UNIVERSITY OF CALIFORNIA, SAN DIEGO

Mitigation Techniques for Severe Narrowband Interference

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

in

Electrical Engineering (Communications Theory and Systems)

by

Arun Batra

Committee in charge:

Professor James R. Zeidler, Chair
Professor Robert R. Bitmead
Professor William S. Hodgkiss
Professor Laurence B. Milstein
Professor John G. Proakis

2009

The dissertation of Arun Batra is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2009

To my parents, Mridul and Anand.

Research is what I'm doing when I don't know what I'm doing

– Wernher Von Braun

Don't ask me why, cause I don't know
Don't ask me how, I'm gonna solve this on my own

– Tim McIlrath of *Rise Against*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. James R. Zeidler, for his guidance, patience, and encouragement during the research phase of my graduate studies. Without Dr. Zeidler, this dissertation would not be nearly as complete as it is today. I would also like to thank my committee members: Dr. Laurence B. Milstein, Dr. John G. Proakis, Dr. Robert R. Bitmead, and Dr. William S. Hodgkiss, for taking the time out of their busy schedules and for providing me with additional guidance, which greatly improved the dissertation. I would be remiss in not thanking Dr. Louis Beex from Virginia Tech for his help during the initial stages of my research and for the numerous helpful technical conversations. I would also like to thank my graduate advisor at Cornell, Dr. Toby Berger, as well as Dr. C. Richard Johnson and Dr. Nirmal Keshava for their encouragement and helping me to pursue my education.

I must also thank my fellow graduate students in the Electrical and Computer Engineering Department for their camaraderie during many barbecues, happy hours, and coffee hours. The students in Dr. Zeidler's research group deserve special praise for their constant encouragement and technical expertise, which often times far exceeded my own. I would also like to thank the countless graduate students who I participated with on numerous intramural sports teams, too many to name in this short acknowledgment; participating on those teams was one of the highlights of my graduate school experience.

A heartfelt thank you goes to my roommates over the years: Alvin AuYoung, Allen Chu, Danielle Gaeta, Clement Kam, Marc Krull, Ed Liao, Eric Lin, Aron Lum, Brendan Morris, Caitlin White, and Kia Yang. I consider you guys a part of my family and there is no way I could have done this without you. Thanks for all the food, carpools, movies, road trips, Sundays, gym workouts, and most importantly, the memories. I would also like to thank my friends from my high school and undergraduate days, for sticking with me through this entire process. A special thanks has to go to my best friend, Rohit Malhotra. He has always been by my side and the last seven years were no exception. Finally, thanks to all those who I crossed paths with during my time in San Diego for making it an interesting adventure.

Lastly, and most importantly, I would not have achieved this goal without my family. It was my parents who instilled my desire for knowledge, and have provided me with never ending love and support. To my brother, Anuj, whose path I have closely followed, thank you so much for lending technical support and advice during my entire

tenure at UCSD. His confidence in me and my abilities never wavered, even when I doubted whether I was up to this monumental task. Thank you Bhaiya. Let me end by saying, honestly what choice did I have but to get my Ph.D., it runs in the family.

Chapter 2, in part, is a reprint of material as it appears in A. Batra, T. Ikuma, J. R. Zeidler, A. A. Beex, and J. G. Proakis, "Mitigation of Unknown Narrowband Interference Using Instantaneous Error Updates," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, pp. 115–119, Nov. 2004, A. Batra, J. R. Zeidler, and A. A. Beex, "Mitigation of Narrowband Interference Using Adaptive Equalizers," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sep. 2006, and A. Batra, J. R. Zeidler, and A. A. Beex, "A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Decision-Feedback Adaptive Equalizers in the Presence of Severe Narrowband Interference," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 390102, 13 pages, 2008. doi:10.1155/2008/390102. The dissertation author was the primary investigator and author of these papers.

Chapter 3, in part, is a reprint of material as it appears in A. Batra, T. Ikuma, J. R. Zeidler, A. A. Beex, and J. G. Proakis, "Mitigation of Unknown Narrowband Interference Using Instantaneous Error Updates," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, pp. 115–119, Nov. 2004, A. Batra, J. R. Zeidler, and A. A. Beex, "Initialization Techniques for Improved Convergence of LMS DFEs in Strong Interference Environments," in *Proceedings of the IEEE Global Communications (Globecom) Conference*, Washington, DC, pp. 3068–3073, Nov. 2007, and is currently being prepared for submission for publication of the material. A. Batra, J. R. Zeidler, and A. A. Beex, "Implementation Methods for Data-Aided Initialization," in preparation, 2009. The dissertation author was the primary investigator and author of these papers.

Chapter 4, in part, is a reprint of material as it appears in A. Batra, J. R. Zeidler, and A. A. Beex, "A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Decision-Feedback Adaptive Equalizers in the Presence of Severe Narrowband Interference," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 390102, 13 pages, 2008. doi:10.1155/2008/390102. The dissertation author was the primary investigator of this paper.

Chapter 5, in part, is a reprint of material as it appears in A. Batra, J. R. Zeidler,

J. G. Proakis, and L. B. Milstein, "Interference Rejection and Management," in *New Directions in Wireless Communications Research*, V. Tarokh, Ed. New York: Springer, 2009. The dissertation author was the primary investigator and author of this paper.

Chapter 6, in part, is a reprint of material as it appears in A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems," in *Proceedings of the Military Communications (MILCOM) Conference*, San Diego, CA, Nov. 2008, A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in BICM OFDM systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, pp. 2605-2608, Apr. 2009, A. Batra, J. R. Zeidler, J. G. Proakis, and L. B. Milstein, "Interference Rejection and Management," in *New Directions in Wireless Communications Research*, V. Tarokh, Ed. New York: Springer, 2009 and is currently being prepared for submission for publication of the material. A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems using the Prediction-Error Filter," in preparation, 2009. The dissertation author was the primary investigator and author of these papers.

VITA

2001	Bachelor of Science in Electrical and Computer Engineering Cornell University Ithaca, New York
Summer 2001	Research Assistant MIT Lincoln Laboratory Lexington, Massachussets
2002	Master of Engineering in Electrical and Computer Engineering Cornell University Ithaca, New York
Summer 2003	Systems Engineering Intern Northrop Grumman San Diego, California
Fall 2003	Graduate Teaching Assistant Department of Electrical and Computer Engineering University of California, San Diego La Jolla, California
2003-2009	Graduate Research Assistant Center for Wireless Communications University of California, San Diego La Jolla, California
Summer 2004	Graduate Researcher Space and Naval Warfare Systems Center San Diego, California
2009	Doctor of Philosophy in Electrical and Computer Engineering University of California, San Diego La Jolla, California

PUBLICATIONS

A. Batra, T. Ikuma, J. R. Zeidler, A. A. Beex, and J. G. Proakis, "Mitigation of Unknown Narrowband Interference Using Instantaneous Error Updates," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, pp. 115–119, Nov. 2004.

A. Batra, J. R. Zeidler, and A. A. Beex, "Mitigation of Narrowband Interference Using Adaptive Equalizers," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sep. 2006.

A. Batra, J. R. Zeidler, and A. A. Beex, "Initialization Techniques for Improved Convergence of LMS DFEs in Strong Interference Environments," in *Proceedings of the IEEE Global Communications (Globecom) Conference*, Washington, DC, pp. 3068–3073, Nov. 2007.

A. Batra, J. R. Zeidler, and A. A. Beex, "A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Decision-Feedback Adaptive Equalizers in the Presence of Severe Narrowband Interference," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 390102, 13 pages, 2008. doi:10.1155/2008/390102.

A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems," in *Proceedings of the Military Communications (MILCOM) Conference*, San Diego, CA, Nov. 2008.

A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in BICM OFDM systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, pp. 2605-2608, Apr. 2009.

A. Batra, J. R. Zeidler, J. G. Proakis, and L. B. Milstein, "Interference Rejection and Management," in *New Directions in Wireless Communications Research*, V. Tarokh, Ed. New York: Springer, 2009.

A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems using the Prediction-Error Filter," in preparation, 2009.

A. Batra, J. R. Zeidler, and A. A. Beex, "Implementation Methods for Data-Aided Initialization," in preparation, 2009.

ABSTRACT OF THE DISSERTATION

Mitigation Techniques for Severe Narrowband Interference

by

Arun Batra

Doctor of Philosophy in Electrical Engineering (Communications Theory and Systems)

University of California, San Diego, 2009

Professor James R. Zeidler, Chair

This dissertation examines the effect of severe narrowband interference on wireless communication systems. In single-carrier systems, the interference causes the adaptive equalizer to have an extended convergence time, where convergence is considered in terms of the bit error rate (BER). Two techniques are proposed to improve the convergence. The first method, data-aided initialization (DAI), initializes the Wiener weights from estimates derived directly from the received data and training sequences. This technique is shown to substantially reduce the number of training symbols needed for convergence. Further, two methods for obtaining the DAI weights are investigated. The use of multistage Wiener filters (MSWF) is preferable to a parametric approach to direct matrix inversion in terms of BER performance and number of training symbols needed. The second method is a two-stage system that utilizes a prediction error filter (PEF) as a pre-filter to the equalizer. It is shown that the two-stage system reduces the number of training symbols required to reach a BER of 10^{-2} by approximately two orders of magnitude without substantially degrading the steady-state BER performance as compared to the DFE-only case.

In block-modulated multi-carrier systems the presence of a severe narrowband in-

terference causes the degradation of a large number of subcarriers due to spectral leakage of the interference power after demodulation. Multi-carrier code division multiple access (MC-CDMA) obtains frequency diversity by spreading the data into every subcarrier, thus mitigating the effects of narrowband interference. On the other hand, orthogonal frequency division multiplexing (OFDM) requires the addition of coding and interleaving to obtain frequency diversity. The use of genie inserted erasures provides little to no improvement in BER performance, thus the PEF is proposed as an erasure insertion mechanism that notches out the tones located close to the interference, while leaving the remaining tones unaffected. This technique provides excellent results as compared to the case of no interference.

This work was done at UCSD's Center for Wireless Communication, under the "Bandwidth Efficient Communications" project (CoRe research grant 06-10216) and supported by the Office of Naval Research, Code 313.

1 Introduction

With the proliferation of data-enabled digital mobile phones and technology, there is a large demand for high data rate wireless communications, that includes voice, web, and video content. To accomplish this task, communication systems require large amounts of bandwidth. Unfortunately, the amount of spectrum is finite and may be shared among a number of communication systems. One example of this scenario occurs in the unlicensed bands where there is little restriction on who may operate. This leads to a competition for spectrum and the possibility that these systems may interfere with each other. Similarly, in licensed bands, a new direction in research aims to utilize unused portions of the spectrum using cognitive radios [57]. In this case, a primary user may access any part of the spectrum, while secondary users must modify their spectrum so as not to interfere with the primary user. Interference may also arise when other communications systems radiate in a primary user's frequency band in order to disrupt communications. This is referred to as intentional jamming, and can be seen primarily in military applications. Finally, interference can also occur because of nonlinearities in the mixer.

To maintain reliable communications it is necessary to quickly mitigate the interference, especially when the information is being transferred in small packets or short bursts. This dissertation examines innovative receivers that can quickly mitigate narrowband interference when it is a hundred times as strong as the signal of interest (signal-to-interference ratio (SIR) of -20 dB). The effect of this type of interference is examined in the case of single-carrier systems, as well as multi-carrier systems. Both cases will use similar techniques to mitigate the interference, however, the goals of the two scenarios are different.

For single-carrier systems, the use of linear and nonlinear equalizers are examined in suppressing the interference, however the convergence properties of these systems are

Figure 1.1: Single-carrier complex baseband system model.

limited in the presence of severe narrowband interference. This issue of convergence of adaptive equalizers is investigated in terms of bit error rate (BER), as opposed to the commonly used metric of mean-squared error (MSE). Also of importance is the resulting steady-state BER. The first approach to solve this problem utilizes data-based averages to initialize the equalizers. The second approach employs a two-stage approach using the prediction-error filter (PEF) as pre-filter to the decision-feedback equalizer (DFE). In the case of multi-carrier systems, the overall BER is considered for two block modulated systems. Multi-carrier code division multiplexing (MC-CDMA) is considered along with orthogonal frequency division multiplexing (OFDM). MC-CDMA attains frequency diversity through spreading of each data symbols into all subcarriers. OFDM requires a signal processing technique, such as the PEF to mitigate the narrowband interference. The PEF acts an erasure insertion mechanism in the coded scenario. An analytical upper bound for the coded OFDM BER is provided for this scenario as well.

1.1 System Model

A complex baseband representation of a single-carrier communication system is depicted in Figure 1.1. The signal of interest, d_l, is composed of i.i.d. bits. This discrete data is mapped into symbols from an arbitrary constellation with average power equal to E_s. These symbols are passed through a pulse-shaping filter (possibly oversampled by a factor of L_o) that is necessary for bandlimited transmission (see Section 1.2 for further discussion). The discrete-time symbols are converted into continuous-time using a digital-to-analog converter (DAC), given by

$$x(t) = \sum_m d_m g(t - mT_s), \tag{1.1}$$

where T_s is the symbol duration and $g(t)$ is the transmit pulse-shaping filter. This signal is transmitted through the channel and the received signal is given by the convolution

of $x(t)$ and the channel, $h(\tau, t)$, as well as corrupted by additive noise,

$$\begin{aligned} y(t) &= h(\tau; t) * x(t) + n(t) \\ &= \int_{-\infty}^{\infty} h(\tau; t) x(t - \tau)\, \mathrm{d}\tau + n(t), \end{aligned} \tag{1.2}$$

where $*$ represents linear convolution. The receiver converts the continuous-time signal into discrete samples using an analog-to-digital converter (ADC). A matched filter is utilized to maximize the SNR, noting that the frequency response of the pulse shape and matched filter is assumed to satisfy Nyquist's criterion for no ISI (see Section 1.2).

The noise is a circular symmetric Gaussian random process. Each sample of which is modeled as a zero-mean, complex additive white Gaussian (AWGN) random variable with variance, given by σ_n^2.

1.1.1 Wireless Channel

The channel can prove to be a difficult problem when it causes multipath propagation. This occurs when transmitted signals traverse multiple paths caused by a rich scattering environment. If the objects within the medium are moving, it is possible that the multipath channel changes with time as well. The channel can be represented by its impulse response, $h(\tau; t)$, which is the response of the channel at time t to a unit pulse transmitted at time $t - \tau$. If the channel's maximum delay spread, T_m is greater than the symbol duration, T_s, the channel will induce intersymbol interference (ISI) [97].

The channel can be further modeled statistically in a discrete manner as

$$h(\tau; t) = \sum_{m=0}^{L_h} \alpha_m(t) e^{-j2\pi f_c \tau_n(t)} \delta\left(\tau - \tau_m(t)\right), \tag{1.3}$$

where $\delta(\cdot)$ is the Dirac delta function, $\alpha_m(t)$ is the attenuation factor for the received signal on the m^{th} path, $\tau_m(t)$ is the propagation delay for the m^{th} path, and f_c is the carrier frequency of the system. The number of resolvable multipath components is denoted by $L_h = \left\lfloor \frac{T_m}{T_s} \right\rfloor + 1$, where $\lfloor \cdot \rfloor$ is the function that returns the integer portion of argument. When there are a large number of paths, the channel response defined by (1.3) may be modeled as a complex-valued Gaussian random process using the central limit theorem [71]. This randomness leads to fading of the signal caused by moving scatterers in the environment. This result arises from the phase of the channel changing very rapidly, and causing the received signal to vary rapidly. At times, the multiple paths will add constructively, causing the received signal power to increase. However,

when the paths sum destructively, the received signal power can decrease substantially, making detection of the data symbols virtually impossible.

When $h(\tau;t)$ is modeled as a zero-mean complex Gaussian random variable, the envelope $|h(\tau;t)| = \alpha_m(t)$ is distributed as a Rayleigh random variable at each t value, while the phase $e^{-j2\pi f_c t}$ is distributed as a uniform random variable. The envelope is characterized by the following probability density function [71],

$$p_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}},\ r \geq 0. \tag{1.4}$$

If there happens to be a line-of-sight (LOS) component or if fixed scatterers exist in the environment in addition to the random scatterers, the channel can no longer be modeled as zero-mean. The envelope of this channel has a Rice distribution, characterized by the following probability distribution function,

$$p_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+s^2}{2\sigma^2}} I_0\left(\frac{rs}{\sigma^2}\right),\ r \geq 0, \tag{1.5}$$

where I_0 is the modified Bessel function of the first kind with order zero. Note that when $s = 0$, (1.5) degrades into (1.4).

These channel impairments can be quite severe, causing the need for time-domain equalizers. Ideally, a maximum-likelihood (ML) receiver is employed, however, this is generally impractical due to the complexity of implementation which grows exponentially with the number of symbols interfering with the symbol of interest. Many suboptimal equalizers have been examined with regard to mitigating ISI [100, and references therein], and some of these will be discussed in Chapter 2. In the case of severe fading, diversity techniques over time, frequency, and space are employed. A review of the research in this area can be found in [115, and references therein].

1.2 Band-Limited Transmission

In practical systems, the channel is band-limited such that transmission of data outside a certain frequency band is not allowed, i.e. $H(f) = 0, |f| > W$, where W is the channel bandwidth of the system. For successful transmission, the data must be shaped such that its frequency content lies within the bandwidth of the channel. This is accomplished using the pulse-shaping filter, $g(t)$, as seen in (1.1), and defined as

$$g(t) = \int_{-W}^{W} G(f) e^{j2\pi f t}\, \mathrm{d}t, \tag{1.6}$$

where $G(f)$ is the frequency response of $g(t)$. The goal is to find $g(t)$ such that there is no distortion of the transmission (aside from additive noise). Without loss of generality, it is assumed that the channel has ideal frequency response, $H(f) = 1, |f| \leq W$.

The received observations, given in (1.2), can be rewritten as

$$y(t) = \sum_{m=0}^{\infty} d_m g(t - mT_s) + n(t). \tag{1.7}$$

Sampling the the received signal, (1.7), at times $t = lT_s + \tau_0, l = 0, 1, \ldots,$ gives

$$\begin{aligned} y_l = y(lT_s + \tau_0) &= \sum_{m=0}^{\infty} d_m g((l-m)T_s + \tau_0) + n(lT_s + \tau_0) \\ &= \sum_{m=0}^{\infty} d_m g_{l-m} + n_l, \end{aligned} \tag{1.8}$$

where τ_0 is the delay associated with transmission through the channel. Extracting the symbol of interest, d_l, from (1.8), gives the sample values as

$$\begin{aligned} y_l &= g_0 d_l + \sum_{\substack{m=0 \\ m \neq l}}^{\infty} d_m g_{l-m} + n_l \\ &= g_0 \left(d_l + \frac{1}{g_0} \sum_{\substack{m=0 \\ m \neq l}}^{\infty} d_m g_{l-m} \right) + n_l. \end{aligned} \tag{1.9}$$

The value g_0 can be regarded as an arbitrary scalar value that is set to unity, giving the first term of (1.9) as the desired symbol, the second term representing the ISI, and the third term is the additive noise. The condition then for no ISI is

$$g_l = \begin{cases} 1, & l = 0, \\ 0, & l \neq 0. \end{cases} \tag{1.10}$$

This condition is known as the Nyquist pulse-shaping criterion. The theorem implies that

$$\sum_{m=-\infty}^{\infty} G\left(f + \frac{m}{T_s}\right) = T_s \tag{1.11}$$

must be satisfied such that (1.10) is true. The proof of this theorem can be found in [97, pp. 557-558].

There are three cases for choosing the sampling rate, T_s, given the bandwidth of the channel, W. When $T_s < \frac{1}{2W}$, (1.11) can not be satisfied, and ISI is induced. When

$T_s = \frac{1}{2W}$, or equivalently $\frac{1}{T_s} = 2W$ (this is known as the Nyquist sampling rate), there is one choice that satisfies (1.11), namely

$$G(f) = \begin{cases} T_s, & |f| < W, \\ 0, & \text{otherwise.} \end{cases} \tag{1.12}$$

Taking the inverse Fourier transform of (1.12) gives the impulse response for the pulse-shaping filter to be

$$g(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s} = \text{sinc}\left(\frac{\pi t}{T_s}\right). \tag{1.13}$$

For this case, the choice of a sinc function as the pulse-shaping impulse response provides an ideal rectangular filter in the frequency-domain. To obtain this filter, it is necessary that $g(t)$ be noncausal, thus making this choice of filter unrealizable requiring the use of a delayed version. Another drawback of this choice of filter is that the sinc function decays as $1/t$. If there is a small error in the sampling time, this will cause an unbounded level of ISI components.

Finally, choosing $T_s > \frac{1}{2W}$ provides a scenario where there is no unique choice of filter. One choice of pulse-shaping filter that has been used in practical situations is the raised cosine pulse. The frequency response of the raised cosine filter is given as

$$G_{\text{rc}}(f) = \begin{cases} T_s, & 0 \le |f| \le \frac{1-\beta}{2T_s}, \\ \frac{T_s}{2}\left\{1 + \cos\left[\frac{\pi T_s}{\beta}\left(|f| - \frac{1-\beta}{2T_s}\right)\right]\right\}, & \frac{1-\beta}{2T_s} \le |f| \le \frac{1+\beta}{2T_s}, \\ 0, & |f| > \frac{1+\beta}{2T_s}, \end{cases} \tag{1.14}$$

where $\beta \in [0, 1]$ is the roll-off factor. When $\beta > 0$, $G_{\text{rc}}(f)$ possesses frequencies that are greater than the Nyquist frequency, $\frac{1}{2T_s}$, and called excess bandwidth. When $\beta = 1$, $G_{\text{rc}}(f)$ has twice the bandwidth of the rectangular filter defined in (1.12), and is termed 100% excess bandwidth. The impulse response, $g_{\text{rc}}(t)$, is defined as

$$g_{\text{rc}}(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s}\frac{\cos(\pi\beta t T_s)}{1 - 4\beta^2 t^2 T_s^2}. \tag{1.15}$$

From (1.15), it can be seen that when $\beta = 0$, $g_{\text{rc}}(t) = \text{sinc}\left(\frac{\pi t}{T_s}\right)$, which is simply the case of $T_s = \frac{1}{2W}$. A benefit of using the raised cosine pulse is that its tail decays as $1/t^3$, which reduces the number of ISI components associated with an error in the sampling time.

In practical implementations, the pulse shaping is usually divided between the transmit and receive filters, with the goal of ensuring that the overall frequency response

approximates the response of the raised cosine pulse. Let $g_T(t)$ and $g_R(t)$ be the impulse responses of the transmitter pulse shape and the receiver pulse shape, respectively. Again assuming that the channel is ideal, let the received observations be defined as

$$\begin{aligned} y(t) &= g_T(t) * g_R(t) + n(t) * g_R(t) \\ &= \int_{-\infty}^{\infty} g_T(\tau) g_R(t-\tau)\,\mathrm{d}\tau + \int_{-\infty}^{\infty} n(\tau) g_R(t-\tau)\,\mathrm{d}\tau. \end{aligned} \tag{1.16}$$

For a given $g_T(t)$, $g_R(t)$ is chosen in a manner to maximize the signal-to-noise ratio (SNR) at the sampling time, $t = T_s$, and is called the matched filter. Sampling (1.16), gives

$$\begin{aligned} y(T_s) &= \int_0^{T_s} g_T(\tau) g_R(T_s-\tau)\,\mathrm{d}\tau + \int_0^{T_s} n(\tau) g_R(T_s-\tau)\,\mathrm{d}\tau \\ &= \int_0^{T_s} g_T(\tau) g_R(T_s-\tau)\,\mathrm{d}\tau + y_n(T_s). \end{aligned} \tag{1.17}$$

The SNR is then defined as

$$\mathrm{SNR} = \frac{\left[\int_0^{T_s} g_T(\tau) g_R(T_s-\tau)\,\mathrm{d}\tau\right]^2}{\mathbb{E}\left[y_n^2(T_s)\right]}. \tag{1.18}$$

The denominator of (1.18) is simply the variance of the noise at the output of the receive filter, and calculated as

$$\begin{aligned} \mathbb{E}\left[y_n^2(T_s)\right] &= \int_0^{T_s}\int_0^{T_s} \mathbb{E}\left[n(\tau)n(t)\right] g_R(T_s-\tau) g_R(T_s-t)\,\mathrm{d}\tau\,\mathrm{d}t \\ &= \int_0^{T_s}\int_0^{T_s} \sigma_n^2 \delta(t-\tau) g_R(T_s-\tau) g_R(T_s-t)\,\mathrm{d}\tau\,\mathrm{d}t \\ &= \sigma_n^2 \int_0^{T_s} g_R^2(T_s-\tau)\,\mathrm{d}\tau. \end{aligned} \tag{1.19}$$

Substituting (1.19) in (1.18), gives the SNR as

$$\mathrm{SNR} = \frac{\left[\int_0^{T_s} g_T(\tau) g_R(T_s-\tau)\,\mathrm{d}\tau\right]^2}{\sigma_n^2 \int_0^{T_s} g_R^2(T_s-\tau)\,\mathrm{d}\tau}. \tag{1.20}$$

To maximize (1.20), the numerator must be maximized while the denominator is held constant. This maximization can be performed using the Cauchy-Schwarz inequality, which is defined as

$$\left[\int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x\right]^2 \le \int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x \int_{-\infty}^{\infty} g^2(x)\,\mathrm{d}x. \tag{1.21}$$

Note that equality is obtained when $g(x) = af(x)$ for any arbitrary constant a. Using the Cauchy-Schwarz inequality, an upper bound of the SNR is defined as

$$\begin{aligned}\text{SNR} &= \frac{\left[\int_0^{T_s} g_T(\tau) g_R(T_s - \tau)\,\mathrm{d}\tau\right]^2}{\sigma_n^2 \int_0^{T_s} g_R^2(T_s - \tau)\,\mathrm{d}\tau} \\ &\leq \frac{\int_0^{T_s} g_T^2(\tau)\mathrm{d}\tau \int_0^{T_s} g_R^2(T_s - \tau)\,\mathrm{d}\tau}{\sigma_n^2 \int_0^{T_s} g_R^2(T_s - \tau)\,\mathrm{d}\tau} \\ &= \frac{1}{\sigma_n^2} \int_0^{T_s} g_T^2(\tau)\,\mathrm{d}\tau. \end{aligned} \tag{1.22}$$

The SNR is maximized when $g_R(t) = a g_T(T - t)$, i.e. $g_R(t)$ is matched to $g_T(t)$. Looking at the overall frequency response of the transmit and receive filters and noting that the filters are matched, $G(f) = G_T(f)G_R(f) = |G_T(f)|^2$. Recall the desire for $G(f)$ to have the same characteristics as the raised cosine filter, implies $G_T(f) = G_R(f) = \sqrt{|G_{\text{rc}}(f)|}$. The impulse response of these filters is the root raised cosine pulse, $g_{\text{rrc}}(t)$ defined [123, p.165] as

$$g_{\text{rrc}}(t) = \frac{4\beta}{\pi\sqrt{T_s}} \frac{\cos\left[(1+\beta)\pi t/T_s\right] + \sin\left[(1-\beta)\pi t/T_s\right] / \left(4\beta t/T_s\right)}{1 - \left(4\beta t/T_s\right)^2}. \tag{1.23}$$

It is assumed that the root raised cosine is unit norm and $g_{\text{rrc}}(0)$ is normalized to unity.

To further improve this system, the data can be oversampled prior to transmission. The oversampled version is obtained by placing $L_o - 1$ zeros between each data symbol, where L_o is the oversampling rate. This data stream is then filtered by the root raised cosine pulse, which interpolates the zero values and provides a waveform for transmission. The oversampling increases the space between replicas in the frequency-domain, allowing the anti-aliasing filter to have a larger transition band at the receiver, at the cost of needing faster hardware to operate at the faster sampling rate. This facilitates the design of the reconstruction filter at the receiver.

1.3 Narrowband Interference

The narrowband interference is an additive term of the received signal, making (1.2) now

$$y(t) = \int_{-\infty}^{\infty} h(\tau; t) x(t - \tau)\,\mathrm{d}\tau + i(t) + n(t). \tag{1.24}$$

The interference, $i(t)$, is modeled as a pure complex exponential,

$$i(t) = \sqrt{E_i} e^{j(\Omega_i t + \theta)}, \tag{1.25}$$

where E_i is interferer power, Ω_i is the angular frequency of the interferer measured in radians, and θ is a random phase that is uniformly distributed between 0 and 2π. Note that the angular frequency can be written as $\Omega_i = 2\pi f_i$, where f_i is the frequency of the interference measured in Hertz. The discrete-time interference samples are obtained after the ADC by sampling (1.25) at time instances, $t = lT_s$,

$$i_l = i(t)|_{t=lT_s} = \sqrt{E_i} e^{j(\Omega_i l T_s + \theta)}. \tag{1.26}$$

When narrowband interference is present in the received signal, the techniques mentioned previously for receive filtering are not appropriate. Statistics of the interference are required to design a filter that minimizes the distortion caused by the interference. In the case that this type of knowledge is not available at the receiver, the receive filters must be implemented by an adaptive filter that will take on different forms in this dissertation.

Adaptive filtering for interference mitigation can be separated into four broad groups, transform-domain filtering, noise cancellation, linear prediction, and equalization. Transform-domain techniques aim to mitigate interference in the frequency-domain, following the work done in spectral estimation [20, 82]. Adaptive implementation was examined using surface acoustic waves (SAW) filters to track frequency variations of the interference in the frequency-domain [114]. Further adaptive transform-domain filtering was investigated for use in spread spectrum (SS) systems in [108].

An adaptive noise canceler (ANC) is another method utilized to suppress interference [47,133] when an auxiliary reference containing the interference term is available. This reference signal is filtered in a manner suitable for interference removal from the primary input. More recently, this technique has been examined for use with gradient-based algorithms and higher order statistics (HOS) [113], however, this type of interference information is considered unavailable in this work.

The next technique of interest is linear prediction (LP) [81, 140] and its variant the prediction-error filter (PEF). This structure was first used in speech coding and for spectral estimation [53] to estimate the frequency content of narrowband signals (as opposed to using Fourier techniques [20]). The use of the PEF as a means for suppressing narrowband interference in spread spectrum systems using either the Wiener algorithm [134] or the maximum entropy algorithm [21, 22] was proposed in [62]. [124] further examined the properties of the PEF when rejecting NBI, specifically demonstrating that the performance is dependent on the interference-to-noise ratio (INR) and the

filter aperture. The adaptive nature of the structure was examined in terms of spectral dynamics in [110, 111], while [102] compared the convergence aspects of the linear predictor and the smoothing filter.

Equalization as a means for interference mitigation was introduced in [73–75]. In [73, 75] a comparison of the PEF and a two-sided transversal filter are compared for a stationary interference and a pulsed interferer, respectively. Reference [74] compared the use of a linear equalizer (LE) to the decision-feedback equalizer (DFE) in the context of interference mitigation. The DFE was shown to significantly outperform the LE due to its nonlinear structure.

A number of works make comparisons between the strategies as seen in [73, 75]. For instance, [68] compared the use of Fourier transform techniques and linear prediction. Finally, reviews of interference mitigation techniques for use in spread spectrum systems can be found in [84, 85]. Also available are reviews of interference mitigation using adaptive notch filtering [70] and using nonlinear estimation techniques [96]. A thorough review of adaptive interference suppression not limited to narrowband interference can be found in [60].

1.4 Dissertation Focus

The issue of mitigating narrowband interference has been a well-studied subject matter. However, the aspect of very strong interferers and their affect on the performance of communication systems has not been resolved in sufficient detail to allow the successful operation of systems under conditions of severe interference. It is the goal of this dissertation to provide solutions that allow effective performance (in terms of BER and complexity) in severe interference for both single and multi-carrier communication systems.

Li and Milstein [74] originally looked at the performance of the adaptive decision-feedback equalizer (DFE) in regards to mitigation of narrowband interference. This work examined the optimal weights of the equalizer, as well as the steady-state BER results of the system. Also examined was the convergence of the least-mean square (LMS) algorithm when the signal-to-interference ratio was moderate, i.e. SIR = 0 dB. In this work, a single-carrier system is examined in the presence of a severe narrowband interferer, i.e. SIR = -20 dB. It is shown that the narrowband interference causes an increase in the time required for the adaptive algorithm to converge. This is due to

the fact that the equalizer does not have a true reference for the interference. As a consequence, a large number of training symbols is required to construct a reference from the received interference contaminated signal. This reduces the efficiency of the system and wastes bandwidth. In this dissertation, two techniques are proposed for improving the convergence of the adaptive equalizers. The first technique, entitled data-aided initialization (DAI), aims to initialize the weights of the adaptive algorithm with an estimate of the Wiener weights. This estimate is obtained from averages of the autocorrelation matrix and the cross-correlation vector derived from the received samples and the training data. Further, two techniques to efficiently obtain the estimate of the Wiener weights is examined, in terms of complexity and steady-state BER as compared to the theoretical Wiener filter. A second approach proposes the use of an adaptive prediction-error filter (PEF) as a pre-filter to the DFE to improve the convergence of the overall system. The PEF generates a direct reference for the strong interference from past samples and mitigates it prior to equalization. The convergence time and steady-state BER of this system are compared to the case of the DFE-only, when both systems have the same number of total taps.

In order to satisfy the demands for increased capacity in modern communication systems, the modulation schemes have shifted from single-carrier to multi-carrier modulation approaches. We next examine how to best extend the results for single-carrier systems to the mitigation of narrowband interference in multi-carrier systems. A number of approaches have been proposed for this problem, especially in the uncoded scenario. Saito, et al. [105] proposed the idea of hole-punching, where the transmitter refrains from sending data on subcarriers that are experiencing interference. This obviously requires coordination between the transmitter and receiver and may be overly complex, especially if the interference is non-stationary or the coherence time of the channel exceeds the feedback delay of the system. Another technique utilizes orthogonal codes [23,45,137] to spread the data into every subcarrier thereby providing frequency diversity, as will be described for multi-carrier code division multiple access (MC-CDMA). The receiver must determine the (optimal) combiner weights to successfully recover the data. Spread-spectrum methods [78, 109] have also been investigated in an effort to reduce the interference power seen at the receiver. Nilsson, et al. [91] examined the use of frequency-domain subtractive cancellation using the singular value decomposition in conjunction with silent tones. Another frequency-domain mitigation technique was

proposed by Darsena [34], using successive interference cancellation to remove the interference. This technique amounts to linear prediction of the interference using estimates generated from more reliable tones. Receiver windowing has also been investigated by a number of researchers [32, 33, 101] using the redundancy of the cyclic prefix to minimize the interference power. Finally, Coulson [29] proposed the use of excision-based filtering to remove the interference in the time-domain prior to demodulation. This technique is a well known method, often used in conjunction with spread-spectrum systems [84] when the processing gain does not provide enough immunity to the interference. Coulson utilizes a number of phase locked loops (PLLs) to estimate and track the interference during a signal-free period between data transmissions. A pre-defined excision filter is then employed to excise the interference. Note that this filter is defined such that the impulse response is greater than the duration of the cyclic prefix, causing distortion in the form of intersymbol interference.

More recently, work has been done on examining mitigation strategies in coded orthogonal frequency division multiplexing (OFDM) systems. Wu and Nassar [137] examined the case of coded OFDM with and without orthogonal codes through simulation. Similarly, Coulson [29] briefly mentions the use of coding and excision filtering, demonstrating results through simulation. Li et al. [76] proposed an advanced joint erasure and decoding scheme that exploits the structure of the code. This method can substantially increase the complexity of the system. Snow et al. [118] primarily examined the problem of efficient error rate analysis for coded multi-carrier systems, but also briefly mentioned the use of erasure insertion in a coded OFDM system. These authors proceeded to examine the case of WiMAX interfering with UWB systems [119, 120], using spectrum estimation (similar to linear prediction) techniques to appropriately weight the bit metrics inputted into the Viterbi decoder.

In this dissertation, two multi-carrier systems are investigated in the presence of narrowband interference. It is shown that demodulation of the interference causes a large number of subcarriers to be degraded, due to the spectral leakage of the interference power. MC-CDMA possesses frequency diversity from spreading to mitigate narrowband interference, even without the use of forward error correction coding. This system, however, requires a complex combiner at the receiver to extract the transmitted data symbols. On the other hand, OFDM is investigated due to its simple equalization structure. However, a lack of inherent frequency diversity requires the use of forward

error correction coding and interleaving. The use of coding alone is not able to mitigate the interference. It will be shown that even with the use of genie inserted erasures, little to no improvement is achieved with this technique when the interference is strong. The technique proposed here uses the PEF as an erasure insertion mechanism that concentrates the erasures in the vicinity of the narrowband interference. The length of the PEF is limited to the unused portion of the cyclic prefix. Note that the cyclic prefix is designed to deal with the delay spread of the channel. An adaptive PEF can be used to allow a time-varying interference to be tracked over time.

1.5 Dissertation Overview

The body of this dissertation is organized as follows. In Chapter 2, the system model for the single-carrier case is introduced as well as filtering structures pertinent to mitigating the interference are reviewed. The adaptive algorithms that will be used are detailed and discussed. In Chapter 3, the non-Wiener effect of linear equalizers adapted via the LMS algorithm is demonstrated in an environment corrupted by narrowband interference and multipath, however, the strong narrowband interference causes the adaptive equalizer to require an extremely long time for convergence. Further, data-aided initialization (DAI) is developed and shown to improve the convergence of adaptive equalizers. Two techniques for estimating the Wiener weights are compared in terms of complexity and BER performance relative to the theoretical Wiener filter. In Chapter 4, a two-stage approach to improve the convergence of the DFE is developed. This two-stage approach utilizes the prediction-error filter (PEF) as a pre-filter to the DFE, generating a direct reference for the interference from past samples and mitigates it prior to equalization. In Chapter 5, the basics of multi-carrier modulation are presented. In Chapter 6, the effect of narrowband interference is examined in a multipath channel for bit-interleaved coded modulated (BICM) multi-carrier systems, MC-CDMA and OFDM. Each modulation scheme is viewed in both the uncoded and coded cases. The PEF is utilized in an OFDM system as an erasure insertion mechanism for which an analytical upper bound is provided and compared to the use of conventional erasures. Finally, concluding remarks and future research are provided in Chapter 7.

2 Background

Ideally, the interference term would be known exactly at receiver. If this was the case, the interference could be subtracted from the received signal to give the desired signal simply distorted by additive noise and the problem would be solved. Conversely, if the interference is known at the transmitter, techniques such as dirty-paper coding [28], have been developed to ensure that the received signal is free of the interference. In this case, the receiver must feedback information of the interference to the transmitter which may become outdated at the next transmission interval.

2.1 Minimum Mean-Square Error Filtering

In general, the exact interference term is not known to either the transmitter or receiver, requiring the use of other means to mitigate the interference. Several methods for suppressing narrowband interference have been discussed in the literature (see Section 1.3). A linear equalizer (LE) and a decision-feedback equalizer (DFE) were studied in [74]. It was shown that the performance of the DFE is better than that of the LE. The LE seen in both systems removes the interference, while the additional feedback taps of the DFE enable the cancellation of the post-cursor ISI that is induced by the LE. Linear prediction [81, 140] is another common technique that has been used in direct-sequence code-division multiple access (CDMA) systems [73, 102, 124] when the processing gain does not provide enough immunity to the interference. When the signal of interest is wideband compared to the bandwidth of the interferer, linear prediction estimates the current value of the interference from past samples. When the structure is implemented as a prediction-error filter (PEF), the estimate of the interference is removed at the cost of some signal distortion. A further review of interference suppression techniques can be found in [70, 84].

When the statistics of the interference are known, the weights of these systems

are found by minimizing the mean-squared error [56]. This is equivalent to solving the Wiener-Hopf equations [134]. These equations arise from the minimum mean-squared error (MMSE) criterion as follows. Assume that a linear estimator of order M attempts to estimate the desired response, y, from the data $\mathbf{x} = [x_1, x_2, \ldots, x_M]$,

$$\hat{y} = \sum_{m=1}^{M} w_m^* x_m = \mathbf{w}^H \mathbf{x}, \tag{2.1}$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_M]^T$ is the coefficient vector of the estimator. The error term given by $e = y - \hat{y}$ and the mean-squared error (MSE) is defined as

$$J = \mathbb{E}\left[|e|^2\right], \tag{2.2}$$

This term can be expanded as

$$\begin{aligned} J(\mathbf{w}) &= \mathbb{E}\left[|y|^2\right] - \mathbf{w}^H \mathbb{E}\left[\mathbf{x}y^*\right] - \mathbb{E}\left[y\mathbf{x}^*\right]\mathbf{w} + \mathbf{w}^H \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]\mathbf{w} \\ &= \mathbb{E}\left[|y|^2\right] - \mathbf{w}^H \mathbf{p} - \mathbf{p}^H \mathbf{w} + \mathbf{w}^H \mathbf{R}\mathbf{w}, \end{aligned} \tag{2.3}$$

where $\mathbf{p} = \mathbb{E}\left[\mathbf{x}y^*\right]$ is the cross-correlation vector between the data vector and the desired response, and $\mathbf{R} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$ is the autocorrelation of the input data vector that is Hermitian and non-negative definite. Further, (2.3) can be written in the form of a perfect square,

$$J(\mathbf{w}) = \mathbb{E}\left[|y|^2\right] - \mathbf{p}^H \mathbf{R}^{-1}\mathbf{p} + (\mathbf{R}\mathbf{w} - \mathbf{p})^H \mathbf{R}^{-1} (\mathbf{R}\mathbf{w} - \mathbf{p}), \tag{2.4}$$

where $\mathbf{R}^{-1}$ is the inverse of the autocorrelation matrix. When $\mathbf{R}$ is positive definite, $\mathbf{R}^{-1}$ exists and is also positive definite. This then implies that $\mathbf{p}^H \mathbf{R}^{-1}\mathbf{p}$ is greater than zero for $\mathbf{p} \neq \mathbf{0}$ and decreases the MSE. On the other hand, $(\mathbf{R}\mathbf{w} - \mathbf{p})^H \mathbf{R}^{-1} (\mathbf{R}\mathbf{w} - \mathbf{p})$ is greater than zero and will increase the MSE. This term is then minimized when $\mathbf{R}\mathbf{w} - \mathbf{p} = \mathbf{0}$. Therefore the necessary and sufficient conditions for minimizing the MSE is

$$\mathbf{R}\mathbf{w}_{\text{opt}} = \mathbf{p}. \tag{2.5}$$

The system of equations in (2.5) are referred to as the Wiener-Hopf equations.

