Low Complexity Turbo Equalizations and Lower Bounds on Information Rate for Intersymbol Interference Channels

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Seongwook Jeong

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Prof. Jaekyun Moon, Advisor

October, 2011

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my Ph.D adviser, Prof. Jaekyun Moon. His invaluable insight and experience, guidance, encouragement, and support have made these works in the thesis possible.

I express special thanks to Prof. Zhi-Quan (Tom) Luo, Prof. Emad Ebbini, and Prof. John Kiffer from the Electrical and Computer Engineering department and Prof. Arindam Banerjee and Prof. Yongdae Kim from the Computer Science department at the University of Minnesota for serving as the committee members in my preliminary and/or final oral exams.

I am also very grateful to Jihoon Park, Jaewook Lee, Hakim Alhussein, and Daejung Yoon in the Communications and Data Storage (CDS) lab for their helpful and profound discussions on my research.

This thesis work was possible with the support of my family, especially my parents. I always feel their unconditional love and unlimited devotion through my life. I dedicate this thesis to them.

I also deeply thank my wife, Sangyoun, whose Ph.D diploma still has wet ink on it, for her love, endurance, and sacrifice. She made my life happy and rich and gave lots of helpful assistance. I am also very proud of her to give birth to a baby boy, Heechan (Derek), while she prepared her thesis and final exam. Heechan, you are so lucky to have such a strong and lovely mother.

To my parents, wife Sangyoun, and son Heechan

Abstract

In this research, low complexity turbo equalization algorithms are examined as an alternatives to the optimal, but, much more complex, Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. First, the soft-in soft-out (SISO) decision feedback equalizer (DFE) algorithm with the extrinsic information mapping methods that directly take into account the error propagation effects of DFE is presented. We also utilize a pair of DFE operating in opposite directions in turbo equalization setting to remove the effect of intersymbol interference (ISI) at the receiver with new extrinsic information combining strategy that explores error correlation between the two sets of DFE outputs. When this method is combined with the proposed DFE extrinsic information formulation, the resulting "bidirectional" turbo-DFE achieves excellent performance-complexity tradeoffs compared to the turbo equalization based on the BCJR algorithm. Furthermore, a self-iterating soft equalizer (SISE) consisting of a few relatively weak equalizers is shown to provide robust performance in severe ISI channels. Constituent suboptimal equalizers are allowed to exchange soft information based on the method that are designed to suppress significant correlation among their soft outputs. The resulting SISE works well as a stand-alone equalizer or as the equalizer component of a turbo equalization system. The performance advantages of the proposed algorithms are validated with bit-error-rate (BER) simulations and extrinsic information transfer (EXIT) chart analysis.

In the thesis, provable lower bounds are also presented for the information rate of any finite ISI channels. Let us consider $I(X; X + S + N)$ where X is the symbol drawn independently and uniformly from a fixed, finite-size alphabet, S a discrete-valued random variable (RV) and N a Gaussian RV. Especially, when S represents the precursor ISI after the infinite-length unbiased minimum mean-squared error (MMSE) DFE is applied at the channel output, the mutual information $I(X; X + S + N)$ serves as a tight lower bound for the symmetric information rate (SIR) as well as capacity

of the ISI channel corrupted by Gaussian noise. The new lower bounds are obtained by first introducing a "mismatched" mutual information function that can be proved as a lower bound to $I(X; X + S + N)$ and then further lower-bounding this function with expressions that can be computed via a few single-dimensional integrations with a small computational load. The new bounds provide a similar level of tightness as the well-known conjectured lower bound by Shamai and Laroia for a wide variety of ISI channels of practical interest.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Assuming a channel with particular bandwidth and noise characteristics, it has been demonstrated that error free transmission is possible if information is sent at a rate less than the computable maximum information rate of the channel; this theoretical limit on the rate of error-free transmission is expressed in terms of the channel capacity [1]. More specifically, it was shown that for any given transmission rate lower than the channel's capacity, there are codes that can achieve an arbitrarily small probability of error on a noisy channel by adding redundant information to the original message.

The first practical codes to approach the channel capacity for additive white noise Gaussian noise (AWGN) channel, called turbo codes, were introduced in 1993 [2, 3]. The basic idea of this new coding scheme is to use a parallel concatenation of two recursive systematic convolutional (RSC) codes with an interleaver between the encoders [4]. The codes are decoded iteratively by exchanging the soft information between two constituent soft-in soft-out (SISO) decoders. This key concept is expressed as the "turbo principle" because of the similarity between the decoding algorithm and the cyclic feedback mechanism in Turbo engines [5].

Intersymbol interference (ISI), which occurs in high-speed digital communication when transmitted symbols overlap with one another due to channel distortion, can also make communication less reliable if it is not compensated for. In order to combat the detrimental effect of ISI, the turbo principle has also been applied to the communication system [6]. Analogous to the turbo codes, this receiving scheme is referred as the turbo equalization, wherein a SISO equalizer (or detector) and a SISO error-correction decoder exchange their soft information in an iterative fashion until reliable decisions are generated. In this system, the ISI channel can be viewed as a specific non-recursive convolutional encoder and the channel equalizer (or detector) may be regarded as a corresponding convolutional decoder.

The detector or equalizer portion of the frequently investigated turbo equalizer (TE) system is based on the Bahl-Cocke-Jelinek-Raviv (BCJR) [7] and the soft-output Viterbi algorithms (SOVA) [8]. The BCJR algorithm exactly computes the *a posteriori* probability of the transmitted signal symbols considering the channel response and the *a priori* information of the transmitted symbols and, as such, can be viewed as an optimum SISO equalizer. Meanwhile, the SOVA algorithm is a variation of the original Viterbi algorithm; it incorporates the *a priori* information of the input symbols to the path metrics and generates a soft-output reliability measure of the final path decisions. However, the computational complexities of both algorithms grow exponentially as a function of the channel length and the symbol alphabet set size. Therefore, the trellis search equalization approach becomes impractical when the multipath channel is very long or the symbol alphabet set size is very large [10, 11].

Numerous less complex, but suboptimal, turbo equalization methods have been proposed to mitigate the high computational complexity of the BCJR and SOVA methods; many of these suboptimal equalizers are based on filters utilizing the *a priori* information and can be classified as either SISO linear equalizers (LE), SISO decision feedback equalizers (DFE), or SISO soft-feedback equalizers (SFE) which can be viewed as a hybrid of LE and DFE [12–25]. Specifically, SISO LEs adopt the soft extrinsic information

from the decoder as the *a priori* information about the transmitted symbols, whereas SISO DFEs use the hard estimates of previously equalized or decoded causal symbols as their *a priori* information and the soft extrinsic information from the decoder as the *a priori* information for anticausal symbols; some schemes even use the hard estimates for anticausal symbols [21]. Similarly, SISO SFEs employ the soft *a posteriori* information of the previously equalized causal symbols as their *a priori* information, and the soft information from the decoder as the *a priori* information for anticausal symbols. Regardless of the classification, the exact minimum mean-squared-error (MMSE) filters all have a time-varying structure wherein each tap must be recalculated at each symbol interval, and the resulting complexity increases at least as a quadratic function of the filter length, which creates significant implementation challenges [13]. Furthermore, even with the time-varying MMSE filters, these equalizers are still individually weak when ISI is severe [24]. Thus, the primary objective of this research is the development of an effective alternative equalization structure for communication systems.

Also in this dissertation, we are concerned with provable lower bounds for the maximum information rate in the communication systems that can be easily computed. The computation of the symmetric information rate (SIR) of the classical discrete-time intersymbol interference (ISI) channel is of great interest in digital communication. The SIR represents the mutual information between the channel input and output while the input is constrained to be independently and uniformly distributed (i.u.d.) over the given alphabet. In this sense, the SIR is also known as capacity with uniform, independent input distribution and itself represents a reasonably tight lower bound to unconstrained channel capacity, especially at high coding rates. During recent years, a number of researchers have worked on estimating or bounding the information rate via simulation of the BCJR algorithm [7]. The information rate with a given input distribution can be closely estimated for finite ISI channels with moderate input alphabet size and channel impulse response length, by running the forward-recursion portion of the BCJR algorithm on long (pseudo) randomly generated input and noise samples [26–28].

The simulation-based method has been further generalized, and lower and upper bounds based on auxiliary finite-state channels with reduced states were introduced for long ISI channels, as well as some non-finite state ISI channels in [29]. The tightness of these bounds is highly related to the optimality of auxiliary channels, but the general rule to find the optimal or near-optimal auxiliary channel has not been provided in [29]. The work of [29] has been recently extended in [30] to further tighten the lower and upper bounds by using an iterative expectation-maximization type algorithm to optimize the parameters of the auxiliary finite-state channels. It is noted, however, that the global optimality of the bounds in [30] is neither guaranteed, nor the lower bound is proven to converge to a stationary point as iteration progresses. Another approach based on auxiliary channels is also proposed to obtain a lower bound utilizing a mismatched Ungerboeck-type channel response to achieve improved tightness for a given level of computational complexity [31]. In the context of [31], the Ungerboeck-type response is the channel's response observed at the output of the matched filter front-end. As such, the trellis search detection algorithms driven by the channel observations of the Ungerboeck model must be designed so that they can handle correlated noise samples [32].

An entirely different direction in estimating or bounding the information rate is based on finding an analytical expression that can easily be evaluated or numerically computed (in contrast to the methods based on Monte-Carlo simulation that rely on generating pseudo-random signal and noise samples). An early work in this direction is the lower bound on the SIR by Hirt [33] based on carving a fixed block out of the channel input/output sequences and performing a single multi-dimensional integration (or running Monte-Carlo simulation for estimating the integral) with the dimensionality equal to the block size. However, this method is also computationally intense unless the size of the block gets small. Unfortunately the lower bound of [33] is not tight unless the block size is very large compared to the channel ISI length.

A number of more computationally efficient and analytically evaluated lower bounds

for the SIR have been discussed in [34, 35]. Unfortunately, however, the only bound presented in [35] that is reasonably tight throughout the entire signal-to-noise ratio (SNR) region (i.e., both low and high code rate regimes) is the one that could not be proved as a lower bound. This particular bound is now widely known as the Shamai-Laroia conjecture (SLC) and, although unproven, is a popular tool for quickly estimating the SIR of ISI channels. At high code rates, the SIR is generally very close to capacity, so an easily computed tight SIR lower bound is also useful for quickly estimating channel capacity for high code rate applications, such as data storage channels and optical fiber channels.

Consider the random variable (RV) $Y = X+S+N$, where X is a symbol drawn independently and uniformly from a fixed, finite-size alphabet set symmetrically positioned around the origin, S a zero-mean discrete-valued RV, and N a zero-mean Gaussian RV. It is well known that $I(X; X + S + N)$ lower-bounds the SIR if $S + N$ represents the noise component of unbiased minimum mean-squared error decision feedback equalizer (MMSE-DFE) filter output after ideal postcursor ISI cancellation [35]. The SLC is concerned with the special case where S is a linear sum of symbols drawn independently and uniformly from the same symbol set where X was taken. As the number of symbols forming S grows, finding an analytical expression for the probability density function of $S + N$ (and thus one for $I(X; Y)$) is a long-standing problem [37, 38], as pointed out in [35]. The SLC of [35] can be stated as $I(X; X + S + N) \geq I(X; X + G)$, where G is a Gaussian RV with variance matching that of $S + N$. The information rate $I(X; X + G)$ is easily obtained by numerically calculating a single one-dimensional integral, and is generally observed to be reasonably tight to $I(X; X + S + N)$ in most cases. Unfortunately, $I(X; X + G)$ remains as a conjectured bound with no proof available to date. One difficulty of proving the SLC stems from the fact that for the channels driven by the inputs from a finite alphabet, Gaussian noise is not the worst-case noise in terms of the achievable information rate [35, 36]. Another difficulty is that the power contribution of a single individual weight involved in constructing S could remain a significant portion

of the total power associated with all weights, even if the number of weights approaches infinity. This is to say that the Lindberg condition for the central limit theorem does not hold for this problem, and the Gaussian approximation of S cannot be justified [35]. In this thesis, we also provide the easily computable analytical expressions for lower bounds to the SIR.

1.2 Problem Statement

We assume that the receiver knows the discrete-time baseband channel response accurately. While the methods in this thesis discussed are general, our presentation will be mainly based on binary symbols with $P_X \triangleq \mathrm{E}(x_n^2) = 1$, $x_n \in \{\pm 1\}$, as well as real-valued ISI channel coefficients and noise samples. Although x_n typically represents a coded bit sequence, our analysis will assume that it is equiprobable and independent and identically distributed (i.i.d.). Given the transmitted bit sequence $\{x_k\}$, the channel output at time n is

$$r_n = \sum_{k=0}^{L_h-1} h_k x_{n-k} + w_n \tag{1.1}$$

$$= h_0 x_n + w_n + \underbrace{\sum_{k=1}^{L_h-1} h_k x_{n-k}}_{\text{ISI}} \tag{1.2}$$

where w_n is AWGN with variance N_0 and $\{h_k\}$ is the energy-normalized channel impulse response with length L_h.

As shown in the equation (1.2), ISI can cause the wrong detection of x_n at receiver. Thus, our first question is how to suppress the ISI terms effectively and efficiently in order to estimate x_n with the given channel output sequence r_n of (1.1). The second question is how to quickly estimate the channel capacity (or SIR since the input is assumed to be i.u.d.) for the discrete ISI channel model. The channel capacity C (bits/channel use) for any finite-ISI channel corrupted by Gaussian noise is given as $C \triangleq \lim_{N\to\infty} \frac{1}{2N+1} I\left(\{x_k\}_{-N}^{N}; \{r_k\}_{-N}^{N}\right)$ where $\{u_k\}_{N_1}^{N_2} = \{u_k, k = N_1, N_1+1, \ldots, N_2\}$ [68].

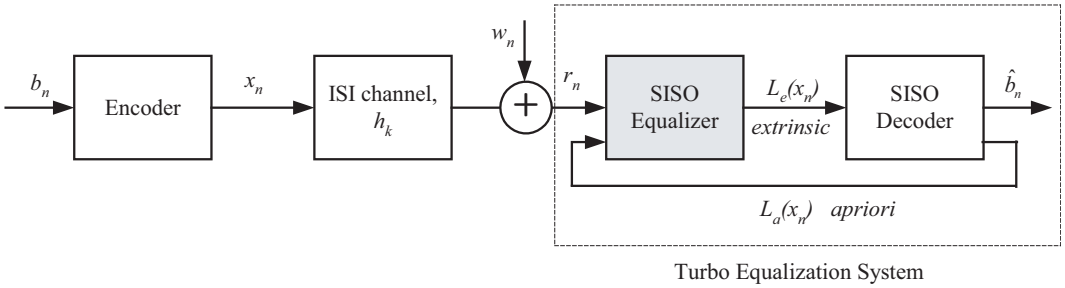


Figure 1.1: Turbo Equalization System.

The first question will be answered in Chapter 2, 3, and 4 while the second question will be answered in Chapter 5.

1.3 Turbo Equalization

Turbo equalization is a well-established technique that is highly effective in combating ISI via iterative exchange of soft decisions between a SISO equalizer and a SISO error-correction decoder separated by an interleaver [6]. The original structure of turbo equalization is described in Fig. 1.1.

In turbo equalization, the equalizer computes the *a posteriori* log-likelihood ratio (LLR) of x_n,

$$L(x_n) \triangleq \ln \frac{\Pr(x_n = +1 \mid \mathbf{r})}{\Pr(x_n = -1 \mid \mathbf{r})}$$

where $\mathbf{r}$ is the entire or partial received sample block utilized for LLR estimation for x_n. Note that this computation requires the knowledge of the *a priori* probabilities of all input bits affecting $\mathbf{r}$. Since these *a priori* probabilities are not available, they are all set to 1/2 initially and then, as the turbo iteration ensues, to the estimated probability values based on the extrinsic information generated and passed back by the outer decoder.

The equalizer then generates its own extrinsic information by subtracting the effect of the probability estimate passed down for the current bit. Write this estimated *a*

priori LLR passed down from the decoder as

$$L_a(x_n) \triangleq \ln \frac{\Pr(x_n = +1)}{\Pr(x_n = -1)}$$

with an understanding that the probabilities in the expression are in reality just estimates.

Then, the equalizer's extrinsic LLR for x_n to be passed to the error-correction code decoder is given by

$$L_e(x_n) \triangleq L(x_n) - L_a(x_n).$$

This equation suggests first computing $L(x_n)$ based on the *a priori* probabilities of all input bits including x_n and then simply subtracting $L_a(x_n)$ to generate the extrinsic LLR $L_e(x_n)$. An alternative way of generating $L_e(x_n)$ is to set $L_a(x_n) = 0$ while computing $L(x_n)$, i.e., suppress the effect of $L_a(x_n)$ in the calculation of $L(x_n)$:

$$L_e(x_n) = L(x_n)|_{L_a(x_n)=0}.$$

The techniques discussed in this dissertation actually use the second method.

1.4 Trellis based Detection

In this section, we briefly review the BCJR algorithm of [7] and SOVA algorithm of [8]. Both algorithms are trellis-based, in the sense that they make use of a trellis diagram showing all possible state transitions at each time step, induced by a finite-state Markov channel model.

1.4.1 BCJR Algorithm

Given the entire received sequence, $\mathbf{r} = \{r_0, r_1, \ldots, r_{N-1}\}$, BCJR algorithm exactly computes the APPs of the transmitted symbols in the presence of the *a priori* information of the transmitted symbols by considering the channel response and the noise

variance. This algorithm is also referred as the maximum *a posteriori* (MAP) symbol detection algorithm since it estimates x_n maximizing the APP of a symbol, $\Pr(x_n \mid \mathbf{r})$.

Let us denote s', s, and s'' as the past, current, and future states of the Markov process respectively. Since $\Pr(x_n \mid \mathbf{r}) = \Pr(x_n, \mathbf{r})/\Pr(\mathbf{r})$, $\Pr(x_n, \mathbf{r})$ is computed instead of $\Pr(x_n \mid \mathbf{r})$. Then,

$$\Pr(x_n, \mathbf{r}) = \sum_{s'} \alpha_{n-1}(s') \cdot \gamma_n(s', s) \cdot \beta_n(s) \tag{1.3}$$

where $\alpha_{n-1}(s') \triangleq \Pr(S_{n-1} = s', \{r_k\}_0^{n-1})$, $\gamma_n(s', s) \triangleq \Pr(r_n \mid S_{n-1} = s', S_n = s)\Pr(x_n)$, and $\beta_n(s) \triangleq \Pr(\{r_k\}_{n+1}^{N-1} \mid S_n = s)$. Since the noise is assumed to be AWGN, $\gamma_n(s', s)$ is simply given as $(2\pi N_0)^{-1/2} \exp(-\frac{(r_n - o_n)^2}{2N_0})\Pr(x_n)$ where o_n is the noiseless channel output corresponding to the state transition from s' to s and $\Pr(x_n)$ is the estimated *a priori* probability. Moreover, the metric $\alpha_n(s)$ and $\beta_n(s)$ can be computed recursively [7]:

$$\alpha_n(s) = \sum_{s'} \alpha_{n-1}(s') \cdot \gamma_n(s', s) \tag{1.4}$$

$$\beta_n(s) = \sum_{s''} \beta_{n+1}(s'') \cdot \gamma_{n+1}(s, s'') \tag{1.5}$$

These equations represent the forward and backward recursions of BCJR algorithm, respectively. Finally, the *a posteriori* LLR of x_n is given as

$$\begin{aligned} L(x_n) &\triangleq \ln \frac{\Pr(x_n = +1 \mid \mathbf{r})}{\Pr(x_n = -1 \mid \mathbf{r})} \\ &= \ln \frac{\Pr(x_n = +1, \mathbf{r})}{\Pr(x_n = -1, \mathbf{r})} \\ &= \ln \frac{\sum_{s', x_n = +1} \alpha_{n-1}(s') \cdot \gamma_n(s', s) \cdot \beta_n(s)}{\sum_{s', x_n = -1} \alpha_{n-1}(s') \cdot \gamma_n(s', s) \cdot \beta_n(s)} \end{aligned} \tag{1.6}$$

Note that the above computations in probability domain can be also implemented in log domain [39]. Furthermore, the approximated version of BCJR algorithm and its improvement have been also discussed in [40–42].

1.4.2 SOVA Algorithm

The classic Viterbi algorithm generates the hard decision of the codeword minimizing the log-likelihood function [9]. It has been, therefore, modified to produce the soft output by Hagenaur et. al [8]. The SOVA computes the probability of making a wrong decision when the survivor paths on the trellis is chosen. Specifically, at time n, the path metrics of a survivor path and its competing path, denoted by M_n^s and M_n^c respectively, are defined as

$$M_n^s = \frac{1}{2N_0}\sum_{k=0}^{n}(r_k - o_k^s)^2, \quad M_n^c = \frac{1}{2N_0}\sum_{k=0}^{n}(r_k - o_k^c)^2$$

where o_k^s and o_k^c are noiseless channel output corresponding to the survivor and competing data path. Notice that $M_n^s \leq M_n^c$. Then, the probability of selecting this path is proportional to

$$\Pr(\text{survivor path}) \sim e^{-M_n^s}, \quad \Pr(\text{competing path}) \sim e^{-M_n^c}$$

Therefore, the probability of selecting the wrong survivor path at time n is

$$P_n^w = \frac{e^{-M_n^c}}{e^{-M_n^s} + e^{-M_n^c}} = \frac{1}{1 + e^{\Delta_n}}$$

where $\Delta_n = M_n^s - M_n^c$.

Now, let us assume that Viterbi algorithm has made errors in e different positions where the decisions of the competing path is different from the decisions of the survivor path at time n. Then, denoting the probability of the previous erroneous decisions on the survivor path as P_j, we can update these probabilities for the e different decisions according to

$$P_j \leftarrow P_j(1 - P_n^w) + (1 - P_j)P_n^w, \qquad j = j_1, j_2, \ldots, j_e.$$

Finally, the SOVA algorithm produce the soft-output for x_n as

$$\begin{aligned} L(x_n) &\triangleq \ln \frac{\Pr(x_n = +1 \mid \mathbf{r})}{\Pr(x_n = -1 \mid \mathbf{r})} \\ &= \hat{x}_n \log\left(\frac{1 - P_n}{P_n}\right). \end{aligned} \tag{1.7}$$

where $\hat{x}_n$ is the estimated hard decision of Viterbi algorithm. Some researchers have also worked to improve the performance of SOVA [43–48].

1.5 Thesis Outline

- Chapter 2 briefly reviews the previously developed filter based turbo equalization algorithms, such as SISO LE and DFE, and then proposes the enhanced DFE based turbo equalization with the modified LLR mapping method. Additionally, the MMSE filters utilizing *a priori* information are summarized and SNR/MI analyses for SISO LE and DFE are provided.

- In Chapter 3, the turbo equalization based on bidirectional DFE (BiDFE) is presented with its SNR figure-of-merit, and the noise correlation coefficient between two DFEs is analyzed under the ideal decision feedback assumption.

- Chapter 4 describes the self-iterating soft equalizers (SISE) which can be employed in turbo equalization systems; more specifically, a proper way to extract the extrinsic information from other constituent equalizers when the information between the equalizers is highly correlated is shown.

- Chapter 5 investigates the analytical and provable lower bounds on the SIR or capacity of the ISI channel corrupted by Gaussian noise. In all finite-ISI channels examined, the proposed bound provides the same level of tightness as the SLC to the SIR with a very reasonable computation load. Moreover, the analytical method is much better in quickly producing a tight bound than the simulation-based method in terms of complexity/accuracy tradeoffs.

- Chapter 6 draws final conclusions and suggests future research.

Chapter 2

SISO LE and DFE

Given the channel observation sequence, what is the best estimate for the transmitted symbols in the sense of minimizing the mean-squared-error (MSE)? This question can be answered with specific estimator structures in mind such as a single finite-impulse response (FIR) filter or a combination of a FIR forward filter with a FIR feedback filter. The former leads to the classical MMSE-LE whereas the latter corresponds to the classical MMSE-DFE of [49].

Naturally, the MMSE solutions depend on the second order statistics involving all underlying random variables - namely, the noise samples and the channel input symbols. In the classical MMSE equalizer design, the assumed mean and variance of the input symbols simply reflect those of the independent, equally-likely symbols since there exists no *a priori* information about these symbols. In turbo equalization setting, however, estimates of the *a priori* symbol probability do become available via the soft decisions made by the outer decoder. As shown by the authors of [14], the classical MMSE equalizer design can be modified by incorporating the mean and variance of the symbols estimated via the extrinsic symbol information generated by the decoder. The resulting structure is a time-varying filter that needs to recalculate the filter taps at every symbol

interval. Unfortunately, this time-varying structure also creates significant implementation challenges because the complexity of the filter tap computation increases at least as a quadratic function of the total filter length.

In the classical, non-turbo setting (i.e., no iterative exchange of soft information between the equalizer and the decoder), it has long been known that the DFE almost always outperforms the LE, despite the fact that the DFE typically suffers from error propagation. This is because when ISI is severe with the channel response showing nulls or deep valleys within the Nyquist band, the LE is subject to large noise enhancement. The work of [14], however, shows that when hard decisions are fed through the feedback filter (to reduce complexity), the SISO DFE performs considerably worse than the SISO LE, presumably due to error propagation.

In this chapter we first briefly review the results of [14] related to the SISO LE and DFE to provide necessary background while establishing notations. We then readdresses the DFE design issue in the turbo equalizer environment and shows that just as in classical non-turbo setting, the DFE outperforms the LE, if extrinsic information is reformulated in a way that combats error propagation more effectively. We also review MMSE filter design methods that utilizes the reliability of the *a priori* information and provide the SNR/MI analyses for previously developed filter based equalizers. This chapter is based on [24, 72].

2.1 SISO Linear Equalizer

The work of [14] has established an effective strategy of utilizing the *a priori* information estimates from the outer decoder in calculating the equalizer tap coefficients. The gist of the approach in [14] is a clever tweaking of the classical MMSE estimation principle where the "mean" of the input symbols are constructed using the available *a priori* information estimates and utilized in the linear estimator weight computation.

Based on the above principle and suppressing the effect of the *a priori* probability

estimate on the current bit x_n (i.e., $\mathrm{E}(x_n) = 0$) in an effort to extract the extrinsic information, the MMSE filter taps (causal filter length L_{f1}, anticausal filter length L_{f2}, and a total length $L_{f1} + L_{f2} + 1$) at time n are derived respectively as:

$$\begin{aligned}\mathbf{f}_n &\triangleq \left[f_{\{n,-L_{f1}\}}, f_{\{n,-L_{f1}+1\}}, \ldots, f_{\{n,L_{f2}\}}\right]^T \\ &= \left\{\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T + (1-z_n)\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \end{aligned} \tag{2.1}$$

where $\mathbf{H}$ is a channel convolution matrix defined as

$$\mathbf{H} \triangleq \begin{bmatrix} h_{L_h-1} & h_{L_h-2} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_{L_h-1} & h_{L_h-2} & \cdots & h_0 & 0 & \cdots \\ & \ddots & & \ddots & & \ddots & \\ 0 & 0 & \cdots & 0 & h_{L_h-1} & h_{L_h-2} & \cdots \end{bmatrix},$$

and $\boldsymbol{\Sigma}_n$ depends on $\mathrm{E}(x_i)$, $i = n - L_{f1}, \ldots, n + L_{f2}$, computed from the decoder output as $\mathrm{E}(x_i) = \tanh(L_a(x_i)/2)$. Specifically, $\boldsymbol{\Sigma}_n \triangleq \mathrm{Diag}(z_{n-L_{f1}}, z_{n-L_{f1}+1}, \ldots, z_{n+L_{f2}})$ with $z_i \triangleq 1 - [\mathrm{E}(x_i)]^2$. Adding the term $(1 - z_n)\mathbf{s}\mathbf{s}^T$ in (2.1) has the same effect of suppressing $\mathrm{E}(x_n)$ to zero in $\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T$. The cross correlation vector is defined as $\mathbf{s} \triangleq \mathbf{H}[\mathbf{0}_{1\times L_{f1}}, 1, \mathbf{0}_{1\times L_{f2}}]^T$.

The equalizer output is obtained as

$$y_n = \mathbf{f}_n^T \cdot (\mathbf{r}_n - \mathbf{H}\bar{\mathbf{x}}_n + \mathrm{E}(x_n)\mathbf{s}) \tag{2.2}$$

where the received vector is defined as $\mathbf{r}_n \triangleq \left[r_{n-L_{f1}}, r_{n-L_{f1}+1}, \ldots, r_{n+L_{f2}}\right]^T$ and the symbols' mean as $\bar{\mathbf{x}}_n \triangleq \left[\mathrm{E}(x_{n-L_{f1}}), \mathrm{E}(x_{n-L_{f1}+1}), \ldots, \mathrm{E}(x_{n+L_{f2}})\right]^T$. The addition of the $\mathrm{E}(x_n)\mathbf{s}$ term is to suppress the effect of $\mathrm{E}(x_n)$ in $\mathbf{H}\bar{\mathbf{x}}_n$.

Define the symbol sequence $\mathbf{x}_n \triangleq \left[x_{n-L_{f1}}, \ldots, x_{n+L_{f2}}\right]^T$ and the noise sequence as $\mathbf{w}_n \triangleq \left[w_{n-L_{f1}}, \ldots, w_{n+L_{f2}}\right]^T$. Then, the linear filter output y_n can be rewritten as

$$\begin{aligned} y_n &= (\mathbf{f}_n^T\mathbf{H}) \cdot \left(\mathbf{x}_n - \mathrm{E}\{\dot{\mathbf{x}}_n\}\right) + \mathbf{f}_n^T\mathbf{w}_n \\ &= p_{\{n,0\}}x_n + \sum_{k=-L_{f1},k\neq 0}^{L_{f2}} p_{\{n,k\}}\left(x_{n+k} - \mathrm{E}(x_{n+k})\right) + \sum_{k=-L_{f1}}^{L_{f2}} f_{\{n,k\}}w_{n+k} \\ &= p_{\{n,0\}}x_n + v_n \end{aligned} \tag{2.3}$$

where $\mathrm{E}\{\dot{\mathbf{x}}_n\} \triangleq \left[\mathrm{E}(x_{n-L_{f1}}), \ldots, \mathrm{E}(x_{n-1}), 0, \mathrm{E}(x_{n+1}), \ldots, \mathrm{E}(x_{n+L_{f2}})\right]^T$. Moreover, $\mathbf{p}_n \triangleq \left[p_{\{n,-L_{f1}\}}, \ldots, p_{\{n,L_{f2}\}}\right] = \mathbf{f}_n^T\mathbf{H}$ and $p_{\{n,0\}} = \mathbf{f}_n^T\mathbf{s}$ especially. The sum of the noise and the remaining ISI terms caused by the neighboring symbols is denoted by v_n: $v_n \triangleq \sum_{k=-L_{f1},k\neq 0}^{L_{f2}} p_{\{n,k\}}\big(x_{n+k} - \mathrm{E}(x_{n+k})\big) + \sum_{k=-L_{f1}}^{L_{f2}} f_{\{n,k\}} w_{n+k}$. The variance of v_n is

$$\begin{aligned}
\mathrm{Var}(v_n) &\triangleq \mathbf{f}_n^T \mathrm{Cov}\{\mathbf{r}_n\mathbf{r}_n^T \mid x_n = x\}\mathbf{f}_n \\
&= \mathbf{f}_n^T \left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}_n \qquad (2.4)\\
&= \mathbf{f}_n^T \left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T + (1-z_n)\mathbf{s}\mathbf{s}^T + N_0\mathbf{I} - \mathbf{s}\mathbf{s}^T\right)\mathbf{f}_n \\
&= \mathbf{f}_n^T \left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T + (1-z_n)\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}_n - \mathbf{f}_n^T\mathbf{s}\mathbf{s}^T\mathbf{f}_n \\
&= \mathbf{f}_n^T\mathbf{s}(1 - \mathbf{s}^T\mathbf{f}_n). \qquad (2.5)
\end{aligned}$$

The equality of (2.4) holds because the variance of the current symbol, z_n, should not be counted for its noise variance and the last equality comes from the definition of $\mathbf{f}_n$ in (2.1).

Then, assuming that v_n is AWGN, the extrinsic LLR is naturally given by

$$\begin{aligned}
L_e(x_n) &\triangleq \ln \frac{\Pr(x_n = +1 \mid y_n)}{\Pr(x_n = -1 \mid y_n)}\bigg|_{L_a(x_n)=0} \\
&= \ln \frac{\Pr(x_n = +1, y_n)}{\Pr(x_n = -1, y_n)}\bigg|_{L_a(x_n)=0} \\
&= \left\{\ln \frac{\Pr(y_n \mid x_n = +1)}{\Pr(y_n \mid x_n = -1)} + \ln \frac{\Pr(x_n = +1)}{\Pr(x_n = -1)}\right\}\bigg|_{L_a(x_n)=0} \\
&= \ln \frac{\Pr(y_n \mid x_n = +1)}{\Pr(y_n \mid x_n = -1)} \qquad (2.6)\\
&= -\frac{\left|y_n - p_{\{n,0\}}\right|^2}{2\mathrm{Var}(v_n)} + \frac{\left|y_n + p_{\{n,0\}}\right|^2}{2\mathrm{Var}(v_n)} \\
&= \frac{2p_{\{n,0\}}y_n}{\mathrm{Var}(v_n)}. \qquad (2.7)
\end{aligned}$$

The equality of (2.6) holds because $L_a(x_n)$ was already suppressed to zero in generating y_n.

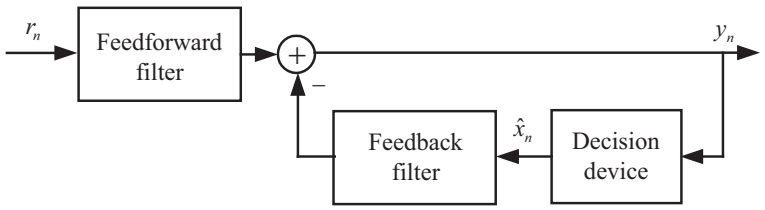


Figure 2.1: Decision Feedback Equalizer (DFE).

2.2 SISO Decision Feedback Equalizer

The structure of the classic DFE is shown in Fig. 2.1. The DFE consists of a linear feedforward filter, a linear feedback filter, and a decision device. The feedforward filter suppresses the pre-cursor ISI, on the other hand, the feedback filter cancels out the post-cursor ISI. The results of two filters are added and then applied to a decision device to determine the symbol of interest.

Like in the case of LE, the MMSE feedforward filter taps (a total of L_c+1) and the feedback filter taps (a total of $L_d = L_h - 1$) at time n are derived respectively as:

$$\begin{aligned}\mathbf{c}_n &\triangleq \left[c_{\{n,0\}}, c_{\{n,+1\}}, \ldots, c_{\{n,L_c\}}\right]^T \\ &= \left\{\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T + (1-z_n)\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \qquad (2.8)\\ \mathbf{d}_n &\triangleq \left[d_{\{n,-L_d\}}, d_{\{n,-L_d+1\}}, \ldots, d_{\{n,-1\}}\right]^T \\ &= \mathbf{M}\mathbf{H}^T\mathbf{c}_n \qquad (2.9)\end{aligned}$$

where $\boldsymbol{\Sigma}_n \triangleq \mathrm{Diag}(\mathbf{0}_{1\times L_d}, z_n, z_{n+1}, \ldots, z_{n+L_c})$ with $z_i \triangleq 1 - [\mathrm{E}(x_i)]^2$. The remaining vector and matrix are defined as $\mathbf{s} \triangleq \mathbf{H}[\mathbf{0}_{1\times L_d}, 1, \mathbf{0}_{1\times L_c}]^T$ and $\mathbf{M} \triangleq [\mathbf{I}_{L_d\times L_d}, \mathbf{0}_{L_d\times(L_c+1)}]$.

The equalizer output is obtained as

$$y_n = \mathbf{c}_n^T \cdot (\mathbf{r}_n - \mathbf{H}\bar{\mathbf{x}}_n + \mathrm{E}(x_n)\mathbf{s}) \qquad (2.10)$$

where the received vector is defined as $\mathbf{r}_n \triangleq [r_n, r_{n+1}, \ldots, r_{n+L_c}]^T$ and the composite vector of the causal symbol decisions and the anticausal symbols' mean as $\bar{\mathbf{x}}_n \triangleq [\hat{x}_{n-L_d}, \ldots, \hat{x}_{n-1}, \mathrm{E}(x_n), \ldots, \mathrm{E}(x_{n+L_c})]^T$ where $\hat{x}_i$ is the available decision for x_i based

on the *a posteriori* LLR of x_i, i.e., if $L(x_i) = L_a(x_i) + L_e(x_i) \geq 0$, then, $\hat{x}_i = +1$; otherwise, $\hat{x}_i = -1$.

Define the anticausal symbol sequence $\mathbf{x}_n \triangleq [x_n, x_{n+1}, \ldots, x_{n+L_c}]^T$, the causal symbol sequence $\mathbf{x}_n^c \triangleq [x_{n-L_d}, x_{n-L_d+1}, \ldots, x_{n-1}]^T$, the available decision sequence $\hat{\mathbf{x}}_n^c \triangleq [\hat{x}_{n-L_d}, \hat{x}_{n-L_d+1}, \ldots, \hat{x}_{n-1}]^T$, and the noise sequence $\mathbf{w}_n \triangleq [w_n, w_{n+1}, \ldots, w_{n+L_c}]^T$. Then, the combined filter output y_n can be rewritten as

$$\begin{aligned} y_n &= (\mathbf{c}_n^T \mathbf{H}_1) \cdot \big(\mathbf{x}_n - \mathrm{E}\{\dot{\mathbf{x}}_n\}\big) + \mathbf{d}_n^T (\mathbf{x}_n^c - \hat{\mathbf{x}}_n^c) + \mathbf{c}_n^T \mathbf{w}_n \\ &= p_{\{n,0\}} x_n + \sum_{k=1}^{L_d} d_{\{n,-k\}} \big(x_{n-k} - \hat{x}_{n-k}\big) + \sum_{k=1}^{L_c} p_{\{n,k\}} \big(x_{n+k} - \mathrm{E}(x_{n+k})\big) + \sum_{k=0}^{L_c} c_{\{n,k\}} w_{n+k} \\ &= p_{\{n,0\}} x_n + i_n + v_n \end{aligned} \tag{2.11}$$

where $\mathrm{E}\{\dot{\mathbf{x}}_n\} \triangleq [0, \mathrm{E}(x_{n+1}), \mathrm{E}(x_{n+2}), \ldots, \mathrm{E}(x_{n+L_c})]^T$ and $\mathbf{H}_1$ is the $(L_c+1) \times (L_c+1)$ submatrix of $\mathbf{H}$ formed by the entire rows of the columns from the (L_d+1)th to the last. Moreover, $\mathbf{p}_n \triangleq \big[p_{\{n,0\}}, p_{\{n,1\}}, \ldots, p_{\{n,L_c\}}\big] = \mathbf{c}_n^T \mathbf{H}_1$ and $p_{\{n,0\}} = \mathbf{c}_n^T \mathbf{s}$. The error propagation caused by the mismatched hard decision feedback is denoted as i_n, i.e., $i_n \triangleq \sum_{k=1}^{L_d} d_{\{n,-k\}} \big(x_{n-k} - \hat{x}_{n-k}\big)$ and v_n is the sum of noise and the remaining ISI terms caused by the neighboring symbols: $v_n \triangleq \sum_{k=1}^{L_c} p_{\{n,k\}} \big(x_{n+k} - \mathrm{E}(x_{n+k})\big) + \sum_{k=0}^{L_c} c_{\{n,k\}} w_{n+k}$. Same as LE, the variance of v_n is

$$\begin{aligned} \mathrm{Var}(v_n) &\triangleq \mathbf{c}_n^T \mathrm{Cov}\{\mathbf{r}_n \mathbf{r}_n^T \mid x_n = x\} \mathbf{c}_n \\ &= \mathbf{c}_n^T \left(\mathbf{H} \boldsymbol{\Sigma}_n \mathbf{H}^T - z_n \mathbf{s}\mathbf{s}^T + N_0 \mathbf{I}\right) \mathbf{c}_n \\ &= \mathbf{c}_n^T \mathbf{s} (1 - \mathbf{s}^T \mathbf{c}_n). \end{aligned} \tag{2.12}$$

Assuming that the feedback decisions are all correct, i.e., $i_n = 0$, and v_n is AWGN, the extrinsic LLR is naturally given by

$$\begin{aligned} L_e(x_n) &= \ln \frac{\Pr(y_n \mid x_n = +1)}{\Pr(y_n \mid x_n = -1)} \\ &= \frac{2 p_{\{n,0\}} y_n}{\mathrm{Var}(v_n)}. \end{aligned} \tag{2.13}$$

However, this LLR mapping method is prone to error propagation. Therefore, SISO DFE has been a less preferred choice as an iterative equalizer structure to eliminate

the ISI although it was observed in [50–52] that DFE with ideal decision feedback can outperform the LE on various ISI channels.

