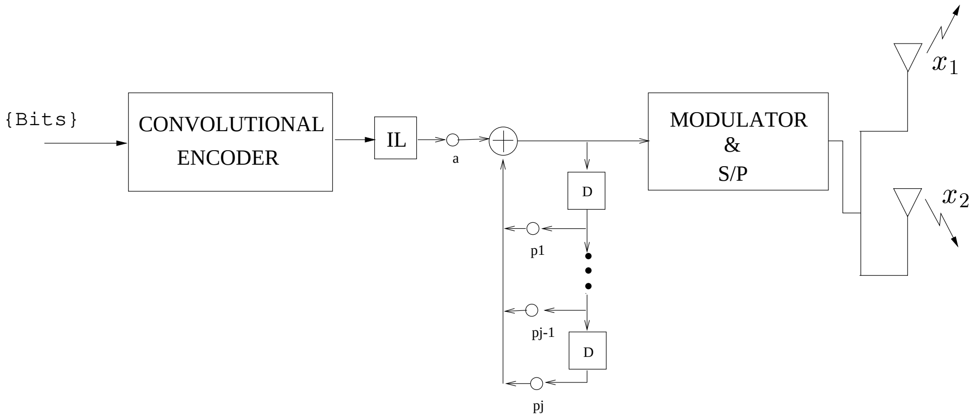

Figure 5.4: Space-time model with precoder over frequency selective channel.

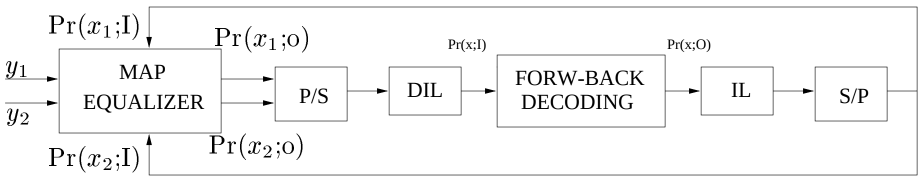

Figure 5.5: Space-time iterative decoding with MAP equalization for precoder case.

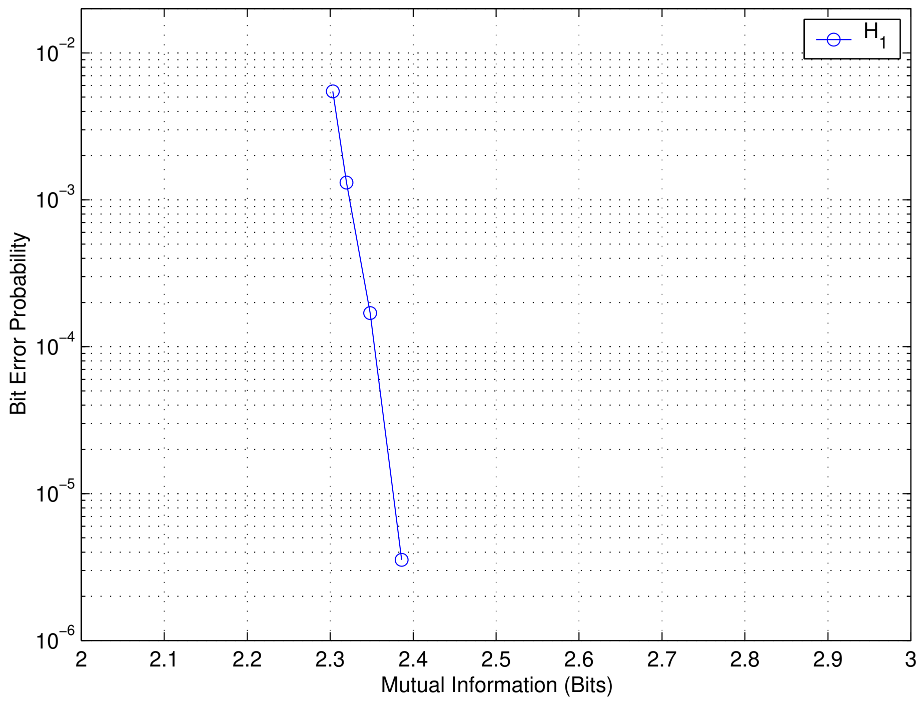

Figure 5.6: BER vs. mutual information for the channel in the expression (5.12).

Chapter 6

Quasiconvexity and Optimal Binary Fusion for Distributed Detection with Identical Sensors in Generalized Gaussian Noise

In wireless sensor networks, the local sensors send information with the finite precisions to fusion center and fusion center makes the final decisions based on these information. To reduce the power consumption in each sensor, the number of bits of the information sent by the sensors to the fusion center need be as small as possible. There is always tradeoff between the precision of the information and the system performance, i.e. the probability of error. The design of local processing in every sensor and the fusion rule in the fusion center to achieve global optimal performance is a challenging research problem. In this chapter, we consider a binary fusion distributed detection with the assumption of identical sensors and present a technique for the system design to achieve the global optimal performance.

Consider distributed detection of $s \in \{-m, m\}$, where the ith of n local sensors observes $x_i = s + z_i$ with i.i.d. noise z_i. The ith sensor compares x_i to a threshold τ to compute a binary decision u_i as

$$u_i = \begin{cases} 0 & \text{when } x_i < \tau \\ 1 & \text{when } x_i \geq \tau\,. \end{cases}$$

Each binary decision u_i is transmitted to a fusion center, which applies a fusion rule F to $k = \sum_{i=1}^{n} u_i$ to produce the final decision $F(k)$. The case with $n = 2$ and Gaussian noise is considered as an example in [33]. This chapter extends this specific case to the general case with any number of sensors n for the family of generalized Gaussian noise and describes a technique for identifying the pair (τ, F) that minimizes the probability of error. For equal *a-priori* probabilities, a sufficient condition of the non-Gaussian noise distribution $g(x)$ for the probability of error to be quasiconvex is given and some non-Gaussian noise distributions which satisfy this condition are listed. This technique extends to Bayes risk and Neyman-Pearson criterion.

The identical threshold τ in local sensors generally does not result in an optimum system. However the local identical threshold assumption reduces the complexity dramatically. For binary hypothesis detection, Irving and Tsitsiklis [34] showed that no optimality is lost with identical local detectors in a two sensor system. Tsitsiklis [35], Chen and Papamarcou [36] showed that identical local detectors are asymptotically optimum when the number of sensors n tends to infinity. These results provide some justification for restricting attention to identical quantizers.

Even with identical local thresholds, the problem is still complicated by the existence of multiple local minima. Furthermore, minimizing probability of error is difficult because the Bayesian error probability is not a smooth function, i.e. its first derivative is a discontinuous function. Hashlamoun and Varshney [33, 37] have overcome this difficulty by using a smooth bound on the Bayesian error probability to approximately determine the optimal pair (τ, F). Avi-Itzhak and Diep [38] provided an even tighter bound leading to a very good approximation. Even so, these techniques still produce a minimization problem that may have local minima, since neither convexity nor quasiconvexity was shown in these references.

In this chapter we show that for any admissible fusion rule F_i (i.e. any F_i that is optimal for at least one τ), $P_e(\tau, i)$ defined as the probability of error for (τ, F_i) is a quasiconvex function of τ. Quasiconvexity [39] means that every sublevel set $S_\alpha = \{\tau : P_e(\tau, i) \leq \alpha\}$ is convex.