In practice, statistical information of the interference may not available a priori. Thus, these systems are best implemented adaptively.

2.2 Adaptive Algorithms

Of the many algorithms available, in this dissertation the focus is placed on the low-complexity method provided by the least-mean square (LMS) algorithm [56]. Although it has been shown that alternate adaptive algorithms, such as the recursive least squares (RLS) algorithm [56] utilizes time averages of the data sequence to provide improved convergence relative to the LMS algorithm in cases of high eigenvalue disparity, there are many reasons why LMS is chosen for practical communications system applications. Hassibi [55] discusses some of the fundamental differences in the performance of gradient based estimators such as the LMS algorithm and time averaged recursive estimators such as the RLS algorithm in the cases of modeling errors and incomplete statistical information concerning the input signal, interference, and noise parameters. Hassibi [55] examines the conditions for which LMS can be shown to be more robust to variations and uncertainties in the signaling environment than RLS. LMS has also been shown to track more accurately than RLS because it is able to base the filter updates on the instantaneous error rather than the time averaged error [14, 79, 80, 130]. The improved tracking performance of LMS over RLS for a linear chirp input is well established [56, 130]. In [58] it is shown that an extended RLS filter that estimates the chirp rate of the input signal can minimize the tracking errors associated with the RLS algorithm and provides performance that exceeds that of LMS. It should be noted however, that the improved tracking performance requires a significant increase in computational complexity and knowledge that the underlying variations in the input signal can be accurately modeled by a linear frequency modulated (FM) chirp. For cases where the input is not accurately represented by the linear chirp model, performance can be expected to be significantly worse than simply using an LMS estimator, for the reasons discussed in [55]. The computational complexity of RLS, in particular for high order systems, favors the use of LMS. The latter is also more robust in fixed-point implementations. In addition the LMS estimator has been shown to provide nonlinear, time-varying weight dynamics that allow the LMS filter to perform significantly better than the time-invariant Wiener filter in several cases of practical interest [13, 104] and especially for the case of narrowband interference that is addressed in this dissertation. It is further shown that the improved performance associated with these non-Wiener effects is difficult to realize for RLS estimators due to the time-averaging that is inherent in the estimation process [12].

2.2.1 Least-Mean Square Algorithm

The LMS algorithm [56] is defined by the following three equations:

$$y_l = \mathbf{w}_l^H \mathbf{x}_l, \tag{2.6}$$

$$e_l = \begin{cases} d_l - y_l, & \text{Training,} \\ \hat{d}_l - y_l, & \text{Decision-Directed,} \end{cases} \tag{2.7}$$

$$\mathbf{w}_{l+1} = \mathbf{w}_l + \mu e_l^* \mathbf{x}_l, \tag{2.8}$$

where $\mathbf{x}_l$ is the input vector to the equalizer, $\mathbf{w}_l$ is the vector of adapted tap weights, d_l is the desired signal, $\hat{d}_l$ is the output of the decision-device when y_l is its input, e_l is the error signal, and μ is the step-size parameter.

Note that there are two stages associated with the adaptive algorithm. The first stage is the training phase, where known training symbols are used to push the filter in the direction of the optimal weights. After the training symbols have been exhausted, the algorithm switches to decision-directed mode. The output of the decision device is used as the desired symbol when calculating the error signal. Ideally, at the end of the training phase the output of the filter is close to the desired signal.

The LMS algorithm is noted for its robustness and improved tracking performance [55, 56]. The drawback of this particular algorithm is its slow convergence when there is a large disparity in the eigenvalues of the input signal [56]. Slow convergence leads to the need for a large number of training symbols. These symbols do not transmit any new information, reducing the overall throughput of the system.

2.2.2 Normalized Least-Mean Square Algorithm

Note that the weight adjustment seen in (2.8) is proportional to input vector, $\mathbf{x}_l$, and when this vector is large, the LMS algorithm suffers from gradient noise amplification. To combat this issue, the normalized LMS (NLMS) [3, 18, 19, 89] algorithm is utilized. The NLMS algorithm [56] is defined by the following three equations:

$$y_l = \mathbf{w}_l^H \mathbf{x}_l, \tag{2.9}$$

$$e_l = \begin{cases} d_l - y_l, & \text{Training,} \\ \hat{d}_l - y_l, & \text{Decision-Directed,} \end{cases} \tag{2.10}$$

$$\mathbf{w}_{l+1} = \mathbf{w}_l + \frac{\mu}{\|\mathbf{x}_l\|^2} e_l^* \mathbf{x}_l. \tag{2.11}$$

Note that the term $\frac{\mu}{\|\mathbf{x}_l\|^2}$ can be thought of as a time-varying step-size that changes based on the squared Euclidean norm of the input to equalizer, $\mathbf{x}_l$.

2.2.3 Mean-Squared Error

The MSE is used as the metric that is minimized by the adaptive algorithms discussed above [56] and is defined in (2.2). The MMSE, defined as $J_{\min}$, of the system is obtained when the system operates in its optimum condition, i.e. when the the weights are found by minimizing the MSE criterion, $\mathbf{w}_{\text{opt}}$. When these weights are utilized it is noted that the optimal output of the filter, $y_{\text{opt},l}$, is orthogonal to its corresponding estimation error, $e_{\text{opt},l} = d_l - y_{\text{opt},l}$ [56, Corollary to the Principle of Orthogonality],

$$\mathbb{E}\left[y_{\text{opt},l} e^*_{\text{opt},l}\right] = 0. \tag{2.12}$$

This arises from the orthogonality principle, that states that the optimum estimation error, $e_{\text{opt},l}$, is orthogonal to each input sample to the filter,

$$\mathbb{E}\left[x_{l-m} e^*_{\text{opt},l}\right] = 0, \quad m = 0, 1, \ldots, M-1. \tag{2.13}$$

2.2.4 Nonlinear Effects

Although the output of the filter in (2.9) is simply a linear combination of the inputs, weighted by the coefficients of the filter, the time varying behavior of the adaptive filter weights produce a nonlinear response as described in detail in [99,104]. A recursive expansion of the weight update equation given in (2.11) can be written as,

$$\mathbf{w}_{l+1} = \mathbf{w}_{l-1} + \mu \left(\frac{\mathbf{u}_{l-1} e^*_{l-1}}{\|\mathbf{u}_{l-1}\|^2} + \frac{\mathbf{u}_l e^*_l}{\|\mathbf{u}_l\|^2} \right). \tag{2.14}$$

This shows that the weights are a function of not only the current input vector and error value, but also of past input vectors and past errors. In fact, if the weights are initialized to all zeros (i.e. $\mathbf{w}_0 = \mathbf{0}$), then (2.14) can be rewritten as

$$\begin{aligned} \mathbf{w}_{l+1} &= \mathbf{w}_0 + \mu \sum_{m=0}^{l} \frac{\mathbf{u}_m e^*_m}{\|\mathbf{u}_m\|^2} \\ &= \mu \sum_{m=0}^{l} \frac{\mathbf{u}_m e^*_m}{\|\mathbf{u}_m\|^2}. \end{aligned} \tag{2.15}$$

This clearly demonstrates that the current weight update is a function of all past input vectors and error terms. At large step-sizes, this property provides the NLMS algorithm with the nonlinear dynamics to outperform the time-invariant (TI) Wiener solution [13].

In the equalization context, the linear adaptive decision-directed equalizer (DDE) (see Section 2.3) is found capable of suppressing the interference better than the equivalent optimal TI linear equalizer. The added performance benefit is a result of the nonlinear instantaneous characteristics of the NLMS algorithm [13]. The instantaneous error signal produces dynamic weight updating that effectively causes the filter to track the interference and allows the performance of the linear DDE to approach that of the nonlinear DFE [104].

2.2.5 Convergence Analysis

In conventional analyses, convergence refers to the asymptotic progress of either the adaptive weights or the MSE toward the optimal solutions ($\mathbf{w}_{\text{opt}}$ and $J_{\min}$, respectively). The convergence (as well as the stability) of the system is dependent on the step-size. The step-size parameter is chosen in a manner to guarantee convergence in the mean-square sense, namely

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{2.16}$$

where $\lambda_{\max}$ is the maximum eigenvalue of the input autocorrelation matrix.

Assuming that the adaptive weights and the input vector are independent, Shensa [110] showed that the convergence of the weight vector can be expressed as

$$\|\mathbf{w}_{\text{opt}} - \mathbb{E}[\mathbf{w}_l]\|^2 = \sum_{i=1}^{M} (1 - \mu\lambda_i)^{2k} \left| \mathbf{v}^{i\,H} \mathbf{w}_{\text{opt}} \right|^2, \tag{2.17}$$

where λ_i are the eigenvalues and $\mathbf{v}^i$ are the eigenvectors of the input autocorrelation matrix. A similar equation arises for the convergence of the MSE [111], when gradient noise (on the order of $\mu M J_{\min}$) is neglected

$$\left\| \mathbb{E}[e_l^2] - J_{\min} \right\|^2 = \sum_{i=1}^{M} (1 - \mu\lambda_i)^{2k} \lambda_i \left| \mathbf{v}^{i\,H} \mathbf{w}_{\text{opt}} \right|^2. \tag{2.18}$$

In general, the learning curve is composed of a number of modes associated with each of the eigenvalues of the input. As discussed in [56], the learning curve can approximated by associating a single exponential with each mode such that a time constant can be defined for each mode as

$$\tau_i \simeq \frac{1}{2\mu\lambda_i}. \tag{2.19}$$

The maximum modal time constant is thus associated with the minimum eigenvalue,

$$\tau_{\max} \simeq \frac{1}{2\mu\lambda_{\min}}. \tag{2.20}$$

This maximal time constant can be seen to be a conservative estimate by examining (2.17) more closely. The convergence will be influenced only by those eigenvalues for which the projection of the corresponding eigenvector on the optimal weights is large. Lastly, it can be seen for the case of $\lambda_i \ll 1$, that it is possible for the convergence of the filter output (MSE) to be faster than the convergence of the filter weights. This is because there may be fewer modes controlling the MSE convergence (i.e. when $\lambda_i |\mathbf{v}^{i\,H} \mathbf{w}_{\text{opt}}| < |\mathbf{v}^{i\,H} \mathbf{w}_{\text{opt}}|$).

The equations above provide excellent insight into the convergence of the LMS algorithm in terms of MSE. In order to evaluate the impact of the processing techniques employed in this application, the convergence must be determined in terms of a metric that directly relates to the performance of the specified communication system. The metric that applies to all the systems considered is the bit error rate (BER). Consequently, the convergence is defined to be the average number of training symbols needed to achieve a BER of 10^{-2}. This value is consistent with the requirement that the BER should be less than 10^{-1} when switching from training to decision-directed mode [5]. Additionally, a convolutional code with an input BER equal to 10^{-2} is equivalent to a BER of 10^{-5} at the output of the decoder [92].

2.2.6 BER Sliding Window

To examine the convergence of the BER here, a sliding window of N_{window} symbols is employed. For example, the first BER value corresponds to the average number of bit errors over symbols 1 through N_{window}; the second value corresponds to the average number of bit errors over symbols 2 through $N_{\text{window}} + 1$; etc. These values are then averaged for N_{runs} trials. A general formula for BPSK modulation can be seen as

$$\text{BER}_l = \frac{1}{N_{\text{runs}}} \sum_{n=1}^{N_{\text{runs}}} \frac{1}{N_{\text{window}}} \sum_{m=k-N_{\text{window}}+1}^{k} \left| d_m^{(n)} - \hat{d}_m^{(n)} \right|, \quad l \geq N_{\text{window}}, \tag{2.21}$$

where $d_m^{(n)}$ is the m^{th} transmitted symbol of the n^{th} packet and $\hat{d}_m^{(n)}$ is the decision of the m^{th} symbol of the n^{th} packet. Note that the minimum non-zero BER value will be equal to $\frac{1}{N_{\text{runs}} N_{\text{window}}}$. Convergence is then defined as the average number of symbols needed to attain a certain BER value.

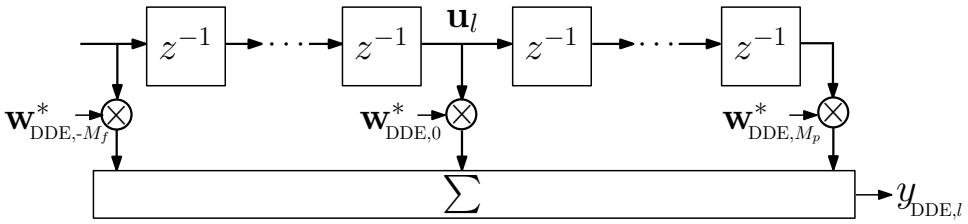


Figure 2.1: Decision-directed equalizer block diagram.

2.3 Decision-Directed Equalizer

2.3.1 Equalizer Structure

Let the discrete-time output of the matched filter be given by u_l, noting that this signal may be oversampled by a factor of L_o. The decision-directed equalizer (DDE) is a linear structure that consists of a transversal filter, as depicted in the block diagram given in Figure 2.1. The filter is composed of the M_p postcursor symbols, the symbol of interest, and M_f precursor symbols.

The output of the filter, $y_{\text{DDE},l}$, is defined as

$$y_{\text{DDE},l} = \mathbf{w}_{\text{DDE}}^H \mathbf{x}_l, \tag{2.22}$$

where $\mathbf{w}_{\text{DDE}}$ is an $(M_f + M_p + 1)L_o \times 1$ vector of equalizer tap weights. The equalizer is fractionally sampled, however it operates at the symbol rate, thus requiring an input of L_o samples at each iteration. The input to the delay line is a vector, $\mathbf{u}_l$, which is formed from L_o samples of u_l,

$$\mathbf{u}_l = \begin{bmatrix} u_{L_o(l+1)-1}, & \cdots, & u_{L_o l+1}, & u_{L_o l} \end{bmatrix}^T. \tag{2.23}$$

The input to the equalizer becomes

$$\mathbf{x}_l = \begin{bmatrix} \mathbf{u}_{l+M_f}^T, & \cdots, & \mathbf{u}_l^T, & \cdots, \mathbf{u}_{l-M_p}^T \end{bmatrix}^T. \tag{2.24}$$

For the remaining properties of the DDE, the case of a one-sided DDE ($M_f = 0$) and input at the symbol rate (whether by $L_o = 1$ or downsampling the oversampled signal) is considered. This implies that (2.23) and (2.24) becomes

$$x_l = u_l = d_l + i_l + n_l. \tag{2.25}$$

These signals were previously introduced in Chapter 1. The output, (2.22), also reduces to,

$$y_{\mathrm{DDE},l} = \mathbf{w}_{\mathrm{DDE}}^H \mathbf{x}_l = \sum_{m=0}^{M_p} w_{\mathrm{DDE},m}^* x_{l-m}. \tag{2.26}$$

Of interest also, is the autocorrelation function of (2.25), where it is assumed that the components of the received signal are independent. The autocorrelation function is then given by

$$\begin{aligned} r_x(m) &= \mathbb{E}\left[x_l x_{l-m}^*\right] \\ &= \left(E_s + \sigma_n^2\right)\delta_m + E_i e^{j\Omega_i m T_s}. \end{aligned} \tag{2.27}$$

2.3.2 Optimal Weights

The optimal weights under the MMSE criterion can be found using the orthogonality principle [56]. $M_p + 1$ equations are obtained, and the weights can be found using the method of undetermined coefficients described in [74, 141]. The optimal DDE tap weights are given by

$$\begin{aligned} w_{\mathrm{DDE},l} &= \frac{\mathrm{SNR}\left[(1+\mathrm{SNR})\sigma_n^2 + M_p E_i\right]}{(1+\mathrm{SNR})\left[(1+\mathrm{SNR})\sigma_n^2 + (M_p+1)E_i\right]} \\ &= C_0, \quad l = 0, \end{aligned} \tag{2.28}$$

$$\begin{aligned} w_{\mathrm{DDE},l} &= \frac{-E_i \mathrm{SNR}}{(1+\mathrm{SNR})\left[(1+\mathrm{SNR})\sigma_n^2 + (M_p+1)E_i\right]} e^{-j\Omega l T_s} \\ &= \frac{C_1}{1+\mathrm{SNR}} e^{-j\Omega l T_s}, \quad l = 1, \ldots, M_p. \end{aligned} \tag{2.29}$$

2.3.3 Minimum Mean-Squared Error

Following the discussion of Section 2.2.3, the MMSE is obtained using the optimal weights given in (2.28) and (2.29). Using these weights makes the output of the filter, $y_{\mathrm{DDE},l}$, optimal in the MMSE sense, and this term is defined in (2.26). The optimal estimation error term is defined as $e_{\mathrm{opt},l} = d_l - y_{\mathrm{DDE},l}$. Based on the orthogonality principle and its corollary, the following relationships are then true:

$$\mathbb{E}\left[x_{l-m} e_{\mathrm{opt},l}^*\right] = 0, \quad m = 0, 1, \ldots, M_p, \tag{2.30}$$

$$\mathbb{E}\left[y_{\mathrm{DDE},l} e_{\mathrm{opt},l}^*\right] = 0. \tag{2.31}$$

From (2.30) and (2.31) the DDE MMSE value is found to be

$$\begin{aligned}
J_{\text{DDE,min}} &= \mathbb{E}\left[|e_{\text{opt},l}|^2\right] = \mathbb{E}\left[e_{\text{opt},l}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[\left(d_l - y_{\text{DDE},l}\right)e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l e^*_{\text{opt},l}\right] - \mathbb{E}\left[y_{\text{DDE},l}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l\left(d_l - y^*_{\text{DDE},l}\right)\right] \\
&= \mathbb{E}\left[d_l d_l^*\right] - \mathbb{E}\left[d_l\left(\sum_{m=0}^{M_p} w^*_{\text{DDE},m}x_{l-m}\right)^*\right] \\
&= E_s\delta_0 - \sum_{m=0}^{M_p} w_{\text{DDE},m}\mathbb{E}\left[d_l x^*_{l-m}\right] \\
&= E_s - \sum_{m=0}^{M_p} w_{\text{DDE},m}\mathbb{E}\left[d_l\left(d_{l-m} + i_{l-m} + n_{l-m}\right)^*\right] \\
&= E_s - \sum_{m=0}^{M_p} w_{\text{DDE},m}\mathbb{E}\left[d_l d^*_{l-m}\right] \\
&= E_s - E_s\sum_{m=0}^{M_p} w_{\text{DDE},m}\delta_m \\
&= E_s - E_s w_{\text{DDE},0} = E_s\left(1 - w_{\text{DDE},0}\right) = E_s\left(1 - C_0\right). \qquad (2.32)
\end{aligned}$$

2.3.4 Signal-to-Interference-Plus-Noise Ratio

The signal-to-interference-plus-noise ratio (SINR) at the input to the decision device of the DDE can be found by examining (2.26) and the optimal weights (2.28) and (2.29). The output of the DDE can be written as

$$\begin{aligned}
y_{\text{DDE},l} &= \sum_{m=0}^{M_p} w^*_{\text{DDE},m}\left(d_{l-m} + i_{l-m} + n_{l-m}\right) \\
&= \underbrace{\sum_{m=0}^{M_p} w^*_{\text{DFE},m}d_{l-m}}_{\zeta_d} + \underbrace{\sum_{m=0}^{M_p} w^*_{\text{DFE},m}i_{l-m}}_{\zeta_i} + \underbrace{\sum_{m=0}^{M_p} w^*_{\text{DFE},m}n_{l-m}}_{\zeta_n}, \qquad (2.33)
\end{aligned}$$

and is composed of the filtered signal (ζ_d), the filtered interference (ζ_i), and the filtered noise (ζ_n). The SINR is defined as,

$$\text{SINR} = \frac{\mathbb{E}\left[|\zeta_d|^2\right]}{\mathbb{E}\left[|\zeta_i|^2\right] + \mathbb{E}\left[|\zeta_n|^2\right]}. \qquad (2.34)$$

Each signal term is determined as follows:

$$\begin{aligned}
\mathbb{E}\left[|\zeta_d|^2\right] &= \mathbb{E}\left[\left(\sum_{m=0}^{M_p} w^*_{\mathrm{DDE},m} d_{l-m}\right)\left(\sum_{k=0}^{M_p} w^*_{\mathrm{DDE},k} d_{l-k}\right)^*\right] \\
&= \sum_{m=0}^{M_p}\sum_{k=0}^{M_p} w^*_{\mathrm{DDE},m} w_{\mathrm{DDE},k} \mathbb{E}\left[d_{l-m} d^*_{l-k}\right] \\
&= \sum_{m=0}^{M_p}\sum_{k=0}^{M_p} w^*_{\mathrm{DDE},m} w_{\mathrm{DDE},k} E_s \delta_{k-m} \\
&= \sum_{m=0}^{M_p} |w_{\mathrm{DDE},m}|^2 E_s \\
&= \left(C_0^2 + \sum_{m=1}^{M_p} \left|\frac{C_1}{1+\mathrm{SNR}} e^{-j\Omega l T_s}\right|^2\right) E_s \\
&= \left(C_0^2 + M_p\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2\right) E_s. \qquad (2.35)
\end{aligned}$$

The noise term follows exactly as the previous derivation and is given by,

$$\begin{aligned}
\mathbb{E}\left[|\zeta_d|^2\right] &= \mathbb{E}\left[\left(\sum_{m=0}^{M_p} w^*_{\mathrm{DDE},m} n_{l-m}\right)\left(\sum_{k=0}^{M_p} w^*_{\mathrm{DDE},k} n_{l-k}\right)^*\right] \\
&= \left(C_0^2 + M_p\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2\right) \sigma_n^2. \qquad (2.36)
\end{aligned}$$

Finally, the interference term is derived to be,

$$\begin{aligned}
\mathbb{E}\left[|\zeta_i|^2\right] &= \mathbb{E}\left[\left|\sum_{m=0}^{M_p} w^*_{\mathrm{DDE},m} i_{l-m}\right|^2\right] \\
&= \mathbb{E}\left[\left|w^*_{\mathrm{DDE},0} i_l + \sum_{m=1}^{M_p} w^*_{\mathrm{DDE},m} i_{l-m}\right|^2\right] \\
&= \mathbb{E}\left[\left|C_0\sqrt{E_i} e^{j(\Omega l T_s+\theta)} + \sum_{m=1}^{M_p} \frac{C_1}{1+\mathrm{SNR}} e^{j\Omega m T_s}\sqrt{E_i} e^{j(\Omega(l-m)T_s+\theta)}\right|^2\right] \\
&= \mathbb{E}\left[\left|\left(C_0 + M_p\frac{C_1}{1+\mathrm{SNR}}\right)\sqrt{E_i} e^{j(\Omega l T_s+\theta)}\right|^2\right] \\
&= \left(C_0 + M_p\frac{C_1}{1+\mathrm{SNR}}\right)^2 E_i. \qquad (2.37)
\end{aligned}$$

Thus the SINR is given by,

$$\mathrm{SINR} = \frac{\left(C_0^2 + M_p\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2\right)\mathrm{SNR}}{\left(C_0 + M_p\frac{C_1}{1+\mathrm{SNR}}\right)^2\frac{E_i}{\sigma_n^2} + C_0^2 + M_p\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2}. \tag{2.38}$$

2.3.5 Autocorrelation Structure

The $(M_p+1)\times(M_p+1)$ input autocorrelation matrix for the DDE is defined as

$$\begin{aligned}\mathbf{R}_{\mathrm{DDE}} &= \mathbb{E}\left[\mathbf{x}_l\mathbf{x}_l^H\right] \\ &= \begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(M_p) \\ r_x^*(1) & r_x(0) & r_x(1) & \cdots & r_x(M_p-1) \\ r_x^*(2) & r_x^*(1) & r_x(0) & \cdots & r_x(M_p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x^*(M_p) & r_x^*(M_p-1) & r_x^*(M_{\mathrm{pef}}-3) & \cdots & r_x(0) \end{bmatrix},\end{aligned} \tag{2.39}$$

where the components of the matrix are given by (2.27).

2.3.6 Eigenvalues

The eigenvalues for the correlation matrix given by (2.39), can be found [102, 110, 127] to be equal to

$$\lambda_{\mathrm{DDE}} = \begin{cases} E_s + \sigma_n^2 + (M_p+1)E_i, & \text{order 1,} \\ E_s + \sigma_n^2, & \text{order } M_p. \end{cases} \tag{2.40}$$

The eigenvalue spread is defined [56] as

$$\chi(\mathbf{R}_{\mathrm{DDE,i}}) = \frac{\lambda_{\mathrm{PEF,max}}}{\lambda_{\mathrm{PEF,min}}} = 1 + \frac{(M_p+1)E_i}{E_s+\sigma_n^2}. \tag{2.41}$$

2.3.7 Convergence Properties

The projection of any of the eigenvectors on the optimal weight vector is nonzero. This implies that the time constant (2.20) is proportional to the minimum eigenvalue (i.e. $\tau_{\mathrm{DDE}} \simeq 1/2\mu(E_s+\sigma_n^2)$. The delay in convergence can be attributed to the fact that the DDE does not have a direct reference for the interferer during adaptation and is thus forced to converge on the basis of the training data only.

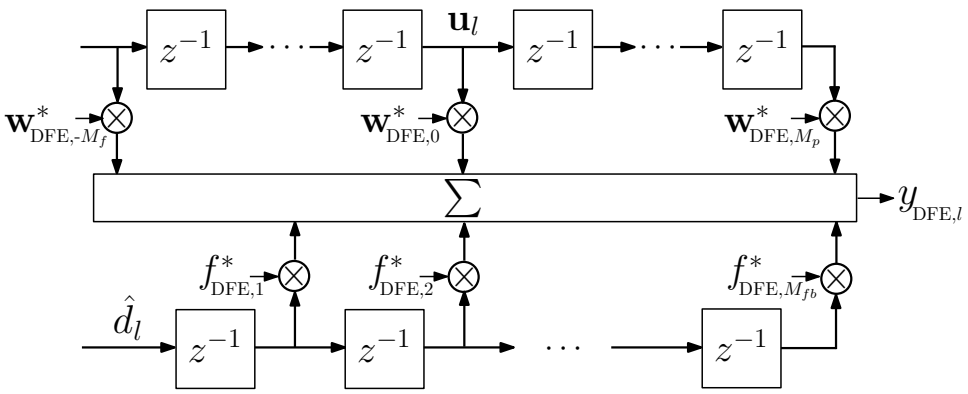


Figure 2.2: Decision-feedback equalizer block diagram.

2.4 Decision-Feedback Equalizer

2.4.1 Equalizer Structure

The decision-feedback equalizer (DFE) is a nonlinear structure that consists of a feedforward filter and a feedback filter, and operates at the symbol rate. The block diagram for this system is shown in Figure 2.2. The feedforward filter is a transversal filter, similar that of the DDE. This filter is composed of M_p postcursor symbols, the symbol of interest, and M_f precursor symbols. The feedback filter feeds back the M_{fb} most recently decided symbols as an input to the equalizer. The input to the equalizer is then given by

$$\mathbf{x}_l = \left[\mathbf{u}^T_{l+M_f}, \quad \cdots, \quad \mathbf{u}^T_l, \quad \cdots, \mathbf{u}^T_{l-M_p} | \hat{d}_{l-1}, \quad \ldots, \quad \hat{d}_{l-M_{fb}}\right]^T. \tag{2.42}$$

where $\mathbf{u}_l$ is given in (2.23) and $\hat{d}_l$ are the fed back decisions. During the training phase, $\hat{d}_l = d_l$. The output of equalizer, $y_{\mathrm{DFE},l}$, is defined to be

$$y_{\mathrm{DFE},l} = \left[\mathbf{w}^T_{\mathrm{DFE}}, \mathbf{f}^T_{\mathrm{DFE}}\right]^* \mathbf{x}_l, \tag{2.43}$$

where $\mathbf{w}_{\mathrm{DFE}}$ is an $(M_f + M_p + 1)L_o \times 1$ vector of tap weights associated with the feedforward filter and $\mathbf{f}_{\mathrm{DFE}}$ is an $M_{fb} \times 1$ vector of tap weights associated with the feedback filter.

The impact of feeding back decisions, is that the noise associated with these symbols has been removed by the decision device and the post-cursor ISI may be removed. However, this situation leads to error propagation through the feedback filter when the decisions that are fed back are incorrect. A BER analysis of the DFE with error

propagation can be accomplished utilizing Markov chains to model the term $[d_{k-l} - \hat{d}_{k-l}]$ as contents of a shift register and the assumption that the fed back decisions are perfect [74, 86, 97, 117]. The number of states in the Markov chain grows exponentially with the number of feedback taps.

For the remaining properties of the DFE, the case of a one-sided DFE ($M_f = 0$) and input at the symbol rate (whether by $L_o = 1$ or downsampling the oversampled signal) is considered, similar to Section 2.3.1. In this case the output of equalizer becomes

$$y_{\mathrm{DFE},l} = \left[\mathbf{w}_{\mathrm{DFE}}^T, \mathbf{f}_{\mathrm{DFE}}^T\right]^* \mathbf{x}_l = y_{\mathrm{DFE},l} = \sum_{m=0}^{M_p} w_{\mathrm{DFE},m}^* x_{l-m} + \sum_{m=1}^{M_{fb}} f_{\mathrm{DFE},m}^* \hat{d}_{l-m}. \tag{2.44}$$

2.4.2 Optimal Weights

The optimal weights under the MMSE criterion can be found using the orthogonality principle [56]. $M_p + M_{fb} + 1$ equations are obtained, and the weights can be found using the method of undetermined coefficients described in [74, 141]. The optimal DFE tap weights are given by

$$\begin{aligned} w_{\mathrm{DFE},l} &= \frac{\mathrm{SNR}\left[(1+\mathrm{SNR})(\sigma_n^2 + M_{fb}E_i) + (M_p - M_{fb})E_i\right]}{(1+\mathrm{SNR})\left[(1+\mathrm{SNR})(\sigma_n^2 + M_{fb}E_i) + (M_p - M_{fb} + 1)E_i\right]} \\ &= C_0, \quad l = 0, \end{aligned} \tag{2.45}$$

$$\begin{aligned} w_{\mathrm{DFE},l} &= \frac{-E_i\mathrm{SNR}}{(1+\mathrm{SNR})(\sigma_n^2 + M_{fb}E_i) + (M_p - M_{fb} + 1)E_i} e^{-j\Omega l T_s} \\ &= C_1 e^{-j\Omega l T_s}, \quad l = 1, \ldots, M_{fb}, \end{aligned} \tag{2.46}$$

$$\begin{aligned} w_{\mathrm{DFE},l} &= \frac{-E_i\mathrm{SNR}}{(1+\mathrm{SNR})\left[(1+\mathrm{SNR})(\sigma_n^2 + M_{fb}E_i) + (M_p - M_{fb} + 1)E_i\right]} e^{-j\Omega l T_s} \\ &= \frac{C_1}{1+\mathrm{SNR}} e^{-j\Omega l T_s}, \quad l = M_{fb}+1, \ldots, M_p, \end{aligned} \tag{2.47}$$

$$\begin{aligned} f_{\mathrm{DFE},l} &= \frac{E_i\mathrm{SNR}}{(1+\mathrm{SNR})(\sigma_n^2 + M_{fb}E_i) + (M_p - M_{fb} + 1)E_i} e^{-j\Omega l T_s} \\ &= -C_1 e^{-j\Omega l T_s}, \quad l = 1, \ldots, M_{fb}. \end{aligned} \tag{2.48}$$

Observe that the weight of the feedback taps (2.48) is the negative of the feedforward side taps (2.46) when $l = 1, \ldots, M_{fb}$. This implies that if the data fed back is

perfect, the ISI caused by the M_{fb} previous data symbols will be completely canceled. Also note that (2.47) is a scaled $\left(\text{by } \frac{1}{1+\text{SNR}}\right)$ multiple of (2.46). This scaling value effectively removes the influence of the associated data symbols that can not be canceled by the feedback taps. For the special case of $M_p = M_{fb}$, it can be seen that if the data fed back is perfect, the ISI caused by the feedforward equalizer will be completely canceled, leaving only the symbol of interest.

Note also that when $M_p = M_{fb} = 0$ in (2.45) and (2.47), the optimal weights for the DDE are obtained, as seen in (2.28) and (2.29).

2.4.3 Minimum Mean-Squared Error

Following the discussion of Section 2.2.3, the MMSE is obtained using the optimal weights given in (2.45)-(2.48). Using these weights makes the output of the filter, $y_{\text{DFE},l}$, optimal, and this term is defined in (2.44). The optimal estimation error term is defined as $e_{\text{opt},l} = d_l - y_{\text{DFE},l}$. Based on the orthogonality principle and its corollary, the following relationships are true,

$$\mathbb{E}\left[x_{l-m}e^*_{\text{opt},l}\right] = 0, \quad m = 0, 1, \ldots, M_p, \tag{2.49}$$

$$\mathbb{E}\left[d_{l-m}e^*_{\text{opt},l}\right] = 0, \quad m = 1, \ldots, M_{fb}, \tag{2.50}$$

$$\mathbb{E}\left[y_{\text{DFE},l}e^*_{\text{opt},l}\right] = 0. \tag{2.51}$$

The DFE MMSE value is calculated assuming that the fed back decisions are

correct, $\hat{d}_l = d_l$, and found to be

$$\begin{aligned}
J_{\text{DFE,min}} &= \mathbb{E}\left[|e_{\text{opt},l}|^2\right] = \mathbb{E}\left[e_{\text{opt},l}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[\left(d_l - y_{\text{DFE},l}\right)e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l e^*_{\text{opt},l}\right] - \mathbb{E}\left[y_{\text{DFE},l}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[d_l\left(d_l - y^*_{\text{DFE},l}\right)\right] \\
&= \mathbb{E}\left[d_l d_l^*\right] - \mathbb{E}\left[d_l\left(\sum_{m=0}^{M_p} w^*_{\text{DFE},m}x_{l-m} + \sum_{m=1}^{M_{fb}} f^*_{\text{DFE},m}d_{l-m}\right)^*\right] \\
&= E_s\delta_0 - \sum_{m=0}^{M_p} w_{\text{DFE},m}\mathbb{E}\left[d_l x^*_{l-m}\right] + \sum_{m=1}^{M_{fb}} f_{\text{DFE},m}\mathbb{E}\left[d_l d^*_{l-m}\right] \\
&= E_s - \sum_{m=0}^{M_p} w_{\text{DFE},m}\mathbb{E}\left[d_l\left(d_{l-m} + i_{l-m} + n_{l-m}\right)^*\right] + E_s\sum_{m=1}^{M_{fb}} f_{\text{DFE},m}\delta_m \\
&= E_s - \sum_{m=0}^{M_p} w_{\text{DFE},m}\mathbb{E}\left[d_l d^*_{l-m}\right] \\
&= E_s - E_s\sum_{m=0}^{M_p} w_{\text{DFE},m}\delta_m \\
&= E_s - E_s w_{\text{DFE},0} = E_s\left(1 - w_{\text{DFE},0}\right) = E_s\left(1 - C_0\right). \qquad (2.52)
\end{aligned}$$

2.4.4 Signal-to-Interference-Plus-Noise Ratio

The SINR at the input to the decision device of the DFE can be found by examining (2.44) and the optimal weights (2.45)-(2.48). Again, given that the fed back decisions are correct, $\hat{d}_l = d_l$, the output of the DFE can be written as

$$\begin{aligned}
y_{\text{DFE},l} &= \sum_{m=0}^{M_p} w^*_{\text{DFE},m}\left(d_{l-m} + i_{l-m} + n_{l-m}\right) + \sum_{m=1}^{M_{fb}} f^*_{\text{DFE},m}d_{l-m} \\
&= \underbrace{w^*_{\text{DFE},0}d_l + \sum_{m=M_{fb}+1}^{M_p} w^*_{\text{DFE},m}d_{l-m}}_{\zeta_d} + \underbrace{\sum_{m=0}^{M_p} w^*_{\text{DFE},m}i_{l-m}}_{\zeta_i} + \underbrace{\sum_{m=0}^{M_p} w^*_{\text{DFE},m}n_{l-m}}_{\zeta_n}
\end{aligned} \qquad (2.53)$$

Note that the second line of (2.53) arises from the fact that $f_{\text{DFE},l} = -w_{\text{DFE},l}$, when $l = 1, \ldots, M_{fb}$. The output of the DFE is composed of the filtered signal (ζ_d), the filtered

interference (ζ_i), and the filtered noise (ζ_n). The SINR is defined in (2.34), and each term is determined as follows:

$$\begin{aligned}
\mathbb{E}\left[|\zeta_d|^2\right] &= \mathbb{E}\left[\left(w^*_{\mathrm{DFE},0}d_l + \sum_{m=M_{fb}+1}^{M_p} w^*_{\mathrm{DFE},m}d_{l-m}\right)\left(w^*_{\mathrm{DFE},0}d_l + \sum_{k=M_{fb}+1}^{M_p} w^*_{\mathrm{DFE},k}d_{l-k}\right)^*\right] \\
&= w^*_{\mathrm{DFE},0}w_{\mathrm{DFE},0}\mathbb{E}\left[d_l d_l^*\right] + \sum_{m=M_{fb}+1}^{M_p}\sum_{k=M_{fb}+1}^{M_p} w^*_{\mathrm{DFE},m}w_{\mathrm{DFE},k}\mathbb{E}\left[d_{l-m}d^*_{l-k}\right] \\
&= \left|w_{\mathrm{DFE},0}\right|^2 E_s\delta_0 + \sum_{m=M_{fb}+1}^{M_p}\sum_{k=M_{fb}+1}^{M_p} w^*_{\mathrm{DFE},m}w_{\mathrm{DFE},k}E_s\delta_{k-m} \\
&= \left(\left|w_{\mathrm{DFE},0}\right|^2 + \sum_{m=M_{fb}+1}^{M_p}\left|w_{\mathrm{DFE},m}\right|^2\right)E_s \\
&= \left(C_0^2 + \sum_{m=M_{fb}+1}^{M_p}\left|\frac{C_1}{1+\mathrm{SNR}}e^{-j\Omega l T_s}\right|^2\right)E_s \\
&= \left(C_0^2 + (M_p - M_{fb})\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2\right)E_s,
\end{aligned} \tag{2.54}$$

$$\begin{aligned}
\mathbb{E}\left[|\zeta_i|^2\right] &= \mathbb{E}\left[\left|\sum_{m=0}^{M_p} w^*_{\mathrm{DFE},m}i_{l-m}\right|^2\right] \\
&= \mathbb{E}\left[\left|w^*_{\mathrm{DFE},0}i_l + \sum_{m=1}^{M_{fb}} w^*_{\mathrm{DFE},m}i_{l-m} + \sum_{m=M_{fb}+1}^{M_p} w^*_{\mathrm{DFE},m}i_{l-m}\right|^2\right] \\
&= \mathbb{E}\left[\left|C_0\sqrt{E_i}e^{j(\Omega l T_s+\theta)} + \sum_{m=1}^{M_{fb}} C_1 e^{j\Omega m T_s}\sqrt{E_i}e^{j(\Omega(l-m)T_s+\theta)}\right.\right. \\
&\qquad \left.\left. + \sum_{m=M_{fb}+1}^{M_p}\frac{C_1}{1+\mathrm{SNR}}e^{j\Omega m T_s}\sqrt{E_i}e^{j(\Omega(l-m)T_s+\theta)}\right|^2\right] \\
&= \mathbb{E}\left[\left|\left(C_0 + M_{fb}C_1 + (M_p - M_{fb})\frac{C_1}{1+\mathrm{SNR}}\right)\sqrt{E_i}e^{j(\Omega l T_s+\theta)}\right|^2\right] \\
&= \left(C_0 + \left(M_{fb} + \frac{M_p - M_{fb}}{1+\mathrm{SNR}}\right)C_1\right)^2 E_i,
\end{aligned} \tag{2.55}$$

$$
\begin{aligned}
\mathbb{E}\left[|\zeta_n|^2\right] &= \mathbb{E}\left[\left(\sum_{m=0}^{M_p} w^*_{\mathrm{DFE},m} n_{l-m}\right)\left(\sum_{k=0}^{M_p} w^*_{\mathrm{DFE},k} n_{l-k}\right)^*\right] \\
&= \sum_{m=0}^{M_p}\sum_{k=0}^{M_p} w_{\mathrm{DFE},k} w^*_{\mathrm{DFE},m}\mathbb{E}\left[n_{l-m} n^*_{l-k}\right] \\
&= \sum_{m=0}^{M_p}\sum_{k=0}^{M_p} w_{\mathrm{DFE},k} w^*_{\mathrm{DFE},m}\sigma_n^2\delta_{k-m} \\
&= \sigma_n^2 \sum_{m=0}^{M_p} |w_{\mathrm{DFE},m}|^2 \\
&= \sigma_n^2\left(|C_0|^2 + \sum_{m=1}^{M_{fb}} \left|C_1 e^{-j\Omega m T_s}\right|^2 + \sum_{m=M_{fb}+1}^{M_p}\left|\frac{C_1}{1+\mathrm{SNR}} e^{-j\Omega m T_s}\right|^2\right) \\
&= \sigma_n^2\left(C_0^2 + \left(M_{fb} + \frac{M_p - M_{fb}}{(1+\mathrm{SNR})^2}\right) C_1^2\right). \qquad (2.56)
\end{aligned}
$$

Therefore the SINR is given by,

$$
\mathrm{SINR} = \frac{\left(C_0^2 + (M_p - M_{fb})\left(\frac{C_1}{1+\mathrm{SNR}}\right)^2\right)\mathrm{SNR}}{\left(C_0 + \left(M_{fb} + \frac{M_p - M_{fb}}{1+\mathrm{SNR}}\right) C_1\right)^2 \frac{E_j}{\sigma_n^2} + C_0^2 + \left(M_{fb} + \frac{M_p - M_{fb}}{(1+\mathrm{SNR})^2}\right) C_1^2}. \qquad (2.57)
$$

Again, the SINR for the DDE can be obtained from the (2.57) by setting $M_p = M_{fb} = 0$.