2.3 Proposed SISO Decision Feedback Equalizer

While the MAP estimation of i_n is equal to zero, we observe that the chance of $i_n \neq 0$ is relatively high for severe ISI channels. In order to mitigate the error propagation in DFE, many techniques [53–56] have been investigated in the classical equalization system. On the other hand, our strategy in turbo equalization system is to estimate i_n and utilize the statistical parameters associated with this estimate in the formulation of the extrinsic information. Since i_n is to be estimated on the basis of the observation $\mathbf{y}_n^c \triangleq [y_{n-L_d}, y_{n-L_d+1}, \ldots, y_{n-1}]^T$, $\mathrm{E}(i_n)$ and $\mathrm{Var}(i_n)$ can be defined as the conditional mean and variance, i.e.,

$$\begin{aligned}\mathrm{E}(i_n) &\triangleq \mathrm{E}(i_n|\mathbf{y}_n^c) = \mathrm{E}\left\{\mathbf{d}_n^T(\mathbf{x}_n^c - \hat{\mathbf{x}}_n^c) \mid \mathbf{y}_n^c\right\} \\ &= \mathbf{d}_n^T\left(\tanh(L(\mathbf{x}_n^c)/2) - \hat{\mathbf{x}}_n^c\right) \qquad (2.14)\\ \mathrm{Var}(i_n) &\triangleq \mathrm{Var}(i_n \mid \mathbf{y}_n^c) = \mathrm{Var}\left\{\mathbf{d}_n^T(\mathbf{x}_n^c - \hat{\mathbf{x}}_n^c) \mid \mathbf{y}_n^c)\right\} \\ &= \mathbf{d}_n^T\acute{\mathbf{\Sigma}}_n^c\mathbf{d}_n \qquad (2.15)\end{aligned}$$

where $L(\mathbf{x}_n^c) = [L(x_{n-L_d}), \ldots, L(x_{n-1})]^T$, $\acute{\mathbf{\Sigma}}_n^c \triangleq \mathrm{Diag}\left(\acute{z}_{n-L_d}, \ldots, \acute{z}_{n-1}\right)$, and $\acute{z}_n = 1 - \tanh(L(x_n)/2)^2$.

Now, let us consider the possible causal error sequence $\mathbf{e}_{\{n,j\}}^c \triangleq \mathbf{x}_{\{n,j\}}^c - \hat{\mathbf{x}}_n^c$ for $j = 1, 2, \ldots, 2^{L_d}$. Then, we can compute the extrinsic information for the given causal error sequence $\mathbf{e}_{\{n,j\}}^c$:

$$\begin{aligned}L_e(x_n|\mathbf{e}_{\{n,j\}}^c) &\triangleq \ln\frac{\Pr(y_n \mid x_n = +1, \mathbf{e}_{\{n,j\}}^c)}{\Pr(y_n \mid x_n = -1, \mathbf{e}_{\{n,j\}}^c)} \\ &= \frac{2p_{\{n,0\}}(y_n - \mathbf{d}_n^T\mathbf{e}_{\{n,j\}}^c)}{\mathrm{Var}(v_n)}. \qquad (2.16)\end{aligned}$$

To compute the extrinsic information of x_n taking into account the probabilities of

possible error sequences, we write

$$\begin{aligned}\Pr(y_n \mid x_n = +1) &= \sum_{j=1}^{2^{L_d}} \Pr(y_n \mid x_n = +1, \mathbf{e}^c_{\{n,j\}}) \Pr(\mathbf{e}^c_{\{n,j\}}) \\ &= \sum_{j=1}^{2^{L_d}} \frac{\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right) \Pr(\mathbf{e}^c_{\{n,j\}})}{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)} \qquad (2.17) \\ \Pr(y_n \mid x_n = -1) &= \sum_{j=1}^{2^{L_d}} \Pr(y_n \mid x_n = -1, \mathbf{e}^c_{\{n,j\}}) \Pr(\mathbf{e}^c_{\{n,j\}}) \\ &= \sum_{j=1}^{2^{L_d}} \frac{\Pr(\mathbf{e}^c_{\{n,j\}})}{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)}. \qquad (2.18)\end{aligned}$$

Accordingly, the extrinsic information of x_n considering the distribution of i_n is given as

$$L_e(x_n) = \ln\left\{\sum_{j=1}^{2^{L_d}} \frac{\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right) \Pr(\mathbf{e}^c_{\{n,j\}})}{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)}\right\} - \ln\left\{\sum_{j=1}^{2^{L_d}} \frac{\Pr(\mathbf{e}^c_{\{n,j\}})}{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)}\right\}. \qquad (2.19)$$

In principle, the extrinsic information of (2.19) can be evaluated using (2.16) and approximating $\Pr(\mathbf{e}^c_{\{n,j\}})$ or $\Pr(\mathbf{e}^c_{\{n,j\}}|\mathbf{y}^c_n)$ by $\prod_{k=1}^{L_d} \Pr(e_{\{n-k,j\}}|y_{n-k})$, which can be computed based on the *a posteriori* LLRs of $\mathbf{x}^c_n$.

However, since the computational complexity of (2.19) increases exponentially according to the length of feedback filter, L_d, we seek a more practical modification. A possible solution is to apply the Bayes' rule only for the two mutually exclusive cases of $i_n = 0$ and $i_n \neq 0$. Then,

$$\Pr(y_n \mid x_n = +1) = \frac{\exp\left(L_e(x_n|i_n=0)\right)\Pr(i_n=0)}{1+\exp\left(L_e(x_n|i_n=0)\right)} + \frac{\exp\left(L_e(x_n|i_n\neq 0)\right)\Pr(i_n\neq 0)}{1+\exp\left(L_e(x_n|i_n\neq 0)\right)} \qquad (2.20)$$

$$\Pr(y_n \mid x_n = -1) = \frac{\Pr(i_n=0)}{1+\exp\left(L_e(x_n|i_n=0)\right)} + \frac{\Pr(i_n\neq 0)}{1+\exp\left(L_e(x_n|i_n\neq 0)\right)}. \qquad (2.21)$$

The extrinsic information of x_n for each case of i_n can be estimated as

$$L_e(x_n|i_n=0) = \frac{2p_{\{n,0\}}y_n}{\mathrm{Var}(v_n)} \tag{2.22}$$

$$\begin{aligned}
L_e(x_n|i_n\neq 0) =& \ln\left\{\sum_{j=1,\mathbf{e}^c_{\{n,j\}}\neq\mathbf{0}}^{2^{L_d}} \frac{\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)\Pr(\mathbf{e}^c_{\{n,j\}})}{\left\{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)\right\}\Pr(i_n\neq 0)}\right\} \\
&- \ln\left\{\sum_{j=1,\mathbf{e}^c_{\{n,j\}}\neq\mathbf{0}}^{2^{L_d}} \frac{\Pr(\mathbf{e}^c_{\{n,j\}})}{\left\{1+\exp\left(L_e(x_n|\mathbf{e}^c_{\{n,j\}})\right)\right\}\Pr(i_n\neq 0)}\right\} \\
\simeq& \ln\left\{\sum_{j=1,\mathbf{e}^c_{\{n,j\}}\neq\mathbf{0}}^{2^{L_d}} \left(\frac{1}{2}+\frac{L_e(x_n|\mathbf{e}^c_{\{n,j\}})}{4}\right)\frac{\Pr(\mathbf{e}^c_{\{n,j\}})}{\Pr(i_n\neq 0)}\right\} \\
&- \ln\left\{\sum_{j=1,\mathbf{e}^c_{\{n,j\}}\neq\mathbf{0}}^{2^{L_d}} \left(\frac{1}{2}-\frac{L_e(x_n|\mathbf{e}^c_{\{n,j\}})}{4}\right)\frac{\Pr(\mathbf{e}^c_{\{n,j\}})}{\Pr(i_n\neq 0)}\right\} \qquad (2.23) \\
=& \ln\left\{\mathop{\mathrm{E}}_{i_n}\left(\frac{1}{2}+\frac{2p_{\{n,0\}}\left(y_n-i_n\right)}{4\mathrm{Var}(v_n)}\middle| i_n\neq 0\right)\right\} \\
&- \ln\left\{\mathop{\mathrm{E}}_{i_n}\left(\frac{1}{2}-\frac{2p_{\{n,0\}}\left(y_n-i_n\right)}{4\mathrm{Var}(v_n)}\middle| i_n\neq 0\right)\right\} \\
=& \ln\left\{1+\frac{p_{\{n,0\}}\left(y_n-\mathrm{E}(i_n|i_n\neq 0)\right)}{\mathrm{Var}(v_n)}\right\} \\
&- \ln\left\{1-\frac{p_{\{n,0\}}\left(y_n-\mathrm{E}(i_n|i_n\neq 0)\right)}{\mathrm{Var}(v_n)}\right\} \\
\simeq& \begin{cases} 2\varphi_n/(1-\varphi_n) \text{ if } \varphi_n<0 \\ 2\varphi_n/(1+\varphi_n) \text{ otherwise} \end{cases} \qquad (2.24) \\
=& \frac{2\varphi_n}{1+|\varphi_n|} \qquad (2.25)
\end{aligned}$$

where $\varphi_n \triangleq p_{\{n,0\}}\left(y_n-\mathrm{E}(i_n|i_n\neq 0)\right)/\mathrm{Var}(v_n)$. The approximation of (2.23) is from the first order Taylor expansion at zero, i.e, $e^x/(1+e^x)\simeq 0.5+0.25x$ and $1/(1+e^x)\simeq 0.5-0.25x$. We also use $\ln\{1+\varphi_n\}-\ln\{1-\varphi_n\}=\ln\{1+2\varphi_n/(1-\varphi_n)\}=-\ln\{1-2\varphi_n/(1+\varphi_n)\}$ and $\ln(1+x)\simeq x$ in (2.24): $\ln\{1+2\varphi_n/(1-\varphi_n)\}\simeq 2\varphi_n/(1-\varphi_n)$ is used for $\varphi_n<0$ while $-\ln\{1-2\varphi_n/(1+\varphi_n)\}\simeq 2\varphi_n/(1+\varphi_n)$ is used for

$\varphi_n \geq 0$. Furthermore, the conditional mean $\mathrm{E}(i_n|i_n \neq 0)$ can be estimated by the *a posteriori* LLRs of L_d causal symbols. In other words, $\mathrm{E}(i_n|i_n \neq 0) = \mathrm{E}(i_n)/\Pr(i_n \neq 0)$ where $\mathrm{E}(i_n)$ is given in (2.14) and $\Pr(i_n \neq 0) = 1 - \Pr(i_n = 0)$ with $\Pr(i_n = 0) = \prod_{k=1}^{L_d} \exp(|L(x_{n-k})|)/(1 + \exp(|L(x_{n-k})|))$. Finally, the extrinsic information of x_n is given as

$$L_e(x_n) = \ln\left\{\frac{\exp\left(L_e(x_n|i_n=0)\right)\Pr(i_n=0)}{1+\exp\left(L_e(x_n|i_n=0)\right)} + \frac{\exp\left(L_e(x_n|i_n\neq 0)\right)\Pr(i_n\neq 0)}{1+\exp\left(L_e(x_n|i_n\neq 0)\right)}\right\} - \ln\left\{\frac{\Pr(i_n=0)}{1+\exp\left(L_e(x_n|i_n=0)\right)} + \frac{\Pr(i_n\neq 0)}{1+\exp\left(L_e(x_n|i_n\neq 0)\right)}\right\}. \quad (2.26)$$

While this gets passed to the outer decoder as equalizer's extrinsic information, hard decisions that propagate down the feedback filter are generated by slicing $L_e(x_n) + L_a(x_n)$ where $L_a(x_n)$ is the extrinsic information from the decoder. Note that the original structure of DFE is not modified, only LLR mapping method is newly adopted. Another improved LLR mapping methods for DFE also exists [23].

2.4 Low Complexity Filter Derivations

As also discussed in [14], the filter tap values derived above are time-varying and creates significant implementation challenges because the filter taps need to be recalculated at each time index n. In an effort to reduce the complexity associated with time-varying filter implementation, low complexity filter computations have been introduced in [13, 14, 16–18].

2.4.1 Time-invariant Filters

A low-complexity variation would be to simply assume the variane of x_n is equal to 1 ($z_n = 1$) for all n during the filter derivations. These time-invariant-filters based approaches can reduce the complexity to a linear function of the filter length since the filter computation is not necessary at each time index n.

MMSE Linear Equalizer

The time-invariant filter of LE is given as

$$\begin{aligned}\mathbf{f} &\triangleq \left[f_{-L_{f1}}, f_{-L_{f1}+1}, \ldots, f_{L_{f2}}\right]^T \\ &= \left\{\mathbf{H}\mathbf{H}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s}. \end{aligned} \tag{2.27}$$

Note that, even though the time-invariant filters are used, the *a priori* information is exploited to subtract the mean values of the channel observation vector $\mathbf{r}_n$ (see (2.2)). Then, as shown in [14], the MSE (or the variance of v_n) of the equalized output, y_n, is

$$\mathrm{Var}(v_n) = \mathbf{f}^T\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}.$$

The extrinsic information can be generated by using (2.7) with a Gaussian RV assumption of v_n.

MMSE Decision Feedback Equalizer

The time-invariant filters of DFE are also given as

$$\begin{aligned}\mathbf{c} &\triangleq [c_0, c_{+1}, \ldots, c_{L_c}]^T \\ &= \left\{\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \end{aligned} \tag{2.28}$$

$$\begin{aligned}\mathbf{d} &\triangleq [d_{-L_d}, d_{-L_d+1}, \ldots, d_{-1}]^T \\ &= \mathbf{M}\mathbf{H}^T\mathbf{c} \end{aligned} \tag{2.29}$$

where $\boldsymbol{\Sigma} \triangleq \mathrm{Diag}(\mathbf{0}_{1\times L_d}, \mathbf{1}_{1\times(L_c+1)})$.

Again, even though the time-invariant filters are used, the *a priori* information is exploited to subtract the mean values of the channel observation vector $\mathbf{r}_n$ (see (2.10)) and enhance *a posteriori* LLR computation: $L_e(x_n) + L_a(x_n)$. Then, the mean and variance of i_n and the noise variance of v_n with the time-invariant filters are also given

by

$$\mathrm{E}(i_n) = \mathbf{d}^T \left(\tanh(L(\mathbf{x}_n^c)/2) - \hat{\mathbf{x}}_n^c\right) \tag{2.30}$$

$$\mathrm{Var}(i_n) = \mathbf{d}^T \acute{\mathbf{\Sigma}}_n^c \mathbf{d} \tag{2.31}$$

$$\mathrm{Var}(v_n) = \mathbf{c}^T \left(\mathbf{H}\mathbf{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{c}. \tag{2.32}$$

Also, the extrinsic information can be obtained in a similar way to (2.26).

2.4.2 Iteration-varying Filters

While the overall complexity is reduced, the time-invariant filters do not provide good performance as the maximum achievable output SNR by this equalizer cannot approach the matched filter bound (MFB), the attainable output SNR when no ISI is introduced. In other words, when the perfect *a priori* information is available for all x_n, i.e., $\mathrm{E}(x_n) = x_n$ for all n, the output SNR should be equal to the MFB ideally, but time-invariant-filter-based equalizers cannot achieve the MFB since the filters have not been derived on the premise of the perfect *a priori* information.

A hybrid scheme that switches between two time-invariant filter settings - one corresponding to the case of no *a priori* information and one to the case of the perfect *a priori* information - is also discussed in [14]. Unfortunately, this switching method requires accurate timing of when the switch must take place. Predicting the switching time based on the extrinsic information transfer (EXIT) chart of [63] is suggested in [64] for the LE. However, accurately-timed switching is hard to achieve in practice. This is true especially with the DFE and BiDFE methods for which the EXIT characteristic is not easy to be characterized due to the non-negligible effect of error propagation. This switching method is extended to the adaptive LE based on the classification of the *a priori* information block [65].

Another solution, as discussed in [13, 16–18, 22, 25], would be the filter tap updating for each iteration rather than for each time index which also leads linearly growing complexity with the filter length same as the time-invariant filter. Specifically, the filter

taps are recomputed only once for each iteration, by replacing the variance of x_n with the "iteration-varying" quantity $z^{(k)}$, which is the time-averaged variance of x_n at the kth turbo iteration, for all n, i.e., $z_n = z^{(k)}$, [13]. Although this method gives rise to suboptimal filters for each individual received sample block, the resulting filters are optimal in the sense of minimizing the averaged MSE of the overall equalized outputs when the filters are constrained to be time-invariant. Note that these iteration-varying filters can be also utilized for DFE although the iteration-varying filters are introduced only for LE in [13].

Linear Equalizer

The iteration-varying filter for the LE at the kth turbo iteration is given by

$$\begin{aligned}\mathbf{f}^{(k)} &\triangleq \left[f_{-L_{f1}}^{(k)}, f_{-L_{f1}+1}^{(k)}, \ldots, f_{L_{f2}}^{(k)}\right]^T \\ &= \left\{\mathbf{H}\boldsymbol{\Sigma}^{(k)}\mathbf{H}^T + (1-z^{(k)})\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \end{aligned} \tag{2.33}$$

where $\boldsymbol{\Sigma}^{(k)} \triangleq z^{(k)}\mathbf{I}$ and $z^{(k)} \triangleq \overline{z_n} = 1 - 1/N\sum_{n=0}^{N-1}\tanh(L_a(x_n)/2)^2$. Notice that, unlike in [25], the *a priori* information from the decoder is directly employed, i.e., no need to generate a Gaussian random variable. Then, the MSE (or the variance of v_n) of the equalized output, y_n, is

$$\mathrm{Var}(v_n) = \mathbf{f}^{(k)T}\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}^{(k)}.$$

It can be shown that this filter actually corresponds to the minimum time-averaged MSE solution. Let us prove this. Let $\mathbf{f}^{(k)}$ denote the time-invariant LE filter at the kth turbo iteration. Then, the optimal filter, $\mathbf{f}^{(k)}$, minimizing the time-averaged MSE of

the overall equalized outputs is given by

$$\begin{aligned}
\mathbf{f}^{(k)} &\triangleq \arg\min_{\mathbf{f}^{(k)}} \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{Var}(v_n) \\
&= \arg\min_{\mathbf{f}^{(k)}} \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{f}^{(k)T}\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}^{(k)} \\
&= \arg\min_{\mathbf{f}^{(k)}} \mathbf{f}^{(k)T}\left(\mathbf{H}\overline{\boldsymbol{\Sigma}_n}\mathbf{H}^T - \overline{z_n}\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}^{(k)} \\
&= \arg\min_{\mathbf{f}^{(k)}} \mathbf{f}^{(k)T}\left(\mathbf{H}\boldsymbol{\Sigma}^{(k)}\mathbf{H}^T - z^{(k)}\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{f}^{(k)} \\
&= \left\{\mathbf{H}\boldsymbol{\Sigma}^{(k)}\mathbf{H}^T + (1-z^{(k)})\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \qquad (2.34)
\end{aligned}$$

where $\boldsymbol{\Sigma}^{(k)} \triangleq z^{(k)}\mathbf{I}$ and $z^{(k)} \triangleq \overline{z_n} = 1/N\sum_{n=0}^{N-1} z_n$. The last equality comes from the orthogonal principle. This expression is identical to the iteration-varying filter of (2.33), and the proof is completed.

Just as in the time-invariant filters, the available individual *a priori* information is still exploited via the subtraction of the mean values of the channel observation vector $\mathbf{r}_n$ (see (2.2)). Finally, the extrinsic information is obtained using (2.7), assuming a Gaussian v_n.

Decision Feedback Equalizer

The iteration-varying filters of the DFE at the kth turbo iteration are given by

$$\begin{aligned}
\mathbf{c}^{(k)} &\triangleq \left[c_0^{(k)}, c_{+1}^{(k)}, \ldots, c_{L_c}^{(k)}\right]^T \\
&= \left\{\mathbf{H}\boldsymbol{\Sigma}^{(k)}\mathbf{H}^T + (1-z^{(k)})\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right\}^{-1}\mathbf{s} \qquad (2.35) \\
\mathbf{d}^{(k)} &\triangleq \left[d_{-L_d}^{(k)}, d_{-L_d+1}^{(k)}, \ldots, d_{-1}^{(k)}\right]^T \\
&= \mathbf{M}\mathbf{H}^T\mathbf{c}^{(k)} \qquad (2.36)
\end{aligned}$$

where $\boldsymbol{\Sigma}^{(k)} \triangleq z^{(k)} \cdot \mathrm{Diag}(\mathbf{0}_{1\times L_d}, \mathbf{1}_{1\times(L_c+1)})$. Again, the available individual *a priori* information is utilized in calculation of the equalized output (2.10).

The mean and variance of i_n and the noise variance of v_n with the iteration-varying filters are also given by

$$\mathrm{E}(i_n) = \mathbf{d}^{(k)T} \left(\tanh(L(\mathbf{x}_n^c)/2) - \hat{\mathbf{x}}_n^c\right) \tag{2.37}$$

$$\mathrm{Var}(i_n) = \mathbf{d}^{(k)T} \acute{\mathbf{\Sigma}}_n^c \mathbf{d}^{(k)} \tag{2.38}$$

$$\mathrm{Var}(v_n) = \mathbf{c}^{(k)T} \left(\mathbf{H}\mathbf{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right) \mathbf{c}^{(k)}. \tag{2.39}$$

The extrinsic information is obtained using (2.26) assuming a Gaussian v_n but taking into account the possibility of $i_n \neq 0$. The optimality of this DFE filter in the sense of minimum time-averaged MSE can be established using a similar proof as the LE case above.

2.4.3 Complexity Comparison

Let us consider the computational complexity of different filter types. Assuming binary symbols, the total required numbers of multiplications per iteration for the classical time-invariant filter, the time-varying filter, and the iteration-varying filter are roughly NL, $N(L + L^2)$, and $NL + L^3$, respectively, where N is the total number of symbols per codeword and L is the filter length. It is assumed that the matrix inversion in filter tap computation requires complexity $O(L^3)$, but the complexity of the time-varying filter derivation can be reduced to $O(L^2)$ using a recursive method [13]. It is easily seen that if the codeword length is much longer than the filter length, which is commonly the case, then the per-iteration, per-symbol complexity of the time-invariant filter, the time-varying filter, and the iteration-varying filter goes as $O(L)$, $O(L^2)$, and $O(L)$, respectively.

2.5 Performance Comparison

In this section, the performance of an equalizer with various filter types are considered. When no *a priori* information is available, i.e., $\mathrm{E}(x_n) = 0$ (or $L_a(x_n) = 0$) for all n, the

time-varying filter, the time-invariant filter and the iteration-varying filter all obviously become identical.

Now, let us consider the ideal condition for the equalizer where the perfect *a priori* information is available, i.e., $\mathrm{E}(x_n) = x_n$ (or $L_a(x_n) = \pm\infty$) for all n. This condition may be satisfied when several iterations are performed at high SNRs in turbo equalization. Then, the remaining ISI terms in the equalized output for the LE and DFE as well as the error propagation variable for the DFE will disappear subsequently. Moreover, the time-varying filter and the iteration-varying filter will converge to the normalized matched filter regardless of the equalization algorithm used. Accordingly, the unbiased output SNR associated with the LE and DFE with the time-varying and the iteration-varying filters will be equal to that of the matched filter. Finally, the LE and DFE will produce the same equalized output and the unbiased output SNRs with the time-varying and the iteration-varying filters are given by

$$\mathrm{SNR}_{\infty,\mathrm{LE}} = \mathrm{SNR}_{\infty,\mathrm{DFE}} = \mathrm{MFB} = \frac{P_X}{N_0} \tag{2.40}$$

where SNR_∞ denotes the unbiased output SNR with the perfect *a priori* information.

On the other hand, the maximum achievable unbiased output SNRs with the time-invariant filters are given as

$$\mathrm{SNR}_{\infty,\mathrm{LE}} = \frac{p_0^2 \mathrm{E}(x_n^2)}{\mathrm{Var}(v_n)} = \frac{P_X}{N_0}\left(\frac{(\mathbf{f}^T\mathbf{s})^2}{\mathbf{f}^T\mathbf{f}}\right) \tag{2.41}$$

$$\mathrm{SNR}_{\infty,\mathrm{DFE}} = \frac{p_0^2 \mathrm{E}(x_n^2)}{\mathrm{Var}(v_n)} = \frac{P_X}{N_0}\left(\frac{(\mathbf{c}^T\mathbf{s})^2}{\mathbf{c}^T\mathbf{c}}\right) \tag{2.42}$$

where p_0 is the bias weight on the transmitted symbol x_n in the equalized output y_n, generated by the time-invariant filters (see (2.2) and (2.10)). Specifically, $p_0 = \mathbf{f}^T\mathbf{s}$ for the LE and $p_0 = \mathbf{c}^T\mathbf{s}$ for the DFE. Since the residual ISI terms disappear in the noise v_n, $\mathrm{Var}(v_n) = N_0\mathbf{f}^T\mathbf{f}$ for the LE while $\mathrm{Var}(v_n) = N_0\mathbf{c}^T\mathbf{c}$ for the DFE.

The values inside the parentheses of (2.41) and (2.42) depend on the channel response and the channel input SNR, but they are less than 1 unless the channel is the ideal AWGN channel or the channel input SNR is zero, i.e., $P_X/N_0 = 0$; thus, these output

SNRs are also less than the MFB. Moreover, $\mathrm{SNR}_{\infty,\mathrm{DFE}}$ is higher than $\mathrm{SNR}_{\infty,\mathrm{LE}}$ since the time-invariant feedforward filter of DFE can be viewed as the LE filter derived by assuming the *a priori* information for the causal symbols is perfect at least. Finally,

$$\mathrm{SNR}_{\infty,\mathrm{LE}} \leq \mathrm{SNR}_{\infty,\mathrm{DFE}} \leq \mathrm{MFB} = \frac{P_X}{N_0} \tag{2.43}$$

where the equalities only hold on the ideal AWGN channel or at zero channel input SNR.

In order to incorporate the above output SNR analysis with the EXIT chart, we further compute the analytical MI with these output SNRs. Because the Gaussian noise term only remains in the equalized output under the perfect *a priori* information, the MI is simply given as

$$C_b(R) \triangleq 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log_2 \left\{ 1 + e^{-2\sqrt{R}\tau - 2R} \right\} d\tau \tag{2.44}$$

where $C_b(R)$ is the SIR of the binary-input Gaussian channel with SNR given by R. Note that the MI computed by substituting $R = \mathrm{SNR}_\infty$ in (2.44) is the maximum attainable MI by an equalizer. The corresponding numerical results will be presented in conjunction with the discussion of the EXIT chart analysis results.

2.6 Numerical Results

In this section, simulation results of several iterative equalization schemes are presented. The transmitted symbols are encoded with a recursive rate-1/2 convolutional code encoder with parity generator $(1+D^2)/(1+D+D^2)$ with 2^{11} message bits and are modulated by binary phase-shift keying (BPSK) so that $x_n \in \{\pm 1\}$. We also assume that the noise is AWGN, and the noise variance and the channel information are perfectly known to the receiver. The impulse response of the ISI channel $\mathbf{h_1} = (1/\sqrt{19})[1 \quad 2 \quad 3 \quad 2 \quad 1]^T$ investigated in [14] is used for evaluating the performance of iterative equalizers. This channel is considered a severe ISI channel as the channel spectra possess nulls over the

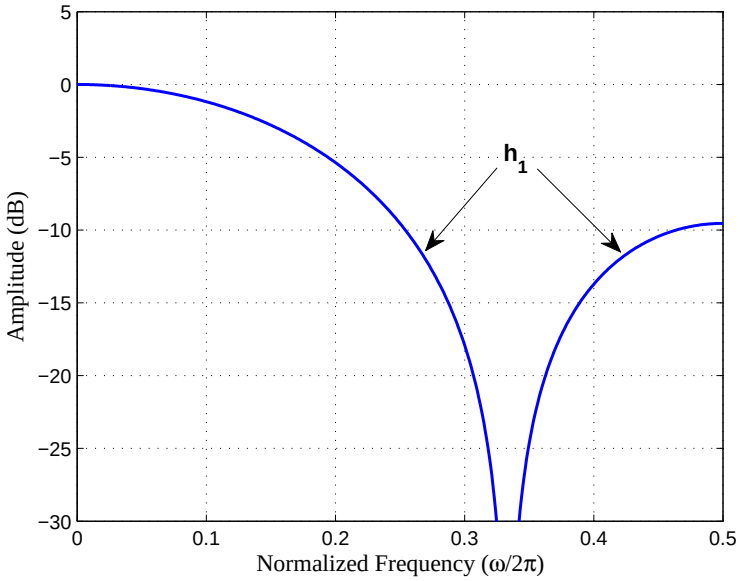


Figure 2.2: Frequency Response of the ISI Channel $\mathbf{h_1} = (1/\sqrt{19})[1 \quad 2 \quad 3 \quad 2 \quad 1]^T$.

Nyquist band, as shown in Fig. 2.2. The decoder is implemented using the BCJR algorithm. Only the SISO equalizer changes from one scheme to another. The DFE with 17 feedforward taps and 4 feedback taps ($L_c = 16$ and $L_d = 4$) is used for both the normal and the time-reversed DFEs while the LE uses 21 taps ($L_{f1} = 10$ and $L_{f2} = 10$). In the figures, the "Proposed DFE" uses the proposed LLR mapping of (2.26) while "DFE" uses the conventional LLR mapping of (2.13) (as used in [14]). Furthermore, the symbols "TV-", "TI-", and "IV-" in the legend box denote equalizers with time-varying filters, time-invariant filters, and iteration-varying filters, respectively. For instance, "TV-LE" is the LE with the time-varying filter (or the exact MMSE in [13, 14]) and "IV-LE" is the LE with the iteration-varying filter (or the low-complexity (LC) approximate MMSE in [13]). Finally, the "MAP" is the optimal equalizer implemented via the BCJR algorithm.

First, let us consider the computational complexity of the various equalizers. As shown in [14], the number of multiplications and additions increases as an exponential

function of the channel memory length for the MAP equalizer whereas the number of the same operations is a quadratic function of both the channel memory length and the filter length for the time-varying filter based approaches. The number of operations, on the other hand, increases only linearly for the time-invariant and the iteration-varying filter based approaches.

Fig. 2.3 shows the performance of several turbo equalizers with time-varying filters after 20 iterations. TV-DFE with the conventional LLR mapping shows poor performance but once the proposed LLR generations are used ("Proposed TV-DFE"), the DFE performance becomes clearly better than the TV-LE method of [14], except at very high SNRs where all schemes other than the conventional DFE perform comparably.

Fig. 2.4 shows the bit error rate (BER) performance of time-invariant filter based turbo equalizers. As the figure indicates, the "Proposed TI-DFE" also shows the superior performance to the "TI-DFE" and "IV-LE". Also notice that "Proposed TI-DFE" achieves decision-error-free performance at low BERs, indicating the error propagation effect has been nearly eliminated using the proposed LLR generation method. However, the equalizers with the time-invariant filters have their performance limitations since the time-invariant filters do not fully exploit the *a priori* information from the decoder.

The performance with the iteration-varying filters are shown in Fig. 2.5. Again, the "Proposed IV-DFE" shows the better BER performance than the "IV-DFE" and "IV-LE". Note that the performance with the iteration-varying filters is very close to the performance with the time-varying filters while achieving a low computational complexity nearly equal to the time-invariant filters.

The performance of the turbo equalizers are analyzed by using the EXIT chart [63], a diagram demonstrating the MI transfer characteristics of the two constituent modules which exchange soft information. In the EXIT charts, the behavior of the channel equalizer is described with its input and output on the horizontal and vertical axis, respectively, while the behavior of the decoder is described in the opposite way. The

EXIT chart curves typically define a path for the MI trajectory to move up during iterative processing of soft information. Moreover, the number of stairs that a given MI trajectory (averaged over 1000 sample codeword blocks here) takes to reach the highest value indicates the necessary number of iterations toward convergence.

Figs. 2.6, 2.7, and 2.8 show the EXIT chart corresponding to time-varying-filter-based, time-invariant-filter-based, and iteration-varying-filter-based equalizers for $\mathbf{h_1}$ at a 7 dB SNR respectively. The maximum MI with $\mathrm{SNR}_{\infty,\mathrm{DFE}}$ and $\mathrm{SNR}_{\infty,\mathrm{LE}}$ achieved by the corresponding filter types are also plotted in the figures. As the figures show, the trajectories of "TV-DFE", "TI-DFE", and "IV-DFE" move up for the first couple of iterations, but then quickly fizzle out due to the inadequate extrinsic LLR generations that cannot handle error propagation. However, the trajectories of "Proposed TV-DFE", "Proposed TI-DFE", and "Proposed IV-DFE" keep moving up as the number of iterations increases, clearly indicating the advantage and effectiveness of the proposed LLR generation method. However, when time-invariant filters are used, the trajectory of "Proposed TI-DFE" and "TI-LE" fail to approach 1 bit of MI even though they keep moving up towards its own maximum MI limit as the number of iterations increases. Furthermore, as expected, the maximum MI of TI-DFE is higher than the maximum MI of TI-LE which confirms why the BER performance of TI-DFE is better than that of TI-LE in Fig. 2.4.

2.7 Summary

In this chapter, we reviewed low-complexity turbo equalization methods based on LE and DFE structures of [14] and proposed new SISO DFE structure well-suited to turbo equalization. The proposed LLR generation for DFE designed to reduce error propagation indeed provide its decision-error-free performance in turbo equalizer settings. We also provide the SNR/MI analyses for previously developed filter based equalizers, LE and DFE.

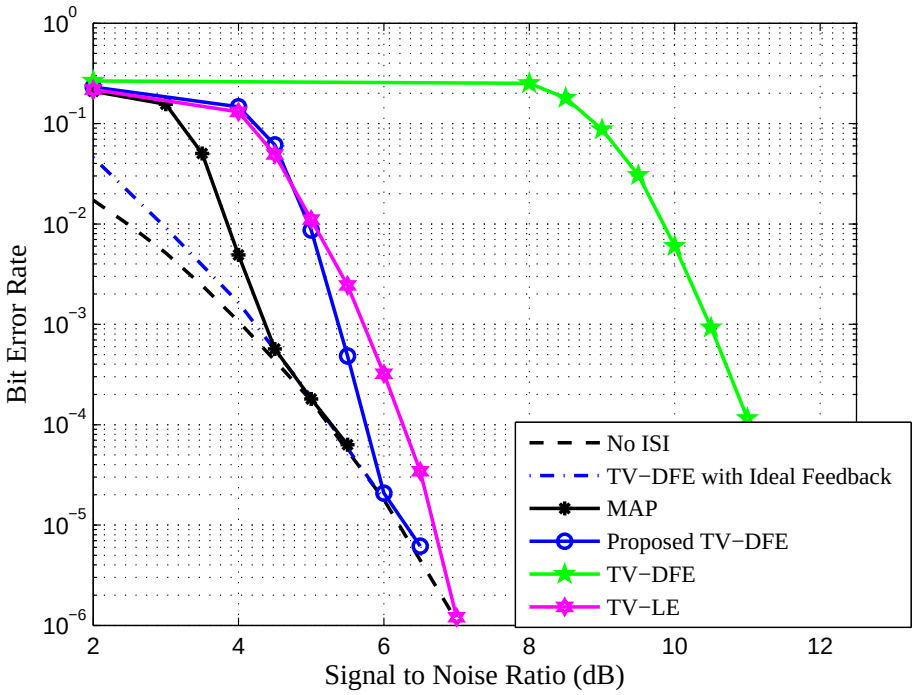


Figure 2.3: BER Curve on the Channel $\mathbf{h_1}$ after 20 Iterations with Time-varying Filters.

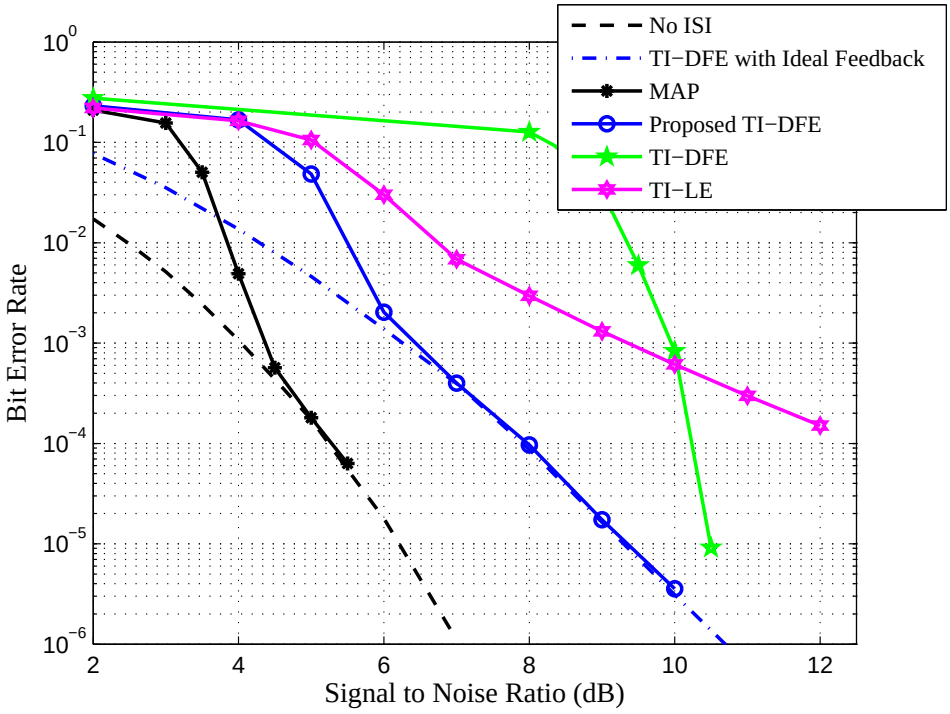


Figure 2.4: BER Curve on the Channel $\mathbf{h_1}$ after 20 Iterations with Time-invariant Filters.