The admissible functions F_i are simply threshold tests of the form

$$s = F_i(k) = \begin{cases} -m & \text{if } k < i \\ m & \text{if } k \geq i\,. \end{cases} \tag{6.1}$$

Hence, the problem decomposes into a series of n quasiconvex optimization problems (one for each F_i) that may be solved exactly using a variety of techniques including, for example, the ellipsoid algorithm [40].

6.1 Generalized Gaussian Noise

6.1.1 Admissible Fusion Rules

This section identifies the admissible fusion rules for generalized Gaussian noise and shows that the probability of error vs. τ is quasiconvex for these rules.

First define

$$A_k(\tau) = p_0 \binom{n}{k} Q^k(\tau + m) Q^{n-k}(-\tau - m),$$

$$B_k(\tau) = p_1 \binom{n}{k} Q^k(\tau - m) Q^{n-k}(-\tau + m),$$

where p_0 is the *a-priori* probability of $s = -m$ and $p_1 = 1 - p_0$ is the *a-priori* probability of $s = m$. $A_k(\tau)$ is the probability that k sensors decide 1 and $n - k$ sensors decide 0 while $s = -m$ and $B_k(\tau)$ is the probability that k sensors decide 1 and $n - k$ sensors decide 0 while $s = m$. $Q(\tau) = \int_\tau^\infty f(x)\mathrm{d}x$, where

$$f(x) = \frac{C}{2\Gamma(\frac{1}{C})[\eta(C)]^{\frac{1}{C}}} \exp\left(-\frac{|\, x \,|^C}{\eta(C)}\right),$$

and

$$\eta(C) = \left[\frac{\sigma^2 \Gamma(\frac{1}{C})}{\Gamma(\frac{3}{C})}\right]^{\frac{C}{2}}.$$

σ^2 is the noise variance and $\Gamma(\cdot)$ is the Gamma function. The noise distribution is Gaussian when the positive parameter $C = 2$ and is Laplacian when $C = 1$.

$Q(\tau + m)$ is the probability of false alarm and $Q(\tau - m)$ is the probability of detection for the local detectors.

Under the maximum *a-posteriori* probability (MAP) criterion, the optimal F for a fixed τ makes the probability of error $P_e(\tau)$ minimal. That is

$$P_e(\tau) = \sum_{k=0}^{n} \min(A_k(\tau), B_k(\tau)) \tag{6.2}$$

Therefore, the optimal F is related to $A_k(\tau)$ and $B_k(\tau)$ as

$$F(k) = \begin{cases} -m & \text{if } A_k(\tau) \geq B_k(\tau) \\ m & \text{if } A_k(\tau) < B_k(\tau)\,. \end{cases} \tag{6.3}$$

Under MAP criterion, the optimal F for a fixed τ has the following property, if $F(k) = -m$, then $F(k-1) = -m$ and if $F(k) = m$, then $F(k+1) = m$.

From (6.3), $F(k) = -m$ implies that

$$A_k(\tau) \geq B_k(\tau). \tag{6.4}$$

and also,

$$\begin{aligned} \frac{A_{k-1}(\tau)}{B_{k-1}(\tau)} &= \frac{A_k(\tau)}{B_k(\tau)} \frac{Q(\tau - m)Q(-\tau - m)}{Q(\tau + m)Q(-\tau + m)} \\ &\geq \frac{A_k(\tau)}{B_k(\tau)}. \end{aligned} \tag{6.5}$$

As a consequence of inequalities (6.4) and (6.5), then $A_{k-1}(\tau) \geq B_{k-1}(\tau)$, which means that $F(k-1) = -m$. Similarly if $F(k) = m$, then $F(k+1) = m$. This property implies that for every τ the MAP F has the form F_i in (6.1) and only fusion rules of the form F_i in (6.1) are admissible, i.e. are MAP for some choice of τ. Actually the admissible fusion rule holds for any noise distribution. The same admissible fusion rule was observed by Chair and Varshney [41] and Reibman and Nolte [42].

6.1.2 Quasiconvexity

Theorem 1 *For a fixed admissible fusion rule F_i, probability of error is a quasiconvex function of τ.*

Proof: $P_e(\tau, i)$ can be expressed as

$$P_e(\tau, i) = p_0 + \sum_{k=0}^{i-1}(B_k(\tau) - A_k(\tau)). \tag{6.6}$$

Using the facts that

$$\binom{n}{k}(n-k) = \binom{n}{k+1}(k+1),$$

and

$$Q'(\tau - m) = Q'(\tau + m)\delta(\tau),$$

where

$$\delta(\tau) = \exp\left(\frac{\mid \tau + m \mid^C - \mid \tau - m \mid^C}{\eta(C)}\right),$$

the derivative of $P_e(\tau, i)$ is a telescoping sum (i.e. has the form $\sum_{k=0}^{i-1}(f_k(\tau) - f_{k-1}(\tau))$ for a specific $f_k(\tau)$) and can be simplified to

$$P_e'(\tau, i) = \alpha(\tau)(\beta(\tau) - \gamma(\tau)), \tag{6.7}$$

where

$$\alpha(\tau) = \binom{n}{i} i|Q'(\tau + m)|Q^{i-1}(\tau + m)Q^{n-i}(-\tau + m),$$

$$\beta(\tau) = p_1\, \delta(\tau) \left(\frac{Q(\tau - m)}{Q(\tau + m)} \right)^{i-1},$$

$$\gamma(\tau) = p_0 \left(\frac{Q(-\tau - m)}{Q(-\tau + m)} \right)^{n-i}.$$

$\alpha(\tau)$ is always positive, $\beta(\tau)$ is a positive and monotonically increasing function and $\gamma(\tau)$ is a positive and monotonically decreasing function. If C is positive and $C \neq 1$, $\beta(\tau)$ intersects $\gamma(\tau)$ since $\beta(-\infty) < \gamma(-\infty)$ and $\beta(\infty) > \gamma(\infty)$. So $P_e'(\tau, i) = 0$ for only one τ^*, for which $\beta(\tau^*) = \gamma(\tau^*)$. For $\tau < \tau^*, P_e'(\tau, i) < 0$ and $\tau > \tau^*, P_e'(\tau, i) > 0$. So $P_e(\tau, i)$ is quasiconvex.

If $C = 1$, $\beta(\tau)$ does not intersect $\gamma(\tau)$ for some specific *a-priori* ratio p_0/p_1, which means that either $\beta(\tau) \geq \gamma(\tau)$ or $\beta(\tau) \leq \gamma(\tau)$ for all τ. In both cases, $P_e(\tau, i)$ is still quasiconvex since either $P_e'(\tau, i) \geq 0$ for all τ or $P_e'(\tau, i) \leq 0$ for all τ.

6.2 Non-Gaussian Noises

In this section, we extend our results to some well known non-Gaussian noises for equal *a-priori* probability case, i.e. $p_1 = p_0 = \frac{1}{2}$. For the noise distribution $g(x)$, it is assumed even and at least twice differentiable. Here again, a single identical threshold τ in the observation space is used as in the generalized Gaussian noise case.