2.4.5 Autocorrelation Structure

The input to the decision-feedback equalizer is the concatenation of the received input to the equalizer and the fed back decisions, given by $\left[\mathbf{x}_l^T, \hat{\mathbf{d}}_l^T\right]^T$. The vector, $\hat{\mathbf{d}}_l$, is composed of the fed back decisions that are assumed to be correct, and is thus defined as

$$
\hat{\mathbf{d}}_l = \mathbf{d}_l \triangleq \left[d_{l-1}, \quad d_{l-2}, \quad \ldots, \quad d_{l-M_{fb}}\right]^T. \qquad (2.58)
$$

The autocorrelation matrix for the M_p+1-tap feedforward and M_{fb}-tap feedback equalizer is defined as

$$\mathbf{R}_{\text{DFE}} = \mathbb{E}\begin{bmatrix}\mathbf{x}_l\mathbf{x}_l^H & \mathbf{x}_l\mathbf{d}_l^H \\ \mathbf{d}_l\mathbf{x}_l^H & \mathbf{d}_l\mathbf{d}_l^H\end{bmatrix}$$

$$= \left[\begin{array}{ccccc|cccc} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(M_p) & 0 & 0 & \cdots & 0 \\ r_x^*(1) & r_x(0) & r_x(1) & \cdots & r_x(M_p-1) & E_s & 0 & \cdots & 0 \\ r_x^*(2) & r_x^*(1) & r_x(0) & \cdots & r_x(M_p-2) & 0 & E_s & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x^*(M_p) & r_x^*(M_p-1) & r_x^*(M_p-2) & \cdots & r_x(0) & 0 & 0 & \cdots & E_s \\ \hline 0 & E_s & 0 & \cdots & 0 & E_s & 0 & \cdots & 0 \\ 0 & 0 & E_s & \cdots & 0 & 0 & E_s & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & E_s & 0 & 0 & \cdots & E_s \end{array}\right]. \tag{2.59}$$

The autocorrelation matrix seen in (2.59) is partitioned into 4 submatrices. The matrices on the diagonal are the autocorrelation matrix of the received input to the equalizer and the autocorrelation matrix of the data symbols, respectively. The values in the upper left submatrix are given by (2.27). The cross-correlation matrix between the received input to the equalizer and the data symbols is located on the off-diagonal.

2.4.6 Eigenvalues

There is no closed form expression for determining the eigenvalues of the correlation matrix defined in (2.59). A method to bound the eigenvalues of positive-definite Toeplitz matrices can be found in [35] and its application to the correlation matrix given in (2.59) can be found in [103]. However, for the case of $M_p \geq 1$ and $M_{fb} \geq 2$, the minimum and maximum eigenvalues are found to be

$$\lambda_{\text{DFE,min}} = \frac{2E_s + \sigma_n^2 - \sqrt{4(E_s)^2 + (\sigma_n^2)^2}}{2}, \tag{2.60}$$

$$\lambda_{\text{DFE,max}} \approx E_s + (M_p+1)E_i + \sigma_n^2, \tag{2.61}$$

and the eigenvalue spread is

$$\chi(\mathbf{R}_{\text{DFE}}) = \frac{\lambda_{\text{DFE,max}}}{\lambda_{\text{DFE,min}}} = \frac{2(E_s + (M_p+1)E_i + \sigma_n^2)}{2E_s + \sigma_n^2 - \sqrt{4(E_s)^2 + (\sigma_n^2)^2}}. \tag{2.62}$$

Note that the eigenvalues given in (2.60) and (2.61) are not a function of M_{fb}.

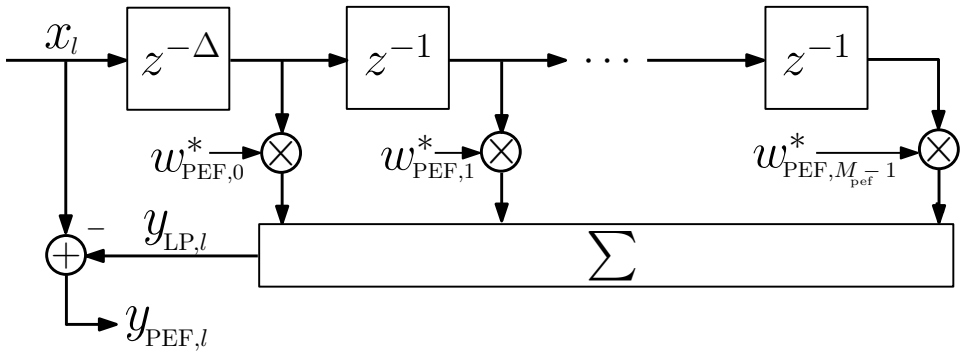


Figure 2.3: Prediction-error filter block diagram.

2.4.7 Convergence Properties

The projection of any of the eigenvectors on the optimal weight vector is nonzero. This implies that the time constant (2.20) is proportional to the minimum eigenvalue (i.e. $\tau_{\mathrm{DFE}} \simeq 1/\mu(2E_s + \sigma_n^2 - \sqrt{4(E_s)^2 + (\sigma_n^2)^2})$). Similar to the DDE, the delay in convergence can be attributed to the fact that the DFE does not have a direct reference for the interferer during adaptation and is thus forced to converge on the basis of the training data only. The feedback taps converge slower than the feedforward taps because the DFE is designed such that the interferer is canceled by the feedforward taps, while the feedback taps attempt to cancel out the signal distortion caused by the feedforward taps [74].

2.5 Prediction-Error Filter

2.5.1 Predictor Structure

The linear predictor (LP) is a structure that uses the correlation between past samples to form an estimate of the current sample [5, 56, 98]. A variant of this filter, the prediction-error filter (PEF) has the property that it removes the correlation between samples, thereby whitening the spectrum. A common example of this property is seen when determining the parameters of an autoregressive (AR) process. The prediction-error filter (assuming a sufficient filter order) of such an input provides both the AR parameters and a white output sequence that is equal to the innovations process.

This technique has also been used to remove narrowband interference in many applications [73, 97, 102, 124, 141]. The filter is able to predict the interferer due to its

narrowband properties. A block diagram of the symbol-spaced prediction-error filter is shown in Fig. 2.3. The PEF is a transversal filter with M_{pef} taps. The decorrelation delay (Δ) ensures that the signal of interest at the current sample is decorrelated from the samples in the filter when calculating the error term. Because the data is i.i.d., $\Delta = 1$ is a sufficient choice, giving the one-step predictor. As discussed in [140] this technique is also applicable for symbols received in correlated noise if the prediction distance is increased to decorrelate the correlated noise when there is sufficient separation between the correlation time of the data of interest and the noise. The linear combination of the weighted input samples, x_l, forms an estimate of the interferer, given by

$$y_{\text{LP},l} = \sum_{m=0}^{M_{\text{pef}}-1} w^*_{\text{PEF},m} x_{l-\Delta-m}, \tag{2.63}$$

where $w_{\text{PEF},m}$ are the tap weights of the predictor. The output of the PEF, $y_{\text{PEF},l}$, is defined as the subtraction of the estimate of the interference given in (2.63) from the current input sample,

$$y_{\text{PEF},l} = x_l - y_{\text{LP},l} = x_l - \sum_{m=0}^{M_{\text{pef}}-1} w^*_{\text{PEF},m} x_{l-\Delta-m}. \tag{2.64}$$

Note that $y_{\text{PEF},l}$ is also the error term of the structure. This implies that the PEF is in fact a blind algorithm. It does not require any training symbols when calculating the error term. This allows the interference to be removed without a specific reference signal.

2.5.2 Predictor Optimal Weights

The optimal tap weights can be found in a way similar to those for the equalizer above using the method of undetermined coefficients [74, 141]. Using the orthogonality principle, M_{pef} equations are obtained and the weights of the PEF are given by

$$w_{\text{PEF},l} = Ke^{-j\Omega(l+\Delta)T_s}, \quad l = 0, \ldots, M_{\text{pef}} - 1, \tag{2.65}$$

where K is equal to

$$K = \frac{E_i}{E_s + \sigma_n^2 + M_{\text{pef}} E_i}. \tag{2.66}$$

Note that (2.65) can be rewritten in vector form as

$$\mathbf{w}_{\text{PEF}} = [1, \underbrace{0, \ldots, 0}_{\Delta-1}, -Ke^{-j\Omega\Delta T_s}, \ldots, -Ke^{-j\Omega(M_{\text{pef}}-1+\Delta)T_s}]. \tag{2.67}$$

This vector of length $M_{\text{pef}} + \Delta$ can be applied directly to a vector of input samples to provide that output of the PEF. Note that only the last M_{pef} weights are adjustable when employed in an adaptive algorithm.

The scenario of interest for this work is when the interference power is much larger than both the signal power and the noise power. Therefore the SIR and the noise-to-interference ratio (NIR) can be assumed to be very small (i.e. SIR $\ll 0$ dB, NIR $\ll 0$ dB [74]) and K can be approximated as

$$K \cong \frac{1}{M_{\text{pef}}}. \tag{2.68}$$

2.5.3 Minimum-Mean Square Error

Following the discussion of Section 2.2.3, the MMSE is obtained using the optimal weights given in (2.65). Using these weights makes the output of the filter, $y_{\text{PEF},l}$, optimal, and this term is defined in (2.64). The optimal estimation error term is defined as $e_{\text{opt},l} = x_l - y_{\text{LP},l}$. Based on the orthogonality principle and its corollary, the following relationships are true,

$$\mathbb{E}\left[x_{l-\Delta-m} e^*_{\text{opt},l}\right] = 0, \quad m = 0, 1, \ldots, M_{\text{pef}} - 1, \tag{2.69}$$

$$\mathbb{E}\left[y_{\text{PEF},l} e^*_{\text{opt},l}\right] = 0. \tag{2.70}$$

The PEF MMSE value is found to be

$$\begin{aligned}
J_{\text{PEF,min}} &= \mathbb{E}\left[\left|e_{\text{opt},l}\right|^2\right] = \mathbb{E}\left[e_{\text{opt},l}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[\left(x_l - \sum_{m=0}^{M_{\text{pef}}-1} w^*_{\text{PEF},m}x_{l-\Delta-m}\right)e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[x_l e^*_{\text{opt},l}\right] - \sum_{m=0}^{M_{\text{pef}}-1} w^*_{\text{PEF},m}\mathbb{E}\left[x_{l-\Delta-m}e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[x_l e^*_{\text{opt},l}\right] \\
&= \mathbb{E}\left[x_l\left(x_l - \sum_{m=0}^{M_{\text{pef}}-1} w^*_{\text{PEF},m}x_{l-\Delta-m}\right)^*\right] \\
&= \mathbb{E}\left[x_l x^*_l\right] - \sum_{m=0}^{M_{\text{pef}}-1} w_{\text{PEF},m}\mathbb{E}\left[x_l x^*_{l-\Delta-m}\right] \\
&= r_x(0) - \sum_{m=0}^{M_{\text{pef}}-1} w_{\text{PEF},m}r_x(m+\Delta) \\
&= E_s + \sigma_n^2 + E_i - \sum_{m=0}^{M_{\text{pef}}-1} Ke^{-j\Omega(m+\Delta)T_s}\left[\left(E_s+\sigma_n^2\right)\delta_{m+\Delta} + E_i e^{j\Omega(m+\Delta)T_s}\right] \\
&= E_s + \sigma_n^2 + E_i - \sum_{m=0}^{M_{\text{pef}}-1} KE_i \\
&= E_s + \sigma_n^2 + \left(1 - M_{\text{pef}}K\right)E_i. \qquad (2.71)
\end{aligned}$$

The MMSE in (2.71) can be approximated using the approximation in (2.68),

$$J_{\text{PEF,min}} \approx E_s + \sigma_n^2. \qquad (2.72)$$

This demonstrates that the interference is approximately removed at the output of the PEF.

2.5.4 Sensitivity to Additive Noise

The PEF has been shown to be sensitive to additive noise when used for channel estimation [1, 116]. An algorithm was proposed in [46] to provide adaptive estimation of unbiased linear predictors with the goal of obtaining a consistent estimate of an ISI single-input multiple-output (SIMO) channel. To examine the effect of the additive noise on the PEF for this problem, we are interested in the noise free predictor weights, given

by

$$\tilde{\mathbf{w}}_{\text{PEF}} = [1, \underbrace{0, \ldots, 0}_{\Delta-1}, -\tilde{K}e^{-j\Omega\Delta T_s}, \ldots, -\tilde{K}e^{-j\Omega(M_{\text{pef}}-1+\Delta)T_s}], \tag{2.73}$$

where $\tilde{K}$ is equal to

$$\tilde{K} = \frac{E_i}{E_s + M_{\text{pef}}E_i}. \tag{2.74}$$

The noise free case given in (2.73) is compared with the biased predictor weights given in (2.67), looking at the norm of the difference (bias),

$$\|\tilde{\mathbf{w}}_{\text{PEF}} - \mathbf{w}_{\text{PEF}}\| = \frac{M_{\text{pef}}\sigma_n^2 E_i}{(E_s + \sigma_n^2 + M_{\text{pef}}E_i)(E_s + M_{\text{pef}}E_i)}. \tag{2.75}$$

This bias can be approximated using the assumptions that the SIR and the NIR are very small to give

$$\|\tilde{\mathbf{w}}_{\text{PEF}} - \mathbf{w}_{\text{PEF}}\| \approx \frac{\left(\sigma_n^2/E_i\right)}{\sqrt{M_{\text{pef}}}} = \frac{\text{NIR}}{\sqrt{M_{\text{pef}}}}. \tag{2.76}$$

The value in (2.76) is quite small due to the assumption that the NIR is small. Thus, in this work, the bias in the linear predictor does not substantially affect the system's performance.

2.5.5 Autocorrelation Structure

The $M_{\text{pef}} \times M_{\text{pef}}$ input autocorrelation matrix for the PEF is defined as

$$\begin{aligned}\mathbf{R}_{\text{PEF,i}} &= \mathbb{E}\left[\mathbf{x}_l \mathbf{x}_l^H\right] \\ &= \begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(M_{\text{pef}}-1) \\ r_x^*(1) & r_x(0) & r_x(1) & \cdots & r_x(M_{\text{pef}}-2) \\ r_x^*(2) & r_x^*(1) & r_x(0) & \cdots & r_x(M_{\text{pef}}-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x^*(M_{\text{pef}}-1) & r_x^*(M_{\text{pef}}-2) & r_x^*(M_{\text{pef}}-3) & \cdots & r_x(0) \end{bmatrix},\end{aligned} \tag{2.77}$$

where the components of the matrix are given by (2.27). Note that this matrix is the same as seen for the DDE in (2.39). The difference between the two structure arises from using different cross-correlation vectors, $\mathbf{p}$. For the case of the DDE, the desired signal is d_l, however, for the PEF the desired signal is x_l as noted in Section 2.5.1.

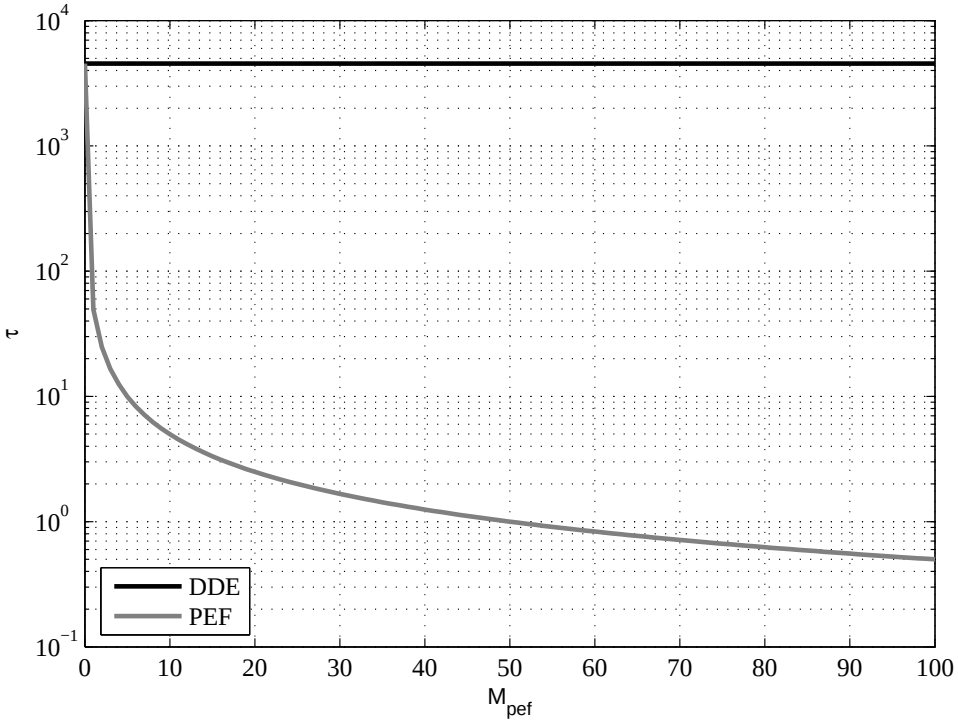


Figure 2.4: Approximate time constants for the DDE and the PEF for SNR = 10 dB, SIR = −20 dB, and $\mu = 0.0001$.

2.5.6 Eigenvalues

The eigenvalues for the correlation matrix given by (2.27) and (2.77), can be found [102, 110, 127] to be equal to

$$\lambda_{\text{PEF}} = \begin{cases} E_s + \sigma_n^2 + M_{\text{pef}} E_i, & \text{order 1,} \\ E_s + \sigma_n^2, & \text{order } M_{\text{pef}} - 1. \end{cases} \tag{2.78}$$

The eigenvalue spread is defined [56] as

$$\chi(\mathbf{R}_{\text{PEF,i}}) = \frac{\lambda_{\text{PEF,max}}}{\lambda_{\text{PEF,min}}} = 1 + \frac{M_{\text{pef}} E_i}{E_s + \sigma_n^2}. \tag{2.79}$$

2.5.7 Convergence Properties

In this case the $M_{\text{pef}} - 1$ eigenvectors corresponding to the minimum eigenvalues are orthogonal to the optimal weight vector, hence these eigenvalues do not affect the convergence [110]. Thus the time constant is dependent only upon the maximum eigenvalue (i.e. $\tau_{\text{PEF}} \simeq 1/2\mu(E_s + \sigma_n^2 + M_{\text{pef}} E_i)$). This demonstrates the improved convergence of the PEF as compared to the DDE. The approximate time constants for each

structure are plotted in Figure 2.4. The time constant for the DDE is not a function of the number equalizer taps and is thus constant. The time constant for the PEF is seen to decrease as the number of prediction taps increases, and for all cases is less than the time constant of the DDE.

2.5.8 Output Autocorrelation

The whitening property of the PEF can be seen more clearly through the autocorrelation function of the output of the PEF, which is derived to be,

$$\begin{aligned} r_{\mathrm{PEF,o}}(m) &= \mathbb{E}\left[y_{\mathrm{PEF},l} y^*_{\mathrm{PEF},l-m}\right] \\ &= (1-KM_{\mathrm{pef}})^2 E_i e^{j\Omega m T_s} \\ &\quad + (E_s+\sigma_n^2)\begin{cases} (1+K^2 M_{\mathrm{pef}}), & m=0, \\ K^2(M_{\mathrm{pef}}-|m|)e^{j\Omega m T_s}, & |m|=1,\cdots,\Delta-1, \\ K(K(M_{\mathrm{pef}}-|m|)-1)e^{j\Omega m T_s}, & |m|=\Delta,\cdots,M_{\mathrm{pef}}-1, \\ -Ke^{j\Omega m T_s}, & |m|=M_{\mathrm{pef}},\cdots,M_{\mathrm{pef}}+\Delta-1. \end{cases} \end{aligned} \tag{2.80}$$

An approximation for the output autocorrelation function in (2.80) can be found using the assumption given in (2.68),

$$r_{\mathrm{PEF,o}}(m) \cong (E_s+\sigma_n^2)\begin{cases} 1+\frac{1}{M_{\mathrm{pef}}}, & m=0, \\ (\frac{1}{M_{\mathrm{pef}}}-\frac{|m|}{M_{\mathrm{pef}}^2})e^{j\Omega m T_s}, & |m|=1,\cdots,\Delta-1, \\ -\frac{|m|}{M_{\mathrm{pef}}}e^{j\Omega m T_s}, & |m|=\Delta,\cdots,M_{\mathrm{pef}}-1, \\ -\frac{1}{M_{\mathrm{pef}}}e^{j\Omega m T_s}, & |m|=M_{\mathrm{pef}},\cdots,M_{\mathrm{pef}}+\Delta-1. \end{cases} \tag{2.81}$$

Finally, letting the filter order increase toward infinity shows that the output spectrum is approximately white,

$$\lim_{M_{\mathrm{pef}}\to\infty} r_{\mathrm{PEF,o}}(m) \cong (E_s+\sigma_n^2)\delta_m. \tag{2.82}$$

2.5.9 Eigenvalue Spread

The effect of the PEF is that the interference is removed, which then results in the reduction of the eigenvalue spread. This can be seen in Fig. 2.5 for SNR = 10 dB, SIR = -20 dB, and $\Omega = \frac{\pi}{6}$. Also in the plot is the eigenvalue spread of the received

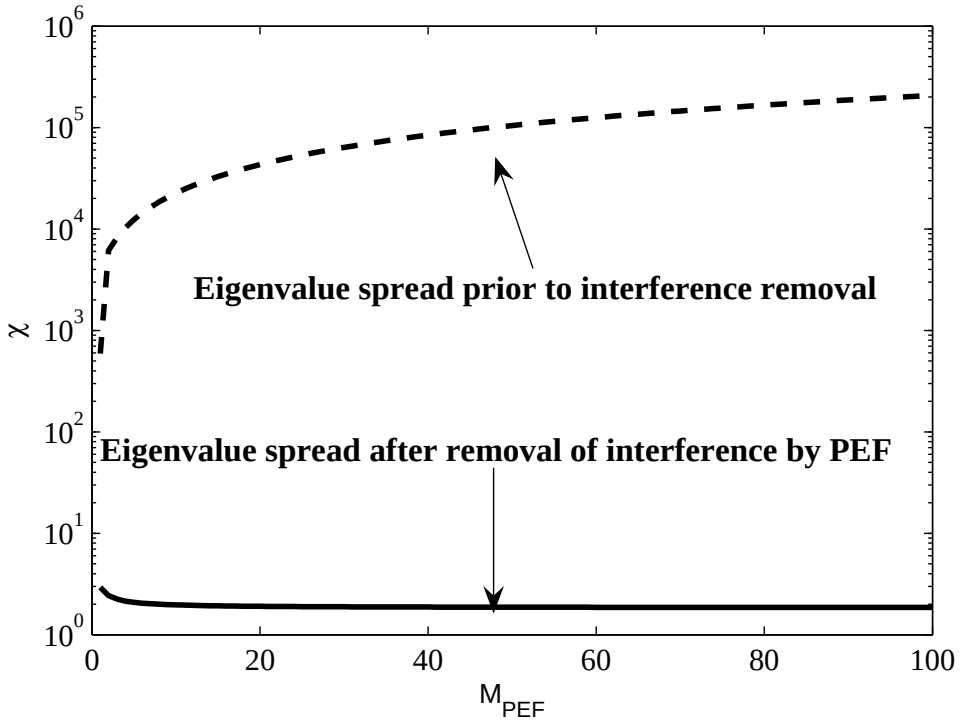


Figure 2.5: Eigenvalue spread of input to DFE-only and output of PEF for SNR = 10 dB, SIR = −20 dB, and $\Omega = \pi/6$.

data given by (2.79). Note that it is assumed that $M_{\text{pef}} = M_p$. It is clearly seen that the spread has been reduced, and the modes of this input to the LMS DFE will converge in similar amounts of time.

2.5.10 Summary

Chapter 2, in part, is a reprint of material as it appears in A. Batra, T. Ikuma, J. R. Zeidler, A. A. Beex, J. G. Proakis, "Mitigation of Unknown Narrowband Interference Using Instantaneous Error Updates," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, pp. 115–119, Nov. 2004, A. Batra, J. R. Zeidler, and A. A. Beex, "Mitigation of Narrowband Interference Using Adaptive Equalizers," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sep. 2006, and A. Batra, J. R. Zeidler, and A. A. Beex, "A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Decision-Feedback Adaptive Equalizers in the Presence of Severe Narrowband Interference," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article

ID 390102, 13 pages, 2008. doi:10.1155/2008/390102. The dissertation author was the primary investigator and author of these papers.

3 Data-Aided Initialization

3.1 Introduction

As discussed in Chapter 2, a technique commonly used to mitigate narrowband interference and multipath distortion is equalization. These equalizers are implemented via time varying adaptive filters that also possess the ability to locate and track any interferers. The performances of such filters are judged based on the approximation to the time-invariant (TI) Wiener filter of the same structure. It has been shown [11, 13, 104] that the normalized least-mean squares (NLMS) algorithm can possibly provide better performance than this corresponding Wiener filter as a result of information in the instantaneous error signal that is used to update the filter. This is especially true when narrowband interference is present in the received data.

In particular, the BER performance benefit gained by adapting in the non-Wiener region is investigated. The equalizers are evaluated, processing short bursts of an interference-contaminated and multipath-distorted communication signal. The channel considered here, is assumed to have a direct path and one resolvable multipath component delayed by one symbol period. The channel impulse response is thus given by

$$h_l = \begin{cases} 1, & \text{if } l = 0, \\ \alpha, & \text{if } l = L_o, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

where α is the multipath coefficient. The transmitter and the receiver are assumed not to have a priori knowledge of the channel. The BER performance of the adaptive linear equalizer is compared to that of the adaptive decision-feedback equalizer (DFE) to assess the performance benefit of the non-Wiener adaptation over the corresponding nonlinear

(optimal) system. Another result of this investigation is that the narrowband interference causes the adaptive equalizers to possess long convergence times. To improve the convergence rate of the equalizer, a new data-aided initialization procedure is proposed that initializes the adaptive weights with estimates of the Wiener weights.

Finally, two techniques are evaluated to initialize the Wiener weights of the adaptive DFE in order to improve the convergence. These weights are obtained from averages of the autocorrelation matrix and the cross-correlation vector derived from the received samples and the training data. These two techniques are compared in terms of relative BER performance and complexity in an environment corrupted solely by narrowband interference (i.e. no multipath, $\alpha = 0$). This examination considers the one-sided DFE with an $M+1$-tap feedforward filter, an M-tap feedback filter, and the input at symbol rate,

3.2 Non-Wiener Results

In the equalization context, the NLMS-DDE is found capable of suppressing the interference better than the equivalent optimal TI linear equalizer. The added performance benefit is a result of the nonlinear instantaneous characteristics of the NLMS algorithm [13] demonstrated in Section 2.2.4. The instantaneous error signal produces dynamic weight updating that effectively causes the filter to track the interference. Fig. 3.1 demonstrates the effect of such nonlinear dynamics [104]. The MSEs of the TI Wiener solutions and the simulated NLMS-DDE using only training data are computed for various step-size parameters and for the different cases of multipath coefficient. Other simulation parameters for this case are described in Section 3.3.1. It is readily observed that NLMS-DDE performs better in an MSE sense than the TI Wiener solution over a wide range of step-sizes under each of the multipath scenarios and the MSE improvement is less for stronger multipath.

3.3 Convergence Simulation

3.3.1 Simulation Setup

In all experiments, the input signal to the equalizer (i.e., the output of the matched filter (MF)) has the same statistical characteristics. The communication signal is a quadrature phase shift keyed (QPSK) signal in which the independent in-phase and

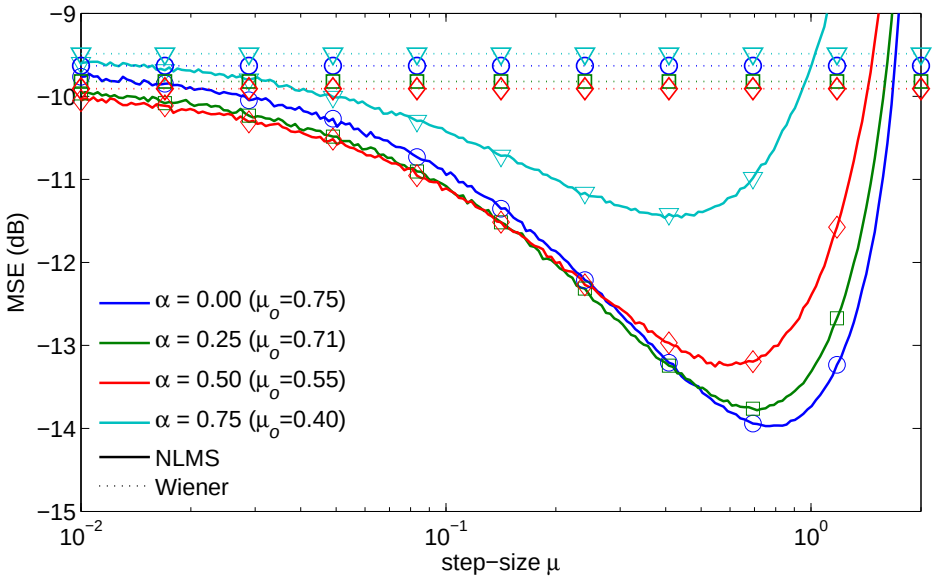


Figure 3.1: NLMS-DDE MSE versus step-size using all training data. Also plotted are the equivalent Wiener MSEs (dotted lines). The optimal step-size for each multipath value is noted in parenthesis.

quadrature components take values of $+1/\sqrt{2}$ and $-1/\sqrt{2}$ with equal probability. This QPSK signal is up-sampled by L_o (=5) and shaped with a square-root raised cosine pulse with roll-off factor of $\beta = 0.25$ and time delay of 3 symbols. The multipath coefficient of the channel is assumed to take on a value from the discrete set: $\alpha = \{0, 0.25, 0.5, 0.75\}$. The complex sinusoidal interference is located at frequency, $\Omega_i = \pi/30$, the SIR = -20 dB, and SNR = 20 dB.

Similar to the received signal setup, the equalizer configurations, other than step-size, are kept the same for all simulations. The input tap configuration is fixed for each filter structure. The DDE is configured so that $M_p = M_f = 3$. The DFE is designed in a manner where the $L_o M_p$ samples associated with the postcursor symbols in the DDE are instead fed back with the M_{fb} most recently decided symbols. Thus, the DFE is configured so that $M_f = M_{fb} = 3$, and $M_p = 0$. The total number of taps for each system is then given by, $M_{\text{DDE}} = 35$ and $M_{\text{DFE}} = 23$. Note that the DDE and DFE use essentially the same number of samples of the MF output, i.e. the same information, while the DFE uses a smaller number of taps. The weights for all filters are initialized to zero, and the input taps are soft-initialized by pre-filling with data, unless noted otherwise.

Finally, a non-adaptive linear equalizer is also introduced as a third equalizer to show the advantage of nonlinear large-μ adaptation. The linear equalizer has the same structure as the DDE but without the adaptation mechanism. The same number of taps is used as the DDE, and the weights are initialized by estimating the Wiener solution over the training sequence. Hereafter, we refer to this linear equalizer as the Wiener Equalizer. Note that in Figure 3.1 the Wiener Equalizer is approached by the NLMS-DDE with small-μ adaptation.

3.3.2 NLMS-DDE Convergence Analyses

To apply an adaptive equalizer in a realistic communication system, fast convergence is essential. Most critically, the training must be completed during the known preamble sequence so that the subsequent decision-directed or decision-feedback mode does not break down due to poor initial BER. Moreover, at the end of the training interval the BER must have been reduced to an acceptable level.

The convergence rates of the NLMS-DDE operating only in the training mode for different multipath coefficients are observed in Figure 3.2 which shows the resulting MSE learning curves and sliding-windowed BERs. In all cases, the NLMS-DDE step-size parameter is set to unity (which corresponds to fastest convergence). The observed convergence rates are unacceptably slow. Even in the multipath-free environment ($\alpha = 0$) the NLMS-DDE requires 650 symbols to attain a BER of 1×10^{-2}. When there is a strong second ray ($\alpha = 0.75$) the NLMS-DDE does not even reach the Wiener performance after $20,000$ symbols. The channel condition severely affects the convergence properties of the NLMS-DDE.

On the other hand, once the NLMS-DDE reaches steady-state, it produces a significant BER performance advantage over the Wiener-based non-adaptive equalizer. The MSE difference of a mere 3 dB boosts the BER by more than an order of magnitude (Wiener $\sim 10^{-3}$, NLMS $< 10^{-4}$).

The observed slow convergence in Figure 3.2 indicates that the training interval must be extremely long in order to obtain lower BER for the cases of $\alpha = 0.5$ and $\alpha = 0.75$. Figure 3.3 illustrates the NLMS-DDE transient behavior for various durations of the training interval for the $\alpha = 0.5$ case. The equalizer breaks down with only 1,000 training symbols. This result corresponds with the notion that an initial BER below $\sim 10^{-1}$ or so is required at the time of the switch for the decision-directed mode to work

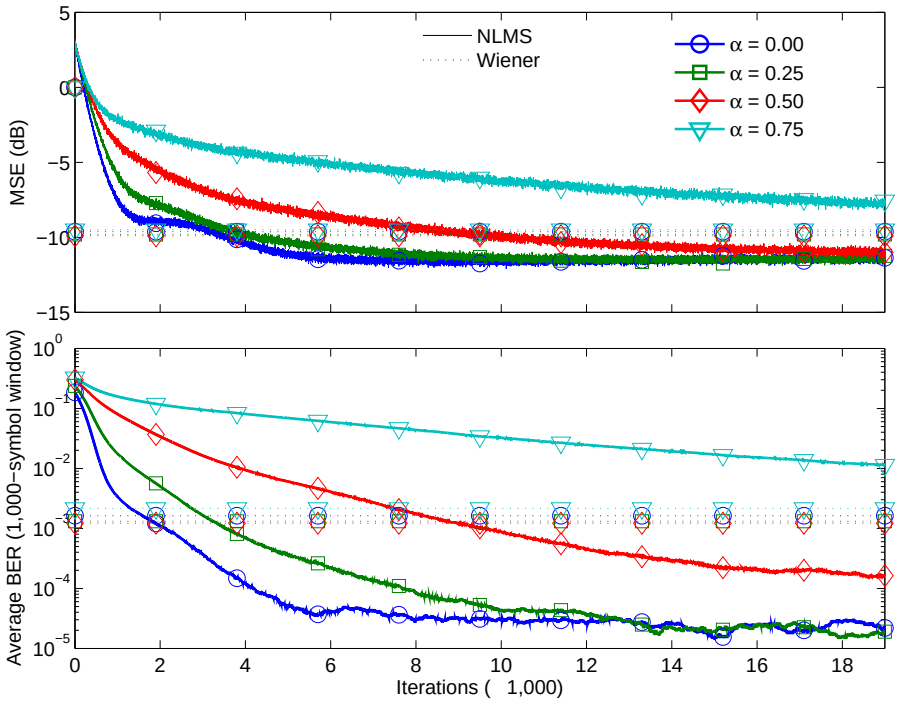


Figure 3.2: Performance of NLMS-DDE exclusively in training mode. A 1,000-symbol window is used to estimate the BER dynamics.

properly [5]. Consequently, poor performance is obtained when the number of training symbols is not sufficient to allow convergence to an MSE value that corresponds to a BER of 10^{-1} or less. Note that the BER performance advantage of the NLMS-DDE over the TI Wiener equalizer is less in the decision-directed mode. Also, the training length affects the NLMS-DDE steady-state BER, as a slightly higher BER obtained for $N_{\text{tr}} = 3,000$ as compared to the case of $N_{\text{tr}} = 7,500$.

3.3.3 Data-Aided Initialization

Despite the slow transient behavior, the BER performance gain obtained with large-μ NLMS adaptation is promising provided that a technique can be developed to accelerate the convergence so that a BER of 10^{-2} or less can be achieved during the training interval. One possible remedy to fight the slow transient is better initialization of the NLMS weights. In the previous section the weights were initialized to zero and allowed to adapt stochastically over the training interval. It is proposed here that the Wiener weights be estimated from the samples received over the training symbols (where

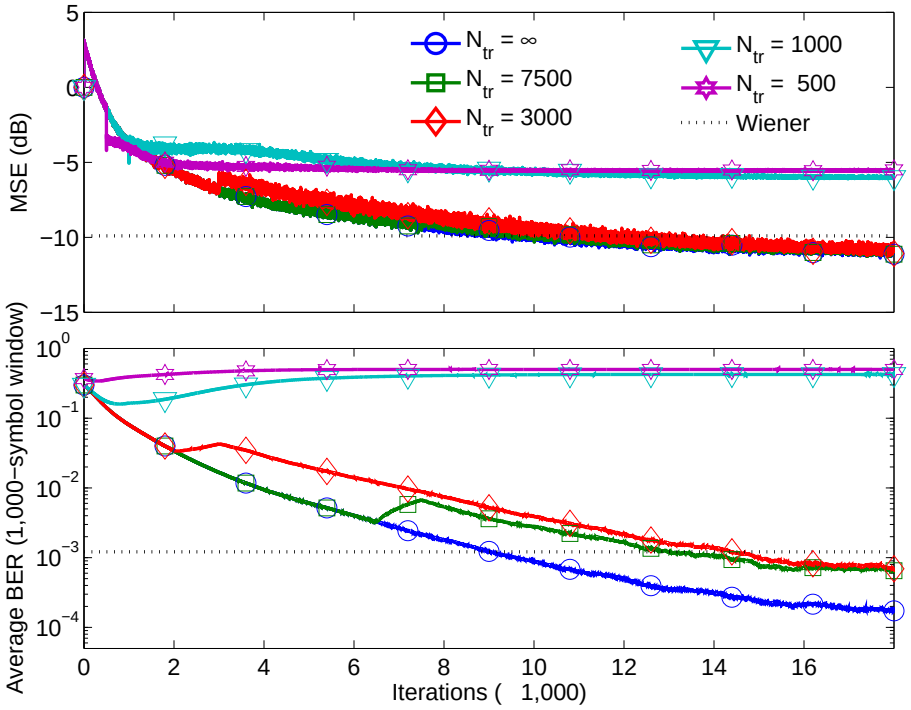


Figure 3.3: Performance of NLMS-DDE for various training durations, $\alpha = 0.5$. A 1,000-symbol window is used to estimate the BER dynamics.

the desired signal is known) and then set to the estimated Wiener weight values. This approach is referred to as data-aided initialization (DAI) of the weights. Upon obtaining these initial weights, the adaptive equalizers operate immediately in decision-directed mode.

Figure 3.4 shows the BER performance of the DAI-NLMS-DDE using different durations of training data. Based on Wiener weights estimated from only 250 training symbols the NLMS-DDE performs at a BER $\sim 10^{-4}$. Dotted lines in the figure correspond to keeping the estimated Wiener weights fixed. The equalizer breaks down when only 50 symbols are used to compute the DAI weights. Note that, similar to what is seen in Figure 3.3, the training length affects the performance but the required number of training symbols is reduced by approximately two orders of magnitude. For the adaptive case of $N_{\text{tr}} = 100$, the performance is worse than that of the non-adaptive scenario with $N_{\text{tr}} = 500$.

Clearly, data-aided initialization eliminates the need for the NLMS-DDE to have a long training period, and also provides superior BER performance over a non-adaptive equalizer in which those initial weights are kept fixed. This DAI training method is

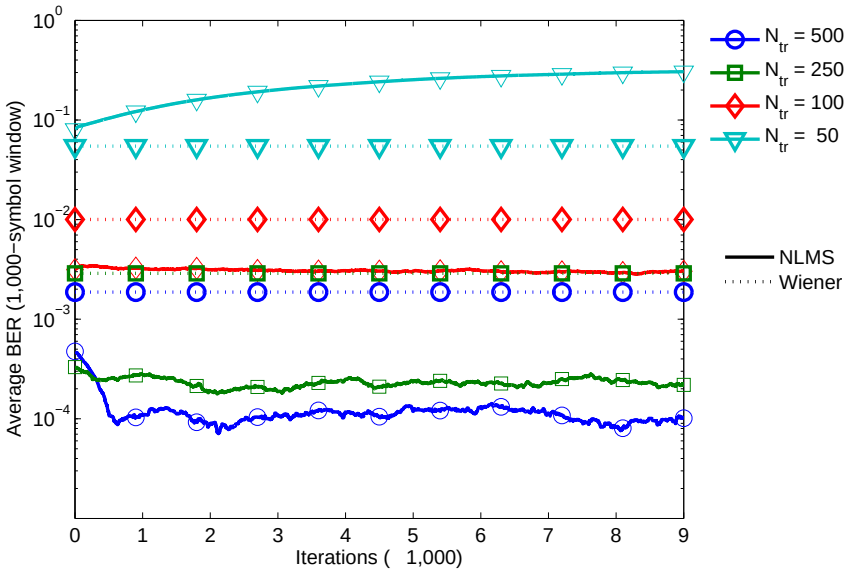


Figure 3.4: BER performance of NLMS-DDE using data-aided initialization, $\alpha = 0.5, \mu = 1$. A 1,000-symbol window is used to estimate the BER dynamics.

used for both the NLMS-DDE and the NLMS-DFE. Also, the optimal step-sizes found in Figure 3.1 are used. Note that, at the start of the decision-directed mode of the DFE, the feedback taps are filled with training symbols.

3.3.4 Equalizer Performance Analysis

The equalizer performance is assessed under the following scenario. The received signal is composed of $5,000$-symbol long QPSK bursts in the interference dominated multipath environment as defined in Section 3.3.1. The first 250 symbols of the QPSK burst are assumed to be the known preamble training symbols. The three equalizer configurations (two adaptive and one non-adaptive) are compared in terms of BER performance over the non-training symbols. Table 3.1 lists step-size parameters for all experiments. The NLMS-DDE step-sizes are chosen according to those found to be optimal in Figure 3.1. As discussed in Chapter 2, the DFE structure does not benefit from NLMS nonlinear effects and therefore a small step-size is used in the data-aided initialization.

