EQUALIZING WITH FRACTIONALLY-SPACED CONSTANT MODULUS AND SECOND-ORDER-STATISTICS BLIND RECEIVERS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Thomas Joseph Endres

May 1997

EQUALIZING WITH FRACTIONALLY-SPACED CONSTANT MODULUS AND SECOND-ORDER-STATISTICS BLIND RECEIVERS

Thomas Joseph Endres, Ph.D.

Cornell University 1997

This dissertation is concerned with the blind, cold-start equalization of a closed-eye, fractionally-spaced channel, so that the error rate is sufficiently reduced for transfer to a decision directed mode. First, a simulated comparison study of recently proposed second-order-statistics (SOS) algorithms and the constant modulus algorithm (CMA) shows a severe lack of robustness for the SOS algorithms to practical situations, while to the contrary, CMA appears quite robust. Next, a class of cyclic all-pass channels is shown to be the only channel class whose cascade with a core channel results in no change in the SOS indicators used by many SOS algorithms. This ambiguity raises robustness concerns for SOS algorithms due to the possibility of channel mis-identification.

The sequel focuses on the robustness properties of fractionally-spaced CMA. Though the literature contains reasonable treatment for CMA's robustness properties to violation of three of the four perfect equalization assumptions, essentially no analytical work exists for the undeniably practical case when the equalizer length is less than that needed for perfect equalization. Original algebraic analysis

is presented for the constant modulus (CM) cost function, describing the deformation of the CM error surface due to undermodeling. The analysis is connected to previous work on CMA misadjustment, suggesting that similar to noisy LMS, a longer fractionally-spaced equalizer (FSE) is not always better than a shorter FSE.

Next, a truncated Taylor series of the binary CM cost function is used to estimate the location of the undermodeled CM minima. A measure is proposed from this estimate to determine the proximity of CM and mean squared error (MSE) minima, suggesting that those CM minima with better MSE performance stay closer to their corresponding Wiener solutions than those CM minima with worse MSE performance.

This dissertation can be used to determine design guidelines for length selection of a FSE updated by CMA (CMA-FSE) given specific signalling formats and channel models. Together, this work and the recent work of others can be used to establish a cohesive behavioral theory of CMA-FSE describing the robustness properties of the CM criterion to practical situations.

Biographical Sketch

Thomas Joseph Endres was born in Youngstown, Ohio on October 8, 1966. He received a B. S. in Electrical Engineering from Cornell University in 1990, and then worked in El Segundo, CA for Hughes Aircraft Company on frequency-hopped communication systems. While working full time at Hughes, he started part time schooling at the University of Southern California, and earned a M. S. in Electrical Engineering in August, 1994 supported by a Hughes Master's Fellowship. He returned to Cornell University in August, 1994 to pursue a Ph.D. in Electrical Engineering, supported by a Hughes Doctoral Fellowship.

To Joe and Viv: my parents, my inspiration, my friends.

To Jerry Garcia: Long ago under a hot Texas sun I heard the strangest sound emanating from my sister's room. What did I know, I thought that tone was normal and that everyone knew about it. It took a number of years (and a seventh grade math teacher) to point out my departure from mainstream. Luckily, I wasn't bothered by my apparent differences, and chalked it up to re-ordered priorities. But these priorities remained–I came to Ithaca in 1985 hoping you would play Barton Hall one more time; I chose electrical engineering hoping to make better tapes; I took a job in California to see more JGB shows; I came back to Ithaca when I saw the end approaching. My hope was to have this dissertation finished and hand you a copy at the GD office. But nothing's for certain, and it can always go wrong, so I'll keep it until we meet again.

Acknowledgements

The author acknowledges the financial support of the Hughes Aircraft Company (El Segundo, CA) in the form of doctoral and master's fellowships. Thanks to those managers who signed on the dotted line and made this education a reality. Thanks also to Applied Signal Technology (Sunnyvale, CA) for their technical assistance and interest in my research, and also for their gifts to the C. U. BERG. Thanks also for assistance from NSF Grant MIP-9509011.

I am truly indebted to Professor Brian D. O. Anderson of the Australian National University for taking me under his wing. The majority of work in this dissertation would not exist without Brian's world-class expertise and general willingness to help. Still, in Brian's absence I could not do it alone; thanks to Dr. Steve Halford formerly of the University of Virginia for invaluable discussions and countless hours of computer babysitting needed for the results in Chapter 3.

Thanks also to those who shared the title "graduate student" with me, and helped in more ways with this dissertation than they will ever know: J. P. LeBlanc, R. A. Casas, and P. B. Schniter–I am lucky to call you peers.

Thanks to sister Cindy for her selfless support when stipend checks were thin. I am deeply indebted to my family for their incredible support and encouragement over the past thirty years. From the starting line to the kickoff, my family never failed to show.

Thanks to my friends who have seen my best and worst–you know who you are. Especially, thanks to Gux, Walther and Esmerelda at GPM Engineering (San Marcos, CA) for motivational support.

When the going gets weird,

the weird turn pro.

–thanks doc

Thanks to SRV: what a loss, what a shame, what a guitar player.

Thanks to Jim Behm (DoD) for editorial help with this manuscript. Thanks also to Professor J. Thorp (Electrical Engineering) and O. Rothaus (Mathematics) for their time and effort in serving on my committee.

Finally, thanks to my advisor and mentor Professor C. R. Johnson, Jr. (Rick). I feel sorry for those graduate students who are unable to experience graduate school in the way which you make possible. To say that this experience has been cultural as well as educational is an understatement. Shucks, it's been fun. If at all this dissertation or my time at C. U. resemble "efficiency," it is solely because Rick knew the right questions to ask. I now believe he saw how all this material fit together in the summer of '95. It just took me this long to sort it all out.

Inspiration, move me brightly

Light the song with sense and color

Hold away despair.

More than this I will not ask

faced with mysteries dark and vast

statements just seem vain at last ...

Table of Contents

List of Tables

List of Figures

List of Abbreviations

The notation used in this dissertation is easily understood. Bold face, lower case letters designate vectors, while bold face, upper case letters designate matrices. All vectors are assumed to be column vectors. A scalar is usually written in lower case italics; for example, a. A scalar element of a vector is denoted by $v_i \in \mathbf{v}$, which is read as "the i^{th} coefficient of vector $\mathbf{v}$," with indexing of vectors starting at $i = 0$ and ranging to the length of the vector minus one. Time is denoted in parenthesis; for example, $\mathbf{v}(k)$ is interpreted as vector $\mathbf{v}$ at time k. Some of the commonly used abbreviations are described below for reference.

$(\cdot)^T$	Transposition
$(\cdot)^*$	Conjugation
$(\cdot)^H$	Transposition and conjugation, or Hermitian transpose
$\lceil\cdot\rceil$, $\lfloor\cdot\rfloor$	Round up, down to nearest integer
$\mathbf{c}$	Channel coefficient vector
$\mathbf{f}$	Equalizer coefficient vector
$\mathbf{s}$	Source vector
$\mathbf{n}$	Additive white gaussian noise vector
$\mathbf{h}$	Combined channel-equalizer coefficient vector

$\mathbf{h}_{\downarrow}$	Decimated (to baud-spaced) combined channel-equalizer vector
$\mathbf{h}_{\delta}$	Pure delay vector; all zeros except a 1 at δ^{th} position
$\mathbf{C}$	Channel convolution matrix
$\mathbf{C}_{\downarrow}$	Row-decimated channel convolution matrix
L_c	Length of channel impulse response
L_f	Length of equalizer
$\mathbf{I}_n$	$n \times n$ Identity matrix
N	Oversample factor (typically 2)
P	Definition depends on section of dissertation
T	Baud interval or symbol spacing
AWGN	Additive white gaussian noise
BSE	Baud-spaced equalizer
CM	Constant Modulus
CMA	Constant Modulus Algorithm
FSE	Fractionally-spaced equalizer
CMA-FSE	Fractionally-spaced equalizer updated by the constant modulus algorithm
i.i.d.	Independent and identically distributed
ISI	Intersymbol interference
MSE	Mean-squared error
P.E.	Perfect equalization
SEP	Somebody else's problem
SNR	Signal to noise ratio
SOS	Second-order-statistics

Chapter 1

Introduction

The chiefs of old are gone. Myself, I take courage.

–Sitting Bull

Perhaps the American Indians were the first to build a digital communications system when they announced the running of the buffalo with smoke signals. Little could they have imagined the evolution of their idea, though today's digital communication systems retain the same basic principles.

Information is encoded into members of a finite alphabet; for example $\mathcal{A} = \{-3, -1, +1, +3\}$. Each member of this alphabet, or *symbol*, is used every T seconds to excite an analog pulse shape, so that the magnitude of the symbol is analogous to the height of the smoke signal. This type of signalling is therefore called pulse amplitude modulation (PAM). The time between adjacent symbols, T, is referred to as the *baud interval*. When the baud interval is small, more smoke signals are sent more often, so that adjacent smoke signals become smeared and perhaps indistinguishable. In modern digital communications systems which transmit millions of symbols per second from more dense constellations than the simple 4-PAM alphabet above, this smearing of adjacent symbols is inevitable due to the dispersive nature of the bandlimited propagation channel, and referred to as *Intersymbol Interference* (ISI).

Bello [7] suggested that the multipath channel be modeled as a linear, discrete-time, finite impulse response (FIR) transfer function, so that with $k = nT$, $n = 0, 1, 2 \ldots$, the current received sample is a finite sum of current and past transmitted symbols, i.e., $r(k) = \sum_{i=0}^{L_c-1} c_i s(k-i)$, where $c_i \in \mathbf{c}$ and $\mathbf{c}$ is the length-L_c vector of channel impulse response coefficients (see Figure 1.1). To remove the ISI resulting from non-trivial channel dynamics, the receiver typically uses a finite-length, linear, tapped-delay-line called an *equalizer*, $\mathbf{f}$ in Figure 1.1.

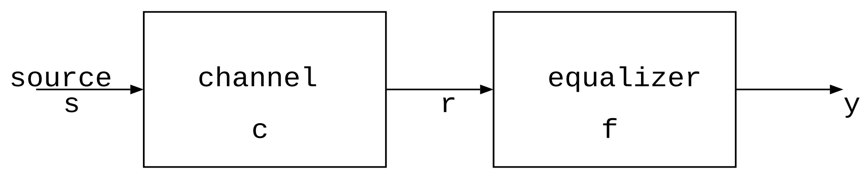


Figure 1.1: Simplified Block Diagram

To see that a linear, FIR equalizer can aid in removing the ISI introduced by a noiseless, FIR channel, consider the following example. The white, source sequence is drawn uniformly from a 16-Quadrature Amplitude Modulation (QAM) alphabet (which may be thought of as $\mathcal{A} \chi j\mathcal{A}$, where χ denotes Cartesian product [35] and $\mathcal{A}$ is the 4-PAM alphabet above) and filtered through the length-8, complex channel whose coefficients are shown in magnitude in the left plot of Figure 1.2. A length-8 equalizer operates on the distorted, received samples (center plot of Figure 1.3) to produce reasonable estimates (right plot of Figure 1.3) of the transmitted symbols (left plot of Figure 1.3). The equalizer output does not exactly equal the transmitted sequence since the equalizer does not remove all ISI from the received sequence. Equivalently, the combined channel-equalizer response (right plot of Figure 1.2) is not precisely a pure delay, but a reasonable approximation to a pure delay.

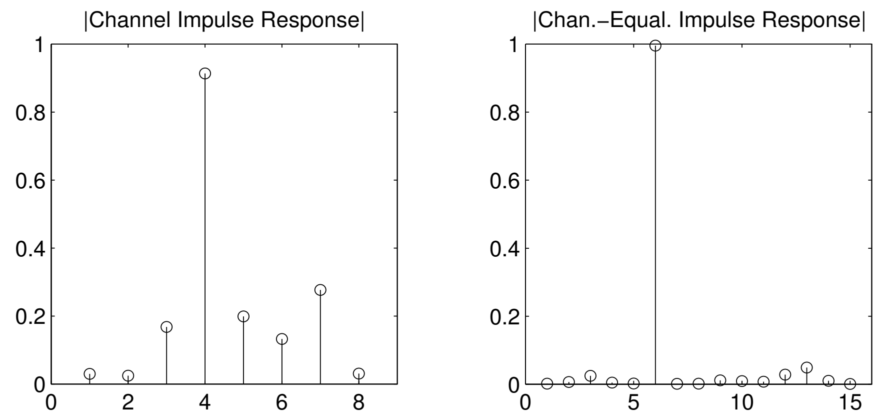


Figure 1.2: Illustrative Example: Impulse Response Magnitudes for Channel and Combined Channel-Equalizer

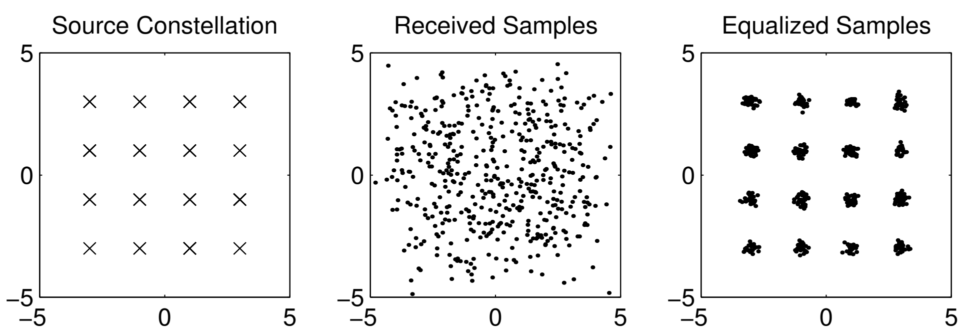


Figure 1.3: Illustrative Example: Source, Received, and Equalized Samples

In most high speed, high capacity applications, the propagation channel is not known *a priori* and is possibly time varying, requiring adaptive solutions in the receiver design. A classical approach in an adaptive receiver is the transmission of a pre-arranged training sequence [94] which is known *a priori* at the transmitter and receiver. This information can be used at the receiver to estimate the source sequence, though periodic re-training may be necessary. Certain applications where a training sequence is either too costly in usable bandwidth or where a training sequence is impractical require a receiver design which operates on the received signal and possibly some statistics of the source, but not the actual source sequence itself. Such an approach is termed *blind*. The goal of the blind algorithm may be viewed as a reduction in the error rate to an acceptable threshold from a cold-start initialization, so that a quantized version (nearest-element decision device) of the equalizer output provides a reliable estimate of the source sequence. When the source estimate is sufficiently reliable, it may replace the training signal and the blind, adaptive algorithm is successful. This second mode of operation which uses the source estimate instead of the source itself is thus termed *decision directed* (DD) and was proposed in [60].

In the preceding example, the equalizer is found from the known channel dynamics and approximates a delayed inverse of the channel. The received sequence is synchronously sampled at the baud rate; the adjacent taps in both $\mathbf{c}$ and $\mathbf{f}$ therefore have a spacing corresponding to the baud interval, T, and the approach is termed *baud-spaced* equalization. Originally introduced to aid in synchronization efforts (see [91]), *fractionally-spaced equalizers* (FSEs) sample the received waveform at a rate higher than the baud rate and have received considerable attention in the recent literature due to their ability to perfectly equalize (zero the ISI of) a possibly non-minimum phase channel with a finite length equalizer under certain assumptions to be discussed in the sequel (see [78]). Such a feat is not possible in the baud-spaced case–the equalizer which achieves perfect equalization is, in general, infinitely long (see [29]).

One recent result for the fractionally-sampled case is the ability to identify a linear, possibly non-minimum phase channel using purely the second-order-statistics (SOS) of the received signal–this feat is not possible in the baud-spaced case. Gardner [26] and then Tong et al. [75] showed that when sampling the received signal at a rate corresponding to $T/2$ or greater, then a system of equations could be solved exactly for the unknown channel (within a gain factor) under certain assumptions. Since the introduction of [26] and [75], the literature is replete with new, improved SOS algorithms and supposed convergence benefits ([62], [70], [78], [79], [96], [28], and numerous others). Robustness studies for these SOS algorithms are rare; [15], [17], and [80] are exceptions.

The focus of this dissertation is the robustness properties of fractionally-spaced, blind, adaptive algorithms which are used from a cold-start initialization to reduce the error rate to a prescribed threshold for transfer to a DD mode of operation. SOS algorithms are compared to the Constant Modulus Algorithm (CMA), which was originally proposed by Godard in [30] and developed independently by Treichler and Agee in [88]. CMA uses a third-order moment of the received signal in directly updating the equalizer coefficients. Since the amount of tolerable mean squared error (MSE) which defines the Transfer Level is a function of the signalling scheme used, noise power present and in general the specific application, it is difficult to produce meaningful, generic results and design guidelines. Consequently, throughout this dissertation we rely on channel models derived from empirically measured microwave radio signals [32] which are now available in a database described in

§3.2 in the sequel. Further, our examples typically use 16-QAM signalling when possible, since this alphabet is representative of the dense, complex alphabets often used in practice. With these admissions, the reader may view this dissertation as a template or manual to address algorithm robustness properties, and replace our database models, signalling format, SNR, etc. with the particulars of his/her own problem and thereby (hopefully) produce meaningful design guidelines and intuition. Appendix A, for example, provides a case study of a cable TV channel based on our analysis assuming 256-QAM signalling.

This dissertation is organized into eight chapters. Chapter 2 formulates the blind equalization problem and discusses optimal solutions, with a brief overview of existing adaptive equalization strategies. It is shown that under a set of assumptions, an equalization setting exists which removes all ISI. These assumptions, however, are impractical for dense, high-data-rate signalling.

Chapter 3 provides a comparison study between CMA and recently-proposed SOS algorithms focused on practical situations. The results suggest severe robustness concerns for the SOS algorithms to situations which are essentially inevitable in high-data-rate communications. Conversely, CMA appears quite robust to practical situations.

Chapter 4 identifies a class of cyclic all-pass channels whose cascade with a core channel results in no change in the SOS indicators used for channel identification by many SOS algorithms. It is shown by construction how to produce example channels which pose serious robustness concerns for these SOS algorithms. Next, the channel identification task based on root-extraction from SOS indicators is briefly studied for the specific case where the length of the channel estimate is less than that of the true channel.

The remainder of the dissertation focuses on CMA's robustness properties. Chapter 5 reviews CMA's robustness properties in a pedagogical way: the existing literature is reviewed for CMA robustness results to each of the four assumptions required for global convergence to a perfect equalization setting. (These assumptions are discussed in Chapter 2.) It is shown that the literature contains reasonable treatment for three of the four assumptions, but essentially no treatment of the length condition imposed on the FSE. Chapters 6 and 7 address CMA's robustness properties to violation of this length condition.

Chapter 6 provides an algebraic analysis of the real Constant Modulus (CM) cost function when the length condition is not satisfied, which suggests that when channel coefficients outside the time span of the FSE are "small," the CM local minimum of best MSE performance stays in a tight neighborhood of the optimal Wiener setting. The analysis is evaluated for the database channel models and a design rule for FSE length selection is proposed for these channels.

Chapter 7 uses calculus to derive a quadratic approximation of the binary CM cost function. When the length condition is not satisfied, the CM local minima are approximated from the quadratic cost, and found to be perturbations to the Wiener solutions. A measure based on these results is proposed to study CM performance for sub-optimal system delays. When evaluated for the database channel models, this measure suggests that the CM local minima with better MSE performance stay in closer proximity to their corresponding Wiener solutions than those CM minima with poorer MSE performance. Chapter 8 contains concluding remarks and some possible directions of future work.

Appendix A contains a short case study. We stray from the database channels and consider a cable TV channel model. Many of our robustness results from the preceding chapters are reproduced for this different channel model. Based on these results, a design guideline for FSE length selection is proposed and tested assuming 256-QAM signalling. This appendix is included to show that our analysis which is applied to the database channels throughout this dissertation can be easily extended to other channel models and signalling formats.

Chapter 2

The Blind Equalization Problem

This chapter discusses the system theoretic description assumed throughout the sequel, including a discussion of the mean squared error (MSE) criterion and adaptive solutions typically used in achieving certain performance measures. The majority of material in this chapter is not original work, though provides a fuller understanding of the blind equalization/identification problem statement, as well as information needed for the implementation of the indirect adaptive algorithms to be discussed in Chapters 3 and 4. Indeed, once a channel is identified, Tong et al. claim in [75] "the estimation of the information symbols ... becomes more or less a classical problem, and various equalization techniques ... can be applied." Much of the information in this chapter is our interpretation of this "classical problem."

2.1 Problem Description

Though carrier and baud synchronization are essential functions in a successful receiver design, we assume they are somebody else's problem (SEP) (see pp. 24 of [2] for further discussion on SEPs) so that a baseband-equivalent block diagram from transmitter to source-estimate emerges (Figure 2.1). An excellent discussion of classical synchronization strategies, however, can be found in [59], and general phaselock techniques can be found in [25]. Figure 2.1 is the genesis of our discussion on equalization/identification strategies. Should the reader find this starting point lacking motivation from the communications task or alignment with the current open literature, he/she is referenced to [85], which describes "...the data communications problem, the rationale for introducing fractionally-spaced equalizers, ... and their implications," as well as further discussions regarding Figure 2.1.

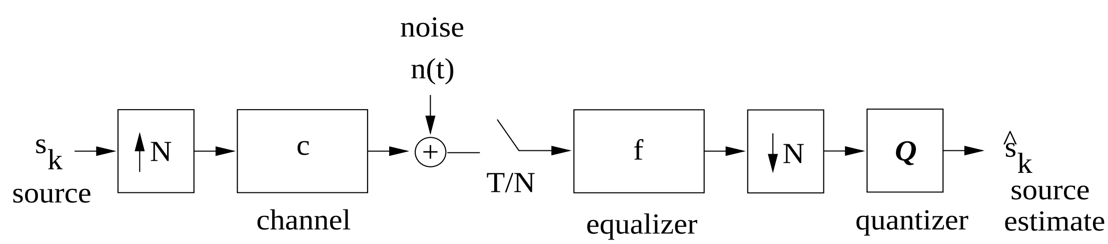


Figure 2.1: System Block Diagram

The received sequence is the superposition of a distorted version of the source sequence, $s(k)$, due to non-trivial channel dynamics and of an additive noise sequence, $n(k)$, which is typically assumed to be drawn independently and identically distributed (i.i.d.) from a gaussian distribution. The equalizer coefficient vector, $\mathbf{f}$, is assumed to be a linear, tapped-delay-line structure, since this implementation is commonly used in practice (see [85]) due to the linear dependence of the equalizer output on the equalizer coefficients. Other equalizer structures, such as decision feedback equalizers (DFEs), which feed back a filtered version of the source estimate to the received sequence, generally lack robustness from a cold-start initialization (see [10] and [47]) and are beyond the scope of this dissertation.

The scalar N defines the oversample ratio, i.e., how much faster the received signal is sampled relative to the baud spacing, T. Though non-integer values may be used in implementation, $N = 2$ is typical and fractional sampling in the sequel may be assumed to be $T/2$-spaced unless otherwise stated. The quantizer is a nearest-element decision device.

Observe that the simple block diagram in Figure 2.1 assumes that the channel model, $\mathbf{c}$, contains those dynamics which are in the actual propagation channel (for example, due to multipath) as well as any filtering at the transmitter which may or may not be known at the receiver (for example, pulse shaping filters).

The ultimate goal in receiver design is usually the minimization of the probability of the occurrence of an incorrect decision, or symbol error rate (SER). The average error rate, however, is usually a highly non-linear function which prohibits the simple, adaptive solutions desired for modern high data rate, dense signalling formats. For example, with additive white gaussian noise (AWGN), analytic evaluation of the SER involves a complementary error function, *erfc* (see [63], [8], or [69]). Maximum likelihood sequence estimation using the Viterbi algorithm [93] is typically too complex and too slow for these applications [29]. Other, more simple structures and performance measures are thus usually used in hopes that their minima closely align with the minimum SER setting.

2.2 MMSE Equalizer Design

Though complexity and speed typically limit implementation of a receiver design which in general minimizes SER (namely the Viterbi algorithm), under certain assumptions about the noise statistics, the Wiener receiver which minimizes the MSE also minimizes the maximum likelihood (ML) probability of error. With further assumptions on the source statistics, the minimum MSE (MMSE) receiver also minimizes the maximum *a posteriori* (MAP) probability of error (see [69] or [55]). We therefore study MMSE equalizer design and adopt MSE as the performance measure to be used in the sequel.

2.2.1 Baud-Spaced Equalization

When $N = 1$ in Figure 2.1, the received waveform is assumed to be synchronously sampled at the same rate as the (baud-spaced) source sequence. The channel and equalizer models are completely described by their vectors of impulse response coefficients, $\mathbf{c}$ and $\mathbf{f}$, respectively. Adjacent elements of these vectors therefore have a relative delay equal to the baud interval, T. The combined channel-equalizer response, $\mathbf{h}$, is the convolution of the two vectors, $\mathbf{c}$ and $\mathbf{f}$. This convolution can be represented as the inner product of a channel convolution matrix, $\mathbf{C}$, associated with vector $\mathbf{c}$, and the equalizer vector, $\mathbf{f}$; i.e., $\mathbf{h} = \mathbf{Cf}$. For example, let the length-L_c channel impulse response be $\mathbf{c} = [c_0\ c_1 \dots c_{L_c-1}]^T$ and the length-L_f equalizer impulse response be $\mathbf{f} = [f_0\ f_1 \dots f_{L_f-1}]^T$. With $P = (L_c + L_f - 1)$, the length-P combined channel-equalizer response, $\mathbf{h}$, is described by the system of equations

$$\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_{P-1} \end{bmatrix} = \begin{bmatrix} c_0 & & & \\ c_1 & c_0 & & \\ \vdots & c_1 & \ddots & \\ c_{L_c-1} & \vdots & \ddots & c_0 \\ & c_{L_c-1} & \ddots & c_1 \\ & & \ddots & \vdots \\ & & & c_{L_c-1} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{L_f-1} \end{bmatrix} \tag{2.1}$$

or

$$\mathbf{h} = \mathbf{C}\mathbf{f} \tag{2.2}$$

where the $P \times L_f$ channel convolution matrix, $\mathbf{C}$, is Töeplitz (see [39]). Adjacent elements of $\mathbf{h}$ also have a relative delay of the baud interval, T. Observe that the rows of the system of equations in (2.1) (and hence the rows of $\mathbf{C}$) correspond to different delays, or powers of z^{-1}.

For most digital communications applications, the desired system response from transmitter to receiver is a pure delay, $z^{-\delta}$, for some integer δ. Define $\mathbf{h}_\delta$ as the length-P unit column vector with zeros in all positions but index δ, which has a 1; $\mathbf{h}_\delta$ thus represents a pure delay. Observe that for the baud-spaced case, the system of equations in (2.1) is always over-determined with respect to the equalizer coefficients, so that in general, a finite-length equalization setting does not exist which gives zero ISI. In other words, a pure delay is not achievable. In general, a pure delay in the absence of noise for the baud-spaced case requires an infinite-length equalizer (see [29]).

In the presence of zero-mean AWGN and a zero-mean, equi-probable, i.i.d. source, the variance in the error between the source and equalizer output can be written as (see [42] or [8])

$$\sigma_e^2 = (\mathbf{h}_\delta - \mathbf{C}\mathbf{f})^H(\mathbf{h}_\delta - \mathbf{C}\mathbf{f})\sigma_s^2 + \mathbf{f}^H\mathbf{f}\sigma_n^2 \tag{2.3}$$

Observe the two contributors to the MSE; one term is the resulting ISI between the actual and desired system responses, the second term is enhancement of the noise power by the squared ℓ_2-norm of the equalizer, $\mathbf{f}^H\mathbf{f}$. The squared ℓ_2-norm of the equalizer is therefore usually referred to as the *noise gain*.

The design objective is the choice of a finite number of equalizer coefficients, $\mathbf{f}$, which minimize the MSE in (2.3). Solution to this length-constrained least squares formulation may be found from pseudo-inversion of (2.1) or gradient and Hessian calculation of (2.3), or simply "completing the square." The solution for these methods is the MMSE equalizer parameterization (see [42])

$$\mathbf{f}^{\dagger} \quad = \quad (\mathbf{C}^H \mathbf{C} + \frac{\sigma_n^2}{\sigma_s^2}\, \mathbf{I}_{L_f})^{-1}\, \mathbf{C}^H \mathbf{h}_\delta \tag{2.4}$$

In the absence of noise, the MMSE solution approximates a delayed inverse of the channel. In the presence of noise, the MMSE equalizer compromises between zeroing the ISI of the combined channel-equalizer response and enhancing the noise from an equalizer setting which has large noise gain. Channels which have zeros near the unit circle have deep nulls in their frequency response and a relatively long inverse; it is well known that these channels therefore pose conditioning problems to baud-spaced equalizer (BSE) design (see pp. 373 of [8], for example).

2.2.2 Fractionally-Spaced Equalization

When $N > 1$ in Figure 2.1, the received waveform is sampled at a rate higher than the baud rate, and the system is fractionally-spaced. We restrict our attention to a $T/2$-spaced system. We next derive a multi-channel block diagram from the polyphase decomposition of the fractionally-spaced system. For more details of the resulting multi-channel system diagram, see [70],[22] or [90] to name just a few.

The analog source signal is described by

$$s(t) \quad = \quad \sum_{i=-\infty}^{\infty} a(i)\delta(t - iT). \tag{2.5}$$

where $a_i \in \mathcal{A}$ for source alphabet $\mathcal{A}$. In the absence of noise, the received analog waveform is the convolution of the channel analog impulse response (including propagation dynamics and transmit filters), $c(t)$, with the source impulse sequence in (2.5)[1],

$$\begin{aligned} r(t) &= \int_{-\infty}^{\infty} s(\tau)c(t-\tau)d\tau & (2.6)\\ &= \sum_{i=-\infty}^{\infty} \int_{-\infty}^{\infty} a(i)\delta(\tau-iT)c(t-\tau)d\tau & (2.7)\\ &= \sum_{i=-\infty}^{\infty} a(i)c(t-iT) & (2.8)\end{aligned}$$

where perfect synchronization is assumed. The received waveform is sampled every $T/2$ seconds yielding the sequence

$$r(k\frac{T}{2}) = \sum_{i=-\infty}^{\infty} a(i)c(k\frac{T}{2}-iT) \qquad (2.9)$$

Even and odd k define two baud-spaced subsequences of the fractionally-sampled received waveform. Specifically, for k even (or $k = 2n, n = 0, 1, 2, \ldots$),

$$\begin{aligned} r_n^{even} &= r(nT) & (2.10)\\ &= \sum_{i=-\infty}^{\infty} a(i)c((n-i)T) & (2.11)\end{aligned}$$

and for k odd (or $k = 2n-1$),

$$\begin{aligned} r_n^{odd} &= r(nT-\frac{T}{2}) & (2.12)\\ &= \sum_{i=-\infty}^{\infty} a(i)c((n-i)T-\frac{T}{2}). & (2.13)\end{aligned}$$

Observe that these two subsequences are precisely the result of a polyphase decomposition on the $T/2$-spaced channel impulse response (see [92]). Let the $T/2$-spaced channel be described by $\mathbf{c} = [c_0\ c_1\ \ldots\ c_{L_c-1}]^T$. Define two baud-spaced vectors

[1]Strictly speaking, we cannot interchange the order of summation and integration in forming $r(t)$ (see [73] or [5]). However, practically speaking, the number of source symbols is large, but finite, and the channel is FIR, and therefore absolutely integrable. Hence, (2.7) is justified.

from $\mathbf{c}$ with a relative delay of $T/2$ seconds from one another,

$$\begin{aligned} \mathbf{c}_{even} &:= [c_0\ c_2\ c_4\ \ldots\ c_{L_c-1}]^T \\ \mathbf{c}_{odd} &:= [0\ c_1\ c_3\ \ldots\ c_{L_c-1}]^T \end{aligned} \tag{2.14}$$

This formulation suggests the multichannel model in Figure 2.2.