Similar to the generalized Gaussian noise case, we define

$$A_k(\tau) = \frac{1}{2} \binom{n}{k} \tilde{Q}^k(\tau + m)\tilde{Q}^{n-k}(-\tau - m),$$

$$B_k(\tau) = \frac{1}{2}\binom{n}{k}\tilde{Q}^k(\tau - m)\tilde{Q}^{n-k}(-\tau + m),$$

where $\tilde{Q}(\tau) = \int_\tau^\infty g(x)\mathrm{d}x$ and $g(x)$ is the pdf of the non-Gaussian noise distribution.

As in the generalized Gaussian noise case, fusion rules of the form F_i in (6.1) are MAP for some choice of τ.

In order to prove the quasiconvexity of the probability of error $P_e(\tau, i)$ w.r.t. τ, a sufficient condition for the non-Gaussian noise distribution $g(x)$ is derived in the following theorem.

Theorem 2 *Given a noise distribution $g(x)$, if it is even, twice differentiable and satisfies:*

$$2(g'(x))^2 \geq g''(x)g(x). \tag{6.8}$$

then for a fixed admissible fusion rule F_i and equal a-priori *probability, i.e. $p_1 = p_0 = \frac{1}{2}$, the probability of error is a quasiconvex function of τ.*

Proof: The derivative of $\tilde{Q}''(x)/\tilde{Q}'(x)^2$ is the l.h.s. of (6.9) below, which is nonnegative as a result of (6.8):

$$\frac{2(g'(x))^2 - g''(x)g(x)}{(g(x))^3} \geq 0. \tag{6.9}$$

Thus, $\tilde{Q}''(x)/\tilde{Q}'(x)^2$ is monotonically increasing for $-\infty < x < \infty$. Assuming $g(x) = g(-x)$, $D(x, i) = \tilde{Q}^{i-1}(x)\tilde{Q}^{n-i}(-x)\tilde{Q}'(x)$ is a quasiconvex function of x, since

$$D'(x, i) = \tilde{Q}^{i-1}(x)\tilde{Q}^{n-i}(-x)(\tilde{Q}'(x))^2(S(x) - T(x)). \tag{6.10}$$

where

$$S(x) = \frac{i-1}{\tilde{Q}(x)} + \frac{\tilde{Q}''(x)}{\tilde{Q}'(x)^2},$$

$$T(x) = \frac{n-i}{\tilde{Q}(-x)},$$

and i=1, 2,..., n. $S(x)$ is monotonically increasing and $T(x)$ is monotonically decreasing. Furthermore, $S(-\infty) < T(-\infty)$ and $S(\infty) > T(\infty)$. So there exists only one x^*, for which $D'(x^*, i)$=0. For $x < x^*$,$D'(x, i) < 0$ and $x > x^*$,$D'(x, i) > 0$. Therefore $D(x, i)$ is a quasiconvex function w.r.t. x.

The first derivative of the probability error of $P_e'(\tau, i)$ can be expressed as:

$$P_e'(\tau, i) = \frac{i}{2}\binom{n}{i}[D(\tau+m, i) - D(\tau-m, i)]. \tag{6.11}$$

With the fact that $D(\tau + m, i)$ is a shifted-by-$2m$ copy of $D(\tau - m, i)$, there is only one τ^* for which $P_e'(\tau, i) = 0$. For $\tau < \tau^*$, $P_e'(\tau, i) < 0$ and $\tau > \tau^*$, $P_e'(\tau, i) > 0$. So $P_e(\tau, i)$ is quasiconvex. Λ

There are several well known non-Gaussian noise distributions [43, 44] satisfying Theorem 2. These distributions are listed in Table 1, where again $\Gamma(.)$ is the Gamma function, a is a *scale* parameter related to the common variance σ^2 of the quadrature components, and ν is a *shape* parameter ruling the rate of decay of the noise pdf. The generalized Cauchy distribution reduces to the Gaussian distribution in the case $\nu \to \infty$.

Lemma 1 *$Pe(\tau, i)$ with hyperbolic secant noise is quasi-convex.*

Proof: For the hyperbolic secant noise distribution,

$$g(\tau) = \frac{1}{2\sigma}\text{sech}(\frac{\pi\tau}{2\sigma}) = \frac{1}{\sigma}\frac{1}{e^{\frac{\pi\tau}{2\sigma}} + e^{-\frac{\pi\tau}{2\sigma}}}.$$

It can be shown that

$$2(g'(\tau))^2 - g''(\tau)g(\tau) = \frac{64\sigma^4}{\pi^2(e^{\frac{\pi\tau}{2\sigma}} + e^{-\frac{\pi\tau}{2\sigma}})^2} > 0.$$

Lemma 2 *$Pe(\tau, i)$ with hyperbolic secant squared noise is quasi-convex.*

Proof: Let $k(\tau) = \frac{\tilde{Q}''(\tau)}{[\tilde{Q}'(\tau)]^2}$. For the hyperbolic secant squared noise distribution, we have $k'(\tau) = \phi(\tau)\psi(\tau)$, where

$$\phi(\tau) = \frac{e^{\frac{\pi\tau}{2}}}{4\pi}[(\pi\tau - 2)^2 + 4],$$

$$\psi(\tau) = \frac{1}{\tau^3} - \frac{e^{-\pi\tau}[(\pi\tau + 2)^2 + 4]}{\tau^3[(\pi\tau - 2)^2 + 4]}.$$

$\phi(\tau)$ is always positive and $\psi(\tau)$ can be shown positive for any τ by algebraic manipulation. So $k(\tau)$ is monotonically increasing. Hence, $Pe(\tau, i)$ is quasi-convex from Theorem 2.

Lemma 3 *$Pe(\tau, i)$ with t-distribution noise is quasi-convex.*

Proof:

$$2(g'(\tau))^2 - g''(\tau)g(\tau) = \frac{A^2(n^2-1)}{n^2}x^2(1 + \frac{x^2}{n})^{-n-3} + \frac{A^2(n+1)}{n}(1 + \frac{x^2}{n})^{-n-2} > 0,$$

where

$$A = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}.$$

When $n = 1$, the t-distribution yields a Cauchy distribution, and it reduces to a Gaussian as $n \to \infty$.

Lemma 4 *$Pe(\tau, i)$ with generalized Cauchy noise is quasi-convex.*

Proof: It can be shown that

$$2(g'(\tau))^2 - g''(\tau)g(\tau) = A(a^2+\tau^2)^{-2\nu-3}(4\nu\tau^2+2a^2),$$

where

$$A = \frac{a^{4\nu}[\Gamma(\nu+1/2)]^2(\nu+1/2)}{\pi[\Gamma(\nu)]^2}.$$

So $2(g'(\tau))^2 - g''(\tau)g(\tau)$ is positive.

6.3 Examples

Figures 6.1-6.5 present numerical examples showing the quasiconvex probability of error vs. τ curves for the admissible fusion rules F_i for $n = 7$ and 8 with $p_0 = 0.5$ and 0.8. The noise z_i is i.i.d. Gaussian distributed ($C = 2$) and Laplacian distributed ($C = 1$). All the examples consider the cases with $m = 1$ and zero mean noise with unit variance.

Figure 6.6 shows the number of sensors needed to achieve error probability of 10^{-5} in the two cases where the sensors communicate their observations to the fusion center with infinite and single-bit precision respectively. For the binary sensors, only odd numbers of sensors are considered for every SNR and the optimal pair is always $(0, F_{n/2+1})$.