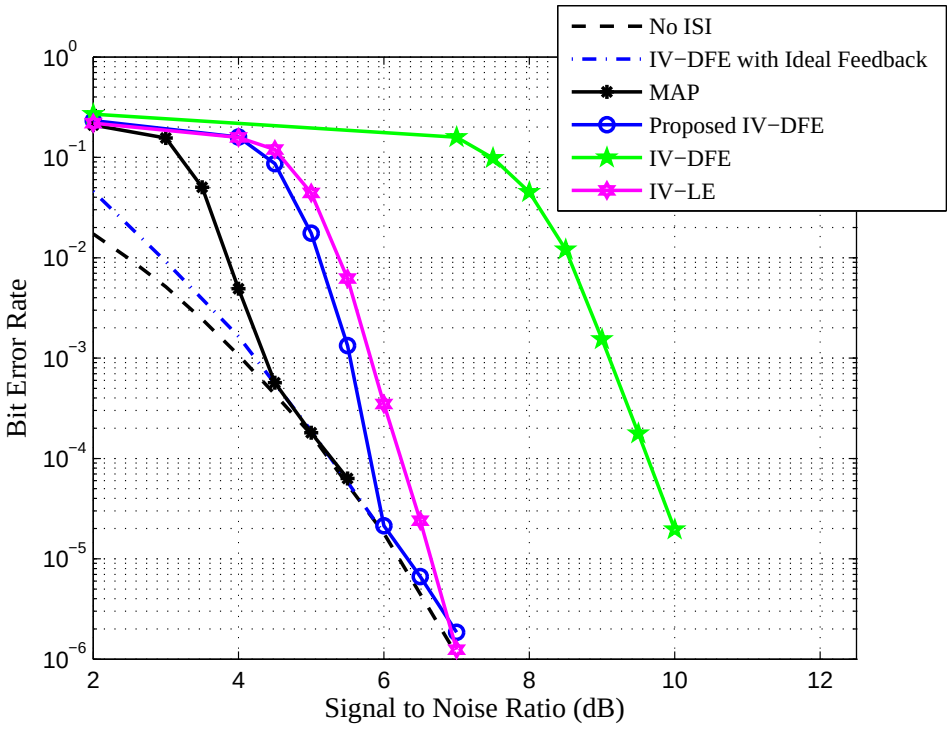


Figure 2.5: BER Curve on the Channel $\mathbf{h_1}$ after 20 Iterations with Iteration-varying Filters.

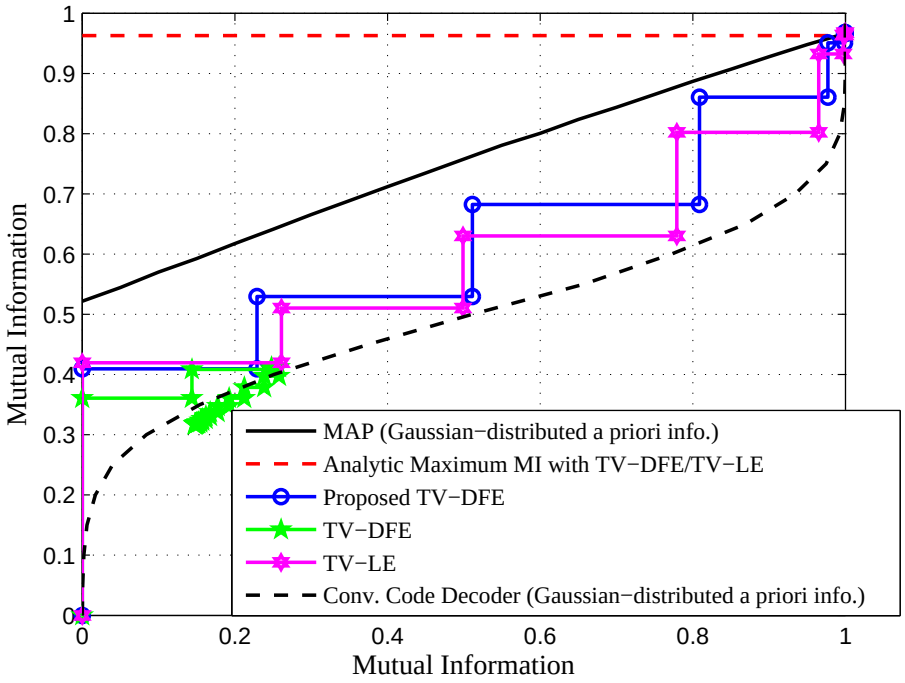


Figure 2.6: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Time-varying Filters.

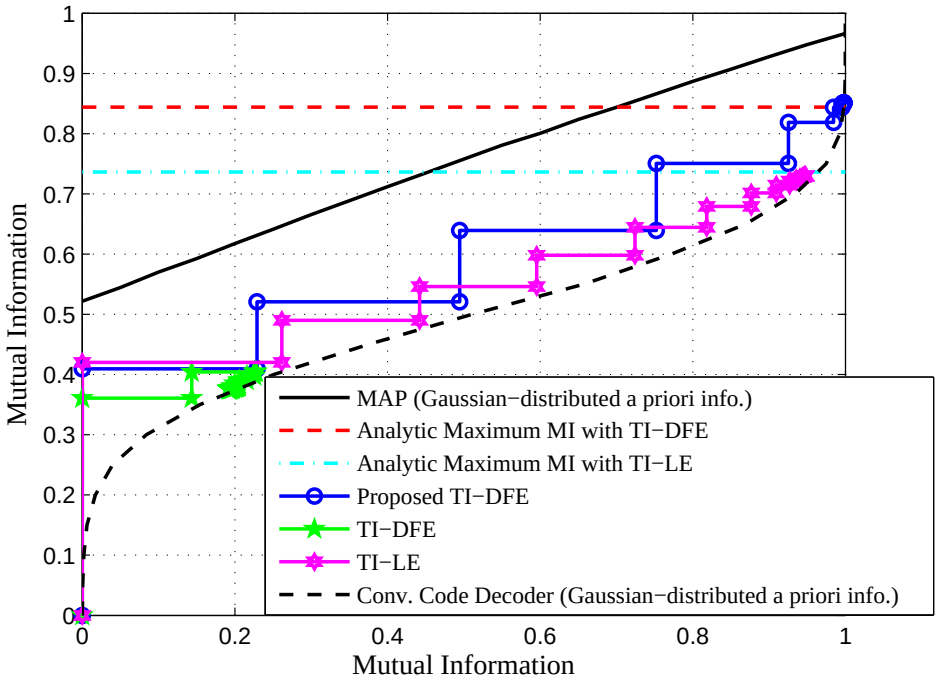


Figure 2.7: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Time-invariant Filters.

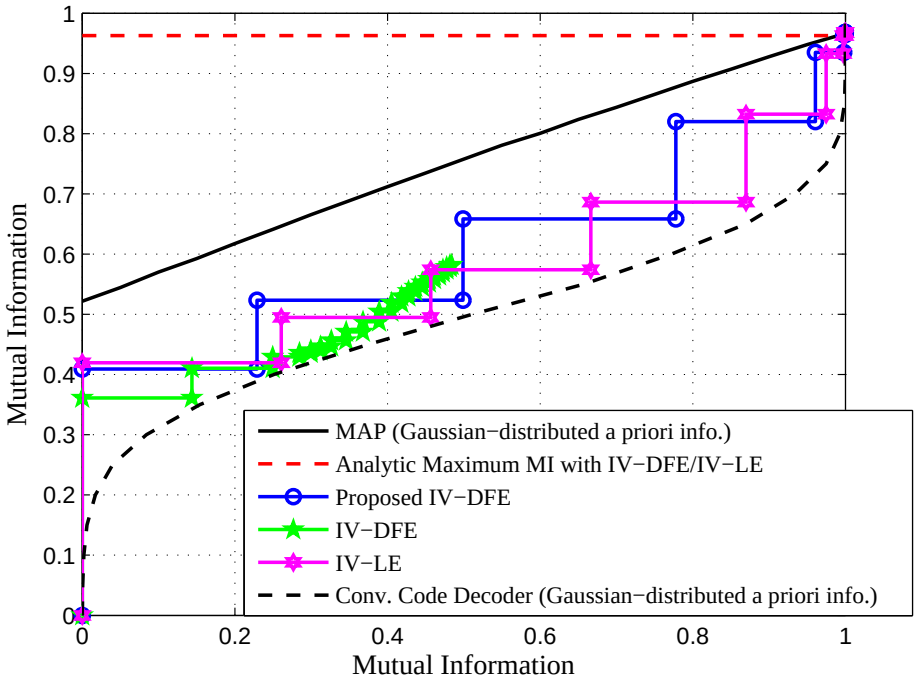


Figure 2.8: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Iteration-varying Filters.

Chapter 3

SISO Bidirectional DFE

Recently, it has been shown [58–60], that conducting both normal and time-reversed equalization of the received data sequence with two DFEs (or SFEs) running in opposite directions and combining two DFE (or SFE) outputs is very effective in reducing error propagation and improving BER performance. This "bi-directional" DFE (called BiDFE) algorithm takes advantage of the different decision error and noise distributions at the outputs of the forward and time-reversed DFEs [59, 60].

In this chapter, a specific DFE extrinsic information combining strategy applied to a BiDFE scheme is addressed to suppress statistical correlation between the outputs of two opposite direction DFEs. We show that the resulting turbo BiDFE performance approaches the performance of the BCJR-based turbo equalizer in a fairly severe ISI environment, easily outperforming the turbo equalizer based on the SISO LE and SISO DFE. Moreover, unlike the LE and DFE, the BiDFE algorithm can be designed to avoid performance degradation even when the filter taps are constrained to be time-invariant [24]. This chapter is written based on [24, 72].

3.1 SNR Advantage of BiDFE

The idea of BiDFE is already motivated in [59,60] by the fact that DFE can be performed on the reversed received sequence using the time-reversed channel response. Here we derive the SNR figure-of-merit for BiDFE assuming ideal feedback in both ways and allowing infinitely long filter lengths. We then compare the result with those of the usual, single-sided DFE as well as the matched filter detector (i.e., ideal detector under zero-ISI condition). As will be seen, the ideal BiDFE SNR is significantly better than the ideal DFE SNR especially at high channel SNRs, further motivating a turbo BiDFE scheme.

3.1.1 Unbiased DFE

It is well known that the D-transforms of the feedforward and feedback MMSE-DFE filter coefficients are, respectively [52]:

$$c(D) = \frac{P_X}{P_0 g^*(D^{-*})},\ d(D) = g(D) \tag{3.1}$$

where P_0 is such that $\log P_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log R_{ss}(e^{-j\theta})d\theta$ and $g^*(D^{-*})$ is obtained from spectral factorization: $R_{ss}(D) = P_X R_{hh}(D) + N_0 = P_0 g(D) g^*(D^{-*})$ where $R_{hh}(D) = h(D)h^*(D^{-*})$ and $h(D)$ is the D-transform of the channel impulse response.

The unbiased equalized outputs of the normal MMSE-DFE in the forward direction, $Y_f(D)$, are given by

$$Y_f(D) = x(D) + \frac{P_0}{P_0 - N_0} e'_f(D) \tag{3.2}$$

where

$$e'_f(D) \triangleq \frac{N_0}{P_0}\left(1 - \frac{1}{g^*(D^{-*})}\right) x(D) + \frac{P_X w'(D)}{P_0 g^*(D^{-*})} \tag{3.3}$$

with $w'(D)$ denoting a complex-valued Gaussian noise sequence with autocorrelation function $R_{w'w'}(D) = N_0 R_{hh}(D)$. Then, the MSE and SNR of the unbiased normal

MMSE-DFE are given by

$$\mathrm{MSE}_{UDFE} = \left(\frac{P_0}{P_0 - N_0}\right)^2 \mathrm{E}(|e'_{f,n}|^2) = \frac{P_X N_0}{P_0 - N_0} \tag{3.4}$$

$$\mathrm{SNR}_{UDFE} \triangleq \frac{P_X}{\mathrm{MSE}_{UDFE}} = \frac{P_0 - N_0}{N_0}. \tag{3.5}$$

3.1.2 Unbiased Time-Reversed DFE

Now, let us assume that the transmitted data sequence x_n is of a finite length so that the MMSE-DFE can be performed on the time-reversed received signals using the time-reverse of the original channel impulse response [61]. Denoting the time-reversed ISI channel coefficients as $\tilde{h}_n = h^*_{L_h-1-n}$, its D-transform is given as $\tilde{h}(D) = D^{L_h-1}h^*(D^{-*})$. Therefore, the D-transform of the autocorrelation function of the time-reversed channel is given by $R_{\tilde{h}\tilde{h}}(D) = \tilde{h}(D)\tilde{h}^*(D^{-*}) = R_{hh}(D)$. Accordingly, the feed-forward and feedback filters of the time-reversed MMSE-DFE, denoted by $\tilde{c}(D)$ and $\tilde{d}(D) - 1$ respectively, are identical to the normal MMSE-DFE filters, i.e.,

$$\tilde{c}(D) = c(D) = \frac{P_X}{P_0 g^*(D^{-*})},\ \tilde{d}(D) = d(D) = g(D). \tag{3.6}$$

The unbiased output of the time-reversed MMSE-DFE can be expressed similarly to the case of the normal, forward MMSE-DFE except that the unbiased output sequence right after the time-reversed MMSE-DFE should also be time-reversed, in order to get the unbiased equalized output $Y_b(D)$ matched to the input sequence $x(D)$. Therefore,

$$Y_b(D) = x(D) + \frac{P_0}{P_0 - N_0} e'_b(D) \tag{3.7}$$

where

$$e'_b(D) \triangleq \frac{N_0}{P_0}\left(1 - \frac{1}{g(D)}\right)x(D) + \frac{P_X}{P_0}\left(\frac{w'(D)}{g(D)}\right). \tag{3.8}$$

Then, the MSE and SNR of the unbiased time-reversed MMSE-DFE are given by

$$\mathrm{MSE}_{URDFE} = \left(\frac{P_0}{P_0 - N_0}\right)^2 \mathrm{E}(|e'_{b,n}|^2) = \frac{P_X N_0}{P_0 - N_0} \tag{3.9}$$

$$\mathrm{SNR}_{URDFE} \triangleq \frac{P_X}{\mathrm{MSE}_{URDFE}} = \frac{P_0 - N_0}{N_0}. \tag{3.10}$$

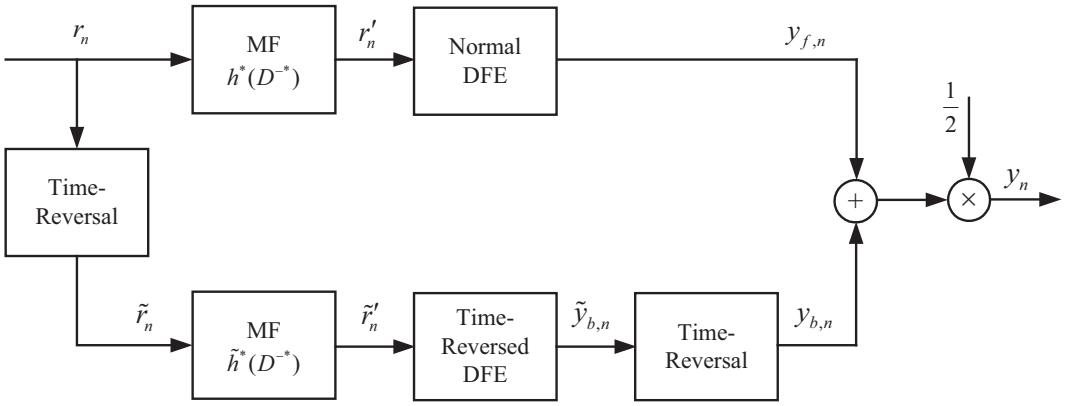


Figure 3.1: Bidirectional Decision Feedback Equalizer (BiDFE).

3.1.3 Unbiased BiDFE

The structure of the BiDFE is shown in Fig. 3.1. If we assume that the feedback sequence is correct, the outputs of two unbiased DFEs are:

$$Y_{f,n} = X_n + V_{f,n} \tag{3.11}$$

$$Y_{b,n} = X_n + V_{b,n} \tag{3.12}$$

where $V_{f,n}$ and $V_{b,n}$ have D-transforms $V_f(D)$ and $V_b(D)$ as given by (from (3.2), (3.3), (3.7), and (3.8))

$$V_f(D) = \frac{N_0}{P_0 - N_0}\left(1 - \frac{1}{g^*(D^{-*})}\right)x(D) + \frac{P_X}{P_0 - N_0}\left(\frac{w'(D)}{g^*(D^{-*})}\right) \tag{3.13}$$

$$V_b(D) = \frac{N_0}{P_0 - N_0}\left(1 - \frac{1}{g(D)}\right)x(D) + \frac{P_X}{P_0 - N_0}\left(\frac{w'(D)}{g(D)}\right). \tag{3.14}$$

Assuming stationary random processes, we drop time index n for notational simplicity and write: $Y_f = X + V_f$ and $Y_b = X + V_b$. From (3.4) and (3.9), the variance of V_f and V_b are also given as:

$$\mathrm{Var}(V_f) = \mathrm{Var}(V_b) = \frac{P_X N_0}{P_0 - N_0}.$$

The variables V_f and V_b are correlated with the correlation coefficient given by

$$\begin{aligned}
\rho &\triangleq \frac{\mathrm{E}(V_f V_b^*)}{\sqrt{\mathrm{Var}(V_f)\mathrm{Var}(V_b)}} \\
&= \frac{P_0 - N_0}{P_X N_0} \mathrm{E}\left[V_f(D) V_b^*(D^{-*})\right]_0 \\
&= \frac{P_X}{N_0(P_0 - N_0)} \mathrm{E}\left[\left(\frac{1}{g^*(D^{-*})}\right)^2 w'(D) w'^*(D^{-*})\right]_0 \qquad (3.15) \\
&= \frac{P_0^2}{P_X N_0(P_0 - N_0)} \left[\{c(D)\}^2 R_{w'w'}(D)\right]_0 \\
&= \frac{P_0^2}{P_X(P_0 - N_0)} \left[\{c(D)\}^2 R_{hh}(D)\right]_0 \qquad (3.16)
\end{aligned}$$

where $[z(D)]_0 = z_0$ with $z(D) = \sum_k z_k D^k$. The equality in (3.15) holds due to the assumption that X_n is an i.i.d random variable and the self-interference term is removed from the expression $1 - 1/g^*(D^{-*})$.

Since $\mathrm{Var}(V_f) = \mathrm{Var}(V_b)$, the linear MMSE combiner of [59, 62] becomes $Y = \frac{1}{2}(Y_f + Y_b)$. Naturally, the MSE and SNR of the unbiased BiDFE are given as

$$\begin{aligned}
\mathrm{MSE}_{UBiDFE} &= \frac{(1 + \mathrm{Re}[\rho])}{2} \mathrm{MSE}_{UDFE} \\
&= \frac{(1 + \mathrm{Re}[\rho]) P_X N_0}{2(P_0 - N_0)} \qquad (3.17) \\
\mathrm{SNR}_{UBiDFE} &\triangleq \frac{P_X}{\mathrm{MSE}_{UBiDFE}} = \frac{2}{(1 + \mathrm{Re}[\rho])} \mathrm{SNR}_{UDFE} \\
&= \frac{2(P_0 - N_0)}{(1 + \mathrm{Re}[\rho]) N_0} \qquad (3.18)
\end{aligned}$$

where $\mathrm{Re}[\rho]$ denotes the real part of ρ.

Note that the infinite-length normal/time-reversed MMSE-DFE and BiDFE analyzed here do not exploit the *a priori* information of X_n. In other words, the feedforward and feedback filters of DFE are derived by assuming $\mathrm{E}(X_n) = 0$ for all n, meaning that the calculated SNR performance would reflect the non-turbo ideal-decision BiDFE performance with time-invariant filter taps of Section 2.4.1.

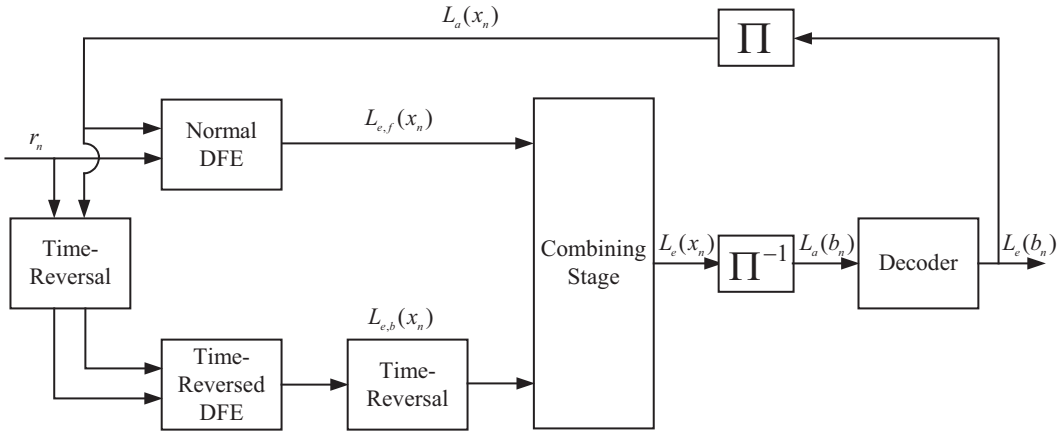


Figure 3.2: Turbo BiDFE Scheme.

3.2 Iterative BiDFE Algorithm

We now discuss an iterative BiDFE algorithm. Iterative equalization schemes based on BiDFE are shown in Fig. 3.2. Basically, the channel equalizer is a SISO equalizer which employs the normal forward DFE, the time-reversed DFE and an LLR combining block. The received data sequence is equalized in both directions by the two DFEs, and the extrinsic information from two DFEs are combined and passed to the error correction code decoder. We show that a proper combining of the two sets of extrinsic information can suppress error propagation and noise further and generate more reliable extrinsic information for the outer decoder.

3.2.1 Combining Extrinsic Information

Similarly to the finite-length time-varying feedforward and feedback filter of the normal DFE at time index n, which are previously defined as $\mathbf{c}_n$ in (2.8) and $\mathbf{d}_n$ in (2.9), we also define the finite-length time-varying feedforward and feedback filter of the time-reversed DFE at time index n as $\tilde{\mathbf{c}}_n$ and $\tilde{\mathbf{d}}_n$ with the same lengths as $\mathbf{c}_n$ and $\mathbf{d}_n$ respectively. Note that $\tilde{\mathbf{c}}_n$ and $\tilde{\mathbf{d}}_n$ are defined in a similar way as (2.8) and (2.9) except that the

channel convolution matrix $\tilde{\mathbf{H}}$ for the time-reversed channel is given as

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} h_0 & h_1 & \cdots & h_{L_h-1} & 0 & \cdots & & 0 \\ 0 & h_0 & h_1 & \cdots & h_{L_h-1} & 0 & \cdots & 0 \\ & \ddots & & \ddots & & \ddots & & \\ 0 & 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{L_h-1} \end{bmatrix}.$$

The unbiased equalizer output [52] corresponding to the transmitted coded symbol from the the normal (forward) and the time-reversed (backward) DFE can be represented respectively as

$$Y_{f,n} = X_n + I_{f,n} + V_{f,n} \tag{3.19}$$

$$Y_{b,n} = X_n + I_{b,n} + V_{b,n} \tag{3.20}$$

where $X_n \triangleq x_n$, $V_{f,n} \triangleq v_{f,n}/p_{\{n,0\}}$ and $I_{f,n} \triangleq i_{f,n}/p_{\{n,0\}}$. Also, $V_{b,n} \triangleq v_{b,n}/\tilde{p}_{\{n,0\}}$ and $I_{b,n} \triangleq i_{b,n}/\tilde{p}_{\{n,0\}}$ where $v_{b,n}$ and $i_{b,n}$ are defined similarly to the normal DFE and $\tilde{p}_{\{n,0\}} = \tilde{\mathbf{c}}_n^T \tilde{\mathbf{s}}$ where $\tilde{\mathbf{s}} \triangleq \tilde{\mathbf{H}}[\mathbf{0}_{1\times L_d}, 1, \mathbf{0}_{1\times L_c}]^T$. For notational simplicity, we further drop time index n with an understanding that processing remains identical as n progresses: $Y_f = X + I_f + V_f$ and $Y_b = X + I_b + V_b$.

Now, we discuss the problem of how to combine the extrinsic information from two DFEs. Initially, let us consider two unbiased equalizer outputs, which are corrupted by AWGN, corresponding to the transmitted coded symbol X:

$$Y_f = X + U_f$$

$$Y_b = X + U_b$$

where the noise U_f and U_b are assumed to be zero mean Gaussian random variables which are independent of the coded data X but correlated with each other with correlation coefficient ρ.

In order to combine the extrinsic information, it is beneficial to whiten the noise U_f

and U_b before combining. The noise correlation matrix $\mathbf{R}$ is defined as

$$\mathbf{R} \triangleq \begin{bmatrix} \mathrm{Var}(U_f) & \mathrm{E}(U_f U_b) \\ \mathrm{E}(U_f U_b) & \mathrm{Var}(U_b) \end{bmatrix} = \begin{bmatrix} N_f & \rho\sqrt{N_f N_b} \\ \rho\sqrt{N_f N_b} & N_b \end{bmatrix}$$

where $N_f \triangleq \mathrm{Var}(U_f)$ and $N_b \triangleq \mathrm{Var}(U_b)$. Then, the eigenvalues of the noise correlation matrix, λ_1 and λ_2, with their corresponding normalized eigenvectors $\mathbf{g}_1$ and $\mathbf{g}_2$ are given by

$$\lambda_1 = \frac{(N_f + N_b) + \sqrt{(N_f - N_b)^2 + 4\rho^2 N_f N_b}}{2}$$
$$\lambda_2 = \frac{(N_f + N_b) - \sqrt{(N_f - N_b)^2 + 4\rho^2 N_f N_b}}{2}$$
$$\mathbf{g}_1 = \frac{1}{\sqrt{g_{11}^2 + g_{21}^2}} \begin{bmatrix} g_{11} \\ g_{21} \end{bmatrix}, \; \mathbf{g}_2 = \frac{1}{\sqrt{g_{12}^2 + g_{22}^2}} \begin{bmatrix} g_{12} \\ g_{22} \end{bmatrix}$$

where $g_{21} = g_{22} = \rho\sqrt{N_f N_b}$, $g_{11} = \frac{1}{2}\Big[(N_f - N_b) + \sqrt{(N_f - N_b)^2 + 4\rho^2 N_f N_b}\Big]$, and $g_{12} = \frac{1}{2}\Big[(N_f - N_b) - \sqrt{(N_f - N_b)^2 + 4\rho^2 N_f N_b}\Big]$. It is easy to see that the noise correlation matrix $\mathbf{R}$ is non-singular unless $\rho = \pm 1$. If $\mathbf{R}$ is non-singular, $\mathbf{R}$ can be expanded as $\mathbf{R} = \mathbf{G}\mathbf{\Lambda}\mathbf{G}^{-1}$ where $\mathbf{G} \triangleq [\mathbf{g}_1 \, \mathbf{g}_2]$ and $\mathbf{\Lambda} \triangleq \mathrm{Diag}(\lambda_1, \lambda_2)$. Since $\mathbf{G}$ is a unitary matrix, the noise whitening matrix is $\mathbf{A} \triangleq [\mathbf{a}_1 \, \mathbf{a}_2] = \mathbf{G}^{-1} = \mathbf{G}^T$ where $\mathbf{a}_1 \triangleq [a_{11} \, a_{21}]^T$ and $\mathbf{a}_2 \triangleq [a_{12} \, a_{22}]^T$. So, given the equalized output vector $\mathbf{Y} \triangleq [Y_f, Y_b]^T$, the whitened vector is $\mathbf{Y}' \triangleq [Y_f', Y_b']^T = \mathbf{A}\mathbf{Y}$ with the new noise correlation matrix $\mathbf{R}' = \mathbf{A}\mathbf{R}\mathbf{A}^T = \mathbf{\Lambda}$.

Finally, the extrinsic information of X can be expressed as

$$\begin{aligned}
L_e(X) &= \ln \frac{\Pr(Y_f, Y_b \mid X=+1)}{\Pr(Y_f, Y_b \mid X=-1)} \\
&= \ln \frac{\Pr(Y'_f, Y'_b \mid X=+1)}{\Pr(Y'_f, Y'_b \mid X=-1)} \\
&= \ln \frac{\Pr(Y'_f \mid X=+1)}{\Pr(Y'_f \mid X=-1)} + \ln \frac{\Pr(Y'_b \mid X=+1)}{\Pr(Y'_b \mid X=-1)} \\
&= \frac{2(a_{11}+a_{12})Y'_f}{\lambda_1} + \frac{2(a_{21}+a_{22})Y'_b}{\lambda_2} \\
&= \frac{2\left(N_b - \rho\sqrt{N_f N_b}\right) Y_f}{(1-\rho^2)\, N_f N_b} + \frac{2\left(N_f - \rho\sqrt{N_f N_b}\right) Y_b}{(1-\rho^2)\, N_f N_b} \\
&= \frac{\left(N_b - \rho\sqrt{N_f N_b}\right)}{(1-\rho^2)\, N_b} L_{e,f}(X) + \frac{\left(N_f - \rho\sqrt{N_f N_b}\right)}{(1-\rho^2)\, N_f} L_{e,b}(X). \qquad (3.21)
\end{aligned}$$

For the singular noise correlation matrix $\mathbf{R}$ (i.e., $\rho = +1$), $N_f = N_b = N$ and $Y_f = Y_b = Y$ so that $L_{e,f}(X) = L_{e,b}(X)$. Consequently, the extrinsic information of X becomes $L_e(X) = 2Y/N = (L_{e,f}(X) + L_{e,b}(X))/2$. Note that the mean combiner of [58], $L_e(X) = (L_{e,f}(X) + L_{e,b}(X))/2$, can be considered as the proposed combiner with $\rho = +1$. If $\rho = -1$, $U_f = -U_b$ and we can cancel out the noise perfectly by averaging the outputs: $(Y_f + Y_b)/2$. The extrinsic information of X in this case is $L_e(X) = +\infty$ when $(Y_f + Y_b)/2 \geq 0$ while $L_e(X) = -\infty$ when $(Y_f + Y_b)/2 < 0$.

3.2.2 Reducing the Combiner Sensitivity to the Estimation Error

Let us consider the effect of errors in estimating ρ on extrinsic information. Write $\hat{\rho} = \rho + \varepsilon$ where ε is the estimation error. Then, the sensitivity of the combiner in (3.21) to the estimation error can be defined as

$$\begin{aligned}
\mathrm{S}(\rho) &\triangleq \left| \frac{\partial L_e(X)}{\partial \rho} \right| \\
&= \left| \frac{\left(2\rho N_b - (1+\rho^2)\sqrt{N_f N_b}\right)}{(1-\rho^2)^2\, N_b} L_{e,f}(X) + \frac{\left(2\rho N_f - (1+\rho^2)\sqrt{N_f N_b}\right)}{(1-\rho^2)^2\, N_f} L_{e,b}(X) \right|
\end{aligned}$$

which approaches infinity as $\rho \to \pm 1$. This means that the combiner of (3.21) is unfortunately very sensitive to the correlation estimator error, as the magnitude of the

correlation becomes large.

The sensitivity of the combiner can be reduced if we assume that the variance of U_f and U_b are the same, i.e., $N_f = N_b$. This assumption is reasonable when the same feedforward and feedback filter length is used in both DFEs. Then, from (3.21), the combined extrinsic information of X for non-singular $\mathbf{R}$ is simply given as

$$L_e(X) = \frac{1}{(1+\rho)}\Big(L_{e,f}(X) + L_{e,b}(X)\Big) \tag{3.22}$$

with the sensitivity to the correlation estimation error

$$\mathrm{S}(\rho) = \left|\frac{1}{(1+\rho)^2}\Big(L_{e,f}(X) + L_{e,b}(X)\Big)\right|.$$

Although the sensitivity of this combiner to the estimation error also goes to infinity as $\rho \to -1$, it is more robust as $\rho \to +1$ since $\lim_{\rho\to+1} \mathrm{S}(\rho) = |(L_{e,f}(X) + L_{e,b}(X))/4|$.

3.2.3 Application to the BiDFE Algorithm

Although the composite noise $I_{f,n}+V_{f,n}$ and $I_{b,n}+V_{b,n}$ are not Gaussian, we exploit the combiner of (3.22) in order to produce the combined extrinsic information to be passed to the convolutional decoder. The noise correlation coefficient between $I_{f,n} + V_{f,n}$ and $I_{b,n} + V_{b,n}$ is naturally defined as

$$\rho_n \triangleq \frac{\mathrm{E}\left\{\left(I_{f,n} - \mathrm{E}(I_{f,n}) + V_{f,n}\right)\left(I_{b,n} - \mathrm{E}(I_{b,n}) + V_{b,n}\right)\right\}}{\sqrt{\left(\mathrm{Var}(I_{f,n}) + \mathrm{Var}(V_{f,n})\right)\left(\mathrm{Var}(I_{b,n}) + \mathrm{Var}(V_{b,n})\right)}}.$$

Unfortunately, it is difficult to compute the correlation coefficient analytically in the presence of decision feedback errors. However, assuming that the noise is stationary, we have $\rho_n = \rho$ and the correlation coefficient can be estimated through time-averaging such as (3.23) where the summations are over some reasonably large finite window. Note that the hard decisions for the transmitted symbols in normal and time-reversed DFEs might be different; in estimating the correlation coefficient, we only consider those noise samples for which $\hat{X}_{f,n}$ and $\hat{X}_{b,n}$ are identical.

$$\hat{\rho} = \frac{\sum \left\{ (Y_{f,n} - \hat{X}_{f,n} - \mathrm{E}(I_{f,n}))(Y_{b,n} - \hat{X}_{b,n} - \mathrm{E}(I_{b,n})) \right\}}{\sqrt{\sum (Y_{f,n} - \hat{X}_{f,n} - \mathrm{E}(I_{f,n}))^2} \sqrt{\sum (Y_{b,n} - \hat{X}_{b,n} - \mathrm{E}(I_{b,n}))^2}} \tag{3.23}$$

Let us summarize our LLR combining method: 1) The extrinsic information $L_{e,f}(X_n)$ and $L_{e,b}(X_n)$ for $n = 0, 1, \ldots, N-1$ are acquired according to (2.26) in the normal and time-reversed MMSE-DFE settings. 2) Estimate the noise correlation coefficient, $\hat{\rho}$, between $I_{f,n}+V_{f,n}$ and $I_{b,n}+V_{b,n}$ by (3.23). 3) Generate the combined extrinsic information $L_e(X_n)$ according to (3.22) with $\rho_n = \hat{\rho}$.

3.2.4 Correlation Analysis under Ideal Feedback

We provide correlation analysis in the following. The analysis will allow validation of (3.23) in different scenarios. The observation of how the simulated correlation coefficient (3.23) converges to the analytically computed one under the assumptions of ideal feedback and perfect *a priori* information will also provide useful insights into the iterative behaviour of the proposed turbo BiDFE.

First of all, the noise variance of $V_{f,n}$ and $V_{b,n}$ from the time-varying filters are:

$$\mathrm{Var}(V_{f,n}) = (1 - \mathbf{s}^T \mathbf{c}_n)/\mathbf{c}_n^T \mathbf{s}$$
$$\mathrm{Var}(V_{b,n}) = (1 - \tilde{\mathbf{s}}^T \tilde{\mathbf{c}}_n)/\tilde{\mathbf{c}}_n^T \tilde{\mathbf{s}}.$$

When we assume ideal decision feedback, $\Pr(I_f = 0) = \Pr(I_b = 0) = 1$ so that $I_{f,n} =$

$I_{b,n} = 0$, the noise correlation coefficient ρ_n between $V_{f,n}$ and $V_{b,n}$ becomes

$$\begin{aligned}
\rho_n &\triangleq \frac{\mathrm{E}(V_{f,n}V_{b,n})}{\sqrt{\mathrm{Var}(V_{f,n})\mathrm{Var}(V_{b,n})}} \\
&= \frac{\mathrm{E}\left[\sum_{j=0}^{L_c} c_{\{n,j\}} w_{n+j} \sum_{k=0}^{L_c} \tilde{c}_{\{n,k\}} w_{n-k+L_h-1}\right]}{\sqrt{\mathbf{c}_n^T\mathbf{s}(1-\mathbf{s}^T\mathbf{c}_n)}\sqrt{\tilde{\mathbf{c}}_n^T\tilde{\mathbf{s}}(1-\tilde{\mathbf{s}}^T\tilde{\mathbf{c}}_n)}} \qquad (3.24) \\
&= \frac{\sum_{j=0}^{L_c}\sum_{k=0}^{L_c} c_{\{n,j\}}\tilde{c}_{\{n,k\}}\mathrm{E}\left[w_{n+j}w_{n-k+L_h-1}\right]}{\sqrt{\mathbf{c}_n^T\mathbf{s}(1-\mathbf{s}^T\mathbf{c}_n)}\sqrt{\tilde{\mathbf{c}}_n^T\tilde{\mathbf{s}}(1-\tilde{\mathbf{s}}^T\tilde{\mathbf{c}}_n)}} \\
&= N_0\left(\frac{\sum_{j=0}^{L_c}\sum_{k=0}^{L_c} c_{\{n,j\}}\tilde{c}_{\{n,k\}}\delta(j+k+1-L_h)}{\sqrt{\mathbf{c}_n^T\mathbf{s}(1-\mathbf{s}^T\mathbf{c}_n)}\sqrt{\tilde{\mathbf{c}}_n^T\tilde{\mathbf{s}}(1-\tilde{\mathbf{s}}^T\tilde{\mathbf{c}}_n)}}\right) \qquad (3.25)
\end{aligned}$$

where $\delta(t)$ is defined as: if $t = 0$, $\delta(t) = 1$; otherwise, $\delta(t) = 0$. The equality in (3.24) holds because X_n is an i.i.d random variable.