The first set of results, shown in Figure 3.5, is the I/Q plot for each equalizer based on a single observation given the exact same received signal. A multipath coefficient $\alpha = 0.5$ is used for this illustration. The associated BER is also displayed for each plot. While both DAI-NLMS-DDE and DAI-NLMS-DFE show zero bit errors (which

Table 3.1: Step-size parameter configuration for different multipath conditions.

α	0	0.25	0.5	0.75
NLMS-DDE	0.75	0.71	0.55	0.40
NLMS-DFE	0.01	0.01	0.01	0.01

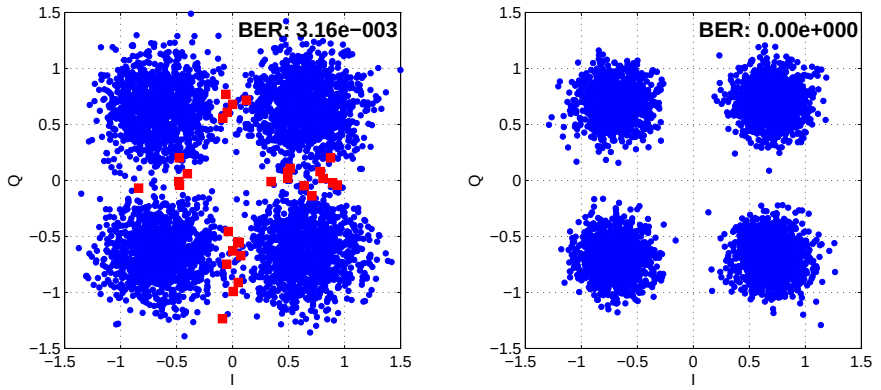
are indicated by squares in Figure 3.5), the tightness of the symbol clusters illustrates the relative BER performance of each equalizer. Clearly NLMS-DFE shows the most compact clusters among the three. At the same time, the performance improvement attained by exploiting the non-Wiener NLMS behavior is also visible when Wiener and NLMS-DDE results are contrasted.

Table 3.2 contains the second set of results for which the experiment is repeated 500 times for each receiver configuration and the observed BER values are averaged. Zero bit errors are recorded for several configurations because the experimental BER values are lower than the experimental BER resolution (anything less than 1 error in $4,750 \times 2 \times 500 = 4,750,000$ bits is reported as 0 BER).

Table 3.2: Mean BER for all structures and all α values.

α	0	0.25	0.5	0.75
Wiener Equalizer	3.81e−3	3.13e−3	2.90e−3	4.64e−3
DAI-NLMS-DDE	8.47e−6	7.29e−6	2.59e−5	4.88e−4
DAI-NLMS-DFE	0	0	0	0

With the static interference and resolvable multipath channel, all DFE-based equalizers perform better than the equivalent DDE-based structures, even when using a large step-size in the latter. However, utilization of the nonlinear NLMS behavior boosts the NLMS-DDE performance by a significant margin (by as much as three orders of magnitude) when compared to the Wiener Equalizer (or equivalently the small-μ NLMS-DDE). With the optimal large step-size selection, the NLMS-DDE performance approaches the NLMS-DFE. Note that the DFE essentially uses the previous values of the desired signal for feedback, whereas the NLMS-DDE only utilizes the corresponding



(a) Wiener equalizer.

(b) Data-aided initialized NLMS-DDE.

(c) Data-aided initialized NLMS-DFE.

Figure 3.5: I/Q plots and associated BERs for all receiver structures, $\alpha = 0.5$, $N_{\text{tr}} = 250$.

input signal.

3.4 Implementation of Data-Aided Initialization

In this portion of the chapter, under consideration is the one-sided DFE composed of an $M+1$-tap feedforward filter, an M-tap feedback filter, and input at symbol rate. That is $M = M_p = M_{fb}$ and $M_f = 0$.

3.4.1 Data-Based Averages

The autocorrelation matrix estimate, based on the received data, is

$$\hat{\mathbf{R}} = \frac{1}{N_{\text{tr}}} \sum_{k=1}^{N_{\text{tr}}} \mathbf{u}_k \mathbf{u}_k^H, \tag{3.2}$$

where N_{tr} is the number of training symbols. The received vector is defined as,

$$\mathbf{u}_k = \begin{bmatrix} x_k, & \ldots, & x_{k-M}, & \hat{d}_{k-1}, & \ldots, & \hat{d}_{k-M} \end{bmatrix}^T. \tag{3.3}$$

Note that because there are training symbols available, $\hat{d}_k = d_k$. The cross-correlation estimate is given by

$$\hat{\mathbf{p}} = \frac{1}{N_{\text{tr}}} \sum_{k=1}^{N_{\text{tr}}} \mathbf{u}_k d_k^*. \tag{3.4}$$

3.5 Direct Matrix Inversion

3.5.1 Wiener Filter

The Wiener solution filter weights can be estimated from the estimated correlation matrices by implicitly solving

$$\hat{\mathbf{R}} \mathbf{w}_{\text{DMI}} = \hat{\mathbf{p}}. \tag{3.5}$$

If the inverse is found using Gaussian elimination, $\mathcal{O}((2M+1)^3)$ multiplications for an $(2M+1)\times(2M+1)$ matrix are required. Instead a method that requires less complexity is evaluated here. An expression for the inverse can be obtained more efficiently using direct formulas, such as the Levinson algorithm [72], the Gohberg-Semencul formula [48, 122], Schur's complement [87], the matrix inversion lemma [56], and the Toeplitz structure of the autocorrelation matrix. It is assumed that $\hat{\mathbf{p}}$ is given by its theoretical value,

$$\mathbf{p} = \begin{bmatrix} E_s, & 0, & \cdots, & 0 \end{bmatrix}^T, \tag{3.6}$$

based on the fact that the components of (2.25) are uncorrelated. Note that the signal power is assumed normalized to unity.

3.5.2 Complexity for Toeplitz Matrices

A simple procedure has been proposed for efficiently finding the inverse of an $(M+1)\times(M+1)$ Toeplitz matrix [50, and references therein] due to its special structure. The first step uses the Levinson algorithm [72] to obtain an estimate for the autoregressive (AR) parameters. This technique requires M^2+2M+1 complex multiplications [50]. These AR parameters are then used with the Gohberg-Semencul formula [48,122] to obtain the inverse of the matrix. This final step uses $M^2/2+3M/2$ complex multiplications. In all, $3M^2/2 + 7M/2 + 1$ complex multiplies are necessary to write down the inverse of an $(M + 1) \times (M + 1)$ Toeplitz matrix. The complexity for finding the inverse of the DFE correlation matrix is found in the next section.

3.5.3 Complexity for the DFE

The theoretical autocorrelation matrix can be partitioned as follows,

$$\mathbf{R} = \mathbb{E}\left[\mathbf{u}_k\mathbf{u}_k^H\right] = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{Q}_{dx}^H \\ \mathbf{Q}_{dx} & E_s\mathbf{I}_M \end{bmatrix}, \tag{3.7}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\mathbf{R}_{xx} = \mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^H\right]$ is the autocorrelation of the feedforward section, and $\mathbf{Q}_{dx} = \mathbb{E}\left[\mathbf{d}_k\mathbf{x}_k^H\right]$ is the cross-correlation of the feedforward section (received signal, $\mathbf{x}_k$) and the feedback section (fed back training symbols, $\mathbf{d}_k$) and is given by

$$\mathbf{Q}_{dx} = \begin{bmatrix} \mathbf{0}_{M,1} & E_s\mathbf{I}_M \end{bmatrix}, \tag{3.8}$$

where $\mathbf{0}_{M,1}$ is the $M \times 1$ zero vector. A technique using Schur's complement [87, pp. 264-265] can be used to invert $\mathbf{R}$,

$$\begin{aligned} \mathbf{R}^{-1} &= \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{Q}_{dx}^H \\ \mathbf{Q}_{dx} & \mathbf{I_M} \end{bmatrix}^{-1}, \\ &= \begin{bmatrix} \mathbf{S_C^{-1}} & -\mathbf{S_C^{-1}}\mathbf{Q}_{dx}^H \\ -\mathbf{Q}_{dx}\mathbf{S_C^{-1}} & \mathbf{I}_M + \mathbf{Q}_{dx}\mathbf{S_C^{-1}}\mathbf{Q}_{dx}^H \end{bmatrix}, \end{aligned} \tag{3.9}$$

where $\mathbf{S_C}$ is Schur's complement and is given by

$$\mathbf{S_C} = \mathbf{R}_{xx} - \mathbf{Q}_{dx}^H\mathbf{Q}_{dx}. \tag{3.10}$$

Using the matrix inversion lemma [56],

$$\left[\mathbf{R}_{xx} - \mathbf{Q}_{dx}^H \mathbf{Q}_{dx}\right]^{-1} = \mathbf{R}_{xx}^{-1} - \mathbf{R}_{xx}^{-1}\mathbf{Q}_{dx}^H \left(\mathbf{Q}_{dx}\mathbf{R}_{xx}^{-1}\mathbf{Q}_{dx}^H - \mathbf{I}_M\right)^{-1} \mathbf{Q}_{dx}\mathbf{R}_{xx}^{-1}. \tag{3.11}$$

Now we are interested only in finding the inverse of Schur's complement which is an $(M+1)\times(M+1)$ matrix. The reduction in complexity for this method arises from the assumption of knowledge of both $\mathbf{Q}_{dx}$ and $\mathbf{I}_M$ of (3.7). Note that this assumption will result in a degradation in the approximated Wiener filter when there is not an adequate number of training symbols. Conversely, the assumption holds for a large number of training symbols because the estimated correlation matrix will be closer to the theoretical matrix.

The algorithm requires finding the inverse of Schur's complement and it is found through the following steps (ignoring multiplications by 1):

- Let $\mathbf{Z} = \mathbf{R}_{xx}^{-1}$. The resulting matrix ($\mathbf{Z}$) happens to be Toeplitz and is found using the procedure discussed in the previous section, thus requiring $3M^2/2 + 7M/2 + 1$ complex multiplications. Note that it is not always the case that the inverse of a Toeplitz matrix is Toeplitz [52].

- Let $\mathbf{Y} = \mathbf{Q}_{dx}\mathbf{R}_{xx}^{-1}\mathbf{Q}_{dx}^H$. This step does not involve any multiplications; instead, $\mathbf{Y}$ is equal to the lower-right $M \times M$ submatrix of $\mathbf{Z}$. Note that $\mathbf{Y}$ is still Toeplitz.

- Let $\mathbf{X} = \left(\mathbf{Q}_{dx}\mathbf{R}_{xx}^{-1}\mathbf{Q}_{dx}^H - \mathbf{I}_M\right)^{-1}$. This inverse is found in the same manner as the first step, since the term in parentheses is Toeplitz. This step requires $3M^2/2 + M/2$ complex multiplications. The matrix $\mathbf{X}$ is again Toeplitz.

- Let $\mathbf{W} = \mathbf{Q}_{dx}\mathbf{X}\mathbf{Q}_{dx}^H$. This step does not require any multiplications. $\mathbf{W}$ is equal to $\mathbf{X}$ with an additional row of zeros at the top of the matrix and an additional column of zeros at the front of the matrix.

- Finally, $\mathbf{V} = \mathbf{R}_{xx}^{-1}\mathbf{W}\mathbf{R}_{xx}^{-1} = \mathbf{Z}\mathbf{W}\mathbf{Z}$. The matrix $\mathbf{V}$ has the desirable properties that the lower-right $M \times M$ submatrix is Hermitian Toeplitz and the top row and first column are related via Hermitian symmetry. $\mathbf{V}$ can be found by performing $M(4M+1)$ complex multiplications.

Finally, the components of $\mathbf{R}^{-1}$ are found with the following steps:

- $\mathbf{S}_{\mathbf{C}}^{-\mathbf{1}}$ is found by performing $7M^2 + 5M + 1$ complex multiplications as discussed in the previous steps.

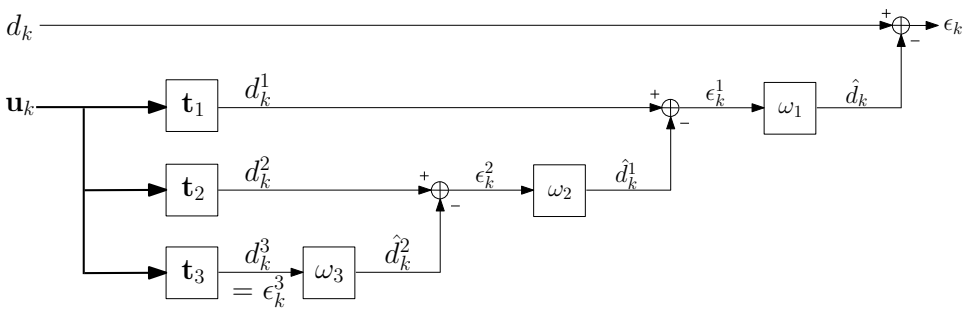


Figure 3.6: Multistage Wiener filter as a filterbank.

- $\mathbf{S_C^{-1}Q}_{dx}^H$ is the inverse of Schur's complement (found in the previous step) with the first column eliminated.
- $\mathbf{Q}_{dx}\mathbf{S_C^{-1}}$ is the inverse of Schur's complement with the first row eliminated.
- $\mathbf{Q}_{dx}\mathbf{S_C^{-1}Q}_{dx}^H$ is the inverse of Schur's complement with both the first row and first column eliminated.

A total of $7M^2 + 5M + 1$ complex multiplications are needed to determine the inverse. The approximation for the Wiener solution is then found by multiplying this inverse by the cross-correlation vector. Recalling that the signal power is normalized to unity, $\mathbf{p}$ simply selects the first column of the inverse. No additional multiplications are necessary to find the approximation of the Wiener filter.

3.5.4 Eigenvalues of Schur's Complement

The minimum eigenvalue of (3.10) for the case of $M > 1$ is found to be $\lambda_{\min} = \sigma_n^2$ with multiplicity $M - 1$. When this eigenvalue approaches zero (i.e SNR approaches infinity), the inverse may diverge as shown in Section 3.7.

3.6 Multistage Wiener Filter

3.6.1 Wiener Filter

The next approach used for DAI is based on the concept of Multi-Stage Wiener Filters (MSWFs) as a method that decomposes the conventional Wiener filter into a nested chain of scalar Wiener filters based on orthogonal projections [49]. They can be

used to approximate the desired Wiener filter in a lower dimension (i.e a form of rank reduction), thus reducing the complexity of the algorithm.

Honig and Xiao [61] observed that the Wiener solution provided by the MSWF is found in the Krylov subspace of the correlation matrix, $\mathbf{R}$, and the cross-correlation vector, $\mathbf{p}$. This observation led to the use of the Arnoldi algorithm [50] as a means to determine the orthonormal basis vectors. Using the fact that the correlation matrix is Hermitian, Joham, et al. [66] noted that the Lanczos algorithm [50] can replace the Arnoldi algorithm when finding the basis vectors. Finally, Dietl, et al. [36] derived the relationship between the Conjugate Gradient (CG) algorithm [50, 59, 112] and the Lanczos algorithm for use with MSWFs. This formulation allows the filter weights and the MSE to be updated as each stage is added. The CG implementation of the MSWF also reduces the required complexity by one matrix-vector product as compared to the Lanczos algorithm. The reduced complexity of this approach is illustrated for the implementation of the equalizer for the high speed downlink packet access (HSDPA) receiver in [37], where it was shown that a Krylov equalizer allows a reduction in the computational complexity and storage requirements with almost no loss in performance.

The CG algorithm is a method of the Conjugate Direction family of iterative techniques [112]. It utilizes $\mathbf{R}$-orthogonal (or conjugate) search directions, where exactly one step is taken in each direction toward the solution. The solution is guaranteed to be found in m steps, where m is the dimension of $\mathbf{R}$. This particular algorithm has been shown to be especially useful for solving problems of the type, $\mathbf{Rw} = \mathbf{p}$, when the dimensionality of $\mathbf{R}$ is large and $\mathbf{R}$ is sparse [112, 126].

An example of a MSWF as a filter bank can be seen in Figure 3.6. Note that $\mathbf{t}_i$ are the orthonormal basis vectors for the observation space, ω_i are the scalar Wiener filters, and d_l is the desired value. The error values are defined as $\epsilon_k^i = d_k^i - \hat{d}_k^i$.

The basis vectors can be found [66] according to

$$\mathbf{t}_i = \frac{\mathbf{P}_{i-1}\mathbf{P}_{i-2}\mathbf{R}\mathbf{t}_{i-1}}{\|\mathbf{P}_{i-1}\mathbf{P}_{i-2}\mathbf{R}\mathbf{t}_{i-1}\|}, \tag{3.12}$$

where $\|\cdot\|_2$ is the 2-norm and $\mathbf{P}_k$ is the projection operation onto the space orthogonal to $\mathbf{t}_k$, i.e.,

$$\mathbf{P}_k = \mathbf{I}_{2M+1} - \mathbf{t}_k\mathbf{t}_k^H, \tag{3.13}$$

where $\mathbf{I}_{2M+1}$ is the $(2M+1) \times (2M+1)$ identity matrix. The algorithm is initialized with $\mathbf{P}_0 = \mathbf{I}_{2M+1}, \mathbf{t}_0 = \mathbf{0}$, and $\mathbf{t}_1 = \mathbf{p}/\|\mathbf{p}\|$.

Let $\mathbf{T}^{(D)} = \left[\mathbf{t}_1, \quad \mathbf{t}_2, \quad \ldots, \quad \mathbf{t}_D\right]$ be a set of D basis vectors. The number of basis vectors (D) is chosen dependent upon the desired complexity and approximation to the Wiener solution. The application of (3.12) and (3.13) in an interference limited environment returns $D = 3$ basis vectors:

$$\mathbf{T}^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{M}} e^{-j\Omega_i T} & 0 \\ \vdots & \vdots & \vdots \\ 0 & \frac{1}{\sqrt{M}} e^{-jM\Omega_i T} & 0 \\ 0 & 0 & \frac{1}{\sqrt{M}} e^{-j\Omega_i T} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \frac{1}{\sqrt{M}} e^{-jM\Omega_i T} \end{bmatrix}. \tag{3.14}$$

The observation vector, $d_l^{(3)}$, is found using the basis vectors given in (3.14),

$$d_k^{(3)} = \begin{bmatrix} d_k^1 \\ d_k^2 \\ d_k^3 \end{bmatrix} = \mathbf{T}^{(3),H}\mathbf{u}_k = \begin{bmatrix} x_k \\ \frac{1}{\sqrt{M}} \sum_{m=1}^{M} x_{k-m} e^{j\Omega m T} \\ \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \hat{d}_{k-m} e^{j\Omega m T} \end{bmatrix}. \tag{3.15}$$

Note that this algorithm is run using training data, so $\hat{d}_k$ can be replaced with the actual transmitted symbols, d_k.

The first basis vector is chosen to maximize the correlation between the desired signal, d_k, and the first observed signal, d_k^1. The second basis vector is orthogonal to the first and is associated with the feedforward side taps. This can be seen by noting that the second observed value, d_k^2, is a linear combination of past received samples. Notice also that the components of this basis vector contain the phase shifts found in (2.46) that are needed to cancel the interference. The third basis vector which is orthogonal to the first two, is associated with the feedback taps because the third observed value, d_k^3, is a linear combination of the fed back symbols. The components have again the phase shifts needed, such that the post-cursor ISI caused by the feedforward side taps can be canceled.

Finally, the scalar Wiener filters can be found as

$$\omega_i = \frac{r_{\epsilon^i, d^{i-1}}}{\sigma_{\epsilon^i}^2}, \tag{3.16}$$

where $r_{\epsilon^i, d^{i-1}} = \mathbb{E}\left[\epsilon_k^i \left(d_k^{i-1}\right)^*\right]$ and $\epsilon_k^D = d_k^D$. Using the autocorrelation function given

in (2.27) and the above equations, the scalar Wiener filters are found to be

$$\omega_3 = 1, \tag{3.17}$$

$$\omega_2 = E_s + \sigma_n^2 + ME_i, \tag{3.18}$$

$$\omega_1 = \frac{\left(\sigma_n^2 + ME_i\right) E_s}{(E_s + \sigma_n^2)(\sigma_n^2 + ME_i) + \sigma_n^2 E_i}. \tag{3.19}$$

The intermediate estimated observations ($\hat{d}_k^i$) are

$$\hat{d}_k^2 = d_k^3, \tag{3.20}$$

$$\hat{d}_k^1 = \frac{E_i\sqrt{M}}{\sigma_n^2 + ME_i}\left[d_k^2 - d_k^3\right]. \tag{3.21}$$

The estimate of the desired signal ($\hat{d}_k$) is then found to be

$$\begin{aligned}\hat{d}_k &= \omega_1 d_k^1 + \omega_1\omega_2 d_k^2 - \omega_1\omega_2\omega_3 d_k^3 \\ &= \kappa_{\text{DFE}} \begin{bmatrix} \sigma_n^2 + ME_i \\ -E_i e^{-j\Omega_i T} \\ \vdots \\ -E_i e^{-jM\Omega_i T} \\ E_i e^{-j\Omega_i T} \\ \vdots \\ E_i e^{-jM\Omega_i T} \end{bmatrix}^H \quad \mathbf{u}_k = \mathbf{w}_{\text{DFE}}^H \mathbf{u}_k,\end{aligned} \tag{3.22}$$

where $\kappa_{\text{DFE}} = \frac{\text{SNR}}{(1+\text{SNR})(\sigma_n^2+ME_i)+E_i}$ and $\mathbf{w}_{\text{DFE}}$ is the desired Wiener filter. This implies that when the theoretical correlation matrices and the frequency of the interferer are available, the iterative algorithm requires $D = 3$ stages to obtain the Wiener filter.

The three subspaces arise from the structure of the DFE, as seen earlier in the basis determination. For the special structure seen in (2.45)-(2.48), the feedforward side taps have the same magnitude and the feedback side taps also have the same magnitude. This implies that three values need to be determined: the main tap, the feedback side taps, and the feedback taps. From the first line of (3.22), ω_1 is the weight needed to scale the current received sample, $d_k^1 = x_n$, while the product of ω_1 and ω_2 provides the weight for the feedforward side taps. Recall that the phase shifts are contained in the basis vector. Finally, noting that the feedforward side taps and the feedback taps are negatives of each other, gives $\omega_3 = 1$, which allows the cancellation of the fed back data symbols.

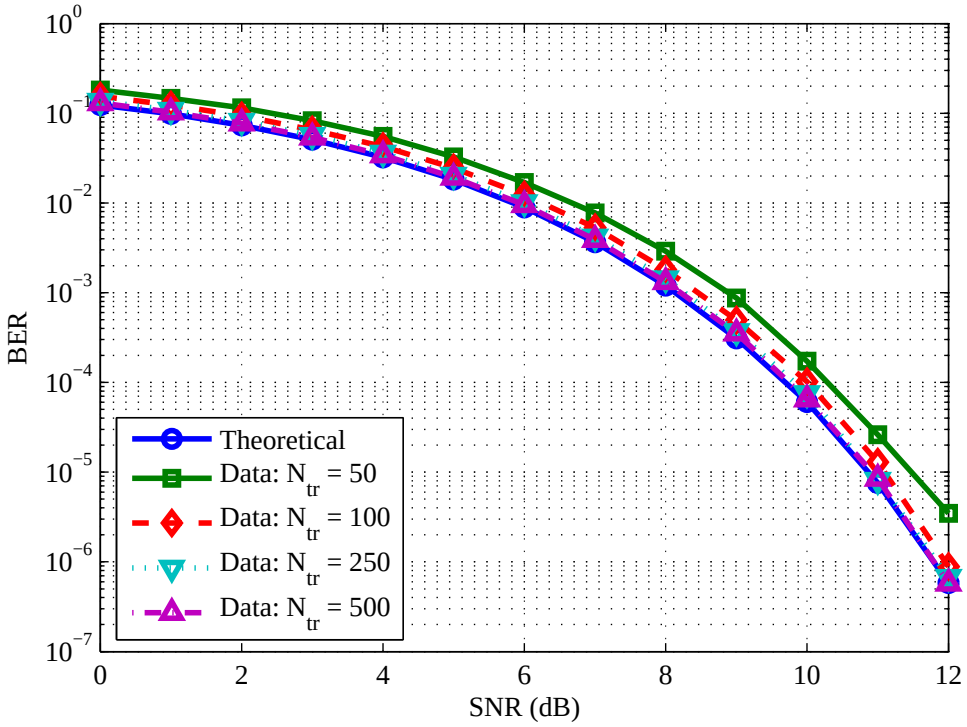


Figure 3.7: BER performance for the theoretical Wiener filter and the full-rank CG MSWF for a varying number of training symbols, SIR = -20 dB.

3.7 Results

In the simulation results to follow, a BPSK constellation is considered, the SIR = -20 dB, and the interferer frequency is located at DC ($\Omega_i = 0$). The equalizer order is set to $M = 3$, where the DFE is formed by a 4-tap feedforward filter and a 3-tap feedback filter. Each packet is made up of 10,000 symbols. Note that the received samples and the training symbols are used to form the estimates of the correlation matrices given in (3.2) and (3.4). The estimated weights are then used to calculate the BER in a decision-directed mode over the remaining symbols.

Figure 3.7 demonstrates the BER for the theoretical Wiener filter and the full-rank MSWF as a function of the number of training symbols. The MSWF DFE suffers when the estimated correlation matrices are poor, due to the lack of training symbols. When this number is increased, the performance of the CG MSWF approaches that of the theoretical Wiener filter. For this scenario, both 250 and 500 training symbols provide a good approximation, which entails a 2.5-5% of overhead for this scenario.

Figure 3.8 is a plot of the BER for the theoretical Wiener filter and the reduced-

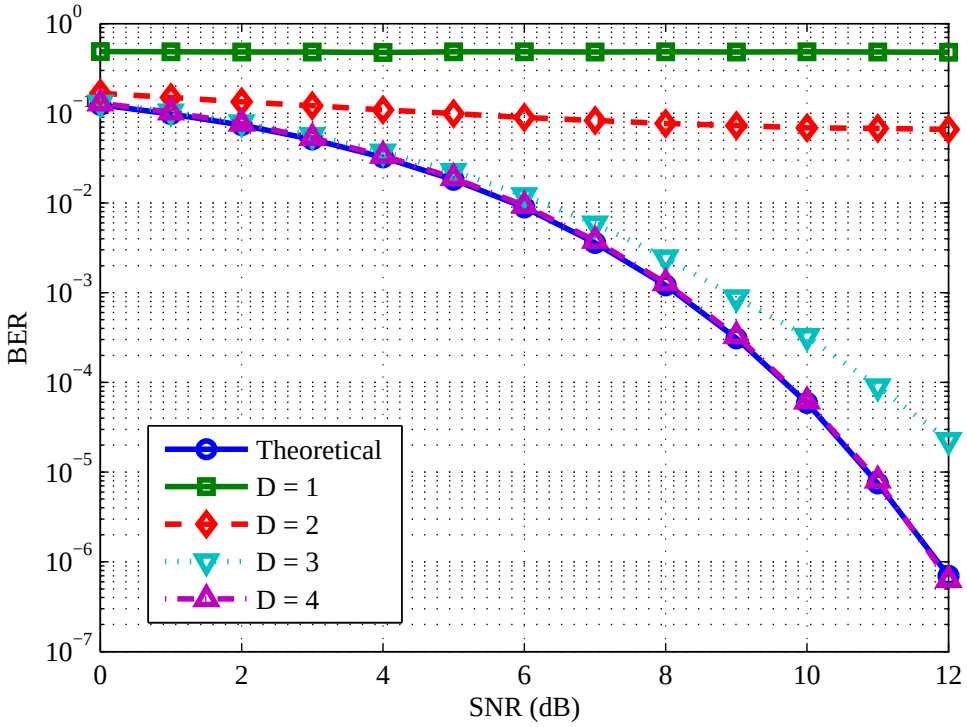


Figure 3.8: BER performance for the theoretical Wiener filter and the reduced-rank CG MSWF for a varying number of stages, SIR = -20 dB and $N_{\text{tr}} = 500$.

rank MSWF as a function of the number of implemented stages, with $N_{\text{tr}} = 500$. When employing the theoretical correlation matrices with the MSWF algorithm, it was previously shown that 3 stages are needed to obtain the Wiener solution. In this plot for $D = 1, 2$, the performance is very poor, indicating that the MSWF solution is far from the desired. An improvement is obtained when using $D = 3$ stages, however toward the high SNR region, the performance deviates from the ideal. An additional stage ($D = 4$) improves the performance and it is comparable to that of the theoretical Wiener filter. Increasing the number of stages further does not provide any further gains as $D = 4$ has reached the limit of the theoretical Wiener filter.

Figure 3.9 shows the BER for the theoretical Wiener filter and the parametric approximation to the DMI solution as a function of the number of training symbols. The parametric approximation to the DMI solution provides a good approximation to the desired correlation matrices at low SNR for all sizes of training symbols. However, as the SNR increases, two forms of degradation for the parametric approximation to DMI solution are noticed. The first deviation from the theoretical curve occurs because of the assumptions made on the structure of the correlation matrix. Recall that we are

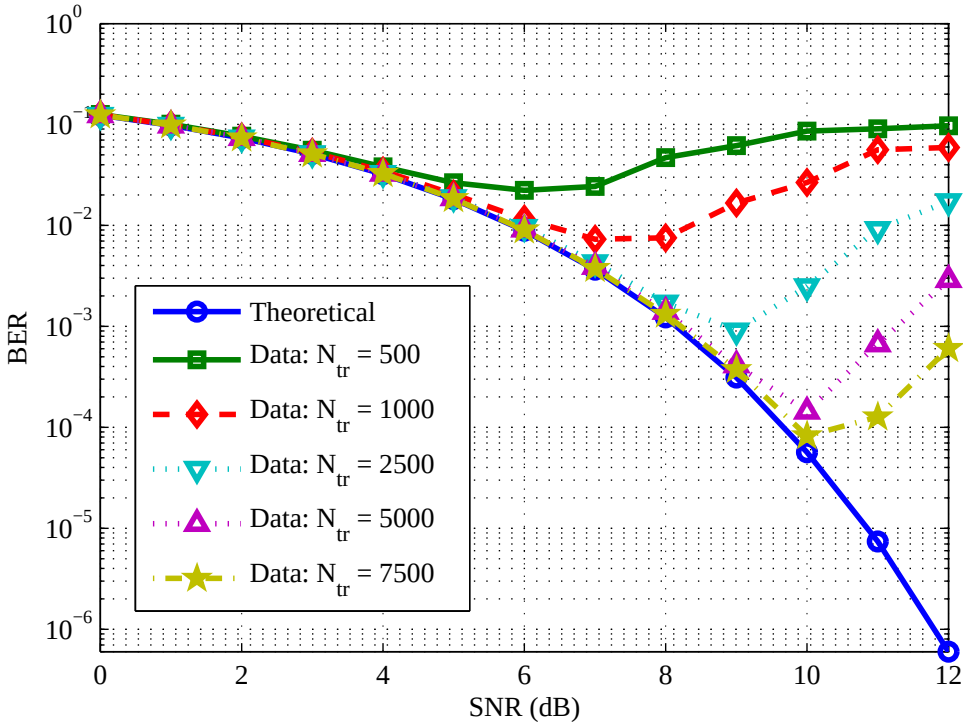


Figure 3.9: BER performance for the theoretical Wiener filter and the parametric approximation to the DMI solution for a varying number of training symbols, SIR = -20 dB.

attempting to invert, $\hat{\mathbf{S}}_{\mathbf{C}} = \hat{\mathbf{R}}_{xx} - \hat{\mathbf{Q}}_{dx}^{H}\hat{\mathbf{Q}}_{dx}$, however, because we assume that $\hat{\mathbf{Q}}_{dx} = \mathbf{Q}_{dx}$ and $\hat{\mathbf{I}}_{\mathbf{M}} = \mathbf{I}_{\mathbf{M}}$ in (3.7), we are actually solving $\hat{\mathbf{S}}_{\mathbf{C}} = \hat{\mathbf{R}}_{xx} - \mathbf{Q}_{dx}^{H}\mathbf{Q}_{dx}$. These assumptions, the forced Toeplitz structure of $\mathbf{R}_{xx}$, and the use of the ideal cross-correlation vector ($\mathbf{p}$) lead to a loss of information when inverting the matrix and finding the approximate Wiener solution. Note that the initial deviation from the theoretical curve occurs earlier for smaller sets of training symbols. As more training is employed, the estimate of the correlation matrix approaches the theoretical matrix, which allows the assumptions made to hold and provide better performance. The second degradation occurs at higher SNR values and is due the minimum eigenvalue of the data-based estimate of Schur's complement. This eigenvalue given in Section 3.5.4 is close to zero in this SNR region causing the inversion process to suffer. The MSWF algorithm seen in Figures 3.7 and 3.8 is not affected in a similar manner because the CG implementation is less sensitive to eigenvalue spread [112, 126] of $\mathbf{R}$.

For the parameters used, the parametric approximation to the DMI solution is less complex, needing 79 multiplications, while the CG MSWF requires 347 multi-

plications and 8 scalar divisions. Note, however, that the impact of these additional computations for the CG MSWF on performance is shown to be significant since the requirement of additional training symbols and reduced reliability is a major system limitation.

The CG MSWF algorithm proves to be a better option than the parametric approximation to the DMI solution even though its complexity is larger because it needs fewer training symbols to obtain performance comparable to that of the theoretical DFE Wiener filter for all SNR values.

3.8 Summary

In the first part of the chapter, the nonlinear dynamics of the NLMS equalizer are studied in the presence of severe narrowband interference and multipath, and in comparison with other equalizer structures. The previous findings that an NLMS equalizer can provide lower MSE than the corresponding time-invariant Wiener equalizer in the presence of narrowband interference, are shown to translate into a significant BER performance advantage. The BER performance of the decision-directed NLMS equalizer is shown to approach the BER performance of the decision-feedback NLMS equalizer when both structures use essentially the same information. The drawback of this approach is the need for a long training period. Finally, it is shown that the extended training period is not necessary when data-aided initialization (DAI) is employed to initialize the Wiener weights from estimates generated from the training data.

The second part of this chapter is dedicated to the investigation of two different computationally efficient techniques for obtaining the estimate of the DFE Wiener weights necessary for DAI. The two methods are compared in terms of complexity and performance relative to that of the theoretical Wiener filter for the DFE. The parametric approximation to the DMI solution is the lower complexity method ($7M^2 + 5M + 1$ multiplications, where M is the order of the feedback filter), however, the performance is degraded due to assumptions on the correlation matrix, $\mathbf{R}$, that are only met when a large number of training symbols is used. Further degradation occurs when the minimum eigenvalue gets small (at high SNR) causing the inverse to possibly diverge. The CG MSWF algorithm proves to be a better option than the parametric approximation to the DMI solution even though its complexity is larger ($4DM^2 + 2(7D + 1)M + 7D + 1$ multiplications and $2D$ scalar divisions, where D is the number of stages implemented). The

benefit arises because it needs fewer training symbols to obtain performance comparable to that of the theoretical DFE Wiener filter for all SNR values.

Chapter 3, in part, is a reprint of material as it appears in A. Batra, T. Ikuma, J. R. Zeidler, A. A. Beex, and J. G. Proakis, "Mitigation of Unknown Narrowband Interference Using Instantaneous Error Updates," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, pp. 115–119, Nov. 2004, A. Batra, J. R. Zeidler, and A. A. Beex, "Initialization Techniques for Improved Convergence of LMS DFEs in Strong Interference Environments," in *Proceedings of the IEEE Global Communications (Globecom) Conference*, Washington, DC, pp. 3068–3073, Nov. 2007, and is currently being prepared for submission for publication of the material. A. Batra, J. R. Zeidler, and A. A. Beex, "Implementation Methods for Data Aided Initialization," in preparation, 2009. The dissertation author was the primary investigator and author of these papers.

4 A Two-Stage Approach for Improving Convergence

In this chapter, an alternate technique to reduce the required training period and convergence time is evaluated. The results of Chapter 3 illustrate that an increasing amount of training is required to achieve a given bit error rate (BER) when the interference is strong. The required training period increases as the strength of the multipath coefficient increases as well. To reduce the convergence time and the number of training symbols needed, a two-stage system that uses a least-mean square (LMS) prediction-error filter (PEF) as a pre-filter to the LMS decision-feedback equalizer (DFE) is proposed. In this approach the PEF generates a direct reference for the interference from past samples and mitigates it prior to equalization.

A two-stage system employing a linear predictor has been previously investigated [51, 107] in combination with the constant modulus algorithm (CMA). The prediction filter is employed to mitigate the interference and ensure that the CMA locks on to the signal of interest. The prediction filter is not used specifically for its convergence properties. The two-stage structure in this chapter uses a supervised algorithm for the adaptation of the second structure and is developed with the goal of improving the convergence of the overall system.

It will be shown that the two-stage system reduces the number of training symbols required to reach a BER of 10^{-2} by approximately two orders of magnitude without substantially degrading the steady-state BER performance as compared to the LMS DFE-only case. All comparisons will be made under the condition that the LMS DFE-only and the two-stage structure have the same numbers of total taps.

It will be further shown that the two-stage system may be implemented with a blind algorithm that does not require any training symbols. The PEF is inherently a

blind algorithm because the error signal is determined from the current sample and the past samples. A relationship between the PEF weights and the DFE feedback weights is obtained, allowing the DFE to be operated in decision-directed mode after convergence of the PEF weights. This technique outperforms the non-blind decision-directed implementation when a small number of training symbols is used. The non-blind decision-directed implementation suffers because the feedback weights lie far from their steady-state values prior to the switch to decision-directed mode. This blind method also allows for a reduction in the complexity of the system (i.e. fewer weights that need to be adapted), at the cost of a slight decrease in steady-state BER.

4.1 Two-Stage System

As discussed in Section 2.5.8, the PEF provides an approximately white output spectrum when an infinite number of filter taps is used. Each additional tap provides an increase in spectral resolution when notching out the narrowband interference. However, the implementation of a large number of taps increases both the complexity and convergence time and is not generally feasible, consequently some distortion in the form of post-cursor ISI will be present. To combat the distortion induced by the PEF, the DFE provides a simple structure that removes the ISI without enhancing the noise. This leads to a simple two-stage structure that uses the PEF for rapid convergence and the DFE for removing post-cursor ISI as a system to mitigate narrowband interference. A similar approach is discussed in [5, pp. 364-365] when deriving the zero-forcing decision-feedback equalizer. Barry, et al., demonstrate that the optimal DFE precursor equalizer is related to optimal linear prediction. Consider transmitting data through a channel that induces ISI. This distortion can be removed by employing a linear zero-forcing equalizer, while causing the noise samples at the output of the equalizer to be correlated. This correlation can be subsequently removed with a PEF, at the expense of post-cursor ISI. Finally, a zero-forcing feedback post-cursor equalizer removes the ISI without enhancing the noise.

The performance of the PEF followed by the DFE is now considered, which will be abbreviated as PEF+DFE. A block diagram of the two-stage structure is shown in Fig. 4.1. Recall that the PEF is tasked with whitening the spectrum by removing the interference, but due to its limited length it will introduce post-cursor ISI; this ISI is then removed by the DFE. The DFE is designed to have a one tap feedforward section and an M_{fb}-tap feedback section. In general, there is no need for a feedforward section, because

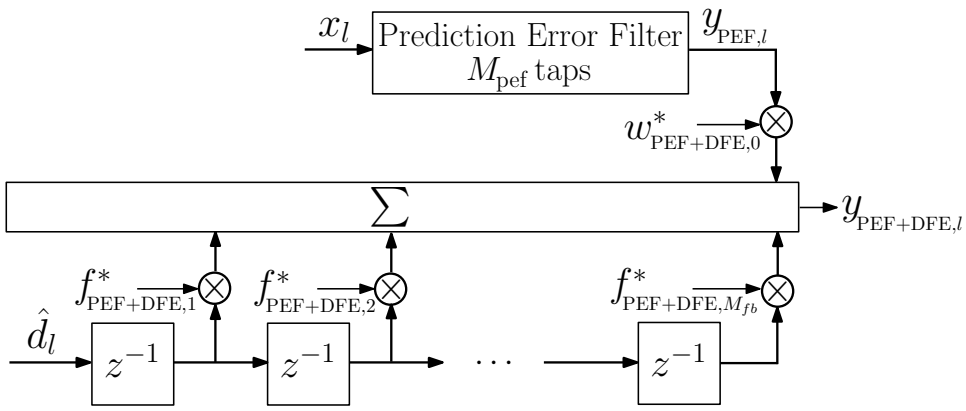


Figure 4.1: Two-stage structure (PEF+DFE) block diagram.

the input is distorted with only post-cursor ISI that can be resolved by the feedback equalizer portion. The single tap is included to compensate for any phase shifts that might exist because of phase errors, and/or gain mismatch between the transmitter and receiver.