For infinite precision sensors, the fused observation x at the fusion center is essentially the maximum ratio combining of $x_i, i = 1, \cdots, n$. With identical SNR at each sensor, the fused observation at the fusion center is $x = \frac{1}{n}\sum_{i=1}^{n} x_i$ with noise variance σ^2/n, where σ^2 is Gaussian noise variance for each sensor. Thus the error probability is $Q(\sqrt{n} \cdot m)$. Observe in Figure 6.6 that the number of binary sensors needed for every SNR is fewer than twice the number of infinite-precision sensors, which can make the binary sensor a better choice from a practical or economic point of view.

Figure 6.7 shows two curves for the minimum probability of error vs. the threshold τ. The solid line is for $n = 7$ and the thick line is for $n = 8$. The result indicates that the performance is improved by increasing n to $n+1$, where n is odd, provided the threshold τ is shifted from zero. If the threshold τ is assumed to be zero, then there is no performance improvement due to increasing n by 1 for odd number n as observed by Blum [45].

6.4 Further Extensions

6.4.1 Bayesian Analysis

The quasiconvexity of error probability $P_e(\tau, i)$ can be extended to Bayes risk $\Re(\tau, i)$.

If we define $A(\tau)$ and $B(\tau)$ as

$$A_k(\tau) = p_0(C_{10} - C_{00}) \begin{pmatrix} n \\ k \end{pmatrix} Q^k(\tau + m) Q^{n-k}(-\tau - m),$$

$$B_k(\tau) = p_1(C_{01} - C_{11}) \begin{pmatrix} n \\ k \end{pmatrix} Q^k(\tau - m) Q^{n-k}(-\tau + m),$$

then for every τ the optimal F has the form F_i in (6.1) for any noise distribution, where $C_{ij}, i = 0, 1, j = 0, 1$ represents the cost of declaring H_i when H_j is actually present. Here we assume that $C_{10} > C_{00}$ and $C_{01} > C_{11}$.

Then for every fusion rule F_i, the Bayes risk can be written as

$$\Re(\tau, i) = p_0 C_{10} + p_1 C_{11} + \sum_{k=0}^{i-1} (B_k(\tau) - A_k(\tau)).$$

The Bayes risk $\Re(\tau, i)$ becomes the probability of error $P_e(\tau, i)$ when $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$. Considering only the generalized Gaussian noise case, similar to the $P_e(\tau, i)$, $\Re(\tau, i)$ is a quasiconvex function of τ.

6.4.2 Neyman-Pearson Criterion

The approach of optimizing τ for each admissible fusion rule also succeeds under the Neyman-Pearson criterion. According to the Neyman-Pearson criterion, the detection probability P_D needs to be maximized under the constraint that the false alarm probability P_F is less than or equal to an acceptable value α_0. For every τ the optimal F has the form F_i in (6.1) for any noise distribution since the Neyman-Pearson test employs the likelihood ratio test. Thus as before we can restrict our attention to these admissible fusion rules.

For every admissible fusion rule F_i and generalized Gaussian noise distribution, we have

$$P_F(\tau, i) = \sum_{k=i}^{n} Q^k(\tau + m) Q^{n-k}(-\tau - m),$$

$$P_D(\tau, i) = \sum_{k=i}^{n} Q^k(\tau - m) Q^{n-k}(-\tau + m).$$

Both $P_F(\tau, i)$ and $P_D(\tau, i)$ are quasiconvex. Moreover they are both strictly monotonically decreasing functions w.r.t. τ. So for every admissible fusion rule

F_i, define τ_i to be the smallest τ such that $P_F(\tau, i) \leq \alpha_0$. Due to the monotonically decreasing property, this τ_i must satisfy $P_F(\tau_i, i) = \alpha_0$ and the corresponding $P_D(\tau_i, i)$ is the maximized detection probability for the admissible fusion rule F_i. Define

$$i^* = \arg\max_i P_D(\tau_i, i),$$

then F_{i^*} is the optimal fusion rule and the maximum detection probability is $P_D(\tau_{i^*}, i^*)$. This method also applies to the case with non-Gaussian noise distributions.

6.4.3 More than one bit quantization

Finally, consider the quantizer design for multiple bit output at each sensor. Similar to the one bit quantization, each sensor has identical quantizers and the thresholds are $\tau_1 < \tau_2 < \cdots < \tau_l$, where l is some positive integer. u_i takes values in the set of $\{0, 1, \cdots, l\}$. Is $k = \sum_{i=1}^n u_i$ a sufficient statistic for U? Or do we need to consider the vector $U = [u_1, u_2, \cdots, u_n]$ instead? The answer is that generally $k = \sum_{i=1}^n u_i$ is not a sufficient statistic and the fusion center needs to make decisions based on U instead of k. Therefore the set of fusion rules of the form F_i in (6.1) is not the same as the set of admissible rules anymore for multiple thresholds.

For n sensors and $l \geq 3$, based on $U = [1, 1, 0, \cdots, 0]$ and MAP detection rule, the fusion center decides H_0 if $(\tau_1, \tau_2) \in Z_0$ and decides H_1 if $(\tau_1, \tau_2) \in Z - Z_0$, where $Z = \{(\tau_1, \tau_2) : \tau_1 \leq \tau_2\}$. Based on $U = [2, 0, \cdots, 0]$, the fusion center decides H_0 if $(\tau_1, \tau_2, \tau_3) \in Z_0'$ and decides H_1 if $(\tau_1, \tau_2, \tau_3) \in Z' - Z_0'$, where $Z' = \{(\tau_1, \tau_2, \tau_3) : \tau_1 \leq \tau_2 \leq \tau_3\}$. Generally $Z_0 \neq Z_0'$.

For $n \geq 3$ and $l = 2$, based on $U = [1, 1, 0, \cdots, 0]$ and MAP detection rule, the fusion center decides H_0 if $(\tau_1, \tau_2) \in Z_0$ and decides H_1 if $(\tau_1, \tau_2) \in Z - Z_0$, where $Z = \{(\tau_1, \tau_2) : \tau_1 \leq \tau_2\}$. Based on $U = [2, 0, \cdots, 0]$, the fusion center decides H_0 if $(\tau_1, \tau_2) \in Z'_0$ and decides H_1 if $(\tau_1, \tau_2) \in Z' - Z'_0$, where $Z' = \{(\tau_1, \tau_2) : \tau_1 \leq \tau_2\}$. Here again, generally $Z_0 \neq Z'_0$.

However as a specific case, the fusion rules of the form F_i in (6.1) are admissible when $n = 2, l = 2$ and $p_0 = \frac{1}{2}$. The following lemma justifies that k is a sufficient statistic of U.

Lemma 5 *For $n = 2, l = 2$ and $p_0 = 1/2$, based on $k = 2$ and MAP detection rule, the fusion center decides H_0 if $(\tau_1, \tau_2) \in \mathrm{Z}_0$ and decides H_1 if $(\tau_1, \tau_2) \in \mathrm{Z} - \mathrm{Z}_0$, where $Z_0 = \{(\tau_1, \tau_2)$: $\tau_1 < 0$ and $\tau_1 \leq \tau_2 \leq -\tau_1\}$ and $Z = \{(\tau_1, \tau_2) : \tau_1 \leq \tau_2\}$.*

For the case of 2 thresholds and 2 sensors, only the fusion rules of the form F_i in (6.1) are admissible, i.e. are MAP for some choice pair (τ_1, τ_2), where F_1 and F_4 fusion rules reduce to single threshold case. So only F_2 and F_3 are considered for two threshold design. Because of the symmetry, only one of the fusion rules needs to be considered.