If the time-invariant filters are used instead of the time-varying filters, the variances of $V_{f,n}$ and $V_{b,n}$ become

$$\begin{aligned}
\mathrm{Var}(V_{f,n}) &= \mathbf{c}^T\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{c}/\left(\mathbf{c}^T\mathbf{s}\right)^2 \\
\mathrm{Var}(V_{b,n}) &= \tilde{\mathbf{c}}^T\left(\tilde{\mathbf{H}}\tilde{\boldsymbol{\Sigma}}_n\tilde{\mathbf{H}}^T - z_n\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T + N_0\mathbf{I}\right)\tilde{\mathbf{c}}/\left(\tilde{\mathbf{c}}^T\tilde{\mathbf{s}}\right)^2.
\end{aligned}$$

Then, the noise correlation coefficient can be also obtained as below.

$$\rho_n = N_0\left(\frac{\sum_{j=0}^{L_c}\sum_{i=0}^{L_c} c_j\tilde{c}_i\delta(j+i+1-L_h)}{\sqrt{\mathbf{c}^T\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{c}}\sqrt{\tilde{\mathbf{c}}^T(\tilde{\mathbf{H}}\tilde{\boldsymbol{\Sigma}}_n\tilde{\mathbf{H}}^T - z_n\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T + N_0\mathbf{I})\tilde{\mathbf{c}}}}\right) \qquad (3.26)$$

Similarly, if the iteration-varying filters are used, the variances of $V_{f,n}$ and $V_{b,n}$ become

$$\begin{aligned}
\mathrm{Var}(V_{f,n}) &= \mathbf{c}^{(k)T}\left(\mathbf{H}\boldsymbol{\Sigma}_n\mathbf{H}^T - z_n\mathbf{s}\mathbf{s}^T + N_0\mathbf{I}\right)\mathbf{c}^{(k)}/\left(\mathbf{c}^{(k)T}\mathbf{s}\right)^2 \\
\mathrm{Var}(V_{b,n}) &= \tilde{\mathbf{c}}^{(k)T}\left(\tilde{\mathbf{H}}\tilde{\boldsymbol{\Sigma}}_n\tilde{\mathbf{H}}^T - z_n\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T + N_0\mathbf{I}\right)\tilde{\mathbf{c}}^{(k)}/\left(\tilde{\mathbf{c}}^{(k)T}\tilde{\mathbf{s}}\right)^2.
\end{aligned}$$

Accordingly, the noise correlation coefficient is given as

$$\rho_n = N_0 \left(\frac{\sum_{j=0}^{L_c} \sum_{i=0}^{L_c} c_j^{(k)} \tilde{c}_i^{(k)} \delta(j+i+1-L_h)}{\sqrt{\mathbf{c}^{(k)T} \left(\mathbf{H}\mathbf{\Sigma}_n \mathbf{H}^T - z_n \mathbf{s}\mathbf{s}^T + N_0 \mathbf{I} \right) \mathbf{c}^{(k)}} \sqrt{\tilde{\mathbf{c}}^{(k)T} (\tilde{\mathbf{H}}\tilde{\mathbf{\Sigma}}_n \tilde{\mathbf{H}}^T - z_n \tilde{\mathbf{s}}\tilde{\mathbf{s}}^T + N_0 \mathbf{I}) \tilde{\mathbf{c}}^{(k)}}} \right) \tag{3.27}$$

Now, let us consider some special cases.

No *A Priori* Information

When no *a priori* information is available, i.e., $\mathrm{E}(X_n) = 0$ for all n, the feedforward and feedback filters are the same as the time-invariant filters and the noise variances are stationary:

$$\mathrm{Var}(V_{f,n}) = \mathrm{Var}(V_f) = (1 - \mathbf{s}^T \mathbf{c})/\mathbf{c}^T \mathbf{s}$$
$$\mathrm{Var}(V_{b,n}) = \mathrm{Var}(V_b) = (1 - \tilde{\mathbf{s}}^T \tilde{\mathbf{c}})/\tilde{\mathbf{c}}^T \tilde{\mathbf{s}}.$$

Therefore, the noise correlation coefficient is given by

$$\rho_n = \rho = N_0 \left(\frac{\sum_{j=0}^{L_c} \sum_{k=0}^{L_c} c_j \tilde{c}_k \delta(j+k+1-L_h)}{\sqrt{\mathbf{c}^T \mathbf{s}(1 - \mathbf{s}^T \mathbf{c})} \sqrt{\tilde{\mathbf{c}}^T \tilde{\mathbf{s}}(1 - \tilde{\mathbf{s}}^T \tilde{\mathbf{c}})}} \right). \tag{3.28}$$

We observed that the noise correlation coefficient of the infinite-length BiDFE in (3.16) is almost identical to that of the finite-length BiDFE in (3.28) when L_c is chosen to be long enough.

Time-varying/Iteration-varying Filters with Perfect *A Priori* Information

When several iterations are performed at high SNRs in turbo equalization, the perfect *a priori* information could be available, i.e., $\mathrm{E}(X_n) = X_n$ for all n. When $\mathrm{E}(X_n) = X_n$ for all n, the time-varying and the iteration-varying feedforward filters of two DFEs become

the normalized matched filters corresponding to the forward and reverse channel impulse responses:

$$\mathbf{c}_n = \mathbf{c}^{(k)} = A\left[h_0, h_1, \ldots, h_{L_h-1}, \mathbf{0}_{1\times L_c-L_h+1}\right]^T$$
$$\tilde{\mathbf{c}}_n = \tilde{\mathbf{c}}^{(k)} = A\left[h_{L_h-1}, h_{L_h-2}, \ldots, h_0, \mathbf{0}_{1\times L_c-L_h+1}\right]^T$$

where A is a real-valued constant depending on SNR, i.e., $A \triangleq P_X/(N_0+P_X\sum_{k=0}^{L_h-1}|h_k|^2)$. Moreover, since the first terms of $V_{f,n}$ and $V_{b,n}$ disappear, the noise variances are simply:

$$\mathrm{Var}(V_{f,n}) = \mathrm{Var}(V_f) = \frac{N_0\mathbf{c}_n^T\mathbf{c}_n}{(\mathbf{c}_n^T\mathbf{s})^2} = \frac{N_0A^2}{(\mathbf{c}_n^T\mathbf{s})^2}\sum_{k=0}^{L_h-1}|h_k|^2$$
$$\mathrm{Var}(V_{b,n}) = \mathrm{Var}(V_b) = \frac{N_0\tilde{\mathbf{c}}_n^T\tilde{\mathbf{c}}_n}{(\tilde{\mathbf{c}}_n^T\tilde{\mathbf{s}})^2} = \frac{N_0A^2}{(\tilde{\mathbf{c}}_n^T\tilde{\mathbf{s}})^2}\sum_{k=0}^{L_h-1}|h_k|^2.$$

Accordingly, the noise correlation coefficient is

$$\rho_n = \rho = 1. \tag{3.29}$$

Note that the noise correlation coefficient ρ with perfect *a priori* information converges to 1 regardless of the SNR value. As will be shown shortly, the measured correlation coefficient using simulated turbo BiDFE outputs indeed approaches 1, as turbo iteration progresses. This indicates that both assumptions - ideal decision feedback and perfect *a priori* information - are reasonable.

Time-invariant Filters with Perfect *A Priori* Information

When the time-invariant filters are used with perfect *a priori* information, the time-invariant DFEs yield the noise variances as

$$\mathrm{Var}(V_{f,n}) = \mathrm{Var}(V_f) = N_0\mathbf{c}^T\mathbf{c}/(\mathbf{c}^T\mathbf{s})^2$$
$$\mathrm{Var}(V_{b,n}) = \mathrm{Var}(V_b) = N_0\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}/(\tilde{\mathbf{c}}^T\tilde{\mathbf{s}})^2.$$

The noise correlation coefficient is also simply given by

$$\rho_n = \rho = \frac{\sum_{j=0}^{L_c}\sum_{k=0}^{L_c}c_j\tilde{c}_k\delta(j+k+1-L_h)}{\sqrt{\mathbf{c}^T\mathbf{c}}\sqrt{\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}}}. \tag{3.30}$$

As will be discussed in the next section, in the simulation of turbo BiDFE with time-invariant taps it is observed that the BiDFE output correlation does indeed converge to (3.30), indicating again that the assumptions of error-free decisions and perfect *a priori* information are reasonable.

3.2.5 Complexity Comparison

The BiDFE equalizers, including the proposed BiDFE methods, require roughly twice as many operations as the DFE counterparts, due to the presence of the time-reversed filter components. Most notably, while the complexity of the proposed BiDFE with time-invariant filters is considerably lower than that of the MAP equalizer as well as the LE and DFE with time-varying filter taps, the performance is significantly better than the LE and DFE with time-varying filters as shown shortly.

3.2.6 Performance Comparison

Same as the previous SNR/MI analysis for LE and DFE, let us consider the ideal condition for the equalizer where the perfect *a priori* information is available, i.e., $\mathrm{E}(X_n) = X_n$ (or $L_a(X_n) = \pm\infty$) for all n. Then, the BiDFE with the time-varying and the iteration-varying filters also achieves the MFB since the noise correlation coefficient between the normal DFE and the time-reversed DFE is equal to 1, which means they produce the same equalized output and there is no SNR advantage of combining. Finally, the unbiased output SNRs with the time-varying and the iteration-varying filters are given by

$$\mathrm{SNR}_{\infty,\mathrm{LE}} = \mathrm{SNR}_{\infty,\mathrm{DFE}} = \mathrm{SNR}_{\infty,\mathrm{BiDFE}} = \mathrm{MFB} = \frac{P_X}{N_0} \tag{3.31}$$

However, although all equalizer schemes can achieve the MFB with the ideal condition, their realized performance is different [24].

On the other hand, the maximum achievable unbiased output SNRs with the time-invariant filters is given as

$$\mathrm{SNR}_{\infty,\mathrm{BiDFE}} = \left(\frac{2}{1+\rho}\right)\mathrm{SNR}_{\infty,\mathrm{DFE}} = \frac{P_X}{N_0}\left(\frac{2(\mathbf{c}^T\mathbf{s})^2}{(1+\rho)\mathbf{c}^T\mathbf{c}}\right) \tag{3.32}$$

where ρ is the noise correlation coefficient between the normal DFE and the time-reversed DFE under perfect *a priori* information and is given in (3.30).

The value inside the parentheses of (3.32) is also less than 1 unless the channel is the ideal AWGN channel or the channel input SNR is zero, i.e., $P_X/N_0 = 0$; thus, these output SNRs are also less than the MFB. However, since the output SNR by the BiDFE is close to the MFB even when time-invariant filters are used [24], its performance is not degraded as badly as in the LE or the DFE. In other words,

$$\mathrm{SNR}_{\infty,\mathrm{LE}} \le \mathrm{SNR}_{\infty,\mathrm{DFE}} \le \mathrm{SNR}_{\infty,\mathrm{BiDFE}} \le \mathrm{MFB} = \frac{P_X}{N_0} \tag{3.33}$$

where equalities only hold on the ideal AWGN channel or at zero channel input SNR. Again, since the Gaussian noise term only remains in the equalized output, the analytical maximum attainable MI of BiDFE with ideal condition is given as

$$C_b(R) \triangleq 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log_2\left\{1 + e^{-2\sqrt{R}\tau - 2R}\right\} d\tau \tag{3.34}$$

where $R = \mathrm{SNR}_{\infty,\mathrm{BiDFE}}$.

3.3 Numerical Results

In this section, simulation results of iterative BiDFE schemes are presented under the same simulation conditions of Section 2.6. We also evaluate the performance of the iterative equalizers over the ISI channel $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$ investigated in [20]. The second ISI channel, $\mathbf{h_2}$, introduces more severe ISI than $\mathbf{h_1}$, as shown in Fig. 3.3. For the ISI channel $\mathbf{h_2}$, the DFE with 21 feedforward taps and 6 feedback taps ($L_c = 20$ and $L_d = 6$) is used for both the normal and the time-reversed DFEs while the LE uses 27 taps ($L_{f1} = 13$ and $L_{f2} = 13$).

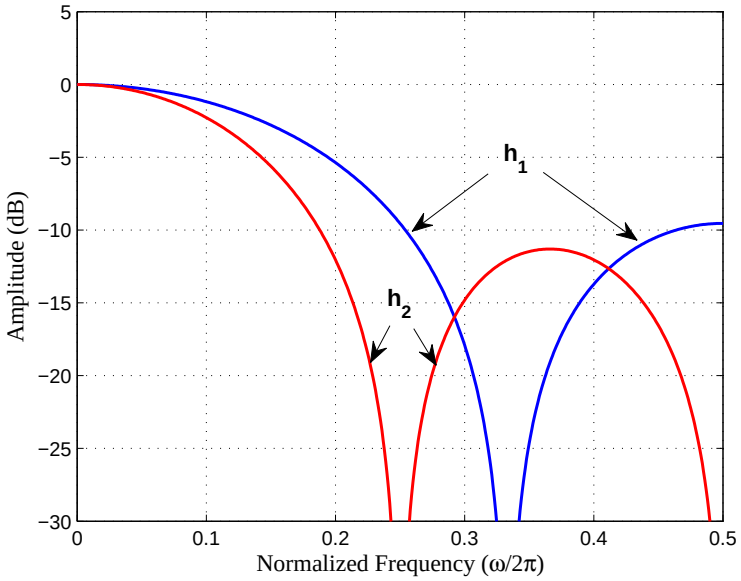


Figure 3.3: Frequency Response of the ISI Channels: $\mathbf{h_1} = (1/\sqrt{19})[1 \quad 2 \quad 3 \quad 2 \quad 1]^T$ and $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$.

The notations in Section 2.6 are also adopted here, but, the "Proposed BiDFE" is the iterative BiDFE algorithm which is described in Section 3.2.3. In other words, "Proposed BiDFE" uses the the proposed LLR generation for both normal and time-reversed DFEs along with the proposed extrinsic information combiner of (3.22) in conjunction with the noise correlation coefficient of (3.23). The "BiDFE (mean combiner)" is the iterative BiDFE algorithm with the conventional LLR mapping and the mean combiner, $L_e(X) = (L_{e,f}(X) + L_{e,b}(X))/2$ (of [58]), simulated for performance comparison purposes.

Figs. 3.4, 3.5, and 3.6 show the performance of several turbo equalizers with time-varying filters, time-invariant, and iteration-varying filters after 20 iterations. The "Proposed BiDFE" is considerably better than the BiDFE based on the mean combiner, the proposed DFE, and the LE, approaching the performance of the MAP scheme. Especially, as the Fig. 3.5 indicates, the performance of "Proposed TI-BiDFE" is very close

to the performance of the MAP equalizer while achieving a low computational complexity based on the use of time-invariant filters. Also notice that "Proposed TI-BiDFE" achieve decision-error-free performance at low BERs, indicating the error propagation effect has been nearly eliminated using the proposed LLR generation method and LLR combining method. It is noteworthy that the proposed BiDFE algorithm still provides near-optimal performance even with the time-invariant filter taps. While the TI-BiDFE based on the existing mean combiner appears to perform almost as well, the EXIT chart analysis to be discussed below indicate that with a smaller number of turbo iterations, its performance is distinctly inferior to the proposed TI-BiDFE based on the new combining method.

Figs. 3.7, 3.8, and 3.9 show the same set of simulation results now applied to the more severe ISI channel $\mathbf{h_2}$. While all filter-based schemes lag clearly behind the BCJR-based scheme at the error rates simulated, the proposed BiDFE scheme with the time-varying, time-invariant, and iteration-varying filter cases outperform the DFE and LE scheme by a significant margin. In fact, in this severe channel the BER curve of the DFE and LE scheme, even with time-varying filters, appears to diverge considerably from the ideal no-ISI curve. Overall, the proposed BiDFE based on time-invariant filter taps offer excellent performance-complexity trade-off.

When the performance of LE with different filter types are compared for channel $\mathbf{h_2}$, the iteration-varying filters sometimes provide even better BER performance than the time-varying filters which is contradictory to the simulation results of [13] (the exact MMSE shows a slightly better performance than the LC approximate MMSE as shown in Fig. 3 of [13]). This is mainly due to the fact that when inaccurate *a priori* information arrives, the optimal time-varying filters more easily fail to produce reliable extrinsic information than the IV filters. In other word, the time-varying (or the exact MMSE) filter solutions are designed regardless of the quality of the *a priori* information so that an equalizer with time-varying filters can erroneously maximize the extrinsic LLRs (or minimize the MSEs) based on the optimally but incorrectly derived MMSE filters

by the wrong *a priori* LLRs. Note that the MSE comparison in [13] is based on the premise that the incoming *a priori* information is reliable, but, this assumption is not always guaranteed especially on the severe ISI channels, i.e., there are high possibilities that the decoder generates incorrect *a priori* LLRs to the equalizer when ISI is severe. Furthermore, a few erroneously maximized extrinsic LLRs by the equalizer can degrade the overall turbo equalization performance since they can remove the possibility of the correct signal path in the decoder. This phenomenon can be also explained with the EXIT chart of Fig. 3.17. As the trajectory of the TV-LE shows for each stair, the TV filter solutions tend to maximize the extrinsic LLRs quality based on the given *a priori* LLRs. However, its side effect is also to generate the maximized wrong extrinsic LLRs with the potential inaccuracies in the *a priori* information by the decoder. On the other hand, the iteration-varying (or the LC approximate MMSE) filter solutions generate the extrinsic LLRs moderately, thus, the erroneously generated LLRs are also moderately controlled. In short, the LLRs with the time-varying filters ascend steeply but riskily while the LLRs with the iteration-varying filters advance slowly but steadily. Note that an equalizer with iteration-varying filters can also achieve MFB performance if enough turbo iterations are performed. In the case of the BiDFE, this sensitivity plays an even bigger role in determining the overall performance. As compared, "TI-BiDFE" shows better performance than BiDFE with any other filter types since it would be better not to update the filter taps at all when the *a priori* information is unreliable at low SNRs.

The noise correlation in one block of coded data bits is described in Fig. 3.10, at different iteration numbers at a 6 dB SNR on $\mathbf{h_1}$. The correlation coefficient of "Proposed TV-BiDFE" and "Proposed IV-BiDFE" go to 1 as the number of iterations increases because the *a priori* information from the decoder becomes reliable, and the time-varying and iteration-varying filters in the normal and the time-reversed DFEs produce essentially the same equalized output sequences. This phenomenon of Fig. 3.10 accurately reflects the result of (3.29). On the other hand, the correlation coefficient of "Proposed TI-BiDFE" actually decreases as the number of iterations increases, and the

noise correlation coefficient converges to that of "TI-BiDFE with Ideal Feedback" or the correlation coefficient of (3.30). This is because the decision feedback errors disappear and the perfect *a priori* information is available from decoder. Note that the filter coefficients in both DFEs do not change with the *a priori* information.

Figs. 3.11, 3.12, and 3.13 show the EXIT chart corresponding to time-varying-filter-based, time-invariant-filter-based, and iteration-varying-filter-based equalizers for $\mathbf{h_1}$ at a 7 dB SNR while Figs. 3.14, 3.15, and 3.16 show the similar EXIT charts for $\mathbf{h_2}$ at a 10 dB SNR. As the figures show, the trajectories of the "Proposed BiDFE" indicate that this scheme moves from 0 bit of mutual information to 1 bit with a less number of iteration runs than "Proposed DFE" and "LE". We notice, however, that the proposed BiDFE scheme requires more iterations in achieving the full performance, relative to the MAP equalizer (whose trajectory is not shown to avoid cluttering). Nevertheless, the proposed BiDFE method offers a reasonable tradeoff among complexity, performance, and latency.

Finally, Fig. 3.18 shows the SNR comparison at the output of the unbiased DFE and BiDFE assuming ideal feedback on the channel $\mathbf{h_1}$ when the *a priori* information is not available. As the figure shows, the output SNR of BiDFE is considerably higher than the output SNR of DFE but with a certain gap to the MFB.

3.4 Summary

In this chapter, we proposed turbo equalization methods based on BiDFE structures. When employing the previously developed LLR mapping method for DFE and an LLR combining method that estimates the correlation between the forward and backward DFE outputs and whitens them, the resulting performance is remarkably good given the simple structure of the BiDFE, relative to that of the BCJR equalizer. The proposed LLR generation and combining methods remain effective even when a time-invariant constraint is imposed on the feedforward and feedback filters of the DFEs. Overall,

the proposed BiDFE method based on time-invariant filter taps provides the excellent performance-complexity tradeoff for severe ISI channels where other filter based approaches fail to operate adequately.

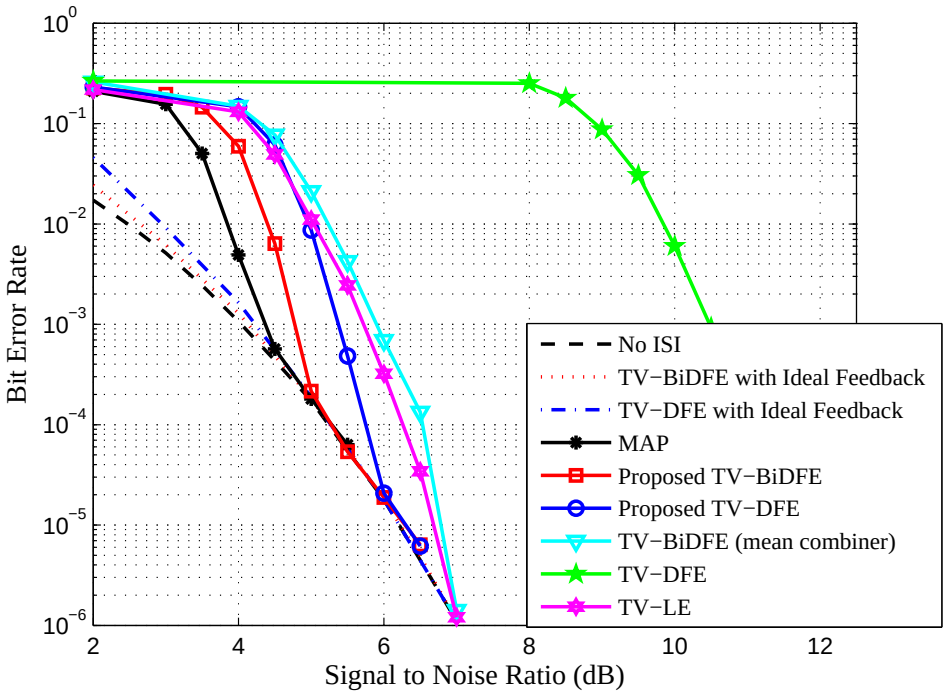


Figure 3.4: BER Curve on the Channel $\mathbf{h_1}$ after 20 Iterations with Time-varying Filters.

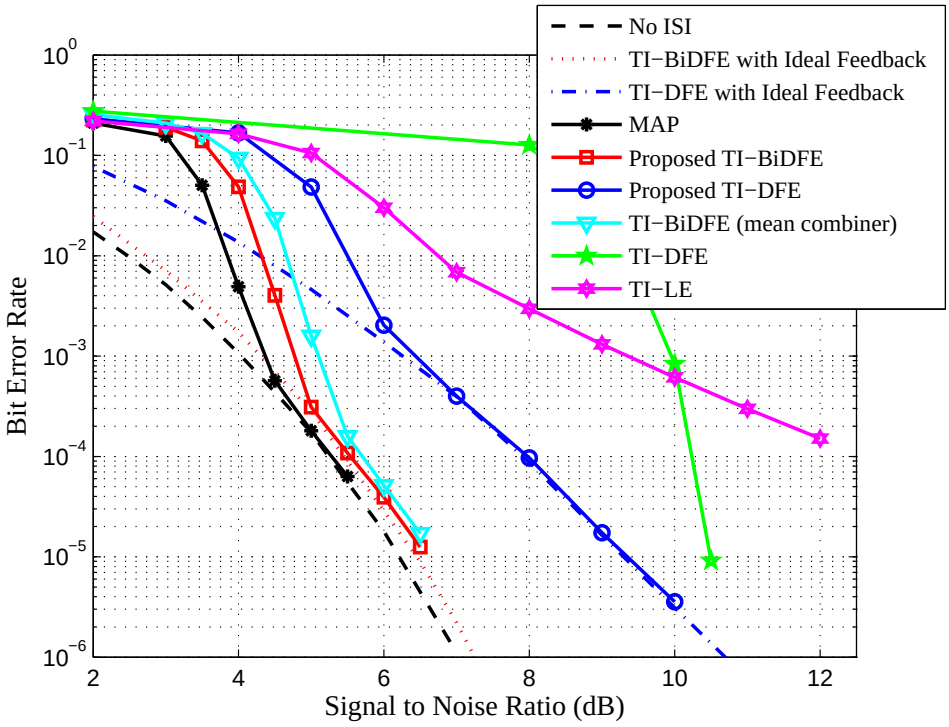


Figure 3.5: BER Curve on the Channel $\mathbf{h_1}$ after 20 iterations with Time-invariant Filters.

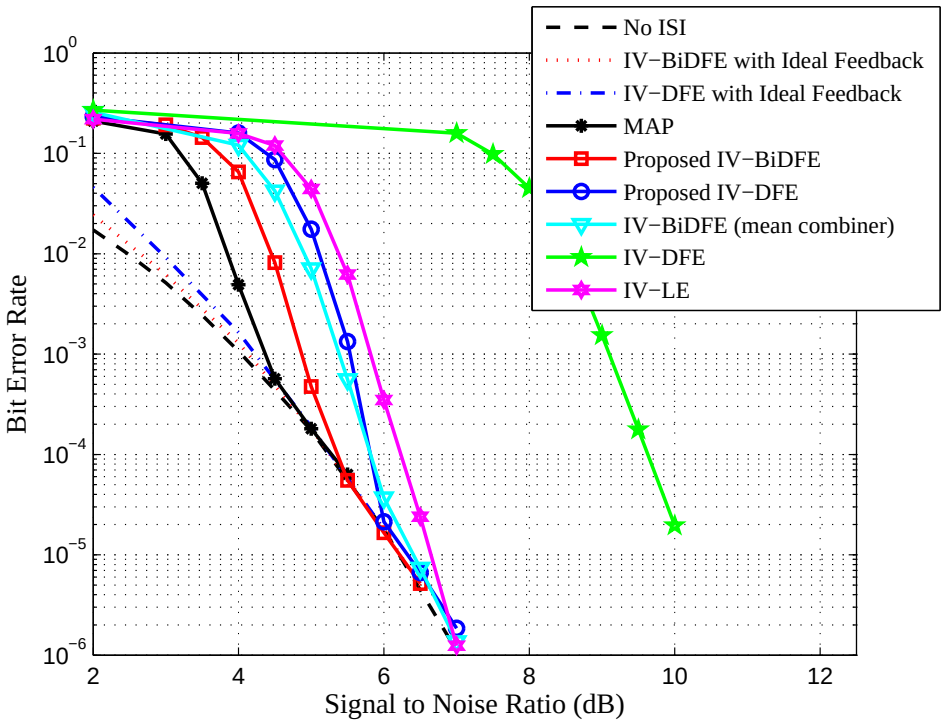


Figure 3.6: BER Curve on the Channel $\mathbf{h_1}$ after 20 iterations with Iteration-varying Filters.

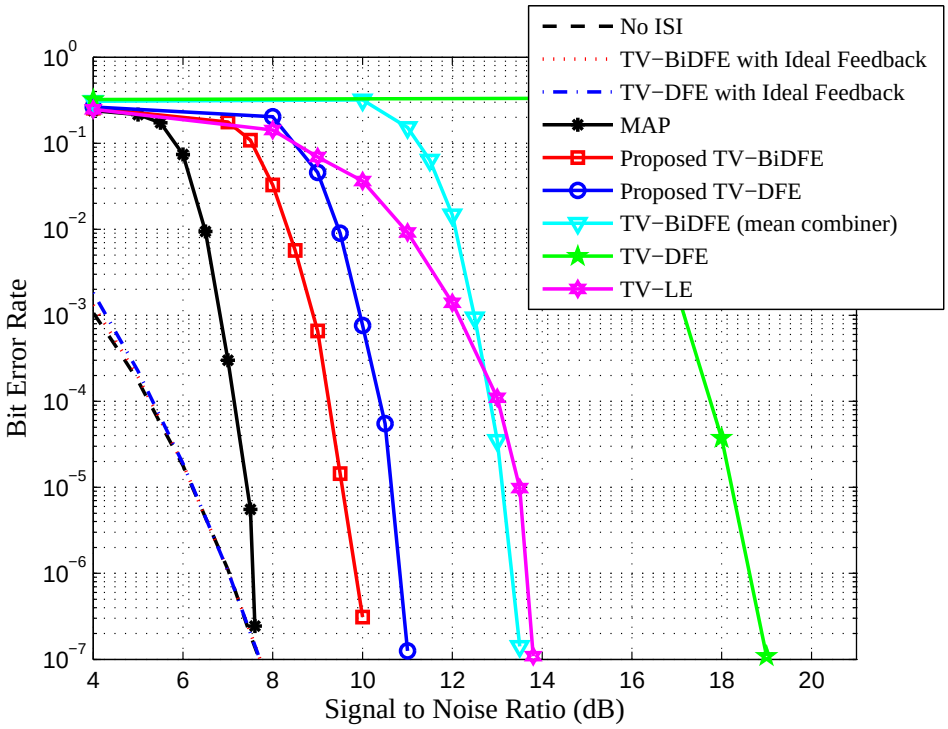


Figure 3.7: BER Curve on the Channel $\mathbf{h_2}$ after 20 Iterations with Time-varying Filters.

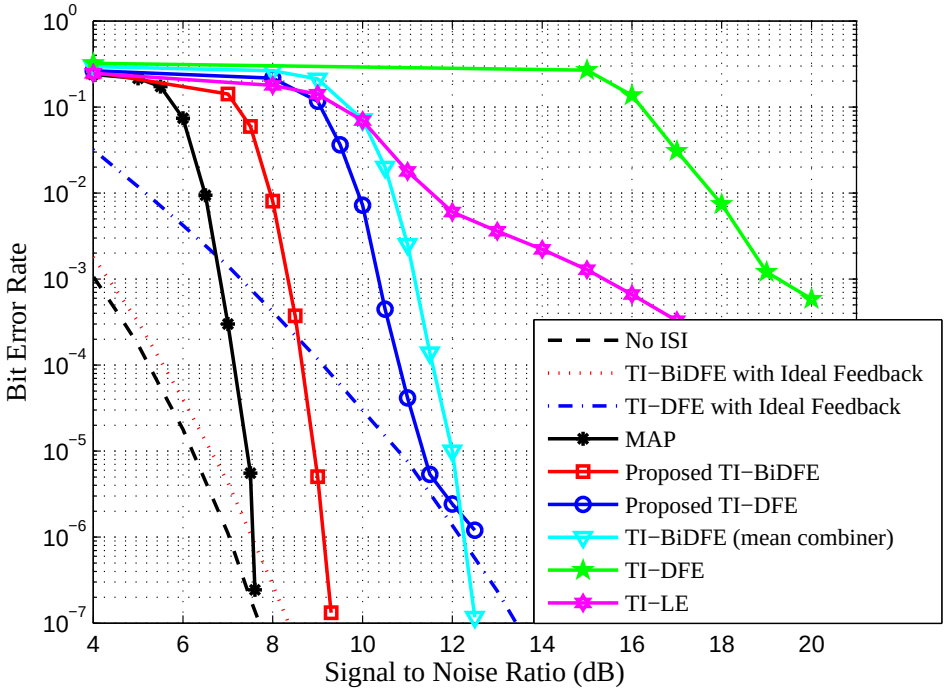


Figure 3.8: BER Curve on the Channel $\mathbf{h_2}$ after 20 iterations with Time-invariant Filters.

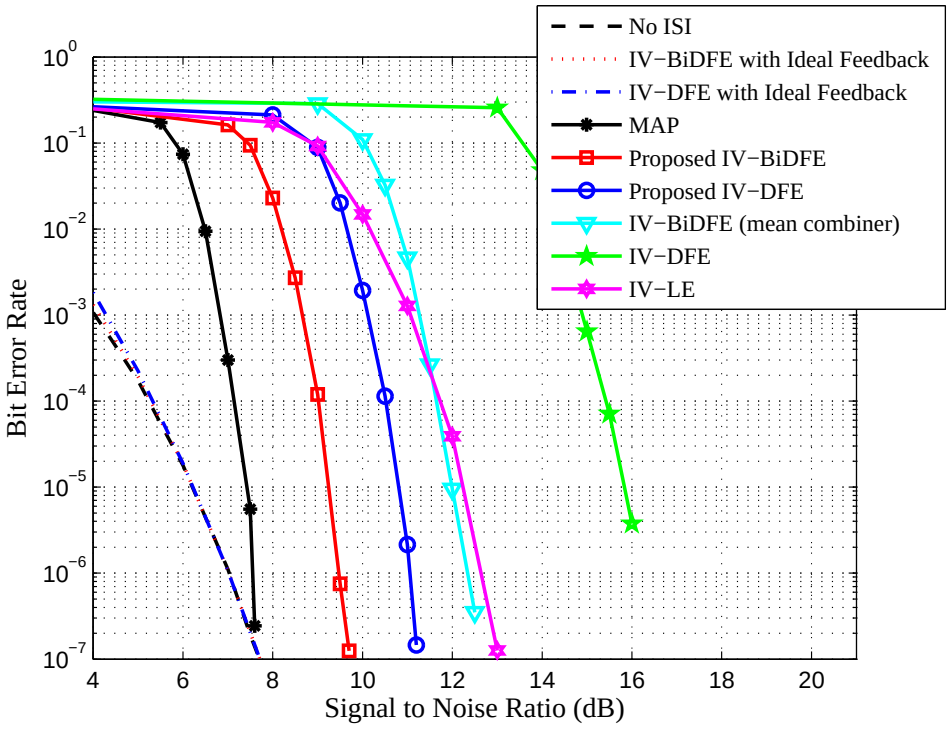


Figure 3.9: BER Curve on the Channel $\mathbf{h_2}$ after 20 iterations with Iteration-varying Filters.

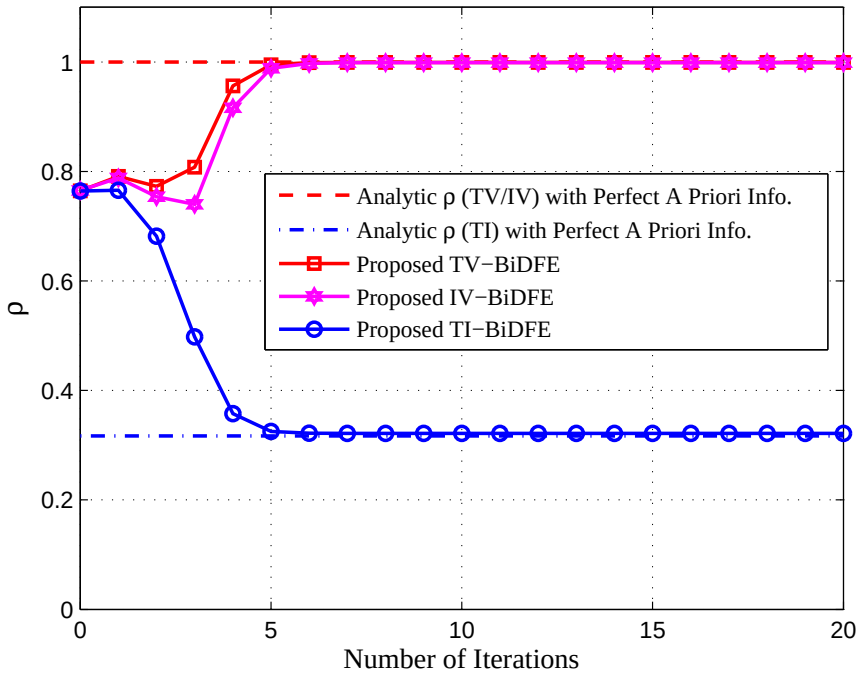


Figure 3.10: Noise Correlation of "Proposed BiDFE" on the Channel $\mathbf{h_1}$.

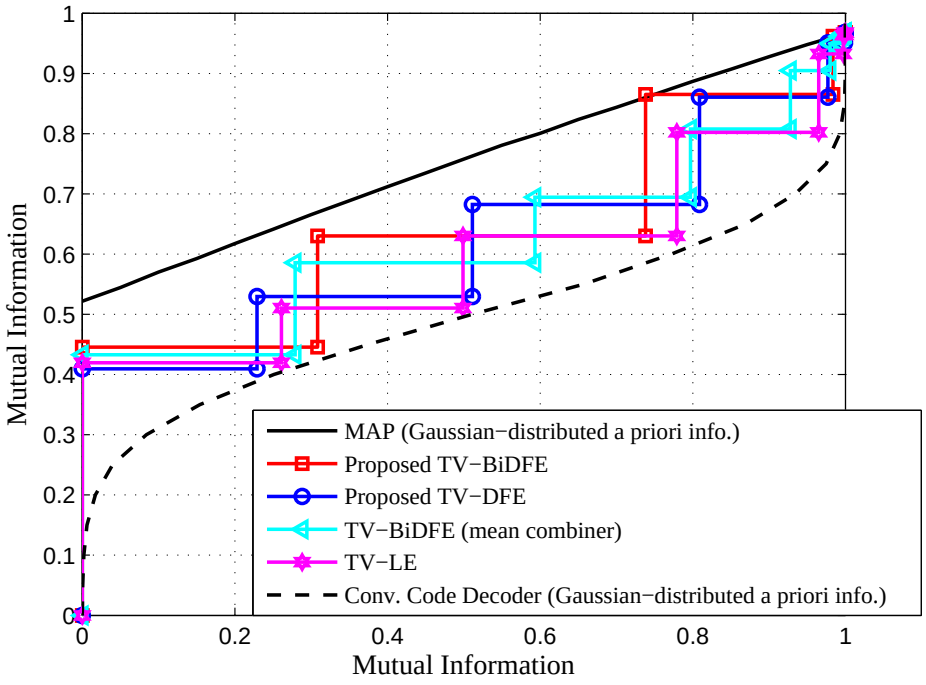


Figure 3.11: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Time-varying Filters.

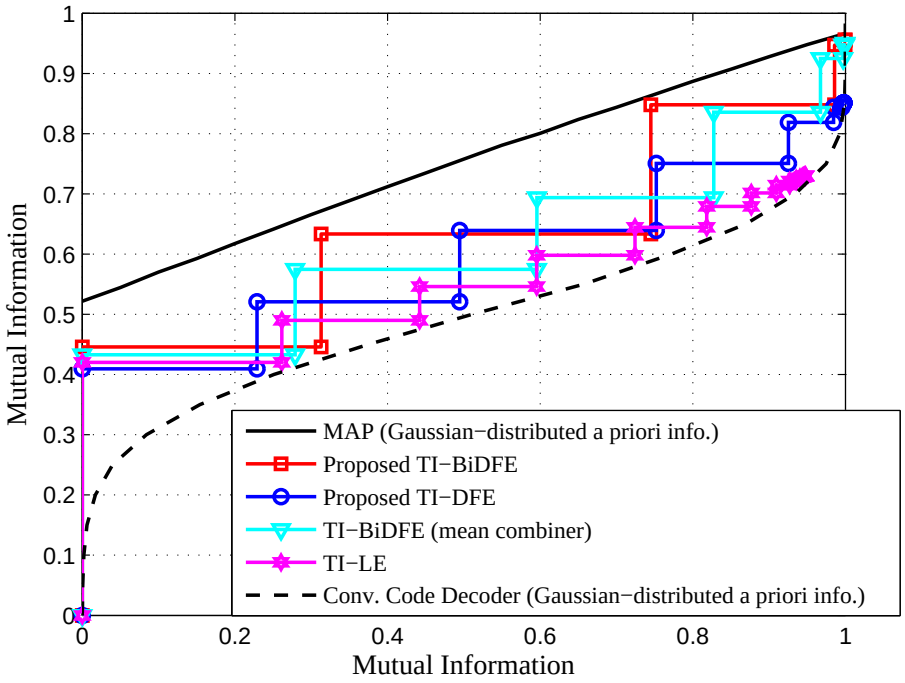


Figure 3.12: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Time-invariant Filters.