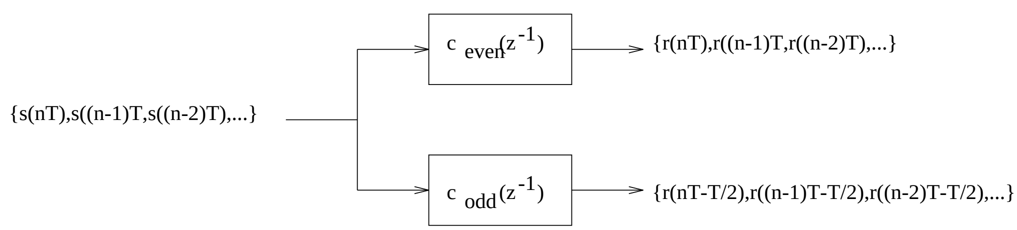


Figure 2.2: SubChannel Model Resulting From Polyphase Decomposition

Consider the $T/2$-spaced output of the equalizer,

$$y(k\frac{T}{2}) = \sum_{i=0}^{L_f-1} f(i\frac{T}{2})r((k-i)\frac{T}{2}) \tag{2.15}$$

Either the even or odd samples correspond to "on-baud" symbols; without loss of generality, we assume the even output samples ($k = 2n$) correspond to the "on-baud" symbols and are retained after decimation to a baud-spaced output sequence. (For an analogous derivation which assumes that the odd samples correspond to "on-baud" signalling, see [41].) The baud-spaced equalizer output therefore becomes

$$y(nT) = \sum_{i=0}^{\frac{L_f-1}{2}} f(iT)r((n-i)T) + f(iT+\frac{T}{2})r((n-i)T-\frac{T}{2}) \tag{2.16}$$

Notice that the sum is broken into two *baud-spaced* convolutions of the even and odd received sequences with even and odd subequalizers, defined as

$$\mathbf{f}_{even} = f(nT) \tag{2.17}$$

$$= \quad [f(0), f(T), f(2T), \cdots] \tag{2.18}$$

$$\mathbf{f}_{odd} \quad = \quad f(nT - \frac{T}{2})] \tag{2.19}$$

$$= \quad [f(\frac{T}{2}), f(3\frac{T}{2}), f(5\frac{T}{2}), \cdots] \tag{2.20}$$

These definitions suggest the multi-input, single-output split equalizer shown in Figure 2.3.

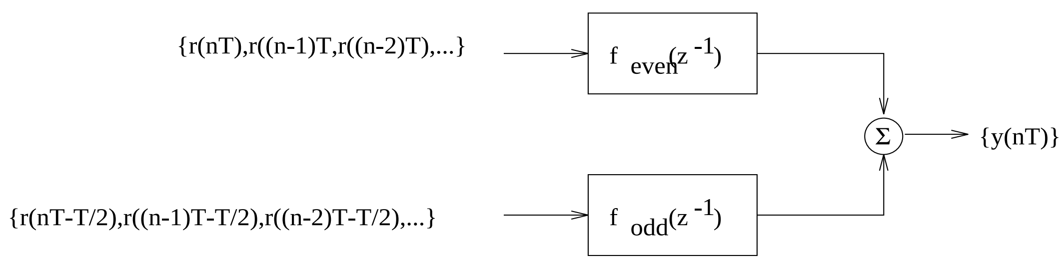


Figure 2.3: Split Equalizer Model Resulting From Polyphase Decomposition

Combining Figures 2.2 and 2.3 suggests the T-spaced Input-Output model in Figure 2.4, which is characterized by the transfer function,

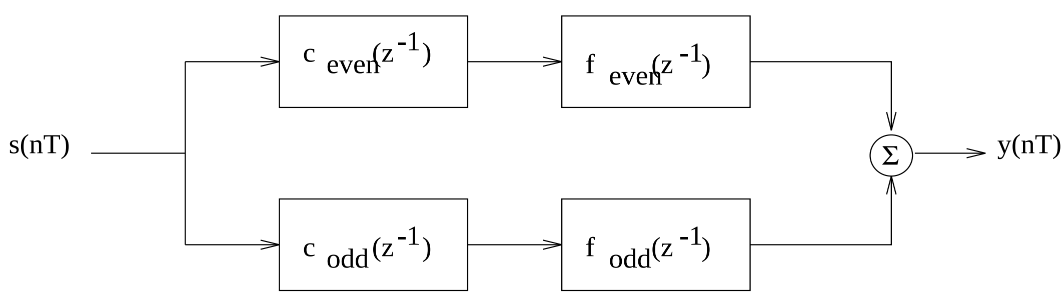


Figure 2.4: T-spaced Input-Output Model of Oversampled Channel and FSE

$$\mathbf{h} \quad = \quad \mathbf{c}_{even}\mathbf{f}_{even} + \mathbf{c}_{odd}\mathbf{f}_{odd} \tag{2.21}$$

Recognize this form as mathematically equivalent to the integer-valued Diophantine equation, which when restricted to a polynomial ring as in our case is usually attributed to Bezout (see [46]). Contrast the design equation in (2.21) with that for the baud-spaced case in (2.1); where the BSE approximates the inverse of the channel, the FSE is in general *not* the inverse of the $T/2$-spaced channel.

Solution of the Bezout identity follows in an analogous way to the baud-spaced case. Let the $T/2$-spaced channel and equalizer impulse response coefficients be described by $\mathbf{c} = [c_0\ c_1 \ldots c_{L_c-1}]^T$ and $\mathbf{f} = [f_0\ f_1 \ldots f_{L_f-1}]^T$, respectively. Before decimation, the combined channel-equalizer response can be written identically to that in (2.1), the distinction being that the relative delay between adjacent rows in the system of equations now corresponds to half the baud interval or $T/2$, whereas in (2.1) the relative delay is T. However, since symbols are transmitted at the baud rate and not every $T/2$ seconds, the design objective now becomes the choice of $\mathbf{f}$ to solve (2.1) with every other row removed, or a row-decimated version of (2.1). For example, retaining the even rows which correspond to the "on-baud" samples and (without loss of generality) assuming L_c is even, (2.1) now becomes

$$\begin{bmatrix} h_0 \\ h_2 \\ h_4 \\ \vdots \\ h_{P-1} \end{bmatrix} = \begin{bmatrix} c_0 & & & & & & \\ c_2 & c_1 & c_0 & & & & \\ \vdots & c_3 & c_2 & c_1 & c_0 & & \\ c_{L_c-2} & \vdots & \vdots & \ddots & \ddots & & \\ & c_{L_c-1} & c_{L_c-2} & \ddots & \ddots & & \\ & & & \ddots & \ddots & c_1 & c_0 \\ & & & \ddots & \ddots & c_3 & c_2 \\ & & & \ddots & \ddots & \vdots & \vdots \\ & & & & & c_{L_c-1} & c_{L_c-2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{L_f-1} \end{bmatrix}$$

or

$$\mathbf{h}_{\downarrow} \;=\; \mathbf{C}_{\downarrow}\mathbf{f} \tag{2.22}$$

The baud-spaced combined channel-equalizer, $\mathbf{h}_{\downarrow}$, and the channel convolution matrix, $\mathbf{C}_{\downarrow}$, are appropriately row-decimated versions of $\mathbf{h}$ and $\mathbf{C}$, respectively. In general, the number of equations in (2.22) is $\lceil\frac{L_f+L_c-1}{2}\rceil$ if the even rows correspond to the "on-baud" samples, or $\lfloor\frac{L_f+L_c-1}{2}\rfloor$ if the odd rows correspond to the "on-baud" samples, where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ denote round up or down to the nearest integer, respectively. The presence of the last column in $\mathbf{C}_{\downarrow}$ depends if L_f is even or odd–as it is drawn above, L_f is assumed odd.

It should be noted that $\mathbf{C}_{\downarrow}$ is a generalized Sylvester matrix of the two subchannels (see [45]) and therefore can be written in different forms. Another form of $\mathbf{C}_{\downarrow}$ commonly seen in the literature is found with a re-ordering of the equalizer coefficients as if from stacking the two subequalizer vectors; $\mathbf{f} = [f_0\ f_2 \ldots\ f_1\ f_3\ \ldots]^T$. In this case, the alternative version of $\mathbf{C}_{\downarrow}$ (=$[\mathbf{C_e} \mid \mathbf{C_o}]$) is block Töeplitz, where the two blocks, $\mathbf{C_e}$ and $\mathbf{C_o}$, are the convolution matrices associated with the two subchannels and hence each look like the matrix in (2.1) (for example, see [62]). This alternative version perhaps reflects the Bezout identity in (2.21) more closely than the row-decimated version in (2.22), though the two forms result from elementary row and column operations and are therefore equivalent.

The fact that the number of equations to be satisfied is reduced due to row-decimation (for a given FSE length) admits the possibility that the system of equations becomes exactly determined, i.e., the equalizer length can be chosen so that the row-decimated channel convolution matrix is square and the system of equations can be exactly solved when this matrix is invertible (full-column rank).

Hence, the FSE when appropriately long (and under other assumptions to be discussed in §2.2.4) can be chosen to achieve a pure delay, and is said to achieve *perfect equalization* (P.E.). The appropriate length equalizer to achieve P.E. depends on whether the even or odd samples correspond to the "on-baud" samples. It can be shown that if the even rows are retained upon decimation, the P.E.-length FSE is $L_f = L_c - 1$, while if the odd rows are retained upon decimation, the P.E.-length FSE is $L_f = L_c - 2$.

When the FSE is not long enough to exactly equal the multichannel inverse needed to solve (2.22), we seek an approximate solution. The design objective is the choice of FSE which minimizes the MSE, which in the presence of AWGN for a white, symmetric source is expressed as

$$\arg\min_{\mathbf{f}} \left\{ \sigma_e^2 \right\} = \arg\min_{\mathbf{f}} \left\{ (\mathbf{h}_\delta - \mathbf{C}_\downarrow \mathbf{f})^H (\mathbf{h}_\delta - \mathbf{C}_\downarrow \mathbf{f}) \sigma_s^2 + \mathbf{f}^H \mathbf{f} \sigma_n^2 \right\} \tag{2.23}$$

which has solution (see [42])

$$\mathbf{f}^\dagger = (\mathbf{C}_\downarrow{}^H \mathbf{C}_\downarrow + \frac{\sigma_n^2}{\sigma_s^2} \, \mathbf{I}_{L_f})^{-1} \, \mathbf{C}_\downarrow{}^H \mathbf{h}_\delta \tag{2.24}$$

One might argue that in the presence of noise, the FSE be chosen longer than the channel, in order that the extra degrees of freedom be used to find a minimum norm solution and therefore minimize the noise gain. Such schemes exist which propose over-estimation of the FSE length ([1], for example), though such a choice of equalizer length is impractical for the high data rates of interest in this dissertation. For these applications, the FSE length is always chosen to violate the length condition needed for P.E. due to hardware constraints.

2.2.3 Optimum Delay Choice

The design variables for design of an equalizer according to (2.4) or (2.24) include the FSE length, L_f, and the delay, δ, in the baud-spaced, combined channel-equalizer response. The MSE performance is admittedly sensitive to this delay choice; [8] says on pp. 373, "...the performance of a finite equalizer does depend on the choice of (δ)." More recently, [97] has demonstrated MSE performance dependence on the overall system delay. A general guideline for the choice of δ based on Cramer-Rao bounds is suggested in [27], though the optimum choice is admittedly not guaranteed. A technique for adaptively adjusting the delay to minimize the MSE was proposed in [64], though its convergence is not guaranteed. A rule is proposed in [8] which requires simple implementation, but achieves sub-optimum performance.

We present here a closed form solution for the optimum system delay which minimizes the MSE for a given channel and SNR estimate for both baud-spaced and fractionally-spaced cases. This derivation[2] first appeared in the appendix of [16].

Let the optimum equalizer setting (in a MMSE sense) be described by

$$\mathbf{f}^{\dagger} \quad = \quad (\mathbf{C}^H\mathbf{C} + \frac{\sigma_n^2}{\sigma_s^2}\,\mathbf{I}_{L_f})^{-1}\,\mathbf{C}^H\mathbf{h}_{\delta} \tag{2.25}$$

where with little fear of confusion, $\mathbf{C}$ is the channel convolution matrix with a relative delay between adjacent rows equal to the baud interval; i.e., $\mathbf{C}$ may be a row-decimated version of the matrix described in (2.1) if using a FSE, but we drop the down-arrow subscript.

[2]Thanks to P. Bert Schniter for suggesting simplifications from the original derivation.

Define $\mathbf{M} := (\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})^{-1}$ where $\lambda = \frac{\sigma_n^2}{\sigma_s^2}$ is the noise-to-signal ratio, and recognize that we drop the subscript on $\mathbf{I}$ for simplicity. Observe that $\mathbf{M}^{-1}$, and hence $\mathbf{M}$, are Hermitian symmetric (see pp. 48 of [72]). The resulting MSE can be written as

$$\begin{aligned}
\sigma_e^2(\delta) = & \qquad (2.26) \\
& (\mathbf{h}_\delta - \mathbf{C}\mathbf{f}^\dagger)^H(\mathbf{h}_\delta - \mathbf{C}\mathbf{f}^\dagger)\sigma_s^2 + (\mathbf{f}^\dagger)^H(\mathbf{f}^\dagger)\sigma_n^2 \\
= & \; (\mathbf{h}_\delta - \mathbf{CMC}^H\mathbf{h}_\delta)^H(\mathbf{h}_\delta - \mathbf{CMC}^H\mathbf{h}_\delta)\sigma_s^2 \qquad (2.27) \\
& + (\mathbf{MC}^H\mathbf{h}_\delta)^H(\mathbf{MC}^H\mathbf{h}_\delta)\sigma_n^2 \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[(\mathbf{I} - \mathbf{CMC}^H)^H(\mathbf{I} - \mathbf{CMC}^H) + \lambda\mathbf{CM}^H\mathbf{MC}^H\right]\mathbf{h}_\delta \qquad (2.28) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - 2\mathbf{CM}^H\mathbf{C}^H + \mathbf{CM}^H\mathbf{C}^H\mathbf{CMC}^H + \lambda\mathbf{CM}^H\mathbf{MC}^H\right]\mathbf{h}_\delta \qquad (2.29) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - 2\mathbf{CM}^H\mathbf{C}^H + \mathbf{CM}^H(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})\mathbf{MC}^H\right]\mathbf{h}_\delta \qquad (2.30) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - 2\mathbf{CMC}^H + \mathbf{CM}(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})\mathbf{MC}^H\right]\mathbf{h}_\delta \qquad (2.31) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - 2\mathbf{CMC}^H + \mathbf{CMM}^{-1}\mathbf{MC}^H\right]\mathbf{h}_\delta \qquad (2.32) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - 2\mathbf{CMC}^H + \mathbf{CMC}^H\right]\mathbf{h}_\delta \qquad (2.33) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - \mathbf{CMC}^H\right]\mathbf{h}_\delta \qquad (2.34) \\
= & \; \sigma_s^2\mathbf{h}_\delta{}^H\left[\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})^{-1}\mathbf{C}^H\right]\mathbf{h}_\delta \qquad (2.35)
\end{aligned}$$

Recall that the desired response, $\mathbf{h}_\delta$, is a pure delay. The effect of this response is therefore to select the main diagonal element of the matrix $\left[\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})^{-1}\mathbf{C}^H\right]$ corresponding to row and column δ. Since the objective is selection of δ corresponding to a MMSE setting, the optimum delay choice corresponds to the minimum main diagonal element of the matrix $\left[\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})^{-1}\mathbf{C}^H\right]$, or

$$\delta^\dagger = \arg\min_\delta \left\{\left[\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I})^{-1}\mathbf{C}^H\right]_{\delta,\delta}\right\} \qquad (2.36)$$

Table 2.1: Optimum System Delay in Baud Intervals for Channels 1 and 3

Channel	32-tap FSE	64-tap FSE
channel 1	26	28
channel 3	17	13

To see the sensitivity of the MSE on the design variable, δ, consider two $T/2$-spaced channel models from the database which is described in §3.2 of the sequel. Figures 2.5 and 2.6 (channels 1 and 3, respectively) each contain three plots. The top plot is the magnitude of the complex valued, $T/2$-spaced channel impulse response coefficients. The middle and bottom plots show MSE performance calculated according to (2.3) (where $\mathbf{C}$ is appropriately row-decimated) versus the possible system delays for length-32 and length-64 FSE's found according to (2.25). Note that both of these examples correspond to the case where the FSE length is too short to perfectly solve the Bezout identity for a pure delay, so there necessarily exists a non-zero MSE.

Other examples from the database look similar to these figures. The "trough" appears to be the approximate width of the FSE time span and positioned approximately with the center of gravity of the channel vector. (This same behavior is observed in the cable TV channel of Appendix A.) The optimum system delays for the different cases predicted by (2.36) when the even channel coefficients correspond to "on-baud" signalling are shown in Table 2.1 and agree with the figures. These examples demonstrate strong MSE dependence on the system delay.

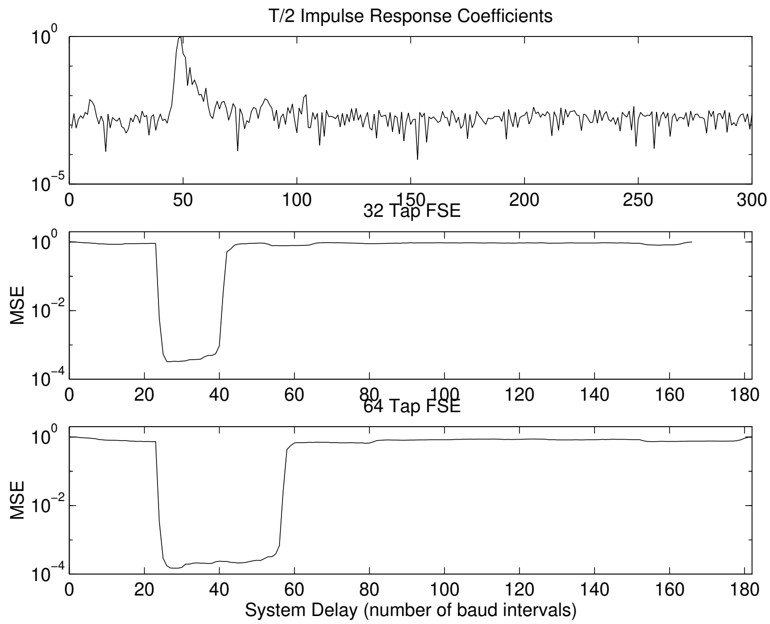


Figure 2.5: MMSE Sensitivity to System Delay for Channel 1

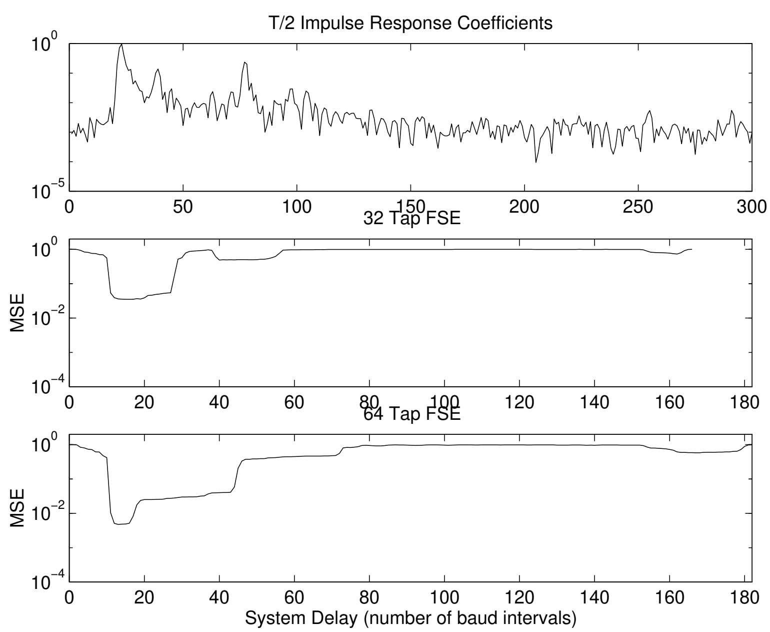


Figure 2.6: MMSE Sensitivity to System Delay for Channel 3

2.2.4 Perfect Equalization Assumptions

As mentioned in §2.2.2, the fact that the FSE length may be chosen to exactly solve the FSE design equation, (2.21) or (2.22), admits the possibility of a combined channel-equalizer response which is a pure delay. This constraint on the FSE length can be seen as one of four assumptions which are required for P.E. using a blind receiver (see [13], [89], [77], [79], [21], [71] [62], and others). We formalize the P.E. assumptions below, specifically for the $T/2$-sampled case.

P.E. Assumptions:

1. $\mathbf{A_1}$ (Source Distribution):

   The source sequence is zero-mean, i.i.d. and equi-probable.

2. $\mathbf{A_2}$ (Additive Noise):

   There is no additive channel noise.

3. $\mathbf{A_3}$ (Channel Disparity):

   There exist no roots of the $T/2$-spaced channel which are symmetric with respect to reflection through the origin. It is shown in [45] that this condition is equivalent to roots shared by the baud-spaced subchannels, i.e., reflected $T/2$-spaced channel roots $\Longleftrightarrow$ common subchannel roots.

4. $\mathbf{A_4}$ (FSE Length):

   The FSE length is chosen (i) greater than or equal to $L_c - 1$ if "on-baud" signalling corresponds to the even channel coefficients, or (ii) greater than or equal to $L_c - 2$ if "on-baud" signalling corresponds to the odd channel coefficients (see §2.2.2).

Various chapters in the sequel study by simulation and analysis the robustness properties of popular blind equalization/identification algorithms when these P.E. assumptions are not satisfied. Our main interest, however, is reserved specifically for algorithm robustness to violation of the length condition in $\mathbf{A_4}$, since the number of FSE coefficients in a tapped-delay-line implementation translates directly into hardware cost (see [85]).

2.2.5 Sensitivity to Equalizer Length

Faster is better today, and this view is unlikely to change too soon, so that the data rates employed in a digital communications system are typically much faster than the rate at which a bulk of equalizer coefficients can be reliably updated at the receiver. This condition necessarily implies the violation of the length condition in $\mathbf{A_4}$ for practical applications. We therefore study the sensitivity of channel order undermodeling on the MMSE performance of the length-constrained Wiener solution found via (2.25) with system delay chosen according to (2.36). The $T/2$-spaced channel model used in this experiment is Channel 3 from the database which is described in §3.2 of the sequel. Figure 2.7 contains two plots. The top plot shows the magnitude of the $T/2$-spaced channel coefficients. The bottom plot shows the MMSE achievable from the Wiener solution versus FSE length for various SNR's.

This figure suggests a staircase-like dependence on FSE length. Note that similar results for the database channels can be found in [49], where the authors plot a measure derived from SER performance versus FSE length. Our goal is a blind, adaptive algorithm which is robust to channel order undermodeling, that can be used from a cold-start to reduce the MSE to a prescribed threshold for transfer to a DD mode. The behavior in Figure 2.7 suggests that dependent on the Transfer

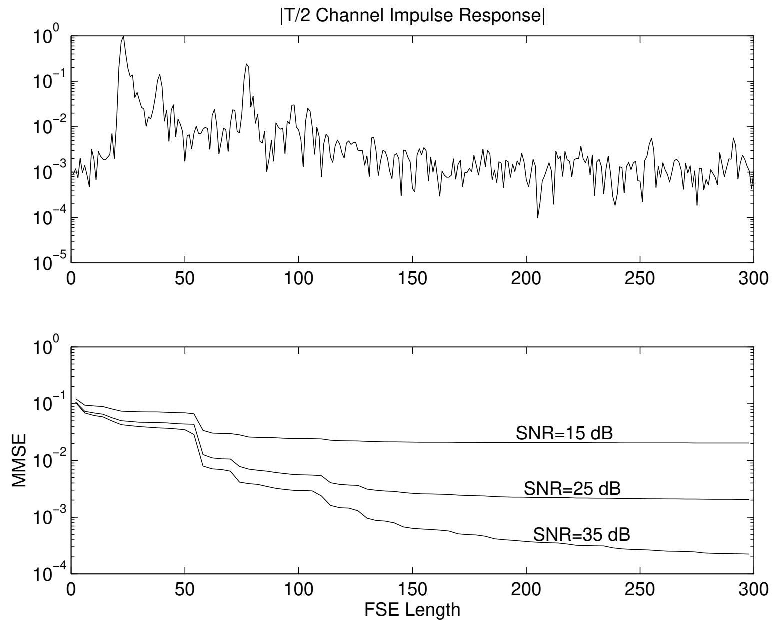


Figure 2.7: MMSE Sensitivity to FSE Length for Channel 3

Level MSE (which is a function of the signalling and SNR, see Appendix A,) an equalization setting exists to achieve this threshold which requires far fewer taps than that needed for P.E. in the absence of noise. In fact, by considering sub-optimal system delays, there exist many equalization settings which can achieve a prescribed Transfer Level MSE. For example, Figure 2.8 displays the MMSE surface as function of the two variables, L_f and δ with 100 dB SNR for Channel 3 of the database. With a Transfer Level MSE of $0.076 = -11.2$ dB which corresponds to 16-QAM signalling (see Appendix A), Figure 2.8 suggests the choice of FSE length and system delay is forgiving–there exist many settings which give acceptable performance. Another way to see this fact is to consider the contour line of constant MSE (corresponding to -11.2 dB) drawn in the $L_f - \delta$ plane (see Figure 2.9). The many settings (FSE length and system delay) "below" this contour line yield acceptable MSE performance with respect to the prescribed threshold.

We next review some existing adaptive equalization strategies which attempt to find such desirable settings, and then test some of these ideas in the next chapter.

2.3 Adaptive Solutions

The optimum (MMSE) equalizer solution found via (2.25) is predicated on knowledge of the channel impulse response. Typical applications attempt iterative estimation of this channel impulse response (hence called *indirect adaptive equalization*) or iterative estimation of the equalizer coefficients themselves (hence called *direct adaptive equalization*). We present here a brief discussion of those algorithms which are studied in the sequel. For fuller discussions on equalization/identification strategies, [65] is an award winning survey on non-blind techniques, [37] is the first compilation-style book devoted specifically to blind deconvolution, and [14] is an excellent tutorial on blind equalization.

2.3.1 Trained LMS

With little doubt, the most popular adaptive algorithm is the least mean squares (LMS) algorithm, which uses a stochastic gradient search to transcend the MSE cost surface (see [60], [61], or [87] for fuller discussions on LMS in an equalization context, and see [36], [3], or [95] for fuller discussions on LMS not necessarily in an equalization context).

The gradient descent scheme to update the equalizer coefficients is given in some generality by

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu \nabla_{\mathbf{f}} J \tag{2.37}$$

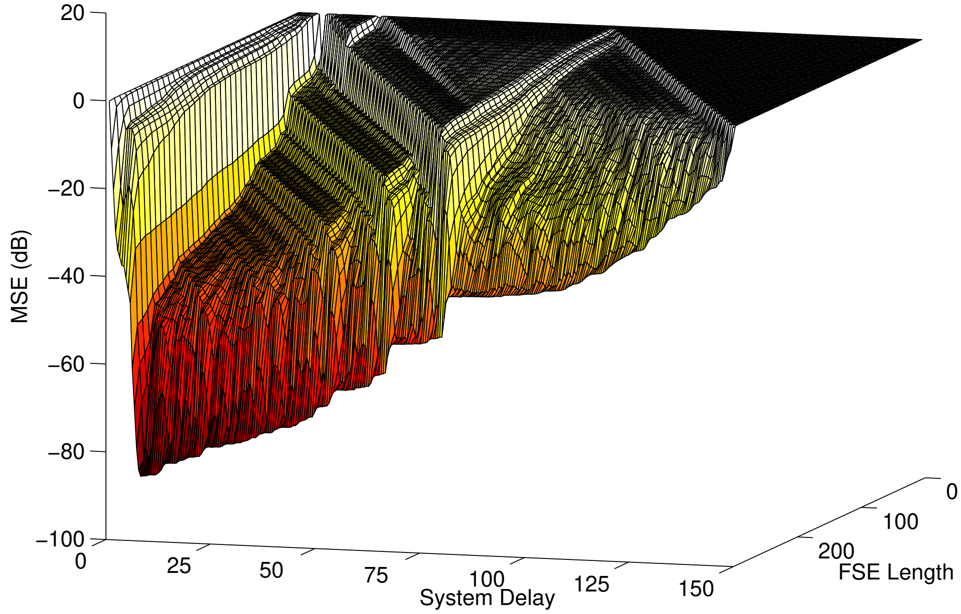


Figure 2.8: MMSE as a Function of L_f and δ, Channel 3

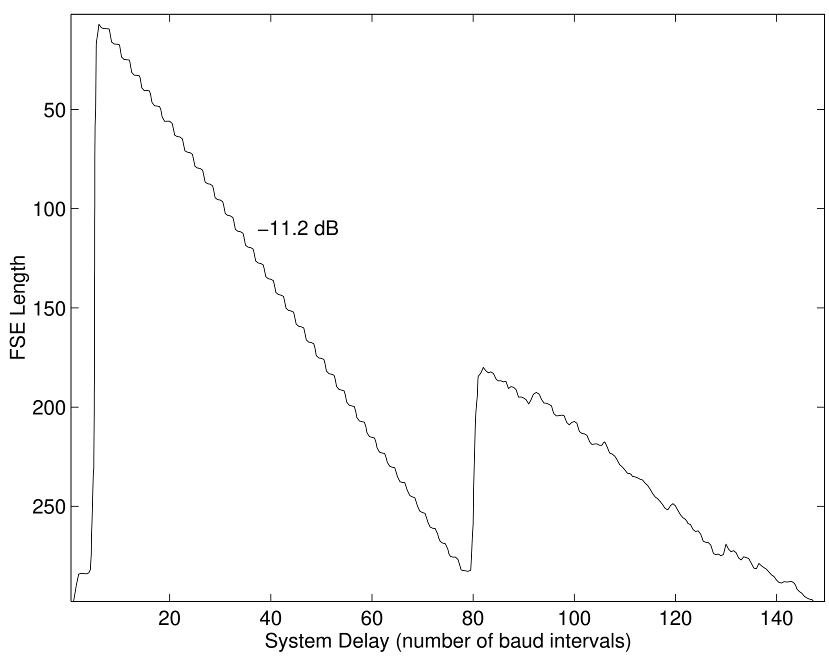


Figure 2.9: Line of Constant MSE in $L_f - \delta$ Plane, Channel 3

where k is the time index, μ is a small, positive number, $\nabla_{\mathbf{f}}$ is the gradient operator with respect to the length-L_f equalizer coefficient vector at time k, and J is the cost function to be minimized. Further define the vector of data in the equalizer tapped-delay-line at time k (commonly called the regressor vector) as

$$\mathbf{r}(k) \quad := \quad [r(k)\ r(k-1) \ldots r(k-L_f+1)]^T \tag{2.38}$$

so that the equalizer output can be written as the inner product

$$y(k) \quad = \quad \mathbf{r}^T(k)\mathbf{f}(k) \tag{2.39}$$

For the MSE criterion, $J = \frac{1}{2}E\{|s(k) - y(k)|^2\}$. For simplicity we assume no delay between transmitter and demodulator output, or $\delta = 0$. (See [28] for discussions of LMS and LMS-like algorithms with non-zero δ.) It can be shown that the gradient of the MSE cost with respect to the equalizer vector becomes

$$\nabla_{\mathbf{f}} J \quad = \quad -E\left\{(s(k) - y(k))\mathbf{r}^*(k)\right\} \tag{2.40}$$

The adaptive algorithm is obtained by replacing this true gradient expression with its instantaneous approximation, (equivalently, dropping the expectation operator), so that the LMS algorithm emerges

$$\mathbf{f}(k+1) \quad = \quad \mathbf{f}(k) + \mu(s(k) - y(k))\mathbf{r}^*(k) \tag{2.41}$$

$$= \quad \mathbf{f}(k) + \mu e_{MSE}(k)\mathbf{r}^*(k) \tag{2.42}$$

where $e_{MSE}(k) = s(k) - y(k)$ is the instantaneous error.