6.5 Discussions

This chapter presented a technique in wireless sensor network to achieve the global minimal probability of error for binary hypothesis detection. We assumed n local identical binary sensors and one fusion center, where the fusion rule of k-out-of-n is the admissible fusion rule. This chapter showed that the probability of error is a quasiconvex function of local threshold τ for each admissible fusion

Table 6.1: Non-Gaussian Noise Distributions

	$g(x)$
Hyperbolic Secant	$\frac{\text{sech}(\pi x/2\sigma)}{2\sigma}$
Hyperbolic Secant Squared	$\frac{\pi x}{2}\text{cosech}(\frac{\pi x}{2})$
t-distribution	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, n \geq 1$
Generalized	$\frac{a^{2\nu}\Gamma(\nu+1/2)}{\sqrt{\pi}\Gamma(\nu)}(a^2+x^2)^{-\nu-1/2}$
Cauchy	$a > 0; \nu > 1; a^2 = 2\sigma^2(\nu-1)$

rule. Therefore the global minimal probability of error can be found among the set of n minimal probability of errors for all the possible k-out-of-n fusion rules, which allows the optimal system design for the pair of optimal fusion rule and optimal local threshold to minimize the probability of error.

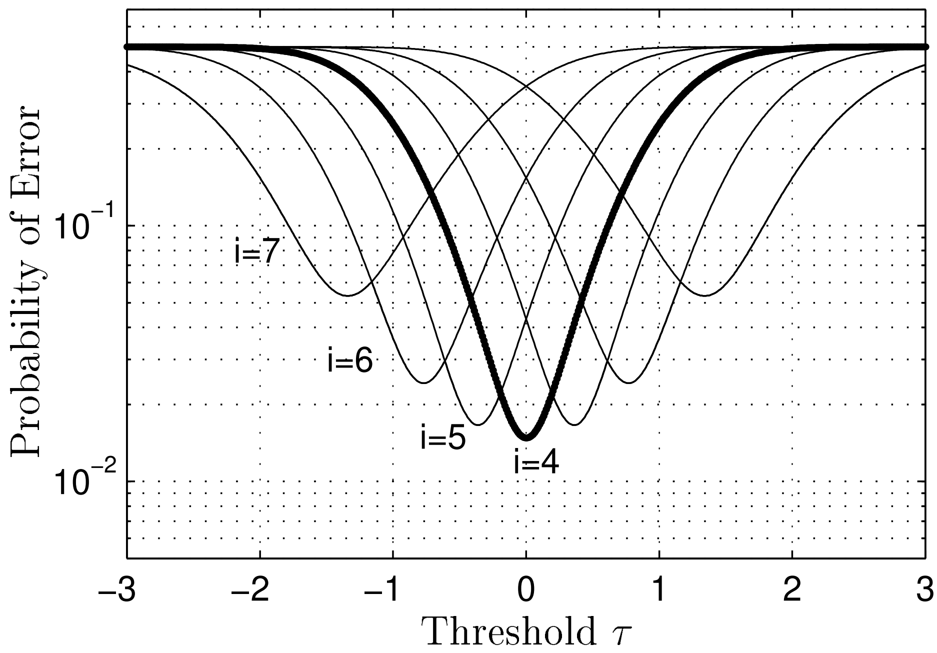

Figure 6.1: $P_e(\tau, i)$ vs. τ for various i with 7 sensors for $p_0 = 0.5$ and Gaussian noise. The optimal pair is $(0, F_4)$.

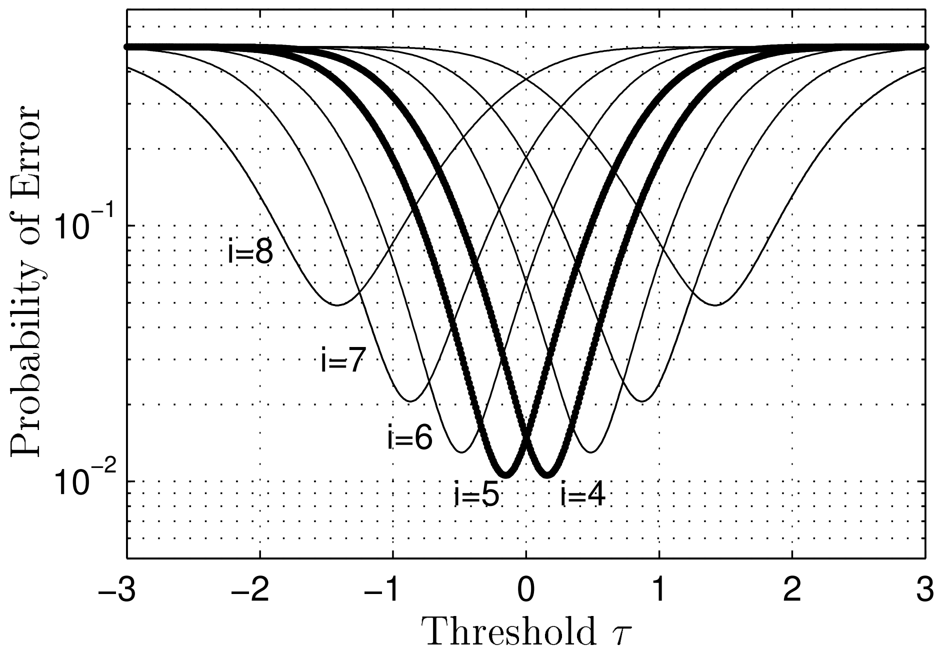

Figure 6.2: $P_e(\tau, i)$ vs. τ for various i with 8 sensors for $p_0 = 0.5$ and Gaussian noise. The optimal pairs are $(-\epsilon, F_5)$ and (ϵ, F_4), where $\epsilon = 0.1570$.

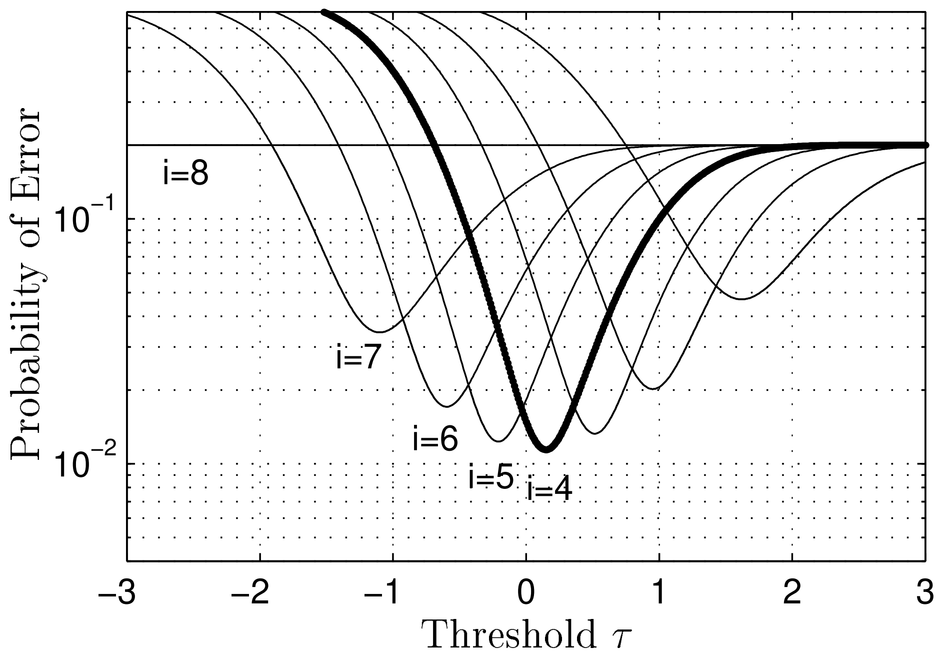

Figure 6.3: $P_e(\tau, i)$ vs. τ for various i with 7 sensors for $p_0 = 0.8$ and Gaussian noise. The optimal pair is (ϵ, F_4) where $\epsilon = 0.1471$.