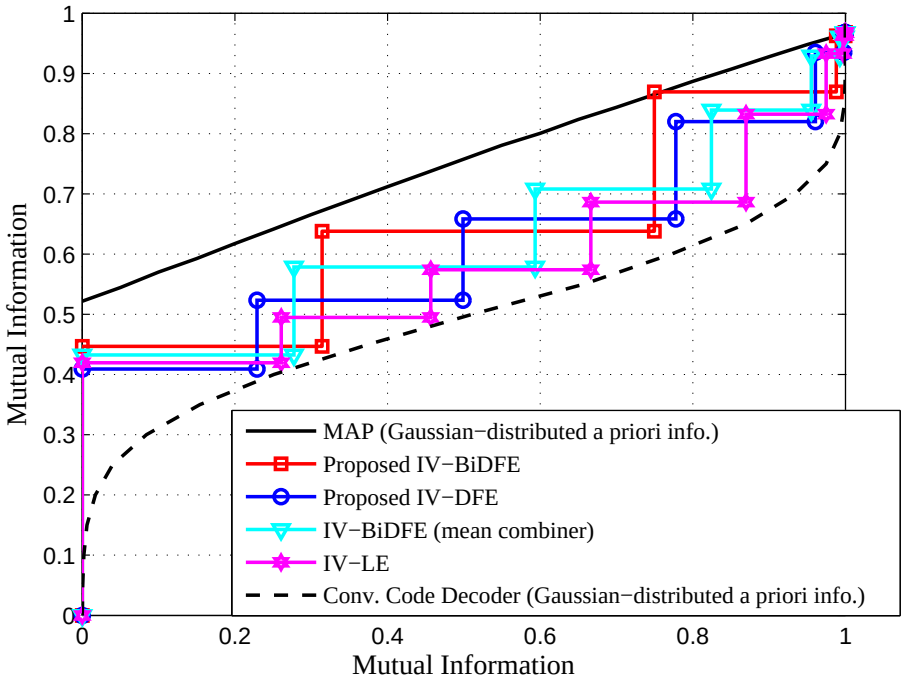


Figure 3.13: EXIT Chart on the Channel $\mathbf{h_1}$ at a 7 dB with Iteration-varying Filters.

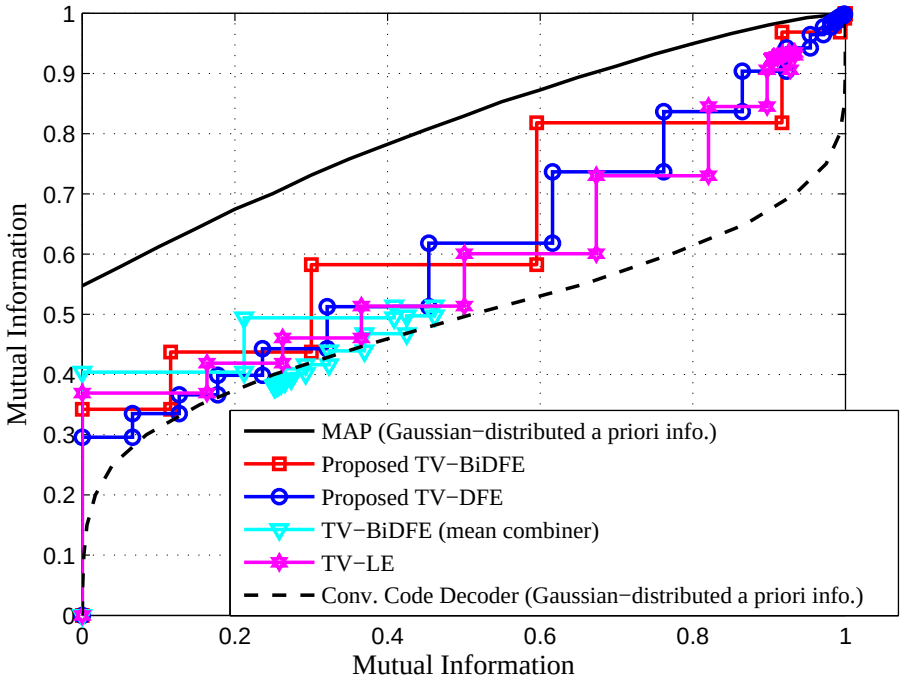


Figure 3.14: EXIT Chart on the Channel $\mathbf{h_2}$ at a 10 dB with Time-varying Filters.

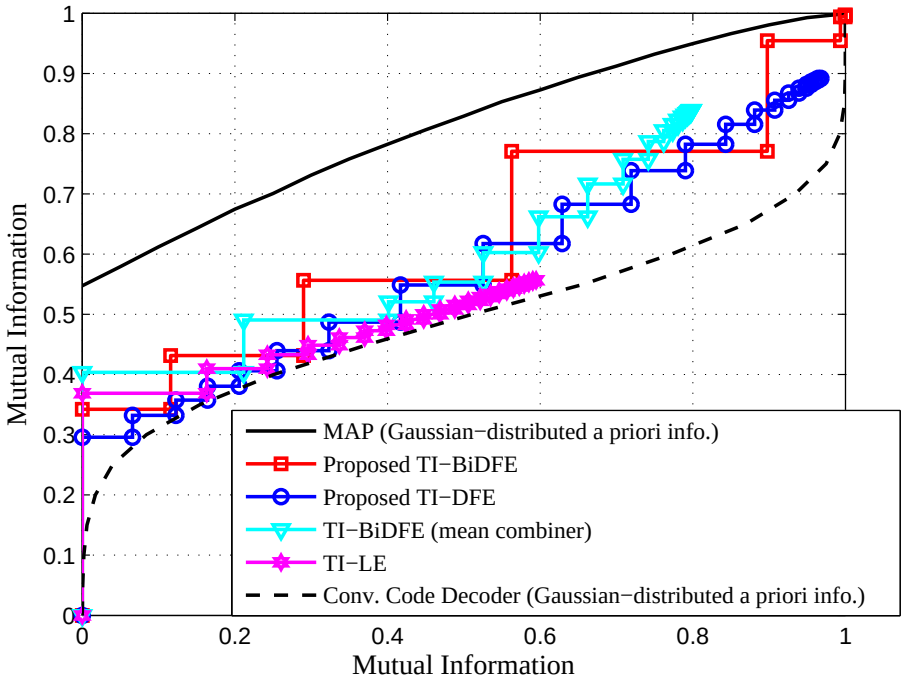


Figure 3.15: EXIT Chart on the Channel $\mathbf{h_2}$ at a 10 dB with Time-invariant Filters.

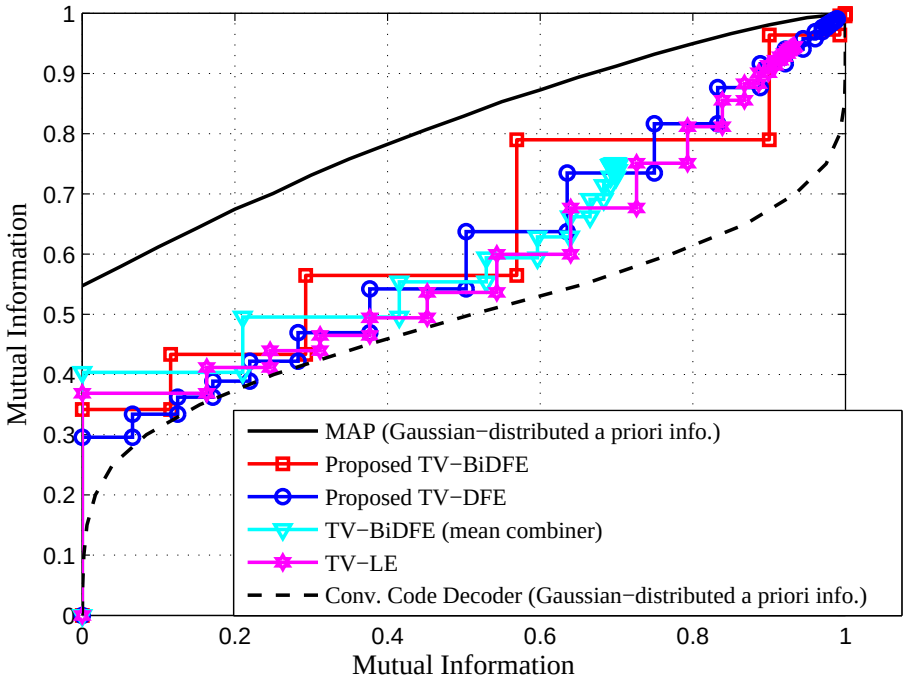


Figure 3.16: EXIT Chart on the Channel $\mathbf{h_2}$ at a 10 dB with Iteration-varying Filters.

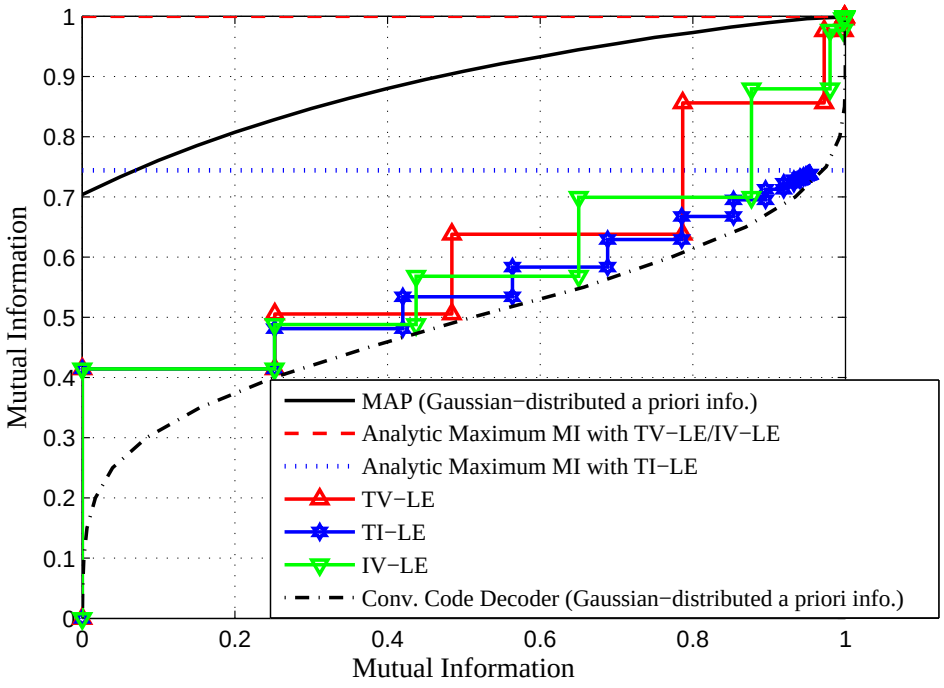


Figure 3.17: EXIT Chart of LE with Various Filters on the Channel $\mathbf{h_2}$ at a 12 dB.

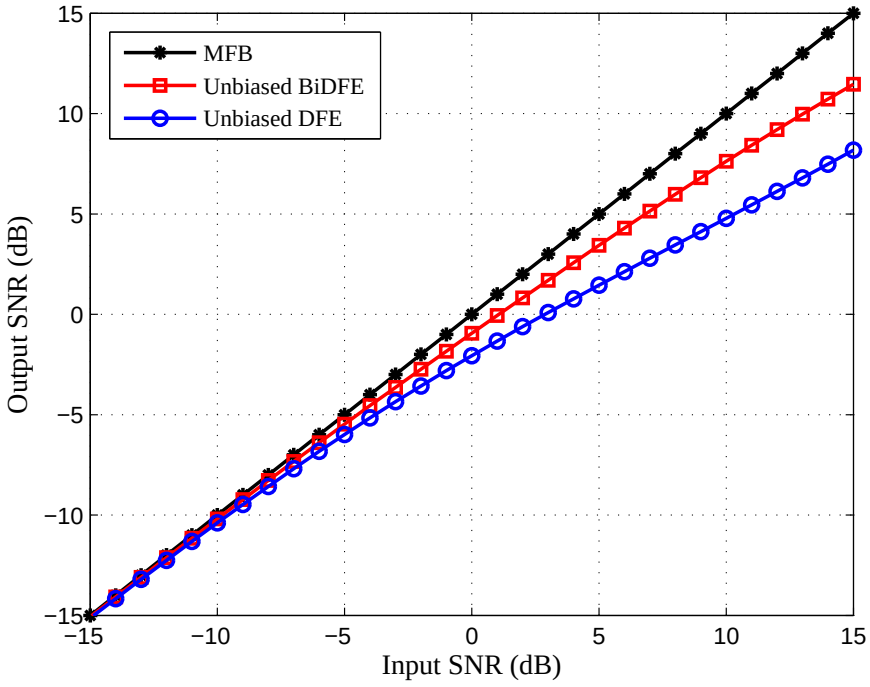


Figure 3.18: SNR plot on the Channel $\mathbf{h_1}$.

Chapter 4

Self-Iterating Soft Equalizer

The previous BiDFE algorithm can be considered as a parallel concatenated scheme with two suboptimal DFEs producing somewhat correlated yet significantly different extrinsic information. The time-reversal operation applied to the reverse DFE can be viewed as a type of interleaving that attempts to make two input streams going into the forward and reverse DFEs appear independent.

In this chapter, we focus on a new equalizer structure that employs LE, DFE or BiDFE as constituent modules, devising a strategy that allows iterative exchange of soft information among the constituent equalizers. Unlike typical turbo processing methods, no interleaver exists between the SISO equalizer modules, and a special strategy to combat the correlation between successive module outputs must be devised. It is shown that the extrinsic information of one module becomes the *a priori* information for the next module connected in serial concatenation via a specific scaling law that depends on the correlation between the two sets of information. This equalizer is viewed as a self-iterating soft equalizer (SISE) consisting of several suboptimal constituent equalizers which are serially concatenated with no interleavers placed between them. The rationale behind this particular equalizer structure is that the suboptimal equalizers such as the LE, DFE, and BiDFE all have their own advantages and disadvantages, and one should

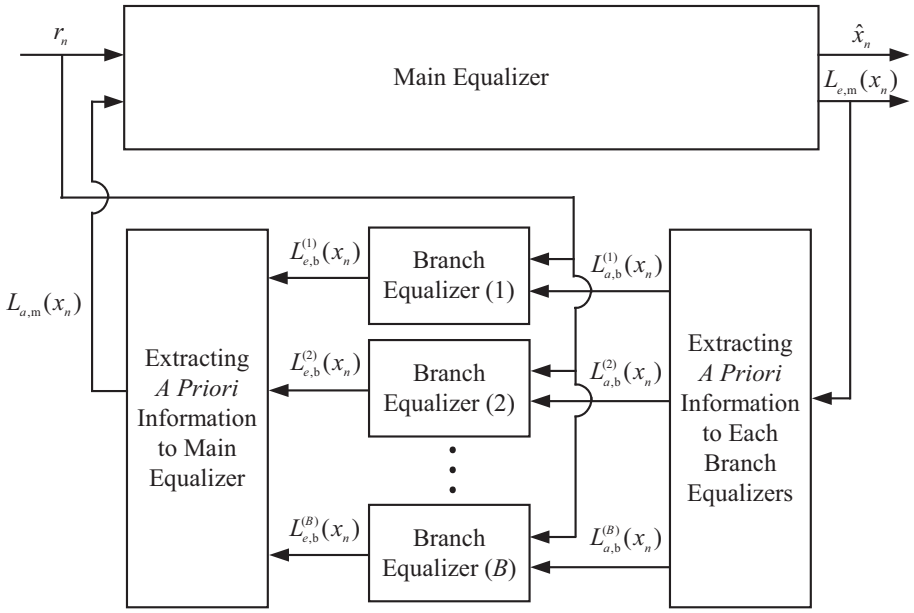


Figure 4.1: Self-Iterating Soft Equalizer (SISE).

be able to benefit from the presence of the other equalizers. For example, the LE does not have the error propagation problem which the DFE suffers from, whereas the DFE often shows significantly better performance than the LE when feedback decisions are correct; and the BiDFE provides solid performance even with time-invariant filters, although its complexity is roughly double the complexity of the DFE. This chapter is based on [71, 72].

4.1 Self-Iterating Soft Equalizer Algorithm

In this section, we discuss the SISE algorithm. Basically, the SISE is a SISO equalizer which consists of one main suboptimal SISO equalizer and B branch suboptimal SISO equalizers. The proposed SISE is illustrated in Fig. 4.1.

The key procedure in this algorithm is that the received data sequence is equalized by the main equalizer and its extrinsic information is passed to the branch equalizers as their *a priori* information. The extrinsic information generated in the branch equalizers

is also passed back to the main equalizer to be used as its *a priori* information for the next stage. Note that since this equalization algorithm can perform iteratively without the decoder (hence the name "self-iterating" equalizer), it can be used in uncoded systems as well. The terms "main" and "branch" here do not necessarily imply the difference in the complexity levels or computational powers between the constituent equalizers. Rather, the distinction simply indicates the scheduling strategy, i.e., the main equalizer is the one that makes the initial decision in the serial concatenation of the constituent equalizers. In fact, different arrangements of the constituent equalizers are possible, including full parallel concatenation, full serial concatenation and combined parallel/serial concatenation, along with many different scheduling strategies. For the particular SISE structure shown in Fig. 4.1, for example, it can be seen that while there is "self-iteration" between the main equalizer and the block of branch equalizers, no self-iterations are assumed among the branch equalizers. While the concept and methods developed in this chapter are general, for the performance analysis and simulation results to be presented, we shall focus on a serial concatenation of one main equalizer and one branch equalizer.

Unlike the extrinsic information between the decoder and the equalizer in usual turbo equalization, the extrinsic information between the main equalizer and the branch equalizers have correlation because no interleaving techniques can be used and their equalization processes are all based on the common received data sequence. It has been suggested that the high correlation between the *a priori* information and the extrinsic information of a module in a turbo system can cause the performance degradation [43, 48]. In this section, we show a proper way to construct the extrinsic information from the branch equalizers when their outputs are correlated with the main equalizer output. The same method can be applied in extracting the extrinsic information from the main equalizer when its soft output is correlated with those of the branch equalizers.

4.1.1 Generation of Uncorrelated *A Priori* Information

First let us assume that there is one main equalizer and one branch equalizer in an uncoded system. We will later extend the proposed algorithm to the case of multiple branch equalizers. We assume that the main equalizer passes the extrinsic LLR sequence, $L_{e,\mathrm{m}}(x)$, to the branch equalizer as the latter's *a priori* LLR, $L_{a,\mathrm{b}}(x)$, and the branch equalizer produces the extrinsic LLR, $L_{e,\mathrm{b}}(x)$, with the given $L_{a,\mathrm{b}}(x)$ sequence.

Since the *a priori* information (or extrinsic information) can be modeled as the output of an equivalent AWGN channel [63], we write the two unbiased equalizer outputs corrupted by AWGN, corresponding to the transmitted symbol x, as:

$$y_{\mathrm{m}} = x + u_{\mathrm{m}}$$

$$y_{\mathrm{b}} = x + u_{\mathrm{b}}$$

where the subscripts 'm' and 'b' indicate the main equalizer and the branch equalizer, respectively. Here, time index n is dropped for notational simplicity; the process remains identical as n evolves. The noise terms u_{m} and u_{b} are assumed to be zero mean Gaussian random variables which are independent of the transmitted data x but correlated with each other with correlation coefficient ρ.

Since the *a posteriori* information at the branch equalizer should explore both its own observation y_b as well as the available "side" information y_m, the *a posteriori* LLR

of the branch equalizer can be written as

$$L_{\mathrm{b}}(x) = \ln \frac{\Pr\left(x=+1|y_{\mathrm{b}}, y_{\mathrm{m}}\right)}{\Pr\left(x=-1|y_{\mathrm{b}}, y_{\mathrm{m}}\right)} \tag{4.1}$$

$$= \ln \frac{\Pr\left(y_{\mathrm{b}}, y_{\mathrm{m}}|x=+1\right)}{\Pr\left(y_{\mathrm{b}}, y_{\mathrm{m}}|x=-1\right)} + \ln \frac{\Pr\left(x=+1\right)}{\Pr\left(x=-1\right)}$$

$$= \ln \frac{\Pr\left(y_{\mathrm{b}}, y_{\mathrm{m}}|x=+1\right)}{\Pr\left(y_{\mathrm{b}}, y_{\mathrm{m}}|x=-1\right)} \tag{4.2}$$

$$= \ln \frac{\Pr\left(y'_{\mathrm{b}}, y'_{\mathrm{m}}|x=+1\right)}{\Pr\left(y'_{\mathrm{b}}, y'_{\mathrm{m}}|x=-1\right)}$$

$$= \ln \frac{\Pr\left(y'_{\mathrm{m}}|x=+1\right)}{\Pr\left(y'_{\mathrm{m}}|x=-1\right)} + \ln \frac{\Pr\left(y'_{\mathrm{b}}|x=+1\right)}{\Pr\left(y'_{\mathrm{b}}|x=-1\right)}$$

$$= \frac{\left(N_{\mathrm{b}} - \rho\sqrt{N_{\mathrm{m}}N_{\mathrm{b}}}\right)}{\left(1-\rho^2\right)N_{\mathrm{b}}} L_{e,\mathrm{m}}(x) + \frac{\left(N_{\mathrm{m}} - \rho\sqrt{N_{\mathrm{m}}N_{\mathrm{b}}}\right)}{\left(1-\rho^2\right)N_{\mathrm{m}}} L_{e,\mathrm{b}}(x) \tag{4.3}$$

where $N_{\mathrm{m}} \triangleq \mathrm{Var}(u_{\mathrm{m}})$ and $N_{\mathrm{b}} \triangleq \mathrm{Var}(u_{\mathrm{b}})$, and y'_m and y'_b are the whitened equalized outputs obtained by passing of y_m and y_b through a whitening filter computed via the singular value decomposition (SVD) of the noise correlation matrix $\mathbf{R}$ defined as

$$\mathbf{R} \triangleq \begin{bmatrix} \mathrm{Var}(u_m) & \mathrm{E}(u_m u_b) \\ \mathrm{E}(u_b u_m) & \mathrm{Var}(u_b) \end{bmatrix} = \begin{bmatrix} N_m & \rho\sqrt{N_m N_b} \\ \rho\sqrt{N_m N_b} & N_b \end{bmatrix}.$$

The equality of (4.2) holds because the branch equalizer does not have any *a priori* information other than what it derives from the main equalizer.

Obtaining (4.3) is straightforward; for detailed derivation, see Section 3.2.1. Due to the sensitivity of $L_{\mathrm{b}}(x)$ to the estimation error of ρ in Section 3.2.1, we further assume that the variance of u_{m} and u_{b} are the same, i.e., $N_{\mathrm{m}} = N_{\mathrm{b}}$. Then, we can write

$$L_{\mathrm{b}}(x) = \frac{1}{1+\rho}\Big(L_{e,\mathrm{m}}(x) + L_{e,\mathrm{b}}(x)\Big). \tag{4.4}$$

Now, we discuss how to extract the extrinsic information at the branch equalizer from this *a posteriori* LLR and pass it as the *a priori* LLR of the main equalizer. Notice that if the noise u_{m} and u_{b} are uncorrelated, i.e., $\rho = 0$, the *a posteriori* LLR of the branch equalizer would simply be $L_{\mathrm{b}}(x) = L_{e,\mathrm{m}}(x) + L_{e,\mathrm{b}}(x)$ and the *a priori* LLR to the main equalizer is naturally defined as $L_{a,\mathrm{m}}(x) = L_{\mathrm{b}}(x)|_{L_{e,\mathrm{m}}(x)=0} = L_{e,\mathrm{b}}(x)$.

However, due to the effect of correlation between noise, $L_{\mathrm{b}}(x)$ is reduced here by the factor $1/(1+\rho)$. This suggests that the correlated information between the extrinsic information of the main equalizer and the extrinsic information of the branch equalizer be defined as

$$\begin{aligned} L_{\mathrm{corr}}(x) &\triangleq L_{\mathrm{b}}(x)|_{\rho=0} - L_{\mathrm{b}}(x) \\ &= \frac{\rho}{1+\rho}\Big(L_{e,\mathrm{m}}(x) + L_{e,\mathrm{b}}(x)\Big). \end{aligned} \tag{4.5}$$

We can view the relationship between $L_{\mathrm{b}}(x)|_{\rho=0}$, $L_{\mathrm{b}}(x)$, and $L_{\mathrm{corr}}(x)$ in the following sense: Since the correlated information is added twice in $L_{\mathrm{b}}(x)|_{\rho=0}$, it is necessary to subtract the correlated information, $L_{\mathrm{corr}}(x)$, from $L_{\mathrm{b}}(x)|_{\rho=0}$ to get $L_{\mathrm{b}}(x)$.

Then, it makes sense to define the *a priori* LLR of the main equalizer from the branch equalizer as

$$\begin{aligned} L_{a,\mathrm{m}}(x) &\triangleq \Big(L_{\mathrm{b}}(x) - L_{\mathrm{corr}}(x)\Big)\Big|_{L_{e,\mathrm{m}}(x)=0} \\ &= \left(\frac{1-\rho}{1+\rho}\right) L_{e,\mathrm{b}}(x). \end{aligned} \tag{4.6}$$

Notice that if two noise samples are uncorrelated like the extrinsic information between the decoder and the equalizer, we could fully utilize the extrinsic LLR of the branch equalizer as the *a priori* LLR of the main equalizer, i.e., $L_{a,\mathrm{m}}(x) = L_{e,\mathrm{b}}(x)$. On the other hand, if the noise correlation coefficient is 1, the extrinsic information of the branch equalizer would be completely redundant to the main equalizer, i.e., $L_{a,\mathrm{m}}(x) = 0$. Note that this process is also valid in opposite direction, i.e., when the *a priori* LLR of the branch equalizer is extracted from the *a posteriori* LLR of the main equalizer.

4.1.2 Estimation of Noise Correlation Coefficient

As mentioned before, it is difficult to compute the correlation coefficient between u_{m} and u_{b} (or $L_{e,\mathrm{m}}(x)$ and $L_{e,\mathrm{b}}(x)$) analytically. However, assuming that the noise is stationary, the correlation coefficient can be estimated through time-averaging $L_{e,\mathrm{m}}(x)$ and $L_{e,\mathrm{b}}(x)$

over some reasonably large finite window:

$$\hat{\rho} = \frac{\sum \left\{ \left(L_{e,\mathrm{m}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{m}}(x_n) \right] m_{\mathrm{m}} \right) \left(L_{e,\mathrm{b}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{b}}(x_n) \right] m_{\mathrm{b}} \right) \right\}}{\sqrt{\sum \left(L_{e,\mathrm{m}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{m}}(x_n) \right] m_{\mathrm{m}} \right)^2} \sqrt{\sum \left(L_{e,\mathrm{b}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{b}}(x_n) \right] m_{\mathrm{b}} \right)^2}} \quad (4.7)$$

where, due to the symmetricity of x_n, m_{m} and m_{b} are the respective mean values of $L_{e,\mathrm{m}}(x_n)$ and $L_{e,\mathrm{b}}(x_n)$ for $x_n = +1$, i.e., $m_{\mathrm{m}} = \mathrm{E}(L_{e,\mathrm{m}}(x_n)|x_n = +1)$ and $m_{\mathrm{b}} = \mathrm{E}(L_{e,\mathrm{b}}(x_n)|x_n = +1)$. We can also estimate them through time-averaging:

$$\hat{m}_{\mathrm{m}} = \frac{1}{2} \left\{ \overline{(L_{e,\mathrm{m}}(x_n)|L_{e,\mathrm{m}}(x_n) \geq 0)} - \overline{(L_{e,\mathrm{m}}(x_n)|L_{e,\mathrm{m}}(x_n) < 0)} \right\}$$
$$\hat{m}_{\mathrm{b}} = \frac{1}{2} \left\{ \overline{(L_{e,\mathrm{b}}(x_n)|L_{e,\mathrm{b}}(x_n) \geq 0)} - \overline{(L_{e,\mathrm{b}}(x_n)|L_{e,\mathrm{b}}(x_n) < 0)} \right\}$$

where $\overline{t_n}$ means the time-average of t_n. Note that the hard decisions for the transmitted symbols in the main equalizer and the branch equalizer might be different; in estimating the correlation coefficient, we only consider those LLR samples for which $\mathrm{sign}\left[L_{e,\mathrm{m}}(x_n) \right]$ and $\mathrm{sign}\left[L_{e,\mathrm{b}}(x_n) \right]$ are identical. Also, notice that $L_{e,\mathrm{m}}(x_n)$ in (4.7) can be replaced by $L_{a,\mathrm{b}}(x_n)$.

4.1.3 Extension to the Case of Multiple Branch Equalizers

Let us assume that there is one main equalizer and two branch equalizers. Then, the *a priori* LLR to the main equalizer from each branch equalizer can be defined as

$$L_{a,\mathrm{m}}^{(1)}(x) = \left(\frac{1 - \rho^{(1)}}{1 + \rho^{(1)}} \right) L_{e,\mathrm{b}}^{(1)}(x)$$
$$L_{a,\mathrm{m}}^{(2)}(x) = \left(\frac{1 - \rho^{(2)}}{1 + \rho^{(2)}} \right) L_{e,\mathrm{b}}^{(2)}(x).$$

where superscript (i) points to a specific branch equalizer and $\rho^{(i)}$ is the correlation coefficient between $L_{e,\mathrm{b}}^{(i)}(x)$ and $L_{a,\mathrm{b}}^{(i)}(x)$. Again, since the *a priori* information (or extrinsic information) can be modeled as an equivalent AWGN channel output, under the assumption that the noise variances are the same, the whitened and combined *a priori* information to the main equalizer [24] is given by

$$L_{a,\mathrm{m}}(x) = \frac{1}{1 + \xi} \left(L_{a,\mathrm{m}}^{(1)}(x) + L_{a,\mathrm{m}}^{(2)}(x) \right) \quad (4.8)$$

where ξ is the noise correlation coefficient between $L^{(1)}_{a,\mathrm{m}}(x)$ and $L^{(2)}_{a,\mathrm{m}}(x)$ and it can be also estimated through time-averaging using a similar equation to (4.7). It is straightforward to extend the SISE algorithm to the system consisting of B branch equalizers. Specifically, the whitened and combined *a priori* LLR to the main equalizer, $L_{a,\mathrm{m}}(x)$, can be obtained by passing the *a priori* LLRs generated by B branch equalizers, $L^{(1)}_{a,\mathrm{m}}(x), L^{(2)}_{a,\mathrm{m}}(x), \ldots, L^{(B)}_{a,\mathrm{m}}(x)$, through the whitening filter computed by the SVD of their $B \times B$ noise correlation matrix.

4.1.4 SISE Algorithm

Finally, the proposed SISE algorithm for an uncoded system can be summarized as follows:

- Initialize the *a priori* information of the main equalizer, i.e., $L_{a,\mathrm{m}}(x_n) = 0$ and $L^{(i)}_{a,\mathrm{m}}(x_n) = 0$ for all time index n and branch index i.
- For the specified number of self-iterations,
    1. Generate the extrinsic information of the main equalizer, $L_{e,\mathrm{m}}(x_n)$, with the *a priori* information $L_{a,\mathrm{m}}(x_k)$ for all $k \neq n$ (the process of generating the extrinsic information in a given equalizer is described in Chapter 2 and 3).
    2. Compute the noise correlation coefficients, $\rho^{(i)}_{\mathrm{m}}$, between $L_{e,\mathrm{m}}(x_n)$ and $L^{(i)}_{a,\mathrm{m}}(x_n)$ and set $L^{(i)}_{a,\mathrm{b}}(x_n) = (1 - \rho^{(i)}_{\mathrm{m}})/(1 + \rho^{(i)}_{\mathrm{m}}) \cdot L_{e,\mathrm{m}}(x_n)$ for all i.
    3. Generate the extrinsic information of each branch equalizer, $L^{(i)}_{e,\mathrm{b}}(x_n)$, with the given *a priori* information $L^{(i)}_{a,\mathrm{b}}(x_k)$ for all $k \neq n$, for all i.
    4. Compute the noise correlation coefficient, $\rho^{(i)}_{\mathrm{b}}$, between $L^{(i)}_{e,\mathrm{b}}(x_n)$ and $L^{(i)}_{a,\mathrm{b}}(x_n)$ and set $L^{(i)}_{a,\mathrm{m}}(x_n) = (1 - \rho^{(i)}_{\mathrm{b}})/(1 + \rho^{(i)}_{\mathrm{b}}) \cdot L^{(i)}_{e,\mathrm{b}}(x_n)$ for all i.
    5. Generate the *a priori* information for the main equalizer, $L_{a,\mathrm{m}}(x_n)$, from $L^{(i)}_{a,\mathrm{m}}(x_n)$ via the extended equation of (4.8).

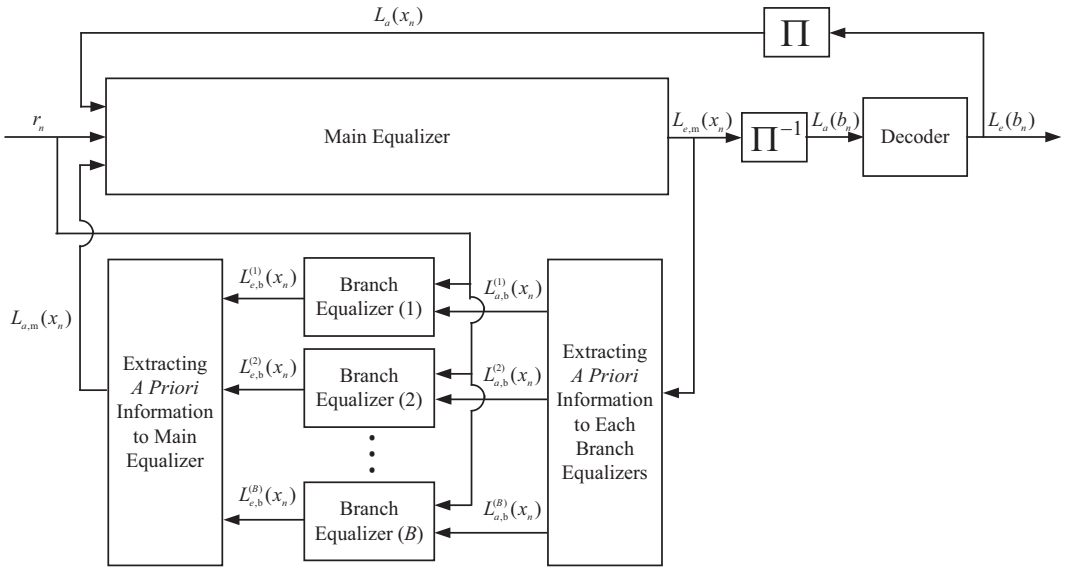


Figure 4.2: Turbo Self-Iterating Soft Equalization Scheme.

4.2 Iterative Self-Iterating Soft Equalizer Algorithm

In this section, we propose turbo equalization based on the previously developed SISE algorithm. The structure of the turbo equalization scheme based on SISE is shown in Fig. 4.2. Various iterative equalization algorithms are possible based on this structure, but two main algorithms are considered here.

4.2.1 SISE 1 Algorithm

The first algorithm passes the uncorrelated extrinsic information of the main equalizer to the branch equalizers and then to the decoder in turn. The information flow of SISE 1 algorithm is shown in Fig. 4.3(a). The detail steps of SISE 1 are described as

- Initialize the *a priori* information from the decoder and the branch equalizers, i.e., $L_a(x_n) = 0$, $L_{a,\mathrm{m}}(x_n) = 0$, and $L_{a,\mathrm{m}}^{(i)}(x_n) = 0$ for all time index n and branch index i.

- For the specified number of outer iterations,

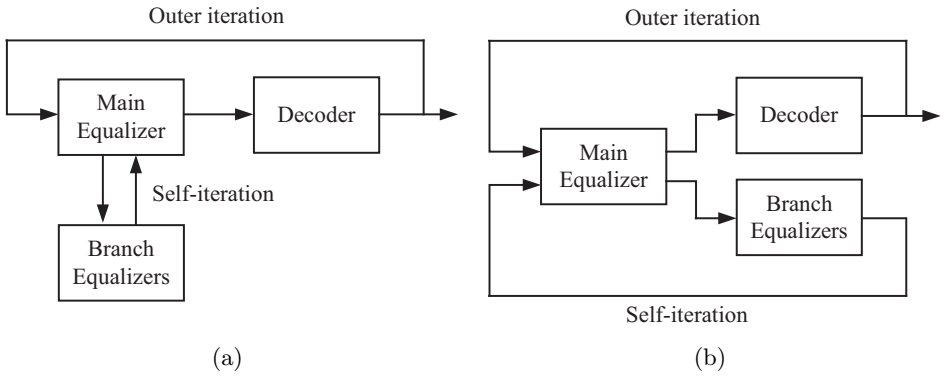


Figure 4.3: Information Flow: (a) Turbo SISE 1 (b) Turbo SISE 2.

1. Generate the extrinsic information of the main equalizer, $L_{e,\mathrm{m}}(x_n)$, utilizing the *a priori* information from the decoder, $L_a(x_k)$ for all $k \neq n$.
2. Take the same steps of $2) - 5)$ in 4.1.4.
3. Reproduce the extrinsic information of the main equalizer, $L_{e,\mathrm{m}}(x_n)$, with the combined *a priori* information from the decoder and the branch equalizers, $L_a(x_k) + L_{a,\mathrm{m}}(x_k)$ for all $k \neq n$.
4. Pass $L_{e,\mathrm{m}}(x_n)$ back to the decoder.

4.2.2 SISE 2 Algorithm

Due to the sequential nature of the self-iteration steps, the SISE 1 algorithm has long processing latency issue; the second algorithm (SISE 2) is also proposed to get around this issue. Different from the first algorithm, SISE 2 passes the correlation-compensated extrinsic information of the main equalizer to the branch equalizers and to the decoder *simultaneously*. Thus, in this case, the self-iteration step is performed in parallel with the outer turbo iteration. The information flow of SISE 2 algorithm is summarized in Fig. 4.3(b). The detailed algorithm of SISE 2 is described as

- Initialize the *a priori* information from the decoder and the branch equalizers, i.e., $L_a(x_n) = 0$, $L_{a,\mathrm{m}}(x_n) = 0$, and $L^{(i)}_{a,\mathrm{m}}(x_n) = 0$ for all time index n and branch

index i.

- For the specified number of outer iterations,

  1. Generate the extrinsic information of the main equalizer, $L_{e,\mathrm{m}}(x_n)$, with the combined *a priori* information from the decoder and the branch equalizer, $L_a(x_k) + L_{a,\mathrm{m}}(x_k)$ for all $k \neq n$.
  2. Pass $L_{e,\mathrm{m}}(x_n)$ to the decoder.
  3. While the decoder computes its own extrinsic information and passes it as $L_a(x_n)$ to the main equalizer,
     - Take the same steps of $2) - 5)$ in 4.1.4.