Immediately apparent from (2.42) is the necessity of a pre-arranged pilot or training sequence between the transmitter and receiver, since the LMS algorithm requires $s(k)$ to form $e_{MSE}(k)$. This training sequence is typically a fixed portion of the source sequence which is known *a priori* at the transmitter and receiver, but does not itself contain information. This training sequence may be sent solely during a start-up phase to establish a link, or periodic re-training may be necessary in other applications. In many emerging applications, the penalty in usable bandwidth is too high (as for broadcast formats), or a training sequence is simply impractical (as for "uncooperative" environments) so that a receiver design which does not require the source signal itself is desirable. Still, LMS enjoys widespread maturity both analytically and empirically and is the standard to which other algorithms are compared.

2.3.2 Blind Equalization

Blind equalization/identification algorithms operate on the received sequence and possibly some information about the source sequence, but not the actual source sequence itself. Dropping the requirement of a reference signal has spawned applications which previously existed only in a James Bond film.

Decision Directed LMS

LMS can be shown to adapt the equalizer coefficients on average to the MMSE solution in the absence of measurement noise from any equalizer initialization. This global convergence property and LMS's simplicity account for its analytic and empirical maturity, despite its need for a reference signal. Lucky [60] recognized, however, that if the error rate is sufficiently small so that the decisions are correct "most of the time," the source sequence in (2.42) can be replaced with its estimate, $\hat{s}$, the output of the decision device. The blind algorithm is thus termed Decision Directed LMS (DD-LMS). Reliability of the algorithm clearly depends on the reliability of the source estimate, $\hat{s}$. Typically, another blind algorithm operates from a cold start to reduce the error rate sufficiently and then is transferred to DD-LMS [85]. The amount of acceptable error which defines the Transfer Level depends on the signalling used (see Appendix A).

Second-Order Statistics

It is easily shown that for the baud-spaced case, second-order statistics (SOS) of the output signal are, in general, insufficient to identify the phase of a linear time-invariant system. One exception is the impractical case where the input signal is purely gaussian. Since most practical channel models are mixed phase (neither maximum nor minimum phase) (see [32]) many blind algorithms traditionally rely on higher-order moments of the received signal. Indeed, Benveniste et al. claim in [9] in regards to identification of a mixed phase channel, "hence, second-order statistics are irrelevant to our problem."

Fractional spacing, as it turns out, offers more than just help in synchronization efforts. Gardner [26] first showed that fractional sampling induced a cyclostationarity[3] in the received sequence which could be used to resolve the phase of an unknown channel model using purely the SOS of the received signal. (The received signal sampled at the baud rate is shown to produce a wide sense *stationary* process.) Tong et al. then developed numerous blind, indirect equalization algorithms [75], [76], [78], [79] based on estimation of the autocorrelation function of the received signal. Moulines et al. [62] exploit the orthogonality of signal and noise subspaces of the received correlation matrix. This subspace technique is of lower computational burden than Tong et al., though still requires an esti-

[3]By cyclostationary, we mean that the fractionally-spaced received signal satisfies $E\{r(k_1)r^*(k_2)\} = E\{r(k_1+T)r^*(k_2+T)\}$.

mate of the time-varying, received-signal correlation matrix and subsequent eigen-decomposition for channel estimation. Slock et al. [70], [71] relate the equalization problem to perfect-reconstruction filter bank theory and show that the channel can be identified from the received signal SOS by linear prediction. Xu et al. [96] propose an indirect scheme which treats the input sequence as a deterministic, rather than stochastic (but still unknown) sequence, in hopes that the algorithm be robust to arbitrary source statistics. Baccalá and Roy [4] propose a reduced-complexity algorithm in the time domain. Giannakis and Halford [28] propose a computationally simple, LMS-like blind algorithm based on a stochastic gradient descent structure, among other algorithms of various complexities. More recently, [33] proposes a modification of [62] which accounts for correlated noise. Hardly an issue of *IEEE Transactions on Signal Processing* goes by that is untouched by the SOS pen.

The swell in the recent open literature on blind equalization/identification algorithms based on SOS is attributable to the idea that estimates of second-order moments converge with fewer number of observations than estimates of higher-order moments. Hence, an algorithm based purely on SOS should provide convergence rate benefits over the algorithm based on higher-order statistics (HOS). Though the literature is replete with new and improved SOS algorithms, it is surprisingly sparse in addressing the robustness of these algorithms ([12], [15], [17], [16], [80], and [89] are exceptions). Chapters 3 and 4 of this dissertation specifically address the natural robustness concerns of some of the above SOS algorithms and show a severe lack of robustness to practical situations.

Constant Modulus Algorithm (CMA)

The Constant Modulus (CM) criterion was first proposed by Godard in [30] and developed independently by Treichler and Agee in [88]. The stochastic gradient descent implementation, or Constant Modulus Algorithm (CMA), is widely used in practice (see [85]). CMA is arguably the most popular blind algorithm for cold start-up of a tapped-delay-line equalizer structure (see [63]). Godard's original intention was to develop an algorithm for phase and amplitude-modulated signals (for example, QAM) which decoupled equalization and carrier recovery so that carrier phase tracking could be accomplished at the equalizer output in a DD mode. Treichler and Agee's original intention was to develop a criterion which sensed multipath-induced AM on an otherwise constant envelope FM signal.

The CM criterion attempts to fit a power of the modulus of the equalizer output to a constant. This constant is chosen to essentially project all constellation points onto a circle. Mathematically, from [30] this criterion is expressed as

$$J_{CM} = \frac{1}{4}E\{(|y(k)|^p - R_p)^2\} \tag{2.43}$$

for some integer, p. The case $p = 1$ is usually attributed to Sato [66]. We study solely the case $p = 2$ in the sequel and henceforth CM and CMA refer to the $p = 2$ case, which is expressed as

$$J_{CM} = \frac{1}{4}E\{\left(|y(k)|^2 - \gamma\right)^2\} \tag{2.44}$$

The CM dispersion constant is defined as $\gamma = \frac{E\{|s|^4\}}{E\{|s|^2\}}$. Godard shows that this value of γ minimizes the CM cost.

The objective is a gradient search implementation of the CM cost in (2.44) according to the rule in (2.37). It is shown in [88] that the gradient of J_{CM} with respect to the equalizer coefficient vector is

$$\nabla_{\mathbf{f}} J_{CM} \quad = \quad E\{\left(|y(k)|^2 - \gamma\right) \cdot y(k)\mathbf{r}^*(k)\} \tag{2.45}$$

where $\mathbf{r}$ is the regressor vector. CMA is obtained in an analogous way to LMS– by replacing the true gradient with its instantaneous approximation (equivalently, dropping the expectation operator),

$$\mathbf{f}(k+1) \quad = \quad \mathbf{f}(k) + \mu\left(\gamma - |y(k)|^2\right) y(k)\mathbf{r}^* \tag{2.46}$$

Observe that upon defining $e_{CM}(k) = (\gamma - |y(k)|^2)\, y(k)$, the algorithm is precisely the form of LMS in (2.42) when $e_{MSE}(k)$ $(= s(k) - y(k))$ is replaced with $e_{CM}(k)$. The connection between the CM criterion and the MSE criterion is strong. Treichler and Agee [88] show that "...in the vicinity of convergence, minimizing the constant modulus performance function is equivalent to minimizing the mean squared error." Godard [30] shows that for an infinite length equalizer in the absence of noise, "...perfect equalization is one steady-state solution of the adaptation process." As shown later in [6] and [24], in the absence of noise with an independent and uniformly distributed source, an *infinitely* long BSE adapted by CMA is globally convergent to a P.E. setting. For the fractionally-spaced case, CMA is globally convergent to a P.E. setting with a *finite* length equalizer under the P.E. assumptions discussed in §2.2.4 (see [56] or [20]). We will address CMA's robustness properties to violation of the four P.E. assumptions in Chapters 5, 6 and 7.

Both [30] and [88] recognize the multi-modality of the CM error surface. LeBlanc [51] counts the fractionally-spaced CM stationary points (including minima, saddles, and maximum) under an i.i.d. source assumption as $3^{L_c+L_f-1}$. The multi-modal surface results from the decoupling of magnitude and phase in the CM criterion. Since the CM criterion operates purely on the magnitude of the equalizer output, not the phase, the phase shift of the adaptive filter is not uniquely determined. Since most modern systems use differential encoding, this phase ambiguity is easily accountable. However, the multi-modality of the CM error surface results from this indiscernible phase–the different minima of the CM error surface correspond to the possible choices of system delay, δ, with $\pm$ polarity.

For example, consider the example of low enough dimension so that the CM surface is easily drawn. The 4-tap channel is described by $\mathbf{c} = [\ 0.2 \ \ 1 \ \ -0.7 \ \ 0.15\ ]^T$. If we assume the odd samples correspond to "on-baud" signalling, the row-decimated channel convolution matrix for a 2-tap FSE becomes

$$\mathbf{C}_{\downarrow} = \begin{bmatrix} c_1 & c_0 \\ c_3 & c_2 \end{bmatrix} \tag{2.47}$$

$$= \begin{bmatrix} 1 & 0.2 \\ 0.15 & -0.7 \end{bmatrix} \tag{2.48}$$

which is full rank since the channel does not contain reflected roots. Hence, in the absence of noise, a P.E. setting is achievable. We examine the CM cost for BPSK signalling (± 1's) as a function of the two possible equalizer taps. The CM error surface in dB is plotted above equalizer parameter space (a plane for two FSE coefficients) in Figure 2.10. The surface is not a simple hyper-ellipsoid, as is the case for the MSE criterion for a specific system delay. Instead, four distinct minima are clearly visible, all of equal depth. (Finite resolution with the dB scale tends to hide this last fact–the CM cost at all minima is in fact zero for binary signalling.) Crude contour lines are projected onto the parameter plane. To further study the surface, Figure 2.11 shows the contour lines in the equalizer parameter plane with

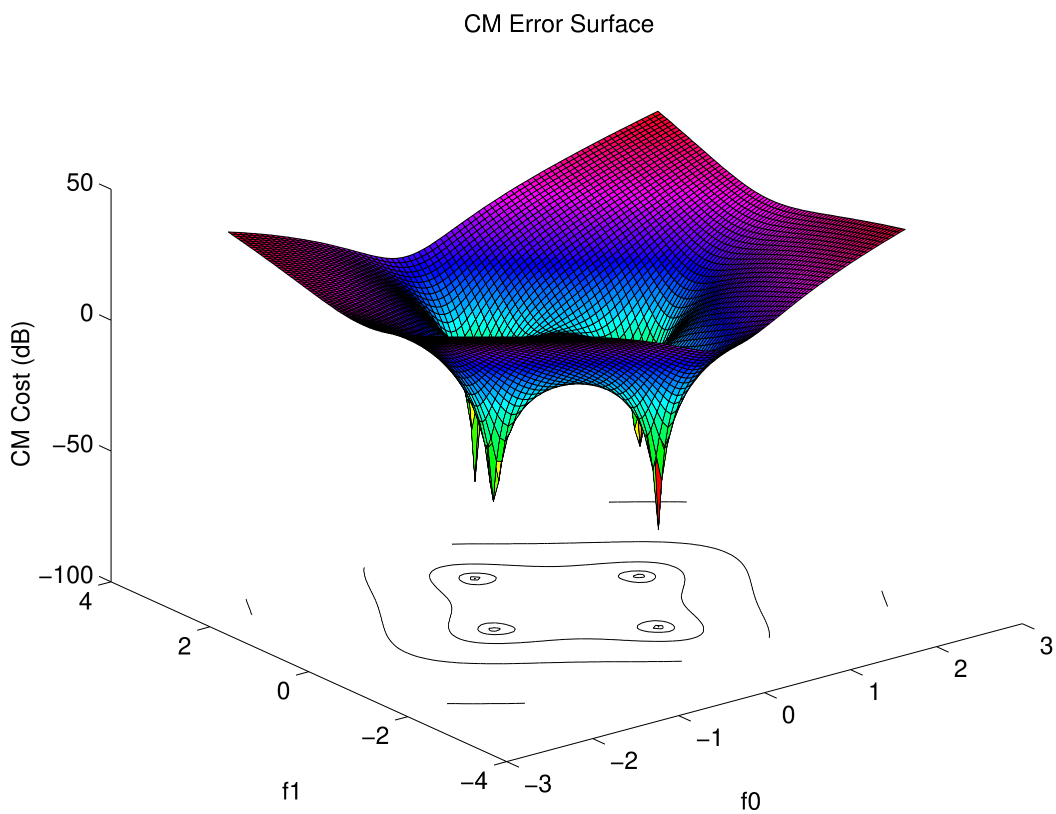


Figure 2.10: CM Error Surface Above Equalizer Plane

a finer resolution than those in Figure 2.10.

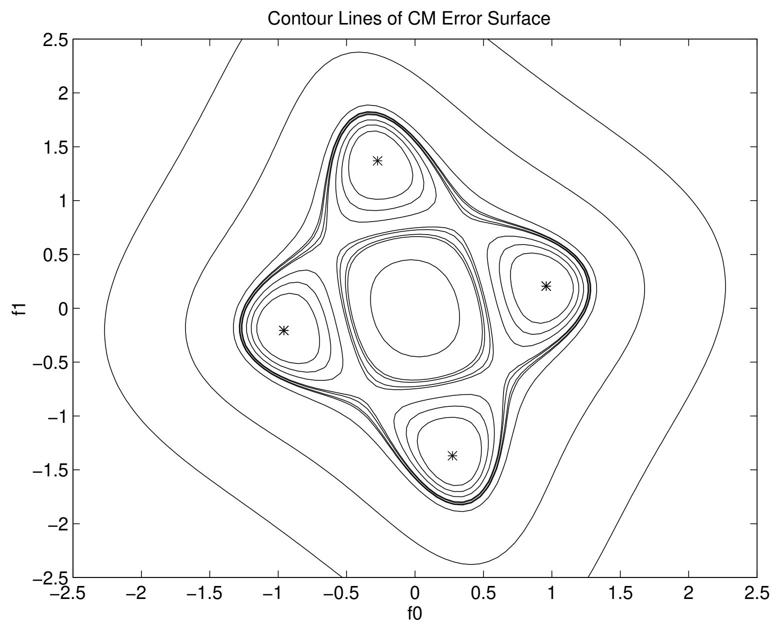


Figure 2.11: Contour Lines of CM Error Surface

The asterisks ($*$) are the locations of the Wiener FSE solutions which minimize the MSE criterion for the four desired channel-equalizer combinations – $[\ 1\ 0\ ]^T$, $[-1\ 0\ ]^T$, $[\ 0\ 1\ ]^T$, and $[\ 0\ \ -1\ ]^T$. The four CM minima are exactly the MSE minima of various delay and polarity; i.e., the CM and MSE minima correspond to the same equalization settings, all of which achieve P.E.

Observe also the existence of a local maximum at the origin. It is shown in [51] and [43] that the origin is the only maximum of the CM surface. Treichler and Agee [88] recognize the local maxima and suggest it can be avoided by proper equalizer initialization, "...the traditional all-zero initial vector should not be used." "Proper" equalizer initialization remains an open research issue, with various ideas in the literature – [30], [88], [81], [74], [31], [48] for example – but is beyond the scope of this dissertation.

Chapters 3, 5, 6 and 7 in the sequel specifically address the robustness properties of the fractionally-spaced CM criterion and its gradient search implementation, CMA-FSE.

Chapter 3

Simulated Comparisons

This chapter provides a simulation-based comparison study focused on the robustness of SOS blind identification/equalization methods versus CMA when applied to various classes of channels under noisy conditions and channel-order undermodeling. We view the blind algorithms tested here as start-up algorithms, whose sole job is to open the channel eye sufficiently (given a certain number of observations) so that a DD mode can provide further error rate reduction and necessary tracking (though we consider strictly time-invariant channel models). The SOS algorithms of [79], [62], [96], and [28] are compared to CMA, which itself accumulates a third-order moment of the received signal. This comparison focuses on each algorithm's behavior across practical classes of channels of interest to both the practitioner and researcher in achieving acceptably rapid and acceptably thorough ISI reduction. Earlier versions of this work can be found in [16] and [17].

We study algorithm robustness by considering the MSE trajectories of the achieved channel-equalizer combination as the additive noise power at the equalizer input is increased and/or the estimated channel order/equalizer order is decreased relative to the true channel order. The primary observations are that the decomposition-based SOS algorithms' performance degradation due to channel-order undermodeling is not graceful: these algorithms fail in all cases when the order of the channel estimate is less than that of the true channel, even by two $T/2$-spaced taps out of 16. These algorithms' degradation to additive noise power, however, is sometimes graceful. Further, the direct algorithms' sensitivity to initialization is evident: CMA fails when randomly initialized (further explanation in §3.1.2), though is overwhelmingly successful with a unity center spike initialization and all other taps zero. Moreover, simulations suggest CMA's sensitivity to source kurtosis can be graceful, though potentially less so than that of SOS schemes.

The study is organized as follows. §3.1 discusses the indirect and direct equalization strategies tested in this robustness study, with fundamental descriptions and associated nuances. §3.2 describes the channel classes considered, and motivation for their inclusion. §3.3 then describes the assumptions and settings made in implementing the equalization algorithms. §3.4 provides enumerated observations based on the entire data set of simulations.

3.1 Equalization Strategies

3.1.1 Indirect Adaptive Equalization

A logical approach in recovering the source sequence is to first solve for a channel estimate, since the equalization then becomes "...more or less a classical problem..." [75]. The computational penalty of the required matrix decomposition (for channel estimation) and matrix pseudo inversion (for equalizer solution) is (hopefully) outweighed by performance benefits.

The estimates from the three following decomposition-based channel estimators are used to solve for the MMSE-FSE according to (2.25) with system delay found according to (2.36). Note that these indirect algorithms do not have a tunable stepsize to adjust convergence. Instead, channel identification arises from a particular solution of a data-based set of equations. Observe (see below) that the algorithms of Tong et al. [79] and Moulines et al. [62] decompose a matrix whose dimensions are independent of the number of observations, M. Hence, these two algorithms easily admit a non-recursive implementation. The full problem is resolved, just for a larger data set, every 600^{th} received signal sample. Quite to the contrary, the deterministic algorithm of Xu et al. [96] decomposes a matrix whose number of rows is proportional to M when the channel estimate is constrained to unit norm. Hence, our implementation of this algorithm works on the equations for each subsequent data block: the problem is resolved every 600^{th} received sample using only the previous 600 observations.

1. **Tong et al.** The algorithm of [79] estimates the order-L (note, $L = L_c - 1$) channel impulse response coefficients by first building a set of vectors from the even and odd subchannel contributions of the received signal autocorrelation sequence. These vectors then form a $(3L + 1) \times (L + 1)$ matrix; the channel estimate is found from the singular-value decomposition (SVD) of this matrix. Hence, this algorithm estimates the channel to within a gain and phase factor, which is easily overcome by automatic gain control and differential encoding. Observe also that this algorithm assumes the channel order, L, is known (or estimated). This study will consider in part cases where this assumption is broken; i.e., the channel order assumed by the algo-

rithm is less than the actual channel order, since this case is most likely true in practice (see [85]). Moreover, this algorithm assumes the additive noise power is known (or estimated) and hence can be negated in construction of the necessary $(3L+1) \times (L+1)$ matrix. Even though noise estimation is itself an open issue, all subsequent simulations for this algorithm and all other algorithms assume the noise power is known exactly at the receiver.

2. **Moulines et al.** The algorithm of [62] estimates the order-L channel impulse response coefficients by exploiting the orthogonality between the signal and noise subspaces of the received signal time-varying correlation matrix. The algorithm first estimates this $(2(L+1) \times 2(L+1))$ matrix. The channel estimate is found as the eigenvector corresponding to the smallest eigenvalue of this matrix, under the nontrivial constraint that this estimate has unit ℓ_2-norm. This algorithm therefore assumes knowledge of the actual channel order, and also estimates the channel to within gain and phase constants. Though the algorithm solves for a channel estimate without explicit knowledge of the noise power, an estimate of the noise power is offered from the eigenvalues found in the decomposition. Our implementation, however, uses the true noise power when solving (2.25).

3. **Xu et al.** The algorithm of [96] estimates the order-L channel impulse response coefficients by treating the channel input as deterministic rather than stochastic. This approach should therefore apply to input sequences with arbitrary statistical characteristics. The algorithm estimates the channel impulse response vector (constrained to unit ℓ_2-norm) as the eigenvector corresponding to the minimum eigenvalue of a matrix which is $(M - (L +$

$1)/2) \times (L+1)$. Observe that this candidate matrix for SVD is of dimension proportional to the observation length, M. Our recursive implementation of this algorithm mentioned above is due to the computational burden associated with the dependence of this matrix (and hence SVD) on M. Note also that this algorithm assumes knowledge of the actual channel order, and also estimates the channel to within gain and phase constants (as does [79] and [62] above). Though the algorithm solves for a channel estimate without explicit knowledge of the noise power, (2.25) is solved with the true noise power.

3.1.2 Direct Adaptive Equalization

Direct adaptive algorithms estimate the equalizer coefficients without explicitly identifying the channel coefficients. Hence, they do not suffer the computational burden associated with channel estimation via decomposition and subsequent matrix (pseudo) inversion for equalizer solution. The two algorithms (Cyclic-LMS of [28] and CMA) of this class considered in this study use a stochastic gradient descent approach in updating the equalizer coefficients. The stepsize is fixed at $\mu = 1 \times 10^{-3}$, since this value proved a nice compromise between smooth equalizer trajectories and stability for CMA and Cyclic-LMS of [28].

Equalizer initialization is performed according to one of two approaches for each of the direct algorithms. Since a unity center spike initialization is commonly cited in the open literature as desirable and is used in practice, we consider it and also a random initialization for CMA. For the random initialization, each complex-valued tap is chosen independently from a uniform distribution $[-1, 1]$ in both the real and imaginary parts. The initialization vector is not normalized to unit ℓ_2-norm. Cyclic-LMS is initialized either with a unity center spike and all other taps zero, or with a unit-valued tap corresponding to the index of the system delay chosen in J_{LMS} (see below) and all other taps zero.

4. **Cyclic-LMS** The Cyclic-LMS algorithm employed in this study was proposed by Giannakis and Halford in [28]. The blind algorithm uses a stochastic gradient descent method in attempting to traverse the MSE cost surface:

$$J_{LMS} \;:=\; \{|y(k) - s(k-\delta)|^2\} \tag{3.1}$$

where $y(k)$ is the output of the equalizer at time k and $s(k-\delta)$ is the transmitted symbol delayed by δ baud intervals. The $(L_f \times 1)$ equalizer tap vector for $\delta = 0$ is updated according to:

$$\mathbf{f}_0(k+1) \;=\; \mathbf{f}_0(k) - \frac{1}{2}\,\mu\,\left[\mathbf{r}^* \cdot \mathbf{r}^T \cdot \mathbf{f}_0(k) \;-\; \mathbf{e}_0\right] \;, \tag{3.2}$$

where $\mathbf{e}_0 := [\sigma_s^2 \cdot c_0 \; 0 \; \ldots \; 0]^T$.

And for $\delta > 0$

$$\mathbf{f}_\delta(k+1) \;=\; \mathbf{f}_\delta(k) - \frac{1}{2}\,\mu\,\left[\mathbf{r}^* \cdot \mathbf{r}^T \cdot \mathbf{f}_\delta(k) \;-\; \mathbf{e}_\delta\right] \;, \tag{3.3}$$

where $\mathbf{e}_\delta := \mathbf{r}^*(k+1) \cdot \mathbf{r}^T(k+1-\delta) \cdot \mathbf{f}_0(k)$.

In practice, the channel coefficient c_0 and input variance σ_s^2 are not known *a priori* and $\sigma_s^2 \cdot c_0$ is set to 1. In this case, (3.2) minimizes (3.1) to within an unknown magnitude and phase shift, as does [79], [62], and [96]. Also note that the equalizer found for $\delta > 0$ depends on the equalizer estimate for $\delta = 0$. To the best of our knowledge, cyclic-LMS is the only SOS blind algorithm of this simple computational level (which is comparable to CMA). See [34] for fuller discussions on cyclic-LMS and related algorithms.

5. **CMA** CMA employed in this study was first proposed by Godard in [30] and developed independently by Treichler and Agee in [88]. See §2.3.2 of Chapter 2 for a description of CMA and the CM criterion. The equalizer coefficients are adjusted according to (2.46).

3.2 Channel Classes

Specific FIR $T/2$-spaced channel models are chosen to be representative of the following classes of channels. These classes are chosen because they are acknowledged as stressful to one or more of our five algorithms or are thought benign to all or are data-based (and therefore undeniably realistic). Note that for Figures 3.1-3.6, when overlaying the roots of two subchannels, $\times$'s ($\circ$'s) correspond to even (odd) subchannel zeros.

1. Channel Class NRR (Nearly Reflected Roots):

   Most indirect algorithms suffer a penalty in MSE performance due to inversion of a nearly singular matrix when the $T/2$-spaced channel impulse response loses disparity, i.e., has symmetric $T/2$-spaced roots with respect to reflection through the origin [89], [22]. Such channels are therefore typically not admitted in analysis and implementation of these SOS algorithms. [21] showed that CMA effectively solves for the approximate baud-spaced inverse of this common subchannel root, leaving the remaining equalizer taps to solve the remaining Diophantine equation of FSE design (see §5.3).

This channel class therefore contains FIR channel models which have nearly common subchannel roots, but no $T/2$-spaced roots near the unit circle. For example, Figure 3.1 shows Channel NRR-A of Channel Class NRR.

2. Channel Class NUC (Near Unit Circle):

   Since channels with roots near the unit circle contain nulls in their frequency response, a baud-spaced inverse thus suffers noise enhancement in certain frequency ranges. $T/2$-spaced examples are readily found where the FSE solution (either MSE or CM minima) does not resemble the (pseudo) inverse of the $T/2$-spaced channel impulse response vector. In fact, the open literature contains little work addressing the significance (if any) of such channel dynamics for the fractionally sampled case. This channel class contains models with a ring of zeros near the unit circle, but none of the zeros form reflected pairs (see Figure 3.2).

3. Channel Class RC (Raised Cosine):

   The raised cosine pulse shape is commonly used in digital communications since it satisfies the Nyquist criterion (eliminates ISI at T-spaced intervals) and displays relatively fast asymptotic convergence in its time support (see [29]). Moreover, some systems may rely on equalization to remove only the uncertain pulse-shaping dynamics, so channel models which consist of only the pulse-shaping response are of practical interest. Hence, this channel class consists purely of raised cosine and square-root raised cosine pulse shapes in both I and Q channel paths, with 30-50% excess bandwidth. For example, see Figure 3.3.

4. Channel Class B (Benign):

   The channel classes NRR and NUC expose known problematic root patterns for currently popular blind, adaptive equalization algorithms. Channel class B is intended to avoid these known limitations, and therefore does not contain roots near the unit circle nor nearly common subchannel roots nor the symmetric impulse response of the RC class. For example, see Figure 3.4.

5. Channel Class AppSigTec:

   The channel models in [32] are derived from empirically measured $T/2$-spaced digital microwave radio signals. The complex valued impulse responses consist of greater than 200 non-zero taps. The source sequence is 16-QAM for the high (multi-megabaud) rate employed. These channel models are now available over the World Wide Web at the signal processing database maintained at Rice University

   http://spib.rice.edu/spib/microwave.html

   Signal Processing Database

   and henceforth referred to as the "database" in the sequel.

   The channel models in this robustness study are length-16 versions of the channel models available from the database. These shortened versions are derived by linear decimation of the FFT of the "full-length" $T/2$-spaced impulse responses. The resulting length-16 frequency-domain models are inverse transformed to obtain length-16 impulse responses. (For example, see Figure 3.5). This procedure reduces aliasing common in time-domain decimation and retains the original models' frequency characteristics reasonably

Table 3.1: Map Between Length-16 Channels Used in Robustness Study and Database Channels

AppSigTec Channel	Database Designator
A	5
B	3
C	4

well. Interestingly, the shortened impulse responses retain the characteristic of near common subchannel roots, which is evident in the original $T/2$-spaced models (see [42] for "full-length" $T/2$-spaced dynamics), as well as roots near the unit circle. Hence, our NUC and NRR channel classifications are of apparent practical significance. Our length-16 channel models AppSigTec-A, -B and -C originated from the "full length" $T/2$-spaced channels in the database according to Table 3.1.

6. Channel Class R (Random):

   Each channel in this class is a length-16 approximation of a two-ray multipath channel, similar to experiments conducted in [102],

$$c(t) \quad = \quad \sum_{i=0}^{1} a_i \cdot p(t - \tau_i) \tag{3.4}$$

   The two paths are independently fading. That is, there are two raised cosine pulse shapes, $p(t)$, with roll-off factor 0.3 each at different delays, τ_i, and each with a different complex gain, a_i. The delays are chosen i.i.d. and uniform over $[0\ 2T]$ and the gains are complex, zero-mean gaussian with unit variance. A typical channel model for this class is shown in Figure 3.6.