4.1.1 Feedback Filter Order Estimation

The optimal feedback filter order can be estimated from the output of the DFE. When the feedforward filter weight is assumed to be equal to unity and the decisions fed back are perfect, the output is defined as

$$\begin{aligned} y_{\text{PEF+DFE},l} &= \sum_{m=0}^{0} w^*_{\text{PEF+DFE},m} y_{\text{PEF},l-m} + \sum_{m=1}^{M_{fb}} f^*_{\text{PEF+DFE},m} d_{l-m} \\ &= y_{\text{PEF},l} + \sum_{m=1}^{M_{fb}} f^*_{\text{PEF+DFE},m} d_{l-m}. \end{aligned} \tag{4.1}$$

To minimize the error, the optimal DFE weights are found according to

$$\begin{aligned} f_{\text{PEF+DFE},l} &= \arg\min_{f_l} \mathbb{E}\left[\left\| d_k - \left(y_{\text{PEF},k} + \sum_{l=1}^{M_{fb}} f^*_l d_{k-l} \right) \right\|^2\right] \\ &= \arg\min_{f_l} \mathbb{E}\left[\left\| d_k - \left(x_k - K \sum_{m=0}^{M_{\text{pef}}-1} x_{k-\Delta-m} e^{j\Omega_i (m+\Delta) T_s} + \sum_{l=1}^{M_{fb}} f^*_l d_{k-l} \right) \right\|^2\right]. \end{aligned} \tag{4.2}$$

Taking the derivative of the expected value term and setting this result to zero, the optimal weights are given by

$$f_{\text{PEF+DFE},l} = \begin{cases} Ke^{-j\Omega_i l T_s}, & l = \Delta, \ldots, \min(M_{fb}, M_{\text{pef}}), \\ 0, & l = \{1, \ldots, \Delta - 1\} \cup \{M_{\text{pef}} + \Delta, \ldots, M_{fb}\}. \end{cases} \tag{4.3}$$

When $\Delta = 1$, the optimal choice for the feedback filter order is $M_{fb} = M_{\text{pef}}$. This ensures that the ISI caused by the PEF is removed. With these choices and the assumption that the interference is canceled by the PEF, the output of the DFE is given by

$$y_{\text{PEF+DFE},k} = d_k + n_k - K \sum_{m=1}^{M_{\text{pef}}} n_{k-m} e^{j\Omega_i m T_s}. \tag{4.4}$$

4.1.2 Optimal Equalizer Weights after Prediction-Error Filtering

The DFE possesses a 1-tap feedforward section and an M_{fb}-tap feedback section. The optimal weights for the DFE are found by solving the Wiener-Hopf equations [56, 104, 134]. The feedforward weight is equal to $w_{\text{PEF+DFE}} = (\mathbf{R}_{\text{PEF,o}} - \mathbf{Q}^{\mathbf{H}}\mathbf{Q}/E_s)^{-1}\mathbf{p}$. The output autocorrelation matrix $\mathbf{R}_{\text{PEF,o}}$ reduces to a scalar value due to the 1-tap feedforward filter and is defined as

$$\mathbf{R}_{\text{PEF,o}} = r_{\text{PEF,o}}(0). \tag{4.5}$$

The latter term is given in (2.80). $\mathbf{Q}$ is defined as

$$\mathbf{Q} = \mathbb{E}\left[\mathbf{d}_k y^*_{\text{PEF},k}\right], \tag{4.6}$$

where the components of $\mathbf{Q}$ are given by

$$\mathbb{E}\left[d_{k-m} y^*_{\text{PEF},k}\right] = -KE_s e^{-j\Omega_i m T_s}, \quad m = \{\Delta, \ldots, \Delta + M_{\text{pef}} - 1\} \cap \{1, \ldots, M_{fb}\}. \tag{4.7}$$

Finally, $\mathbf{p}$ is defined as

$$\mathbf{p} = \mathbb{E}\left[y_{\text{PEF},k} d^*_k\right] = E_s. \tag{4.8}$$

The feedback weights are defined as $\mathbf{f}_{\text{PEF+DFE}} = -\mathbf{Q} w_{\text{PEF+DFE}}/E_s$.

4.1.3 Steady-State Equivalence

The two-stage structure can be viewed in a different manner when operating in steady-state. Based on linear system theory, two linear time-invariant (LTI) systems can

be combined into one LTI structure [94, pg. 107-108]. For example, the PEF weights given in (2.67) and the feedforward weight of the subsequent DFE ($w_{\text{PEF+DFE}}$) can be combined to form an extended feedforward filter ($\mathbf{w}_{\text{ext}}$) of a DFE with one main tap and $M_p = M_{\text{pef}} + \Delta - 1$ side taps. This is accomplished by

$$\mathbf{w}_{\text{ext}} = \mathbf{c}_{\text{PEF}} * w_{\text{PEF+DFE}} = w_{\text{PEF+DFE}} \times \mathbf{c}_{\text{PEF}}. \tag{4.9}$$

The feedback taps remain the same, that is $\mathbf{f}_{\text{ext}} = \mathbf{f}_{\text{PEF+DFE}}$. Observe that $\mathbf{w}_{\text{ext}}$ and $\mathbf{f}_{\text{ext}}$ are the weights of a DFE operating in steady-state. The case of interest is when $\Delta = 1$ and $M_{\text{pef}} = M_{fb}$ (as postulated in Section 4.1.1).

Solving,

$$w_{\text{PEF+DFE}} = (\mathbf{R}_{\text{PEF}} - \mathbf{Q}_{\text{PEF}}^H \mathbf{Q}_{\text{PEF}}/E_s)^{-1}\mathbf{p}_{\text{PEF}}, \tag{4.10}$$

$$\mathbf{f}_{\text{PEF+DFE}} = -\mathbf{Q}_{\text{PEF}} w_{\text{PEF+DFE}}/E_s, \tag{4.11}$$

for the weights gives

$$w_{\text{PEF+DFE}} = \frac{\text{SNR}}{\text{SNR} + (K^2 M_{fb} + 1) + (1 - KM_{fb})^2 E_i/\sigma_n^2}, \tag{4.12}$$

$$f_{\text{PEF+DFE},l} = \frac{K\text{SNR}}{\text{SNR} + (K^2 M_{fb} + 1) + (1 - KM_{fb})^2 E_i/\sigma_n^2} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}. \tag{4.13}$$

The extended feedforward filter weights can be found according to (4.9),

$$w_{\text{ext},0} = \frac{\text{SNR}}{\text{SNR} + (K^2 M_{fb} + 1) + (1 - KM_{fb})^2 E_i/\sigma_n^2}, \tag{4.14}$$

$$w_{\text{ext},l} = -\frac{K\text{SNR}}{\text{SNR} + (K^2 M_{fb} + 1) + (1 - KM_{fb})^2 E_i/\sigma_n^2} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}, \tag{4.15}$$

$$f_{\text{ext},l} = \frac{K\text{SNR}}{\text{SNR} + (K^2 M_{fb} + 1) + (1 - KM_{fb})^2 E_i/\sigma_n^2} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}. \tag{4.16}$$

Note that the feedback weights remain the same, namely (4.13) is equal to (4.16).

As mentioned previously in Section 2.5.2, the scenario of interest occurs when the interference dominates the signal of interest and the noise. Equations (4.14)-(4.16) can be approximated in this region using (2.68) to give

$$w_{\text{ext},0} \cong \frac{\text{SNR}}{(1 + \text{SNR}) + \frac{1}{M_{fb}}}, \tag{4.17}$$

$$w_{\text{ext},l} \cong -\frac{\text{SNR}}{(1 + \text{SNR})M_{fb} + 1} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}, \tag{4.18}$$

$$f_{\text{ext},l} \cong \frac{\text{SNR}}{(1 + \text{SNR})M_{fb} + 1} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}. \tag{4.19}$$

As a comparison to (4.17)-(4.19), the DFE-only weights described by (2.45)-(2.48) need to be approximated for the assumption of small SIR and NIR as well. Letting $M_p = M_{fb}$, so that there are $M_{fb}+1$ taps in the feedforward section and M_{fb} taps in the feedback section, the DFE-only weights are approximated as

$$w_{\mathrm{DFE},0} \cong \frac{\mathrm{SNR}}{(1+\mathrm{SNR}) + \frac{1}{M_{fb}}}, \tag{4.20}$$

$$w_{\mathrm{DFE},l} \cong -\frac{\mathrm{SNR}}{(1+\mathrm{SNR})M_{fb}+1} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}, \tag{4.21}$$

$$f_{\mathrm{DFE},l} \cong \frac{\mathrm{SNR}}{(1+\mathrm{SNR})M_{fb}+1} e^{-j\Omega_i l T_s}, \quad l = 1, \ldots, M_{fb}. \tag{4.22}$$

Comparing (4.17)-(4.19) and (4.20)-(4.22), it can be seen that combining the two-stage weights approximates the weights of the DFE-only.

4.1.4 Blind Implementation

The previous sections established a relationship between the PEF weights, the feedforward weight and the feedback weights. Note that in Section 4.1.1 the feedback weights are equal to the PEF weights associated with past data symbols scaled by the feedforward tap weighting. Also, recall that the weights of the PEF rapidly converge and the structure does not require knowledge of training symbols. With $\Delta = 1$ and $M_{\mathrm{pef}} = M_{fb}$, the two-stage system in Figure 4.1 can be implemented in a manner where the feedback tap weights are not adapted. After the PEF weights have converged, the multiplication of the PEF weights and the feedforward weight defines the feedback weights. The feedforward tap is initialized to unity and is adapted in decision-directed mode. Thus, no explicit training symbols are required during the adaptation process. This method also reduces the complexity of the system; only $M_{fb}+1$ of the total $2M_{fb}+1$ tap weights are adapted. In the scenario where there is a phase and/or gain error, the system requires the use of either training symbols to adapt the feedforward weight or a phase locked loop (PLL) and automatic gain control (AGC). Observe that these two components can be implemented in a decision-directed manner with no need for training symbols.

4.2 Results

4.2.1 Simulation Parameters

In the simulation results to follow, a QPSK constellation is utilized and the SNR = 9 dB. For convergence results a 100-symbol window was used and the BER values are averaged over 1,000 runs. The interferer frequency is located at DC ($\Omega_i = 0$). All of the data were considered as training data, unless specified otherwise. The step-sizes are chosen to ensure convergence toward the steady-state BER. The DFE steady-state BER results in the convergence plots are given by $Q\left(\sqrt{\text{SINR}}\right)$, where $Q(\cdot)$ is the Q-function [97, p. 40] and the SINR is given in (2.57).

The DFE adapted with the RLS algorithm [56] is also simulated as a benchmark for the LMS DFE and the LMS PEF+DFE. The forgetting factor and the regularization factor were found through trial and error and set to $\lambda = 0.99$, $\delta = 0.001$, respectively, for all simulations.

The adaptive weights are initialized such that the main tap is set to one, resulting in the desired symbol being part of the output of the equalizer. The remaining taps are set to zero.

4.2.2 Convergence Results

In previous works [74], [9] the convergence has been viewed through the adaptive weights, even though they may not be unique [13]. As discussed in Section 2.2.5, the convergence of the weights may lag behind the MSE convergence if the eigenvalues are small. Similarly, the weight convergence does not provide an indication of how the BER behaves during the transient period. Thus, the convergence results are shown in terms of a sliding BER window, discussed in Section 2.2.6.

Figure 4.2 demonstrates the convergence of the LMS DFE, the LMS PEF+DFE, and the RLS DFE in relation to the steady-state BER for SIR = -20 dB. The number of taps is set such that $M_p = M_{\text{pef}} = M_{fb} = 3$, and the step-sizes for each structure are $\mu_{\text{DFE}} = 0.0001, \mu_{\text{PEF}} = 0.0001, \mu_{\text{PEF+DFE}} = 0.01$. The LMS PEF+DFE is seen to converge significantly faster than the LMS DFE. Specifically, the LMS PEF+DFE converges to a BER of 10^{-2} in approximately 450 symbols (or iterations, as adaptation takes place at the symbol rate), while the LMS DFE converges in approximately 20,000 symbols. In the case of the RLS DFE, convergence to a BER of 10^{-2} occurs in 150

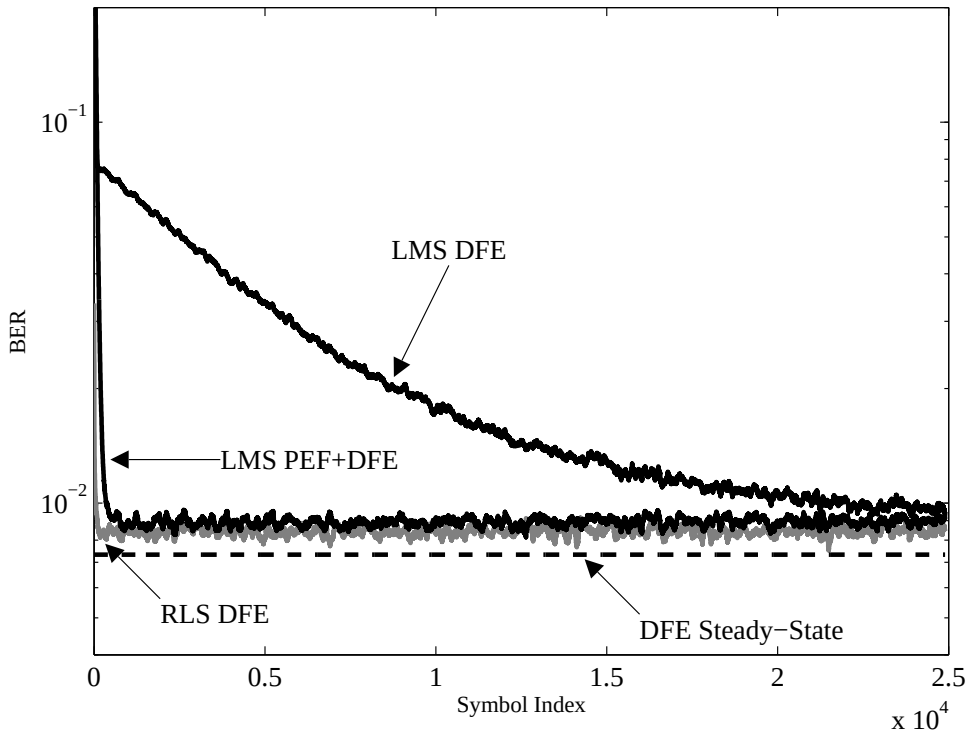


Figure 4.2: Convergence comparison of the LMS DFE, the LMS PEF+DFE, and the RLS DFE for SNR = 9 dB, SIR = -20 dB, $M_p = M_{\text{pef}} = M_{fb} = 3, \Omega_i = 0, \mu_{\text{DFE}} = 0.0001, \mu_{\text{PEF}} = 0.0001, \mu_{\text{PEF+DFE}} = 0.01, \lambda = 0.99, \delta = 0.001$.

symbols. As expected, RLS provides faster convergence because it whitens the input by using the inverse correlation matrix. This improved convergence comes at the cost of higher complexity. For example, in the context of echo cancellation, it has been shown that the implementation of RLS in floating point on the 32 bit, 16MIPS, 1 serial port, TMS320C31 requires 20 times the number of machine cycles that LMS does [106].

Figure 4.3 is a plot of the convergence for the above scenario when the SIR = -30 dB. The step-sizes for this case are $\mu_{\text{DFE}} = 0.00001, \mu_{\text{PEF}} = 0.00001, \mu_{\text{PEF+DFE}} = 0.001$. Again, the time required for convergence of the LMS PEF+DFE is dramatically less than for the convergence of the LMS DFE. The LMS PEF+DFE converges in 3,000 symbols, while the LMS DFE requires 200,000 symbols. The RLS DFE requires 160 symbols to converge for this case.

Finally, Figure 4.4 shows the convergence of the two systems when the number of filter coefficients for each stage is doubled, namely $M_p = M_{\text{pef}} = M_{fb} = 6$ and SIR = -20 dB. The step-sizes for this scenario are $\mu_{\text{DFE}} = 0.0001, \mu_{\text{PEF}} = 0.00005, \mu_{\text{PEF+DFE}} = 0.01$. The convergence time for both systems is reduced with the addition of more taps.

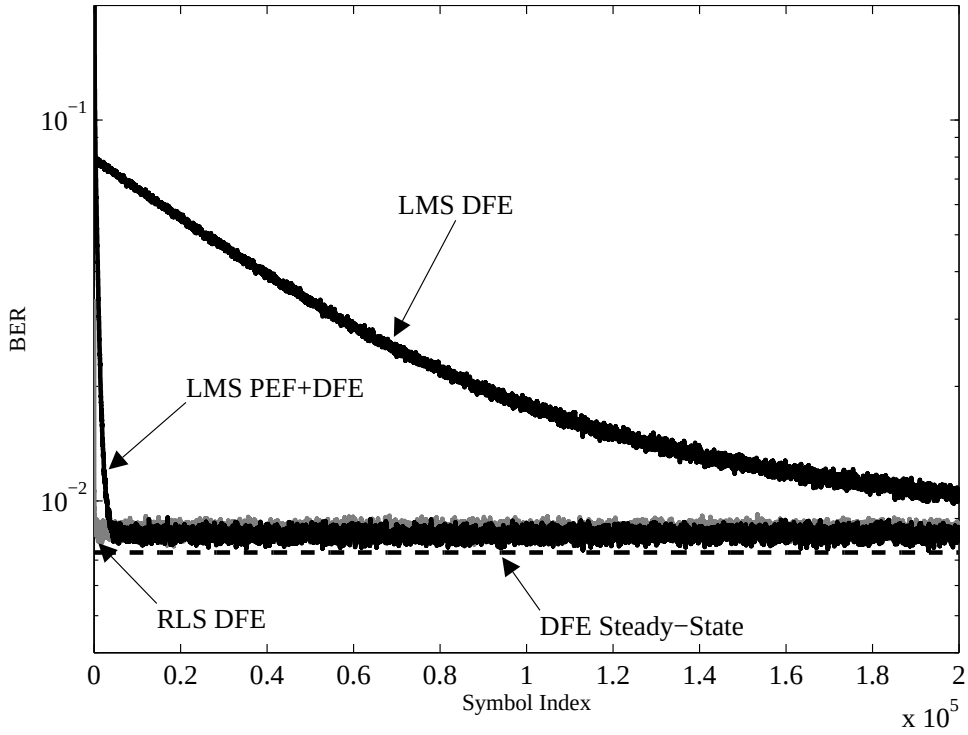


Figure 4.3: Convergence comparison of the LMS DFE, the LMS PEF+DFE, and the RLS DFE for SNR = 9 dB, SIR = -30 dB, $M_p = M_{\text{pef}} = M_{fb} = 3, \Omega_i = 0, \mu_{\text{DFE}} = 0.00001, \mu_{\text{PEF}} = 0.00001, \mu_{\text{PEF+DFE}} = 0.001, \lambda = 0.99, \delta = 0.001$.

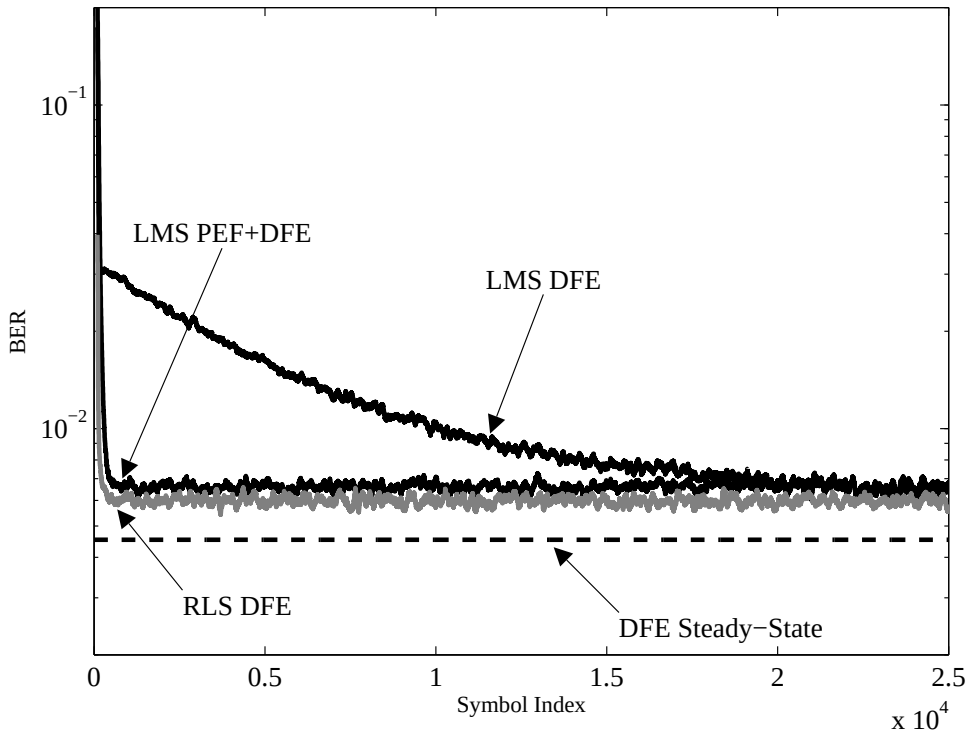


Figure 4.4: Convergence comparison of the LMS DFE, the LMS PEF+DFE, and the RLS DFE for SNR = 9 dB, SIR = -20 dB, $M_p = M_{\text{pef}} = M_{fb} = 6, \Omega_i = 0, \mu_{\text{DFE}} = 0.0001, \mu_{\text{PEF}} = 0.00005, \mu_{\text{PEF+DFE}} = 0.01, \lambda = 0.99, \delta = 0.001$.

The LMS PEF+DFE converges in 300 symbols and the LMS DFE converges in 10,000 symbols. Doubling the complexity, halves the convergence time required. The RLS DFE converges in 130 symbols. Note that increasing the order will eventually lead to a degradation in the performance due to the increase of gradient noise. This degradation is observed when increasing the number of taps from $K = L = M = 3$ (in Figure 4.2) to $M_p = M_{\text{pef}} = M_{fb} = 6$ (in Figure 4.4) by noting the difference in the misadjustment of the adaptive algorithm and DFE steady-state values.

Blind Implementation

In this section, the convergence of the blind implementation discussed in Section 4.1.4 is examined. This algorithm allows the LMS PEF to converge before the LMS DFE that follows it is turned on. Let N_{off} represent the number of symbols that are allocated to allow for PEF convergence. This system is compared to two other cases. The first is the scenario where all the transmitted symbols are considered as training data (similar to the results shown above). The second scenario demonstrates the convergence

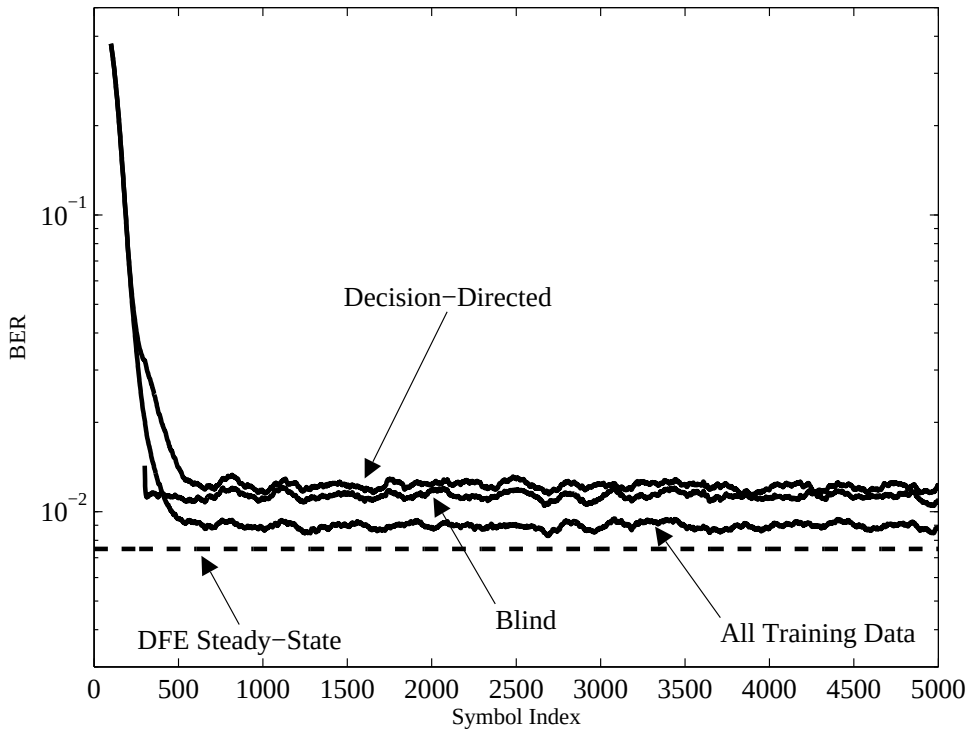


Figure 4.5: Convergence comparison of the different LMS PEF+DFE implementations for SNR = 9 dB, SIR = -20 dB, $M_p = M_{\text{pef}} = M_{fb} = 3, \Omega_i = 0, \mu_{\text{PEF}} = 0.0001, \mu_{\text{PEF+DFE}} = 0.01, N_{\text{off}} = 200$.

when a subset of the symbols is used for training, while the adaptive algorithm operates in decision-directed mode for the remaining symbols. This case is referred to as the decision-directed algorithm. The number of training symbols used for this case will also be equal to N_{off}.

Figure 4.5 demonstrates the BER convergence of the three discussed cases in relation to the steady-state BER for SIR = -20 dB and $N_{\text{off}} = 200$ symbols. The number of taps is set such that $M_p = M_{\text{pef}} = M_{fb} = 3$, and the step-sizes for each structure are $\mu_{\text{PEF}} = 0.0001$ and $\mu_{\text{PEF+DFE}} = 0.01$. The performance of both the blind algorithm and the decision-directed algorithm deviates from the case of using all training data. This is due to propagation of feedback errors that cause more errors. Observe that the blind algorithm produces faster convergence and slightly better BER performance than the decision-directed algorithm.

Figure 4.6 demonstrates the BER convergence of the three discussed cases in relation to the steady-state BER for SIR = -20 dB, however now $N_{\text{off}} = 100$ symbols. The blind algorithm now significantly outperforms the decision-directed algorithm in

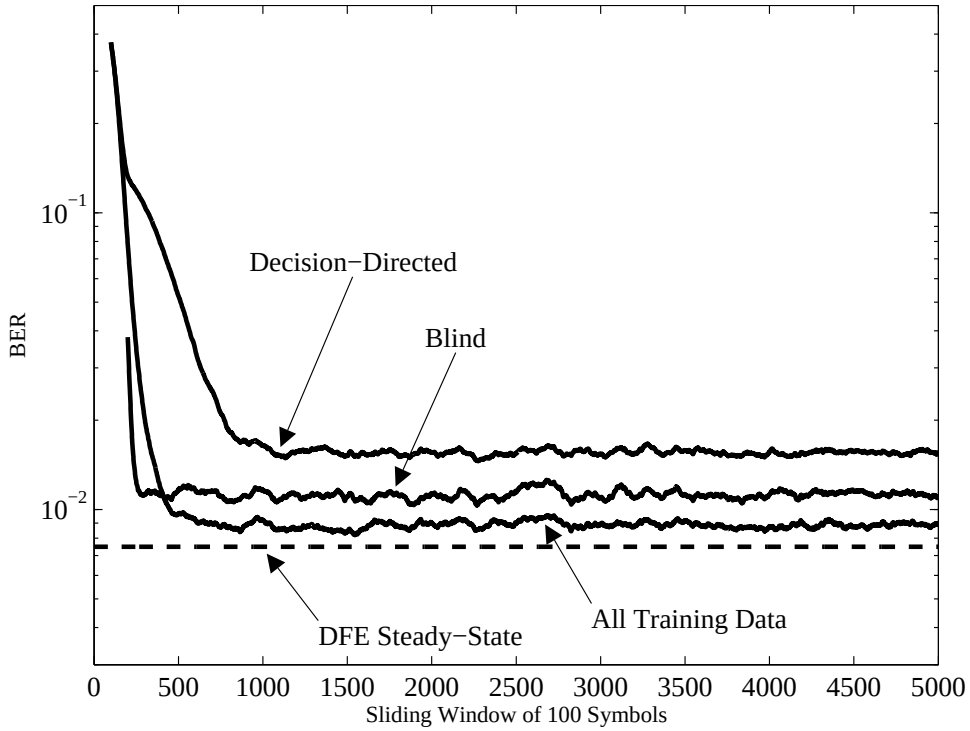


Figure 4.6: Convergence comparison of the different LMS PEF+DFE implementations for SNR = 9 dB, SIR = -20 dB, $M_p = M_{\text{pef}} = M_{fb} = 3, \Omega_i = 0, \mu_{\text{PEF}} = 0.0001, \mu_{\text{PEF+DFE}} = 0.01, N_{\text{off}} = 100$.

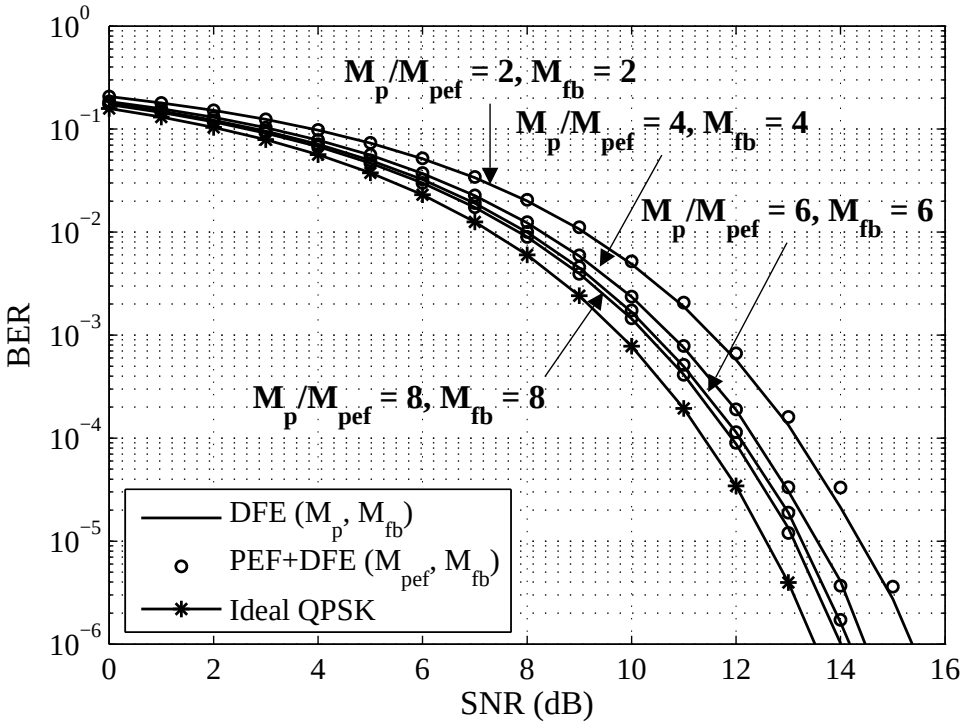


Figure 4.7: Steady-state BER results of the DFE and the PEF+DFE for SIR = -20 dB and $\Omega_i = 0$. DFE results obtained using optimal weights given in (2.45)-(2.48), PEF+DFE results obtained using optimal weights given in (2.67), (4.12), (4.13).

terms of both convergence and BER. The degradation of the decision-directed algorithm arises from the fact that the number of training symbols used does not allow the feedback weights to approach their steady-state values before switching to decision-directed mode.

4.2.3 BER Results

Figure 4.7 is a plot of the steady-state BER results for the DFE and PEF+DFE for SNR = -20 dB and varying filter orders. The performance of ideal QPSK is plotted as a reference. The performance of the PEF+DFE is seen to be approximately the same as the performance of the DFE when both structures are operating in steady-state. This validates the analysis performed in Section 4.1.3. It is also seen that the performance of the systems improves as the number of filter taps is increased, approaching the performance of QPSK. The improvement results from the increased spectral resolution provided by the larger number of taps in the feedforward section of each system.

Fig. 4.8 demonstrates the BER results of the LMS PEF+DFE blind implementation in comparison to the steady-state PEF+DFE results. For the blind implementation,

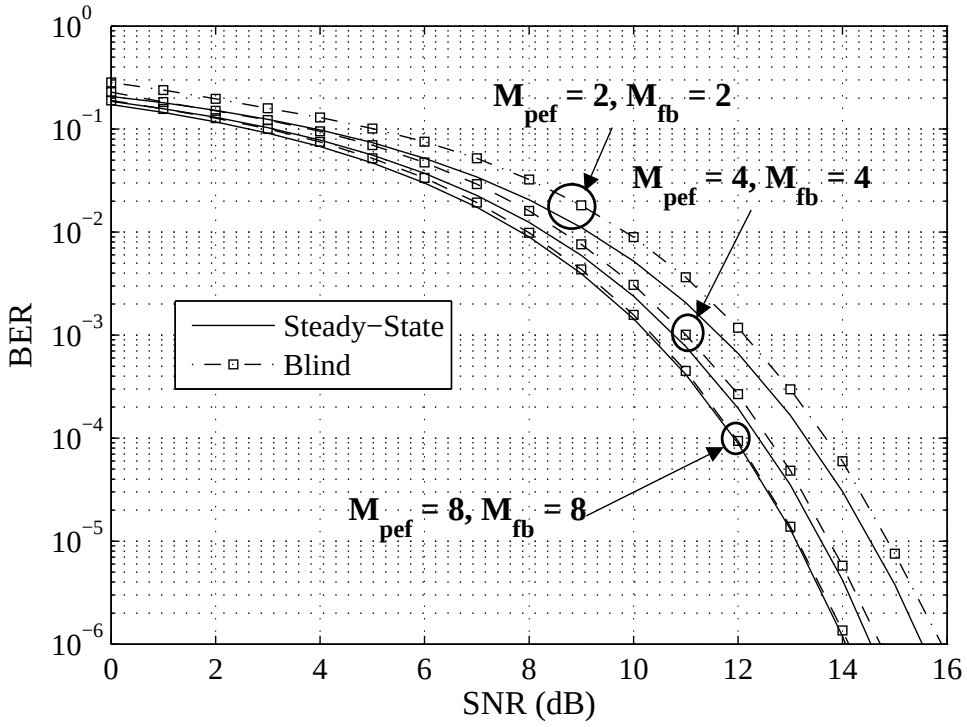


Figure 4.8: Steady-state BER results of PEF+DFE and the BER for the LMS blind implementation for SIR = -20 dB and $\Omega_i = 0$. PEF+DFE steady-state results obtained using optimal weights given in (2.67), (4.12), (4.13).

the DFE is turned on after $N_{\text{off}} = 250$ symbols and the BER is calculated over the last 2,500 symbols. The step-sizes are chosen for convergence to the steady-state BER and are noted in Table 4.1. This table also gives the average number of symbols required to obtain a BER of 10^{-2} for the blind implementation when SNR = 10 dB. A convergence value equal to N_{off} indicates that the blind algorithm has converged to the target BER after the first windowed calculation. It is clear that there is a small degradation in the BER when implementing the blind version of the PEF+DFE algorithm. This degradation is attributed to the combination of the misadjustment of the adaptive algorithm and the presence of uncanceled interference that causes feedback errors. Note that this degradation in BER becomes smaller as the number of parameters is increased. This occurs because a larger number of taps allows for more of the interference to be canceled, thereby reducing the number of feedback errors.

Table 4.1: Step-sizes and Convergence (at SNR = 10 dB) for LMS PEF+DFE Blind Implementation

$M_{\text{pef}} = M_{fb}$	2	4	8
μ_{PEF}	1e-4	5e-5	1e-5
$\mu_{\text{PEF+DFE}}$	0.01		
No. of symbols to BER = 10^{-2}	354	250	555

4.3 Summary

The response of the LMS DFE is investigated in the presence of severe narrowband interference. Due to the absence of a reference for the interference, the convergence time for this equalizer may be unacceptably slow for use in some realistic applications. The proposed system of an LMS PEF as a pre-filter to the equalizer is shown to provide a solution to this problem. The two-stage system's adaptive implementation is superior due to the fact that the prediction-error filter utilizes the narrowband nature of the interference to obtain a beneficial initialization point. On the other hand, the LMS DFE-only employs only the training symbols which have no knowledge of the statistical characteristics of the interference.

This two-stage system was shown to reduce the convergence time, in terms of reaching a BER of 10^{-2}, by approximately two orders of magnitude. An added benefit is that the steady-state BER for the two-stage system approximates that of the LMS DFE-only. Thus, it is possible to improve the convergence results of the LMS DFE, by splitting the system into an LMS prediction-error filter and a separate LMS DFE while not significantly degrading the steady-state BER results. The convergence results were also benchmarked against the DFE adapted with the RLS algorithm, which demonstrated faster convergence at the cost of higher complexity. A blind implementation (i.e. no training symbols are needed) that reduces complexity at the cost of a small degradation in the steady-state BER is also discussed.

Chapter 4, in part, is a reprint of material as it appears in A. Batra, J. R. Zeidler, and A. A. Beex, "A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Decision-Feedback Adaptive Equalizers in the Presence of Severe Narrowband Interference," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 390102, 13 pages, 2008. doi:10.1155/2008/390102. The dissertation author was the

primary investigator and author of this paper.

5 Multi-carrier Background

Multi-carrier transmission schemes [17, 42] are becoming the dominant modulation technique for current communication systems due to their ability to obtain higher data rates that users are demanding. In order to provide the increased data rate with a single carrier transmission scheme generally requires a bandwidth that is larger than the coherence bandwidth of the channel. This causes groups of frequencies to be attenuated and shifted in relative phase to each other thereby severely distorting the symbol. Consequently, in the time-domain, intersymbol interference arises as symbols are smeared with adjacent symbols requiring equalizers of increasing complexity as the ISI magnitude increases.

Multi-carrier schemes split the input data sequence into many low-rate streams that are transmitted on independent channels. Each channel could be considered as a different frequency bin in a frequency division multiplexing (FDM) scheme. However, this wastes bandwidth due to the guard bands needed when filtering at the receiver. In block modulated multi-carrier schemes, the narrowband subcarriers (or subchannels) are allowed to overlap orthogonally in the frequency-domain, thereby improving the spectral efficiency of the system. Orthogonality is obtained be choosing the subcarrier spacing to be the inverse of the symbol duration. This ensures that each subchannel has zeros located at all other subcarriers.

To solve the problem of intersymbol interference (and by duality frequency-selectivity), a guard interval is added to the front of the time-domain sequence. It is possible to simply use zero-padding for the guard interval, however this causes a loss of orthogonality [95] and requires a complex receiver that employs the overlap-and-add technique [88] as is done in multi-band orthogonal frequency division multiplexing (MB-OFDM) ultra-wideband (UWB) systems [38,64]. More commonly, a cyclic prefix is used. The cyclic prefix is a repeat of the end of the signal pre-pended to the beginning of that

signal. This method preserves orthogonality among the subcarriers. Note that this addition is overhead for the overall system, as it does not transmit any useful information. However, as will be shown, the cyclic prefix aids in reducing the complexity required for equalization.

These schemes are easily implemented in hardware as well. Modulation of all subcarriers may seem to require a bank of filters, however, it can be performed using the highly efficient inverse fast Fourier transform (IFFT) algorithm. Correspondingly, at the receiver, the FFT algorithm is utilized as the matched filter. A consequence of using the FFT/IFFT pair is that these matrices form the left and right eigenvectors of the circulant (due to the use of the cyclic prefix) channel matrix. The result is that the data is essentially multiplied by a diagonal matrix, and each subcarrier is distorted by flat fading. Hence, the frequency-selectivity problem is resolved and a one-tap equalizer on each subcarrier is now sufficient. Another consequence of having flat fading on each subcarrier is that the number of bits transmitted per subcarrier can vary depending on the channel gain for that particular subcarrier. This is referred to as bit loading [30], and requires channel knowledge at the transmitter.

For all of the benefits of multi-carrier transmission, there are also drawbacks. Carrier frequency offset (CFO) can be caused when very accurate frequency synchronization is not available. This causes intercarrier interference (ICI) because the sampling no longer occurs at the zeros of the other subcarriers. ICI may also arise when there is mobility in the system, and the transmissions experience a Doppler shift [129]. Another disadvantage is a large peak-to-average-power ratio (PAPR) that occurs when the phases of the subcarriers sum constructively. When passed through a nonlinear amplifier, intermodulation distortion may cause a loss of orthogonality of the subcarriers and the signal will be corrupted with intercarrier interference. To combat this problem, a backoff in the transmitted power is required to remain in the linear region of the amplifier, or have an amplifier with a large linear range. Techniques to compensate for the PAPR are described in [125]. Finally, narrowband interference can severely degrade the system and make reliable communication impossible. This type of interference may arise from radio nonlinearities, intentional jamming, or from sharing of spectrum. Many of the new wireless standards that employ multi-carrier schemes utilize the unlicensed bands, thus there is the possibility that the system will have to share the frequency band with other communication systems. For instance, WiMax is a narrowband interferer for

UWB Systems [119,120]. This is especially true when discussing OFDM-based cognitive radios [6] where it is required to modify the spectrum in order to avoid primary users.

Techniques to mitigate narrowband interference are developed in this chapter for two multi-carrier schemes; multi-carrier code division multiple access (MC-CDMA) [26,41,54,139] and orthogonal frequency division multiplexing (OFDM) [27,131]. These two systems differ in that MC-CDMA employs multi-carrier principles with spreading that has been seen in CDMA-based systems [97]. In MC-CDMA, each subcarrier of a given user is multiplied by a single chip of a spreading sequence. At the receiver, after demodulation of the subcarriers, the chips are correlated with a locally generated chip sequence to provide both the multiple access capability and frequency diversity. Conversely, OFDM does not spread the data symbols across the subcarriers. OFDM is thus a special case of MC-CDMA, in which the spreading codes are simply the columns of the identity matrix. The result of this is that uncoded OFDM does not possess frequency diversity.

These techniques differ from other wideband systems, such as direct-sequence CDMA (DS-CDMA) and multi-carrier direct-sequence CDMA (MC-DS-CDMA) [31,69]. For these modulations, the bandwidth of the signal is greater than the coherence bandwidth of the channel, inducing frequency-selective fading. A Rake receiver [97] can be used in addition to obtain path diversity, thereby reducing the deleterious effects of the frequency-selective multipath fading channel. A comparison of MC-DS-CDMA and MC-CDMA can be found in [42,77].

5.1 Multi-carrier Basics

A block modulated multi-carrier symbol is formed by modulating and summing individual subcarriers. In continuous time, the set of complex sinusoids, $\left\{e^{j2\pi f_k t}\right\}_{k=0}^{N-1}$ are referred to as the subcarriers, where $f_k = k/T_s$ and $1/T_s$ is the subcarrier spacing. This choice of spacing leads to orthogonality between the subcarriers. For example, looking at the k^{th} and m^{th} subcarriers, the correlation is given by

$$\begin{aligned}\frac{1}{T_s}\int_0^{T_s} e^{j2\pi f_k t}\left(e^{j2\pi f_m t}\right)^* \,\mathrm{d}t &= \frac{1}{T_s}\int_0^{T_s} e^{j2\pi (f_k - f_m)t}\,\mathrm{d}t \\ &= \begin{cases} 1, & f_k = f_m, \\ 0, & f_k \neq f_m. \end{cases}\end{aligned} \tag{5.1}$$

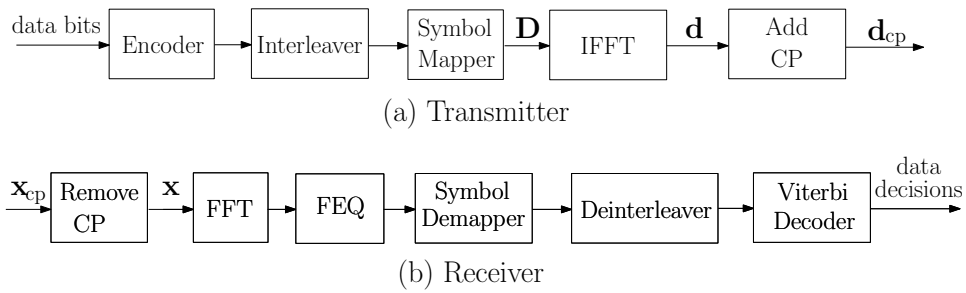


Figure 5.1: BICM OFDM system model.