4.2.3 Comparison of Complexity and Latency

Let the computational complexity of the main equalizer, the branch equalizers, and the decoder be C_M, C_B, and C_D, respectively. For each outer iteration performed, the amount of computation for the conventional turbo equalization is $C_M + C_D$, whereas it is $2C_M + C_B + C_D$ and $C_M + C_B + C_D$ for SISE 1 and SISE 2, respectively. Moreover, assuming the processing time for the main equalizer, the parallel branch equalizers, and the decoder is all equal to T, the total processing time for each outer iteration is $2T$, $4T$, and $2T$ for the conventional system, SISE 1, and SISE 2, respectively. As will be shown later, in addition to having complexity/latency advantage, SISE 2 also has performance advantage over SISE 1 in many channel situations and thus seems to be the preferred choice. Also, while SISE 2 requires higher complexity (by C_B) than existing turbo equalizers when C_M is fixed in both cases, it will be shown that SISE 2 often enables substantial error rate reduction in channel conditions where the existing turbo equalizer cannot provide any performance improvement regardless of how large C_M is allowed to grow.

4.3 Numerical Results

In this section, simulation results of the proposed SISE equalization schemes for both uncoded and coded systems are presented. Same as before, the transmitted symbols are modulated with BPSK so $x_n \in \{\pm 1\}$. The message bit length is 2^{11}. We also assume that the noise is AWGN, and the noise variance and the channel information are perfectly known to the receiver. In the figures, the label "Ideal" indicates the performance of an equalizer with perfect *a priori* information.

4.3.1 Uncoded System

The impulse response of the ISI channel $\mathbf{h_0} = (1/\sqrt{6})[1 \quad 2 \quad 1]^T$ introduced in [10] is used for the uncoded system. This channel has a spectral null over the Nyquist band as shown in Fig. 4.4. Three different equalizer types are simulated for this channel. The SISE method is the self-iterating soft equalizer algorithm described in Section 4.1.4. Specifically, the SISO BiDFE algorithm of Section 3.2 is adopted as the main equalizer which takes into account the possibility that error propagation i_n is not zero by tracking the conditional extrinsic LLR: $L_e(x_n|i_n \neq 0) = 2p_{\{n,0\}}\left(y_n - \mathrm{E}(i_n|i_n \neq 0)\right)/\mathrm{Var}(v_n)$. Moreover, when decisions are made by the BiDFE, the arbitration criterion of [60] with window size 15 is employed; the symbol sequence is decided among the estimated sequence of two DFEs based on which candidate shows the smaller MSE in a window centered around the symbol of interest. Finally, the SISO LE is used for the branch equalizer and 2 self-iterations are applied. The "BAD" method is the BAD algorithm of [60], and the MAP equalizer is the optimal equalizer implemented via the BCJR algorithm. Each DFE in BiDFE consists of 13 feedforward taps and 2 feedback taps ($L_c = 12$ and $L_d = 2$) while the LE uses 15 taps ($L_{f1} = 7$ and $L_{f2} = 7$) for $\mathbf{h_1}$. Fig. 4.5 shows the performance comparison. As the figure shows, the proposed SISE algorithm shows the superior performance to the BAD method of [60] and approaches the performance of "Ideal BAD".

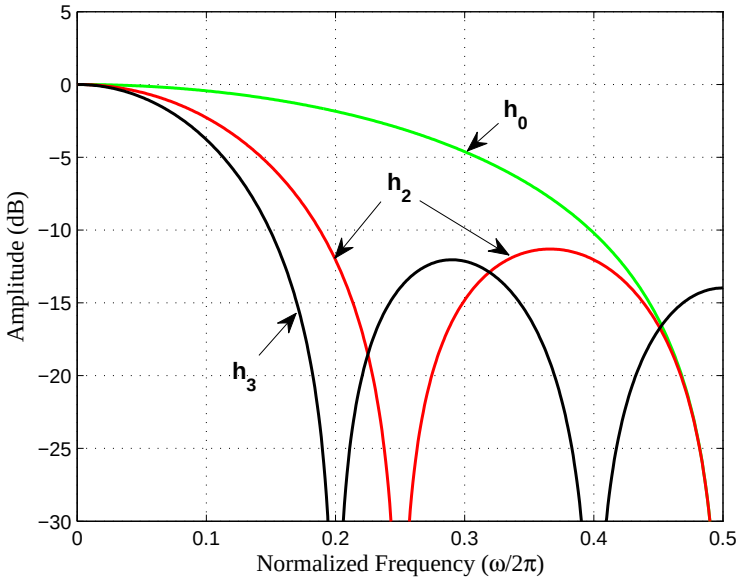


Figure 4.4: Frequency Magnitude Response of the ISI Channels: $\mathbf{h_0} = (1/\sqrt{6})[1 \quad 2 \quad 1]^T$, $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$, and $\mathbf{h_3} = (1/\sqrt{85})[1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 4 \quad 3 \quad 2 \quad 1]^T$.

4.3.2 Coded System

In this subsection, simulation results of the iterative SISE equalization schemes are presented. The transmitted symbols are encoded with a recursive rate-1/2 convolutional code encoder with parity generator polynomials $(1 + D^2)/(1 + D + D^2)$. The impulse response of the severe ISI channel $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$ investigated as well as an extremely severe ISI channel of $\mathbf{h_3} = (1/\sqrt{85})[1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 4 \quad 3 \quad 2 \quad 1]^T$ are used for evaluating and comparing the performances of iterative equalizers. These channels are natural extensions of previously considered ISI patterns by Proakis and their frequency responses are shown in Fig. 4.4.

In the error rate figures, curves labeled by "SISE 1" and "SISE 2" correspond to the iterative SISE algorithms described in Section 4.2. Although various combinations

of the main equalizer and the branch equalizers are possible in the iterative SISE algorithm, some limited combinations are considered here. The label "(M, B)" in the legend denotes a specific 'M' algorithm as the main equalizer and a 'B' algorithm as the branch equalizer. For instance, "SISE 2 (IV-DFE, IV-RDFE)" denotes the iterative SISE 2 algorithm with the normal DFE with iteration-varying filters as the main equalizer and the time-reversed DFE (RDFE) with iteration-varying filters as the branch equalizer. The straightforward LLR mapping method of (2.7) is adopted for the LE while the LLR mapping methods of (2.26) and (3.22) are used for the DFE and BiDFE respectively. Moreover, the DFE (and each DFE in the BiDFE) consists of 21 feedforward taps and 6 feedback taps ($L_c = 20$ and $L_d = 6$) on $\mathbf{h_2}$ while 21 feedforward taps and 8 feedback taps ($L_c = 20$ and $L_d = 8$) on $\mathbf{h_3}$. The LE uses 27 taps ($L_{f1} = 13$ and $L_{f2} = 13$) for $\mathbf{h_2}$ and 29 taps ($L_{f1} = 14$ and $L_{f2} = 14$) for $\mathbf{h_3}$. Again, "MAP" is for the optimal equalizer implemented via the BCJR algorithm. Finally, the decoder is also implemented using the BCJR algorithm.

Figs. 4.6, 4.7, and 4.8 show the performance of several turbo equalizers after 20 outer iterations on the ISI channel $\mathbf{h_2}$. The performance of the proposed iterative SISE algorithms is compared with the performance achieved when the equalizer consists only of a single equalizer.

Fig. 4.6 shows performance comparison with LE-based single equalizers. Among the single LE schemes, the iteration-varying filter design method gives the best performance, outperforming even the time-varying filter method. When the number of filter taps increases to 81 (40 causal and 40 anticausal taps) versus a total of 27, the IV-LE method does not provide any performance gain, indicating that using 27 taps for this channel already realizes IV-LE's full potential. Of the two SISE methods showing superior performance to the single LE, SISE 2 is better and is an obvious winner given its lower complexity and latency compared to SISE 1. Both SISE schemes employ TI-BiDFE as the sole branch equalizer. Given that each DFE in BiDFE has a total of 27 taps (21 feedforward and 6 feedback), each SISE scheme requires an overall total of 81 taps.

Fig. 4.7 tells a similar story but the comparison is now with the DFE schemes. Accordingly, the main equalizer of the SISE methods is also set up with the DFE, under the iteration-varying filter design scheme. Among the single equalizers, this time, the time-varying filter method performs slightly better than the iteration-varying filter scheme (again, increasing the number of the iteration-varying filter taps does not give any performance boost). Still, SISE methods give the best performance, other than the MAP-based turbo equalizer, with SISE 2 once again coming ahead. Fig. 4.8 presents results that compare schemes based on BiDFE. Iteration-varying-filter-based single equalizer is better than time-varying filter method and, again, increasing the filter length does not improve performance. As for the branch equalizer, SISE schemes use the IV-LE. This time, SISE 1 performs better than SISE 2, albeit by a very small margin.

A comparison of the simulation results in Figs. 4.6 and 4.8 reveals that the performance of "SISE (TI-BiDFE, IV-LE)" is considerably better than that of "SISE (IV-LE, TI-BiDFE)". On the other hand, among the single equalizers, TI-BiDFE outperforms IV-LE by a large margin. It turns out that the extrinsic LLR quality of the main equalizer is an important factor determining the overall performance of the SISE algorithms, as the extrinsic LLRs of the main equalizer are passed to the decoder as well as the branch equalizers. Therefore, in the design of the SISE algorithm, the equalizer showing the best BER performance (or LLR quality) should be chosen as the main equalizer.

Fig. 4.9 shows comparison in the extremely severe ISI channel $\mathbf{h_3}$. Among the single equalizers, TI-BiDFE schemes exhibit clear failure because the erroneously generated *a priori* LLRs from the decoder during the iterations cause more errors in the subsequent turbo iterations, while IV-BiDFE is better than TV-BiDFE. Both SISE schemes show robust performance, again lagged in performance only by the optimal MAP-based scheme.

The performance of the turbo equalizers are analyzed by using the EXIT chart [63]. In order to avoid excessive cluttering, only the trajectories of "SISE 1", "SISE 2", and the single equalizer are plotted in Figs. 4.10, 4.11, and 4.12. They describe the EXIT

chart on $\mathbf{h_2}$ at a 12 dB SNR when IV-LE is used for the main equalizer, the EXIT chart on $\mathbf{h_2}$ at a 10 dB SNR when IV-DFE is used for the main equalizer, and the EXIT chart corresponding to $\mathbf{h_3}$ at a 13 dB SNR when TI-BiDFE is used for the main equalizer, respectively. As the figures show, both SISE algorithms widen the EXIT chart tunnels with aid of the branch equalizers while the trajectory of the single main equalizer itself tends to be stuck at the bottleneck region. Accordingly, both SISE algorithms reach the maximum MI value at a considerably smaller number of steps than the single equalizer scheme, indicating a faster convergence for the SISE schemes.

It might be worthwhile to observe that SISE 1 sometimes shows a worse BER performance than SISE 2 as seen in Figs. 4.6 and 4.7, although SISE 1 exhibits as clear an MI trajectory towards the maximum MI as SISE 2 does. This is due to the fact that the MI in the EXIT chart depends on the overall quality of the extrinsic LLRs and some erroneously generated extrinsic LLRs have little effect on the MI, while this is not true for BER performance.

4.4 Summary

In this chapter, we proposed self-iterating soft equalizers which can be further employed in turbo equalization systems to improve performance in very severe ISI channels. The proposed algorithms are designed to utilize the extrinsic information of other serially concatenated suboptimal equalizers by reducing correlation on the information generated by other equalizers. The proposed algorithms show robust performance, even when the constituent suboptimal equalizers are individually weak. The proposed SISE schemes also provide good BER performance in uncoded systems.

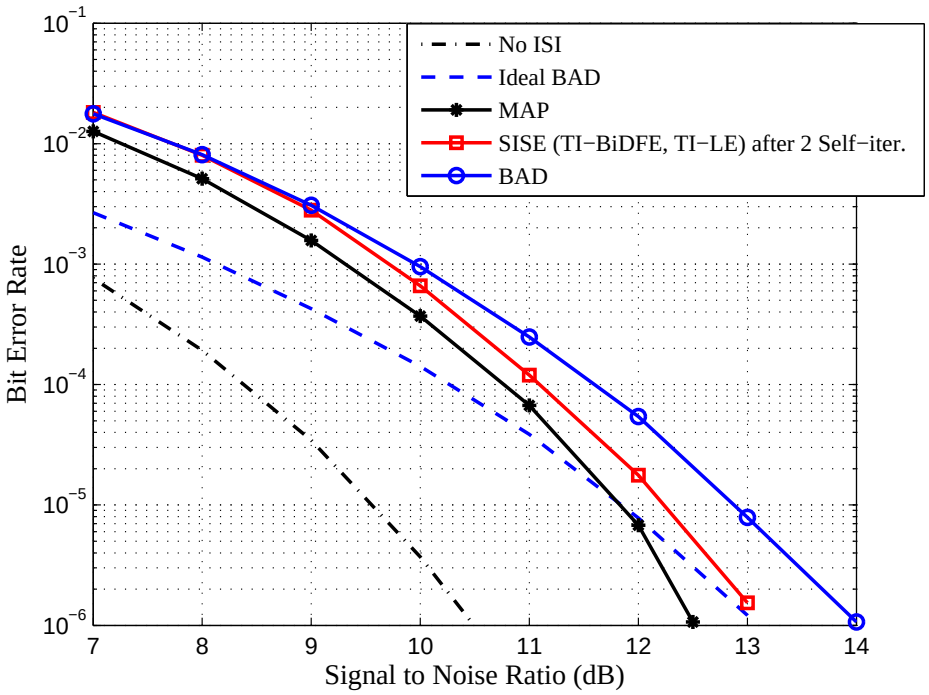


Figure 4.5: BER Curves on the Channel $\mathbf{h_0}$.

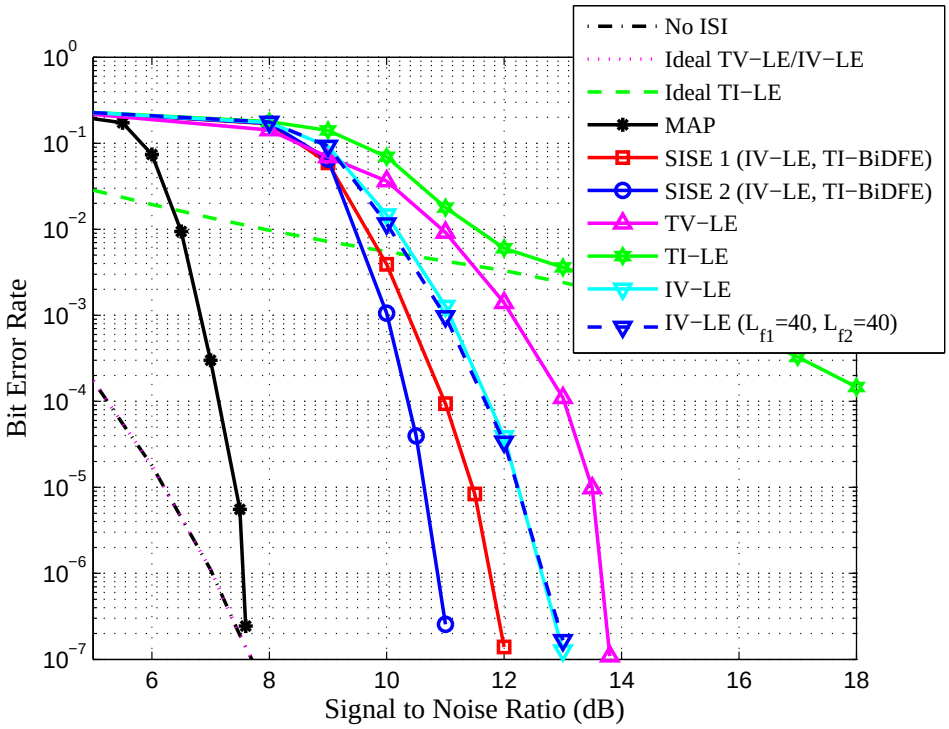


Figure 4.6: LE based BER Curves on the Channel $\mathbf{h_2}$ after 20 outer iterations.

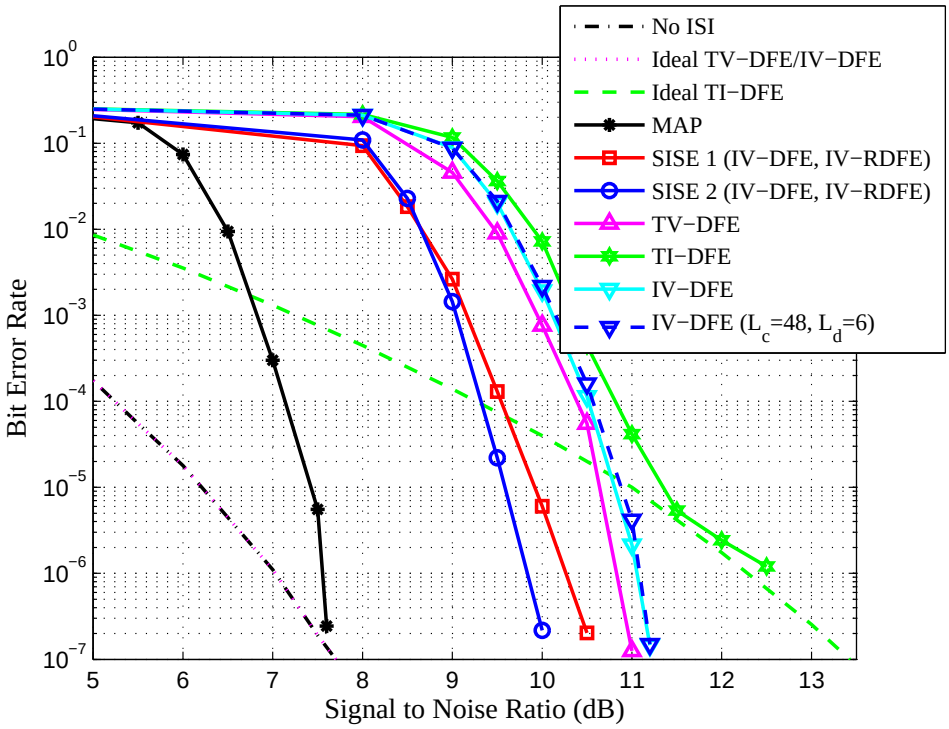


Figure 4.7: DFE based BER Curves on the Channel $\mathbf{h_2}$ after 20 outer iterations.

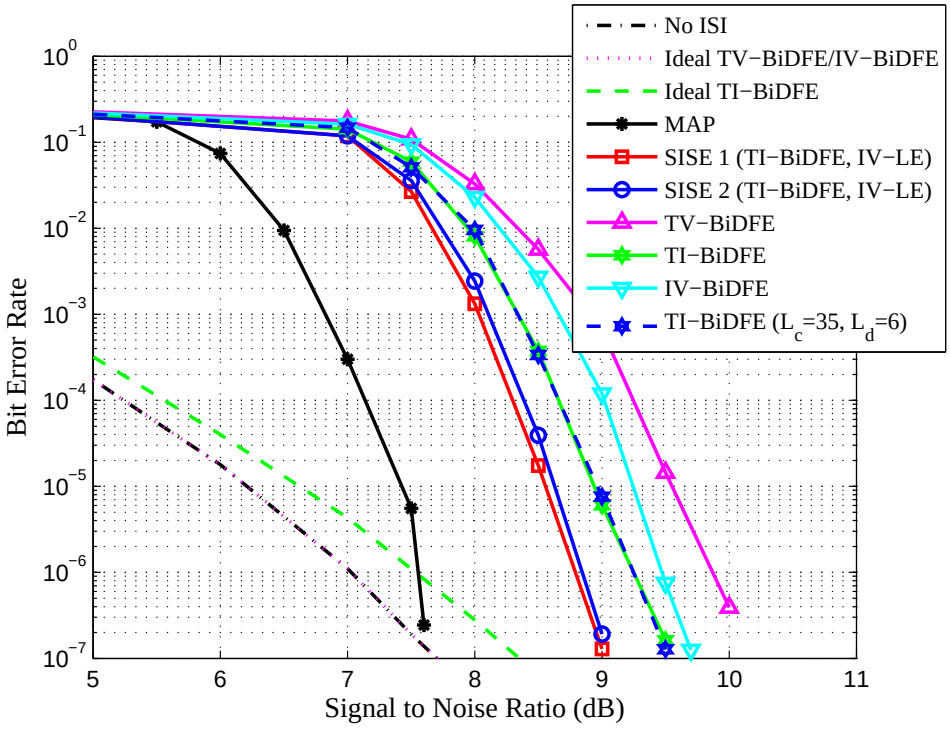


Figure 4.8: BiDFE based BER Curves on the Channel $\mathbf{h_2}$ after 20 outer iterations.

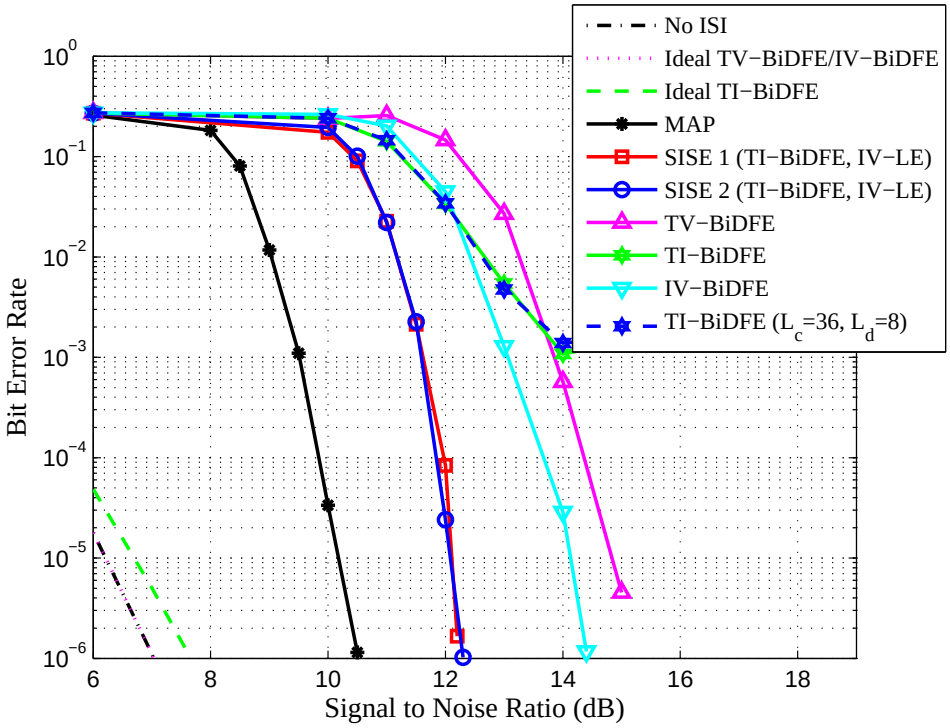


Figure 4.9: BiDFE based BER Curves on the Channel $\mathbf{h_3}$ after 20 outer iterations.

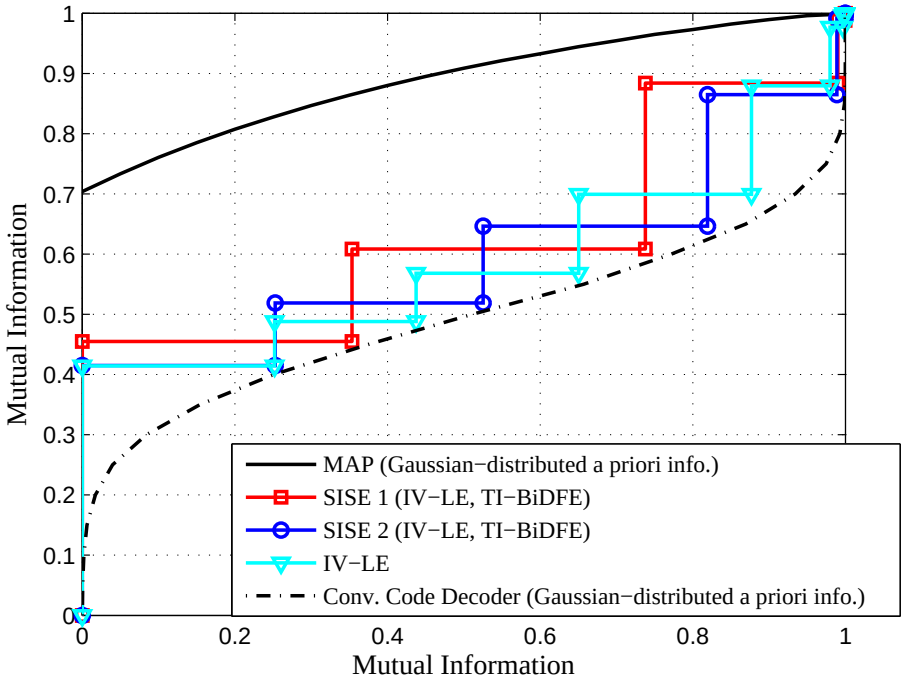


Figure 4.10: LE based EXIT Chart on the Channel $\mathbf{h_2}$ at a 12 dB.

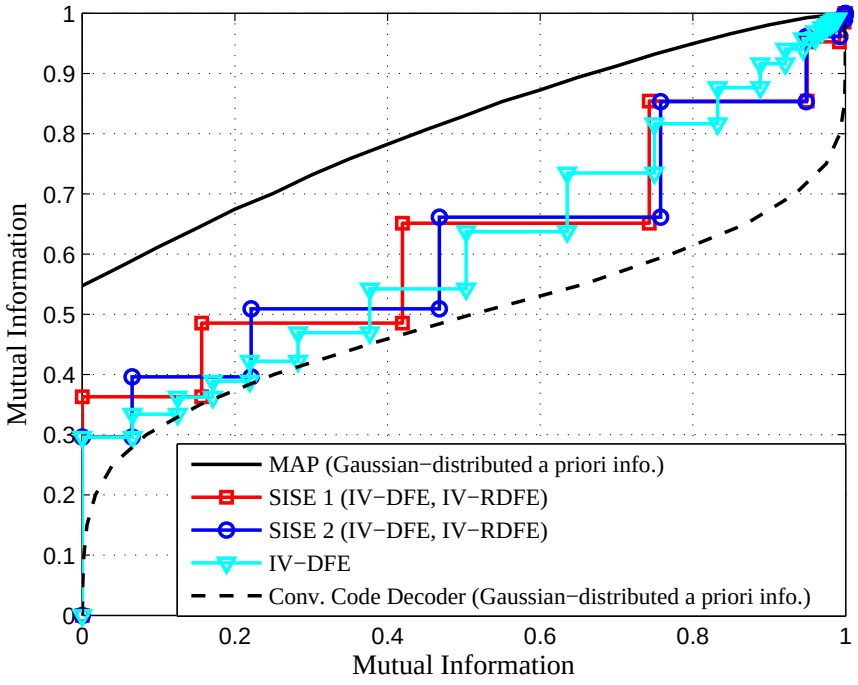


Figure 4.11: DFE based EXIT Chart on the Channel $\mathbf{h_2}$ at a 10 dB.

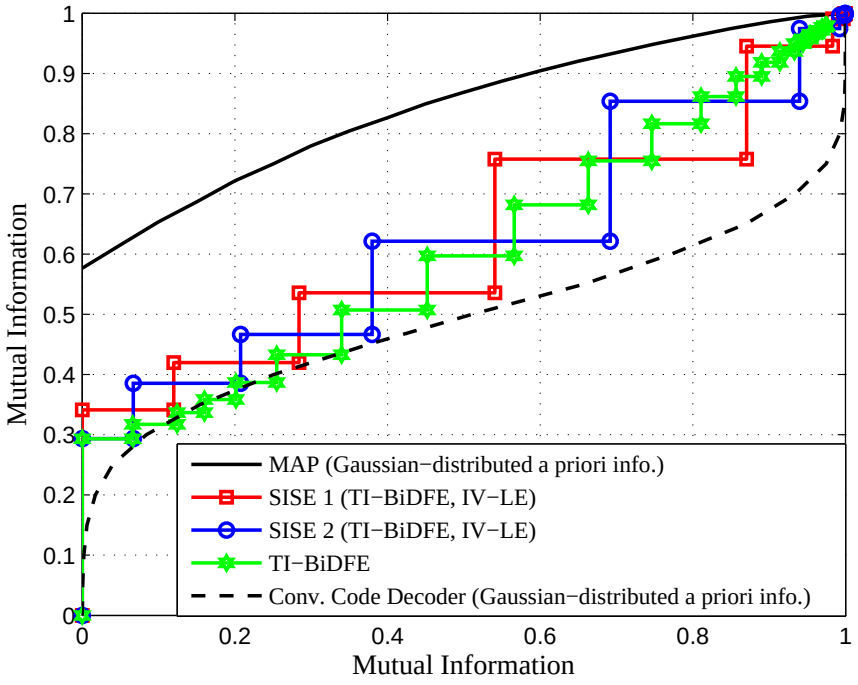


Figure 4.12: BiDFE based EXIT Chart on the Channel $\mathbf{h_3}$ at a 13 dB.

Chapter 5

Computing Lower Bounds on the Information Rate of ISI Channels

In this chapter, we introduce the easily computable analytical expressions for lower bounds to SIR. It is well known that $I(X; X + S + N)$ lower-bounds the SIR if $S + N$ is the noise component of unbiased infinite-length MMSE-DFE filter output after ideal postcursor ISI cancellation [35]. Therefore, we also focus on the easily computable analytical expressions for lower bounds to $I(X; X+S+N)$ and thus to the SIR. The bounds we develop here are fairly tight, with their tightness generally enhanced with increasing computational load (which in the end still remains small). Our approach is to first define a "mismatched" mutual information (MI) function based on the "mismatched" entropy that takes the log operation not on the actual underlying probability density but on the Gaussian density with the same variance. We then prove that this "mismatched" MI is always less than or equal to $I(X; X + S + N)$. We further bound this function from below so that the final bound can be evaluated using numerical integration. The bound is basically evaluated by computing a few single-dimensional integrals. This is in contrast to the Hirt bound that computes a single multi-dimensional integral of very high dimension. Our bound computation also requires the evaluation of sum of the absolute

values of the linear coefficients that form S as well as the identification of dominant coefficient values, if they exist. At a reasonable overall computational load, our bounds are shown to be for all practical purposes as tight as the Shamai-Laroia conjecture for many practical ISI channels. This chapter is written based on [70].

5.1 A Provable Lower Bound to the Symmetrical Information Rate

We first present a provable lower bound to $I(X;Y)$ where $Y = X + \sum_{k=1}^{L} d_{-k}X_k + N = X + S + N$. The symbols X and X_k are all independently and uniformly drawn. The linear coefficients d_{-k}'s are related to the channel impulse response and will be specified in Section 5.3. Let $V = S + N$ so we can write $Y = X + V$. Note that V is a Gaussian mixture. Also let $Z = X + G$ where G is a zero mean Gaussian with variance matching that of V, i.e., $\sigma_G^2 = \sigma_V^2$.

Definition 1 ("Mismatched" MI (MMI) Function) *Define*

$$I'(X;Y) \triangleq H'(Y) - H'(V) \tag{5.1}$$

where

$$H'(Y) \triangleq -\int_{-\infty}^{\infty} f_Y(t) \log f_Z(t) dt,$$
$$H'(V) \triangleq -\int_{-\infty}^{\infty} f_V(t) \log f_G(t) dt$$

and $f_Y(t)$, $f_V(t)$, $f_Z(t)$, and $f_G(t)$ are the probability density functions (pdfs) of the RVs, Y, V, Z, and G, respectively. Note that the "mismatched" entropy functions $H'(Y)$ and $H'(V)$ are defined based the log *operation applied not to the actual underlying pdf $f_V(t)$ but rather to the "mismatched" Gaussian pdf $f_G(t)$.*

Lemma 1 *Given the MMI function defined as above, we have*

$$I'(X;Y) \leq I(X;Y). \tag{5.2}$$

Proof 1 *See Section 5.5.1.*

Let us now take a close look at this MMI function $I'(X;Y)$ and develop some insights into its behavior. Let the variances of V, S, and N be σ_V^2, σ_S^2, and σ_N^2 respectively. Further assume that the RVs, X, V, S, and N are all real-valued. We will also assume a binary input alphabet. These assumptions are not necessary for our development but make the presentation clearer as well as less cluttered. We will simply state the results in Section 5.2.3 for a non-binary/complex-valued example. We also denote $m_i = \sum_{k=1}^{L} d_{-k} X_k$ for $i = 1, 2, \ldots, 2^L$ since $\{X_k\}_{k=1}^{L}$ can have 2^L different sequences. Naturally, the pdfs of RVs V and G can be written as

$$f_V(t) = 2^{-L} \sum_{i=1}^{2^L} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(t-m_i)^2}{2\sigma_N^2}\right)$$
$$f_G(t) = \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{t^2}{2\sigma_V^2}\right).$$

Proposition 1 *Denoting* $\rho_i \triangleq m_i/\sqrt{P_X}$ *and* $\tau \triangleq (t-m_i)/\sigma_N$, *letting* ρ_k^+*'s to mean the positive-half subset of* ρ_i*'s, and defining* $R \triangleq P_X/\sigma_V^2$ *and* $\phi \triangleq \sigma_N/\sigma_V$, *the MMI function can be rewritten as* $I'(X;Y) = \log 2 - F$ *with the new definition*

$$\begin{aligned} F &\triangleq 2^{-L} \sum_{i=1}^{2^L} \mathrm{E}_\tau \left[\log\left\{1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \\ &= \mathrm{E}_{\rho,\tau} \left[\log\left\{1 + e^{-2R\rho} e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \qquad (5.3a) \\ &= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2R\rho_k^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \\ &= \mathrm{E}_{\rho^+,\tau} \left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2R\rho^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right]. \qquad (5.3b) \end{aligned}$$

A detailed derivation is given in Section 5.5.2. The position m_i of the ith Gaussian pdf of the mixture $f_V(t)$ is expressed as a dimensionless quantity: $\rho_i = m_i/\sqrt{P_X}$, with

the normalization by the square root of the input power. Because of the symmetric nature of $f_V(t)$, ρ_i occurs in equal-magnitude, opposite-polarity pairs. The expectation is initially over τ, which is considered a zero-mean unit-variance Gaussian random variable when contained inside the argument of the expectation operator. The expectation operator in this case can simply be viewed as a short-hand notation as in

$$\mathrm{E}_\tau\left[p(\tau)\right] = \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} p(\tau) d\tau.$$

In (5.3a) and (5.3b), however, ρ (or ρ^+) is also treated as a RV and the expectation is over both τ and ρ (or τ and ρ^+) as the double subscripts indicate. Given the pdfs of τ, ρ and ρ^+, the computation of the expectation now involves numerical evaluation of a double integral. Note that in (5.3a) ρ is a discrete-valued random variable distributed according to $f_\rho(t)$, which denotes the probability distribution of $\rho = (1/\sqrt{P_X}) \sum_{k=1}^{L} d_{-k} X_k$ and ρ^+ is a discrete-valued random variable distributed according to $2f_\rho(t)u(t)$ where $u(t)$ is a step function. Also, notice that $\cosh(2R\rho^+) \geq 1$ and $\phi \leq 1$. Since it is not easy to find $f_\rho(t)$ when L is large, evaluating (5.3a) or (5.3b) is difficult in general.

It is insightful to compare F with

$$\begin{aligned}
F_{SLC} &\triangleq \log 2 - C_{SLC}(R) \\
&= \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log\left\{1 + e^{-2\sqrt{R}\tau - 2R}\right\} d\tau \\
&= \mathrm{E}_\tau\left[\log\left\{1 + e^{-2\sqrt{R}\tau - 2R}\right\}\right] \qquad (5.4a) \\
&= \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2e^{-2\sqrt{R}\tau - 2R} + e^{-4\sqrt{R}\tau - 4R}\right\}\right] \qquad (5.4b)
\end{aligned}$$

where $C_{SLC}(R)$ is the SIR of the binary-input Gaussian channel with SNR given by $R \triangleq P_X/\sigma_V^2$ and is the well-known SLC. The function F_{SLC} quantifies the gap between the SLC and the maximum attainable capacity for any binary channel with infinite SNR, namely, 1 bit/channel use. Comparing the expressions for F in (5.3b) and F_{SLC} in (5.4b), we see that if $\rho^+ = 0$ so that $\phi = 1$, then $F = F_{SLC}$, and $I'(X;Y)$ and the SLC both become equal to $I(X;Y)$. Also, if the discrete RV ρ converges to a

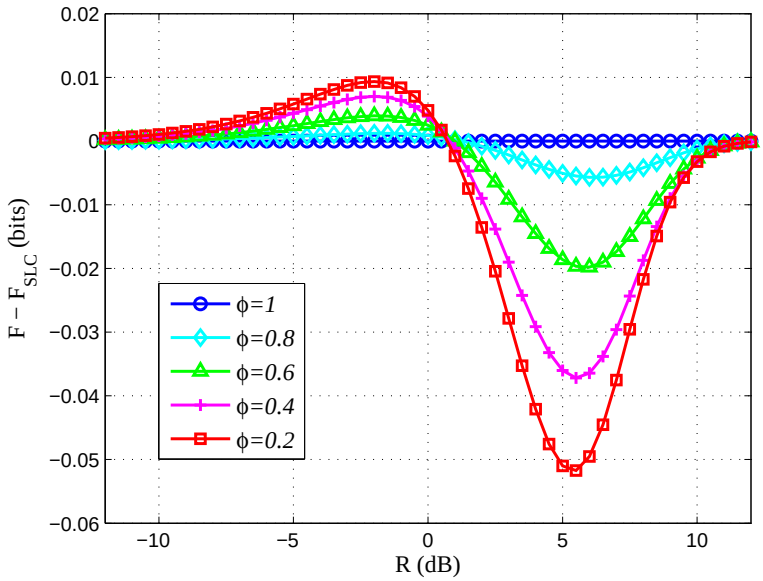


Figure 5.1: $F - F_{SLC}$ as a function of R for a uniform ρ.

Gaussian random variable (in cumulative distribution), then again we get $F = F_{SLC}$ and $I'(X;Y) = C_{SLC}(R) = I(X;Y)$.

Furthermore, that $\rho^+ \geq 0$ in $(5.3b)$ makes F larger while the factor ϕ being less than 1 has an effect of decreasing F as it increases. If $I'(X;Y) = \log 2 - F$ is to be a tight lower bound to $I(X;Y)$, then F needs to be small. The important question is: how does F overall compare with F_{SLC}, over all interested range of SNR? Since it is already proved that $I'(X;Y) = \log 2 - F$, if $F \leq F_{SLC}$ for some R values, then clearly $C_{SLC}(R) = \log 2 - F_{SLC} \leq I(X;Y)$ at those SNRs, i.e., the SLC holds true at least at these SNRs.

While exact computation of $(5.3b)$ requires in general obtaining all possible positive-side values of $\rho = (1/\sqrt{P_X}) \sum_{k=1}^{L} d_{-k} X_k$ and thus can be computationally intense for large L, in the cases where we know the functional form of the distribution for ρ, evaluation of $(5.3a)$ or $(5.3b)$ is easy; the behavior of F under different ρ distributions offers useful insights.