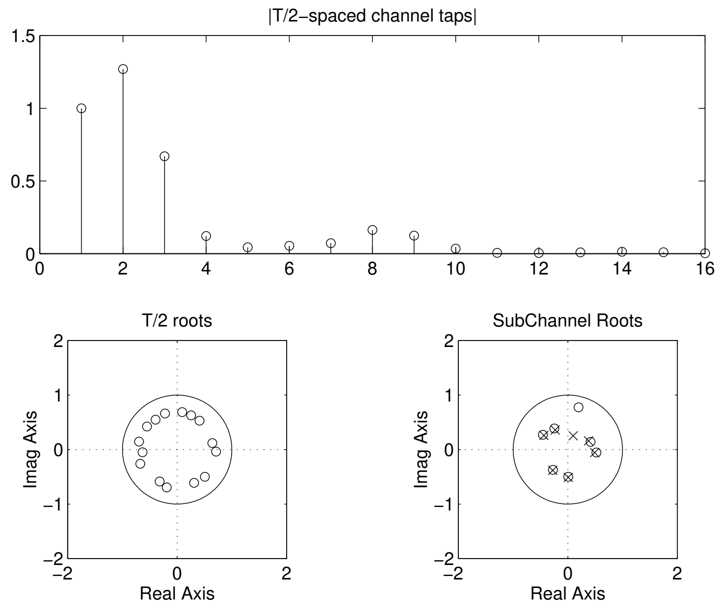


Figure 3.1: Channel NRR-A Dynamics

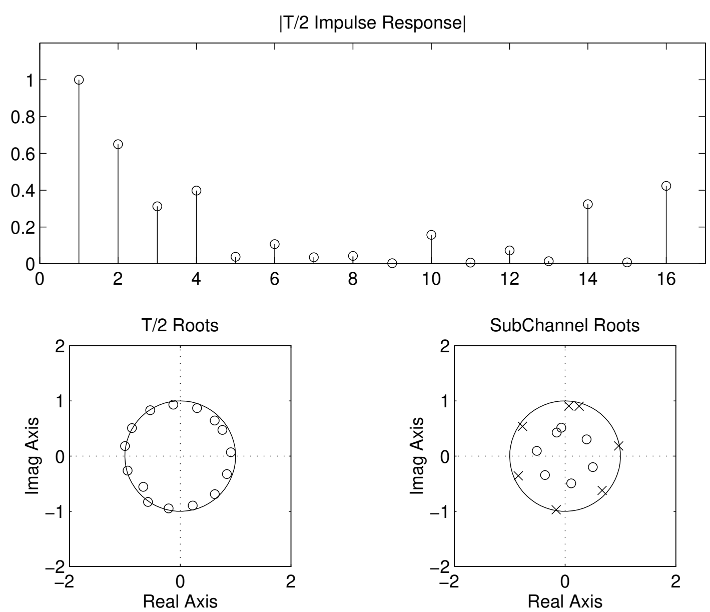


Figure 3.2: Channel NUC-A Dynamics

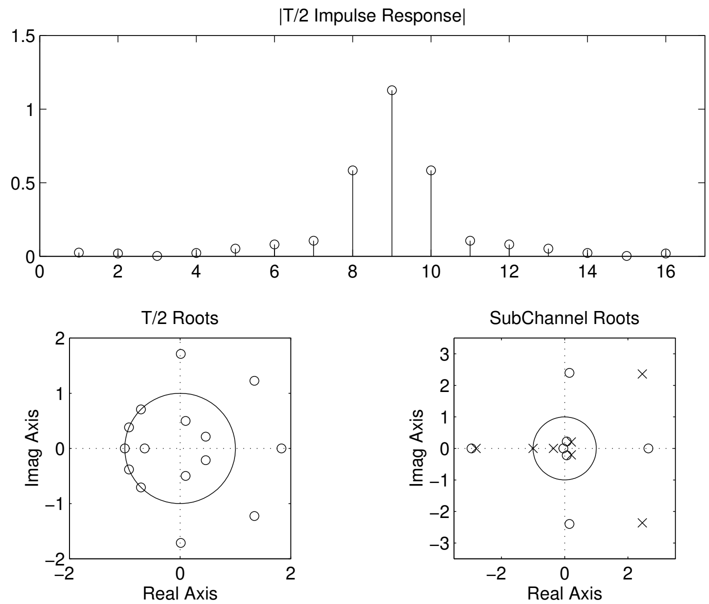


Figure 3.3: Channel RC-A Dynamics

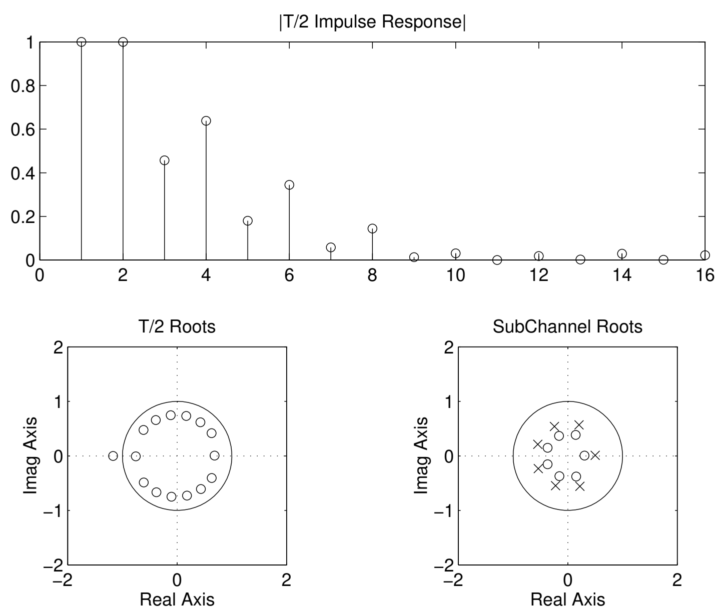


Figure 3.4: Channel B-A Dynamics

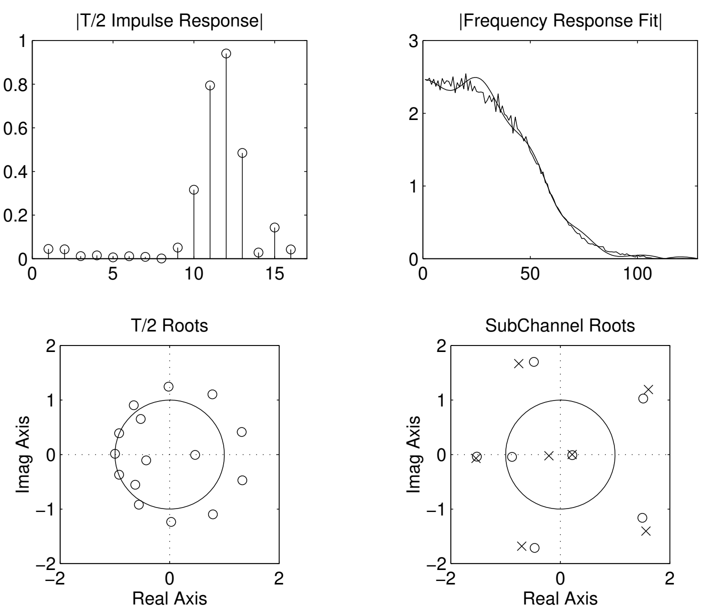


Figure 3.5: Channel AppSigTec-C Dynamics

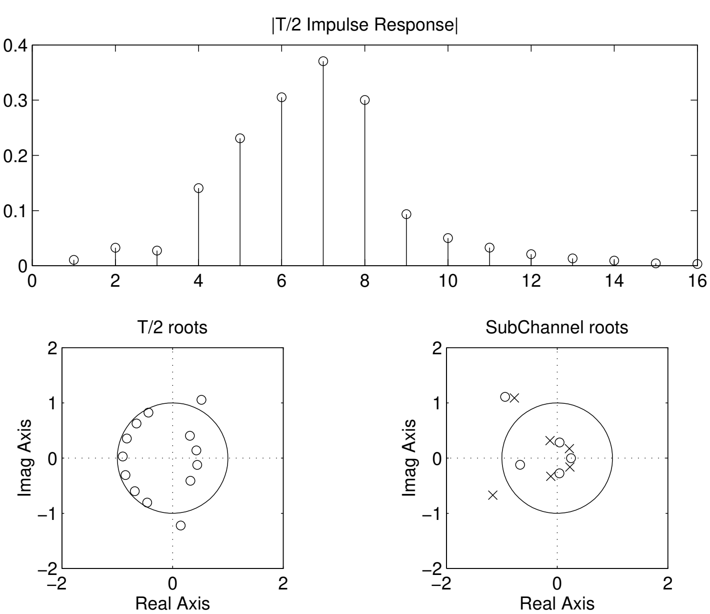


Figure 3.6: Typical Random Channel Dynamics

3.3 Assumptions/Settings

This section describes assumptions and settings employed in implementing the blind equalization/ identification algorithms discussed in §3.1.

- All fractional spacing is at a rate of $T/2$ since this case is of current practical interest.

- Each source symbol is drawn from a white source distribution which assumes values from a 16-QAM unit-variance alphabet before upsampling. This modulation is representative of higher-order constellations used in practice. Since constellation shaping is proposed to maximize channel throughput in some digital communications applications (see [38] or [50]) but at the same time is shown to flatten the CM error surface (see [52] or [51] for a fuller description,) non-equi-probable as well as equi-probable (but temporally uncorrelated) source sequences are considered. This shaping is such that symbols closer to the origin are chosen more frequently.

- For comparisons across channels and SNRs, the received signal variance $\sigma_r^2 = \mathbf{c}'\mathbf{c}\frac{\sigma_s^2}{2} + \sigma_n^2$ is set to unity by a normalized channel response $\mathbf{c}$ such that for $\sigma_s^2 = 1$

$$\mathbf{c}'\mathbf{c} \quad = \quad 2(1 - \sigma_n^2) \tag{3.5}$$

  with $\sigma_n^2 < 1$. The ambiguous gain of the channel estimates of [79], [62] and [96] is thus set according to (3.5).

- The view of the blind algorithm as a start-up algorithm implies the "convergent" eye is "sufficiently open" for transfer to a DD mode. Practitioners define this condition in terms of a symbol error rate (SER) in the neighborhood of $10^{-1.5}$ or between 10^{-1} and 10^{-2} [86]. We take this condition to be a SER of approximately 0.04. This SER with 16-QAM and channel normalization according to (3.5) corresponds to a MSE of approximately 0.076 (see [63] for underlying mathematics and Appendix A for Transfer Levels of other modulation formats). Hence, any MSE less than 0.076 achieved by the blind algorithm is considered successful for transfer to DD mode. The algorithm is said to fail if the achieved MSE is greater than the Transfer Level MSE of 0.076.

- All channel lengths are $L_c = 16$. SNRs (at the equalizer input) considered are 100, 35, 25 and 15 dB. Channel estimate lengths and equalizer lengths considered are 16, 14, 12 and 10. MSE performance is averaged over at least 10 independent observations, each of length 30,000 $T/2$-spaced samples.

3.4 Observations

This section describes observations based on the 100+ MSE trajectories drawn from the simulations. The intent is to provide relevant and useful observations based on the numerous simulations, without an exhaustive show of figures and plots. More simulation results can currently be found over the World Wide Web at

www.backhoe.ee.cornell.edu/BERG.

These results are similar to Figures 3.7-3.10, which show MSE performance for channels AppSigTec-A and AppSigTec-C, SNR of 100 dB, when the channel length is correctly modeled and undermodeled by two $T/2$-spaced taps.

1. All SOS decomposition-based algorithms fail for all channel classes at all SNRs considered when the estimated channel order is less than the actual channel order. Hence, the degradation in performance of these algorithms with channel-length undermodeling is not graceful. See Figures 3.7-3.10. Figures 3.7 and 3.9 also show that the decomposition-based algorithms are not always successful under "ideal" conditions, i.e., correct model order and 100db SNR.

2. The decomposition-based algorithms' sensitivity to noise is, however, sometimes graceful. For example, compare Figures 3.11 and 3.12 which show performance for AppSigTec-A when the channel length is correctly modeled, with additive noise powers 35 and 15 dB, respectively.

3. For channel class AppSigTec, CMA when initialized with a unity center spike and all other taps zero with a white, equi-probable source is successful for all FSE lengths and SNRs considered.

4. For a non-equi-probable source and center spike initialization, however, CMA's performance degrades due to a flattening of its error surface (see [52], [51], or §5.1). This degradation, though, appears graceful; acceptable performance is observed up to a source kurtosis[1] in the range of 2.2-2.5 (see Figure 3.13, showing achieved MSE versus source kurtosis for AppSigTec channels).

5. The SOS algorithms appear to be unaffected by constellation shaping.

6. Both gradient algorithms are sensitive to initialization. Though CMA is overwhelmingly successful for AppSigTec with a center spike initialization, it

[1]Kurtosis of s is calculated as $\kappa_s = \frac{E\{|s|^4\}}{E\{|s|^2\}^2}$

fails when randomly initialized (see §3.1.2 and Figures 3.7-3.12).

7. For RC channels, no algorithm is successful for any case considered. The gradient-based algorithms do achieve $\sigma_e^2 < 1$ when initialized with a unity center spike, however. We attribute this behavior to the (nearly) common subchannel roots indicative of models from this channel class.

8. [28] is the only algorithm which is successful (center spike initialization and zero system delay in J_{LMS}) for NRR channel with correct channel length at 35 dB SNR. [62] is successful at 100 dB SNR.

9. For the NUC channel, all algorithms are successful when the channel order is known and SNR is either 35 or 100 dB, though the recursive implementation of [96] causes fluctuations in the achieved MSE about the Transfer Level MSE. Further, [96] is more sensitive to noise than the other algorithms. The other four algorithms are successful when the SNR drops, and the gradient-based algorithms continue to succeed when the length is undermodeled. This behavior suggests that (non-repeated) $T/2$ roots near the unit circle do not present a problem to these blind algorithms.

10. [79] fails for a Benign channel under “ideal” conditions (channel order known and 100 dB SNR). Both [62] and [96] are sensitive to noise. The gradient-based algorithms continue to succeed when the length is undermodeled and are less sensitive to noise than [62] and [96].

11. Experiments using the two-ray channels from the channel class R were conducted slightly differently than the aforementioned experiments. For this experiment, 500 length-16 channels of the Random channel class were first

generated. Next, a single input sequence consisting of 20,000 $T/2$-spaced samples was used to generate a received sequence from each of the 500 channels for the blind algorithms to process. The gradient descent algorithms are both initialized with a unity center spike, and all other taps are zero. In all cases for these two-ray experiments, the true channel order is used in implementing the blind algorithms.

Figures 3.14-3.16 show the percentage of the 500 channels which achieve a given MSE level for additive noise powers 100 dB, 35 dB, and 25 dB, respectively. Observe that for the 100 dB case (Figure 3.14,) similar performance is seen between CMA, [96] and [62]. However, [96] and [62] performance severely degrades with additive noise power, while CMA's degradation is quite graceful (Figures 3.15 and 3.16). The algorithms of [79] and [28], though, struggle for all experiments using this channel class.

12. The algorithm of [28] is particularly sensitive to initialization. Compare Figures 3.14 and 3.17. The difference is the initialization of the cyclic-LMS algorithm of [28]; Figure 3.14 corresponds to a unity-center-spike initialization, while Figure 3.17 corresponds to an initialization with the first tap set to $\frac{1}{c_0}$. The performance of [28] improves significantly with the latter initialization, leaving only the algorithm of [79] with much different performance than the other four algorithms.

Despite the greater computational burden of decomposition and (pseudo) inversion associated with the indirect equalization methods considered, the SOS schemes do not seem to be gaining performance relative to CMA, in terms of more efficient information extraction from the data or in achievable MSE within a reasonable time window. CMA's superiority is striking, but not uncommon from other earlier studies, for example [40] and [68]. Incorporating other algorithms (for example, the Cyclic-RLS proposed in [28] which is of medium complexity relative to the algorithms in this robustness study), clever initialization schemes, and channel types may alter this conclusion. At the very least, these tests argue for a deeper study of CMA's robustness properties and SOS shortcomings.

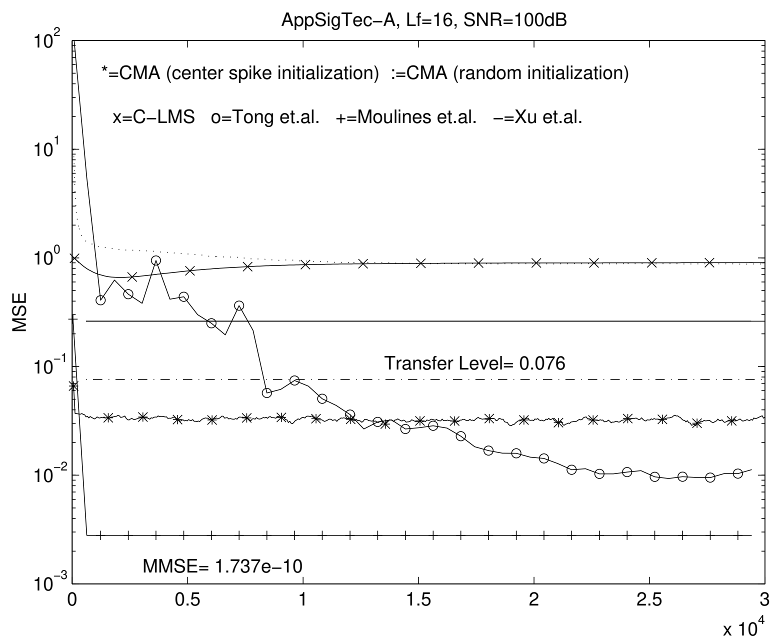


Figure 3.7: MSE Perf.: AppSigTec-A, SNR=100 dB, $L_f = 16$, equi-prob. source

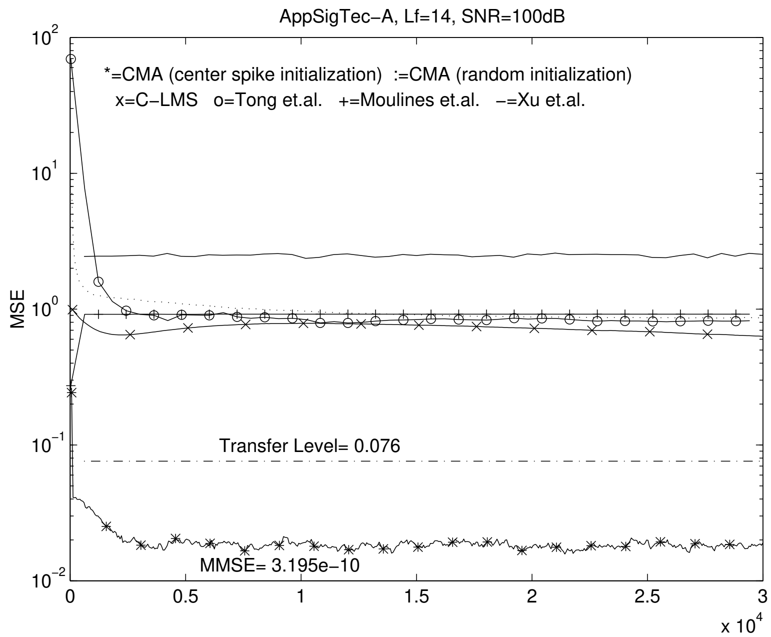


Figure 3.8: MSE Perf.: AppSigTec-A, SNR=100 dB, $L_f = 14$, equi-prob. source

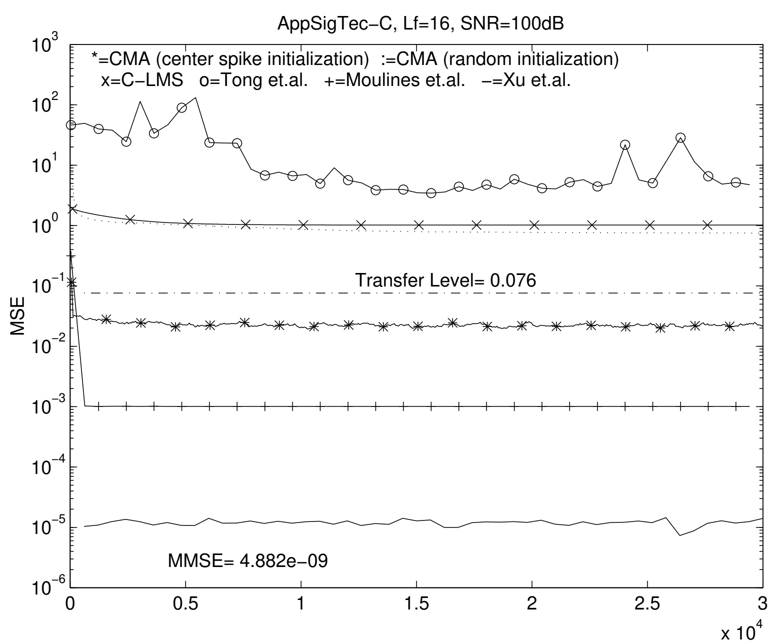


Figure 3.9: MSE Perf.: AppSigTec-C, SNR=100 dB, $L_f = 16$, equi-prob. source

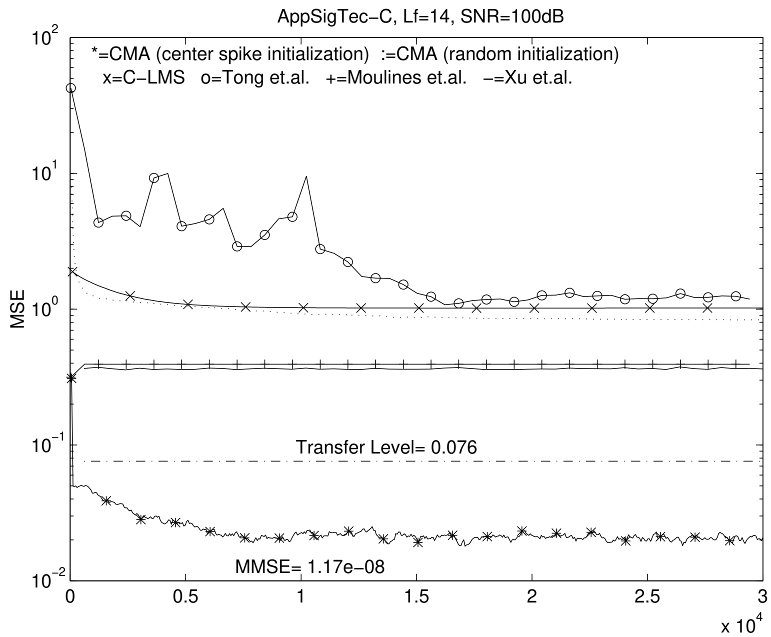


Figure 3.10: MSE Perf.: AppSigTec-C, SNR=100 dB, $L_f = 14$, equi-prob. source

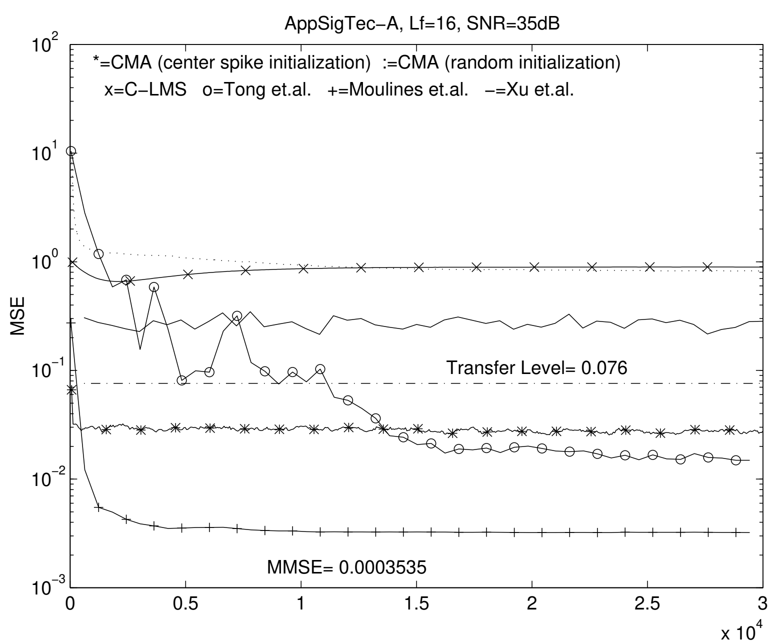


Figure 3.11: MSE Perf.: AppSigTec-A, SNR=35 dB, $L_f = 16$, equi-prob. source

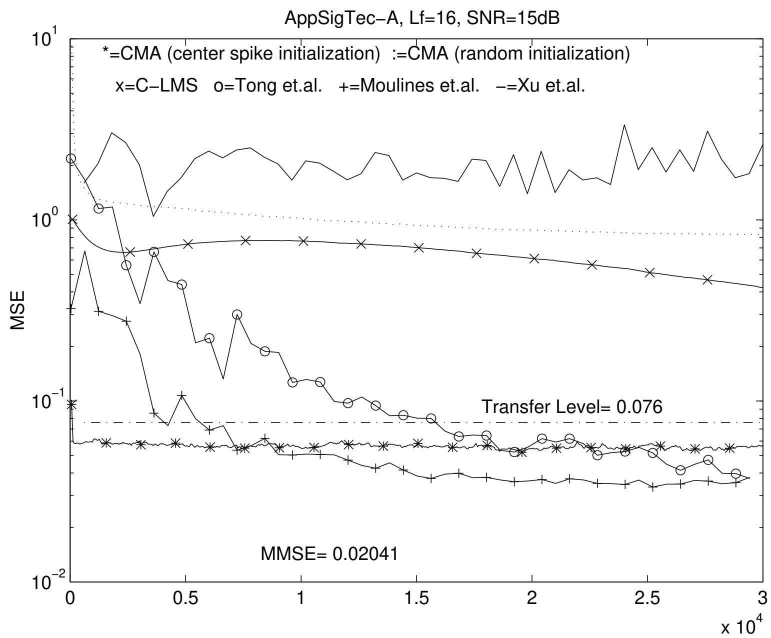


Figure 3.12: MSE Perf.: AppSigTec-A, SNR=15 dB, $L_f = 16$, equi-prob. source

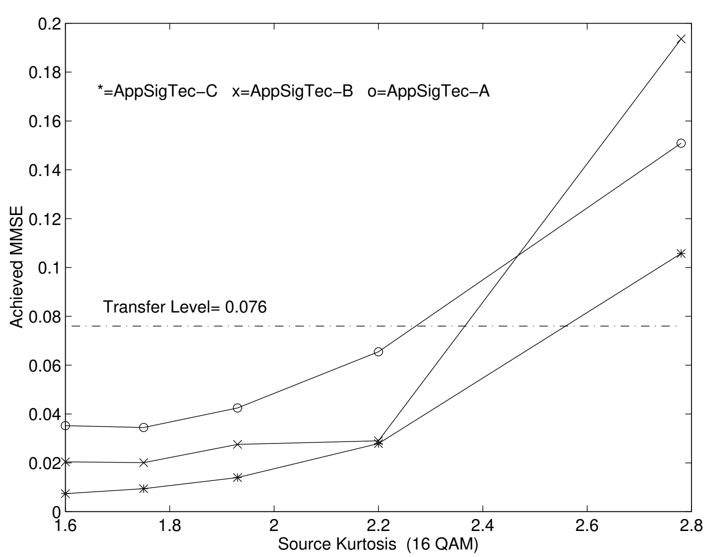


Figure 3.13: CMA Performance as a Function of κ_s: AppSigTec-channels, SNR=100 dB, $L_f = 16$

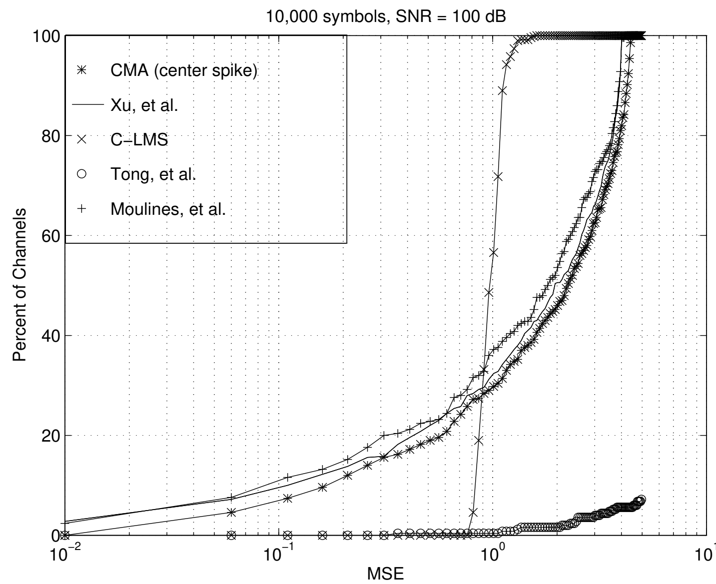


Figure 3.14: Random Channel Results, 100 dB SNR

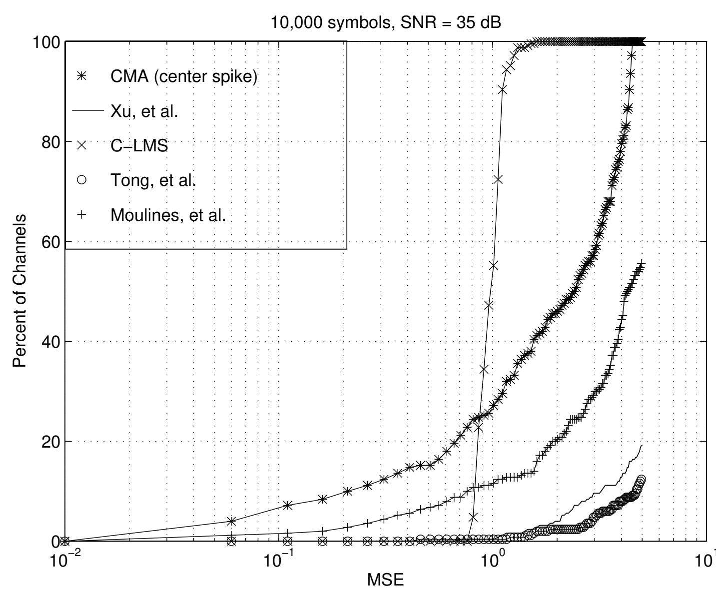


Figure 3.15: Random Channel Results, 35 dB SNR

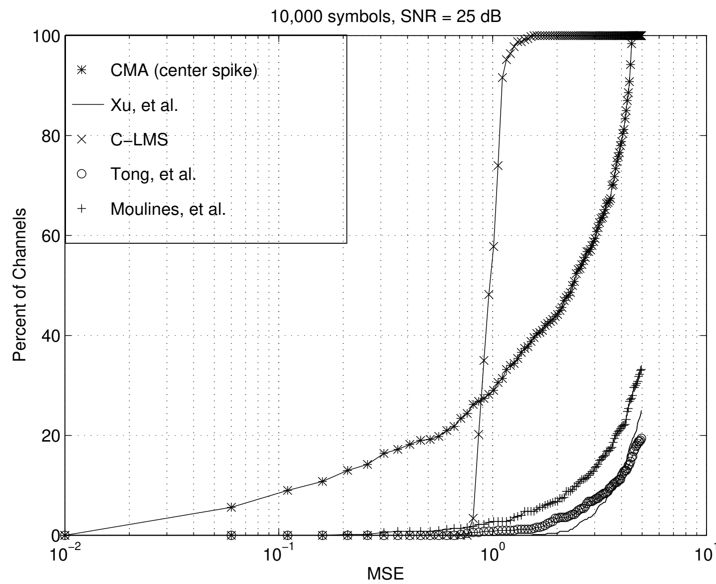


Figure 3.16: Random Channel Results, 25 dB SNR

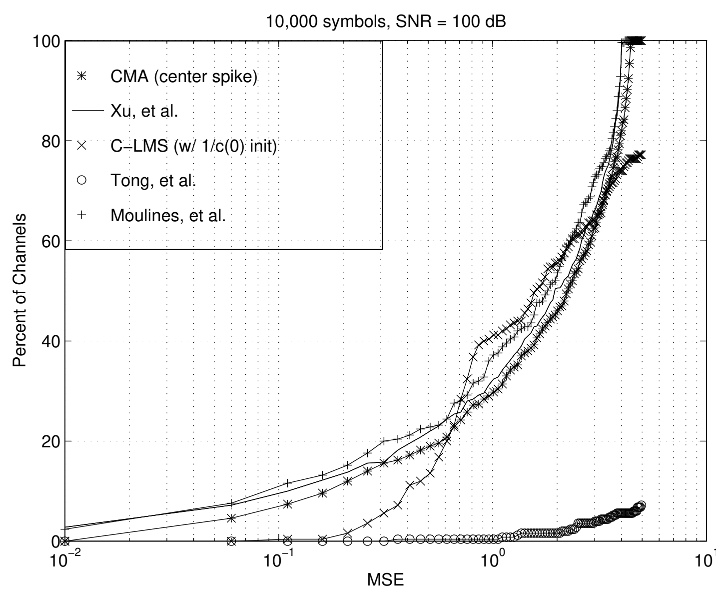


Figure 3.17: Cyclic-LMS Sensitivity to Initialization

Chapter 4

SOS Shortcomings

This chapter presents a modest amount of analysis of one of the decomposition-based SOS algorithms studied in the previous chapter. First, a class of FIR channels which are nearly un-identifiable using SOS, but still satisfying channel identifiability conditions, is presented by construction. This work is then connected with Fijalkow's work [22] on channel disparity. The fundamental premise of channel identification on which the algorithm of Tong et al. [79] is based is then studied for the undermodeled case.