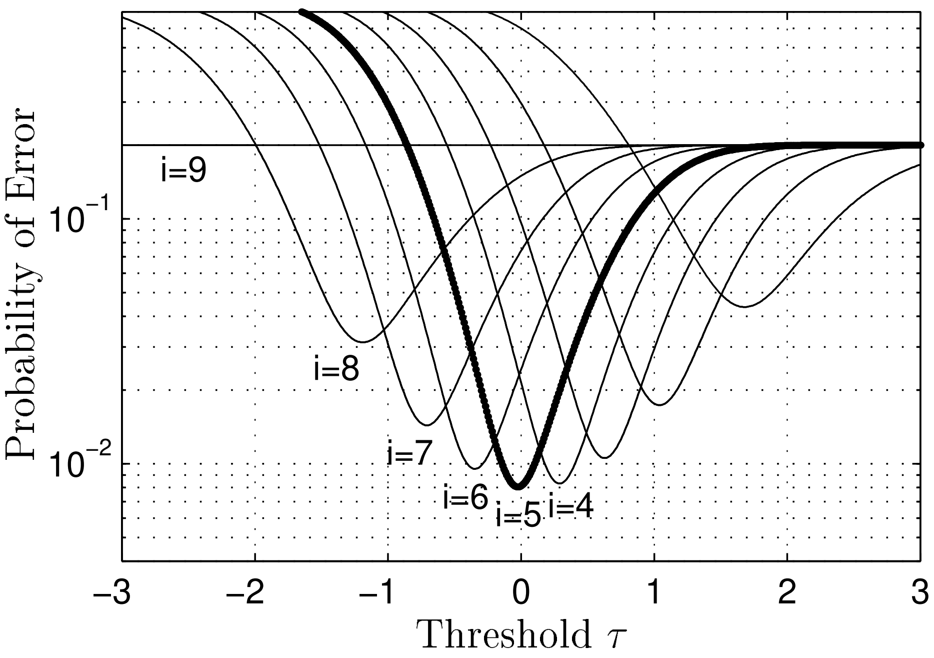

Figure 6.4: $P_e(\tau, i)$ vs. τ for various i with 8 sensors for $p_0 = 0.8$ and Gaussian noise. The optimal pair is $(-\epsilon, F_5)$, where $\epsilon = 0.0264$.

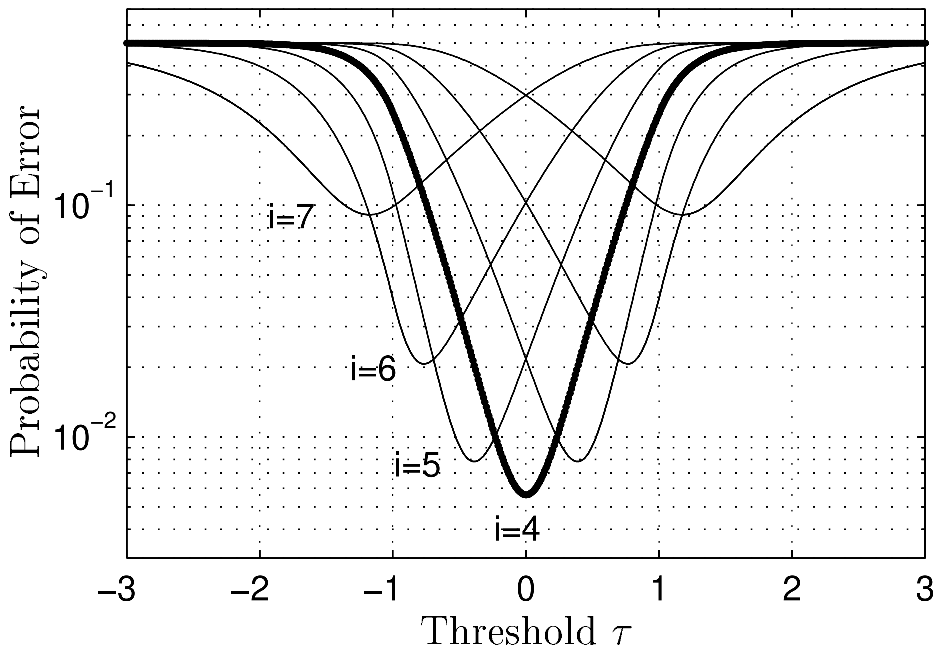

Figure 6.5: $P_e(\tau, i)$ vs. τ for various i with 7 sensors for $p_0 = 0.5$ and Laplacian noise. The optimal pair is $(0, F_4)$.

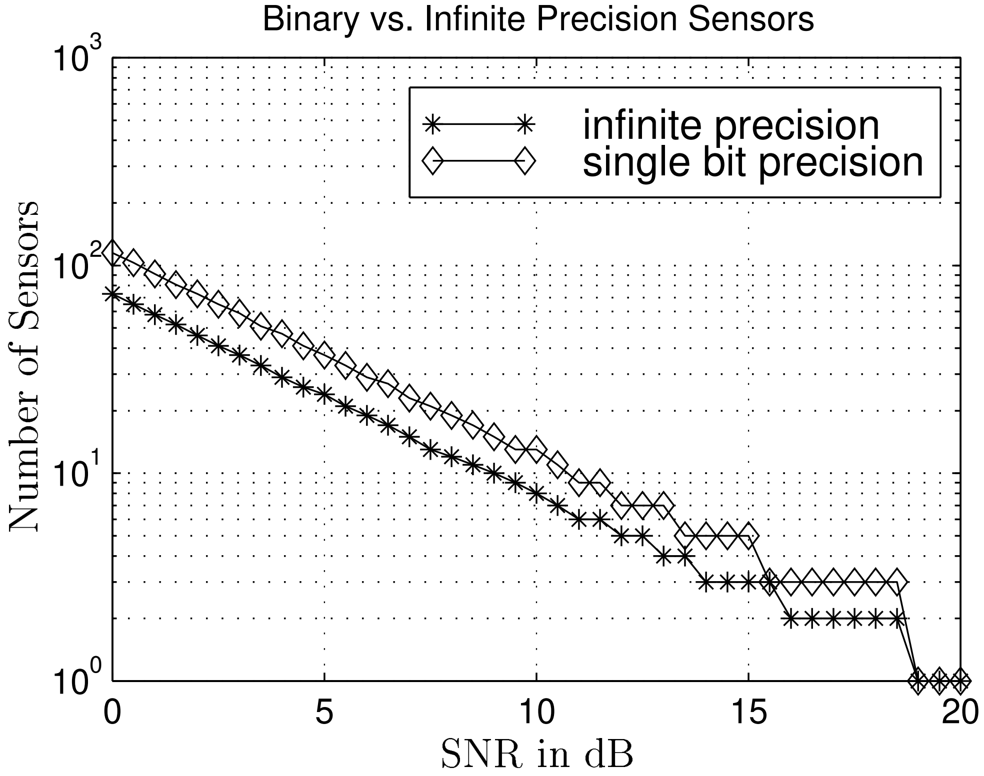

Figure 6.6: The minimum number of sensors needed vs. the SNR to achieve the error probability 10^{-5}.