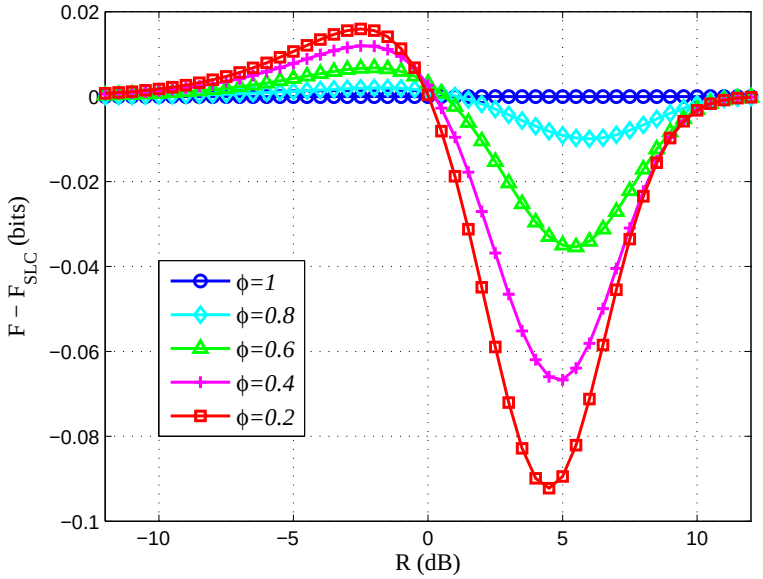


Figure 5.2: $F - F_{SLC}$ as a function of R for a two-valued ρ.

First try a uniform distribution for ρ. For a uniformly distributed discrete random variable ρ from $-K\Delta = -|\rho|_{\max}$ to $K\Delta = |\rho|_{\max}$ with a gap Δ between delta functions in the pdf, we have

$$\sigma_S^2 = \frac{2P_X\Delta^2}{2K+1}\sum_{i=1}^{K} i^2 = \frac{P_X\Delta^2 K(K+1)}{3} = \frac{P_X|\rho|_{\max}(|\rho|_{\max}+\Delta)}{3}$$

which makes

$$\phi^2 = \frac{\sigma_N^2}{\sigma_N^2+\sigma_S^2} = 1 - \frac{\sigma_S^2}{\sigma_V^2} = 1 - \frac{R\Delta^2 K(K+1)}{3} = 1 - \frac{R|\rho|_{\max}(|\rho|_{\max}+\Delta)}{3}.$$

Fig. 5.1 shows F and F_{SLC} plotted with $K = 1000$ as functions of R for various values of ϕ. We also consider a simple case involving only a single coefficient d_{-1}, in which case ρ takes only two possible values, e.g., $\rho = \pm\sqrt{(1-\phi^2)/R}$. The plots of F and F_{SLC} for this case are shown against R for different values of ϕ in Fig. 5.2. Figs. 5.1 and 5.2 point to similar behaviors of F versus F_{SLC}. Namely, F becomes smaller than F_{SLC} as ϕ decreases for a range of R values. At these R values, the provable lower bound $I'(X;Y)$ is apparently tighter than the SLC, with respect to the SIR.

5.2 Bounding F

Exact computation of F in general is not easy, especially when L goes to infinity. We thus resort to bounding F with expressions that can easily be computed. An upper bound on F will provide a lower bound on $I'(X;Y)$ and thus on $I(X;Y)$. Lower bounds on F are also derived to see if they can get smaller than F_{SLC}. If so, this would mean $I'(X;Y) = \log 2 - F$ is larger than $C_{SLC}(R) = \log 2 - F_{SLC}$, i.e., our bound is tighter than the SLC.

5.2.1 Simple Bounds

Since $\log\left(1+2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)$ is convex in ρ^+, its integral function with respect to τ, $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$, is also convex in ρ^+. Moreover, this function increases as ρ^+ increases. Accordingly, we can develop bounds on F. The first simple upper bound is

$$F^{u1} \triangleq T\left(|\rho|_{\max},\theta\right)\Big|_{\theta=\sigma_\rho} \tag{5.5}$$

where, for a given $|\rho|_{\max}$, the function $T\left(|\rho|_{\max},\theta\right)$ represents a straight line passing through two points of the function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\cosh(2R\theta)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ at $\theta = 0$ and at $\theta = |\rho|_{\max}$. Note that $|\rho|_{\max} \triangleq \max|\rho_i| = \sum_{k=1}^{L}|d_{-k}|$ and σ_ρ is the standard deviation of RV ρ.

Similarly, $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\alpha e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ is a concave and increasing function of $\alpha \triangleq \cosh(2R\rho^+)$. Based on this property, we can develop another upper bound.

$$F^{u2} \triangleq \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2(s\sigma_\rho+1)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \tag{5.6}$$

where $s = \left(\cosh(2R|\rho|_{\max})-1\right)/|\rho|_{\max}$, the slope of a straight line connecting two points $(0,1)$ and $(|\rho|_{\max},\cosh(2R|\rho|_{\max}))$.

A lower bound on F can also be obtained that can help shed lights on how tight the upper bounds on F are. Using the convexity of $\mathrm{E}_\tau\left[\log\left(1+e^{-2R\rho}e^{-2\phi\sqrt{R}\tau-2R}\right)\right]$ in ρ,

the simple lower bound of F is

$$F^l \triangleq \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]. \tag{5.7}$$

Detailed derivations of (5.5), (5.6), and (5.7) are given in Section 5.5.3.

5.2.2 Tightened Bounds Based on Cluster Identification

The above bounds can be tightened up by identifying clusters in the Gaussian mixture $f_V(t)$. In practical ISI channels, $f_V(t)$ often consists of clusters. This is due to the fact that the coefficient set d_{-k}'s typically contains a few dominating coefficients plus many small terms. Assuming there are M dominating coefficients among d_{-k}'s, we can let $\rho_k = \lambda_n + \mu_i$ where $n = 1, 2, \ldots, 2^M$, $i = 1, 2, \ldots, 2^{L-M}$, and $k = (n-1)2^{L-M} + i$. Since X_k is an i.u.d. RV, λ and μ are independent so that $\sigma_\rho^2 = \sigma_\lambda^2 + \sigma_\mu^2$ where σ_λ^2 and σ_μ^2 denote the variance of RVs λ and μ, respectively. Notice that λ_n can be viewed as the position of a specific cluster while μ_i points to a specific Gaussian pdf out of 2^{L-M} Gaussian pdf's symmetrically positioned around λ_n.

Therefore, assuming there are 2^M clusters of Gaussian pdfs, the upper bound F^{u1} can be tightened as

$$F_M^{u1} \triangleq 2^{-M}\sum_{n=1}^{2^M} T_n\left(|\mu|_{\max}, \theta\right)\Big|_{\theta=\sigma_\mu} \tag{5.8}$$

where, for a given $|\mu|_{\max}$, the function $T_n(|\mu|_{\max}, \theta)$ is a straight line that passes through the two points of the convex function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\theta\right)e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]$ at $\theta = 0$ and $\theta = |\mu|_{\max}$, σ_μ is the standard deviation of RV μ defined as $\sigma_\mu = \sqrt{\sigma_\rho^2 - \sigma_\lambda^2}$, and $|\mu|_{\max} = |\rho|_{\max} - |\lambda|_{\max}$.

Another form of tightened upper bound based on F^{u2} is obtained as

$$F_M^{u2} \triangleq 2^{-M}\sum_{n=1}^{2^M}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\left(s_M\sigma_\mu+1\right)e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \tag{5.9}$$

where $s_M = \left(\cosh(2R|\mu|_{\max}) - 1\right)/|\mu|_{\max}$.

The lower bound F^l can also be tightened similarly based on the cluster identification:

$$F_M^l \triangleq 2^{-(M-1)} \sum_{k=1}^{2^{M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2R\lambda_k^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \tag{5.10}$$

where λ_k^+'s form the positive-half subset of λ_n's. The detail derivations of (5.8), (5.9), and (5.10) can be found in Section 5.5.4.

5.2.3 Bounds for Complex Channels with QPSK Inputs

In the previous subsections, ISI coefficients and noise samples are assumed to be real-valued with the channel inputs being the binary phase shift keying (BPSK) signal. In this subsection, we provide a complex-valued example along with the channel inputs taken from a quadrature phase shift keying (QPSK) quaternary alphabet, i.e., $X_k \in \left\{ \sqrt{\frac{P_X}{2}}(\pm 1 \pm j) \right\}$. The extension to larger alphabets should be straightforward.

Denoting the real and imaginary parts of complex number a by $a^{(r)}$ and $a^{(i)}$ respectively, i.e., $a = a^{(r)} + ja^{(i)}$, and $m_i = \sum_{k=1}^{L} d_{-k} X_k$ for $i = 1, 2, \ldots, 4^L$, the pdf's of complex random variables V and G are given as

$$\begin{aligned}
f_V(t) &= 4^{-L} \sum_{i=1}^{4^L} \frac{1}{\pi\sigma_N^2} \exp\left(-\frac{|t - m_i|^2}{\sigma_N^2} \right) \\
&= 4^{-L} \sum_{i=1}^{4^L} \left\{ \frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(r)} - m_i^{(r)}\right)^2}{\sigma_N^2} \right) \frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(i)} - m_i^{(i)}\right)^2}{\sigma_N^2} \right) \right\} \\
f_G(t) &= \frac{1}{\pi\sigma_V^2} \exp\left(-\frac{|t|^2}{\sigma_V^2} \right) \\
&= \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(r)}\right)^2}{\sigma_V^2} \right) \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(i)}\right)^2}{\sigma_V^2} \right).
\end{aligned}$$

Then, for the SLC, we write

$$\begin{aligned} F_{SLC} &\triangleq \log 4 - C_{SLC}(R) \\ &= 2\int_{-\infty}^{\infty} \frac{e^{-\tau^2}}{\sqrt{\pi}} \log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\} d\tau \\ &= 2\mathrm{E}_\tau\left[\log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\}\right] \qquad (5.11a) \\ &= 2\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2e^{-2\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2R}\tau - 4R}\right\}\right] \qquad (5.11b) \end{aligned}$$

where $\mathrm{E}_\tau\left[p(\tau)\right] = \int_{-\infty}^{\infty} \pi^{-1/2} e^{-\tau^2} p(\tau) d\tau$.

The function F is also given as

$$\begin{aligned} F &\triangleq 4^{-L} \sum_{i=1}^{4^L} \left(\mathrm{E}_\tau\left[\log\left\{1 + e^{-2\sqrt{2}R\rho_i^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R}\right\}\right] \right. \\ &\qquad \left. + \mathrm{E}_\tau\left[\log\left\{1 + e^{-2\sqrt{2}R\rho_i^{(i)}} e^{-2\phi\sqrt{2R}\tau - 2R}\right\}\right] \right) \\ &= 2\mathrm{E}_{\rho^{(r)},\tau}\left[\log\left\{1 + e^{-2\sqrt{2}R\rho^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R}\right\}\right] \qquad (5.12a) \\ &= 4^{-(L-1)} \sum_{k=1}^{4^{L-1}} 2\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\rho_k^{(r)+}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \\ &= 2\mathrm{E}_{\rho^{(r)+},\tau}\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\rho^{(r)+}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \qquad (5.12b) \end{aligned}$$

where $\rho_i^{(r)} \triangleq m_i^{(r)}/\sqrt{P_X}$, $\rho_i^{(i)} \triangleq m_i^{(i)}/\sqrt{P_X}$, and $\rho_k^{(r)+}$'s and $\rho_k^{(i)+}$'s denote the positive-half subset of $\rho_i^{(r)}$'s and $\rho_i^{(i)}$'s respectively. The equality of $(5.12a)$ holds because the pdf of $\rho^{(r)}$ is identical to the pdf of $\rho^{(i)}$.

Then, the upper bound based on F^{u1} can be derived in a similar way as

$$F_M^{u1} \triangleq 4^{-M} \sum_{n=1}^{4^M} 2T_n^{(r)}\left(|\mu^{(r)}|_{\max}, \theta\right)\Big|_{\theta = \frac{\sigma_\mu}{\sqrt{2}}} \qquad (5.13)$$

where, for a given $|\mu^{(r)}|_{\max}$, $T_n^{(r)}(|\mu^{(r)}|_{\max}, \theta)$ denotes a straight line that passes through the two points of the function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\theta\right) e^{-2\sqrt{2}R\lambda_n^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_n^{(r)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and at $\theta = |\mu^{(r)}|_{\max}$. Note that $|\mu^{(r)}|_{\max} = |\mu|_{\max}/\sqrt{2}$ and the variance of $\mu^{(r)}$ is equal to $\sigma_\mu^2/2$ since the pdfs of $\lambda^{(r)}$ and $\mu^{(r)}$ are identical to the pdfs of $\lambda^{(i)}$ and $\mu^{(i)}$, respectively.

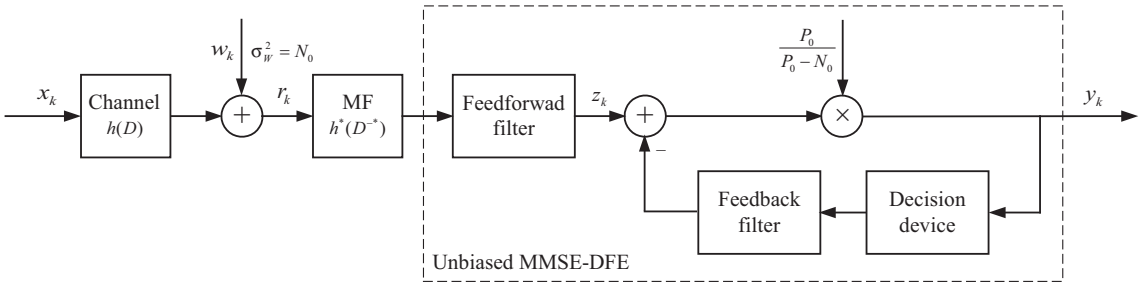


Figure 5.3: System Model of ISI channels.

A second upper bound on F is given as

$$F_M^{u2} \triangleq 4^{-M} \sum_{n=1}^{4^M} 2\mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\left(\frac{s_M^{(r)} \sigma_\mu}{\sqrt{2}} + 1 \right) e^{-2\sqrt{2}R\lambda_n^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_n^{(r)}} e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \tag{5.14}$$

where $s_M^{(r)} = \left(\cosh(2\sqrt{2}R|\mu^{(r)}|_{\max}) - 1\right) / |\mu^{(r)}|_{\max}$.

Finally, a lower bound to F can be shown to be

$$F_M^{l} \triangleq 4^{-M} \sum_{k=1}^{4^M/2} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2\sqrt{2}R\lambda_k^{(r)+}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \tag{5.15}$$

where $\lambda_k^{(r)+}$'s form the positive-half subset of $\lambda_n^{(r)}$'s.

5.3 Application to ISI Channels and Numerical Examples

5.3.1 The ISI Channel and MMSE-DFE

Fig. 5.3 shows the discrete-time equivalent system model of the finite-ISI channel with the infinite-length feedforward filter of the unbiased MMSE-DFE preceded by the matched filter (MF) for the channel. The discrete-time MF output of Fig. 5.3 is identical to the baud-rate sampled output of the continuous-time MF applied to the

continuous-time channel, under the assumption that the channel is strictly limited to the Nyquist band.

We also assume that the receiver knows the D-transform of the finite-ISI channel response, $h(D)$, x_k is an i.u.d. input sequence and w_k is additive white Gaussian noise (AWGN) with variance $\sigma_W^2 = N_0$. Furthermore, r_k is the channel output sequence, z_k is the output sequence of the infinite-length MMSE-DFE feedforward filter and y_k is the unbiased MMSE-DFE output after ideal postcursor ISI cancellation.

Denoting $X = x_0$, $X_k = x_k$, and $Y = y_0$, the output of the the unbiased MMSE-DFE with ideal feedback [52] is

$$Y = X + \sum_{k=1}^{\infty} d_{-k} X_k + N = X + S + N = X + V$$

where N is the Gaussian noise sample observed at the DFE forward filter output and $d_{-k}X_k$ is the precursor ISI sequence. Note we are assuming stationary random processes. It is well-known that the D-transform of the precursor ISI taps d_{-k} is given by [52]

$$d(D) = \frac{N_0}{P_0 - N_0}\left(1 - \frac{1}{g^*(D^{-*})}\right) \tag{5.16}$$

where P_0 is such that $\log P_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log R_{ss}(e^{-j\theta})d\theta$ and $g^*(D^{-*})$ is obtained from spectral factorization: $R_{ss}(D) = P_X R_{hh}(D) + N_0 = P_0 g(D) g^*(D^{-*})$ with $R_{hh}(D) = h(D)h^*(D^{-*})$. Notice that a convenient numerical spectral factorization algorithm exists for recursively computing the coefficients of $g^*(D^{-*})$ [66, 67].

Accordingly, the variances of V, N, and S are given as

$$\begin{aligned}
\sigma_V^2 &= \frac{P_X N_0}{P_0 - N_0} \\
\sigma_N^2 &= \frac{P_X P_0 N_0}{2\pi\,(P_0 - N_0)^2}\int_{-\pi}^{\pi} \frac{R_{hh}(e^{-j\theta})}{R_{hh}(e^{-j\theta}) + N_0/P_X} d\theta \\
\sigma_S^2 &= \sigma_V^2 - \sigma_N^2.
\end{aligned}$$

We can obtain $|\rho|_{\max}$ by the absolute summation of the inverse D-transform of $d(D)$ if the feedforward filter of MMSE-DFE is stable, i.e., $\sum_{k=1}^{\infty} |d_{-k}| < \infty$. Let us first

consider the case where $d(D)$ has P multiple first-order poles, p_i for $i = 1, 2, \ldots, P$. Then, $|\rho|_{\max}$ can be obtained by the partial fraction method since $d(D)$ is a rational function. In other words, the inverse D-transform of individual fraction terms can be found and then added together to form d_{-k}. Denoting $a(D) = \frac{1}{g^*(D^{-*})} = \sum_{i=1}^{P} \frac{c_i}{1-p_i D^{-1}}$, the sequence a_{-k} is given as $a_{-k} = \sum_{i=1}^{P} c_i p_i^k$. Therefore,

$$\begin{aligned}
|\rho|_{\max} &= \frac{1}{\sqrt{P_X}} \sum_{k=1}^{\infty} |d_{-k} X_k| = \sum_{k=1}^{\infty} |d_{-k}| \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} |a_{-k}| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} \left| \sum_{i=1}^{P} c_i p_i^k \right| \right) \\
&\leq \frac{N_0}{(P_0 - N_0)} \left(\sum_{i=1}^{P} \sum_{k=1}^{\infty} \left| c_i p_i^k \right| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{i=1}^{P} \frac{|c_i p_i|}{1 - |p_i|} \right). \qquad (5.17)
\end{aligned}$$

The upper bound of $|\rho|_{\max}$ can be also tightened by identifying the first K dominant taps:

$$\begin{aligned}
|\rho|_{\max} &= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} \left| \sum_{i=1}^{P} c_i p_i^k \right| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{K} \left| \sum_{i=1}^{P} c_i p_i^k \right| + \sum_{k=K+1}^{\infty} \left| \sum_{i=1}^{P} c_i p_i^k \right| \right) \\
&\leq \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{K} \left| \sum_{i=1}^{P} c_i p_i^k \right| + \sum_{i=1}^{P} \sum_{k=K+1}^{\infty} \left| c_i p_i^k \right| \right) \\
&= \sum_{k=1}^{K} |d_{-k}| + \frac{N_0}{(P_0 - N_0)} \left(\sum_{i=1}^{P} \frac{\left| c_i p_i^{K+1} \right|}{1 - |p_i|} \right). \qquad (5.18)
\end{aligned}$$

For the case of the multiple-order poles of $d(D)$, the upper bound of $|\rho|_{\max}$ can be also obtained in a similar way using the triangle inequality $|a + b| \leq |a| + |b|$.

The SIR or the i.u.d. capacity (bits/channel use) for any finite-ISI channel corrupted

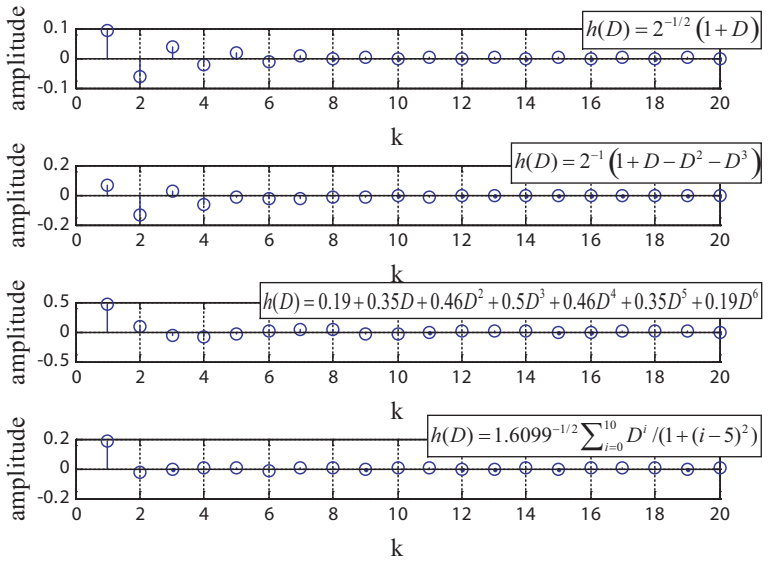


Figure 5.4: First 20 Precursor taps after unbiased MMSE-DFE at SNR=10 dB for four example channels.

by Gaussian noise is given [68] as

$$\begin{aligned}
\text{SIR} &\triangleq \lim_{N\to\infty} \frac{1}{2N+1} I\left(\{x_k\}_{-N}^{N}; \{r_k\}_{-N}^{N}\right) \\
&\geq \lim_{N\to\infty} \frac{1}{2N+1} I\left(\{x_k\}_{-N}^{N}; \{z_k\}_{-N}^{N}\right) \qquad (5.19) \\
&\geq I\left(x_0; z_0 | \{x_k\}_{-\infty}^{-1}\right) \qquad (5.20) \\
&= I\left(X;Y\right) \qquad (5.21)
\end{aligned}$$

where $\{u_k\}_{N_1}^{N_2} = \{u_k, k = N_1, N_1+1, \ldots, N_2\}$. The inequality in (5.19) holds due to the data processing theorem (equality holds if the MMSE-DFE feedforward filter is invertible). The inequality of (5.20) can be obtained by applying the chain rule of mutual information and assuming stationarity [35]. The equality (5.21) is valid because known post-cursor ISI can simply be subtracted out without affecting capacity.

5.3.2 Numerical Results

Now, let us examine the particular ISI channels, $h(D) = 2^{-1/2}(1+D)$, $h(D) = 2^{-1}(1 + D - D^2 - D^3)$ and $h(D) = 0.19 + 0.35D + 0.46D^2 + 0.5D^3 + 0.46D^4 + 0.35D^5 + 0.19D^6$, which are well-known and previously investigated in [26, 34, 35], and $h(D) = 1.6099^{-1/2} \sum_{i=0}^{10} D^i/(1 + (i-5)^2)$, which was considered in [30]. The first 20 precursor ISI tap values are computed and shown in Fig. 5.4 for these example channels. In addition, we consider a complex-valued partial response channel: $h(D) = 2^{-1}\{(1+j) + (1-j)D\}$. The channel inputs are binary, except the complex-valued channel for which the inputs are assumed quaternary.

Since the infinite-length MMSE-DFE is used, i.e., $L = \infty$, the probability distribution of ρ is not available generally. Hence the lower bounds $C_{L1,M} = \log 2 - F_M^{u1}$ and $C_{L2,M} = \log 2 - F_M^{u2}$ along with $C_{SLC} = \log 2 - F_{SLC}$ are considered as functions of $\text{SNR} = P_X/N_0$ for different values of M. When no clustering is used, we set $M = 0$. In computing $|\rho|_{\max}$ (and thus $|\mu|_{\max}$), which was needed to calculate F_M^{u1} or F_M^{u2}, we were able to run numerical recursive spectral factorization to find all non-negligible d_{-k} coefficients relatively quickly for all channels considered, without resorting to the bounds of (5.17) or (5.18). We observed that the lower bounds, $C_{L1,M}$ and $C_{L2,M}$, produced similar results, so only $C_{L1,M}$ were chosen and plotted as $C_{L,M}$ in Figs. 5.5 through 5.9. The SIR of each channel is also obtained using the simulation-based approach [26–28].

For each capacity figure, we first plotted the SIR and C_{SLC}. We then plotted C_L for $M = 0$ and then another C_L by choosing an M value for which the C_L bound is almost as tight as the C_{SLC} conjecture (this is why the C_{SLC} curve is almost overwritten and indistinguishable in some figures). We also show for each channel how the upper and lower bounds of F close on each other as M increases. The bounds on F are shown with F_{SLC} subtracted from them in the part (b) figure. In this way, it should be clear that for those SNR values where $F^u - F_{SLC}$ becomes less than zero eventually, F is less than F_{SLC}, guaranteeing that $I'(X:Y) = \log 2 - F$ is larger than C_{SLC}. In fact,

it can be seen from the part (b) plots of Figs. 5.5 through 5.8 that this is true for the high SNR range corresponding to all rates higher than roughly 0.6 in each of the first four channels considered, although the difference $F - F_{SLC}$ is small (note the very small scale of the vertical axis in each part-b figure). In the fifth example shown in Fig. 5.9, $F^u - F_{SLC}$ is actually less than zero at all SNR values, meaning that for this channel our bound is tighter than the SLC in the entire SNR range.

A significant implication here is that whenever $I'(X:Y) = \log 2 - F$ is larger than C_{SLC}, there is an assurance that the SLC holds true. The curves of $F^l - F_{SLC}$ for different M values also provide a detailed picture of how large M should be in order for C_L to get close enough to the C_{SLC}.

Note that the computational load for evaluating the integral of (5.8) and (5.9) to obtain the bound depends exponentially on M, the number of clusters in the pdf $f_V(t)$. The computational load in computing the dominant precursor ISI taps and their magnitude sum is minimal. The results summarized in the figures indicate that in each channel considered, a relatively small value of M (and thus a reasonably low computational load) yields a bound as tight as the SLC. As a case in point, comparison of Fig. 5.8 with the results of [30] (Figure 6 of [30], specifically) gives a good idea on the usefulness of an easily computable bound such as the one presented here. At a rate 0.9, for example, one can observe from a close examination of Fig. 6 of [30] that the lower bound of [30] approaches the SIR within about 0.88 dB with 2 iterations, which would require basically running the BCJR algorithm twice on a reduced channel trellis of 64 states. In contrast, our bound based on just two clusters is about 0.84 dB away from the SIR at the same rate, as estimated from Fig. 5.8. This bound requires computation of $2^2 = 4$ single-dimensional integrals, the complexity of which amounts to virtually nothing relative to that associated with two BCJR simulation runs in the method of [30]. The simulation-based bound of [30] does narrow the gap to about 0.65 dB with five iterations, but at the expense of much more computational time.

We stress that the value of the simulation-based SIR estimation methods is not in

their ability to provide easily obtained bounds; rather they play a critical role in estimating the SIR (or capacity) with a very high accuracy, given the ample computational resources. As for providing convenient and easily computed SIR estimates or bounds, the need for analytically evaluated bounds such as the one developed in this chapter continues to be high. In particular, the question remains as to how does the proposed method perform on very long ISI channels with no dominant taps. A good example of this type of channel is the indoor wireless channel in a highly scattered multi-path setting. Unfortunately, our analysis indicates the lower bounds developed here are not very effective in this case, with their gaps to the SIR bigger than that of the SLC when computational loads are kept at reasonable levels. Easily-computed tight bounds for this type of channel remain elusive.

5.4 Summary

In this chapter, we derived a lower bound to the SIR of the ISI channel driven by discrete and finite-amplitude inputs. The approach taken was to introduce a "mismatched" mutual information function that acts as a lower bound to the symmetric information rate between the channel input and the ideal-feedback MMSE-DFE filter output. This function turns out to be tighter than the Shamai-Laroia conjecture for a practically significant range of SNR values for some example channels. We then further lower-bounded this function by another function that can be evaluated via numerical integration with a small computational load. The final computation also requires finding a few large precursor ISI tap values as well as the absolute sum of the remaining ISI terms, which can be done easily. The final lower bounds are demonstrated for a number of well-known finite-ISI channels, and the results indicate that the new bounds computed at a fairly low computational load are as tight as the SLC.

5.5 Proofs and Derivations

5.5.1 Proof of Lemma 1

We show below that $I'(X;Y) \leq I(X;Y)$. Start by writing

$$\begin{aligned} & I(X;Y) - I'(X;Y) \\ & \quad = \big(H(Y) - H'(Y)\big) - \big(H(V) - H'(V)\big) \\ & \quad = -\int_{-\infty}^{\infty} f_Y(t) \log\left(\frac{f_Y(t)}{f_Z(t)}\right) dt + \int_{-\infty}^{\infty} f_V(t) \log\left(\frac{f_V(t)}{f_G(t)}\right) dt \\ & \quad = -D\Big(f_Y(t)||f_Z(t)\Big) + D\Big(f_V(t)||f_G(t)\Big) \end{aligned} \tag{5.22}$$

where $D\Big(p(t)||q(t)\Big)$ is the Kullback-Leibler (K-L) divergence defined as

$$D\Big(p(t)||q(t)\Big) \triangleq \int_{-\infty}^{\infty} p(t) \log\left(\frac{p(t)}{q(t)}\right) dt.$$

The K-L divergence is always greater than or equal to zero and convex in pair $(p(t)||q(t))$, [69], i.e, assuming $p_1(t), p_2(t), q_1(t)$, and $q_2(t)$ are all pdfs, for $0 \leq \lambda \leq 1$, we have

$$\begin{aligned} & D\Big(\lambda p_1(t) + (1-\lambda)p_2(t)||\lambda q_1(t) + (1-\lambda)q_2(t)\Big) \\ & \quad \leq \lambda D\Big(p_1(t)||q_1(t)\Big) + (1-\lambda)D\Big(p_2(t)||q_2(t)\Big). \end{aligned} \tag{5.23}$$

For the sake of clarity, we assume that X is from the BPSK alphabet, i.e., $X \in \{\pm\sqrt{P_X}\}$. Then,

$$\begin{aligned} f_Y(t) &= \frac{1}{2}\left\{f_V(t-\sqrt{P_X}) + f_V(t+\sqrt{P_X})\right\} \\ f_Z(t) &= \frac{1}{2}\left\{f_G(t-\sqrt{P_X}) + f_G(t+\sqrt{P_X})\right\}. \end{aligned}$$

Substituting $p_1(t) = f_V(t-\sqrt{P_X})$, $p_2(t) = f_V(t+\sqrt{P_X})$, $q_1(t) = f_G(t-\sqrt{P_X})$, $q_2(t) = f_G(t+\sqrt{P_X})$, and $\lambda = 0.5$ in (5.23), we get

$$\begin{aligned} D\Big(f_Y(t)||f_Z(t)\Big) \leq \frac{1}{2}\Big\{ & D\Big(f_V(t-\sqrt{P_X})||f_G(t-\sqrt{P_X})\Big) \\ & + D\Big(f_V(t+\sqrt{P_X})||f_G(t+\sqrt{P_X})\Big)\Big\} \\ = D\Big(f_V(t)||f_G(t)\Big). & \end{aligned}$$

Accordingly, (5.22) is always greater than or equal to zero or $I'(X;Y) \leq I(X;Y)$. While this proof is for the binary alphabet, it is easy to see that the application of the pair-wise convexity of (5.23) for any i.u.d. input leads to the same conclusion.

5.5.2 Derivation of the Proposition 1

From the pdfs of RVs V and G, we can write

$$\begin{aligned} H'(V) &= -\int_{-\infty}^{\infty} f_V(t) \log f_G(t) dt \\ &= \frac{1}{2}\log\left(2\pi\sigma_V^2\right) + \int_{-\infty}^{\infty} \frac{t^2}{2\sigma_V^2} f_V(t) dt. \end{aligned} \tag{5.24}$$

Moreover, we have

$$\begin{aligned} f_Y(t) &= \frac{1}{2}\left\{ f_V(t-\sqrt{P_X}) + f_V(t+\sqrt{P_X}) \right\} \\ f_Z(t) &= \frac{1}{2}\left\{ f_G(t-\sqrt{P_X}) + f_G(t+\sqrt{P_X}) \right\} \\ &= \frac{1}{2}\left\{ \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{\left(t-\sqrt{P_X}\right)^2}{2\sigma_V^2} \right) + \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{\left(t+\sqrt{P_X}\right)^2}{2\sigma_V^2} \right) \right\} \\ &= \frac{1}{2\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t-\sqrt{P_X})^2}{2\sigma_V^2} \right) \left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2} \right) \right\} \\ &= \frac{1}{2\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t+\sqrt{P_X})^2}{2\sigma_V^2} \right) \left\{ 1 + \exp\left(\frac{2\sqrt{P_X}t}{\sigma_V^2} \right) \right\}. \end{aligned}$$

We can write $-\log f_Z(t)$ in two different ways:

$$\begin{aligned} -\log f_Z(t) &= \log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right) + \frac{(t-\sqrt{P_X})^2}{2\sigma_V^2} - \log\left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2} \right) \right\} \\ &= \log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right) + \frac{(t+\sqrt{P_X})^2}{2\sigma_V^2} - \log\left\{ 1 + \exp\left(\frac{2\sqrt{P_X}t}{\sigma_V^2} \right) \right\}. \end{aligned}$$

Thus, we have

$$
\begin{aligned}
&-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t-\sqrt{P_X})\log f_Z(t)dt \\
&\quad=\frac{1}{2}\left\{\log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)\right\} \\
&\qquad+\frac{1}{2}\int_{-\infty}^{\infty}\frac{(t-\sqrt{P_X})^2}{2\sigma_V^2}f_V(t-\sqrt{P_X})dt \\
&\qquad-\frac{1}{2}\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2}\right)\right\}f_V(t-\sqrt{P_X})dt \\
&\quad=\frac{1}{2}\left\{\log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)\right\}+\frac{1}{2}\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt \\
&\qquad-\frac{1}{2}\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t+\sqrt{P_X})\log f_Z(t)dt \\
&\quad=\frac{1}{2}\left\{\log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)\right\}+\frac{1}{2}\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt \\
&\qquad-\frac{1}{2}\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt.
\end{aligned}
$$

Accordingly,

$$
\begin{aligned}
H'(Y)&=-\int_{-\infty}^{\infty} f_Y(t)\log f_Z(t)dt \\
&=-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t-\sqrt{P_X})\log f_Z(t)dt-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t+\sqrt{P_X})\log f_Z(t)dt \\
&=\log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)+\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt \\
&\qquad-\int_{-\infty}^{\infty}\frac{1}{2}\left[\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}\right. \\
&\qquad\qquad\left.+\log\left\{1+\exp\left(\frac{2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}\right]f_V(t)dt \\
&=\log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)+\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt \\
&\qquad-\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt. \qquad (5.25)
\end{aligned}
$$

The last equality in (5.25) holds because $f_V(t)$ is an even function. Finally, from (5.24) and (5.25), we arrive at

$$\begin{aligned} I'(X;Y) &= H'(Y) - H'(V) \\ &= \log 2 - \int_{-\infty}^{\infty} f_V(t) \log\left\{1 + \exp\left(\frac{-2\sqrt{P_X}t - 2P_X}{\sigma_V^2}\right)\right\} dt. \end{aligned} \tag{5.26}$$

Now write $I'(X;Y) = \log 2 - F$ with the new definition

$$\begin{aligned} F &\triangleq \int_{-\infty}^{\infty} f_V(t) \log\left\{1 + \exp\left(\frac{-2\sqrt{P_X}t - 2P_X}{\sigma_V^2}\right)\right\} dt \\ &= 2^{-L} \sum_{i=1}^{2^L} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(t-m_i)^2}{2\sigma_N^2}\right) \log\left\{1 + \exp\left(\frac{-2\sqrt{P_X}t - 2P_X}{\sigma_V^2}\right)\right\} dt \\ &= 2^{-L} \sum_{i=1}^{2^L} \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log\left\{1 + \exp\left(\frac{-2\sqrt{P_X}\left(\tau\sigma_N + m_i\right) - 2P_X}{\sigma_V^2}\right)\right\} d\tau \\ &= 2^{-L} \sum_{i=1}^{2^L} \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log\left\{1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R}\right\} d\tau \\ &= 2^{-L} \sum_{i=1}^{2^L} \mathrm{E}_\tau\left[\log\left\{1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \end{aligned} \tag{5.27a}$$

where the third equality is obtained with a variable change $(t - m_i)/\sigma_N = \tau$ and $\rho_i \triangleq m_i/\sqrt{P_X}$, $R \triangleq P_X/\sigma_V^2$, and $\phi \triangleq \sigma_N/\sigma_V$. The expression (5.27$a$) can also be written as

$$\begin{aligned} F &= 2^{-L} \sum_{i=1}^{2^L} \mathrm{E}_\tau\left[\log\left\{1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \\ &= 2^{-L} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau\Big[\log\left\{1 + e^{-2R\rho_k^+} e^{-2\phi\sqrt{R}\tau - 2R}\right\} + \log\left\{1 + e^{2R\rho_k^+} e^{-2\phi\sqrt{R}\tau - 2R}\right\}\Big] \\ &= 2^{-L} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau\Big[\log\left\{1 + \left(e^{-2R\rho_k^+} + e^{2R\rho_k^+}\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\Big] \\ &= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2R\rho_k^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \end{aligned} \tag{5.27b}$$

where ρ_k^+'s is the positive-half subset of ρ_i's.

5.5.3 Derivation of the Simple Bounds

Due to the convexity of $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ in ρ^+, the upper bound of F can be found as

$$\begin{aligned}
F &= 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\rho_k^+\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&\le 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\left\{T\left(|\rho|_{\max},\theta\right)\Big|_{\theta=\rho_k^+}\right\} \\
&= T\left(|\rho|_{\max},\theta\right)\Big|_{\theta=2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\rho_k^+} \\
&= T\left(|\rho|_{\max},\theta\right)\Big|_{\theta=|\rho|_{\mathrm{avg}}} \\
&\le T\left(|\rho|_{\max},\theta\right)\Big|_{\theta=\sigma_\rho} \triangleq F^{u1}
\end{aligned} \tag{5.28}$$

where, for a given $|\rho|_{\max}$, $T(|\rho|_{\max},\theta)$ represents a straight line passing through the two points of $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\cosh(2R\theta)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$: at $\theta=0$ and at $\theta=|\rho|_{\max}$ and $|\rho|_{\mathrm{avg}} \triangleq 2^{-L}\sum_{i=1}^{2^L}|\rho_i| = 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\rho_k^+$. The last inequality is obtained from the Cauchy-Schwarz inequality: $|\rho|_{\mathrm{avg}} \le \sigma_\rho$.