4.1 A Class of Nearly Un-Identifiable Channels

A class of cyclic all-pass channels first associated with the equalization problem in [80] is shown to be the *only* class of IIR channels whose cascade with a core channel (IIR or FIR) results in no change in the SOS indicators used by several popular SOS algorithms. FIR approximations to the poles of the resulting combination (core channel and cyclic all-pass) result in robustness concerns for the SOS algorithms even when the FIR identifiability conditions of [78] are well satisfied. An earlier version of this work can be found in [15]. These results are connected to the recent work of Fijalkow et al. [22].

4.1.1 SOS Indicators

The algorithm in [79], and numerous others ([62], [75], [78], [70], for example) rely for the identification of a $T/2$-spaced channel transfer function $C(z)$ on the uniqueness of the $\Gamma^0(z)$, $\Gamma^1(z)$ pair, where the $\Gamma^i(z)$ are derived from the received auto-correlation sequence, yielding the identification equations

$$\Gamma^0(z) = C(z)C(z^{-1}) \text{ and } \Gamma^1(z) = C(z)C(-z^{-1}) \tag{4.1}$$

Note that the form of the SOS indicators in (4.1) requires for clarity that we keep the z dependence, where previous (and future) chapters drop this notation for convenience. Excluding $T/2$-spaced channels with zeros reflected through the origin, a FIR channel may be identified from the common zeros of the $\Gamma^0(z)$, $\Gamma^1(z)$ pair, and the channel is said to satisfy the identifiability conditions discussed in [78]. The issue addressed here is whether or not the exclusion of just FIR channels with such undesirable reflected zero patterns (or their close approximation) guarantees a robust estimation task in ascertaining the roots of $C(z)$ as the common roots of $\Gamma^0(z)$ and $\Gamma^1(z)$, even if the true channel order is known. We will show by simple construction that even when such FIR identifiability conditions are well satisfied, serious robustness problems may arise due to close approximation of non-identifiable IIR models by an identifiable FIR model with no nearly reflected zeros. A general method to construct channels with such sensitivity concerns is presented by example.

4.1.2 Example

Consider an example with two length-22, unit ℓ_2-norm, $T/2$-spaced channels, $A(z)$ and $B(z)$, both of which are such that their T-spaced subchannels contain no roots which are within a distance of 0.1 to any other root in that subchannel. The right-most plots of Figure 4.3 show the zero locations of the $T/2$-spaced channels. This non-zero proximity implies the channels' true roots are correctly extracted from the intersection of zeros in application of (4.1), since the identifiability conditions of [78] are well satisfied. With $\approx$ denoting approximately equals, the fact that $\Gamma^0(z)$ and $\Gamma^1(z)$ have to be estimated from data and, more generally, good engineering practice should demand the following property of the blind identification algorithm:

Conjecture 1

$$\Gamma^0_A(z) \approx \Gamma^0_B(z) \quad \textit{and} \quad \Gamma^1_A(z) \approx \Gamma^1_B(z) \implies A(z) \approx B(z)$$

where the A, B subscript acts as a channel designator. Specifically, forming $\Gamma^0(z)$ and $\Gamma^1(z)$ as described in (4.1) for channels $A(z)$ and $B(z)$ yields the unit ℓ_2-norm polynomials whose coefficients are such that

$$||\Gamma^0_A(z) - \Gamma^0_B(z)||_2^2 \;<\; 8 \times 10^{-3} \tag{4.2}$$

$$||\Gamma^1_A(z) - \Gamma^1_B(z)||_2^2 \;<\; 1 \times 10^{-12} \tag{4.3}$$

Conjecture 1 therefore implies that these two channels are similar, when in fact the channels are grossly different as suggested by their impulse responses in Figure 4.1. To study the effect of this sensitivity on the equalization task, length-22 FSE's

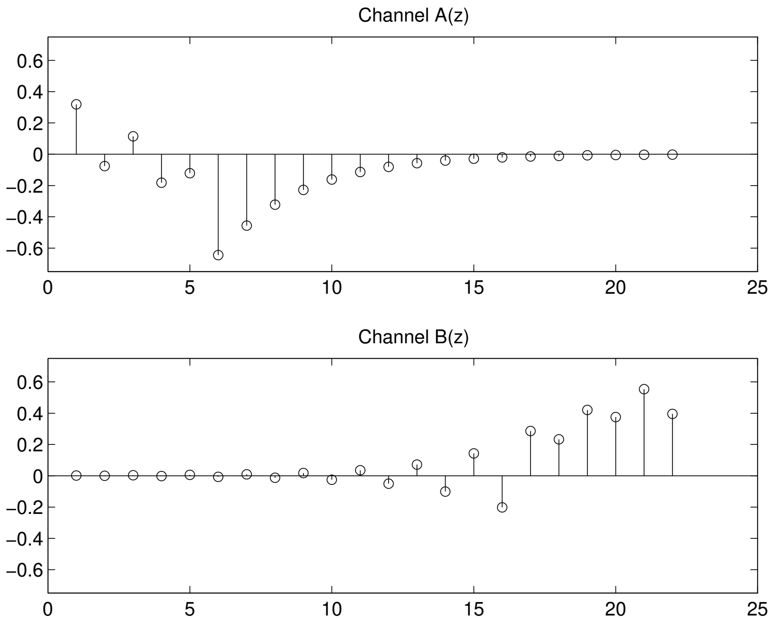


Figure 4.1: Counter Example Channels to Conjecture 1

are found according to (2.25) for 100 dB SNR. The system delay of the desired response is arbitrarily set to the center position. Though the SOS indicators are nearly identical, when the equalizer is applied to the incorrect channel, the channel-equalizer combination is closed eye (see Table 4.1[1]). Thus, Conjecture 1 appears false even though FIR identifiability conditions are well satisfied.

4.1.3 IIR Insertion

Under the identifiability premise of [78], there exists no reflected roots in the $T/2$-spaced FIR channel models for which the SOS algorithms are applicable. It is easily shown that this condition implies that there is no FIR cascade with a core channel which leaves the Γ's (and subsequent equalizers) unaltered. Consider, however, the

[1] ISI is computed as $ISI = \dfrac{\sum_i |h_i| - \max_i |h_i|}{\max_i |h_i|}$

Table 4.1: Performance Sensitivity Resulting From SOS Channel Estimation

Channel	FSE For Which Channel	ISI	Comments
$A(z)$	$A(z)$	1.96×10^{-10}	Perfectly Equalizable
$A(z)$	$B(z)$	2.32	Closed Eye
$B(z)$	$B(z)$	3.37×10^{-10}	Perfectly Equalizable
$B(z)$	$A(z)$	1.66	Closed Eye

cascade of an IIR structure with a core channel and its effects on the Γ functions.

Theorem 1 *Let $K(z)$ be an IIR transfer function and $C(z)$ be the transfer function for a core channel. Then $C(z)$ and $C(z)K(z)$ have the same $\Gamma^0(z), \Gamma^1(z)$ pair* if and only if *$K(z)$ is all-pass with $K(z) = K(-z)$.*

Proof: Suppose the two channels have the same $\Gamma^0(z), \Gamma^1(z)$ pair. Using (4.1) it follows that

$$K(z)K(z^{-1}) = 1 \text{ and } K(z)K(-z^{-1}) = 1 \tag{4.4}$$

which implies $K(z) = K(-z)$.

The other direction[2] of the proof follows directly from construction. □□□

The class of $K(z)$ thus consists of all-pass channels with pole patterns which are symmetric with respect to reflection through the origin, referred to as cyclic all-pass structures in [80] since they are obtained by all-pass rotations. For example, Figure 4.2 shows an example of a cyclic all-pass channel root pattern. Though the

[2]Note that the "if" direction of this theorem is detailed in [80] for arbitrary integer oversampling rates. That is, the cyclic all-pass class is shown to yield identical Γ sets. The above theorem shows the cyclic all-pass class to be the only such class of IIR cascades yielding equivalent Γ sets.

$\Gamma^0(z)$ and $\Gamma^1(z)$, and equalizer computed from them, will be unaltered with the cascade insertion of this cyclic all-pass structure, these all-pass insertions ruin P.E. and can even close the eye of the total channel-equalizer combination.

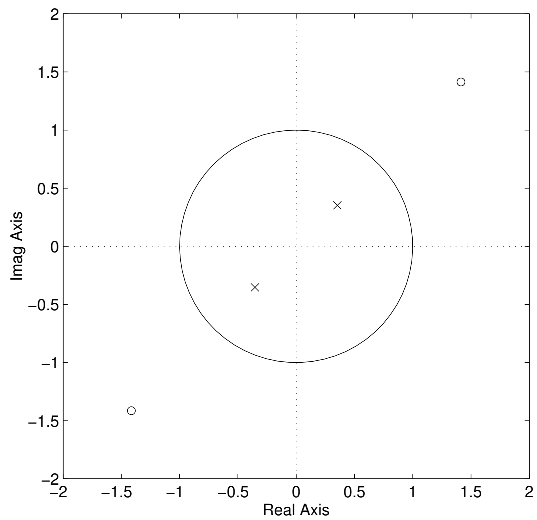


Figure 4.2: Cyclic All-Pass Root Pattern Example

4.1.4 Example Revisited

The example in Section 4.1.2 and its construction can be understood in light of the cyclic all-pass effect on the SOS indicators. Consider first the all-pass belonging to the class described in Theorem 1 with zeros at $\pm 1/\sqrt{2}$ and poles reflected about the origin at $\pm\sqrt{2}$, i.e., the cyclic all-pass transfer function is $K(z) = \frac{(z-\frac{1}{\sqrt{2}})(z+\frac{1}{\sqrt{2}})}{(z-\sqrt{2})(z+\sqrt{2})}$. A core channel is described by

$$C(z) \quad = \quad (1 + 0.47z^{-1} + 1.19z^{-2} + 0.86z^{-3} + 1.24z^{-4}) \times \frac{(z-\sqrt{2})}{(z-\frac{1}{\sqrt{2}})} \tag{4.5}$$

and the cascade of the all-pass and $C(z)$ is

$$D(z) \quad = \quad C(z)K(z) \tag{4.6}$$

$$\begin{aligned} &= (1 + 0.47z^{-1} + 1.19z^{-2} + 0.86z^{-3} + 1.24z^{-4}) \\ &\quad \times \frac{(z - \sqrt{2})}{(z - \frac{1}{\sqrt{2}})} \frac{(z + \frac{1}{\sqrt{2}})}{(z + \sqrt{2})} \frac{(z - \frac{1}{\sqrt{2}})}{(z - \sqrt{2})} \qquad (4.7) \\ &= (1 + 0.47z^{-1} + 1.19z^{-2} + 0.86z^{-3} + 1.24z^{-4}) \times \frac{(z + \frac{1}{\sqrt{2}})}{(z + \sqrt{2})} \qquad (4.8) \end{aligned}$$

See Figure 4.3 for pole-zero plots of $C(z)$ and $D(z)$. Note that here, unlike Chapter 3, $\times$'s designate poles and $\circ$'s designate zeros.

The FIR channels used in §4.1.2 are related to these IIR channels. Channel $A(z)$ is formed from an FIR approximation of $C(z)$. A length-17 FIR approximation to the pole at $z = \frac{1}{\sqrt{2}}$ results in a ring of 16 zeros evenly spaced around a circle of radius $\frac{1}{\sqrt{2}}$, with an absent zero in place of the pole. This approximation minimizes the ℓ_2-norm of the difference of impulse responses between the pole and its FIR approximation. Channel $B(z)$ is formed in a similar fashion from an FIR approximation of $D(z)$, the cascade of the all-pass and core channel. The positive pole and zero of the all-pass are canceled by the core channel, and a length-17 FIR approximation to the remaining all-pass pole results in a ring of 16 zeros evenly spaced around a circle of radius $\sqrt{2}$ with an absent zero at $z = -\sqrt{2}$. The root locations for the rational and resulting FIR approximations are shown in Figure 4.3.

Despite the ring of zeros in $A(z)$ and $B(z)$, none are reflections of any other across the origin. (Recall that no two T-spaced subchannels' zeros are closer than 0.1 to any other subchannel zero.) Thus, the identifiability condition of [78] for this $T/2$-spaced example is well satisfied. Note that both FIR channels, $A(z)$ and $B(z)$, have 21 zeros. These two FIR channels are precisely the ones used in the example of §4.1.2.

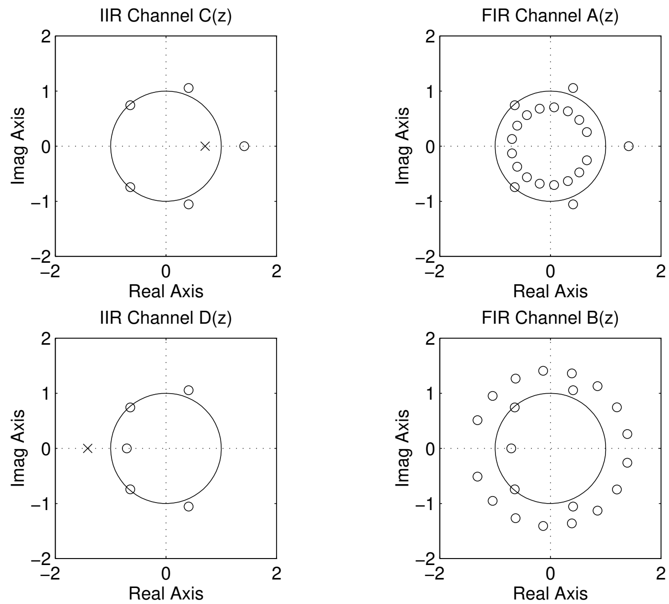


Figure 4.3: Rational and FIR Channel Model Roots

This example thus demonstrates how identifiable FIR approximations to IIR channel pairs described in Theorem 1 can yield a counter-example to Conjecture 1. In fact, arbitrarily close $\Gamma^0(z)$ and $\Gamma^1(z)$ approximation is possible with a longer FIR approximation, and only in the limit of an infinite length approximation does the example collapse into the special case of reflected roots (given odd-length approximations). The result of the all-pass insertion is to yield virtually equal Γ sets for the two channels that can produce grossly different channel estimates. This example motivates the search for other root patterns which can result in excessive sensitivity.

4.1.5 Connection To Related Work

Recently, Fijalkow [22] has proposed a measure of channel disparity based on the SNR and subchannel root locations. Her motivation is to address the proximity of nearly common subchannel roots, and show that there exists a neighborhood where the nearly common subchannel roots are close enough that they should be treated as exactly common, so that a baud-spaced inverse of these common roots is formed in the equalizer. The channel disparity measure proposed in [22] is

$$\psi = \frac{\sigma_n^2}{\sigma_s^2} \cdot \frac{\left(\prod_i |\mathbf{c}_{even}(z_o^i)| \cdot \prod_j |\mathbf{c}_{odd}(z_e^j)|\right)^{-2}}{\prod_{i,j} \left|z_e^i - z_o^j\right|^2} \tag{4.9}$$

where z_e are the roots of $\mathbf{c}_{even}$ and z_o are the roots of $\mathbf{c}_{odd}$. There exists some typos relating this measure to the determinant of $\mathbf{C}_{\downarrow}{}^T\mathbf{C}_{\downarrow}$ in the original work [22], though [11] extends the results of [22] with illuminating decomposition of the MSE design objective (and corrects for the typos). Fijalkow (see [22]) observes that the measure depends on the determinant of the channel autocorrelation matrix $(\mathbf{C}_{\downarrow}{}^T\mathbf{C}_{\downarrow},)$ but recognizes that "...a small value for ...(the determinant) ...does not necessarily imply a large value for ...(the MSE)." The measure assumes the value $+\infty$ when the subchannels exactly share a root. Apparently, a "large" value of the performance measure ψ flags a problematic channel.

We evaluate (4.9) for channels $A(z), B(z)$ and length-32 channels derived from the database channels via frequency domain decimation discussed in §3.2. These channels are designated as c_{32}^i to denote a length-32 version of database channel i. The results are shown in Table 4.2 for a modest SNR of 30 dB

Fijalkow's disparity measure clearly flags channel $A(z)$ as problematic (though apparently not $B(z)$), despite the fact that these channels satisfy the identifiabil-

Table 4.2: Channel Disparity Measure for Various Channel Models

Channel	$A(z)$	$B(z)$	c_{32}^1	c_{32}^2	c_{32}^3	c_{32}^4	c_{32}^5
ψ	10^{+75}	10^{-97}	10^{-70}	10^{-95}	10^{-24}	10^{-10}	10^{-30}

ity conditions of [78]. The performance measure offers insight into the relationship between ill-conditioned channels, identification with SOS, and the cyclic all-pass class. The measure in (4.9) indicates not only the presence of a *single* common-zero pair, but considers the proximity of *all possible* subchannel pairs, $\prod_{i,j} |z_e^i - z_o^j|^2$. Loosely speaking, this method induces an averaging effect over all possible subchannel pairs. The cyclic all-pass class is now seen as an intuitive way of inducing the same type of averaging effect by placing the channel zeros in a symmetric fashion and thereby "minimizing" $\prod_{i,j} |z_e^i - z_o^j|^2$, without having any single subchannel pair "too close."

This connection motivates further study of channel $A(z)$. The left plot of Figure 4.4 shows the achievable MSE versus system delay for a length-16 FSE, which indicates a severe sensitivity. However, it also suggests that there do, in fact, exist equalization settings corresponding to specific delays which give quite good performance. CMA, however, appears unable to find one of these settings. For example, we initialize a length-16 FSE[3] updated by CMA with a unity-valued tap at position $\delta = 6$ and all other taps set to zero. This initialization starts the adaptation process at system delay[4] $\delta = 6$, which is in the "trough" of good MSE performance (see left plot of Figure 4.4). The stepsize is 1×10^{-3} and we use 16-

[3] Observe that $L_c = 22$ so that our FSE is slightly undermodeled

[4] By approximating the channel cursor at position 5, the system delay at the start of adaptation is approximated as $\delta = \lceil \frac{5+6}{2} \rceil = 6$.

QAM signalling with SNR of 100 dB. The MSE trajectory is shown in the right plot of Figure 4.4. CMA heads in the correct direction and then diverges. In fact, it is easily demonstrated that CMA diverges when initialized at the Wiener solution for this system delay. Perhaps a longer FSE is needed? The disturbing implication here is not necessarily the poor equalization of a recognized "problematic" channel like $A(z)$, but the possible mis-identification of a channel due to nearly indistinguishable SOS indicators for channels like $A(z)$ and $B(z)$ (see Table 4.1).

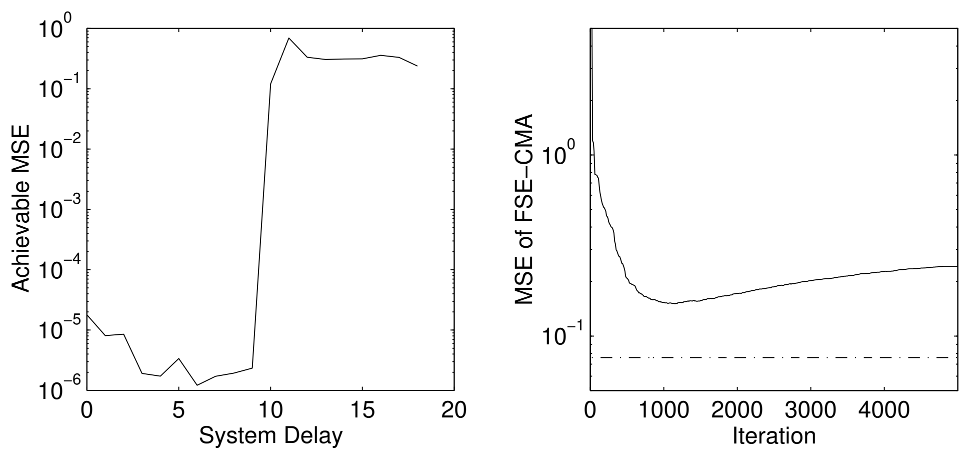


Figure 4.4: Poor Equalization of Channel $A(z)$

4.2 Undermodeling Robustness Concerns

Some would say that the simulation results in Chapter 3 and the analysis in §4.1 are testimony enough that SOS-based algorithms are not the right tool for the job of cold, start-up equalization using dense, high data rate signalling. Still, we crudely examine the channel-root-extraction task given the true SOS spectra in (4.1) and the notion that the channel estimate is of order less than the true channel. See [34], however, for channel-order estimation techniques.

The premise of the identification task as discussed in [79] is the correct root extraction from intersecting the roots of the Γ pair in (4.1) when the $T/2$-spaced channel does not contain reflected roots. Suppose the true statistics in (4.1) are known, though the estimated channel order, $\hat{L}$, is constrained to less than the true channel order, L. Presumably, a subset of the true channel roots are extracted from the application of (4.1). Admittedly, one should recognize that there exist roots which satisfy our intersection criterion, but were not extracted due to the length constraint. True enough, but the fact that these statistics must be estimated in practice requires that a thresholding of root proximity somehow governs the root-extraction task; i.e., the "closest" $\hat{L}$ roots are extracted from the application of (4.1).

With the idea that only a subset of the true channel roots are extracted in application of (4.1), we demonstrate that the combined channel equalizer can still contain substantial ISI. Let $\hat{L} = L - k$, where $\hat{L}$ is the order of the channel estimate, L is the order of the true channel, and k is the number of true channel zeros which are not extracted in application of (4.1). Then the true channel can be written as

$$C(z) = \prod_{i=1}^{k}(z - \alpha_i)\hat{C}(z) \tag{4.10}$$

where the α_i are those k channel zeros not extracted, and $\hat{C}(z)$ is the order-$\hat{L}$ channel estimate. Further assume that the equalizer, $f(z)$, found from this channel estimate perfectly equalizes $\hat{C}(z)$, so that the combined channel-equalizer response can be written as

$$h(z) \;=\; \prod_{i=1}^{k}(z-\alpha_i)\left[\hat{C}(z)f(z)\right] \qquad (4.11)$$

which upon decimation becomes

$$h(z)_{\downarrow} \;=\; \left[\prod_{i=1}^{k}(z-\alpha_i)\right]_{\downarrow} z^{-\delta} \qquad (4.12)$$

It is easy to imagine channels for which the decimated product of non-extracted channel roots in (4.12) can result in a closed eye channel-equalizer response, despite the fact that the equalizer perfectly equalizes the channel estimate.

4.3 Comments on Using SOS

Though it may be possible to perform some type of sensitivity analysis on each of the SOS algorithms and thereby produce performance bounds or even algorithm "band-aids" for specific cases, we leave that as SEP, and instead focus on the robustness properties of the CM criterion and CMA in the sequel. With this admission, we are recognizing the questionable robustness properties of the SOS algorithms specifically for the problem outlined in this dissertation–namely, in a cold, start-up application for high data rate, dense signalling. Performance for different applications, and hence under different assumptions, (for example, lower data rates and hence shorter channel models) may produce different conclusions.

Chapter 5

Historical CMA Robustness Results

Chapters 3 and 4 demonstrate by example and analysis a severe lack of robustness for recently proposed, mathematically clever, blind algorithms which rely purely on the received signal SOS. At the same time, Chapter 3 suggests attractive robustness properties for CMA. This chapter reviews existing robustness results in the open literature for CMA-FSE. Together, this chapter and Chapters 6 and 7 organize CMA-FSE's robustness properties in a pedagogical nature centered around violation of each of the P.E. assumptions in §2.2.4 separately. This chapter discusses the main existing robustness treatments of CMA-FSE for P.E. assumptions $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$. Chapters 6 and 7 provide original analysis for violation of $\mathbf{A_4}$ (length condition).

5.1 Source Distribution and Correlation

The analyses which demonstrates the global convergence of CMA to an open-eye setting ([6], [30], [24] for the baud-spaced case, and [56], [20] for the fractionally-spaced case) rely on the assumption that the source is independent and uniformly distributed. This assumption is academically amenable, but unrealistic in practice. LeBlanc studies the CM error surface when the source distribution is not uniform (known as shaping), but still independent, and also when the source is temporally correlated (or colored). His most complete work can be found in [51].

LeBlanc's analysis is organized around the system of equations resulting from gradient and Hessian expressions carried out in the combined channel-equalizer parameter space. Under the independent source assumption, the CM cost function is written as a function of the source kurtosis, $\kappa_s = \frac{E\{|s|^4\}}{E^2\{|s|^2\}}$. The number of stationary points (minima, maximum, and saddles) is counted as $(3^{L_c+L_f-1})$ and classified according to the source kurtosis. For example, when the combined channel-equalizer response has a single non-zero coefficient, which is the desired setting corresponding to P.E., LeBlanc classifies the stationary points of the CM criterion according to the source kurtosis. He shows that when the source kurtosis is platykurtic (or $\kappa_s < 3$), there exist only stable, local minima in the CM surface corresponding to the P.E. combined channel-equalizer setting. When the source kurtosis is leptokurtic (or $\kappa_s > 3$), these combined channel-equalizer parameterizations correspond to saddle points of the CM surface. (In this case, CMA is shown to converge to combined channel-equalizer settings which correspond to maximal ISI.) Thus, the CM error surface exhibits a bifurcation specifically when the source kurtosis corresponds to a gaussian distribution ($\kappa_s = 3$). Further, [51] demonstrates a flattening of the curvature in the CM error surface as the source kurtosis approaches 3. The disturbing implication here is the connection with constellation shaping. In [50] it is shown that for large two dimensional signalling (i.e., QAM), the optimal source distribution is essentially gaussian. LeBlanc recognizes these conflicting issues: "The relevancy of the results here point to the tradeoffs between coding gain (a function of the source distribution) and the error surface curvature (with its effects on convergence)."

When the source sequence is temporally correlated, LeBlanc uses Groebner Basis tools to develop algebraic solution methods for the CM stationary points under arbitrary source correlations. These solutions are extremely computationally intensive. Numerical techniques based on continuation methods are proposed for specific forms of correlation, including periodic and Markov sources. Simulations suggest a large variance in equalizer performance for a given "amount" of correlation. See [44], [53] or [54] for further results concerning temporal source correlation.

5.2 Additive Channel Noise

The presence of additive channel noise, possibly due to interference in the propagation channel or introduced by measurements in the receiver, is unavoidable in practice. Historically, analysis is performed in the absence of noise and behavior in the presence of noise is predicted based on the noiseless analysis. Such inference is usually acceptable with modest noise levels. However, it is only recently that the literature satisfactorily studies fractionally-spaced CMA in the presence of noise. For the baud-spaced case, see [57].

Recently, Tong has steered his focus away from SOS schemes and concentrated on the CM criterion. Zeng and Tong's first work on CMA appeared in [98], with numerous revisions following ([100], [101], [97], for example). Their most complete work to date can be found in [99]. A geometric approach is proposed in the presence of noise (but satisfying the length, source, and channel equalizability assumptions) to detect the presence of a CM minima in the neighborhood of a Wiener solution of unknown system delay. Under certain assumptions, the specification of regions which contain the CM minima and the corresponding upper and lower bounds on MSE performance of these minima may be possible. The approach relies on the equivalence of the minima settings for the two criteria (CM and MSE) in the absence of noise, so that when gaussian noise is added, the deformation in the CM error surface is small enough so that the CM minima stays in proximity with the MSE minima. A boundary region, $\mathcal{B}$, in the equalizer coefficient space is projected onto the CM cost surface, which defines a contour on the CM error surface. If the CM cost along all points of this contour is greater than the CM cost associated with a reference point which lies in $\mathcal{B}$, then there exists a CM local minima also in the region $\mathcal{B}$. This fact is true due to the Weierstrass theorem [73] which states that a continuous function (the CM cost function) on a compact set (the region $\mathcal{B}$) must have a minimum. The region $\mathcal{B}$ and the reference point in $\mathcal{B}$ are found by a signal space and matched filter interpretation of the CM receiver. It should be noted that unlike other analyses of CMA in the presence of noise ([83], for example) the analysis in [99] is exact in that it involves no approximations. Simulation results suggest that when the approach can in fact detect the presence of a CM local minima, the performance is tight relative to the true MSE performance.

Another recent treatment of the fractionally-spaced CM criterion in the presence of noise (but satisfying the length, source, and channel equalizability assumptions) is [83] and its journal version [23]. The CM cost in the presence of noise is written as a linear combination of the noiseless CM cost and a regularized term proportional to the inverse of the SNR and also proportional to the FSE noise gain (squared ℓ_2-norm of the equalizer coefficient vector). Since the regularized term is proportional to the equalizer noise gain, a tradeoff exists between ISI removal and noise gain enhancement. The authors refer to this robustness property as a "smoothing effect" and state the minima of the CM criterion "realize a balance between the noise-free ... equalization settings and the noise enhancement due to the equalizer norm." (Observe that this behavior is precisely the type of trade-off discussed for the MSE criterion in §2.2.1.) Since this regularized term also has fourth-order dependence on the equalizer coefficient vector, the subsequent analysis is based on a quadratic approximation of the regularized term. Under this approximation, the combined channel-equalizer response in the presence of noise is written as a perturbation to the pure delay which represents P.E. The perturbation depends on the SNR, channel, and system delay. The MSE of the CM receiver is approximated based on this perturbation. Reasonable accuracy to true performance is obtained for modest noise levels with a low dimensional example.