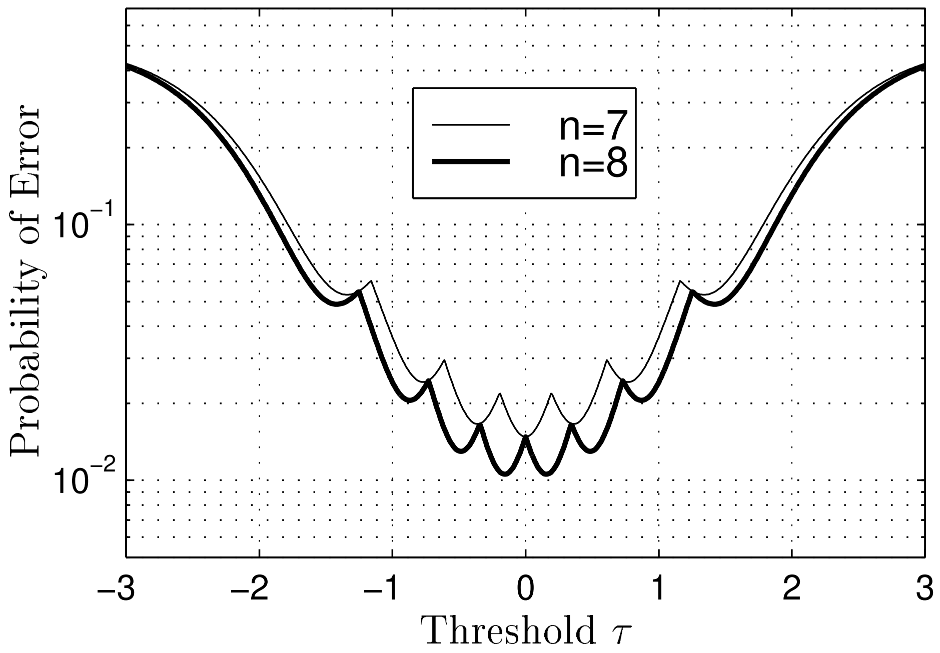

Figure 6.7: The minimum probability of error vs. threshold τ for $n = 7$ and $n = 8$.

Chapter 7

Conclusion

This thesis presented several new results in wireless communications.

Chapter 2 showed that FIR DFE does not require feedback terms for some channels and some specific input SNRs. For the T-spaced case, we derived the conditions that the channel impulse response should satisfy to have the linear equalizer with finite length as the optimal solution. The feedback term is sometimes not useful in the case of relatively small number of feedforward taps. However we proved that, for those channels where the FIR LE is optimal, it is beneficial to process blocks backwards in time using an FIR DFE. Finally, we examined the output SNR versus the input SNR in an example $\frac{T}{2}$-spaced FIR DFE and compared the performance of $\frac{T}{2}$-spaced FIR DFE with that of T-spaced FIR DFE. For any n-tap channel in the T-spaced case, $\frac{T}{2}$-spaced FIR DFE did not have the expected IIR DFE behavior because the subchannels have common zeros.

Chapter 3 examined DFE and THP under mismatch. For the DFE with mismatch, we computed the output SNR under mismatch. For THP, we computed lower bounds on the information rate for the two cases: mismatch at both transmitter and receiver and mismatch at the receiver only. For mismatch at both transmitter and receiver, we also computed a tight upper bound for uniform inputs and a two-tap channel. The output SNR (both biased and unbiased) is bounded even as input SNR increases without bound for DFE with channel mismatch. This is the direct result of residual ISI in the mismatched system. Similarly for THP with mismatch at both transmitter and receiver, residual ISI limits maximum achievable rate to be bounded even as SNR increases without bound. With only the transmitter suffering channel estimation error, a correctly designed feedforward filter allows the information rate to increase without bound as SNR increases. However severe transmitter mismatch can still cause non-negligible loss.

The robustness of space-time turbo-codes was presented in Chapter 4. The BER vs. mutual information (MI) performance is determined by the 2 singular values (SV's). When the 2 SV's are close to each other, the performance is robust in the family of channels with the same SV's. In this case all the channels in the family perform very close to each other and to channel capacity.

When the 2 SV's differ significantly, there is a noticeable gap between the diagonal channel and the worst channel with the most interference between transmit antennas. Also the channels in the family need more additional mutual information to get the desired BER and therefore perform farther from the channel capacity.

By using the time-varying linear transformation technique, we observed a slight performance loss for the best channels and much more performance gain for the bad channels. This time-varying linear transformation has the same effect even for channels with singular values close to each other.

Chapter 5 showed the robustness of space-time turbo codes in terms of proximity to channel capacity for frequency selective fading channels.

Chapter 6 considered distributed detection of $s \in \{-m, m\}$, where the ith of n local sensors observes $x_i = s + z_i$ with i.i.d. additive noise z_i. The ith sensor makes a binary decision u_i based on a threshold τ. A fusion center uses these decisions to produce the global decision using a fusion rule F.

When all admissible rules have the probability of error as a quasiconvex function of τ, the problem decomposes into a series of n quasiconvex optimization problems that may be solved using well-known techniques. We showed this quasiconvexity property for generalized Gaussian noise. For some non-Gaussian noise distributions we showed this quasiconvexity property when the hypotheses have equal *a-priori* probability.

We also used the quasiconvexity perspective to provide solution techniques for Bayes risk and Neyman-Pearson formulations of the sensor data fusion problem.

Zhang et al. [46] generalize these results by showing quasiconvexity in the likelihood ratio function for any distribution on the i.i.d. observations x_i.

Bibliography

[1] A. B. Sesay and M. R. Gibbard. Asymmetric Signal Processing for Indoor Wireless LANs. In *Sixth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 6–10, Toronto, Canada, September 1995.

[2] R. E. Badra and B. Daneshrad. Asymmetric physical layer design for high-speed wireless digital communications. *submitted to IEEE J. Select. Areas Commun.*

[3] R. E. Badra and B. Daneshrad. Performance of a precoded high speed wireless communication system over time-varying indoor channels. In *Proceedings of GLOBECOM*, November 1998.

[4] G. Bauch and A. F. Naguib. MAP Equalization of Space-Time Coded Signals over Frequency Selective Channels. In *IEEE Wireless Communications and Networking Conference(WCNC), New Orleans, LA, USA*, Sept. 1999.

[5] J. M. Cioffi, G. P. Dudevoir, M. Vedat Eyuboğlu, and G. David Forney, Jr. MMSE Decision-Feedback Equalizers and Coding – Parts I and II. *IEEE Trans. on Comm.*, 43(10):2582–2604, October 1995.