The subcarriers can be examined in discrete-time allowing for modulation and demodulation to be performed using the DFT (implemented with the FFT). This is seen by sampling the subcarriers at N equally spaced instances,

$$\begin{aligned} e^{j2\pi f_k t}|_{t=lT_s/N} &= e^{j2\pi \frac{k}{T_s} l \frac{T_s}{N}} \\ &= e^{j\frac{2\pi}{N}kl}, \quad 0 \leq k, l \leq N-1. \end{aligned} \tag{5.2}$$

The discrete-time subcarriers, seen in (5.2), are the basis functions of the IDFT. A block diagram of the system is depicted in Figure 5.1.

A block of N data symbols, $\{D_k\}_{k=0}^{N-1}$ are defined in the frequency-domain. The energy per modulated symbol, $E_s = R_c R_m E_b$ where R_c is the code rate of the encoder (see Section 5.4.1), R_m is the number of bits per subcarrier that depends on the constellation size and E_b is the energy per information bit. These data symbols are modulated onto the subcarriers using the IDFT, giving the time-domain signal as

$$d_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} D_k e^{j\frac{2\pi}{N}kl}, \quad 0 \leq l \leq N-1. \tag{5.3}$$

Let $\mathbf{F}$ be the N-point DFT matrix, whose elements are given by

$$[\mathbf{F}]_{kl} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}kl}, \qquad 0 \leq k, l \leq N-1. \tag{5.4}$$

Note that $\mathbf{F}$ is unitary, i.e. $\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}_N$. Using this notation, (5.3) can be written in matrix form as

$$\mathbf{d} = \mathbf{F}^H \mathbf{D}, \tag{5.5}$$

where $\mathbf{D}$ is the vector of frequency-domain symbols, defined as

$$\mathbf{D} = \begin{bmatrix} D_0, & D_1, & \ldots, & D_{N-1} \end{bmatrix}^T, \tag{5.6}$$

and $\mathbf{d}$ are the time-domain samples given by

$$\mathbf{d} = \begin{bmatrix} d_0, & d_1, & \ldots, & d_{N-1} \end{bmatrix}^T. \tag{5.7}$$

The time-domain samples given in (5.3) can be transformed back to the frequency-domain using the DFT,

$$\begin{aligned} D_k &= \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} d_l e^{-j\frac{2\pi}{N}kl}, \quad 0 \leq k \leq N-1 \\ &= \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} D_m e^{j\frac{2\pi}{N}ml} e^{-j\frac{2\pi}{N}kl} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} D_m \sum_{l=0}^{N-1} e^{j\frac{2\pi}{N}(m-k)l} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} D_m N \delta_{m-k} \\ &= D_k. \end{aligned} \tag{5.8}$$

This can also be viewed in vector form when the time-domain samples are transformed using the DFT matrix, $\mathbf{F}$,

$$\begin{aligned} \mathbf{D} &= \mathbf{F}\mathbf{d} \\ &= \mathbf{F}\mathbf{F}^H\mathbf{D} \\ &= \mathbf{D}. \end{aligned} \tag{5.9}$$

As mentioned above, the cyclic prefix is utilized to mitigate the effects of ISI. The cyclic prefix is composed of the last N_g samples of $\mathbf{d}$, and is pre-pended to $\mathbf{d}$ to give,

$$\mathbf{d}_{\text{cp}} = \begin{bmatrix} d_{N-N_g}, & \cdots, & , d_{N-1}, & d_0, & \cdots, & d_{N-1} \end{bmatrix}^T. \tag{5.10}$$

The vector given in (5.10) is transmitted over a time-invariant multipath fading channel, $h(t)$. The channel samples are obtained by sampling at $t = lT_s/N$,

$$h_l = h(t)|_{t=lT_s/N}, \quad 0 \leq l \leq L_h - 1, \tag{5.11}$$

where L_h is the number of resolvable paths. Each tap, h_l, is modeled as an i.i.d. zero-mean, circularly Gaussian random variable with variance equal to σ_h^2. Without loss of generality, the total multipath power is normalized to unity,

$$\sum_{l=0}^{L_h-1} \mathbb{E}\left[|h_l|^2\right] = \sum_{l=0}^{L_h-1} \sigma_h^2 = 1. \tag{5.12}$$

The received time-domain samples are found by convolving, (5.10) and (5.11),

$$\mathbf{x}_{\text{cp}} = \mathbf{h} * \mathbf{d}_{\text{cp}} + \mathbf{n}_{\text{cp}}, \tag{5.13}$$

where $\mathbf{h}$ is the $1 \times L_h$ vector of channel taps given in (5.11) and $\mathbf{n}_{\text{cp}}$ is the $(N + N_g) \times 1$ vector of time-domain additive noise samples. If $N_g \geq L_h$, the guard interval can be neglected, noting that the cyclic prefix converts linear convolution into circular convolution. This allows (5.13) to be rewritten as,

$$\mathbf{x} = \mathbf{H}\mathbf{d} + \mathbf{n}, \tag{5.14}$$

where $\mathbf{H}$ is the $N \times N$ channel convolution matrix and $\mathbf{n}$ is the now the $N \times 1$ vector of noise samples after cyclic prefix removal.

Applying the DFT matrix to (5.14), transforms the received time-domain samples, into the frequency-domain,

$$\begin{aligned} \mathbf{X} = \mathbf{F}\mathbf{x} &= \mathbf{F}\mathbf{H}\mathbf{d} + \mathbf{F}\mathbf{n} \\ &= \mathbf{F}\mathbf{H}\mathbf{F}^H\mathbf{D} + \mathbf{N}, \end{aligned} \tag{5.15}$$

where $\mathbf{N}$ is the vector of frequency-domain noise samples that have the same distribution as the time-domain samples, due to $\mathbf{F}$ being unitary. Noting that the use of the cyclic prefix causes $\mathbf{H}$ to be circulant and that the eigenvectors of a circulant matrix are the columns of the DFT matrix [52], leads (5.15) to be rewritten as

$$\mathbf{X} = \tilde{\mathbf{H}}\mathbf{D} + \mathbf{N}, \tag{5.16}$$

where $\tilde{\mathbf{H}}$ is the $N \times N$ diagonal matrix of the eigenvalues of $\mathbf{H}$ [52], and is given by

$$\tilde{\mathbf{H}} = \text{diag}\left(\sqrt{N}\mathbf{F}\left[\mathbf{h}, \quad \mathbf{0}_{1,N-L_h}\right]^T\right). \tag{5.17}$$

Because $\tilde{\mathbf{H}}$ is diagonal, (5.16) can be decomposed and viewed on a per-subcarrier basis, thus the output of the DFT on the k^{th} subcarrier is given as

$$X_k = \tilde{H}_k D_k + N_k. \tag{5.18}$$

This decomposition, allows estimates of the transmitted data symbols, $\hat{D}_k$, to be obtained by scaling (5.18) by the frequency-domain channel value, $\tilde{H}_k$,

$$\begin{aligned} \hat{D}_k = \frac{1}{\tilde{H}_k}X_k &= \frac{1}{\tilde{H}_k}\left(\tilde{H}_k D_k + N_k\right) \\ &= D_k + \frac{N_k}{\tilde{H}_k}. \end{aligned} \tag{5.19}$$

This is the definition of one-tap frequency-domain equalization (FDE).

5.2 Mulitcarrier Systems in the Presence of Narrowband Interference

Recall the interference term given in (1.25), can be written as

$$i(t) = \sqrt{E_i} e^{j(2\pi f_i t + \theta)}. \tag{5.20}$$

Sampling the interference term at the N equally spaced instances as in (5.2) and letting the frequency of the interference term be defined as $f_i = (m + \alpha)/T_s$ gives

$$\begin{aligned} i_l = i(t)|_{t=lT_s/N} &= \sqrt{E_i} e^{j\left(2\pi \frac{m+\alpha}{T_s} l \frac{T_s}{N} + \theta\right)} \\ &= \sqrt{E_i} e^{j\left(\frac{2\pi}{N}(m+\alpha)l + \theta\right)}, \end{aligned} \tag{5.21}$$

where m is the subcarrier closest to the interference and α is the offset position of the interferer from tone m, distributed uniformly as $\mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$.

This interference term is part of the received signal, previously defined in (5.13) and (5.14), where it is now

$$\mathbf{x}_{\text{cp}} = \mathbf{h} * \mathbf{d}_{\text{cp}} + \mathbf{i}_{\text{cp}} + \mathbf{n}_{\text{cp}}, \tag{5.22}$$

where $\mathbf{i}_{\text{cp}}$ is the $(N + N_g) \times 1$ vector of interference samples. Note that $\mathbf{i}_{\text{cp}}$ and $\mathbf{n}_{\text{cp}}$ are not cyclically extended as the data is. After cyclic prefix removal, the time-domain samples are transformed into the frequency-domain using the DFT,

$$\begin{aligned} \mathbf{X} = \mathbf{F}\mathbf{x} &= \mathbf{F}\mathbf{H}\mathbf{d} + \mathbf{F}\mathbf{i} + \mathbf{F}\mathbf{n} \\ &= \tilde{\mathbf{H}}\mathbf{D} + \mathbf{I} + \mathbf{N}, \end{aligned} \tag{5.23}$$

where $\mathbf{I}$ is the vector of interference samples in the frequency-domain. Of interest is the interference that is located on subcarrier k, denoted as I_k. When the interferer is non-orthogonal to the subcarriers (i.e. not located directly on a tone, $\alpha \neq 0$), the interference is spread across all the tones, which is described as

$$I_k = \sqrt{\frac{E_i}{N}} \frac{\left(1 - e^{j2\pi\alpha}\right) e^{j\theta}}{1 - e^{j\frac{2\pi}{N}(m+\alpha-k)}}. \tag{5.24}$$

This effect is termed spectral leakage and arises from the fact that the DFT is of finite length. The interference is multiplied by a rectangular window[1] which is equivalent to

[1] Other windows may be chosen to reduce the impact of the spectral leakage, however in this work, other means are examined to mitigate the interference.

convolution by a sinc function in the frequency-domain. This representation is similar to the case when ICI is present. When the interference is orthogonal only to tone m (i.e. $\alpha = 0$), the interference after the DFT reduces to

$$I_k = \begin{cases} \sqrt{NE_i}e^{j\theta}, & k = m, \\ 0, & k \neq m. \end{cases} \tag{5.25}$$

The interference impacts only subcarrier m in this case.

5.2.1 Definition of Signal-to-Interference Ratio

It is necessary to examine the definition of the term, signal-to-interference ratio (SIR) in the context of a multi-carrier system that is corrupted by narrowband interference. Specifically, a resolution is needed since the data is defined in the frequency-domain and the interference is defined in the time-domain. The uncoded case (i.e. $R_c = 1$) is examined here for simplicity.

The frequency-domain data vector given in (5.9) is composed of QAM symbols, with each subcarrier having a symbol of average energy, E_s. The average power of the received signal over the spectrum (i.e. across all subcarriers) is equal to $N \times E_s/N = E_s$, using $\mathbb{E}|H_k|^2 = 1$ (from (5.12) and Parseval's Theorem [97]) and (5.17).

Both interference cases must be examined. Looking first at the orthogonal case ($\alpha = 0$), the interference power per subcarrier is given by

$$\sigma^2_{I,k} = \mathbb{E}\left[I_k I_k^*\right] = \begin{cases} NE_i, & k = m, \\ 0, & k \neq m. \end{cases} \tag{5.26}$$

The average power of the interference is equal to $\frac{1}{N}\left(1 \times NE_i + (N-1) \times 0\right) = E_i$. Note that the multiplication by the factor of N can be viewed as a form of processing gain, allowing weaker interferences to be detected more easily through DFT processing.

The case of a non-orthogonal interferer is more involved. The interference power per subcarrier is given by,

$$\sigma^2_{I,k} = \mathbb{E}\left[I_k I_k^*\right] = \frac{E_i}{N} \frac{1 - \cos 2\pi\alpha}{1 - \cos \frac{2\pi}{N}\left(m + \alpha - k\right)}. \tag{5.27}$$

The average interference power over the spectrum is found by summing over all the subcarriers and normalizing by the DFT length. It can be shown (see Appendix) that

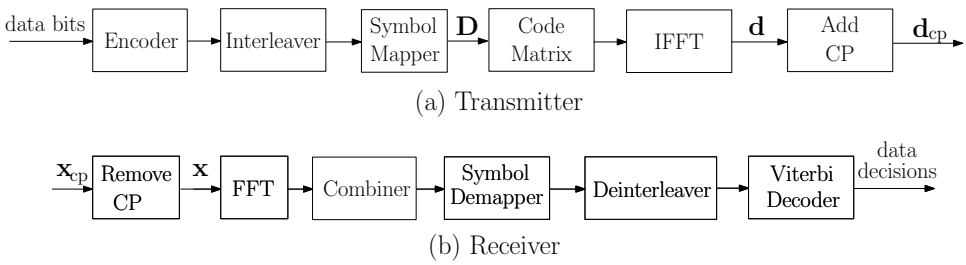


Figure 5.2: BICM MC-CDMA system model.

the following expression is true,

$$\frac{1}{N}\sum_{k=0}^{N-1}\frac{E_i}{N}\frac{1-\cos 2\pi\alpha}{1-\cos\frac{2\pi}{N}(m+\alpha-k)} = E_i. \tag{5.28}$$

Thus, for both cases, the SIR can be reliably defined as E_s/E_i.

5.3 MC-CDMA

The above discussion relates to general block modulated multi-carrier schemes, however, it precisely describes the implementation of OFDM systems. For the case of MC-CDMA systems, each data symbol is spread into all of the subcarriers. This is accomplished by using a code (spreading) matrix, $\mathbf{B}$, which is an $N \times N$ unitary matrix whose columns contain the N spreading sequences. The k^{th} column is responsible for spreading the k^{th} data symbol, D_k. A block diagram for this system can be seen in Figure 5.2. Note that when $\mathbf{B}$ is the identity matrix (i.e. $\mathbf{B} = \mathbf{I}_N$), the system is identical to OFDM. Thus, OFDM is a special case of MC-CDMA.

Taking into account the code matrix, the received signal given in (5.15) can be be written as

$$\begin{aligned}\mathbf{X} &= \mathbf{FHF}^H\mathbf{BD} + \mathbf{N} \\ &= \tilde{\mathbf{H}}\mathbf{BD} + \mathbf{N}.\end{aligned} \tag{5.29}$$

The design of the code matrix requires the columns of $\mathbf{B}$ to be orthogonal to successfully recover the transmitted data symbols with no cross-talk between subcarriers. One choice for the orthogonal spreading codes is the Walsh-Hadamard code matrix [97]. However, a drawback of this choice is that these codes are only defined for lengths of $N = 2^n$, where n is an integer greater than zero [42]. Another choice that is utilized in this work, is

using the DFT basis vectors as the columns of $\mathbf{B}$ [90,137] and is defined component-wise as,

$$[\mathbf{B}]_{kl} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}kl}, \quad 0 \leq k, l \leq N-1. \tag{5.30}$$

Note that (5.30) is the same as (5.4) and thus $\mathbf{B}$ is a unitary matrix. This choice allows more flexibility because it can be determined for any choice of N.

Spreading each data symbol into all the subcarriers requires a combiner at the receiver to extract the data from the received signal. The combiner weights are derived under the MMSE criterion [137,138] and obtained by solving the Wiener-Hopf equations,

$$\mathbf{W} = \mathbf{R}_X^{-1}\mathbf{R}_{XD}, \tag{5.31}$$

where $\mathbf{R}_X$ is the autocorrelation matrix of the received signal and $\mathbf{R}_{XD}$ is the cross-correlation vector between the received signal and the data vector. Assuming that the channel is known, the autocorrelation matrix can be found to be

$$\begin{aligned}
\mathbf{R}_X &= \mathbb{E}\left[\left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{N}\right)\left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{N}\right)^H\right] \\
&= \tilde{\mathbf{H}}\mathbf{B}\mathbb{E}\left[\mathbf{D}\mathbf{D}^H\right]\mathbf{B}^H\tilde{\mathbf{H}}^H + \mathbb{E}\left[\mathbf{N}\mathbf{N}^H\right] \\
&= E_s\tilde{\mathbf{H}}\mathbf{B}\mathbf{B}^H\tilde{\mathbf{H}}^H + \sigma_n^2\mathbf{I}_N \\
&= E_s\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \sigma_n^2\mathbf{I}_N \\
&= E_s\left|\tilde{\mathbf{H}}\right|^2 + \sigma_n^2\mathbf{I}_N.
\end{aligned} \tag{5.32}$$

It can be seen from (5.32) that $\mathbf{R}_X$ is an $N \times N$ diagonal matrix, where the k^{th} diagonal component is given by

$$R_X(k,k) = E_s\left|\tilde{H}_k\right|^2 + \sigma_n^2, \quad 0 \leq k \leq N-1. \tag{5.33}$$

The cross-correlation vector can be found in similar way,

$$\begin{aligned}
\mathbf{R}_{XD} &= \mathbb{E}\left[\left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{N}\right)\mathbf{D}^H\right] \\
&= \tilde{\mathbf{H}}\mathbf{B}\mathbb{E}\left[\mathbf{D}\mathbf{D}^H\right] \\
&= E_s\tilde{\mathbf{H}}\mathbf{B}.
\end{aligned} \tag{5.34}$$

Solving (5.31), gives the combiner weights to be

$$\mathbf{W} = \text{diag}\left(\left[\frac{E_s\tilde{H}_0}{E_s\left|\tilde{H}_0\right|^2+\sigma_n^2}, \quad \frac{E_s\tilde{H}_1}{E_s\left|\tilde{H}_1\right|^2+\sigma_n^2}, \quad \cdots, \quad \frac{E_s\tilde{H}_{N-1}}{E_s\left|\tilde{H}_{N-1}\right|^2+\sigma_n^2}\right]\right)\mathbf{B}. \tag{5.35}$$

Finally, the estimates of the transmitted data symbols are given by

$$\hat{\mathbf{D}} = \mathbf{W}^H \mathbf{X}. \tag{5.36}$$

Note that a consequence of MMSE combining in frequency-selective channels is that there is a loss of orthogonality among the subcarriers,

$$\begin{aligned}\hat{\mathbf{D}} &= \mathbf{W}^H \mathbf{X} = \mathbf{W}^H \left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{N}\right) \\ &= \mathbf{W}^H \tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{W}^H \mathbf{N} \\ &= \mathbf{D} + \left(\mathbf{W}^H \tilde{\mathbf{H}}\mathbf{B} - \mathbf{I}_N\right)\mathbf{D} + \mathbf{W}^H \mathbf{N},\end{aligned} \tag{5.37}$$

where the second term represents the ICI. When the channel is AWGN or flat fading, equal gain combining (i.e. $\mathbf{W} = \mathbf{B}$) can be utilized to preserve orthogonality,

$$\begin{aligned}\hat{\mathbf{D}} &= \mathbf{D} + \left(\mathbf{W}^H \rho \mathbf{I}_N \mathbf{B} - \mathbf{I}_N\right)\mathbf{D} + \mathbf{W}^H \mathbf{N} \\ &= \mathbf{D} + \left(\rho \mathbf{W}^H \mathbf{B} - \mathbf{I}_N\right)\mathbf{D} + \mathbf{W}^H \mathbf{N} \\ &= \rho \mathbf{D} + \mathbf{W}^H \mathbf{N},\end{aligned} \tag{5.38}$$

where ρ is dependent on the channel (i.e. $\rho = 1$ for an AWGN channel).

5.4 Convolutional Coding

Reliable wireless communication links require the use of forward error correction in practical scenarios. In this work, convolutional codes are examined in conjunction with multi-carrier modulation. These codes convert an input stream (possibly infinite) into a single codeword. This differs from block codes that segment the input stream into a fixed block length and output a codeword of a fixed length.

Convolutional coding was first introduced by Elias [39], and it was shown that choosing random codes based on linear shift registers worked well. Ideally, a maximum-likelihood (ML) estimate could be formed by examining all possible codewords with the received vector. Unfortunately, the complexity of this approach increases exponentially with $k_c L$, where k_c is the number of bits inputted into the encoder and L is the number of bits in the sequence. This required the research into new decoding algorithms. A class of sequential algorithms evolved through a number of iterations [40, 65, 83, 136]. Finally, Viterbi [128] discovered a maximum-likelihood decoding algorithm [43] by minimizing a path through a weighted, directed graph [93]. This dynamic programming algorithm was shown to be asymptotically optimal [128].

5.4.1 Encoder

The convolutional encoder is composed of a linear finite-state shift register. In general, the shift register consists of K_c stages, where K_c is the constraint length of the code. This term relates the number of output bits that are affected by one input bit. The binary input stream is inputted k_c bits at a time, producing an output of n_c bits. The code rate is then defined as $R_c = k_c/n_c$, which is similar to the case for block coding. The minimum distance of the code, $d_{\min}$, is calculated as the minimum Hamming distance between all pairs of codewords. Note that these encoders can be considered as finite impulse response (FIR) digital filters or as finite state machines.

The encoder can be described by its generator polynomial. This is obtained by inputting a delta function (i.e. a single one followed by a string of zeros) to provide the impulse response of the encoder at the output. This polynomial is also apparent from the block diagram of the structure, simply by examining the tap sequence of the encoder. Using these polynomials, a generator matrix can be formed as is done in block coding, however, because the length of the input sequence may not be bounded, the generator matrix may be semi-infinite.

It is assumed that the shift register is initialized to the all-zero state. To aid in the decoding of the codeword, it is necessary to assume that the final state of the shift register is also the all-zero state. To accomplish this, an additional $(K_c - 1)k_c$ zeros are appended to the end of the input stream. These input bits result in coded bits that do not contain any useful information and this overhead reduces the effective rate of the code. This reduction can be characterized in terms of a fractional rate loss as

$$\rho = \frac{(K_c - 1)k_c}{L + (K_c - 1)k_c}. \tag{5.39}$$

There are a number of methods that describe the operation of convolutional codes. These are the tree diagram, the state diagram, transfer functions, and the trellis diagram. In this work the trellis diagram is utilized. A review of the alternate methods can be found in [97, 132].

5.4.2 Decoder

The decoder is implemented using Viterbi's algorithm and uses the trellis diagram to determine the ML estimate. The trellis diagram depicts the state of the shift registers as a function of time. The branches are labeled with the output bits that correspond

with the associated state transitions. For an input sequence of $k_c L$ input bits, there are $2^{k_c L}$ distinct paths.

Consider the information sequence, $\mathbf{d}$, encoded to form the codeword, $\mathbf{x}$, which is then transmitted across the channel. The received codeword, $\mathbf{y}$ is decoded to give an estimate of the transmitted codeword, $\hat{\mathbf{x}}$. The ML decoder looks to maximize the probability, $p(\mathbf{y}|\hat{\mathbf{x}})$. With the assumption of a memoryless channel (i.e. each received bit is independent from the other bits), the ML estimate can be decomposed into the ML estimate for each bit,

$$\begin{aligned} p(\mathbf{y}|\hat{\mathbf{x}}) &= \prod_{i=0}^{L+(K_c-1)-1} \left[p\left(y_i^{(0)}|\hat{x}_i^{(0)}\right) p\left(y_i^{(1)}|\hat{x}_i^{(1)}\right) \ldots \left(y_i^{(n_c-1)}|\hat{x}_i^{(n_c-1)}\right) \right] \\ &= \prod_{i=0}^{L+(K_c-1)-1} \left(\prod_{j=0}^{n_c-1} p\left(y_i^{(j)}|\hat{x}_i^{(j)}\right) \right), \end{aligned} \tag{5.40}$$

where the first product is over the block numbers (set of k_c bits inputted to the encoder) and the second product is over the number of output bits, n_c, for each block. Taking the logarithm of (5.40) also maximizes the ML estimate,

$$\log p(\mathbf{y}|\hat{\mathbf{x}}) = \sum_{i=0}^{L+(K_c-1)-1} \left(\sum_{j=0}^{n_c-1} \log p\left(y_i^{(j)}|\hat{x}_i^{(j)}\right) \right). \tag{5.41}$$

The bit metric can be defined as,

$$\mathcal{M}\left(y_i^{(j)}|\hat{x}_i^{(j)}\right) = \log p\left(y_i^{(j)}|\hat{x}_i^{(j)}\right), \tag{5.42}$$

and the path metric is the sum over all bit metrics,

$$\mathcal{M}(\mathbf{y}|\hat{\mathbf{x}}) = \sum_{i=0}^{L+(K_c-1)-1} \left(\sum_{j=0}^{n_c-1} \mathcal{M}\left(y_i^{(j)}|\hat{x}_i^{(j)}\right) \right). \tag{5.43}$$

Recall that at each node there are 2^{k_c} branches at each node. The path with the best metric at each node is chosen and referred to as the survivor. If the best metric is not unique, the survivor is chosen randomly. Once the algorithm reaches the $L + (K_c - 1)$ iteration, the ML path is the one with the largest metric that ends in the all-zero (final) state and is found by tracing back through the trellis.

This decoding algorithm requires keeping track of $2^{(K_c-1)k_c}$ surviving paths and $2^{(K_c-1)k_c}$ metrics, as well as computing 2_c^k metrics at each node. Thus the number of computations at each stage increases exponentially with k_c and K_c.

5.4.3 Log-Likelihood Ratio

The Viterbi decoder experiences improved performance when the input is composed of soft-decisions [97, 132]. These decisions provide the decoder with a reliability assessment for each bit, unlike hard decisions which are limited to binary values.

The soft-decisions out of the symbol demapper, are simply the log-likelihood ratios (LLRs). For simplicity, the LLR for a BPSK system is calculated as follows,

$$\Lambda(x_l|y_l) = \log \frac{p(x_l = +1|y_l)}{p(x_l = -1|y_l)}. \tag{5.44}$$

This is noted as the ratio of the maximum a posteriori (MAP) probabilities. Using Bayes rule, (5.44) can be rewritten in term of the ML probabilities,

$$\begin{aligned}
\Lambda(x_l|y_l) &= \log \frac{p(x_l = +1|y_l)}{p(x_l = -1|y_l)} \\
&= \log \frac{p(y_l|x_l = +1)p(x_l = +1)}{p(y_l|x_l = -1)p(x_l = -1)} \\
&= \log \frac{p(y_l|x_l = +1)}{p(y_l|x_l = -1)} + \log \frac{p(x_l = +1)}{p(x_l = -1)} \\
&= \log \frac{p(y_l|x_l = +1)}{p(y_l|x_l = -1)}.
\end{aligned} \tag{5.45}$$

The last line of (5.45) comes from the assumption that the a priori probabilities of the encoded bits are equal (i.e. $p(x_l = +1) = p(x_l = -1)$). Evaluation of (5.45) requires the conditional probability density function of the received signal, y_l. This term is used as the input to Viterbi decoder. For larger constellations, refer to [4] for exact and approximate calculations.

5.4.4 Puncturing

The rate of a convolutional code can be increased by puncturing the code. This entails removing bits from the codeword prior to transmission. This increased rate comes at the cost of redundancy and therefore error protection. In particular, there exists codes that are rate-compatible, allowing a number of code rates to be obtained by following a certain puncturing pattern. For example, rate-compatible punctured codes are seen in the standards for WLAN [63] and UWB [38]. At the decoder, erasures (LLRs of zero) are inserted at the locations of the punctured bits. The decoding follows the process as described in Section 5.4.2.

5.4.5 Interleaving

An interleaver is generally used in conjunction with forward error correction coding to prevent a burst of errors from entering the decoder. Convolutional codes are sensitive to burst errors, whereas they are well suited for handling well-spaced errors. These errors can occur from fading channels that have deep fades over a number of bits. This also implies that the received bits are not usually independent as assumed in (5.40). The interleaver can make the received signal look independent to the decoder. This can improve the performance by providing diversity to the system. At the receiver, a deinterleaver is used to invert the interleaving process.

In this work an $(n \times m)$ block interleaver is considered, which consists of n rows and m columns. A similar block interleaver is used in WLAN standard [63]. The coded data is read row by row. The output of the interleaver is read out column-wise. Each symbol is then separated from neighboring symbols by $n - 1$ other symbols. Note that the structure of the deinterleaver is exactly the same as that of the interleaver.

5.4.6 Summary

Chapter 5, in part, is a reprint of material as it appears in A. Batra, J. R. Zeidler, J. G. Proakis, and L. B. Milstein, "Interference Rejection and Management," in *New Directions in Wireless Communications Research*, V. Tarokh, Ed. New York: Springer, 2009. The dissertation author was the primary investigator and author of this paper.

6 Mitigating Narrowband Interference in Block Modulated Multi-carrier Systems

In this chapter, the cases of a multi-carrier code division multiple access (MC-CDMA) and orthogonal frequency division multiplexing (OFDM) system are considered in the presence of strong NBI (i.e., SIR $\ll$ 0 dB) operating in a frequency-selective channel. The optimal combining weights for MC-CDMA are first provided. It is then shown that coding with erasure insertion (using a Bayesian method) is not adequate for a severe non-orthogonal interferer in an OFDM system because a large number of erasures are required when the interference power leaks (i.e. spectral leakage) into a large number of subcarriers [7]. The prediction-error filter (PEF) is proposed as an erasure insertion mechanism that localizes erasures around the interference by placing a notch in the frequency spectrum of the received signal [8]. As discussed earlier, the PEF uses the narrowband nature of the interference (as compared to the OFDM spectrum) to remove the interference and can be easily implemented adaptively using the low complexity least-mean square (LMS) algorithm. This is an example of excision filtering as proposed by Coulson [29], however this implementation considers a filter that can be obtained adaptively as well as limited to the portion of the cyclic prefix that is not allocated toward mitigating intersymbol interference (ISI). This work also differs in that frequency-selective fading is examined with explicit forward error correction coding. An upper bound on the bit error rate (BER) is determined for the case of coding and prediction-error filtering. The simulation demonstrates excellent results between the case of no interference and when interference is present and mitigated by the PEF. Finally, a hybrid system [7] that uses both erasure insertion and the PEF depending on the signal-to-

interference ratio (SIR) is examined and it is seen that the PEF outperforms erasure insertion over almost all SIR values considered.

6.1 Optimal Combining Weights for MC-CDMA

When narrowband interference is present, the received signal in the frequency-domain after cyclic prefix removal is given by

$$\mathbf{X} = \tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{I} + \mathbf{N}, \tag{6.1}$$

where $\mathbf{I}$ is the $N \times 1$ frequency-domain vector of interference samples. The autocorrelation matrix of (6.1) is defined as

$$\begin{aligned}\mathbf{R}_X &= \mathbb{E}\left[\left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{I} + \mathbf{N}\right)\left(\tilde{\mathbf{H}}\mathbf{B}\mathbf{D} + \mathbf{I} + \mathbf{N}\right)^H\right] \\ &= \tilde{\mathbf{H}}\mathbf{B}\mathbb{E}\left[\mathbf{D}\mathbf{D}^H\right]\mathbf{B}^H\tilde{\mathbf{H}}^H + \mathbb{E}\left[\mathbf{I}\mathbf{I}^H\right] + \mathbb{E}\left[\mathbf{N}\mathbf{N}^H\right] \\ &= E_s\left|\tilde{\mathbf{H}}\right|^2 + \mathbf{R}_I + \sigma_n^2\mathbf{I}_N,\end{aligned} \tag{6.2}$$

where $\mathbf{R}_I$ is the autocorrelation matrix of the interference samples in the frequency-domain defined as

$$\mathbf{R}_I = \mathbf{F}\mathbf{R}_i\mathbf{F}^H, \tag{6.3}$$

and where $\mathbf{R}_i$ is the time-domain correlation matrix of the interference, its entries given by

$$r_i(l) = E_i e^{j\frac{2\pi}{N}(m_i+\alpha_i)l}. \tag{6.4}$$

When the interference is orthogonal to the subcarriers (except for the m_i^{th} subcarrier), $\alpha_i = 0$, then $\mathbf{I}$ will have one non-zero component, at index m_i. Thus, $\mathbf{R}_I$ will also only have one non-zero value, equal to NE_i at position (m_i, m_i). Let $\mathbf{w} = [w(0), \ldots, w(N-1)]$, where $w(l)$ is given by

$$w(l) = \begin{cases} \frac{E_s\tilde{H}_l}{E_s\left|\tilde{H}_l\right|^2+\sigma_n^2}, & l \neq m_i, \\ \frac{E_s\tilde{H}_l}{E_s\left|\tilde{H}_l\right|^2+\sigma_n^2+NE_i}, & l = m_i. \end{cases} \tag{6.5}$$

The MMSE weights are then given as

$$\mathbf{W} = \text{diag}\left(\mathbf{w}\right)\mathbf{B}. \tag{6.6}$$

It can be seen from (6.6) that when interference is not present on the tone, the weight is the same as for the no interference case. Note that when $l = m$ and the narrowband interference is strong, then $W_k(l)$ is essentially zero, thus ignoring the subcarrier that is unreliable and providing a form of erasure insertion.

For the case of a nonorthogonal interferer, the components of the autocorrelation matrix for the interference are given by

$$R_I(k,l) = \frac{E_i}{N} \frac{2 - 2\cos(2\pi\alpha_i)}{1 - e^{j\frac{2\pi}{N}(m_i+\alpha_i-k)} - e^{-j\frac{2\pi}{N}(m+\alpha-l)} + e^{j\frac{2\pi}{N}(l-k)}}, \quad k,l = 0,\ldots,N-1. \tag{6.7}$$

To find the weights one could use (6.7) and invert (5.31). However, since $\mathbf{R}_I$ is rank-1, having one dominant eigenvalue, $\mathbf{R}_I$ can be approximated using the singular value decomposition (SVD) [56] as

$$\begin{aligned} \mathbf{R}_I &= \mathbf{U}\mathbf{S}\mathbf{V}^H \\ &\approx \lambda_1 \mathbf{u}\mathbf{v}^H, \end{aligned} \tag{6.8}$$

where $\lambda_1 = NE_i$ is the non-zero eigenvalue of $\mathbf{R}_I$, $\mathbf{u}$ is the first column of $\mathbf{U}$, and $\mathbf{v}$ is the first column of $\mathbf{V}$. Then using the Sherman-Morrison formula [56], the MMSE combiner weights are derived to be

$$\mathbf{W} = \left(\mathbf{A}_1 - \frac{\lambda_1 \mathbf{A}_1 \mathbf{u}\mathbf{v}^H \mathbf{A}_1}{1 + \lambda_1 \mathbf{v}^H \mathbf{A}_1 \mathbf{u}} \right) E_s \tilde{\mathbf{H}} \mathbf{B}, \tag{6.9}$$

where $\mathbf{A}_1 = \text{diag}\left(\left[\frac{1}{E_s|\tilde{H}_0|^2+\sigma_n^2}, \quad \frac{1}{E_s|\tilde{H}_1|^2+\sigma_n^2}, \quad \cdots, \quad \frac{1}{E_s|\tilde{H}_{N-1}|^2+\sigma_n^2} \right] \right)$.

Again, the estimates of the transmitted data symbols are given by

$$\hat{\mathbf{D}} = \mathbf{W}^H \mathbf{X}. \tag{6.10}$$

6.2 Conventional Erasure Insertion for OFDM

A technique that can make a coded OFDM system more robust to narrowband interference is erasure insertion. The idea is to insert an erasure, which is a LLR representing no information, on the tones that are suspected of experiencing interference prior to decoding. A similar approach is discussed in [118], while a more robust approach to detection and decoding is proposed in [76].

The Bayesian erasure insertion rule proposed by Baum and Pursely [10] is considered in this work. An erasure is inserted if the following expression is satisfied:

$$\frac{\max_j \pi_j f(X_k|D_j)}{\sum_{i=0}^{M-1} \pi_i f(X_k|D_i)} \leq (1-\gamma), \tag{6.11}$$

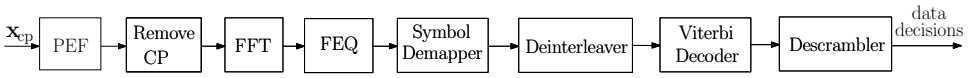


Figure 6.1: BICM OFDM system model with the PEF.

where π_i is the a priori probability of transmitting D_i, $f(X_k|D_i)$ is the conditional probability density function (pdf) of X_k given that D_i was transmitted, M is the size of the constellation, and $\gamma \in [0, 1]$ is a threshold. Note that $\gamma = 1$ is the case of no erasure insertion.

A genie-aided system is considered, where the interference power on each tone, $\sigma^2_{I,k}$, is known at the receiver. The sum of the noise and interference is then assumed to be distributed as a zero-mean Gaussian random variable with variance given by $\sigma^2_k = \sigma^2_n + \sigma^2_{I,k}$. The conditional probability density function per tone in (6.11) is given by

$$f(X_k|D_i) = \frac{1}{\sqrt{2\pi\sigma^2_k}} e^{-\frac{(X_k - D_i)^2}{2\sigma^2_k}}. \tag{6.12}$$

Note that the genie provides the information of the interference only to the erasure insertion rule. The LLRs calculated by the decoder do not have any knowledge of the interference statistics.

6.3 Erasure Insertion using the Prediction-Error Filter for OFDM

6.3.1 Receiver with Prediction-Error Filter

The PEF is implemented before the removal of the cyclic prefix as seen in Figure 6.1. Let the one-step predictor weights from (2.64) be defined in (2.67) and the convolution of the filter and channel be defined as $\mathbf{a} = \mathbf{w}_{\text{PEF}} * \mathbf{h}$. It is assumed that the overall length of $\mathbf{a}$ is less than the length of the cyclic prefix (i.e. $L_h + M_{\text{pef}} - 1 \leq N_g$) to ensure that there is no ISI or intercarrier interference (ICI). Therefore, the effective channel matrix, $\mathbf{A}$, is circulant. Then, the filtered signal in the frequency-domain can be written as

$$\begin{aligned}\mathbf{X} &= \mathbf{F}\mathbf{A}\mathbf{F}^H\mathbf{D} + \mathbf{F}\mathbf{Z}\left(\mathbf{i}_{\text{cp}} + \mathbf{n}_{\text{cp}}\right) \\ &= \tilde{\mathbf{A}}\mathbf{D} + \mathbf{F}\mathbf{Z}\left(\mathbf{i}_{\text{cp}} + \mathbf{n}_{\text{cp}}\right), \end{aligned} \tag{6.13}$$

where $\tilde{\mathbf{A}}$ is the diagonal matrix of the eigenvalues of $\mathbf{A}$, $\mathbf{i}_{\text{cp}}$ and $\mathbf{n}_{\text{cp}}$ are length-$(N+N_g)$ vectors of interference and noise samples, respectively that are not cyclically extended and $\mathbf{Z}$ is the $N \times (N+N_g)$ filtering matrix for the noise and interference that is defined as

$$\mathbf{Z} = \left[\mathbf{0}_{N,N_g-M_{\text{pef}}},\ \text{Toeplitz}\left(\left[w^*_{\text{PEF},M_{\text{pef}}}, \mathbf{0}_{1,N-1}\right]^T, \left[\mathbf{P}\mathbf{w}^*_{\text{PEF}}, \mathbf{0}_{1,N-1}\right]\right)\right], \tag{6.14}$$

where the Toeplitz operator, Toeplitz $(column, row)$, generates a Toeplitz matrix from a column vector and a row vector and $\mathbf{P}$ is an $(M_{\text{pef}}+1) \times (M_{\text{pef}}+1)$ permutation matrix that has unity on the anti-diagonal, while the remaining entries are zero. Note that $\tilde{\mathbf{A}}$ in (6.13) can also be defined as

$$\tilde{\mathbf{A}} = \tilde{\mathbf{H}}\tilde{\mathbf{W}}, \tag{6.15}$$

where $\tilde{\mathbf{W}} = \sqrt{N}\mathbf{F}\left[\mathbf{w}^*_{\text{PEF}} \quad \mathbf{0}_{1,N-(M_{\text{pef}}+1)}\right]^T$ is the sampled frequency response of the notch filter and $\tilde{\mathbf{H}}$ is given by (5.17).

Let the uncanceled interference and noise be grouped into one general noise term,

$$\tilde{\mathbf{N}} = \mathbf{F}\mathbf{Z}\left(\mathbf{i}_{\text{cp}} + \mathbf{n}_{\text{cp}}\right). \tag{6.16}$$

It is clear from (6.16) that the noise samples, $\tilde{N}_k$, are correlated due to the PEF matrix, $\mathbf{Z}$. It is also noted that the noise samples $\tilde{N}_k$ are assumed to be Gaussian random variables. As stated in [74], this system is difficult to analyze when the noise samples are not strictly independent, however, due to the fact that the noise power in the main tap ($w_{\text{PEF},0} = 1$) is much larger than in the remaining taps (approximately $\frac{1}{M_{\text{pef}}}$ for the scenario of interest in this paper), it is reasonable to assume that the noise samples are independent, especially when M_{pef} is large. Note that K reaches a maximum of $1/M_{\text{pef}}$ and as the SIR increases, K decreases toward zero. Therefore, let $\sigma^2_{\tilde{N},k}$ be the variance

of the noise on tone k, given by

$$\begin{aligned}\sigma^2_{\tilde{N},k} &= \mathbb{E}\left[\tilde{N}_k \tilde{N}_k^H\right] \\ &= \left[\mathbf{FZ}\left(\mathbf{R}_i + \sigma_n^2 \mathbf{I}_N\right)\mathbf{Z}^H\mathbf{F}^H\right]_{kk} \\ &= \left[\mathbf{FZR}_i\mathbf{Z}^H\mathbf{F}^H\right]_{kk} + \left[\sigma_n^2\mathbf{FZZ}^H\mathbf{F}^H\right]_{kk} \\ &= \sigma^2_{\tilde{i},k} + \sigma^2_{\tilde{n},k}. \end{aligned} \tag{6.17}$$

Note that the variances given in (6.17) are scaled according to the notch filter that is used to suppress the interference.