5.3 Common SubChannel Roots

When the $T/2$-spaced channel model contains roots which are reflected through the origin, or equivalently the subchannels share a common root (see [89], [78], or [45]) the channel is said to undergo a loss of disparity and P.E. is no longer achievable with a finite length FSE. Fijalkow et al. show in [21] that the achievable performance (when length, noise, and source assumptions are satisfied) depends on the part of the multichannel transfer function which lacks disparity. The Diophantine equation (see §2.2.2) $\mathbf{h} = \mathbf{c}_{even}\mathbf{f}_{even} + \mathbf{c}_{odd}\mathbf{f}_{odd}$ can be re-written in the presence of common roots as $\mathbf{h} = \mathbf{c}_{com}(\mathbf{c}'_{even}\mathbf{f}_{even} + \mathbf{c}'_{odd}\mathbf{f}_{odd})$, where $\mathbf{c}_{com}$ contains those zero locations common to both subchannels. Since the desired response is a pure delay, the design objective involves inverting $\mathbf{c}_{com}$, which is not perfectly achievable with a finite length equalizer. The recently proposed SOS algorithms in this case are therefore not applicable. The authors [21] recognize this concern, and state that the SOS algorithms "...proposed to solve the multichannel equalization problem can no longer be used." It is shown in [43] that CMA, however, has no problem providing the approximate baud-spaced inverse of $\mathbf{c}_{com}$ provided the equalizer length is "long enough". Hence, CMA effectively solves for the *baud-spaced* inverse of the common part of the channel response, and leaves the remaining FSE taps to solve the remaining Diophantine equation of FSE design. Simulation results in [21] show that for low dimensional examples and an equalizer sufficiently long to approximate the inverse of the common roots, CMA-FSE behaves precisely as predicted, achieving "reasonable equalization." Presumably, this work suggests that a *sufficiently long* CMA-FSE should not struggle on problematic channel $A(z)$ discussed earlier in Chapter 4. Recall that Figure 4.4 used a CMA-FSE which did not satisfy the length condition needed for P.E.

More recently, Zeng et al. [100] have extended the approach of [98] and Touzni et al. [82] have extended the approach of [83] to accommodate singular channels. Both [100] and [82] approximate the MSE of the CM receiver, and both provide simulation results showing reasonable agreement with their approximations.

It should be noted that the disparity assumption is not just an academic game, but is of practical interest. The channel models derived from empirical measurements in [32] and available at the database exhibit numerous *nearly* reflected roots.

Chapter 6

CMA Robustness to Length Condition

Clearly, Chapter 5 shows that the literature contains both reasonable and satisfactory treatment for the robustness of the CM criterion and FSE-CMA to violation of P.E. assumptions $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$. Surprisingly, the literature contains essentially no analysis for the undermodeled case. One exception is [97], which extends the geometrical approach of [98]. Zeng et al. (independently and simultaneous with our original work) provide approximations for the MSE of the undermodeled CM receiver when sufficient conditions are satisfied. This chapter provides original analysis describing robustness properties of the CM criterion and FSE-CMA to violation of $\mathbf{A_4}$. Because our approach is algebraic, the derived result is both simple and illuminating.

We study the CM criterion specifically for the case when the FSE time span is less than that needed for P.E. Hence, there necessarily exists an error in the equalized signal. We take an algebraic approach, which is not wholly unrelated to the geometric approach in [98]. The change in CM cost from a P.E. setting is derived for two cases: i) perturbations to the channel outside the time span of the equalizer, and ii) equalizer truncation. We first consider binary signalling, and later extend these results to real, multi-level signalling. This analysis is connected with the previous work of Fijalkow et al. [20] on excess MSE (which is related to the misadjustment) of a CM receiver. As with classical LMS [36] if the CMA-FSE length is too short, then there remains too much ISI, while if the CMA-FSE length is too long, then excess MSE can dominate ISI reduction. Database examples suggest a finite interval of acceptable FSE length which shows that a longer FSE may not be better than a shorter FSE–in some cases, matching the FSE length to that of the channel may not reduce the MSE to the prescribed Transfer Level, while a shorter FSE may be successful in achieving this threshold. The approach can be used to infer a design guideline for FSE length selection so that the CMA-FSE can achieve the Transfer Level MSE. The design guideline suggests that the FSE length be chosen long enough to cover those "significant" channel coefficients–those whose magnitudes are greater than some percentage of the magnitude of the largest channel tap. For the database channels and 16-QAM signalling, this percentage is about $15-20\%$. Earlier versions of this work can be found in [18] and [19]. See [58] for baud-spaced results concerning the undermodeled case.

6.1 Analysis Approaches

6.1.1 Channel Perturbation

The first approach taken in addressing the robustness of a CM receiver to under-modeling is to consider those channel coefficients that are outside the time span of the FSE as channel perturbations, in order to study the CM cost incurred. Though the proposed approach is easily extended to consider channel perturbations also within the FSE time span, our goal is to infer a design guideline for FSE length selection, so our approach is intuitive. The CM criterion is specifically the one minimized by CMA-FSE in a stochastic gradient descent implementation, though MSE or even BER may be the ultimate performance measure. Let

$$\mathbf{c} \;=\; [c_0\; c_1 \ldots c_{L_c-1}]^T \tag{6.1}$$

be the length-L_c fractionally-sampled channel impulse response vector, which is zero outside this finite time support. Define two length-L_c vectors; vector $\mathbf{c_m}$ contains L_f ($L_f < L_c$) consecutive taps of $\mathbf{c}$ in the same positions as they occurred in $\mathbf{c}$ with zeros in the remaining $L_c - L_f$ positions, and vector $\mathbf{c_p}$ contains the $L_c - L_f$ taps of $\mathbf{c}$ that are not in $\mathbf{c_m}$ in the same positions as they occurred in $\mathbf{c}$, with zeros in the remaining L_f positions. Hence, $\mathbf{c} = \mathbf{c_m} + \mathbf{c_p}$; with a length-$L_f$ FSE, the "full length" channel is composed of a "modeled" portion which may be perfectly equalizable, and a "perturbation" portion which is potentially non-zero outside the time support of the FSE. For example, one such partitioning of the channel coefficients is

$$\begin{aligned} \mathbf{c_m} &:= [c_0\; c_1 \ldots c_{L_f-1}\;\; \underbrace{0\;0\ldots 0}_{L_c-L_f\;\; zeros}\;]^T \\ \mathbf{c_p} &:= [\underbrace{0\;0\ldots 0}_{L_f\;\; zeros}\; c_{L_f}\; c_{L_f+1}\ldots c_{L_c-1}]^T \end{aligned} \tag{6.2}$$

which considers the perturbation as appended channel taps with largest delay from the current symbol.

Further, let $\mathbf{C_m}$, $\mathbf{C_p}$ and $\mathbf{C}$ be the convolution matrices associated with $\mathbf{c_m}$, $\mathbf{c_p}$ and $\mathbf{c}$, respectively, and let $\mathbf{f_m}$ be the equalizer coefficient vector corresponding to a global minimum of the CM cost associated with channel $\mathbf{c_m}$. The combined channel-equalizer before decimation becomes

$$\begin{aligned} \mathbf{h} &= \mathbf{C}\mathbf{f_m} && (6.3) \\ &= (\mathbf{C_m} + \mathbf{C_p})\mathbf{f_m} && (6.4) \\ &= \mathbf{C_m}\mathbf{f_m} + \mathbf{C_p}\mathbf{f_m} && (6.5) \end{aligned}$$

and the decimated (baud-spaced) version can be written as

$$\begin{aligned} \mathbf{h}_{\downarrow} &= \mathbf{C}_{\mathbf{m}\downarrow}\mathbf{f_m} + \mathbf{C}_{\mathbf{p}\downarrow}\mathbf{f_m} && (6.6) \\ &= \mathbf{h_m} + \mathbf{h_p} && (6.7) \end{aligned}$$

where $\mathbf{C}_{\mathbf{m}\downarrow}$ and $\mathbf{C}_{\mathbf{p}\downarrow}$ are appropriately row-decimated versions of $\mathbf{C_m}$ and $\mathbf{C_p}$, respectively. Observe that since $\mathbf{f_m}$ is a global minimum of the CM criterion with respect to channel $\mathbf{c_m}$, there is no error in the equalized signal due to the first term in (6.7) provided the source kurtosis is less than 3 (see [51]). The second term is the effect of the channel perturbations outside the time span of the FSE.

6.1.2 Equalizer Truncation

A related approach to that above is to consider the effect on the CM cost due to equalizer coefficients lost in truncation from the FSE which achieves P.E. Let

$$\mathbf{f}^{\ddagger} = [f_0^{\ddagger}\ f_1^{\ddagger} \dots f_{L_c-1}^{\ddagger}]^T \qquad (6.8)$$

be a CM global minimum for channel $\mathbf{c}$. Define two length-L_c vectors $\mathbf{f_t}$ and $\tilde{\mathbf{f}}$ such that $\mathbf{f_t} = \mathbf{f}^{\ddagger} + \tilde{\mathbf{f}}$. Vector $\mathbf{f_t}$ contains L_f consecutive taps of $\mathbf{f}^{\ddagger}$ in the same positions as they occurred in $\mathbf{f}^{\ddagger}$ with zeros in the remaining $L_c - L_f$ positions, and vector $-\tilde{\mathbf{f}}$ contains the $L_c - L_f$ taps of $\mathbf{f}^{\ddagger}$ that are not in $\mathbf{f_t}$ in the same positions as they occurred in $\mathbf{f}^\ddagger$, with zeros in the remaining L_f positions. For example, one such partitioning is

$$\begin{aligned}
\mathbf{f_t} &:= [f_0^\ddagger \; f_1^\ddagger \ldots f_{L_f-1}^\ddagger \; \underbrace{0\;0\ldots 0}_{L_c-L_f \; zeros}\,]^T \\
\tilde{\mathbf{f}} &:= [\underbrace{0\;0\ldots 0}_{L_f \; zeros} \; -f_{L_f}^\ddagger \; -f_{L_f+1}^\ddagger \ldots - f_{L_c-1}^\ddagger]^T
\end{aligned} \tag{6.9}$$

The baud-spaced, combined channel-equalizer response for the truncated equalizer $\mathbf{f_t}$ can be written as

$$\mathbf{h}_\downarrow = \mathbf{C}_\downarrow \mathbf{f_t} \tag{6.10}$$

$$= \mathbf{C}_\downarrow(\mathbf{f}^\ddagger + \tilde{\mathbf{f}}) \tag{6.11}$$

$$= \mathbf{C}_\downarrow \mathbf{f}^\ddagger + \mathbf{C}_\downarrow \tilde{\mathbf{f}} \tag{6.12}$$

where $\mathbf{C}_\downarrow$ is the appropriately row-decimated version of $\mathbf{C}$. Observe that (6.12) is the same form as (6.7), where the first term is analogous to the "modeled" contribution, and second term is analogous to the "perturbation" contribution. Thus, the first term satisfies the length condition and will achieve P.E. since $\mathbf{f}^\ddagger$ is a global minimum of the CM criterion, provided the source kurtosis is less than 3 (see [51]). Our goal in addressing CMA's robustness properties is to study the effect of the second terms of (6.7) and (6.12) on the CM cost function.

6.2 Binary CM Criterion

Notice that the above analysis approaches do not specifically address the CM cost of the optimal (in CM-sense) length-L_f FSE ($L_f < L_c$). Instead, our analysis approaches change the problem slightly to one that is answerable. The CM-cost for this length-L_f optimum (in CM sense) FSE, however, is bounded due to optimality by the cost we now derive in (6.26).

The CM cost function for the BPSK, white, zero-mean, equiprobable source case can be expanded from its form in (2.44) and written as (see [43])

$$J_{CM|BPSK} = 1 - 2\sum_{i=0}^{P} h_i^2 + 3\left(\sum_{i=0}^{P}\sum_{j=0,j\neq i}^{P} h_i^2 h_j^2\right) + \sum_{i=0}^{P} h_i^4 \tag{6.13}$$

where the h_i are the elements of $\mathbf{h}_{\downarrow}$ and $P = \lceil (L_c + L_f - 2)/N \rceil$, with $\lceil\cdot\rceil$ denoting round-up to the nearest integer, and N is the oversample ratio (typically 2). Further, each h_i can be expressed in terms of the "modeled" and "perturbation" portions of the channel-equalizer combination, $h_i = m_i + p_i$, i.e., $m_i \in \mathbf{C}_{\mathbf{m}\downarrow}\mathbf{f}_{\mathbf{m}}$ or $m_i \in \mathbf{C}_{\downarrow}\mathbf{f}^{\ddagger}$, and $p_i \in \mathbf{C}_{\mathbf{p}\downarrow}\mathbf{f}_{\mathbf{m}}$ or $p_i \in \mathbf{C}_{\downarrow}\tilde{\mathbf{f}}$.

Now consider the three terms of (6.13) separately.

$$\begin{aligned}
1 - 2\sum_{i=0}^{P} h_i^2 &= 1 - 2\sum_{i=0}^{P}(m_i + p_i)^2 && (6.14)\\
&= 1 - 2\sum_{i=0}^{P}(m_i^2 + 2m_i p_i + p_i^2) && (6.15)\\
&= 1 - 2(1 + 2p_\delta + p_\delta^2) - 2\sum_{i\neq\delta,i=0}^{P} p_i^2 && (6.16)\\
&= -1 - 2(2p_\delta + p_\delta^2) - 2\sum_{i\neq\delta,i=0}^{P} p_i^2 && (6.17)
\end{aligned}$$

since the "modeled" part of the channel achieves P.E. Hence, $m_\delta = 1$ and $m_i = 0 \ \forall\ i \neq \delta$. (The case where $m_\delta = -1$ is shown to be equivalent in Appendix B.)

Similarly, express

$$\begin{aligned}
\sum_{i=0}^{P} h_i^4 &= \sum_{i=0}^{P}(m_i + p_i)^4 \\
&= \sum_{i=0}^{P} m_i^4 + 4m_i^3 p_i + 6m_i^2 p_i^2 + 4m_i p_i^3 + p_i^4 && (6.18)\\
&= 1 + 6p_\delta^2 + 4p_\delta + 4p_\delta^3 + \sum_{i=0}^{P} p_i^4 && (6.19)
\end{aligned}$$

And for the double sum in (6.13)

$$
\begin{aligned}
\sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} h_i^2 h_j^2 = \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} (m_i + p_i)^2 (m_j + p_j)^2 && (6.20) \\
= \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} (m_i^2 + p_i^2 + 2m_i p_i)(m_j^2 + p_j^2 + 2m_j p_j) && (6.21) \\
= \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} m_i^2 m_j^2 + m_i^2 p_j^2 + 2m_i^2 m_j h p_j + p_i^2 m_j^2 && (6.22) \\
+ \ p_i^2 p_j^2 + 2p_i^2 m_j p_j + 2m_i p_i m_j^2 + 2m_i p_i p_j^2 + 4m_i p_i m_j p_j && \\
= \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} m_i^2 p_j^2 + p_i^2 m_j^2 + p_i^2 p_j^2 + 2p_i^2 m_j p_j + 2m_i p_i p_j^2 && (6.23) \\
= \sum_{j\neq\delta, j=0}^{P} p_j^2 + \sum_{i\neq\delta, i=0}^{P} p_i^2 + \sum_{i=0}^{P} \sum_{j\neq i, j=0}^{P} p_i^2 p_j^2 && (6.24) \\
+ \ 2p_\delta \sum_{i\neq\delta, i=0}^{P} p_i^2 + 2p_\delta \sum_{j\neq\delta, j=0}^{P} p_j^2 && \\
= 2(1 + 2p_\delta) \sum_{i=0, i\neq\delta}^{P} p_i^2 + \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2 && (6.25)
\end{aligned}
$$

since terms involving cross products of m_i and m_j are zero due to the summation limits and the perfect equalizability assumption of the m_i.

Now collecting terms (6.17), (6.19), and (6.25), the CM cost incurred from undermodeled channel taps is expressed as

$$
\begin{aligned}
\Delta J_{CM|BPSK} &= \left[4||\mathbf{h_p}||_2^2\right] + p_\delta \left[4p_\delta^2 + 12 \sum_{i\neq\delta, i=0}^{P} p_i^2\right] \\
&+ \left[\sum_{i=0}^{P} p_i^4 + 3 \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2\right] \qquad (6.26)
\end{aligned}
$$

Observe that the terms in (6.26) are grouped according to powers of the perturbation elements p_i. Further observe that the cubic terms in the incurred CM cost are proportional to p_δ and therefore depend on the delay sought in $\mathbf{h_m}$ by the CMA-FSE, which is unknown due to the inherent phase invariance in the CM cost function.

6.2.1 Relation to MSE

CMA-FSE's performance as reported in field tests [85] and performance studies [17], [40], [68] suggest broad capabilities of the CM criterion for blind update of a FSE. Recognize, however, a classical compromise: the user of CMA-FSE is minimizing one cost in hopes of minimizing another cost, namely MSE (which is equalivalent to minimizing BER under certain assumptions on the noise and source statistics [63], [69]). However, the CM cost can be related to the MSE cost in the vicinity of a local minimum. We next derive the change in the MSE cost according to our analysis approaches in §6.1, as was done in deriving (6.26).

The MSE cost is expressed as

$$J_{MSE} \;=\; E\{(y-s_\delta)^2\} \;=\; \sigma_s^2\left(\sum_{i=0}^{P} h_i^2 + 1 - 2h_\delta\right) \tag{6.27}$$

where y is the FSE output, s_δ is a source symbol, and σ_s^2 is the source power. Hence, for a sequence of ±1's (or unit variance source), the change in MSE cost is

$$\begin{aligned}\Delta J_{MSE} \;&=\; \sum_{i=0}^{P}(m_i+p_i)^2 + 1 - 2(m_\delta + p_\delta) &(6.28)\\ &=\; ||\mathbf{h_p}||_2^2 &(6.29)\end{aligned}$$

Therefore, when the higher-order terms in (6.26) can be neglected (specifically when the perturbation terms are "small") the change in CM and MSE costs are related by

$$\Delta J_{CM|BPSK} \;\approx\; 4\cdot\Delta J_{MSE} \tag{6.30}$$

In this case, the CM and MSE costs incurred due to undermodeling are "small," and the CM cost is approximately a scaled version of the MSE cost. This behavior suggests a small deformation in both error surfaces due to undermodeling, so that the CM minimum stays in a tight neighborhood of the Wiener solution.

6.2.2 Interpretive Examples and Design Guideline

The results thus far help to explain the robustness of the CM criterion to the nearly unavoidable situation of channel order undermodeling, which prevents P.E. The objective of this section is to offer interpretive examples of these results and infer a design guideline for FSE length selection of the database channels based on this preceding analysis.

Low-Dimensional Example

We first present a low-dimensional example similar to that considered in Figure 2.11 which demonstrates the behavior predicted by the preceding analysis; namely, that when the perturbation coefficients are "small," the deformation in the CM error surface is "small," and the CM minima stay in close proximity with the length-constrained MSE minima.

The channel considered in this experiment is the result of appending a perturbation vector of ten channel coefficients to the four channel coefficients that were used to generate Figure 2.11. Let $\mathbf{c_m}$ and $\mathbf{c_p}$ be defined by

$$\mathbf{c_m} = [\ 0.2\ \ 1\ \ -0.7\ \ 0.15\ \ \underbrace{0\ \ 0\ \ \ldots\ \ 0}_{10\ zeros}]^T \qquad (6.31)$$

$$\mathbf{c_p} = [\ 0\ \ 0\ \ 0\ \ 0\ \ \epsilon_0\ \ \epsilon_1\ \ \ldots\ \ \epsilon_9\ ]^T \qquad (6.32)$$

so that the channel used in this experiment is $\mathbf{c} = \mathbf{c_m} + \mathbf{c_p}$. The channel perturbations are chosen according to $\epsilon_i = a_i \cdot u_i$, where u_i is a random variable chosen uniformly over $[0, \epsilon_{max}]$, and a_i is a binary random variable chosen equally likely from $\{-1, +1\}$ which represents polarity.

Figure 2.11 thus represents the case which is perfectly equalizable by two FSE coefficients when the odd samples correspond to "on-baud"signalling. Our experiment here studies the deformation of the CM error surface in Figure 2.11 due to undermodeling. Figure 6.1 contains six plots, each for a different value of ϵ_{max} (0.05, 0.1, 0.15, 0.2, 0.25, 0.3). These plots show three things; (i) contour lines of the CM error surface in equalizer parameter space corresponding to channel $\mathbf{c}$, (ii) the P.E. settings (designated by $*$'s) corresponding to the modeled channel $\mathbf{c_m}$, and (iii) the length-constrained (to two FSE taps) MSE minima (designated by $\times$'s) for the full channel $\mathbf{c}$. Hence, the distance from the CM minima to their corresponding $*$ represents how far the CM minima move away from the P.E. setting, and the distance from the CM minima to their corresponding $\times$ represents the proximity of the undermodeled Wiener and CM minima. When the perturbation coefficients are "small" ($\epsilon_{max} \leq 0.15$,) the CM minima move little from the P.E. settings, and the CM minima stay in a tight neighborhood around their corresponding Wiener solutions. When the perturbation coefficients are increased with $\epsilon_{max} = 0.2$, the existence of local minima (those two CM minima nearly in the vertical direction) begins to be apparent. When ϵ_{max} is increased to 0.25, these local minima are clear, and do not stay in close proximity with the Wiener solutions. However, the other two CM minima (nearly in the horizontal direction) still remain close to the P.E. settings and Wiener solutions. When ϵ_{max} is further increased to 0.3, all the CM minima have moved substantially from the P.E. settings and away from the Wiener solutions.

This example agrees with the preceding analysis–when the perturbation coefficients are "small," the deformation in the CM error surface is "small," and the CM minima stay in close proximity with the Wiener solutions.

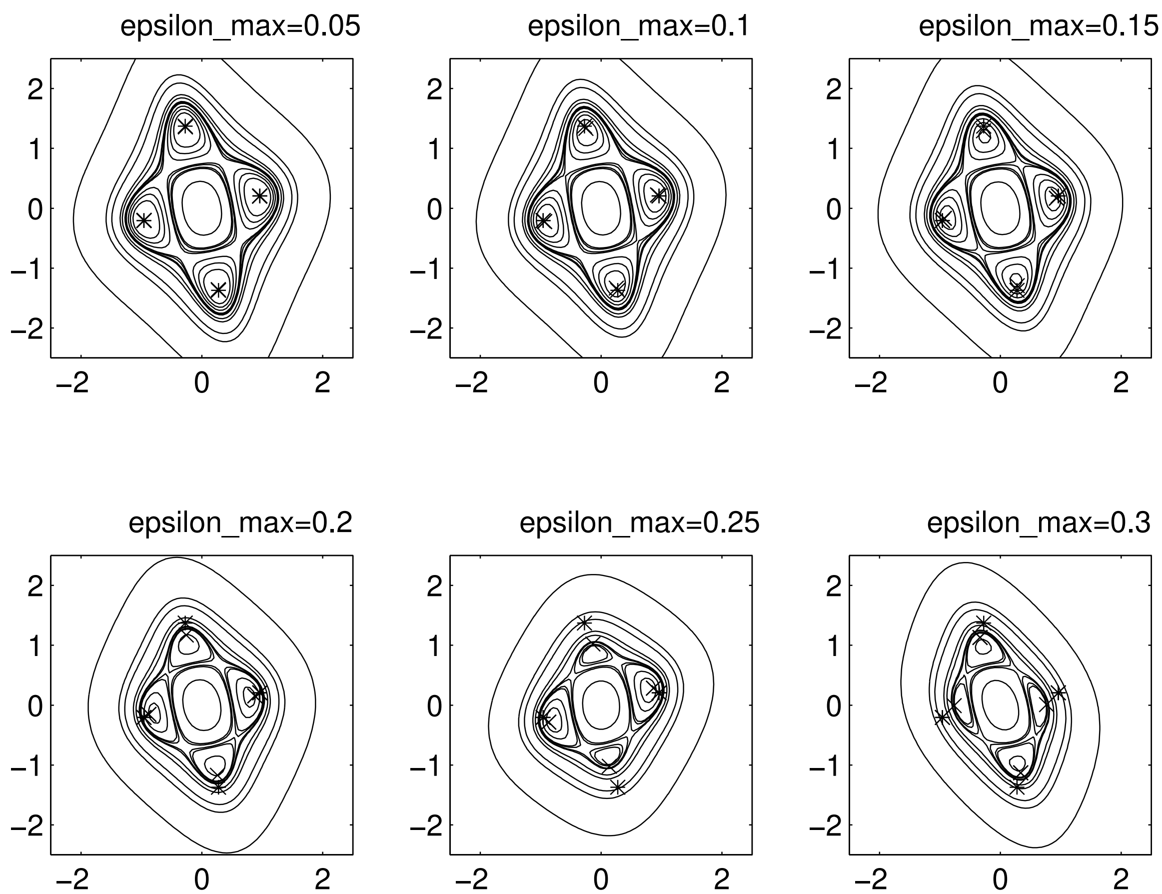


Figure 6.1: CM Error Surface Deformation Due to Undermodeling

Design Guideline

The goal of this section is to infer a design guideline for FSE length selection for the database channels based on 16-QAM signalling. The design guideline should at least attempt to keep the quantity in (6.26) "small". To this end, the "perturbation" coefficients should be small enough so that the cubic and quartic terms in (6.26) can be neglected and the CM cost is approximately a scaled version of the MSE cost. Hence, the FSE should span the "significant" portion of the channel impulse response. The preceding topographical example hints at what "significant" means – perhaps around $\epsilon_{max} = 0.2$ – though we return to the empirically measured channel models from the database for undeniable practicality.

We evaluate (6.26) for the two approaches previously described in §6.1. The channel taps for the channel-perturbation approach of §6.1.1 are partitioned according to (6.2). The partitioning of FSE coefficients for the FSE-truncation approach of §6.1.2 is according to (6.9). The results are scaled (by a factor of $\frac{1}{4}$) to approximate the MSE cost. The system delay is found according to (2.36). Figures 6.2, 6.3 and 6.4 show the results for channels 3, 1 and 4 from the database, respectively. Results for the remaining twelve channels of the database look similar. Each figure contains two plots. The top plot contains the magnitude of the $T/2$-spaced channel impulse response coefficients. The bottom plot contains the graphs of three functions: i) (solid) the approximation of MSE from a scaled version of (6.26) due to channel perturbations outside the FSE time span (§6.1.1), ii) (dashed) the approximation of MSE from a scaled version of (6.26) due to FSE truncation (§6.1.2), and iii) (dotted) the true MSE described by (6.29) according to the approach in §6.1.1 (note that for clarity, we omit the analogous graph for the true MSE according to the approach in §6.1.2)–this quantity is precisely a scaled version of the quadratic contribution of (6.26). The graphs may be referenced to the dashed line of constant MSE which corresponds to a Transfer Level corresponding to 16-QAM signalling for which CMA is typically transferred to DD-LMS (see Appendix A). These figures show that though a FSE length at least as long as the channel is needed for P.E., far fewer CMA-FSE taps are needed for successful transfer to DD-LMS for this practical class of channels. The "significant" portion of the channel appears to be those coefficients whose magnitudes are greater than about 20% of the magnitude of the largest channel tap (agreeing with $\epsilon_{max} = 0.2$ as the approximate threshold in the preceding topological example,) since little improvement in the approximate MSE is observed by increasing the FSE length to span coefficients less than this threshold. Also, in this region, the true MSE as described in iii) above is essentially the same as the approximated MSE as described in i) above, i.e., the CM cost is essentially a scaled version of the MSE cost. Therefore, the CM minimum stays in a tight neighborhood of the MSE minimum. These figures also show that the two different but related approaches described in §6.1 are not order-able (the solid line is not always greater than the dashed line, and vice versa), suggesting the validity of both.

It should be noted that this design guideline is quite different than that typically used for the baud-spaced case, but quite similar in motivation. For example, sidebar C of [85] proposes for the baud-spaced case to first approximate the channel as a two-ray model, so that the channel inverse and hence sum of unequalized terms can be approximated and kept "small." The baud-spaced equalizer length is chosen as a multiple of the channel delay spread. The multiplicity depends on

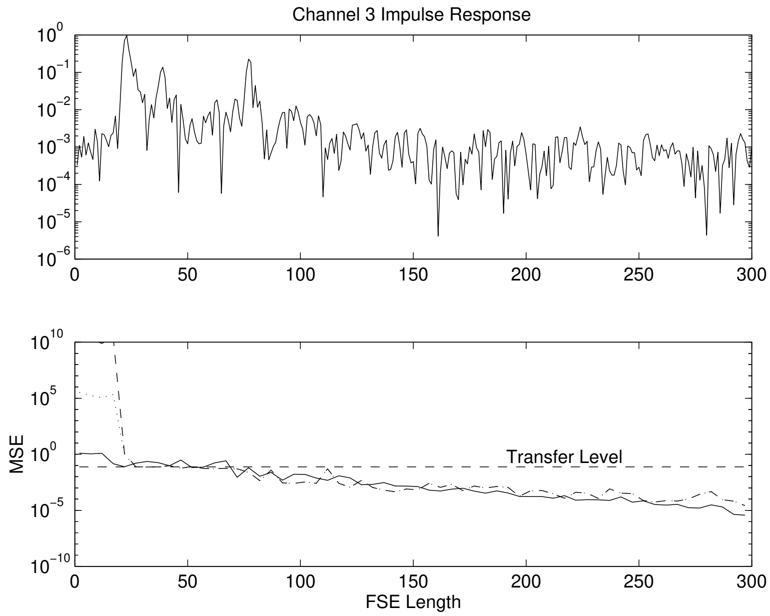


Figure 6.2: CM Robustness to Undermodeling, Channel 3, Binary Signalling

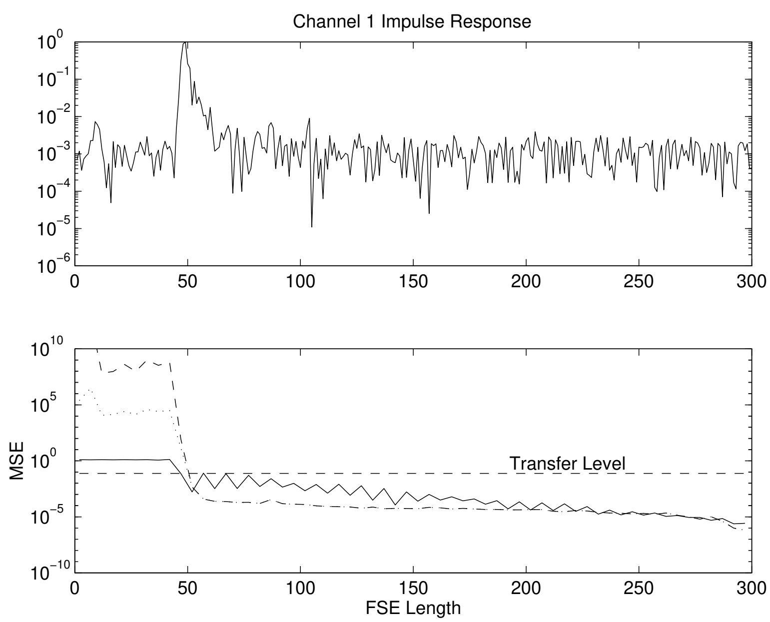


Figure 6.3: CM Robustness to Undermodeling, Channel 1, Binary Signalling

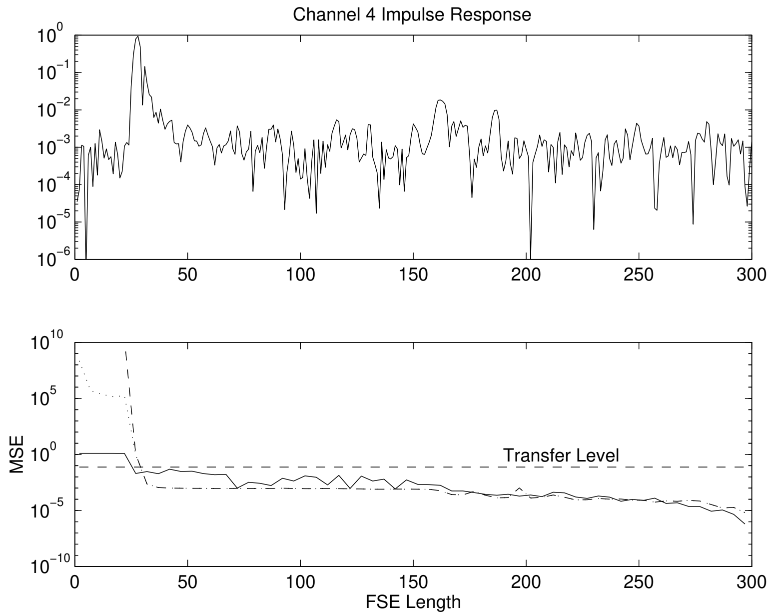


Figure 6.4: CM Robustness to Undermodeling, Channel 4, Binary Signalling

the prescribed threshold.