Another upper bound of F can be also found as

$$\begin{aligned}
F &= 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\alpha_k e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&\le \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\left(2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\alpha_k\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&= \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\alpha_{\mathrm{avg}}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]
\end{aligned} \tag{5.29}$$

where $\alpha_k \triangleq \cosh(2R\rho_k^+)$ and $\alpha_{\mathrm{avg}} \triangleq 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\alpha_k = 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\cosh(2R\rho_k^+)$. The inequality comes from the concavity of $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1+2\alpha e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ in α. Moreover, since it is an increasing function of α, the last expression of (5.29) can be further upper-bounded by replacing α' with $\alpha' \ge \alpha_{\mathrm{avg}}$. For example, note

$$\alpha_{\mathrm{avg}} \le 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\left(s\rho_k^+ + 1\right) = s\,|\rho|_{\mathrm{avg}} + 1 \le s\sigma_\rho + 1 \triangleq \alpha'$$

where $s = \left(\cosh(2R|\rho|_{\max}) - 1\right)/|\rho|_{\max}$, the slope of a straight line connecting two points of the convex function $\cosh(2R\rho)$, $(0,1)$ and $(|\rho|_{\max}, \cosh(2R|\rho|_{\max}))$. This gives

$$F \leq \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2\big(s\sigma_\rho + 1\big)e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \triangleq F^{u2}. \tag{5.30}$$

By using the convexity of the function, $\mathrm{E}_\tau\left[\log\left(1 + e^{-2R\rho}e^{-2\phi\sqrt{R}\tau - 2R}\right)\right]$, in ρ, the lower bound of F is also found as

$$\begin{aligned} F &= 2^{-L}\sum_{i=1}^{2^L}\mathrm{E}_\tau\left[\log\left\{1 + e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \\ &\geq \mathrm{E}_\tau\left[\log\left\{1 + \exp\left(-2R\left(2^{-L}\sum_{i=1}^{2^L}\rho_i\right)\right)e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \\ &= \mathrm{E}_\tau\left[\log\left\{1 + e^{-2\phi\sqrt{R}\tau - 2R}\right\}\right] \\ &= \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \triangleq F^l. \end{aligned} \tag{5.31}$$

5.5.4 Derivation of the Tightened Bounds

The tightened bounds are derived in a similar way using the convexity or concavity of the function except the cluster identification needs be incorporated. Since $\rho_k = \lambda_n + \mu_i$,

we can rewrite F as

$$
\begin{aligned}
F &= 2^{-M} \sum_{n=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \Big[\log \Big\{ 1 + e^{-2R(\mu_i+\lambda_n)} e^{-2\phi\sqrt{R}\tau - 2R} \Big\} \Big] \right) \\
&= 2^{-M} \sum_{n=1}^{2^M} \Bigg(2^{-(L-M-1)} \\
&\qquad \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2R\mu_l^+\right) e^{-2R\lambda_n} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_n} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \Bigg) \\
&\le 2^{-M} \sum_{n=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \left\{ T_n(|\mu|_{\max}, \theta) \Big|_{\theta=\mu_l^+} \right\} \right) \\
&= 2^{-M} \sum_{n=1}^{2^M} \left\{ T_n\left(|\mu|_{\max}, \theta\right) \Big|_{\theta = 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mu_l^+} \right\} \\
&= 2^{-M} \sum_{n=1}^{2^M} \left\{ T_n(|\mu|_{\max}, \theta) \Big|_{\theta=|\mu|_{\mathrm{avg}}} \right\} \\
&\le 2^{-M} \sum_{n=1}^{2^M} \left\{ T_n(|\mu|_{\max}, \theta) \Big|_{\theta=\sigma_\mu} \right\} \triangleq F_M^{u1}
\end{aligned}
\tag{5.32}
$$

where μ_l^+'s form the positive-half subset of μ_i's and, for a given $|\mu|_{\max}$, $T_n(|\mu|_{\max}, \theta)$ is a straight line passing through the function $\mathrm{E}_\tau \Big[\frac{1}{2} \log \Big\{ 1 + 2\cosh\left(2R\theta\right) e^{-2R\lambda_n} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_n} e^{-4\phi\sqrt{R}\tau - 4R} \Big\} \Big]$ at $\theta = 0$ and $\theta = |\mu|_{\max}$. Moreover, $|\mu|_{\mathrm{avg}} \triangleq 2^{-(L-M)} \sum_{i=1}^{2^{L-M}} |\mu_i| = 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mu_l^+$. The last inequality also follows from $|\mu|_{\mathrm{avg}} \le \sigma_\mu$, and note $\sigma_\mu = \sqrt{\sigma_\rho^2 - \sigma_\lambda^2}$ and $|\mu|_{\max} = |\rho|_{\max} - |\lambda|_{\max}$.

Another form of tightened upper bound of F is obtained as

$$
\begin{aligned}
F &= 2^{-M}\sum_{n=1}^{2^M}\Bigg(2^{-(L-M-1)} \\
&\qquad \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\mu_l^+\right)e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]\Bigg) \\
&= 2^{-M}\sum_{n=1}^{2^M}\Bigg(2^{-(L-M-1)} \\
&\qquad \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\beta_l e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]\Bigg) \\
&\le 2^{-M}\sum_{n=1}^{2^M}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\left(2^{-(L-M-1)}\sum_{l=1}^{2^{L-M-1}}\beta_l\right)e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}\right.\right. \\
&\qquad \left.\left.+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&= 2^{-M}\sum_{n=1}^{2^M}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\beta_{\mathrm{avg}}e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&\le 2^{-M}\sum_{n=1}^{2^M}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\beta' e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&= 2^{-M}\sum_{n=1}^{2^M}\mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1+2\left(s_M\sigma_\mu+1\right)e^{-2R\lambda_n}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_n}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\
&\triangleq F_M^{u2}
\end{aligned}
\tag{5.33}
$$

where $\beta_l \triangleq \cosh\left(2R\mu_l^+\right)$ and

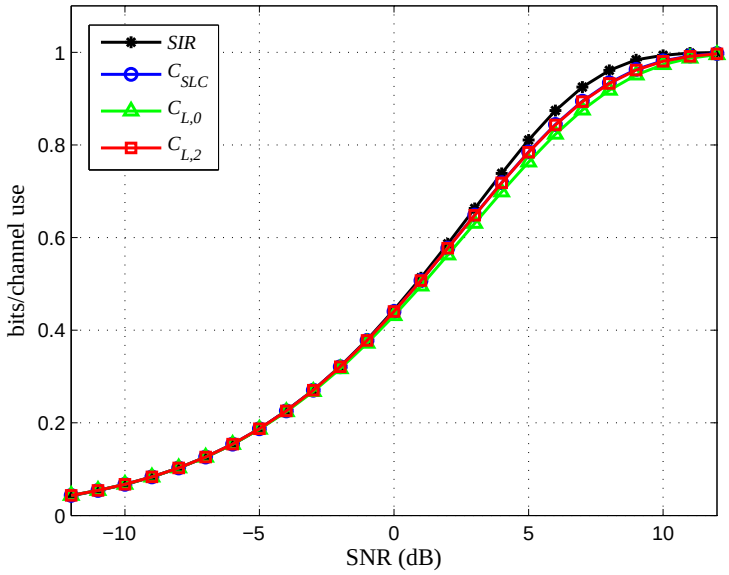
$$
\begin{aligned}
\beta_{\mathrm{avg}} &\triangleq 2^{-(L-M-1)}\sum_{l=1}^{2^{L-M-1}}\beta_l = 2^{-(L-M-1)}\sum_{l=1}^{2^{L-M-1}}\cosh(2R\mu_l^+) \\
&\le 2^{-(L-M-1)}\sum_{k=1}^{2^{L-M-1}}\left(s_M\mu_k^+ + 1\right) = s_M\left|\mu\right|_{\mathrm{avg}} + 1 \\
&\le s_M\sigma_\mu + 1 \triangleq \beta'
\end{aligned}
$$

which is based on a straight line connecting two points of the convex function $\cosh(2R\mu)$, $(0,1)$ and $(|\mu|_{\max}, \cosh(2R|\mu|_{\max})$, having a slope $s_M = \left(\cosh(2R|\mu|_{\max}) - 1\right)/|\mu|_{\max}$.

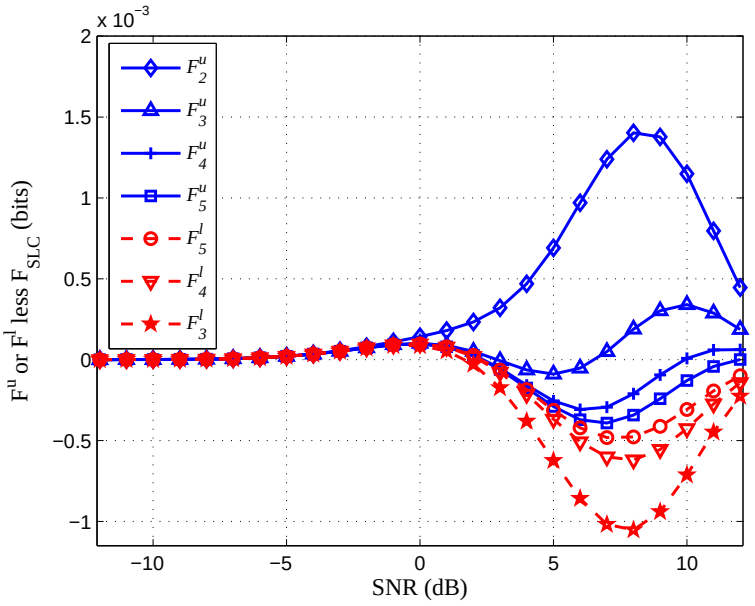
The tightened lower bound of F based on cluster identification is obtained as

$$
\begin{aligned}
F &= 2^{-M} \sum_{n=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \Big[\log \Big\{ 1 + e^{-2R(\mu_i+\lambda_n)} e^{-2\phi\sqrt{R}\tau - 2R} \Big\} \Big] \right) \\
&= 2^{-M} \sum_{n=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \Big[\log \Big\{ 1 + e^{-2R\mu_i} e^{-2R\lambda_n} e^{-2\phi\sqrt{R}\tau - 2R} \Big\} \Big] \right) \\
&\geq 2^{-M} \sum_{n=1}^{2^M} \mathrm{E}_\tau \left[\log \left\{ 1 + \exp \left(-2R \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mu_i \right) \right) e^{-2R\lambda_n} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-M} \sum_{n=1}^{2^M} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R\lambda_n} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-(M-1)} \sum_{k=1}^{2^{M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2R\lambda_k^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \\
&\triangleq F_M^l
\end{aligned}
\tag{5.34}
$$

where λ_k^+'s form the positive-half subset of λ_n's.

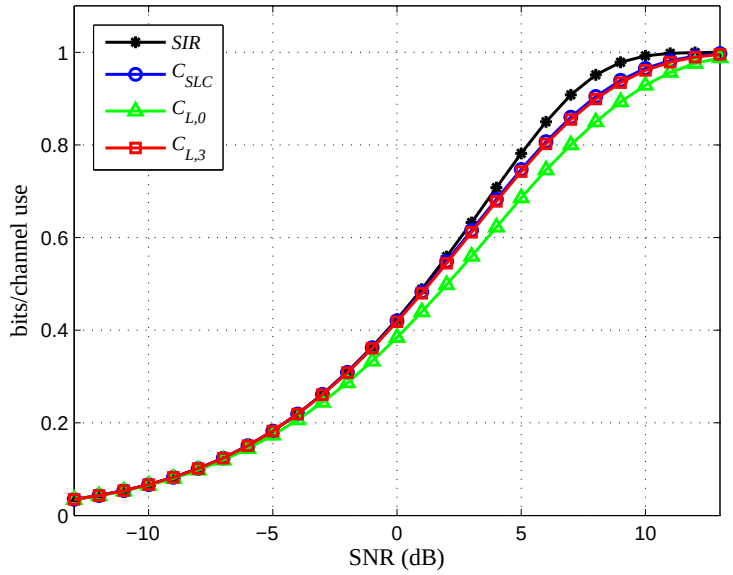


(a)

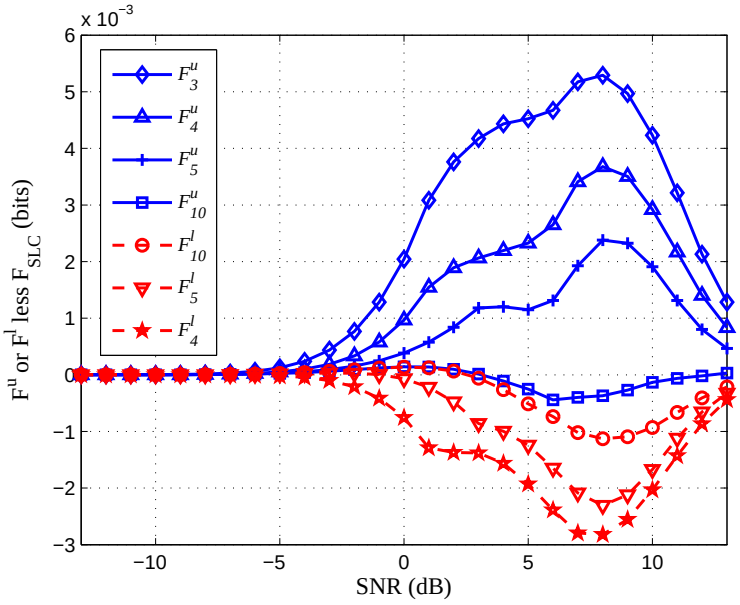


(b)

Figure 5.5: Example channel: $h(D) = 2^{-1/2}(1 + D)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

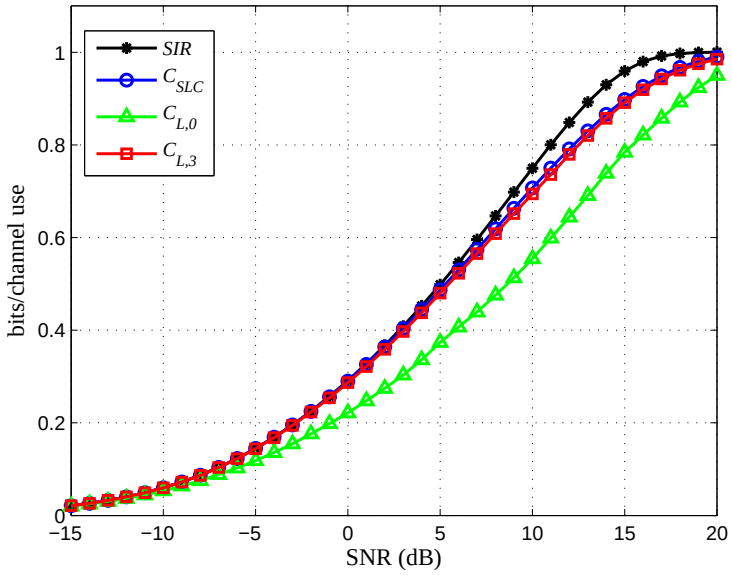


(a)

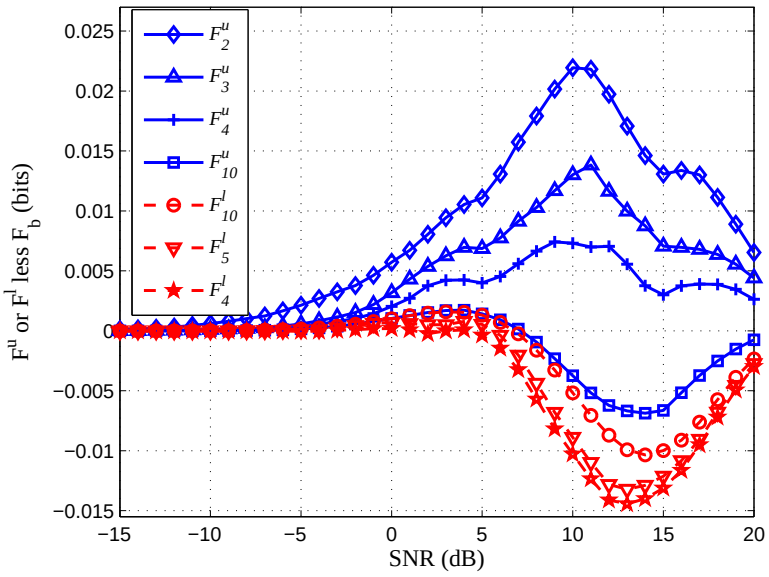


(b)

Figure 5.6: Example channel: $h(D) = 2^{-1}(1 + D - D^2 - D^3)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

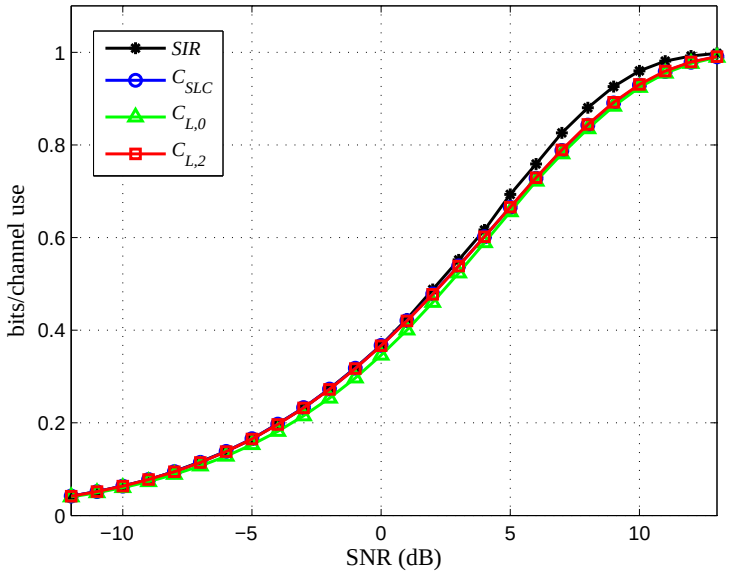


(a)

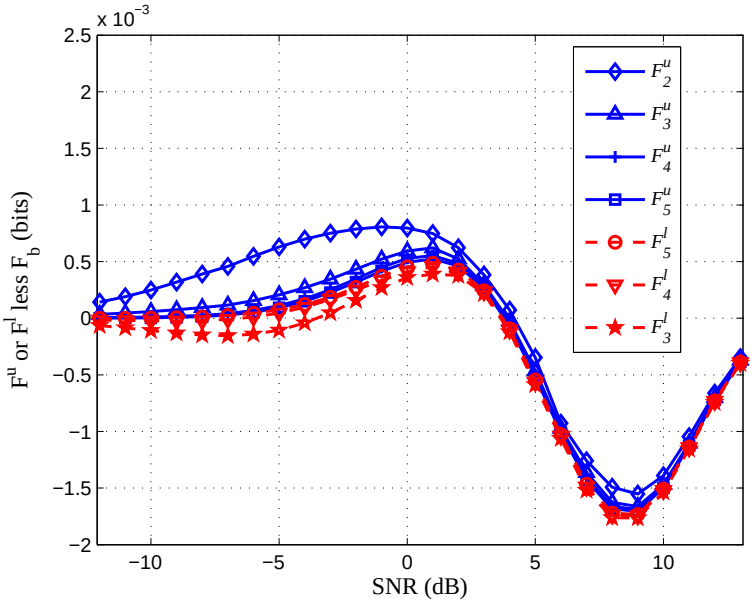


(b)

Figure 5.7: Example channel: $h(D) = 0.19+0.35D+0.46D^2+0.5D^3+0.46D^4+0.35D^5+0.19D^6$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

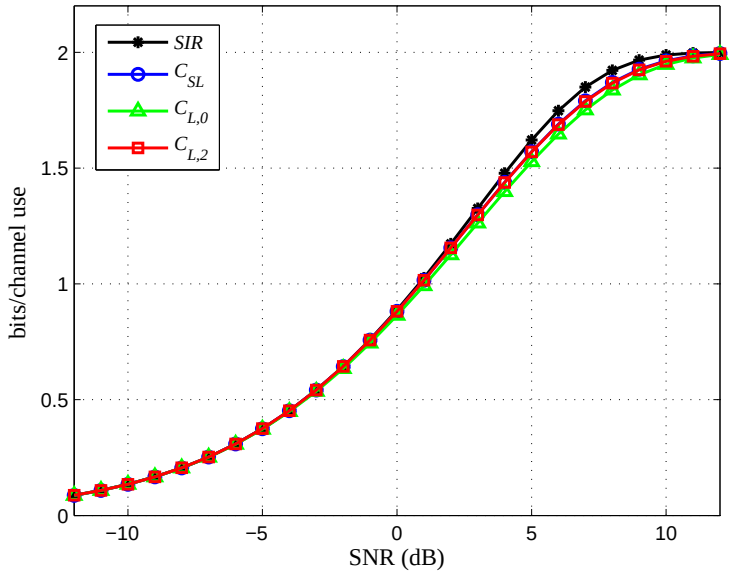


(a)

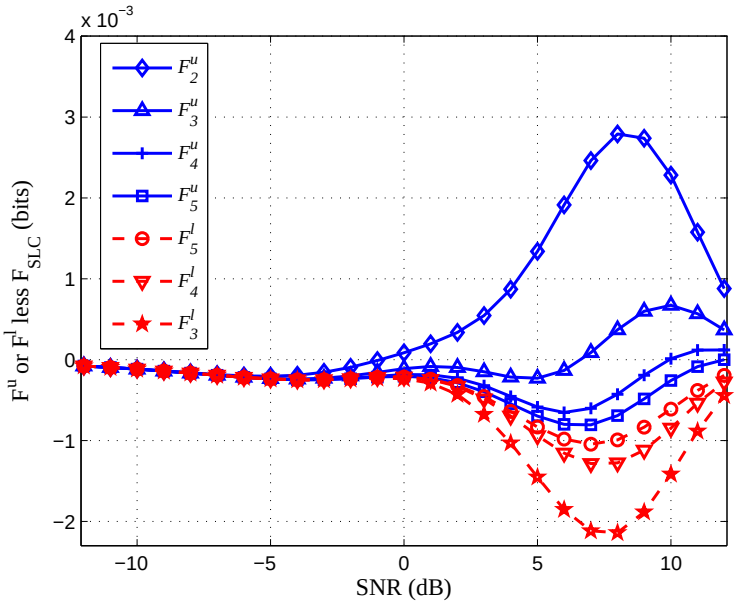


(b)

Figure 5.8: Example channel: $h(D) = 1.6099^{-1/2} \sum_{i=0}^{10} D^i/(1 + (i-5)^2)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.



(a)



(b)

Figure 5.9: Example channel: $h(D) = 2^{-1}\{(1+j)+(1-j)D\}$ with QPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

Chapter 6

Conclusion

6.1 Summary

This thesis has explored low-complexity turbo equalization algorithms as alternatives to the optimal, but much more complex, MAP algorithm across severe ISI channels. Chapter 2 gave a quick review of the SISO equalizer design method established in [14], and presented a new SISO DFE design that relies on extrinsic information formulations and directly accounts for the propagation error effect. With this new approach, both error rate simulation and the EXIT chart analysis indicated that the proposed SISO DFE is superior to the well-known SISO LE. This result contrasts the general understanding today that the error propagation effect of the DFE degrades the overall TE performance to below that of an LE. Finally, various MMSE filters were summarized using the *a priori* information, and the SNR and MI analyses for LE and DFE were provided.

Assuming ideal feedback in both directions, the SNR figure-of-merit for BiDFEs is derived in Chapter 3, allowing infinitely long filler lengths. Additionally, a new extrinsic information combining strategy that investigates statistical noise correlation between the outputs of two opposite direction DFEs was described. When combined

with the previously proposed DFE extrinsic information formulation, this results in a turbo BiDFE with performance approaching that of a BCJR-based TE in a fairly severe ISI environment, significantly outperforming TEs based on SISO LEs and DFEs. This bidirectional turbo DFE achieves excellent performance-complexity tradeoffs compared to the BCJR algorithm-based TE. Remarkably, a time-invariant version of the BiDFE, despite being a lower-complexity method that does not require tap-weight updating as a function of time, nonetheless showed consistently good performance, unlike turbo LEs or DFEs. Lastly, the noise correlation coefficient analysis is presented under the assumption of ideal feedback, and the SNR/MI analyses for LEs, DFEs, and BiDFEs were summarized.

Chapter 4 showed the proper way to extract extrinsic information from other constituent equalizers, when the information between the equalizers could be highly correlated, and then proposed a self-iterating soft equalizer design for uncoded systems. An SISE consists of multiple relatively weak equalizers which are allowed to exchange soft information based on methods designed to suppress significant correlation among their soft outputs. Furthermore, the SISE algorithm is employed in TE systems; the turbo SISE shows robust performance even in severe ISI channels despite the individual weakness of the constituent suboptimal equalizers. The SISE algorithm works well as a stand-alone equalizer or as the equalizer component of a turbo equalization system.

Finally, the analytical and provable lower bounds on the SIR or capacity of the ISI channel corrupted by Gaussian noise are investigated in Chapter 5. Consider the random variable (RV) $Y = X + S + N$, where X is a symbol drawn independently and uniformly from a fixed, finite-size alphabet set symmetrically positioned around the origin, S a zero-mean discrete-valued RV and N a zero-mean Gaussian RV. When S represents the precursor ISI terms after the MMSE-DFE is applied at the channel output, $I(X; X+S+N)$ serves as a tight lower bound for the SIR as well as capacity of the ISI channel corrupted by Gaussian noise. Unfortunately, as the number of symbols forming S grows, finding an analytical expression for the probability density function

of $S + N$ (and thus one for $I(X;Y)$) is not easy. Therefore, the new lower bounds are obtained by first introducing a mismatched mutual information function that can be proved as a lower bound to $I(X;X+S+N)$ and then further lower-bounding this function with expressions that can be computed via a few single-dimensional integrations with a small computational load. The lower bounds on the provable bound itself is developed based on identifying clusters in the distribution of $S + N$. Finding clusters in the $S + N$ distribution is the same as identifying dominant coefficient values from the linear coefficient set that is used to construct S. In all finite-ISI channels examined, the bound defined in this thesis provides the same level of tightness as the SLC to the SIR (while being actually tighter than SLC at high SNRs when viewed closed up) with a very reasonable computation load. In particular, this lower bound was presented on the same channel employed in [30]; the analytical method given here is much better at quickly producing a tight bound than the previous simulation-based method in terms of the complexity/accuracy tradeoff.

6.2 Future work

One potential direction for future research is the examination of low-complexity turbo equalization algorithms for multiple-input multiple-output (MIMO) systems for wireless communications and high track density magnetic recording systems. The turbo equalization schemes in this research were based on the single-input single-output system; thus, these algorithms can naturally be extended to the MIMO systems.

Moreover, this thesis assumes that the noise in an ISI channel model is AWGN; however, in magnetic recording systems the noise is data-dependent. Specifically, medium noise occurs due to the random variations in the geometry of a magnetization transition, so it depends on the data pattern written on the disk. Accordingly, a data dependent (or pattern-dependent) noise prediction/whitening procedure should be incorporated with the detection algorithm, which gives rise to pattern-dependent noise predictive

(PDNP) detectors. It would be beneficial to develop low-complexity turbo equalization algorithms with PDNP processes for ISI channel models with data-dependent noise.

Another direction that warrants investigation is the development of channel estimation algorithms in conjunction with the proposed turbo equalization schemes. Here it was assumed that the impulse responses of ISI channels are perfectly known to the receiver; however, this assumption is impractical, especially for wireless communication systems. Therefore, it would be helpful to research channel estimation algorithms that take advantage of the extrinsic information of other modules, equalizer and decoder, in turbo systems.

One final avenue for future research would be looking for tight and more easily computed analytical forms of SIR or channel capacity, which remains an unsolved problem. In particular, the lower bounds developed here are not very effective on very long ISI channels with no dominant taps, such as the indoor wireless channel in a highly scattered multi-path setting as indicated in Chapter 5. This inquiry could be extended to the various channel models, such as MIMO broadcast channels and 2-dimensional ISI channels.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 and 623-656, 1948.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," *in Proc. IEEE International Conference on Communications (ICC)*, Geneva, Switzerland, May 1993, pp. 1064-1070.

[3] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Communications*, vol. 44, no. 10, pp. 1261-1271 , Oct. 1996.

[4] P. Elias, "Coding for Noisy Channels," *IRE Conv. Rec.*, pt. 4, pp. 37-46, Mar. 1955.

[5] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Information Theory*, vol. 42, no. 2, pp. 429-445, Mar. 1996.

[6] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *European Trans. Telecommunications*, vol. 6, no. 5, pp. 507-511, Sep.-Oct. 1995.

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

[8] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," *in Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dallas, TX, Nov. 1989, pp. 1680-1686.

[9] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Information Theory*, vol. IT-13, pp. 260-269, Apr. 1967.

[10] J. Proakis *Digital Communications*, 4th edition, McGraw-Hill Highter Education, 2000.

[11] S. Haykin *Communication Systems*, 3rd edition, Wiley, 1994.

[12] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Communications*, vol. 47, no. 7, pp. 1046-1061, Jul. 1999.

[13] M. Tüchler, A. Singer, and R. Kötter, "Minimum mean squared error equalization using *a priori* Information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673-683, Mar. 2002.

[14] M. Tüchler, R. Kötter, and A. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 754-767, May 2002.

[15] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: adaptive equalization and channel decoding jointly optimized," *IEEE J. Selected Areas in Communications*, vol. 19, no. 9, pp. 1744-1752, Sep. 2001.

[16] M. Honig, G. Woodward, and Y. Sun, "Adaptive iterative multiuser decision feedback detection," *IEEE Trans. Wireless Communications*, vol. 3, no. 2, pp. 477-485, Mar. 2004.

[17] Y. Sun, V. Tripathi, and M. Honig, "Adaptive turbo reduced-rank equalization for MIMO channels," *IEEE Trans. Wireless Communications*, vol. 4, no. 6, pp. 2789-2800, Nov. 2005.

[18] C. Laot, R. Le Bidan, and D. Leroux, "Low-complexity MMSE turbo equalization: a possible solution for EDGE," *IEEE Trans. Wireless Communications*, vol. 4, no. 3, pp. 965-974, May 2005.

[19] F. R. Rad and J. Moon, "Turbo equalization utilizing soft decision feedback," *IEEE Trans. Magnetics*, vol. 41, no. 10, pp. 2998-3000, Oct. 2005.

[20] J. Moon and F. R. Rad, "Turbo equalization via constrained-delay APP estimation with decision feedback," *IEEE Trans. Communications*, vol. 53, no. 12, pp. 2102-2113, Dec. 2005.

[21] Z. Wu and J. Cioffi, "Low complexity iterative decoding with decision-aided equalization for magnetic recording channels," *IEEE J. Selected Areas in Communications*, vol. 19, no. 4, pp. 699-708, Apr. 2001.

[22] A. Chan and G. Wornell, "A class of block-iterative equalizers for intersymbol interference channels: fixed channel results," *IEEE Trans. Communications*, vol. 49, no. 11, pp. 1966-1976, Nov. 2001.

[23] S. Jeong and J. Moon, "Turbo equalization based on bi-directional DFE," *in Proc. IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[24] S. Jeong and J. Moon, "Soft-in soft-out DFE and bi-directional DFE," to appear in *IEEE Trans. Communications.*

[25] R. Lopes and J. Barry, "The soft-feedback equalization for turbo equalization of highly dispersive channels," *IEEE Trans. Communications*, vol. 54, no. 5, pp. 783-788, May 2006.

[26] D. Arnold and H. -A. Loeliger, "On the information rate of binary-input channels with memory," *in Proc. IEEE International Conference on Communications (ICC)*, Helsinki, Finland, June 2001, pp. 2692-2695.

[27] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," *in Proc. IEEE International Symposium on Information Theory (ISIT)*, Washington, DC, USA, June 2001, pp. 283.

[28] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," *in Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Antonio, TX, Nov. 2001, pp. 2992-2996.

[29] D. Arnold, H. -A. Loeliger, P. Vontobel, A. Kavčić, and W. Zeng "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Information Theory*, vol. 52, no. 8, pp. 3498-3508, Aug. 2006.

[30] P. Sadeghi, P. O. Vontobel, and R. Shams, "Optimization of information rate upper and lower bounds for channels with memory," *IEEE Trans. Information Theory*, vol. 55, no. 2, pp. 663-688, Feb. 2009.

[31] F. Rusek and D. Fertonani, "Lower bounds on the information rate of intersymbol interference channels based on the Ungerboeck observation model," *in Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, pp. 1649-1653, June-July 2009.

[32] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Communications*, vol. 22, pp. 624-636, May 1974.

[33] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1988.

[34] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Information Theory*, vol. 37, no. 6, pp. 1527-1539, Nov. 1991.

[35] S. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. Information Theory*, vol. 42, no. 5, pp. 1388-1404, Sept. 1996.

[36] S. Shamai and S. Verdu, "Worst-case power-constrained noise for binary-input channels," *IEEE Trans. Information Theory*, vol. 38, no. 5, pp. 1494-1511, Sept. 1992.

[37] A. M. Garsia, "Entropy and singularity of infinite convolutions," *Pacific J. Math.*, vol. 13, no. 4, pp. 1159-1169, 1963.

[38] P. H. Wittke, W. S. Smith, and L. L. Campbell, "Infinite series of interference variables with Cantor-type distributions," *IEEE Trans. Information Theory*, vol. 34, no. 6, pp. 1428-1436, Nov. 1988.

[39] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in log domain," *in Proc. IEEE International Conference on Communications (ICC)*, Seattle, WA, June 1995, pp. 1009-1013.

[40] W. Koch and A. Baier, "Optimum and sub-optimum detection of coded data disturbed by time-varying intersymbol interference," *in Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Diego, CA, Dec. 1990, pp. 1679-1684.

[41] J. Erfanian, S. Pasupathy, and G. Gulak, "Reduced complexity symbol detectors with parallel structures for ISI channels," *IEEE Trans. Communications*, vol. 42, no. 2/3/4, pp. 1661-1671, Feb./Mar./Apr. 1994.

[42] J. Vogt and A. Finger, "Improving the max-log-MAP turbo decoder," *IEE Electronic Letters*, vol. 36, no. 23, pp. 1937-1939, Nov. 2000.

[43] L. Papke, P. Robertson, and E. Villerbrun, "Improved decoding with the SOVA in a parallel concatenated (turbo-code) scheme," *in Proc. IEEE International Conference on Communications (ICC)*, Dallas, TX, June 1996, pp. 102-106.

[44] L. Lin and R. Cheng, "Improvements in SOVA-based decoding for turbo codes," *in Proc. IEEE International Conference on Communications (ICC)*, Montreal, Quebec, June 1997. pp. 1473-1478.

[45] M. Fossorier, F. Burkert, S. Lin, and J. Hagenauer, "On the equivalence between SOVA and Max-Log-MAP decodings," *IEEE Communication Letters*, vol. 2, no. 5, pp. 137-139, May 1998.

[46] J. Chen, M. Fossorier, S. Lin, and C. Xu, "Bi-Directional SOVA Decoding for Turbo-Codes," *IEEE Communication Letters*, vol. 4, no. 12, pp.405-407, Dec. 2000.

[47] S. Papaharalabos, P. Sweeney, B. Evans, "Modification of branch metric calculation to improve iterative SOVA decoding of turbo codes," *IEE Electronic Letters*, vol. 39, no. 19, pp. 1391-1392, Nov. 2003.

[48] C. Huang and A. Ghrayeb, "A Simple remedy for the exaggerated extrinsic information produced by the SOVA algorithm," *IEEE Trans. Wireless Communications*, vol. 5, no. 5, pp. 996-1002, May 2006.

[49] M. Austin "Decision-feedback equalization for digital communication over dispersive channels," *M.I.T. RES. Lab. Electron., Tech. Rep.461*, Aug. 1967.

[50] R. Price "Nonlinearity feedback-equalized PAM versus capacity for noisy filter channels," *in Proc. IEEE International Conference on Communications (ICC)*, Philadelphia, PA, 1972.

[51] J. Salz "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, pp.1341-1373, Oct. 1973. pp. 12-16

[52] J. Cioffi, G. Dudevior, M. Eyuboglu, and G. Forney, "MMSE decision-feedback equalizers and coding - Part I and II," *IEEE Trans. Communications*, vol. 43, no. 10, pp. 2582-2604, Oct. 1995.

[53] S. Ariyavisitakul and Y. Li "Joint coding and decision feedback equalization for broadband wireless channels," *IEEE Trans. Selected Areas in Communications*, vol. 16, no. 9, pp.1670 - 1678, Dec. 1998

[54] H. Harashima and H. Miyakawa "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Communications*, vol. 29, no. 4, pp.774-780, Aug. 1972.

[55] A. Duel-Hallen and C. Heegard "Delayed decision feedback sequence estimation," *IEEE Trans. Communications*, vol. 37, no. 5, pp. 428-435, May 1989.

[56] A. Fertner "Improvement of bit-error-rate in decision feedback equalizer by preventing decision-error propagation," *IEEE Trans. Signal Processing*, vol. 46, no. 7, pp.1872-1877, July 1998.

[57] P. Supnithi, R. Lopes, and S. McLaughlin, "Reduced-complexity turbo equalization for high-density magnetic recording systems," *IEEE Trans. Magnetics*, vol. 39, no. 5, pp. 2585-2587, Sept. 2003.

[58] J. Jiang, C. He, E. M. Kurtas, and K. R. Narayanan, "Performance of soft feedback equalization over magnetic recording channels," *in Proc. Intermag*, San Diego, CA, May 2006, pp. 795.

[59] J. Balakrishnan and C. R. Johnson, Jr., "Bidirectional decision feedback equalizer: infinite length results," *in Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2001, pp. 1450-1454.

[60] J. Nelson, A. Singer, U. Madhow, and C. McGahey, "BAD: bidirectional arbitrated decision-feedback equalization," *IEEE Trans. Communications*, vol. 53, no. 2, pp. 214-218, Feb. 2005.

[61] S. Ariyavisitakul, "A decision feedback equalizer with time-reversal structure," *IEEE J. Selected Areas in Communications*, vol. 10, no. 3, pp. 599-613, Apr. 1992.

[62] J. Balakrishnan and C. R. Johnson, Jr., "Time-reversal diversity in decision feedback equalization," *in Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2000.

[63] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Communications*, vol. 49, no. 10, pp. 1727-1737, Oct. 2001.

[64] S. Lee, A. Singer, and N. Shanbhag, "Linear turbo equalization analysis via BER transfer and EXIT charts," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2883-2897, Aug. 2005.

[65] K. Kim, J. Choi, A. Singer, and K. Kim, "A new adaptive turbo equalizer with soft information classification," *in Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, Mar. 2010.

[66] D. G. Messerschmitt, "A geometric theory of intersymbol interference, Part I: Zero-forcing and decision-feedback equalization," *Bell Syst. Tech. J.*, vol. 52, no. 9, pp. 1483-1539, Nov. 1973.

[67] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2384-2415, Oct. 1998.

[68] R. G. Gallager, *Information Theory and Reliable Communication.* New York Wiley, 1968, pp. 97-112, 176-188.

[69] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York Wiley, 1991, pp. 29-31.

[70] S. Jeong and J. Moon, "Easily Computed Lower Bounds on the Information Rate of Intersymbol Interference Channels," to appear in *IEEE Trans. Information Theory.*

[71] S. Jeong and J. Moon, "Self-iterating Soft Equalizer," to appear in *IEEE Global Telecommunications Conference (GLOBECOM) 2011.*

[72] S. Jeong and J. Moon, "Self-iterating Soft Equalizer," to be submitted for *IEEE Trans. Communications.*