Finally, one-tap equalization is performed and the estimates of the transmitted data symbols are given by

$$\hat{\mathbf{D}} = \mathbf{X}/\mathrm{diag}\left(\tilde{\mathbf{A}}\right). \tag{6.18}$$

Note that this equalization does not affect the independence assumption because $\tilde{\mathbf{A}}$ is a diagonal matrix.

An analytic expression for the BER of this OFDM system in the case of BPSK uncoded transmission is simply the average of the per-subcarrier BERs, given by

$$P_b = \frac{1}{N}\sum_{k=0}^{N-1} Q\left(\sqrt{\frac{|\tilde{A}_k|^2 E_b}{\sigma^2_{\tilde{N},k}}}\right), \tag{6.19}$$

where $Q(\cdot)$ is the well known Q-function [97]. Simulation results are shown with the results of (6.19) in Figure 6.2 for an AWGN channel. It is clear that the simulation and theoretical results agree quite well, validating the assumption of independent noise samples. This is true for longer PEF lengths, $M_{\text{pef}} = 16$ (using the entire cyclic prefix), and even when $M_{\text{pef}} = 4$. It is also apparent, that a longer PEF provides improved performance. This arises from the fact the PEF is a whitening filter, and becomes ideal as M_{pef} approaches infinity.

6.3.2 LLR Calculation

Conventional decoding of convolutional codes make the assumption that the noise variance is constant over all inputted symbols. If this is the case, the noise term can be dropped from the LLR calculation. However, the noise variance, (6.16), is not constant for all subcarriers because the notch formed by the PEF attenuates the tones closest to the interference. This results in metric mismatch when decoding [76].

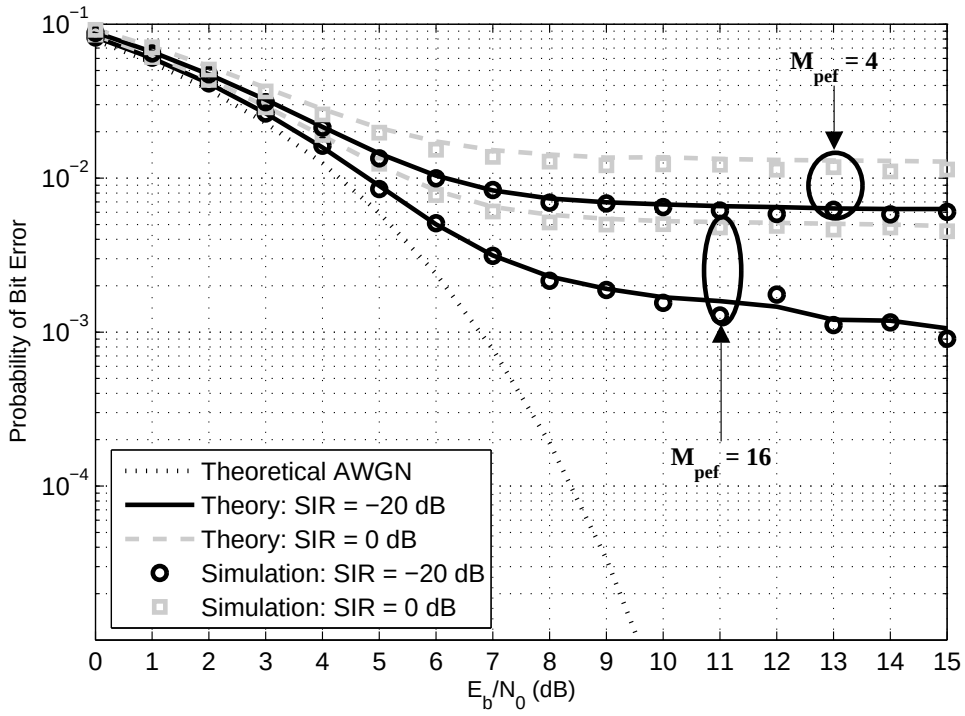


Figure 6.2: Probability of bit error versus E_b/N_0 for the uncoded scenario, comparing theory and simulation for SIR $= 0, -20$ dB and $M_{\text{pef}} = 4, 16$.

In this scenario, it is important to note that the interference statistics (i.e. $\mathbf{R}_i$) are not known when calculating the LLR, thus the term, $\sigma^2_{\tilde{N},k}$, is not available. For example, the LLR inputted into the Viterbi decoder when BPSK modulation is utilized, is given by

$$\Lambda(D_k|\hat{D}_k) = \frac{2\Re\{|\tilde{A}_k|^2\hat{D}_k\}}{\sigma^2_{\tilde{n},k}}, \tag{6.20}$$

where $\Re\{\cdot\}$ is the real operator and $\sigma^2_{\tilde{n},k}$ is simply the variance of the AWGN after filtering, defined in (6.17). The LLR can also be written in an equivalent manner as

$$\Lambda(D_k|X_k) = \frac{2\Re\{A^*_k X_k\}}{\sigma^2_{\tilde{n},k}}. \tag{6.21}$$

Notice that this is the same calculation performed on a per-antenna basis when using maximal-ratio combining [97].

6.3.3 Performance Analysis

The error rate performance of the Viterbi algorithm is considered after deinterleaving the frequency-domain bits. An upper bound on the bit error rate for a convolutional code of rate $R_c = k_c/n_c$ is obtained using the union bound, given as

$$P_b \leq \frac{1}{k_c}\sum_{d=d_{\min}}^{\infty} w_d P_2(d), \tag{6.22}$$

where w_d is the total information weight of all information sequences which produce codewords of weight d. The pairwise error probability (PEP), $P_2(d)$, is the probability of an error event occurring between the all-zero codeword (assumed to be transmitted) and a codeword that differs by d bits. The use of the all-zero codeword arises from the linearity of the code, though may not be possible for certain constellation sizes [118].

For the case of no interference/erasures, the PEP for independent Rayleigh fading channels can be found in closed-form [15, 97] as

$$P_2(d) = \left(\frac{1-\psi}{2}\right)^d \sum_{n=0}^{d-1}\binom{d-1+n}{n}\left(\frac{1+\psi}{2}\right)^n, \tag{6.23}$$

where ψ is defined as

$$\psi = \sqrt{\frac{R_c E_b/N_0}{1+R_c E_b/N_0}}. \tag{6.24}$$

This result is the same as in the case of maximal-ratio combining (as mentioned in Section 6.3.2). This PEP calculation requires averaging out a chi-square random variable with $2d$ degrees of freedom [97, pp. 824-825].

For a general case of channel and coding conditions, the PEP has been determined in [16, 24]. A brief review of the analysis is provided here for completeness. The PEP is the probability of choosing a different codeword ($\hat{\mathbf{C}}$) instead of the transmitted codeword ($\mathbf{C}$) from the received vector ($\mathbf{R}$). This probability is determined when the metric $\mathcal{M}(\mathbf{R}, \hat{\mathbf{C}})$ is greater than $\mathcal{M}(\mathbf{R}, \mathbf{C})$. Letting $\Gamma = \mathcal{M}(\mathbf{R}, \mathbf{C}) - \mathcal{M}(\mathbf{R}, \hat{\mathbf{C}})$, allows the PEP to be written as

$$P_2(d) = \Pr(\Gamma < 0). \tag{6.25}$$

This probability can be obtained by noting that the Laplace transform of the random variable, Γ, is defined as

$$\Phi_\Gamma(s) = \mathbb{E}\left[e^{-s\Gamma}\right] = \int_{-\infty}^{\infty} e^{-s\Gamma} p_\Gamma(x) dx, \tag{6.26}$$

where $p_\Gamma(x)$ is the probability density function of Γ. The PEP is then given as

$$\Pr(\Gamma < 0) = \frac{1}{2\pi j} \int_{a-j\infty}^{a+j\infty} \Phi_\Gamma(s) \frac{ds}{s}, \tag{6.27}$$

where a is a positive real number within the convergence region of $\Phi_\Gamma(s)$.

This integral can be calculated exactly [24] using the residue theorem [121, Appendix A.3], giving the PEP as

$$\Pr(\Gamma < 0) = \begin{cases} 1 + \sum_{\text{LH poles}} \text{Residue}\left[\frac{\Phi_\Gamma(s)}{s}\right], \\ \sum_{\text{RH poles}} \text{Residue}\left[\frac{\Phi_\Gamma(s)}{s}\right], \end{cases} \tag{6.28}$$

where the first summation occurs over the left-hand poles in the complex s-plane and the second summation is over the right-hand poles.

In general, when the system is particularly complicated (as in the case of convolutional coding), (6.28) may not be tractable. For example, if $\Phi_\Gamma(s)$ has multiple poles or essential singularities, the calculation of (6.28) requires a number of derivatives.

To make the calculation of the PEP more tractable, [16] proposes a number of options. The PEP can be upper bounded by the well-known Chernoff bound. This involves minimizing $\Phi_\Gamma(s)$. Another option involves a numerical calculation using Gauss-Chebyshev quadrature rules. The PEP is then given by

$$\Pr(\Gamma < 0) \approx \frac{1}{n} \sum_{k=1}^{n/2} \left[\Re\left\{\Phi_\Gamma\left(c + jc\tau_k\right)\right\} + \tau_k \Im\left\{\Phi_\Gamma\left(c + jc\tau_k\right)\right\}\right], \tag{6.29}$$

where $\tau_k = \tan\frac{(2k-1)\pi}{2n}$, $\Im\{\cdot\}$ is the imaginary operator, and n is an even number. The value, c affects the choice of the value n, and is often chosen such that $\Phi_\Gamma(c)$ is a minimum. See [16] for further discussion.

Further bounding the probability of error for this convolutional coded system using the PEF is very difficult. This arises from the fact that the PEF does not insert true erasures. However, a simplified analysis is obtained if the notched values are assumed to be erased, similar to Li, et al. [76], while the remaining subcarriers are assumed to be unaffected by the filter. As described in [44], the erasures cause a reduction in the weight of the codeword, that is, a codeword of Hamming weight d can have its weight reduced by erasures to $d - d'$, for $0 \leq d' \leq d$. The erasures are considered to be inserted randomly (due to interleaving) with probability, $\delta = N_E/N$, where N_E is the number of erasures per OFDM symbol. The probability of weight reduction to $d - d'$ is given by $\binom{d}{d'}\delta^{d'}(1-\delta)^{d-d'}$. The average PEP for a d-weight codeword with erasures inputted with probability, δ, is then given by

$$\overline{P_2(d)} = \sum_{d'=0}^{d} \binom{d}{d'} \delta^{d'} (1-\delta)^{d-d'} P_2(d-d'). \tag{6.30}$$

After some manipulations, the average probability of error is upper bounded as

$$\overline{P_b} \leq \frac{1}{k} \sum_{d=d_{\min}}^{\infty} w_d \overline{P_2(d)} = \frac{1}{k} \sum_{d=d_{\min}}^{\infty} w_d \sum_{d'=0}^{d} \binom{d}{d'} \delta^{d'} (1-\delta)^{d-d'} P_2(d-d'). \tag{6.31}$$

This equation can be further rewritten as

$$\overline{P_b} \leq \frac{1}{k} \sum_{d=d_{\min}}^{\infty} \hat{w}_d P_2(d), \tag{6.32}$$

where $\hat{w}_d$ is an average weight structure (associated with the erasures) that is based on the original weight structure, w_d, and is defined as

$$\hat{w}_d = \frac{1-\delta^d}{\delta} \sum_{n=d}^{\infty} \binom{n}{d} w_n \delta^n. \tag{6.33}$$

6.4 Results

6.4.1 Simulation Parameters

The OFDM system considered in this paper is equipped with $N = 64$ data-carrying subcarriers and a cyclic prefix of $N_g = 16$ samples. The data on each subcarrier

is BPSK modulated. Note that for larger constellation sizes, Gray encoding is utilized. The encoder uses the industry-standard generator polynomials, $\mathbf{g}_0 = 133_8$ and $\mathbf{g}_1 = 177_8$, giving the rate as $R_c = \frac{1}{2}$. The minimum distance of this code is $d_{\min} = 10$. Note that higher rates can be obtained by puncturing the output of the encoder. One codeword encompasses 64 OFDM symbols. The Viterbi decoder is used to decode the codeword utilizing soft decisions out of the symbol demapper. Specifically, the LLRs, as described in Section 5.4.2 are inputted into the Viterbi decoder.

The channel is modeled as independent complex Gaussian random variables as discussed in Section 5.1. It is assumed that the channel is constant over each OFDM symbol and subsequently changes upon transmission of the next symbol. It is also assumed for this work that the channel is known at the receiver. The narrowband interference is randomly distributed within the spectrum for each OFDM symbol transmission. The PEF weights given in (2.65) are assumed to be known at the receiver, though they can quickly be obtained through adaptive means.

6.4.2 Uncoded Results

MC-CDMA and OFDM with the PEF are compared in an uncoded scenario in Figure 6.3. For reference, the theoretical BER curves for both AWGN and Rayleigh fading are provided. Each system is simulated for the cases of no interference and one nonorthogonal interferer. An 8×8 block interleaver is implemented within each OFDM symbol. The plot for MC-CDMA with no interference indicates that frequency diversity provides a benefit in a frequency-selective channel. The performance of MC-CDMA with a single nonorthogonal interferer is very close to the case of no interference. In this case, the combiner utilizes the statistics of the channel and interference to successfully recombine the transmitted signal. The deviation from the case of no interference is quite small.

The PEF used with OFDM assumes knowledge of the MMSE weights given in (2.65). In the presence of no interference the performance of the system reduces to that of Rayleigh fading, as is expected. When the interference is present and is mitigated using the PEF, the performance approximates the case of Rayleigh fading at low E_b/N_0 values, however at high E_b/N_0 values, an error floor arises. This is due to the notch filter that removes a few tones when mitigating the interference.

For both MC-CDMA and OFDM with PEF, the interference must be estimated

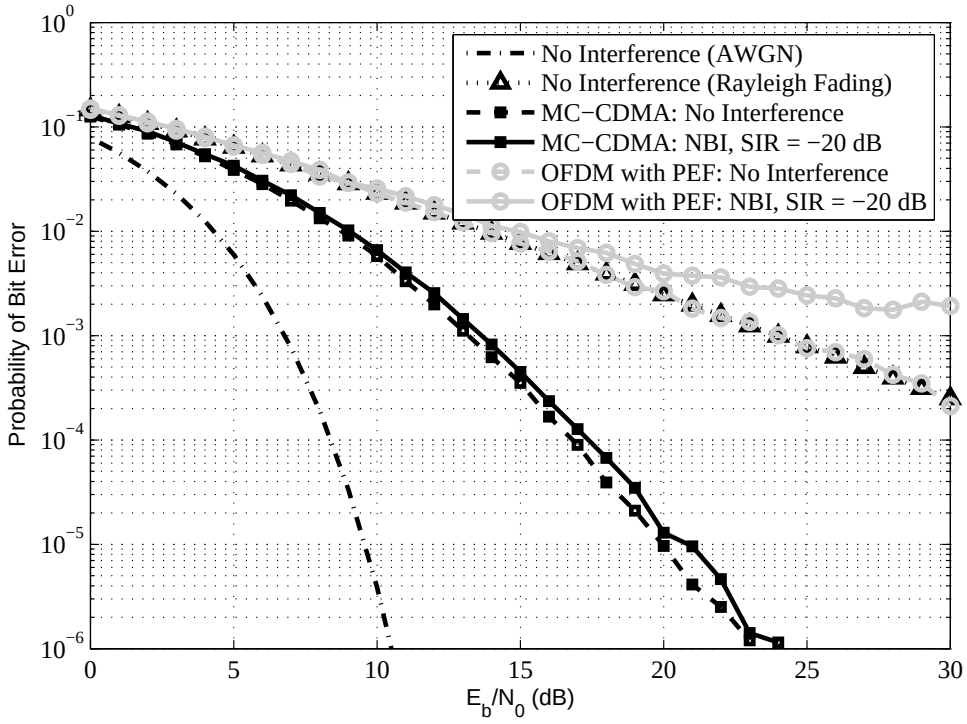


Figure 6.3: Probability of bit error versus E_b/N_0 for MC-CDMA and OFDM with the PEF for the cases of no interference and one non-orthogonal interferer, when $N = 64$, $N_g = 16$, $L_h = 5$, $M_{\text{pef}} = 12$, SIR = -20 dB.

in order to be implemented in a real world communication system. As mentioned earlier, the PEF is easily implemented using the low-complexity LMS algorithm without training symbols. This allows the interference to be adaptively estimated and removed. In the case of MC-CDMA, the correlation matrix of the interference is required when determining the MMSE combiner weights. This can accomplished using training symbols to provide an estimate for the received signal correlation matrix, as discussed in Chapter 3.

Finding the combiner weights requires solving (5.31). This requires inversion of the $N \times N$ Toeplitz matrix, $\mathbf{R}_X$. From Section 3.5.2 efficient matrix inversion requires $3N^2/2+7N/2+1$ complex multiplications, as well as the transmission of training symbols. On the other hand, the complexity of the LMS PEF is $\mathcal{O}(M_{\text{pef}})$ where M_{pef} is generally less than $N/4 = N_g$ and the PEF is a blind algorithm. This demonstrates that mitigation of the narrowband interference is less complex when using OFDM and the PEF, however, it requires the use of forward error correction in all scenarios to attain frequency diversity.

6.4.3 Conventional Erasure Insertion

Using the results of Section 6.2, the case of narrowband interference mitigation is considered using genie-aided erasure insertion as described in [10]. The results are demonstrated in Figure 6.4 for independent fading (see next section). As a reference, the case of no interference is also plotted. The erasure mechanism utilizes two threshold values, $\gamma \in [0.3, 1]$. Recall that $\gamma = 1$ is the case of using no erasures. From the plot it can be seen that when erasures are inserted (i.e. $\gamma = 0.3$), an improvement is achieved for all SIRs. However, the amount of improvement decreases as the SIR decreases. This is due to the increased amount of spectral leakage that occurs with larger interferences. The required number of erasures increases, thereby compromising the code's error correction capability. Note that for SIR = -20 dB, the use of erasures provides little to no improvement in performance.

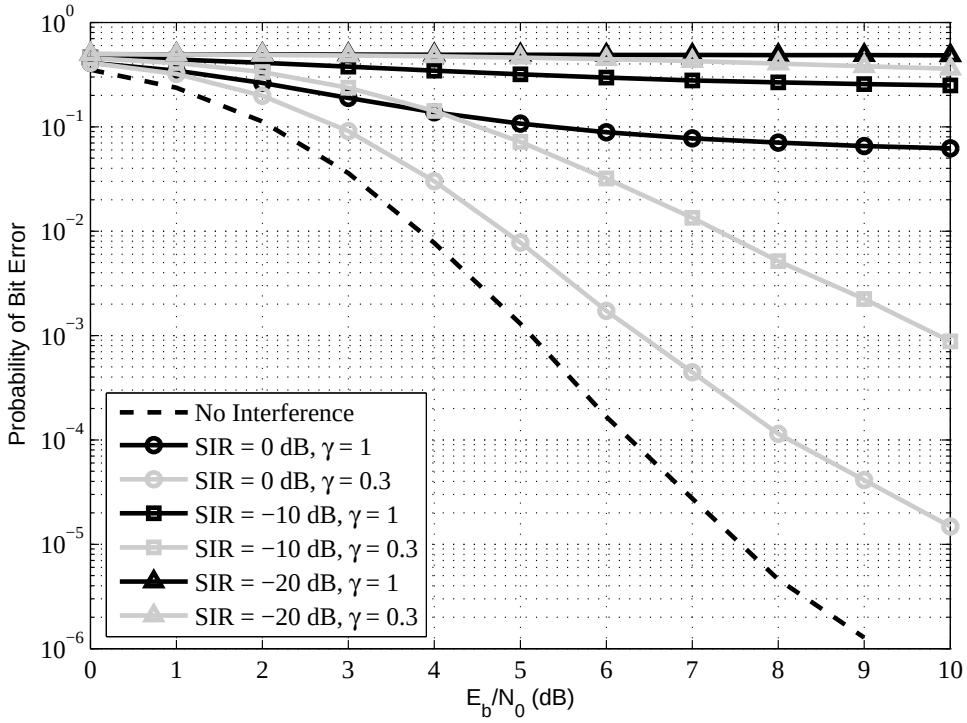


Figure 6.4: Probability of bit error versus E_b/N_0 for the coded scenario and independent fading when $\gamma = 0.3, 1$ and SIR $= 0, -10, -20$ dB.

6.4.4 Prediction-Error Filtering

Independent Fading

Under the assumption of independent fading, the analysis for the approximate PEP is simplified [16]. It is first noted that the overall path metric is defined as

$$\Gamma = \sum_{k \in \mathcal{K}} \Gamma_k, \tag{6.34}$$

where $\mathcal{K}$ is the set of positions where $\mathbf{C}$ and $\hat{\mathbf{C}}$ differ, with cardinality, $|\mathcal{K}| = d$, and $\Gamma_k = \mathcal{M}(R_k, C_k) - \mathcal{M}(R_k, \hat{C}_k)$. The computation of the Laplace transform using the independence assumption becomes

$$\Phi_\Gamma(s) = \mathbb{E}\left(e^{-s\Gamma}\right) = \prod_{k \in \mathcal{K}} \Phi_{\Gamma_k}(s). \tag{6.35}$$

Finally, if the codeword is binary,

$$\Phi_\Gamma(s) = \left[\Phi_{\Gamma_k}(s)\right]^d. \tag{6.36}$$

To obtain independence for this theoretical scenario, an ideal interleaver is necessary. In this work, the interleaver is designed such that neighboring symbols are transmitted in different OFDM symbols. That is the first OFDM symbol consists of symbols $[D_0, D_N, D_{2N}, \dots]^T$, and the second OFDM symbol consists of $[D_1, D_{N+1}, D_{2N+1}, \dots]^T$, etc.

The following figures demonstrate the results for independent fading. Figure 6.5 shows results for the case of $L_h = 5, M_{\text{pef}} = 12$, and both SIR = -20 dB and the case of no interference. Also plotted are the theoretical curves for uncoded AWGN, uncoded Rayleigh fading, and the coded bound for AWGN. It can be seen that for the case of no interference, the simulation and the theoretical results obtained using (6.22) and (6.29) match up quite well. This is also the case when SIR = -20 dB. The theoretical result is obtained by setting $N_E = 1$. This indicates that the PEF essentially inserts one erasure in each OFDM symbol. The bound obtained given in (6.32) provides a good approximation to the simulation. In this case it is seen that the PEF is effective in canceling out the interference, by scaling the LLRs in the codeword around the interferer location. The redundancy provided by the code allows for these erasures to be corrected. The improvement over uncoded Rayleigh fading is due to the frequency diversity that is obtained through the application of forward error correction across the subcarriers.

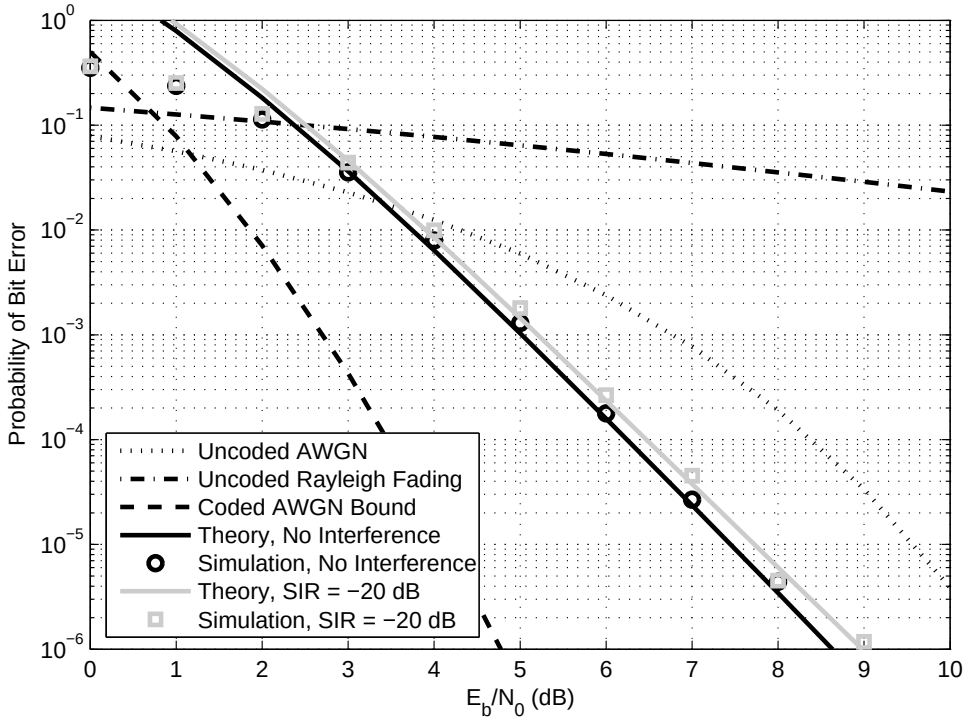


Figure 6.5: Probability of bit error versus E_b/N_0 for the coded scenario and independent fading with $L_h = 5, M_{\text{pef}} = 12$, SIR = -20 dB. The upper bound is obtained using $N_E = 1$.

Figure 6.6 shows results for the case of $L_h = 10, M_{\text{pef}} = 7$ and both SIR = -20 dB and the case of no interference. It is clear that the results seen in Figure 6.6 are identical to the results seen in Fig. 6.5. This result is not surprising because the fading is independent, and thus independent of the the length of the channel (L_h). The change in predictor taps (from $M_{\text{pef}} = 12$ in Figure 6.5 to $M_{\text{pef}} = 7$ in Fig. 6.6) arises from the effort to make the cyclic prefix fully utilized by both the channel and the PEF. Reducing the number of predictor taps changes the characteristics of the notch, which now has a shallower notch. However, notice that the assumption that $N_E = 1$ still holds from the comparison of the simulation and theoretical results.

Correlated Fading

The drawback of the previous interleaver design is that there is a delay associated with the system. The receiver must wait for all the OFDM symbols to be received before the Viterbi decoder can begin to decode the codeword. In many cases, this delay may be unacceptable. To remove the delay constraint associated with independent fading, the

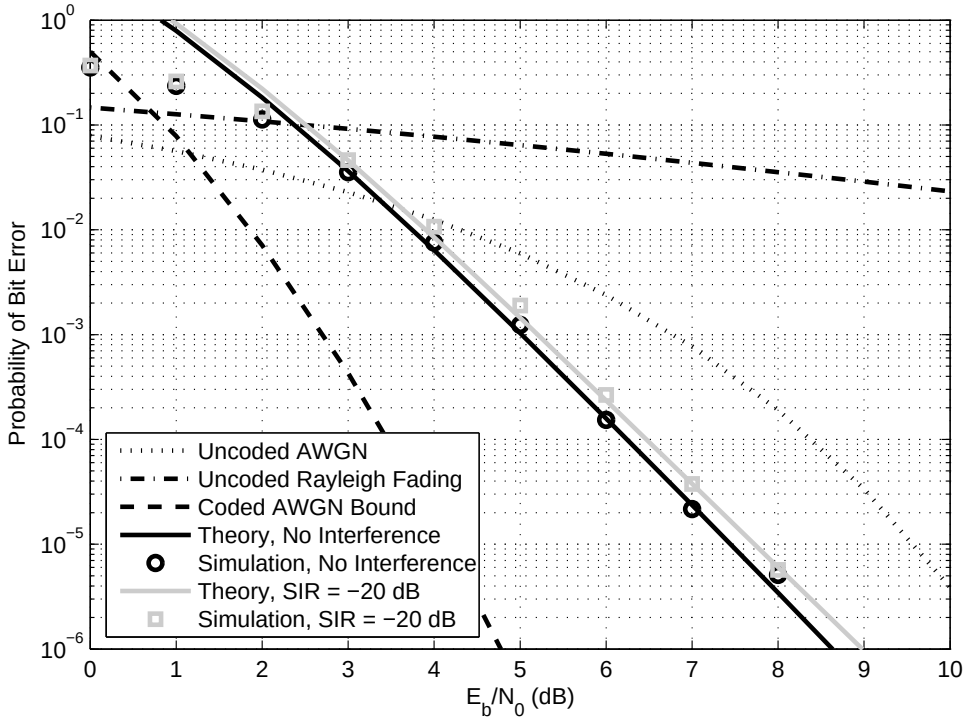


Figure 6.6: Probability of bit error versus E_b/N_0 for the coded scenario and independent fading with $L_h = 10, M_{\text{pef}} = 7$, SIR = -20 dB. The upper bound is obtained using $N_E = 1$.

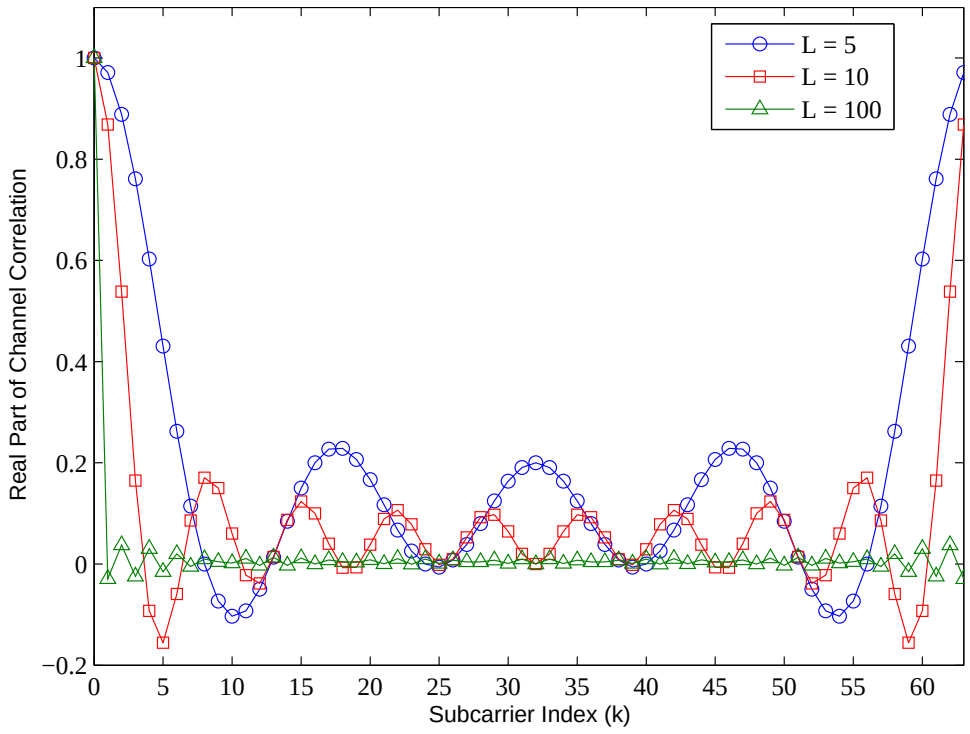


Figure 6.7: Real part of the channel correlation for $L_h = 5, 10, 100$

interleaver is designed such that the symbols within each OFDM symbol are interleaved. This technique is similar to the block interleaver used in the IEEE 802.11a WLAN standard [63]. The drawback of this implementation is that the channel coefficients in the frequency-domain, $\tilde{H}_k$, are correlated. The channel correlation in the frequency-domain [25] is found to be

$$R_{\tilde{H}}(n) = \mathbb{E}\left[\tilde{H}_k \tilde{H}^*_{k-n}\right] = \sum_{l=0}^{L_h-1} \sigma_h^2 e^{-j\frac{2\pi}{N}ln}, \quad n = 0, 1, \ldots, N-1$$
$$= \sigma_h^2 \left(\frac{1 - e^{-j\frac{2\pi}{N}nL_h}}{1 - e^{-j\frac{2\pi}{N}n}}\right). \tag{6.37}$$

This channel correlation is plotted for $L_h = 5, 10, 100$, where the real and imaginary parts are depicted in Figures 6.7 and 6.8, respectively. It is clearly seen that there is substantial correlation, and that as L_h increases, the correlation becomes smaller.

The results demonstrated in Figures 6.9 and 6.10 are for interleaving within OFDM symbols, specifically a 16×4 block interleaver is used. In Figure 6.9, the channel length is set to $L_h = 5$ and the number of predictor taps is varied ($M_{\text{pef}} = 7, 12$) for SIR = -20 dB. This plot clearly demonstrates that as the number of predictor taps decreases,

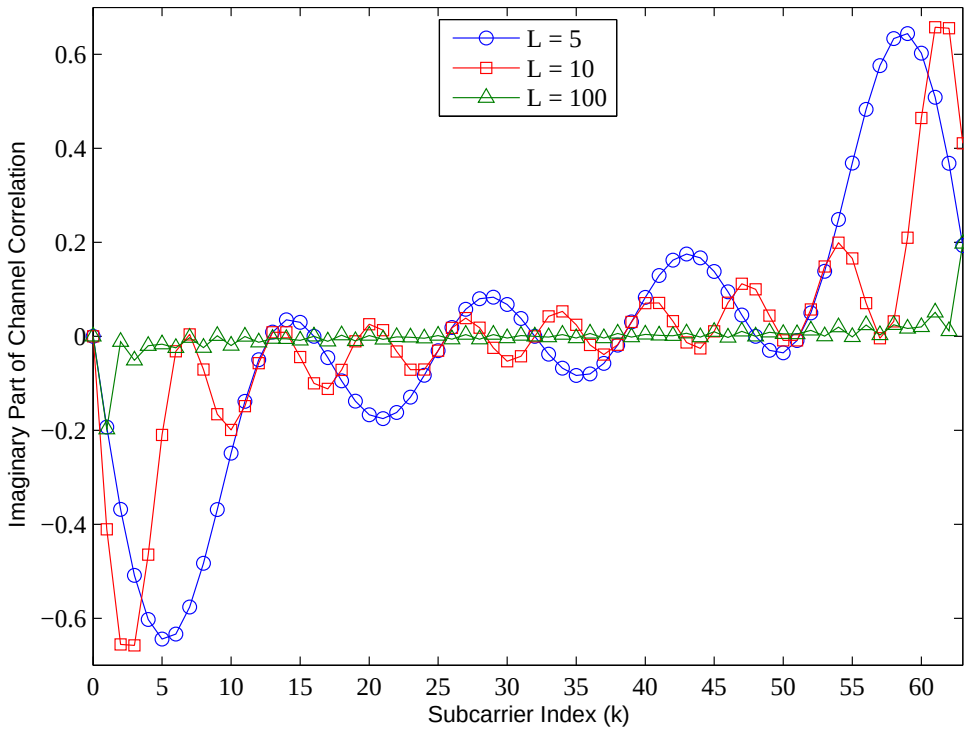


Figure 6.8: Imaginary part of the channel correlation for $L_h = 5, 10, 100$

so does the performance. This fact arises from the decrease in spectral resolution that occurs when fewer taps are used. Recall that the PEF becomes ideal as the number of predictor taps increases toward infinity.

In Figure 6.10, two cases are demonstrated, $L_h = 5, M_{\text{pef}} = 12$ and $L_h = 10, M_{\text{pef}} = 7$ for both SIR = -20 dB and the case of no interference. Also plotted are the independent fading bound with no interference (discussed in the previous section) and the coded AWGN bound. Note that the simulation comparison here examines the case when the whole cyclic prefix is utilized in both cases. It can first be seen that there is a substantial degradation in performance of correlated fading as compared to independent fading. This is due to the fact that the Viterbi algorithm operates optimally when sequential symbols are independent [97]. Secondly, it is seen that the results when using the PEF are very close to the case when no interference is present. Next notice that there is improvement of the case of $L_h = 10$ over the case of $L_h = 5$. This arises from the longer channel that provides an increase in diversity. However, note that there is a limit on the maximum diversity that can be obtained, and has been shown to be equal to $\min(d_{\min}, L_h)$ [2]. Finally, comparing the use of the PEF when $M_{\text{pef}} = 12$ and

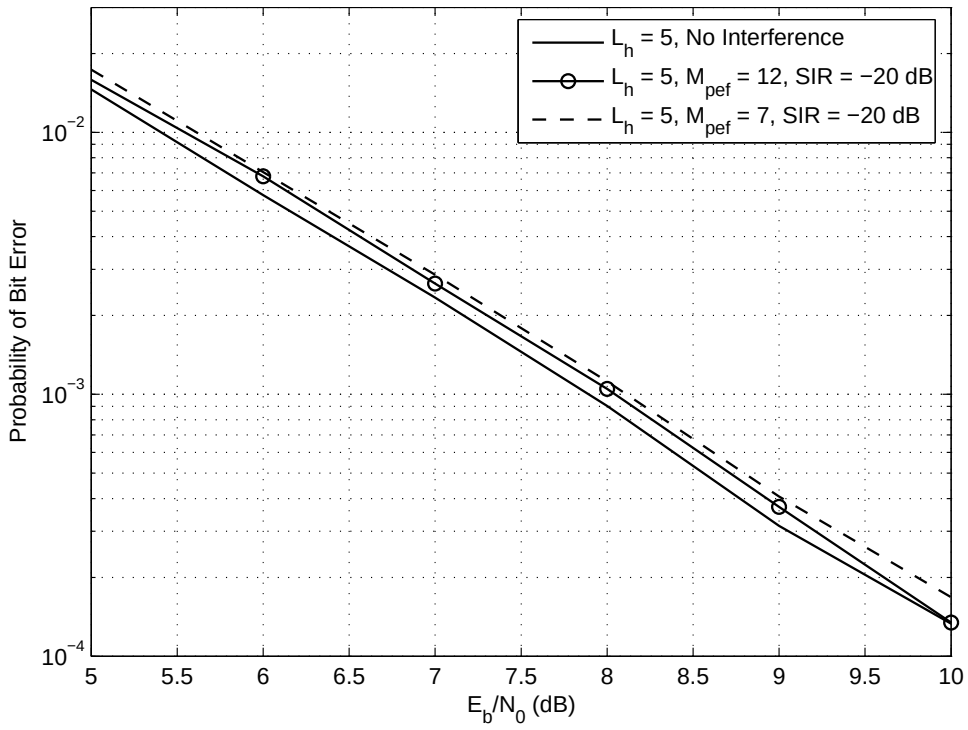


Figure 6.9: Probability of bit error versus E_b/N_0 for the coded scenario and correlated fading with $L_h = 5$, $M_{\text{pef}} = 7, 12$, SIR =-20 dB.

$M_{\text{pef}} = 7$, it can again be seen that there is more degradation between the case of no interference and SIR = -20 dB when the length of the PEF is short.

Figure 6.11 demonstrates the results for the coded simulation of MC-CDMA and OFDM with the PEF for the cases of no interference and one nonorthogonal interferer, for $L_h = 5$, $M_{\text{pef}} = 12$, SIR = -20 dB. Also plotted are the theoretical BER curves for uncoded Rayleigh fading and a BER bound for AWGN. It can be clearly seen that the performance of both systems in the presence of narrowband interference is very close to the case of no interference. Also note that the performance of OFDM with frequency diversity provided by coding and the PEF as an erasure insertion technique is equivalent to that of coded MC-CDMA for which frequency diversity is obtained through spreading. The coding gain is also apparent when comparing the coded results with Figure 6.3

6.4.5 Hybrid System

The erasure insertion method is shown to be useful when the SIR is relatively large (i.e. SIR $>$ 0 dB), while the PEF has been shown to perform well when the

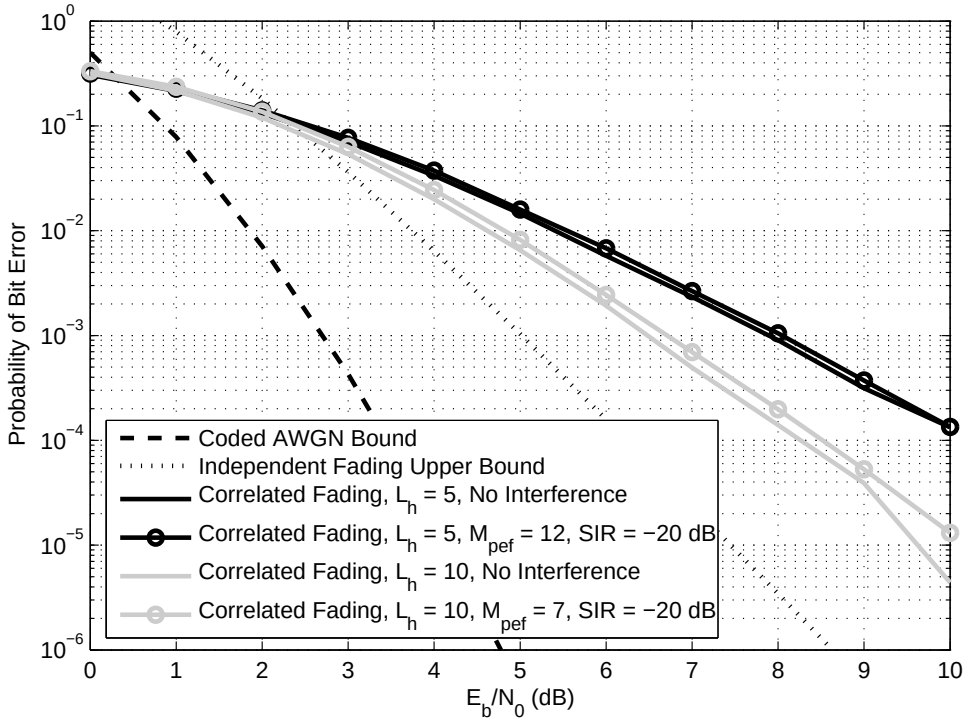


Figure 6.10: Probability of bit error versus E_b/N_0 for the coded scenario and correlated fading with $L_h = 5, M_{\text{pef}} = 12$ and $L_h = 10, M_{\text{pef}} = 7$ for SIR = -20 dB.