16-QAM Example

We next test the FSE length-selection-guideline which was derived for noiseless, BPSK signalling and apply it to a 32-tap channel derived via the frequency-domain decimation of database channel 3 (see §3.2,) whose impulse response magnitudes are shown in the left plot of Figure 6.5. Our experiment, however, uses more realistic 16-QAM signalling with additive white gaussian noise of SNR 30 dB. The design guideline suggests that the FSE span the "significant" channel coefficients– those taps greater in magnitude than approximately $15-20\%$ of the magnitude of the largest tap. This rule implies the FSE span the second and third (dominant) rays of this channel, about $27-12=15$ FSE taps. CMA-FSE is initialized with a unity center spike and the MSE trajectories are averaged over ten independent source sequences of length 10,000 $T/2$-spaced samples. The MSE trajectories for various FSE lengths are shown in the right plot of Figure 6.5.

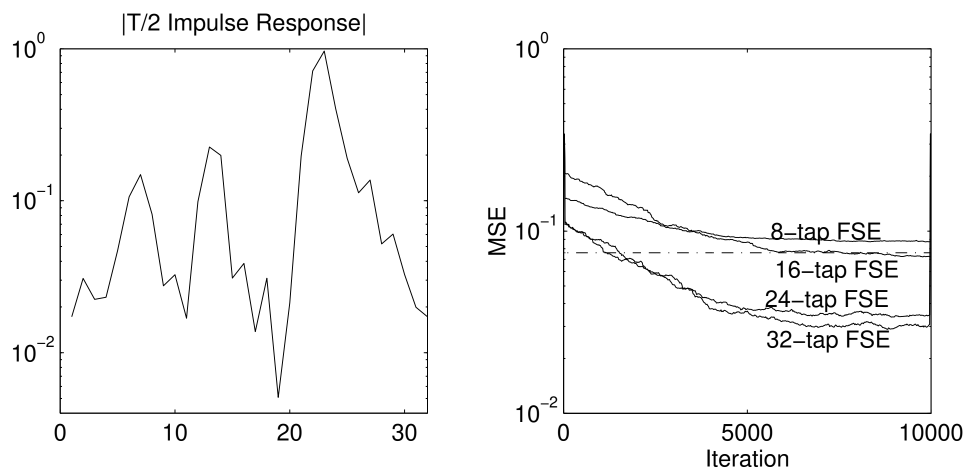


Figure 6.5: CMA-FSE on 32-tap Microwave Channel Model

The results agree remarkably well with our simple guideline; a length-16 FSE is barely sufficient to reach the Transfer Level, while a length-8 FSE fails and length-24 FSE is excessively long. Though our design guideline was derived from analysis assuming noiseless, BPSK signalling, it appears robust when applied to higher-order, complex signalling with modest noise power.

6.2.3 Other Observations

Other observations on the analysis can be made which may (or may not) prove insightful for certain problems.

1. **Quadratic Bound**

   When the error perturbation elements are small,

$$\mathbf{h_p} \quad = \quad [\epsilon_0 \ \epsilon_1 \ldots \ \epsilon_P]^T \tag{6.33}$$

with $|\epsilon_i| \leq \epsilon_{max} \; \forall \; i$, then (6.26) is dominated by the quadratic terms and can be bounded:

$$\begin{aligned} \Delta J_{CM} &\approx 4||\mathbf{h_p}||_2^2 && (6.34) \\ &\leq 4(P+1)\epsilon_{max}^2 && (6.35) \\ &= (\frac{4}{N})(L_c + L_f + N - 2)\epsilon_{max}^2 && (6.36) \end{aligned}$$

This bound may be helpful for analysis of certain known channel classes.

2. **Initialization Dependence and Relative Depth of Local Minima**

When $L_f \approx L_c$, the vector $\mathbf{c_p}$ is predominantly filled with zeros, by definition (see (6.2) and (6.9)). Hence, the corresponding combined channel-equalizer response, $\mathbf{h_p}$, has many zero-valued elements. Depending on the value of δ sought by CMA-FSE, $p_\delta \in \mathbf{h_p}$ is likely to be zero. In this case, the cubic terms in (6.26) disappear, leaving only quadratic and quartic contributions. This dependence suggests a connection between the penalty paid due to under-modeling and "proper" CMA-FSE initialization, which remains unresolved, and gives an indication of the relative CM cost between CM local minima due to undermodeling. For example, let $\alpha_i \in \mathbf{h_p}$ for delay $\delta = a$ and $\beta_i \in \mathbf{h_p}$ for delay $\delta = b$. Note that in general, $\alpha_i \neq \beta_i$, since the equalization settings at the CM minima are different for the different delays. When these equalizer settings have approximately the same ℓ_2-norm (which is evidenced as usually the case when the subchannels do not contain (nearly) reflected roots in our 2-dimensional example of §6.2.2,) the quadratic contributions of (6.26) are approximately the same when evaluated for the different delays. In this case, the difference in the cubic contributions of (6.26) with their initialization dependence approximate the relative cost between the CM local minima. We express the relative CM cost between different CM local minima corresponding to delays $\delta = a$ and $\delta = b$ as

$$
\begin{aligned}
|J_{CM|\delta=a} - J_{CM|\delta=b}| \approx & \qquad\qquad (6.37) \\
& \left| \alpha_a \left(4\alpha_a^2 + 12 \sum_{i \neq a, i=0}^{P} \alpha_i^2 \right) - \beta_b \left(4\beta_b^2 + 12 \sum_{i \neq b, i=0}^{P} \beta_i^2 \right) \right| \\
= & \left| 4(\alpha_a^3 - \beta_b^3) + 12 \left(\alpha_a \sum_{i \neq a, i=0}^{P} \alpha_i^2 - \beta_b \sum_{i \neq b, i=0}^{P} \beta_i^2 \right) \right| \qquad (6.38)
\end{aligned}
$$

3. **Sensitivity to Reflected Roots**

The terms in (6.26) can be expected to suffer a relative increase in magnitude when either the "full-length" channel, $\mathbf{c}$, or the "modeled" portion, $\mathbf{c_m}$, have nearly reflected roots. For example, suppose the length condition is violated according to the approach described in §6.1.1, so that $m_i \in \mathbf{C_{m\downarrow}f_m}$ and $p_i \in \mathbf{C_{p\downarrow}f_m}$. As the "modeled" portion of the channel ($\mathbf{c_m}$) loses disparity, the equalizer suffers noise gain enhancement (an increase in the relative magnitude of the elements of $\mathbf{f_m}$,) which in turn causes a magnification of the p_i and thus a relative increase in (6.26). Compare this idea with the approach discussed in §6.1.2 of truncated FSE coefficients, so that now $m_i \in \mathbf{C}_\downarrow \mathbf{f}^\ddagger$ and $p_i \in \mathbf{C}_\downarrow \tilde{\mathbf{f}}$. In this case, it is reflected roots in $\mathbf{c}$ rather than $\mathbf{c_m}$ which cause a relative increase in the coefficients of $\mathbf{f}^\ddagger$ and hence also in $\tilde{\mathbf{f}}$ and the p_i. The bounds derived in (6.26) may therefore become too large to be practical when $\mathbf{c}$ or $\mathbf{c_m}$ become ill-conditioned. It should be noted, however, that the channel models derived from empirical data in the database and used in the previous examples all have *nearly* reflected roots, and still produce useful results in determining FSE length selection.

6.3 PAM CM Criterion

Despite the binary-signalling assumption used for simplification in deriving our proposed design rule, the preceding 16-QAM example suggests the analysis is robust to multi-level, complex signalling. We now show that the preceding analysis is easily extended to real, multi-level signalling. Presumably, further extension to consider complex signalling is straightforward.

From [43], the CM cost for a real, Non-Constant-Modulus (NCM), white, symmetric source can be expressed as

$$J_{CM|NCM} = \gamma^2 - 2\sigma_s^2\gamma \sum_{i=0}^{P} h_i^2 + \nu \sum_{i=0}^{P} h_i^4 + 3(\sigma_s^2)^2 \left(\sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} h_i^2 h_j^2 \right) \quad (6.39)$$

where $\sigma_s^2 = E\{s^2\}$, $\nu = E\{s^4\}$ and $\gamma = \nu/\sigma_s^2$ is the CM dispersion constant. Hence, the CM cost at a global minimum which achieves P.E. is

$$\begin{aligned} J_{CM|global\ min.} &= \gamma^2 - 2\sigma_s^2\gamma + \nu & (6.40) \\ &= \gamma^2 - \nu & (6.41) \end{aligned}$$

When the length condition is not satisfied, as from channel perturbations outside the FSE time span or from FSE taps lost in truncation as described in §6.1, collecting weighted versions of (6.17), (6.19), and (6.25), the CM cost changes from a P.E. setting to

$$\begin{aligned} J_{CM|length} = & & (6.42) \\ & \gamma^2 - 2\sigma_s^2\gamma \left(1 + 2p_\delta + p_\delta^2 + \sum_{i\neq\delta, i=0}^{P} p_i^2 \right) + \nu \left(1 + 6p_\delta^2 + 4p_\delta + 4p_\delta^3 + \sum_{i=0}^{P} p_i^4 \right) & \\ + & 3(\sigma_s^2)^2 \left(2(1+2p_\delta) \sum_{i=0, i\neq\delta}^{P} p_i^2 + \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2 \right) & \\ = & \left[\gamma^2 - 2\nu + \nu \right] + 4\left[\nu - \nu \right] p_\delta & (6.43) \end{aligned}$$

$$+ \left[(6\nu - 2\nu)p_\delta^2 + (6(\sigma_s^2)^2 - 2\nu)\sum_{i\neq\delta, i=0}^{P} p_i^2\right] + p_\delta \left[4\nu p_\delta^2 + 12(\sigma_s^2)^2 \sum_{i\neq\delta, i=0}^{P} p_i^2\right]$$

$$+ \left[\nu \sum_{i=0}^{P} p_i^4 + 3(\sigma_s^2)^2 \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2\right]$$

$$= \left[\gamma^2 - \nu\right] + \left[4\nu p_\delta^2 + (6(\sigma_s^2)^2 - 2\nu)\sum_{i\neq\delta, i=0}^{P} p_i^2\right] \tag{6.44}$$

$$+ p_\delta \left[4\nu p_\delta^2 + 12(\sigma_s^2)^2 \sum_{i\neq\delta, i=0}^{P} p_i^2\right] + \left[\nu \sum_{i=0}^{P} p_i^4 + 3(\sigma_s^2)^2 \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2\right]$$

Using (6.41) and (6.44), the change in CM cost due to undermodeling is therefore

$$\Delta J_{CM|NCM} = J_{CM|length} - J_{CM|global\ min.} \tag{6.45}$$

$$= \left[4\nu p_\delta^2 + (6(\sigma_s^2)^2 - 2\nu)\sum_{i\neq\delta, i=0}^{P} p_i^2\right] + p_\delta \left[4\nu p_\delta^2 + 12(\sigma_s^2)^2 \sum_{i\neq\delta, i=0}^{P} p_i^2\right] \tag{6.46}$$

$$+ \left[\nu \sum_{i=0}^{P} p_i^4 + 3(\sigma_s^2)^2 \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} p_i^2 p_j^2\right]$$

Note that (6.46) is grouped according to powers of the p_i. For the quadratic contribution, the NCM source weights the p_δ^2 element more heavily than the other p_i^2 elements, since $\nu \geq (\sigma_s^2)^2$. Note that the quadratic terms were equally weighted for the binary case in (6.26). For example, with a 4-PAM unit-variance constellation, $4\nu = 6.56$ while $(6(\sigma^2)^2 - 2\nu) = 2.72$. These factors should be related to the MSE cost incurred using the same analysis methods.

6.3.1 Relation to MSE

The MSE criterion is $J_{MSE} = E\{(y - s_\delta)^2\}$. For a unit variance source, the change in MSE from a P.E. setting due to undermodeling according to the approaches in §6.1 was shown in (6.29) to be $||\mathbf{h_p}||_2^2$. Hence, when the perturbation terms are "small", the change in CM cost in the vicinity of a global minimum is approximately a scaled version of the change in MSE cost

$$\Delta J_{CM|NCM} \approx (6-2\nu) \cdot \Delta J_{MSE} \tag{6.47}$$

This result generalizes that for the binary case in (6.30) where $(6-2\nu)=4$.

6.3.2 Undermodeling Examples

We wish to compare the proposed design rule for the database channels with that which may be suggested by the preceding PAM analysis. Hence, we evaluate (6.46) with a 16-PAM source for the two approaches of channel perturbations and equalizer truncation as described in §6.1 for channels 2 and 3 of the database. The channel taps for the approach of §6.1.1 are partitioned according to (6.2) and the partitioning of FSE coefficients for the approach of §6.1.2 is according to (6.9). The results are scaled (by a factor of $\frac{1}{(6-2\nu)}$) to approximate the MSE cost. The system delay is found according to (2.36). The results for channels 2 and 3 of the database are shown in Figures 6.6 and 6.7, respectively; the results for other channels of the database look similar.

Figures 6.6 and 6.7 each contain two plots. The top plot contains the magnitude of the $T/2$-spaced channel impulse response coefficients. The bottom plot contains the graphs of three functions: i) (solid) the approximation of MSE from a scaled version of (6.46) due to channel perturbations outside the FSE time span (§6.1.1), ii) (dashed) the approximation of MSE from a scaled version of (6.46) due to FSE truncation (§6.1.2), and iii) (dotted) the true MSE described by (6.29) according to the approach in §6.1.1–note that, unlike the BPSK case, this quantity is *not* necessarily a scaled version of the quadratic contribution of (6.46). The graphs may be referenced to the dashed line of constant MSE which corresponds to the

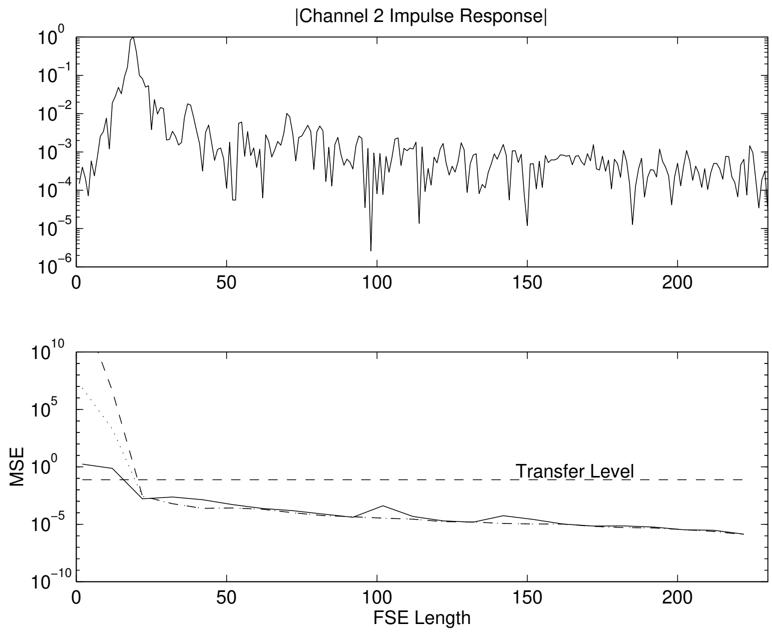


Figure 6.6: CM Robustness to Undermodeling, Channel 2, 16-PAM Signalling

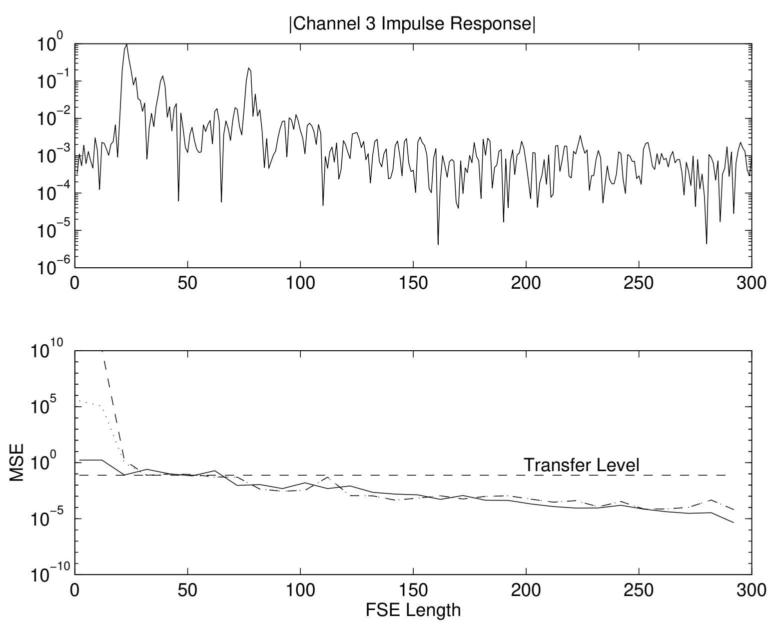


Figure 6.7: CM Robustness to Undermodeling, Channel 3, 16-PAM Signalling

Transfer Level MSE again for 16-QAM, unit variance signalling (see Appendix A).

These figures confirm the conclusions drawn from the binary analysis. Far fewer CMA-FSE taps are needed to reach the Transfer Level than for P.E. The "significant" portion of the channel appears to be consistent with what was found in the binary case–those coefficients greater than approximately $15-20\%$ of the magnitude of the largest channel tap. Though the CM minimum of optimum system delay does not have zero CM cost for a NCM, it still aligns closely with the corresponding Wiener solution provided the perturbations are "small."

6.3.3 Connection to Excess MSE/Misadjustment

The effect of the NCM source is not to disturb the CM criterion robustness properties in the vicinity of a local minimum. Rather, under the P.E. assumptions, the CM error surface sits above the MSE surface (by $\gamma^2 - \nu$), but their minima still precisely align. The deformation in the two surfaces due to undermodeling for a NCM source is similar to that for binary signalling (compare (6.26) and (6.46)) and a similar set of observations to §6.2.3 can be made for the PAM signalling case. The adaptive implementation to descend the CM error surface (CMA-FSE), however, typically uses a gradient descent approach with non-vanishing, but small, stepsize μ. A NCM source causes an excess MSE [1] from the equalizer update equation, since the stepsize is non-vanishing and the instantaneous CMA error is generally nonzero, effectively causing a "rattling around" in both the CM and MSE minima. It is this behavior that usually forces the transfer from CMA-FSE to DD-LMS for further MSE reduction. Note that LMS suffers excess MSE (and hence misadjustment) due to noise, but not due to a NCM source [36]. The excess

[1] Misadjustment is typically defined as the ratio of excess MSE to MMSE.

MSE of a CM receiver updated with CMA-FSE is approximated when the length condition is satisfied and in the absence of noise in [20]:

$$J_{MSE|X} = \lim_{k\to\infty} E\left\{(y(k) - s(k-\delta))^2\right\} \tag{6.48}$$

$$\approx \mu L_f \frac{\frac{E\{s^6\}}{(\sigma_s^2)^3} - \kappa_s}{(3-\kappa_s)} \sigma_s^2 \sigma_r^2 \tag{6.49}$$

where κ_s is the source kurtosis and σ_r^2 is the received signal power. The result mimics that for LMS in its dependence on the FSE length (see [36]), which suggests a classical compromise: the FSE length must be chosen long enough to cover the "significant" portion of the channel so that the undermodeling does not dominate the MSE of the CM receiver, but not too long so that the misadjustment dominates the MSE of the CM receiver. We approximate the MSE due to both violation of the length condition and excess MSE in the receiver implementing CMA-FSE from (6.46) and (6.49) as

$$J_{MSE} \approx \frac{1}{6-2\nu} \cdot \Delta J_{CM|NCM} + J_{MSE|X} \tag{6.50}$$

Figure 6.8 shows this approximation for Channel 2 of the database with a 4-PAM source and stepsizes of $\mu = 5\times 10^{-4}, 10^{-3}$, and 2×10^{-3}. This figure suggests that for a stepsize in the neighborhood of $\mu = 2\times 10^{-3}$, the proper choice of FSE length is much less than that needed for perfect equalization! In fact, this example shows that choosing the FSE length equal to the length of the channel in this case is precisely the wrong thing to do for this stepsize in attempting to reach the MSE Transfer Level.

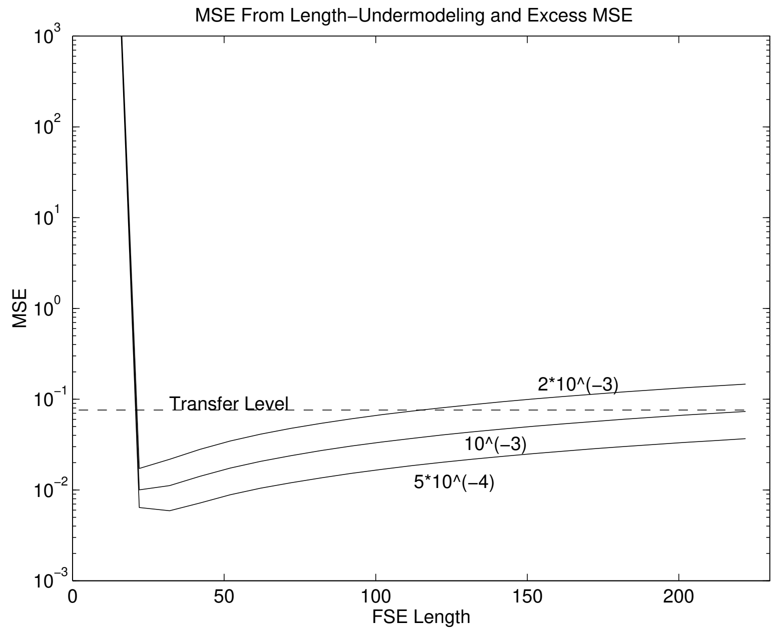


Figure 6.8: Undermodeling and Excess MSE of CM Receiver

6.4 A Bound on MSE of CM Receiver

The analysis approaches of §6.1 can be applied to the CM receiver power constraint first proposed for the baud-spaced binary case in [43] and generalized in [84]. A bound on the ℓ_2-norm of the portion of the combined channel-equalizer arising from violation of the length condition follows easily from this power constraint.

Suppose the length condition is violated according to the channel perturbation approach described in §6.1.1. Define $\mathbf{Q_m} := \mathbf{C_{m\downarrow}}^T\mathbf{C_{m\downarrow}}$. In the absence of noise, the CM receiver satisfies

$$\zeta \leq \mathbf{f_m}^T\mathbf{Q_m}\mathbf{f_m} \leq 1 \tag{6.51}$$

where ζ depends on the source kurtosis (see [84]). An alternative expression for

(6.51) can be written in terms of the minimum and maximum eigenvalues of $\mathbf{Q_m}$.

$$\lambda_{min}(\mathbf{Q_m})\mathbf{f_m}^T\mathbf{f_m} \leq \mathbf{f_m}^T\mathbf{Q_m}\mathbf{f_m} \leq \lambda_{max}(\mathbf{Q_m})\mathbf{f_m}^T\mathbf{f_m} \tag{6.52}$$

Together, (6.51) and (6.52) bound the ℓ_2-norm of the CM equalizer coefficients.

$$\frac{\zeta}{\lambda_{max}(\mathbf{Q_m})} \leq \mathbf{f_m}^T\mathbf{f_m} \leq \frac{1}{\lambda_{min}(\mathbf{Q_m})} \tag{6.53}$$

Define $\mathbf{Q_p} = \mathbf{C_{p\downarrow}}^T\mathbf{C_{p\downarrow}}$ so that a similar expression to (6.52) can be written for the portion of the combined channel-equalizer resulting from violation of the length condition, $\mathbf{C_{p\downarrow}f_m}$,

$$\lambda_{min}(\mathbf{Q_p})\mathbf{f_m}^T\mathbf{f_m} \leq \mathbf{f_m}^T\mathbf{Q_p}\mathbf{f_m} \leq \lambda_{max}(\mathbf{Q_p})\mathbf{f_m}^T\mathbf{f_m} \tag{6.54}$$

Substituting the bounds of the equalizer norm from (6.53) into (6.54) implies that

$$\zeta\frac{\lambda_{min}(\mathbf{Q_p})}{\lambda_{max}(\mathbf{Q_m})} \leq ||\mathbf{C_{p\downarrow}f_m}||_2^2 \leq \frac{\lambda_{max}(\mathbf{Q_p})}{\lambda_{min}(\mathbf{Q_m})} \tag{6.55}$$

These bounds offer some insight into the manifestation of error due to common subchannel roots. As the "modeled" channel loses disparity, $\lambda_{min}(\mathbf{Q_m})$ approaches zero and the upper bound approaches infinity. As the FSE length is increased to match the channel length, $\lambda_{min}(\mathbf{Q_p})$ goes to zero, which forces the lower bound to zero.

Similarly, when FSE taps are truncated as in §6.1.2, $\mathbf{f}^\ddagger$ is the CM equalizer vector associated with channel convolution matrix $\mathbf{C}_\downarrow$, so that with $\mathbf{Q} := \mathbf{C}_\downarrow{}^T\mathbf{C}_\downarrow$,

$$\mathbf{f}^{\ddagger T}\mathbf{f}^\ddagger \leq \frac{1}{\lambda_{min}(\mathbf{Q})} \tag{6.56}$$

Since $\tilde{\mathbf{f}}^T\tilde{\mathbf{f}} \leq \mathbf{f}^{\ddagger T}\mathbf{f}^\ddagger$, the MSE of the term resulting from FSE truncation satisfies

$$||\mathbf{C}_\downarrow\tilde{\mathbf{f}}||_2^2 \;\; \leq \;\; \frac{\lambda_{max}(\mathbf{Q})}{\lambda_{min}(\mathbf{Q})} \tag{6.57}$$

Note that the bound is precisely the condition number of the autocorrelation matrix of the baud-spaced received signal.

Though the bounds in (6.55) and (6.57) offer insight into error due to loss of channel disparity, and therefore may (or may not) merit further study, the bounds do not become tight when applied to the microwave channels from the database.

Chapter 7

CM Cost Function Approximation

The database examples illustrated in Chapter 6 are evaluated for system delay according to (2.36). We wish to study the CM error surface deformation for sub-optimum system delays without drawing an exhaustive set of pictures similar to Figures 6.2-6.4. We turn to simple calculus tools to address CM robustness properties at sub-optimum system delays. When the P.E. conditions are not satisfied, there exists no closed-form solution for the CM minima locations. This chapter derives an approximation of the location of the CM minima settings for the under-modeled case. The approximation results from a second-order approximation of the BPSK CM cost function, and from it, we deduce a measure that indicates the proximity of a CM minimum with a length-constrained Wiener solution. When evaluated for the database channels, this measure suggests that those CM minima corresponding to better MSE performance stay closer to their corresponding Wiener solutions than those CM minima corresponding to system delays which give poor MSE performance.

7.1 Gradient and Hessian Calculation

The previous algebraic analysis in Chapter 6 suggests strong connection between the CM and MSE minima. The fourth-order dependence of the CM cost function on the equalizer coefficient vector, however, prohibits an analytic, closed-form expression for the undermodeled CM minima. We therefore propose a second-order Taylor-series expansion of the CM cost function, for which a closed-form expression of the approximate CM minima can be found. A similar approach (though first-order approximation) exists for the baud-spaced case in [58]. From [73] the truncated (to second-order) Taylor series of a polynomial function g of variable x expanded about the point x_0 is

$$g(x) \;\; = \;\; g(x_0) + g'(x_0)\,(x - x_0) + \frac{1}{2}g''(x_0)\,(x - x_0)^2 \qquad (7.1)$$

where $(\cdot)'$ and $(\cdot)''$ denotes first and second derivatives of $g(x)$ with respect to x, respectively. Since we are interested in a continuous function, J_{CM}, as a function of a vector process, $\mathbf{f}$, we require a gradient vector and Hessian matrix with respect to $\mathbf{f}$. Fortunately, the tedious calculus has been done in [43] for the baud-spaced case, so we need only replicate their approach for the fractionally-spaced vector and matrix expressions. We next find compact descriptions of the gradient and Hessian expressions and then use them to write a second-order Taylor series of the binary CM cost function expanded about the length-L_f Wiener solution.

7.1.1 Gradient Vector

Let the $T/2$-spaced channel and equalizer vectors be described by

$$\mathbf{c} \;\; = \;\; [c_0 \; c_1 \; \ldots \; c_{L_c-1}] \qquad (7.2)$$

$$\mathbf{f} \;=\; [f_0\ f_1\ \ldots\ f_{L_f-1}] \tag{7.3}$$

respectively. The combined $T/2$-spaced channel-equalizer response, $\mathbf{h} = \mathbf{c} \otimes \mathbf{f}$, ($\otimes$ denotes convolution) has elements

$$h_j = \sum_{i=0}^{j} c_{j-i} f_i \tag{7.4}$$

where $j = 0, 1, 2, \ \ldots \ P-1, \ P = L_f + L_c - 1$.