[6] J. Salz. Optimum Mean Square Decision Feedback Equalization. *Bell Syst. Tech. J.*, 52(8):1341–1373, October 1973.

[7] C.A. Belfiore and J.H. Park, Jr. Decision feedback equalization. *Proc. IEEE*, 67:1143–1156, August 1979.

[8] A. Duel-Hallen and C. Heegard. Delayed Decision-Feedback Equalization. *IEEE Trans. on Comm.*, 37:428–36, May 1989.

[9] J. M. Cioffi. Signals and Detection. Stanford University Lecture Notes, 1994.

[10] N. Al-Dhahir and J.M. Cioffi. MMSE Decision-Feedback Equalizers: Finite-length results. *IEEE Trans. on Info. Theory*, 41(4):961–975, July 1995.

[11] J.R. Treichler, I. Fijalkow, and C.R. Johnson, JR. Fractionally spaced equalizers: how long should they be? *IEEE Signal Processing Magazine*, pages 65–81, May 1996.

[12] J. Mai and A.H. Sayed. A feedback approach to the steady-state performance of fractionally-spaced adaptive blind equalizers. *submitted to IEEE Transactions on Signal Processing*, November 1998.

[13] M. Tomlinson. New Automatic Equalizer employing Modulo Arithmetic. *Electronic Letters*, 7:138–139, March 1971.

[14] H. Miyakawa and H. Harashima. Information Transmission Rate in Matched Transmission Systems with Peak Transmitting Power Limitation. In *National Conference Record, Institute of Electronics, Information, and Communication Engineers of Japan*, page 1268, August 1969. Number 1269.

[15] H. Harashima and H. Miyakawa. Matched-Transmission Technique for Channels with Intersymbol Interference. *IEEE Trans. on Comm.*, 20(4):774–780, August 1972.

[16] J. E. Mazo and J. Salz. On the Transmitted Power in Generalized Partial Response. *IEEE Trans. on Comm.*, 24(3):348–352, March 1976.

[17] R. D. Wesel and J. M. Cioffi. Achievable Rates for Tomlinson Harashima Precoding. *IEEE Trans. on Info. Theory*, 44(2):824–831, March 1998.

[18] S. Shamai and A. Dembo. Bounds on the Symmetric Binary Cutoff Rate for Dispersive Gaussian Channels. *IEEE Trans. on Comm.*, 42(1):39–53, January 1994.

[19] G. J. Pottie and M. V. Eyuboğlu. Combined Coding and Precoding for PAM and QAM HDSL Systems. *IEEE J. Select. Areas Commun.*, 9(6):861–870, August 1991.

[20] G. J. Foschini and M. J. Gans. On limits of Wireless Communications in a Fading Environment When Using Multiple Antennas. *Wireless Personal Communications*, 6:311–335, March 1998.

[21] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time Codes for High Data Rate Wireless Communications: Performance Criterion and Code Construction. *IEEE Trans. on Info. Theory*, 44(2):744–765, March 1998.

[22] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time Block Codes for High data Rate Wireless Communications: Performance Results. *IEEE J. Select. Areas Commun.*, pages 451–460, March 1999.

[23] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes (1). In *ICC '93*, pages 1064–1070, 1993.

[24] Y. Liu, M. P. Fitz, and O. Y. Takeshita. QPSK Space-Time Turbo codes. In *ICC '00*, June 2000.

[25] A. Stefanov and T. M. Duman. Turbo Coded Modulation for Wireless Communications with Antenna Diversity. In *VTC-Fall*, 1999.

[26] A. Stefanov and T. M. Duman. Turbo Coded Modulation for Systems with Transmit and Receive Antenna Diversity. In *Globecom '99*, Dec. 1999.

[27] G. J. Foschini. Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas. *Bell Labs Technical Journal*, pages 41–59, Autumn 1996.

[28] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank. Combined Array Processing and Space-Time Coding. *IEEE Trans. on Info. Theory*, 45(4):1121–1128, May 1999.

[29] C. Fragouli and R. D. Wesel. Turbo Encoder Design for Symbol Interleaved Parallel Concatenated Trellis Coded Modulation. *To appear in IEEE Trans. on Comm.*

[30] C. Fragouli and R. D. Wesel. Symbol Interleaved Parallel Concatenated Trellis Coded Modulation. In *ICC '99*, pages 42–46, 1999.

[31] S. Benedetto , D. Divsalar, G. Montorsi, and F. Pollara. Soft-Input Soft-Output Modules for the Construction and Distributed Iterative Decoding of

Code Networks. *European Transactions on Telecommunications*, 9(2):155–172, March-April 1998.

[32] R. D. Wesel, X. Liu, and W. Shi. Trellis Codes for Periodic Erasures. *IEEE Trans. on Comm.*, 48(6):938–947, June 2000.

[33] W.A.Hashlamoun and P.K.Varshney. Near-optimum quantization for signal detection. *IEEE Trans. on Comm.*, 44(3):294–297, March 1996.

[34] W.W.Irving and J.N.Tsitsiklis. Some properties of optimal thresholds in decentralized detection. *IEEE Trans. on Automatic Control*, 39(4):835–838, April 1994.

[35] J.N.Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals, and Systems*, 1:167–182, 1988.

[36] P.Chen and A.Papamarcou. New asymptotic results in parallel distributed detection. *IEEE Trans. on Info. Theory*, 39(6):1847–1863, Nov. 1993.

[37] W.A.Hashlamoun, P.K.Varshney, and V.N.S.Samarasooriya. A tight upper bound on the Bayesian probability of error. *IEEE Trans. on Pattern Anal. Machine Intell.*, 16(2):220–224, February 1994.

[38] H.Avi-Itzhak and T.Diep. Arbitrarily tight upper and lower bounds on the Bayesian probability of error. *IEEE Trans. on Pattern Anal. Machine Intell.*, 18(1):89–91, January 1996.

[39] A.W.Roberts and D.E.Varberg. *Convex Functions.* Academic Press, 1973.

[40] M.Grötschel, L.Lovász, and A.Schrijver. *Geometric algorithms and combinatorial optimization.* Springer-Verlag, 1988.

[41] Z. Chair and P.K.Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. on Aerospace and Electronic Systems*, AES-23(1):98–101, Jan. 1986.

[42] A.R.Reibman and L.W.Nolte. Optimal detection and performance of distributed sensor systems. *IEEE Trans. on Aerospace and Electronic Systems*, AES-23(1):24–30, January 1987.

[43] M.S.Gockenbach and A.J.Hearsley. Optimal signal sets for non-Gaussian detectors. *SIAM Journal on Optimization*, 9(2):316–326, 1999.

[44] P.S.Rao, D.H.Johnson, and D.D.Becker. Generation and analysis of non-Gaussian Markov time series. *IEEE Trans. on Signal Processing*, 40(4):845–856, 1992.

[45] R.S.Blum. Distributed detection for diversity reception of fading signals in noise. *IEEE Trans. on Info. Theory*, 45(1):158–164, January 1999.

[46] Q.Zhang, P.K.Varshney, and R.D.Wesel. Optimal distributed binary hypothesis testing with independent identical sensors. In *CISS*, Princeton University, March 2000.