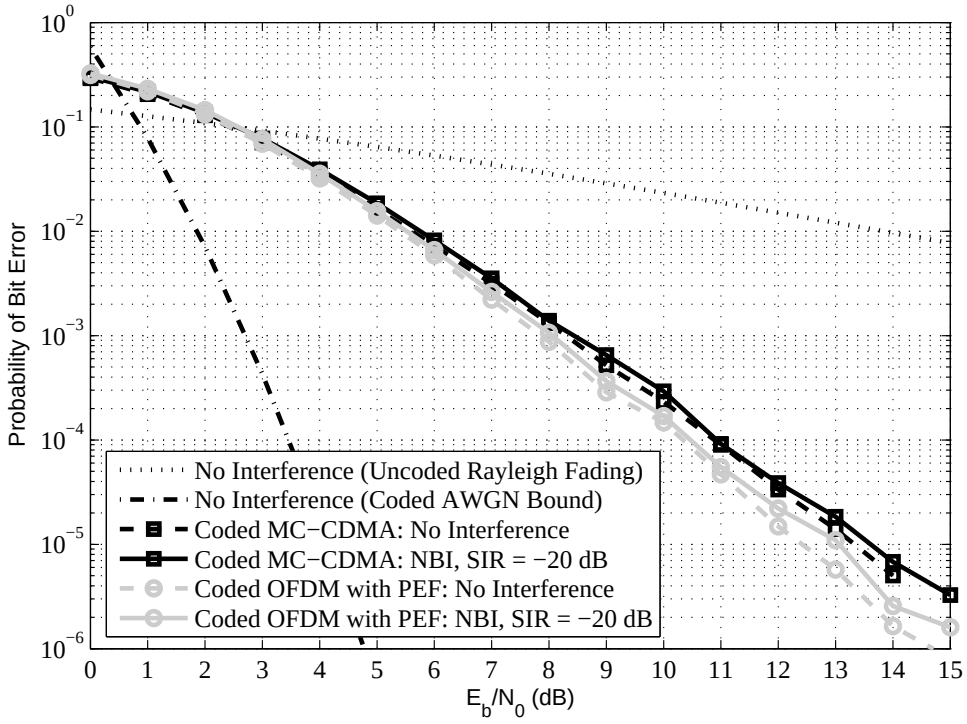


Figure 6.11: Probability of bit error versus E_b/N_0 for MC-CDMA and OFDM with the PEF for the cases of no interference and one non-orthogonal interferer, with $N = 64$, $N_g = 16$, $L_h = 5$, $M_{\text{pef}} = 12$, SIR = -20 dB.

interference is strong (i.e. SIR < 0 dB). These results suggest the use a hybrid approach for mitigating narrowband interference. Such an idea was proposed by Kelleci et al. [67], using a mixed-mode approach to mitigate narrowband interference over a range of SIRs. The authors utilize digital erasure insertion for SIR > 0 dB and an adaptive analog notch filter for SIR < 0 dB. In this work the potential hybrid system is designed in the digital-domain consisting of Bayesian erasure insertion and prediction-error filtering.

Figure 6.12 demonstrates the performance of Bayesian erasure insertion and prediction-error filtering over a range of SIRs when $E_b/N_0 = 7$ dB. Erasures are inserted with a threshold of $\gamma = 0.3$ and the PEF is of length $M_{\text{pef}} = 12$. Also plotted is the case of no interference. From the plot it can be seen that when the fading is independent the PEF outperforms erasures over all SIR values. The erasure insertion method suffers at low SIRs because of the need for a large number of erasures. As the SIR increases, the performance improves, approaching that of the PEF. However, if the fading is in fact correlated, a substantial loss in performance for both techniques is noticed over all SIR values. The PEF is seen to outperform the erasure insertion method over a range of SIR values, however, notice that when the SIR is in the range of 4-14 dB, the PEF performance suffers as compared to the erasure insertion method.

For the independent fading case, it is clear that a hybrid system is unnecessary. The performance of the PEF is excellent over all SIR values. In the case of correlated fading, the use of a hybrid system may be warranted, however, the gain provided by the erasure insertion method over the PEF is not substantial.

6.5 Summary

In this chapter, two block modulated multi-carrier systems, MC-CDMA and OFDM, are compared in the presence of severe narrowband interference and frequency-selective fading. MC-CDMA obtains frequency diversity by utilizing a code matrix to spread each symbol into all the subcarriers and requires a combiner at the receiver to recover the data. This diversity provides robustness against interference in the uncoded scenario as compared to OFDM.

Forward error correction coding and interleaving is used in conjunction with OFDM to provide frequency diversity and obtain results similar to MC-CDMA. The prediction-error filter (PEF) is proposed as an interference mitigation technique in a bit-interleaved coded modulated OFDM system. The PEF effectively places erasures

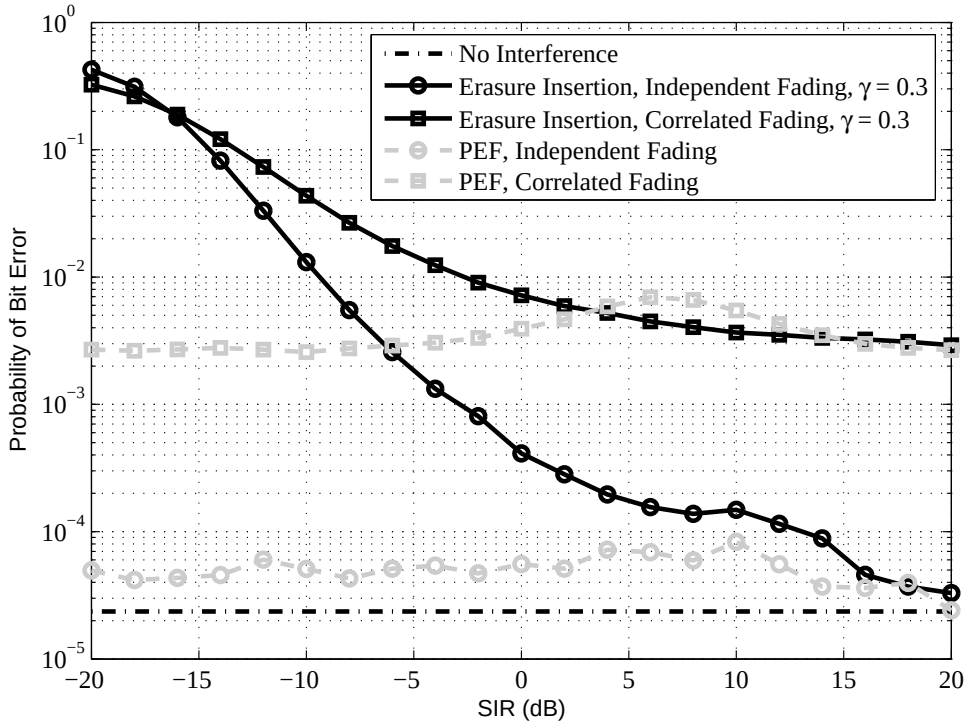


Figure 6.12: Probability of bit error versus SIR for the coded scenario. Comparison of erasure insertion ($\gamma = 0.3$) and the PEF ($M_{\text{pef}} = 12$) for $E_b/N_0 = 7$ and $L_h = 5$.

around the interference in the codeword by scaling the log-likelihood ratios (LLRs) that are inputted into the Viterbi decoder. Results indicate excellent performance of the PEF in conjunction with coding as compared to the case of no interference. A hybrid system of the PEF and erasures is considered, and it is shown that the PEF performs better than erasure insertion over a range of SIRs when the fading between subcarriers is independent. If the fading is correlated, a hybrid system may be warranted.

Chapter 6, in part, is a reprint of material as it appears in A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems," in *Proceedings of the Military Communications (MILCOM) Conference*, San Diego, CA, Nov. 2008, A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in BICM OFDM systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, pp. 2605-2608, Apr. 2009, A. Batra, J. R. Zeidler, J. G. Proakis, and L. B. Milstein, "Interference Rejection and Management," in *New Directions in Wireless Communications Research*, V. Tarokh, Ed. New York: Springer, 2009 and is currently being prepared for submission for publication of the material. A. Batra and J. R. Zeidler, "Narrowband Interference Mitigation in OFDM systems using the Prediction-Error Filter," in preparation, 2009. The dissertation author was the primary investigator and author of these papers.

7 Conclusions

In this dissertation, techniques for mitigating severe narrowband interference are investigated for single-carrier and multi-carrier systems. Interferers that dominate the signal of interest degrade communication links, making them unusable. In many practical cases, a priori knowledge of the interference is not available. This forces receivers to utilize adaptive signal processing methods to suppress the interference, especially in scenarios where the information is transmitted in short packets. Specifically, well known structures that include time-domain equalizers and the prediction-error filter (PEF) are considered as means for removing the interference.

In the case of single-carrier systems, the large interference power causes long convergence times for the adaptive equalizers. These structures use the least-mean square (LMS) or its derivative, the normalized LMS (NLMS) algorithms as the adaptive mechanism. Traditional analysis of these algorithms are based on the metric of mean-squared error (MSE), however, in this work bit error rate (BER) performance is considered. Two techniques are proposed to reduce the number of training symbols needed: data-aided initialization and a two-stage approach utilizing pre-filtering.

In Chapter 3, the adaptive decision-directed equalizer (DDE) and the adaptive decision-feedback equalizer (DFE) are compared in the presence of severe narrowband interference and a two-tap multipath channel. For the case of the DDE, significant BER improvement was obtained over the time-invariant Wiener filter, due to the non-Wiener effects induced by the narrowband interference, even in the presence of a multipath channel. The BER performance of this structure also approaches the BER performance of the DFE when both structures use essentially the same information. Note that the DFE does not experience the non-Wiener effects due to its nonlinear nature. The drawback of suppressing the interference with these structures, is the extended convergence time required. To combat this, a data-aided initialization (DAI) technique is proposed that

generates estimates for the Wiener filter based on correlation matrices formed from the training data. This is shown to substantially reduce the number of training symbols needed for convergence by approximately two orders of magnitude. To obtain the estimates of the DFE Wiener filter, two computationally efficient techniques are compared in terms of complexity and performance relative to that of the theoretical Wiener filter for the DFE. The parametric approach to the direct matrix inversion (DMI) solution has the least complexity, however, the performance degrades due to assumptions made on the structure of the correlation matrix. These assumptions only hold when a large number of training symbols are used, making the technique undesirable. On the other hand the conjugate gradient (CG) multistage Wiener filter (MSWF) has larger complexity, yet requires fewer training symbols to approach the BER performance of the theoretical DFE Wiener filter.

In Chapter 4, a two-stage approach is investigated that aims to improve the convergence of the LMS DFE. The delay in convergence can be attributed to the fact that the DFE does not have a direct reference for the interference. Instead it must adapt based solely on the training data. To improve the convergence, a pre-filter composed of the LMS PEF is proposed for the LMS DFE. This system is shown to improve the convergence of the system by approximately two orders of magnitude. This is accomplished without any substantial degradation in the steady-state BER performance. In fact the BER of the two-stage system is shown to approximate the BER of the DFE-only. This indicates that it is possible to split the LMS DFE into an LMS PEF pre-filter followed by an LMS DFE, improving the convergence and having approximately the same BER performance. Finally, a blind implementation is proposed that reduces the complexity of the overall system, at the cost of a small degradation in BER performance.

The two approaches described in Chapters 3 and 4 can be compared in terms of performance and complexity. The DAI technique using CG MSWFs provides excellent BER performance as compared to the theoretical Wiener DFE using only 250 training symbols when $M = 3$ is the number of side taps of the one-sided feedforward filter (and also equal to the number of feedback taps). This approach suffers from higher complexity which scales as $\mathcal{O}(M^2)$. Conversely, the complexity of two-stage approach scales as $\mathcal{O}(M)$ where M is the number taps of the PEF (and also equal to the number of feedback taps). The BER performance also approaches the steady-state theoretical DFE performance, however, for the case of $M = 3$ this system requires at least 450 training symbols. Thus,

there is a trade-off between complexity and the number of training symbols needed.

In the case multi-carrier systems, two block modulated multi-carrier systems are examined the presence of severe narrowband interference operating in a frequency-selective channel. The spectral leakage of interference power causes a large number of subcarriers to be degraded. This occurs when the interference is nonorthogonal to the subcarriers. When the interference is orthogonal to the subcarriers (other than the m^{th} subcarrier), the interference is limited to single tone.

In Chapter 6, multi-carrier code division multiple access (MC-CDMA) and orthogonal frequency division multiplexing (OFDM) are investigated. MC-CDMA differs from OFDM in that it spreads each data symbol into all the subcarriers using spreading codes. If the spreading matrix is the identity matrix, then MC-CDMA becomes OFDM. The spreading codes provide MC-CDMA with frequency diversity, however, it requires a combiner at the receiver to extract the transmitted data. The optimal combining weights are derived under the MMSE criterion for both cases of interference. In practical situations, determining the optimal weights requires the inversion of a correlation matrix with complexity that scales as $\mathcal{O}(N^2)$, where N is the number of subcarriers. The use of the optimal weights in an uncoded scenario provides a substantial improvement in BER performance over the case of OFDM. To improve the performance of OFDM, it is combined with bit-interleaved coded modulation. In particular, a convolutional code is utilized with a block interleaver at the transmitter, and a corresponding block deinterleaver with a Viterbi decoder at the receiver. The code and interleaver provide OFDM with the frequency diversity it was lacking in the uncoded scenario. This system is compared with a genie-erasure insertion method proposed by Baum and Pursley [10], and it is shown that simply using erasures is inadequate for severe narrowband interference. This is due to the fact that a large number of erasures is needed, thereby compromising the code's error correction capability. The PEF is proposed as an erasure insertion mechanism that localizes erasures around the interference location. This process is performed by the notch of the PEF that utilizes the unused portion of the cyclic prefix (normally designed to deal with delay spread of the channel) and is implemented in the time-domain prior to demodulation. The BER performance of this technique is excellent as compared to the case of no interference and the complexity scales with the number of prediction taps. Analysis for an upper bound on the BER of coded OFDM with the PEF is performed. Finally, a hybrid system of the PEF and erasures is considered over a range of SIR values.

It is determined that PEF performs better than the erasure method over almost all SIR values.

Future work consists of determining an analytical upper bound on the BER specifically for correlated fading. This situation is more relevant due to the delay requirements in processing of the received data. This is seen specifically in 802.11a [63] systems, where the interleaving is performed within each OFDM symbol. Another point of interest, is the length of the PEF that is necessary for mitigating the interference in an OFDM system. It is possible for the PEF to have a large number of taps, thereby increasing the spectral resolution of the notch, to remove the tone closest to the interference. This then causes ISI and ICI, however, it is deterministic and can be removed. In terms of MC-CDMA, methods for obtaining MMSE combiner weights can be investigated. This requires an estimation procedure that could be equivalent to the DAI for the equalizer weights. The performance and complexity must be considered for this situation as well.

A Average Power of a Non-Orthogonal Interferer

The proof for (5.28), can be shown by first making the change of variables, $n = m - k$, giving

$$\frac{1}{N}\sum_{k=0}^{N-1}\frac{E_i}{N}\frac{1-\cos 2\pi\alpha}{1-\cos\frac{2\pi}{N}(m-k+\alpha)} = \frac{E_i}{N^2}(1-\cos 2\pi\alpha)\sum_{n=m-(N-1)}^{m}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)}. \tag{A.1}$$

Separating the sum into two sums,

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\left[\sum_{n=0}^{m}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)} + \sum_{l=m-(N-1)}^{-1}\frac{1}{1-\cos\frac{2\pi}{N}(l+\alpha)}\right]. \tag{A.2}$$

Making another change of variables in the second sum, $n = l + N$,

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\left[\sum_{n=0}^{m}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)} + \sum_{n=m+1}^{N-1}\frac{1}{1-\cos\frac{2\pi}{N}(n-N+\alpha)}\right]. \tag{A.3}$$

Expanding the second sum and simplifying gives

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\left[\sum_{n=0}^{m}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)} + \sum_{n=m+1}^{N-1}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)}\right]. \tag{A.4}$$

Combining the two sums into one,

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\sum_{n=0}^{N-1}\frac{1}{1-\cos\frac{2\pi}{N}(n+\alpha)}. \tag{A.5}$$

Using the double angle formula, i.e. $\cos(2u) = 1 - 2\sin^2(u)$, gives

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\sum_{n=0}^{N-1}\frac{1}{\sin^2\frac{\pi}{N}(n+\alpha)}. \tag{A.6}$$

This can be written as

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\sum_{n=0}^{N-1}\csc^2\frac{\pi}{N}(n+\alpha). \tag{A.7}$$

This finite summation can be simplified [135],

$$\frac{E_i}{N^2}(1-\cos 2\pi\alpha)\frac{N^2}{2}\csc^2\left(N\frac{\pi\alpha}{N}\right). \tag{A.8}$$

After some simple math,

$$E_i(1-\cos 2\pi\alpha)\frac{1}{1-\cos 2\pi\alpha}. \tag{A.9}$$

Finally,

$$\frac{1}{N}\sum_{k=0}^{N-1}\frac{E_i}{N}\frac{1-\cos 2\pi\alpha}{1-\cos\frac{2\pi}{N}(m-k+\alpha)}=E_i. \tag{A.10}$$

Abbreviations

ADC	analog-to-digital converter
AGC	automatic gain control
ANC	adaptive noise canceler
AR	autoregressive
AWGN	additive white Gaussian noise
BER	bit error rate
BICM	bit-interleaved coded modulation
BPSK	binary phase-shift keying
CDMA	code division multiple access
CFO	carrier frequency offset
CG	conjugate gradient
CMA	constant modulus algorithm
CP	cyclic prefix
dB	decibels, $10\log_{10}(\cdot)$
DAC	digital-to-analog converter
DAI	data-aided initialization
DDE	decision-directed equalizer
DFE	decision-feedback equalizer
DFT	discrete Fourier transform
DMI	direct matrix inversion
DS	direct sequence
FDE	frequency-domain equalizer
FDM	frequency division multiplexing
FFT	fast Fourier transform
FIR	finite impulse response
FM	frequency modulation
HSDPA	high speed downlink packet access
Hz	Hertz (1 cycle/s)
ICI	intercarrier interference
IDFT	inverse discrete Fourier transform
IEEE	Institute of Electrical and Electronic Engineers
IFFT	inverse fast Fourier transform
INR	interference-to-noise ratio
ISI	intersymbol interference
LAN	local area network

LE	linear equalizer
LLR	log-likelihood ratio
LMS	least-mean square
LOS	line-of-sight
LP	linear prediction
MAP	maximum a posteriori
MB	multi-band
MC-CDMA	multi-carrier code division multiple access
MF	matched filter
ML	maximum likelihood
MMSE	minimum mean-squared error
MSE	mean-squared error
MSWF	multistage Wiener filter
NIR	noise-to-interference ratio
NLMS	normalized least-mean square
OFDM	orthogonal frequency division multiplexing
PAPR	peak-to-average power ratio
PEF	prediction-error filter
PEP	pairwise error probability
PLL	phase locked loop
QAM	quadrature-amplitude modulation
QPSK	quadrature phase-shift keying
RLS	recursive least square
SAW	surface acoustic wave
SINR	signal-to-interference-plus-noise ratio
SIR	signal-to-interference ratio
SNR	signal-to-noise ratio
SS	spread spectrum
SVD	singular value decomposition
TI	time-invariant
UWB	ultra-wideband
WiMAX	worldwide interoperability for microwave access
WLAN	wireless local area network

Symbols

$\lvert \cdot \rvert$	absolute value
$\arg(\cdot)$	argument
$\binom{\cdot}{\cdot}$	binomial coefficient
$\lceil \cdot \rceil$	ceiling function
$[\cdot]$	closed interval
$(\cdot)^*$	complex conjugate
$\underbrace{x \cdots x}_{m}$	m copies
$\csc(\cdot)$	cosecant
$\cos(\cdot)$	cosine
$\mathrm{diag}(\cdot)$	diagonal matrix returned from vector argument
$\delta(\cdot)$	Dirac delta function
$\mathbb{E}[\cdot]$	expectation operator
$\lfloor \cdot \rfloor$	floor function
$Q(\cdot)$	Gaussian Q-function
$>$	greater than
$\geq$	greater than or equal
$(\cdot)^H$	Hermitian (complex conjugate transpose) operator
j	$\sqrt{-1}$
$J_i(\cdot)$	i^{th}-order modified Bessel function of the first kind
$\Im\{\cdot\}$	imaginary part
$\cap$	intersection operator
$(\cdot)^{-1}$	inverse
δ_m	Kronecker delta function
$<$	less than
$\leq$	less than or equal
$\lim$	limit
$x * y$	linear convolution
$\log$	logarithm
$[\cdot]_{kl}$	matrix value at index (k, l)
$\max$	maximum
$\min$	minimum
$\gg$	much greater than
$\ll$	much less than
$\prod_{n=1}^{N}$	multiple product

$\sum_{n=1}^{N}$	multiple sum
$x \times y$	multiplication
$\Re\{\cdot\}$	real part
$\{x, \ldots, y\}$	set of elements
$\text{sinc}(\cdot)$	sinc function
sin	sine
tan	tangent
$\text{Toeplitz}(c, r)$	Toeplitz matrix: column vector c, row vector, r
$(\cdot)^T$	transpose operator
$\|\cdot\|$	2-norm
$\cup$	union operator
$x \rightarrow y$	x approaches y

Bibliography

[1] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[2] E. Akay and E. Ayanoglu, "Full frequency diversity codes for single input single output systems," in *Proceedings of the 60th Vehicular Technology Conference (VTC-Fall)*, Los Angeles, CA, Sept. 2004, pp. 1870–1874.

[3] A. E. Albert and L. S. Gardener, Jr., *Stochastic Approximation and Nonlinear Regression*, 1st ed. Cambridge, MA: MIT Press, 1967.

[4] S. Allpress, C. Luschi, and S. Felix, "Exact and approximated expression of the log-likelihood ratio for 16-QAM signals," in *Conference Record of the 38th Asilomar Conference on Circuits Systems and Computers*, vol. 1, Pacific Grove, CA, Nov. 2004, pp. 794–798.

[5] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Boston, MA: Kluwer, 2004.

[6] A. Batra, S. Lingam, and J. Balakrishnan, "Multi-band OFDM: A cognitive radio for UWB," in *Conference Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, Island of Kos, Greece, May 2006, pp. 4094–4097.

[7] A. Batra and J. R. Zeidler, "Narrowband interference mitigation in OFDM systems," in *Proceedings of the IEEE Military Communications Conference (MILCOM)*, San Diego, CA, Nov 2008.

[8] ——, "Narrowband interference mitigation in BICM OFDM systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009.

[9] A. Batra, J. R. Zeidler, and A. A. Beex, "Mitigation of narrowband interference using adaptive equalizers," in *Proceedings of the European Signal Processing Conference (EUSIPCO) 2006*, Florence, Italy, Sept. 2006.

[10] C. W. Baum and M. B. Pursley, "Bayesian methods for erasure insertion in frequency-hop communication systems with partial-band interference," *IEEE Trans. Commun.*, vol. 40, pp. 1231–1238, July 1992.

[11] A. A. Beex and J. R. Zeidler, "Non-linear effects in interference contaminated adaptive equalization," in *IASTED International Conference on Signal Processing, Pattern Recognition, and Application*, Crete, Greece, June 2002, pp. 474–479.

[12] ——, "Non-wiener effects in recursive least squares adaptation," in *Seventh International Symposium on Signal Processing and its Applications (ISSPA)*, Paris, July 2003, pp. 595–598.

[13] ——, "Steady-state dynamic weight behavior in (N)LMS adaptive filters," in *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow, Eds. Hoboken, NJ: John Wiley & Sons, 2003, ch. 9, pp. 335–443.

[14] N. J. Bershad and O. M. Macchi, "Adaptive recovery of a chirped sinusoid in noise, Part 2: Performance of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 595–602, Mar. 1991.

[15] E. Biglieri, *Coding for Wireless Channels*, 1st ed. New York, NY: Springer, 2005.

[16] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Computing error probabilities over fading channels: A unified approach," *European Transactions on Telecommunications*, vol. 9, no. 1, pp. 15–25, Jan-Feb 1998.

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

[18] R. R. Bitmead and B. D. O. Anderson, "Lyapunov techniques for the exponential stability of linear difference equations with random coefficients," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 782–787, Aug. 1980.

[19] ——, "Performance of adaptive estimation algorithms in dependent random environments," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 788–794, Aug. 1980.

[20] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York, NY: Dover, 1958.

[21] J. P. Burg, "Maximum entropy spectral analysis," in *37th Annual International Meeting, Society of Exploration Geophysicists*, Oklahoma City, OK, 1967.

[22] ——, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford University, Stanford, CA, 1975.

[23] G. Carron, R. Ness, L. Deneire, L. V. der Perre, and M. Engles, "Comparison of two modulation techniques using frequency domain processing for in-house networks," *IEEE Trans. Consumer Electron.*, vol. 47, pp. 63–72, Feb. 2001.

[24] J. K. Cavers and P. Ho, "Analysis of error performance of trellis-coded modulations in Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, pp. 74–83, Jan. 1992.

[25] A. Chini, M. S. El-Tanany, and S. A. Mahmoud, "On the performance of a coded MCM over multipath Rayleigh fading channels," in *IEEE International Conference on Communications (ICC)*, Seattle, WA, June 1995, pp. 1689–1694.

[26] A. Chouly, A. Brajal, and S. Jourdan, "Orthogonal multicarrier techniques applied to direct sequence spread spectrum CDMA systems," in *Proceedings of the IEEE Global Telecommunications (Globecom) Conference*, Houston, Nov. 1993, pp. 1723–1728.

[27] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–675, July 1985.

[28] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[29] A. J. Coulson, "Bit error rate performance of OFDM in narrowband interference with excision filtering," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2484–2492, Sept. 2006.

[30] A. Czylwik, "Adaptive OFDM for wideband radio channels," in *Proceedings of the IEEE Global Telecommunications (Globecom) Conference*, London, United Kingdom, Nov. 1996, pp. 713–718.

[31] V. M. Da Silva and E. S. Sousa, "Performance of orthogonal CDMA codes for quasi-synchronous communication systems," in *Proceedings of the IEEE International Conference on Universal Personal Communications (ICUPC)*, Ottawa, Oct. 1993, pp. 995–999.

[32] D. Darsena, G. Gelli, L. Paura, and F. Verde, "Joint equalization and interference suppression in OFDM systems," in *Conference Record of the 36th Asilomar Conference on Circuits Systems and Computers*, Pacific Grove, CA, Nov. 2002, pp. 1832–1835.

[33] ——, "NBI-resistant zero-forcing equalizers for OFDM systems," *IEEE Commun. Lett.*, vol. 9, pp. 744–746, Aug. 2005.

[34] D. Darsena, "Successive narrowband interference cancellation for OFDM systems," *IEEE Commun. Lett.*, vol. 11, pp. 73–75, Jan. 2007.

[35] A. Dembo, "Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 34, pp. 352–355, Mar. 1988.

[36] G. Dietl, M. D. Zoltowski, and M. Joham, "Recursive reduced-rank adaptive equalization for wireless communications," *Proc. of SPIE*, vol. 4395, pp. 16–27, Apr. 2001.

[37] C. Dumard, F. Kaltenberger, and K. Freudenthaler, "Low-cost approximate LMMSE equalizer based on Krylov subspace methods for HSDPA," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1610–1614, May 2007.

[38] *High Rate Ultra Wideband PHY and MAC Standard*, ECMA Std. 368, Rev. 3, Dec. 2008. [Online]. Available: http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-368.pdf

[39] P. Elias, "Coding for noisy channels," *IRE Conv. Record*, pp. 37–47, 1955.

[40] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 64–74, Apr. 1963.

[41] K. Fazel and L. Papke, "On the performance of convolutionally-coded CDMA/OFDM for mobile communication systems," in *Proceedings of the IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC) Conference*, Yokahama, Dec. 1993, pp. 468–472.

[42] K. Fazel and S. Kaiser, *Multi-carrier and spread spectrum systems*. New York, NY: John Wiley & Sons, Inc., 2003.

[43] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[44] J. M. Geist and J. B. Cain, "Viterbi decoder performance in Gaussian noise and periodic erasure bursts," *IEEE Trans. Commun.*, vol. COM-28, pp. 1417–1422, Aug. 1980.

[45] D. Gerakoulis and P. Salmi, "An interference suppressing OFDM system for wireless communication," in *IEEE International Conference on Communications (ICC)*, New York, NY, Apr. 2002, pp. 480–484.

[46] D. Gesbert and P. Duhamel, "Unbiased blind adaptive channel identification and equalization," *IEEE Trans. Signal Processing*, vol. 48, pp. 148–158, Jan. 2000.

[47] J. R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 484–491, Dec. 1977.

[48] I. C. Gohberg and A. A. Semencul, "On the inversion of finite Toeplitz matrices and their continuous analogues (in Russian)," *Mat. Issled*, vol. 2, pp. 201–203, 1972.

[49] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2943–2959, Nov. 1990.

[50] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1984.

[51] R. P. Gooch and B. Daellenbach, "Prevention of interference capture in a blind (CMA-based) adaptive receive filter," in *Conference Record of the 23rd Asilomar Conference on Circuits Systems and Computers*, Pacific Grove, CA, Oct. 1989, pp. 898–902.

[52] R. M. Gray, "Toeplitz and Circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, pp. 155–239, 2006.

[53] L. J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 207–222, Apr. 1975.

[54] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, pp. 126–133, Dec. 1997.

[55] B. Hassibi, "On the robustness of LMS filters," in *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow, Eds. Hoboken, NJ: John Wiley & Sons, 2003, ch. 4, pp. 105–144.

[56] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[57] ——, "Cognitive radios: Brain-empowered wireless communication," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 201–220, Feb. 2005.

[58] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Processing*, vol. 45, pp. 1118–1128, May 1997.

[59] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–432, Dec. 1952.

[60] M. L. Honig and H. V. Poor, "Adaptive interference suppression," in *Wireless Communications: Signal Processing Perspectives*, H. V. Poor and G. W. Wornell, Eds. Englewood Cliffs, NJ: Prenctice Hall, 1998, ch. 2, pp. 64–128.

[61] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1928–1946, July 2001.

[62] F. M. Hsu and A. A. Giordano, "Digital whitening techniques for improving spread spectrum communications performance in the presence of narrowband jamming and interference," *IEEE Trans. Commun.*, vol. COM-26, pp. 209–216, Feb. 1978.

[63] *Wireless LAN Medium Access Control (MAC) Physical Layer (PHY) Specifications, Amendment 1: High-Speed Physical Layer in the 5 GHz Band*, IEEE Std. 802.11a, July 1999.

[64] *Multiband OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a (Doc. Number P802.14-03/268r3)*, IEEE Std. 802.15a, Mar. 2004.

[65] F. Jelinek, "An upper bound on moments of sequential decoding effort," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 140–149, Jan. 1969.

[66] M. Joham and M. D. Zoltowski, "Interpretation of the multi-stage nested Wiener filter in the Krylov subspace framework," Purdue University/Munich University of Technology, Tech. Rep. TR-ECE-00-51/TUM-LNS-TR-00-6, 2000. [Online]. Available: http://cobweb.ecn.purdue.edu/~mikedz/research/msnwf_tutorial.pdf

[67] B. Kelleci, T. W. Fischer, K. Shi, Y. Zhou, A. I. Karşilayan, and E. Serpedin, "Narrowband interference suppression in multi-band OFDM ultra wideband communications systems: A mixed-mode approach," in *International Workshop on Digital Signal Processing*, Jackson, WY, Sept. 2006, pp. 55–59.

[68] J. W. Ketchum and J. G. Proakis, "Adaptive algorithms for estimating and suppressing narrow-band interference in PN spread-spectrum systems," *IEEE Trans. Commun.*, vol. COM-30, pp. 913–924, May 1982.

[69] S. Kondo. and L. B. Milstein, "Performance of multicarrier DS CDMA systems," *IEEE Trans. Commun.*, vol. 44, pp. 238–246, Feb. 1996.

[70] J. D. Laster and J. H. Reed, "Interference rejection in digital wireless communications," *IEEE Signal Processing Mag.*, vol. 14, pp. 37–62, May 1997.

[71] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley, 1994.

[72] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *Journal of Mathematical Physics*, vol. 25, pp. 261–278, 1947.

[73] L.-M. Li and L. B. Milstein, "Rejection of narrow-band interference in PN spread-spectrum systems using transversal filters," *IEEE Trans. Commun.*, vol. COM-30, pp. 925–928, May 1982.

[74] ——, "Rejection of CW interference in QPSK systems using decision-feedback filters," *IEEE Trans. Commun.*, vol. COM-31, pp. 473–483, Apr. 1983.

[75] ——, "Rejection of pulsed CW interference in PN spread-spectrum systems using complex adaptive filters," *IEEE Trans. Commun.*, vol. COM-31, pp. 10–20, Jan. 1983.

[76] T. Li, W. H. Mow, V. K. N. Lau, M. Siu, R. S. Cheng, and R. D. Murch, "Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 566–575, Apr. 2007.

[77] A. S. Ling and L. B. Milstein, "Trade-off between diversity and channel estimation errors in asynchronous MC-DS-CDMA and MC-CDMA," *IEEE Trans. Commun.*, vol. 56, pp. 584–597, Apr. 2008.

[78] R. W. Lowdermilk and fredric j. harris, "Interference mitigation in OFDM," in *5th IEEE International Conference on Universal Personal Communication*, Cambridge, MA, Sept. 1996, pp. 623–627.

[79] O. M. Macchi and N. J. Bershad, "Adaptive recovery of a chirped sinusoid in noise, Part 1: Performance of the RLS algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 583–594, Mar. 1991.

[80] O. M. Macchi, N. J. Bershad, and M. Mboup, "Steady-state superiority of LMS over LS for time-varying line enhancer in noisy environment," *Proc. Inst. Elect. Eng., Part F. (Radar and Signal Processing)*, vol. 138, pp. 354–360, Aug. 1991.

[81] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[82] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. Boston, MA: McGraw Hill, 2000.

[83] J. L. Massey, *Threshold Decoding*. Cambridge, MA: MIT Press, 1963.

[84] L. B. Milstein, "Interference rejection techniques in spread spectrum communications," *Proc. IEEE*, vol. 76, pp. 657–671, June 1988.

[85] L. B. Milstein and R. A. Iltis, "Signal processing for interference rejection in spread spectrum communications," *IEEE ASSP Mag.*, vol. 3, pp. 18–31, Apr. 1986.

[86] P. Monsen, "Adaptive equalization of the slow fading channel," *IEEE Trans. Commun.*, vol. COM-22, pp. 1064–1075, Aug. 1974.

[87] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 2000.

[88] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier systems?" *IEEE Trans. Commun.*, vol. 50, pp. 2136–2148, Dec. 2002.

[89] J. I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 282–287, June 1967.

[90] C. R. Nassar, B. Natarajan, Z. Wu, D. Wiegandt, S. A. Zekavat, and S. Shattil, *Multi-carrier technologies for wireless communications*. Boston, MA: Kluwer Academic Publishers, 2002.

[91] R. Nilsson, F. Sjöberg, and J. P. LeBlanc, "A rank-reduced LMMSE canceller for narrowband interference suppression in OFDM-based systems," *IEEE Trans. Commun.*, vol. 51, pp. 2126–2140, Dec. 2003.

[92] J. P. Odenwalder, "Error Control Coding Handbook," Linkabit Corporation, July 1976.

[93] J. K. Omura, "On the Viterbi decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 177–179, Jan. 1969.

[94] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & Systems*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1997.

[95] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Denver, CO, Apr. 1980, pp. 964–967.

[96] H. V. Poor and L. A. Rusch, "Narrowband interference suppression in spread spectrum CDMA," *IEEE Personal Commun. Mag.*, vol. 1, pp. 14–27, Third Quarter 1994.

[97] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw Hill, 2001.

[98] J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing.* New York, NY: MacMillian, 1998.

[99] K. J. Quirk, L. B. Milstein, and J. R. Zeidler, "A performance bound for the LMS estimator," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1150–1158, May 2000.

[100] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.

[101] A. J. Redfern, "Receiver window design for multicarrier communication systems," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 1029–1036, June 2002.

[102] M. J. Reed and B. Liu, "An analysis of LMS adaptive two-sided transversal filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, May 1991, pp. 2145–2148.

[103] F. Reira-Palou, J. M. Noras, and D. G. M. Cruickshank, "Analysis of the decision delay effect on the convergence of gradient recursive decision feedback equalizers," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, May 2002, pp. 2665–2668.

[104] M. Reuter and J. R. Zeidler, "Nonlinear effects in LMS adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 47, pp. 1570–1579, June 1999.

[105] M. Saito, S. Moriyama, and S. Nakahara, "Experimental and simulation results of an OFDM modem for TV broadcasters," *SMPTE Journal*, pp. 39–45, Jan. 1996.

[106] S. G. Sankaran, "Implementation and Evaluation of Echo Cancellation Algorithms," Master's thesis, Virgina Tech, Blacksburg, VA, Dec. 1996.

[107] E. H. Satorius, S. Krishnan, X.-L. Yu, L. J. Griffiths, I. S. Reed, and T.-K. Truong, "Suppression of narrowband interference via signal channel adaptive preprocessing," in *Conference Record of the 23rd Asilomar Conference on Circuits Systems and Computers*, Pacific Grove, CA, Oct. 1988, pp. 270–273.

[108] G. J. Saulnier, "Suppression of narrowband jammers in a spread-spectrum receiver using transform-domain adaptive filtering," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 742–749, May 1992.

[109] G. J. Saulnier, Z. Ye, and M. J. Medley, "Performance of a spread spectrum OFDM system in a dispersive fading channel with interference," in *Proceedings of IEEE Military Communications (MILCOM) Conference*, Boston, MA, Oct. 1998, pp. 679–683.

[110] M. J. Shensa, "Time constants and learning curves of LMS adaptive filters," Naval Ocean Systems Center, San Diego, CA, Tech. Rep. 312, 1978.

[111] ——, "The spectral dynamics of evolving LMS adaptive filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Washington D.C., Apr. 1979, pp. 950–953.

[112] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Aug. 1994, unpublished. [Online]. Available: http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf

[113] D. C. Shin and C. L. Nikias, "Adaptive interference canceler for narrowband and wideband interferences using higher order statistics," *IEEE Trans. Signal Processing*, vol. 42, pp. 2715–2728, Oct. 1994.

[114] D. Shklarsky, P. K. Das, and L. B. Milstein, "Adaptive narrowband interference suppression," in *Conference Record of the National Telecommunications Conference*, Washington, DC, Nov. 1967, pp. 15.2.1 – 15.2.4.

[115] B. Sklar, "Rayleigh fading channels in mobile digital communications systems – Part I: Characterization," *IEEE Commun. Mag.*, vol. 35, pp. 90–100, July 1997.

[116] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter-banks and multichannel linear prediction," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, Adelaide, Apr. 1994, pp. 585–588.

[117] J. E. Smee and N. C. Beaulieu, "Error-rate evaluation of linear equalization and decision feedback equalization with error propagation," *IEEE Trans. Commun.*, vol. 46, pp. 656–665, May 1998.

[118] C. Snow, L. Lampe, and R. Schober, "Error rate analysis for coded multicarrier systems over quai-static fading channels," *IEEE Trans. Commun.*, vol. 55, pp. 1736–1746, Sept. 2007.

[119] ——, "Interference mitigation for coded MB-OFDM UWB," in *Proceedings of the IEEE Radio and Wireless Symposium*, Orlando, FL, Jan. 2008, pp. 17–20.

[120] ——, "Impact of WiMAX interference on MB-OFDM UWB systems: Analysis and mitigation," *IEEE Trans. Commun.*, submitted for publication. [Online]. Available: http://www.ece.ubc.ca/~csnow/publications/preprints/snow_2007_WiMAX_interference.pdf

[121] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[122] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, 1st ed. Philadelphia, PA: Prentice Hall, 1997.

[123] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Boston, MA: Kluwer, 2001.

[124] F. W. Symons, "Narrow-band interference rejection using the complex linear prediction filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 94–98, Feb. 1978.

[125] S. C. Thompson, "CE-OFDM," Ph.D. dissertation, University of California, San Diego, La Jolla, CA, June 2005.

[126] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, 1st ed. Philadelphia, PA: Soc for Industrial & Applied Math, 1997.

[127] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53–62, Feb. 1979.

[128] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

[129] T. Wang, J. G. Proakis, E. Masry, and J. R. Zeidler, "Performance degradation of OFDM systems due to doppler spreading," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1422–1432, June 2006.

[130] P. C. Wei, J. Han, J. R. Zeidler, and W. H. Ku, "Comparative tracking performance of the LMS and RLS algorithms for chirped narrowband signal recovery," *IEEE Trans. Signal Processing*, vol. 50, pp. 1602–1609, July 2002.

[131] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 628–634, Oct. 1971.

[132] S. B. Wicker, *Error Control Systems for Digital Communication Systems*, 1st ed. Upper Saddle River, NJ: Prentice Hall, 1995.

[133] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Heard, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.

[134] N. Wiener and E. Hopf, "On a class of singular integral equations," *Proceedings of the Prussian Academy, Math. Phys. Series*, p. 696, 1931.

[135] Wolfram Research. [Online]. Available: http://functions.wolfram.com/01.10.23.0004.01

[136] J. M. Wozencraft and B. Reiffen, *Sequential Decoding.* Cambridge, MA: MIT Press, 1961.

[137] Z. Wu and C. R. Nassar, "Narrowband interference rejection in OFDM via carrier interferometry spreading codes," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1491–1505, July 2005.

[138] N. Yee and J.-P. M. G. Linnartz, "Wiener filtering of multi-carrier CDMA in a Rayleigh fading channel," in *Proceedings of the IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC) Conference*, The Hague, Sept. 1994, pp. 1344–1347.

[139] N. Yee, J.-P. M. G. Linnartz, and G. Fettweis, "Multi-carrier CDMA in indoor wireless radio," in *Proceedings of the IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC) Conference*, Yokahama, Dec. 1993, pp. 109–113.

[140] J. R. Zeidler, "Performance analysis of LMS adaptive prediction filters," *Proc. IEEE*, vol. 78, pp. 1781–1806, Dec. 1990.

[141] J. R. Zeidler, E. H. Satorius, D. M. Chabries, and H. T. Wexler, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 240–254, June 1978.