The CM cost function in (2.44) for binary, noiseless, symmetric signalling can be expanded in terms of the elements of $\mathbf{h}$ and written as (see [43])

$$J_{CM} = 1 - 2 \sum_{i=0,i-even}^{P-1} h_i^2 + 3 \left(\sum_{i=0,i-even}^{P-1} \ \sum_{j=0,j \neq i,j-even}^{P-1} h_i^2 h_j^2 \right) + \sum_{i=0,i-even}^{P-1} h_i^4 \tag{7.5}$$

where without loss of generality we have assumed that even coefficients correspond to "on-baud" signalling. We wish to find the gradient vector of (7.5) with respect to the equalizer coefficients,

$$\nabla_{\mathbf{f}}(J_{CM}) \;:=\; \begin{bmatrix} \left(\frac{\partial J_{CM}}{\partial f_0}\right) \\ \left(\frac{\partial J_{CM}}{\partial f_1}\right) \\ \left(\frac{\partial J_{CM}}{\partial f_2}\right) \\ \vdots \\ \left(\frac{\partial J_{CM}}{\partial f_{L_f-1}}\right) \end{bmatrix} \tag{7.6}$$

The necessary derivatives to assemble the gradient vector are calculated for the terms comprising (7.5).

$$\frac{\partial (\sum_{j=0,j-even}^{P-1} h_j^4)}{\partial f_i} \;=\; 4 \sum_{j=0,j-even}^{P-1} h_j^3 c_{j-i} \tag{7.7}$$

$$\;=\; 4 \sum_{j=i,j-even}^{L_c-1+i} h_j^3 c_{j-i} \tag{7.8}$$

$$\frac{\partial(\sum_{j=0,j-even}^{P-1} h_j^2)}{\partial f_i} \;=\; 2\sum_{j=i,j-even}^{L_c-1+i} h_j c_{j-i} \tag{7.9}$$

$$\frac{\partial(\sum_{j=0,j-even}^{P-1} \sum_{l=0,l\neq j,l-even}^{P-1} h_j^2 h_l^2)}{\partial f_i} \;=\; 4\sum_{l=0,l-even}^{P-1} h_l^2 \sum_{j=i,j\neq l,j-even}^{L_c-1+i} h_j c_{j-i} \tag{7.10}$$

Collecting terms (7.8), (7.9) and (7.10), the elements of the gradient vector may be written as

$$\partial J_{CM}/\partial f_i \;=\; \sum_{j=i,j-even}^{L_c-1+i} \left[12 \sum_{l=0,l-even}^{P-1} h_l^2 - 4 - 8h_j^2\right] h_j c_{j-i} \tag{7.11}$$

Define $\Lambda_j := \left(12\sum_{l=0,l-even}^{P-1} h_l^2 - 4 - 8h_j^2\right)$, so that

$$\partial J_{CM}/\partial f_i = \sum_{j=i,j-even}^{L_c-1+i} \Lambda_j h_j c_{j-i} \tag{7.12}$$

Now concatenate (7.12) for each $i = 0,\ 1,\ 2,\ \ldots\ L_f - 1$ to form the gradient vector:

$$\nabla_{\mathbf{f}}(J_{CM}) \;=\; \left[\begin{array}{ccccccccc}
c_0 & c_2 & c_4 & \ldots & c_{L_c-1} & & & & \\
 & c_1 & c_3 & c_5 & \ldots & c_{L_c-1} & & & \\
 & c_0 & c_2 & c_4 & \ldots & c_{L_c-1} & & & \\
 & & c_1 & c_3 & c_5 & \ldots & c_{L_c-1} & & \\
 & & c_0 & c_2 & c_4 & \ldots & c_{L_c-1} & & \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\
 & & & & & c_{1,0} & c_{3,2} & \ldots & c_{L_c-1}
\end{array}\right]
\left[\begin{array}{c}
\Lambda_0 h_0 \\ \Lambda_2 h_2 \\ \Lambda_4 h_4 \\ \vdots \\ \Lambda_{P-1} h_{P-1}
\end{array}\right]$$

Observe that the matrix of channel coefficients is simply the transpose of the row-decimated channel convolution matrix described in (2.22). Hence, call this matrix $\mathbf{C}_{\downarrow}{}^{T}$. Further, define Λ as the $\lceil P/2 \rceil \times \lceil P/2 \rceil$ diagonal matrix

$$\Lambda \;:=\; diag([\Lambda_0\ \Lambda_2\ \ldots\ \Lambda_{P-1}]^T) \tag{7.13}$$

where $diag(\cdot): \mathcal{C}^n \mapsto \mathcal{C}^{n\times n}$ by placing the argument vector on the main diagonal of the resulting matrix which otherwise has all zero entries. We now compactly represent the gradient vector as

$$\nabla_{\mathbf{f}}(J_{CM}) = \mathbf{C}_{\downarrow}{}^{T}\Lambda\mathbf{h}_{\downarrow} \tag{7.14}$$

where $\mathbf{h}_{\downarrow}$ is the decimated (baud-spaced) combined channel equalizer response.

7.1.2 Hessian Matrix

We seek a similar compact expression to (7.14) for the necessary second derivatives. Define the Hessian as the $(L_f-1)\times(L_f-1)$ matrix whose i,j^{th} element is $\partial^2 J_{CM}/\partial f_i \partial f_j$,

$$\mathcal{H}_{\mathbf{f}}(J_{CM}) \;\; := \begin{bmatrix} \frac{\partial^2 J_{CM}}{\partial f_0 \partial f_0} & \frac{\partial^2 J_{CM}}{\partial f_0 \partial f_1} & \cdots & \frac{\partial^2 J_{CM}}{\partial f_0 \partial f_{L_f-1}} \\ \frac{\partial^2 J_{CM}}{\partial f_1 \partial f_0} & \frac{\partial^2 J_{CM}}{\partial f_1 \partial f_1} & \cdots & \frac{\partial^2 J_{CM}}{\partial f_1 \partial f_{L_f-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J_{CM}}{\partial f_{L_f-1} \partial f_0} & \frac{\partial^2 J_{CM}}{\partial f_{L_f-1} \partial f_1} & \cdots & \frac{\partial^2 J_{CM}}{\partial f_{L_f-1} \partial f_{L_f-1}} \end{bmatrix} \tag{7.15}$$

The necessary derivatives are found in an analogous way to the derivation of the gradient. When $i=j$, the main diagonal terms are

$$\begin{aligned} \frac{\partial^2 J_{CM}}{\partial f_i^2} &= \\ & \sum_{j=i,j-even}^{L_c-1+i} \frac{\partial}{\partial f_i}\left(-4h_j + 12h_j \sum_{l=0,l-even}^{P-1} h_l^2 - 8h_j^3\right) c_{j-i} \\ &= \sum_{j=i,j-even}^{L_c-1+i}\left(-4c_{j-i} + 12\frac{\partial h_j}{\partial c_i}\sum_{l=0,l-even}^{P-1} h_l^2 + 12h_j \sum_{l=0,l-even}^{P-1}\frac{\partial}{\partial f_i}h_l^2 - 24h_j^2 c_{j-i}\right) c_{j-i} \\ &= \sum_{j=i,j-even}^{L_c-1+i}\left(-4c_{j-i}^2 + 12c_{j-i}^2 \sum_{l=0,l-even}^{P-1} h_l^2 + 24h_j c_{j-i}\sum_{l=i,l-even}^{L_c-1+i} h_l c_{l-i} - 24h_j^2 c_{j-i}^2\right) \end{aligned} \tag{7.16}$$

The off-diagonal terms ($i \neq j$) are found by

$$\frac{\partial^2 J_{CM}}{\partial f_i \partial f_m} = \sum_{j=i,j-even}^{L_c-1+i} \tag{7.17}$$

$$-4c_{j-i}c_{j-m} + 12c_{j-i}c_{j-m} \sum_{l=0,l-even}^{P-1} h_l^2 + 24h_j c_{j-i} \sum_{l=i,l-even}^{L_c-1+i} h_l c_{l-m} - 24h_j^2 c_{j-i}c_{j-m}$$

Collecting diagonal and off-diagonal terms and putting into a matrix expression, the Hessian is expressed as

$$\mathcal{H}_{\mathbf{f}}(J_{CM}) = \mathbf{C}_{\downarrow}{}^T \Psi \mathbf{C}_{\downarrow} \tag{7.18}$$

where

$$\Psi = \left(12 \sum_{l=0,l-even}^{P-1} h_l^2 - 4\right) \mathbf{I}_{\lceil P/2 \rceil} + 24\mathbf{h}_{\downarrow}\mathbf{h}_{\downarrow}^{\mathbf{t}} - 24diag([h_0^2\ h_2^2\ h_4^2 \ldots h_{P-1}^2]^T)$$

7.2 Taylor Expansion of CM Cost Function

Now that we have compact descriptions of the gradient and Hessian expressions, (7.14) and (7.18), we propose a second-order approximation of the CM cost function based on Taylor's Theorem (see [73]).

$$\hat{J}_{CM} = \tag{7.19}$$

$$J_{CM|\mathbf{f}^\dagger} + \nabla_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger}\left(\mathbf{f} - \mathbf{f}^\dagger\right) + \frac{1}{2}\left(\mathbf{f} - \mathbf{f}^\dagger\right)^T \mathcal{H}_{\mathbf{f}}^T(J_{CM})_{|\mathbf{f}^\dagger}\left(\mathbf{f} - \mathbf{f}^\dagger\right)$$

$$\approx J_{CM} \tag{7.20}$$

Though no closed-form solution exists for the minima settings of J_{CM} when the FSE is shorter than that needed for P.E., our approximation $\hat{J}_{CM}$ in (7.20) is quadratic in $\mathbf{f}$, so that a closed-form solution exists for the minimum setting of $\hat{J}_{CM}$. By choosing $\mathbf{f}^\dagger$ equal to the length-constrained Wiener solutions found according to (2.25), solving

$$\hat{\mathbf{f}}_{CM} \quad = \quad \arg\min_{\mathbf{f}} \left\{ \hat{J}_{CM} \right\} \tag{7.21}$$

provides an approximation of the CM minima settings in the neighborhood of the Wiener solutions corresponding to various system delays.

We solve (7.21) by straightforward calculus.

$$\frac{\partial \hat{J}_{CM}}{\partial \mathbf{f}} \quad = \quad \nabla_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger} + \frac{1}{2}\left(\mathcal{H}_{\mathbf{f}}^T(J_{CM})_{|\mathbf{f}^\dagger} + \mathcal{H}_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger}\right)\left(\mathbf{f} - \mathbf{f}^\dagger\right) \tag{7.22}$$

Setting the derivative equal to zero and recognizing that the Hessian is symmetric, the CM minima setting corresponding to system delay δ is approximated by

$$\hat{\mathbf{f}}_{CM} \quad = \quad \mathbf{f}^\dagger - \left[\mathcal{H}_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger}\right]^{-1} \cdot \nabla_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger} \tag{7.23}$$

where $\mathbf{f}^\dagger$ is the Wiener solution corresponding to system delay δ.

This result approximates the CM minima settings as perturbations to the Wiener settings by the perturbation vector $[\mathcal{H}_{\mathbf{f}}(J_{CM})]^{-1} \cdot \nabla_{\mathbf{f}}(J_{CM})$ evaluated at the Wiener settings. Observe that when the length condition is satisfied, the CM and Wiener settings coincide, so that the perturbation vector should contain all zeros. For this P.E. case, the gradient vector is zero (since $\Lambda_{j=\delta} = 0$ and $h_{j\neq\delta} = 0$) and the Hessian reduces to $8{\mathbf{C}_\downarrow}^T\mathbf{C}_\downarrow$ (since Ψ reduces to $8\mathbf{I}_{\lceil P/2 \rceil}$). Since $\mathbf{C}_\downarrow$ is assumed full-column rank (there are no reflected channel roots,) our perturbation vector is well conditioned, so that our approximation equals the Wiener solution under P.E.

7.3 A Measure of the Proximity of CM and Wiener Minima

Though error bounds may be determined for our simple approximation, we propose rather to treat the perturbation vector as an indicator of the proximity between CM and Wiener settings when undermodeled. Hence, we propose as a measure[1] of this proximity the squared ℓ_2-norm of the perturbation vector,

$$\rho \quad = \quad \left\| \left[\mathcal{H}_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger} \right]^{-1} \cdot \nabla_{\mathbf{f}}(J_{CM})_{|\mathbf{f}^\dagger} \right\|_2^2 \tag{7.24}$$

The results of Chapter 6 suggest that when the channel coefficients outside the time span of the FSE are "small", the CM minimum corresponding to optimum system delay stays in a tight neighborhood around the corresponding Wiener solution. We use the measure proposed in (7.24) to study the effect of system delay on the proximity of CM and Wiener solutions for the undermodeled case. We evaluate the MMSE achievable (that associated with $\mathbf{f}^\dagger$) and ρ in (7.24) for Channels 4 and 3 from the database for various FSE lengths over all possible system delays. Figures 7.1-7.4 each contain three plots (and are similar to Figures 2.5 and 2.6). The top plot is the magnitude of the $T/2$-spaced channel coefficients. The middle plot shows the MMSE achievable versus system delay, and the bottom plot shows ρ for the possible system delays.

Observe from these figures that the system delays corresponding to the "trough" of good MSE performance are precisely the same system delays where ρ is relatively "small." Hence, the CM minima corresponding to system delays which give better MSE performance stay closer to their corresponding Wiener solutions than the CM minima corresponding to system delays which give poorer MSE performance.

[1]It is shown in [5] that the (squared) ℓ_2-norm induces a measure.

These figures also confirm the conclusions of the previous chapter–for the optimum system delay, the CM minimum stays in a tight neighborhood around the Wiener solution, since the value of ρ assumed in the "trough" is "small." These results, and those from Chapter 6, suggest that the CMA-FSE length can be chosen substantially less than that needed for P.E. and that "proper" initialization is likely to find a local minima setting which achieves tolerable MSE performance in terms of a prescribed Transfer Level.

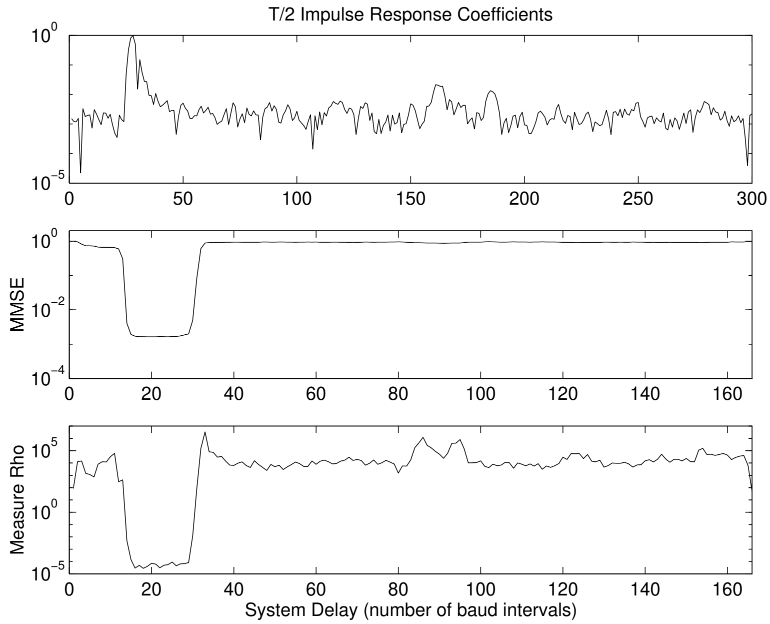


Figure 7.1: MMSE and ρ Sensitivity to System Delay, Channel 4, $L_f = 32$

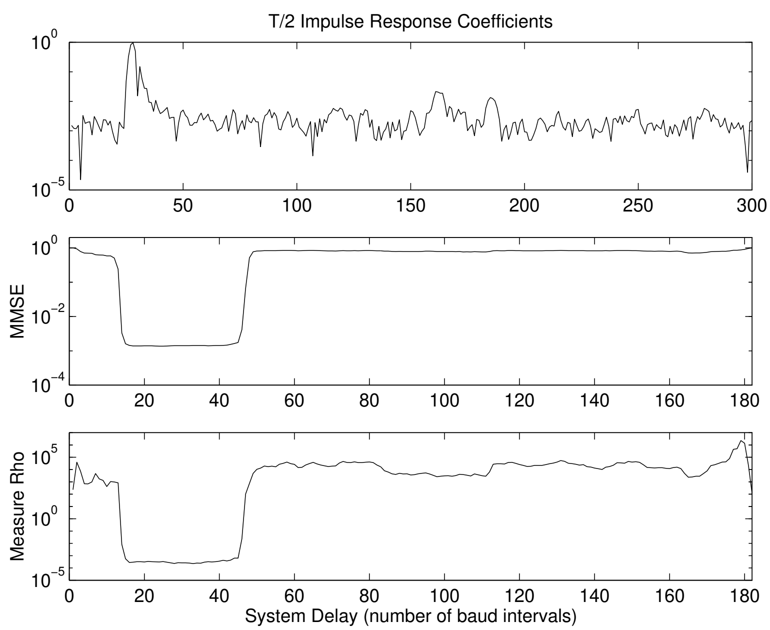


Figure 7.2: MMSE and ρ Sensitivity to System Delay, Channel 4, $L_f = 64$

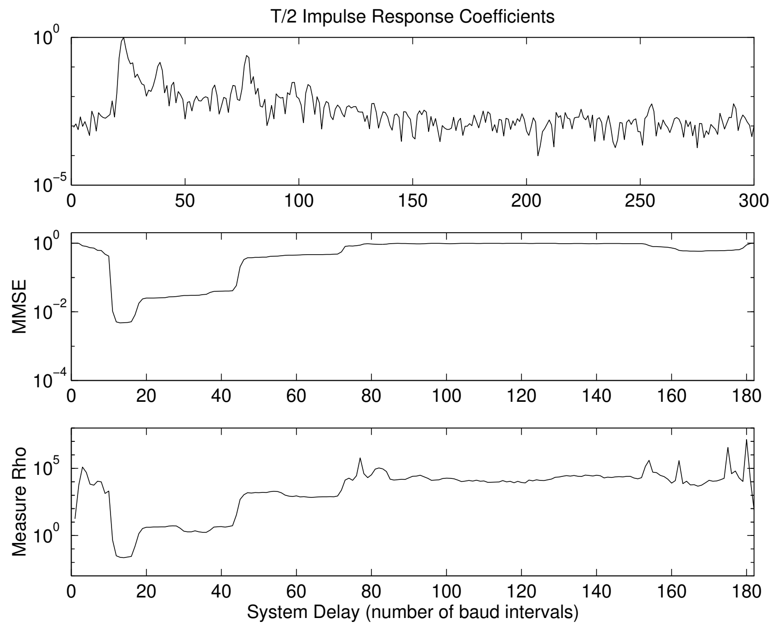


Figure 7.3: MMSE and ρ Sensitivity to System Delay, Channel 3, $L_f = 64$

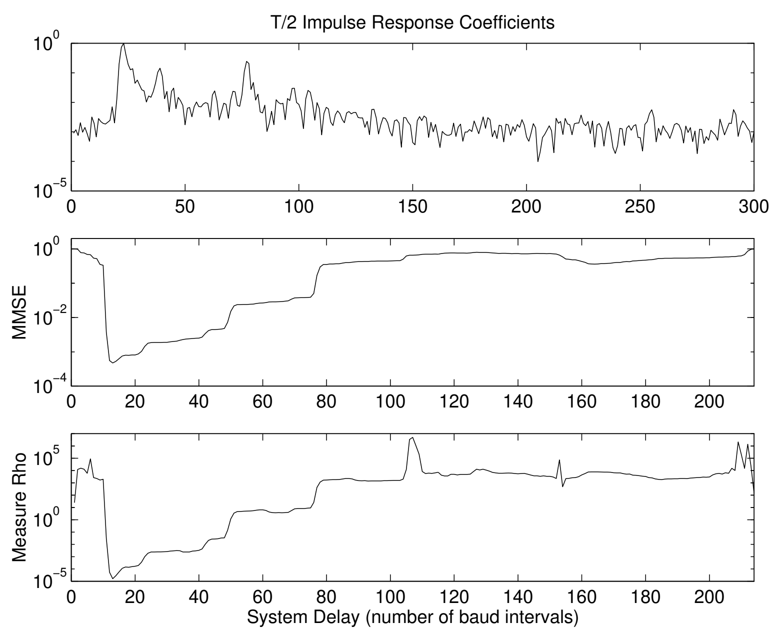


Figure 7.4: MMSE and ρ Sensitivity to System Delay, Channel 3, $L_f = 128$

Chapter 8

Conclusion

8.1 Concluding Remarks

This dissertation has investigated the blind, cold-start equalization of a closed-eye channel so that the error rate is sufficiently reduced in order that a decision directed mode can be employed. We have studied a class of recently introduced algorithms based purely on the SOS of the received signal which are receiving a plethora of attention in the current open literature. Though these algorithms promise convergence benefits compared to higher-order-statistics methods, we have shown that they lack robustness properties which make their implementation questionable for the problem of interest.

Based on these shortcomings, we turned our attention to the older, widely used CMA which accumulates a third-order moment of the received signal. Though this algorithm is computationally simple and easily implemented, its inherent decoupling of magnitude and phase result in a multi-modal cost surface which has kept it from reaching the analytic maturity enjoyed by other algorithms of its age (for example, LMS). Still, the recent literature (past few years) has seen some key results from fellow researchers in understanding CMA's behavior. A literature search, however, has shown a hole in this behavior theory for the case when the FSE length is less than that needed for P.E. Note that this undermodeled case is precisely the scenario in which the SOS algorithms display gross robustness concerns, and precisely the practical case for the high data rates of interest.

We have proposed original algebraic analysis of the CM cost function which describes the deformation of the surface due to undermodeling. Some straightforward calculus led to a truncated Taylor series expansion of the binary CM cost function which resulted in a measure to indicate the proximity of the CM minima and Wiener solutions of various system delays. Our analysis suggests that when the channel coefficients outside the FSE time span are "small" and the delay is chosen (presumably by FSE initialization) close to the optimum system delay, the CM minimum stays in a tight neighborhood about the Wiener solution. Recognizing that the stochastic gradient implementation (CMA) of the CM cost surface suffers an excess MSE term due to a multi-modulus source, we turned our analysis into an approach to infer a design guideline for FSE length selection given the specifics of the application (signalling scheme, channel dynamics, etc.). We evaluated our approach for the undeniably practical channel models in the database, which resulted in useful design rules. Note that Appendix A applies these analysis tools in a case study for a cable TV channel which again produces a useful set of design guidelines.

Together, the analysis in Chapters 6 and 7 with the existing results in Chapter 5 can be used to form a cohesive, pedagogical behavior theory for the CM criterion. This theory suggests that the blind receiver implementing the CM criterion can achieve nearly the performance of a trained or non-blind receiver, and attests to CMA's widespread application.

8.2 Future Work

Though our work is illuminating in understanding CMA's robustness properties when the length condition is not satisfied, many of the results can be extended.

The algebraic analysis in Chapter 6 assumes real signalling. Though somewhat impractical, this assumption simplifies the mathematics while still producing meaningful results. Presumably, the extension to the complex case is straightforward (albeit tedious) and is left as SEP. Note that this algebraic method does not easily lend itself to noisy signalling, since the approach hinges on decomposing $h_i = m_i + p_i$, where the m_i achieve P.E. However, the misadjustment formula proposed in [20] and used in Chapter 6 can be extended to noisy signalling. This analysis is in preparation, but not included in this dissertation.

The calculus approach in Chapter 7 starts with the noiseless, binary CM cost function. Fellow graduate students have expanded the CM cost function for noisy, real signalling of arbitrary source statistics in [67]. Presumably, compact gradient and Hessian expressions can be derived from their expression and the results of Chapter 7 extended. This calculus approach might help to answer the following questions:

1. Can the error in our approximation of the CM minima in (7.23) be bounded?

2. Can a higher than second-order expansion be realized and better, more-illuminating approximations to the CM minima found than (7.23)?

3. What is the relationship between our result for the CM minima assuming binary signalling in (7.23) and the result to be found for multi-level signalling?

4. Can these calculus-based results be extended to include noise?

5. Can a classification of CM stationary points be done when undermodeled, as was done in [43] for the baud-spaced case?

The directions of future work depend on the specifics of the application. Appendix A applies our existing analysis to a Cable TV application.

Appendix A

Case Study: Cable TV

Throughout this dissertation, we have consistently used the database channel models, which are derived from empirically measured digital microwave radio signals, and hence are undeniably practical. This appendix strays from these database channels and considers an empirically measured cable TV channel model. This channel model, designated as Cable TV Channel A, and a second cable TV channel model, are available at the database (see §3.2) as binary, MATLAB-readable files, courtesy of Applied Signal Technology (Sunnyvale, CA). The goal of this appendix is to show that the analysis methods previously discussed in Chapters 6 and 7 can be extended to other channel classes and other signalling schemes to produce usable design guidelines for FSE length selection in a blind receiver using CMA from a cold start.

As previously mentioned, the Transfer Level MSE depends on the signalling scheme used, and practitioners usually define this threshold based on the SER. According to [86], a typical Transfer Level is the MSE corresponding to an SER of approximately $10^{-1.5}$, or between 10^{-1} and 10^{-2}. Figure A.1 shows SER curves for symmetric QAM signalling schemes assuming a unit-variance source, and the dashed line is our interpretation of the prescribed SER threshold. Based on these curves, the corresponding Transfer Level MSE's are computed and tabulated in Table A.1.

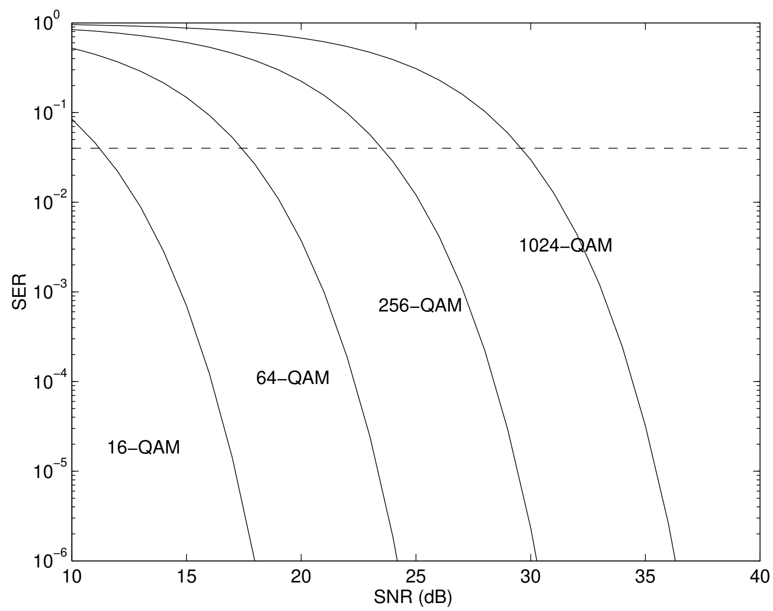


Figure A.1: SER Curves for Symmetric QAM Signalling

Instead of the 16-QAM signalling assumed in most of the preceding examples, we assume a more dense 256-QAM signalling format for Cable TV Channel A. We first evaluate the MMSE achievable from the Wiener solution in (2.25) with system delay found according to (2.36) for various FSE lengths (see Figure A.2). This figure is analogous to Figure 2.7.

Figure A.2 suggests that for modest noise levels (our undermodeling robustness treatment will assume 100 dB SNR) there exists a length-constrained equalization setting which achieves the prescribed Transfer Level for 256-QAM. The goal is the design of a blind FSE using CMA which does not satisfy the length condition

Table A.1: QAM Signalling and Corresponding MSE Transfer Levels

Signalling Format	MSE Transfer Level
16-QAM	0.076
64-QAM	0.0182
256-QAM	0.0045
1024-QAM	0.0011

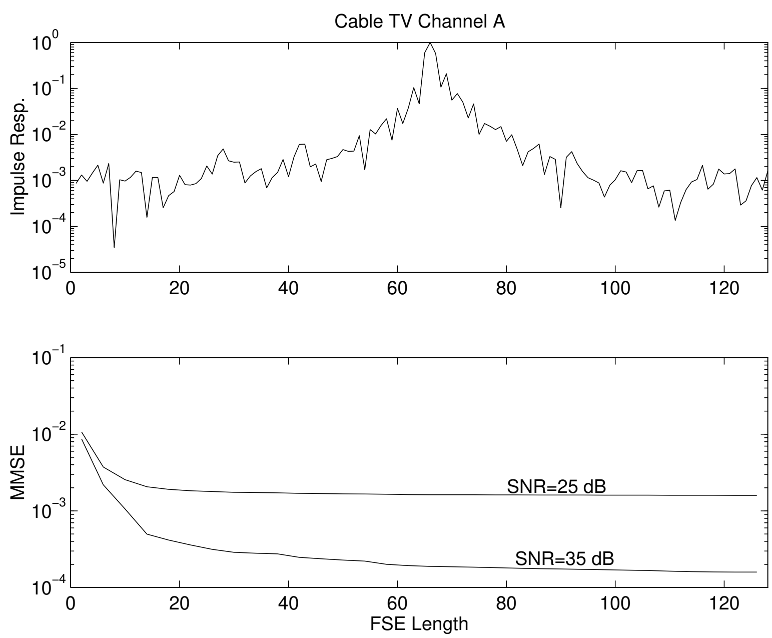


Figure A.2: Cable TV Channel A: MMSE versus FSE Length needed for P.E., but still achieves the prescribed Transfer Level. Hence, we approximate the MSE of the undermodeled CM receiver according to (6.47), where the statistics for the NCM source correspond to 256-QAM, white, unit variance, symmetric signalling. In evaluating (6.47), we use the analysis approaches described in §6.1. For the FSE truncation approach of §6.1.2, the FSE coefficients are partitioned according to (6.9). However, for the channel perturbation approach of §6.1.1, we recognize that Cable TV Channel A exhibits a significant precursor (the channel coefficient of largest magnitude is approximately at the center of the channel impulse response). Rather than considering the channel perturbation as appended channel coefficients with largest delay as is done in (6.2), we introduce a partitioning of the channel coefficients such that the "modeled" channel is the L_f ($L_f < L_c$) center coefficients of the length-L_c cable TV channel to reflect the significant precursor,

$$\begin{aligned} \mathbf{c_m} &:= [\underbrace{0\ 0 \ldots 0}_{P\ zeros}\ \underbrace{c_P \ldots c_{P+L_f}}_{L_f\ coefficients}\ \underbrace{0\ 0 \ldots 0}_{P\ zeros}]^T \\ \mathbf{c_p} &:= [c_0 \ldots c_{P-1}\ \underbrace{0\ 0 \ldots 0}_{L_f\ zeros}\ c_{P+L_f+1} \ldots c_{L_c-1}]^T \end{aligned} \qquad \text{(A.1)}$$

where $P = \lceil \frac{L_c - L_f}{2} \rceil$. With these definitions, $\mathbf{c} = \mathbf{c_m} + \mathbf{c_p}$ and Figure A.3 is analogous to Figures 6.6 and 6.7.

Our accounting for the channel precursor in (A.1) is evident in Figure A.3; the channel perturbation approach gives better results than the FSE truncation approach. For an FSE length greater than about 20 coefficients, the MSE incurred with a 256-QAM source by the CM cost function due to undermodeling is "small", and essentially equal to the MSE cost incurred by the MSE cost function due to undermodeling. This behavior suggests a small deformation in the CM error

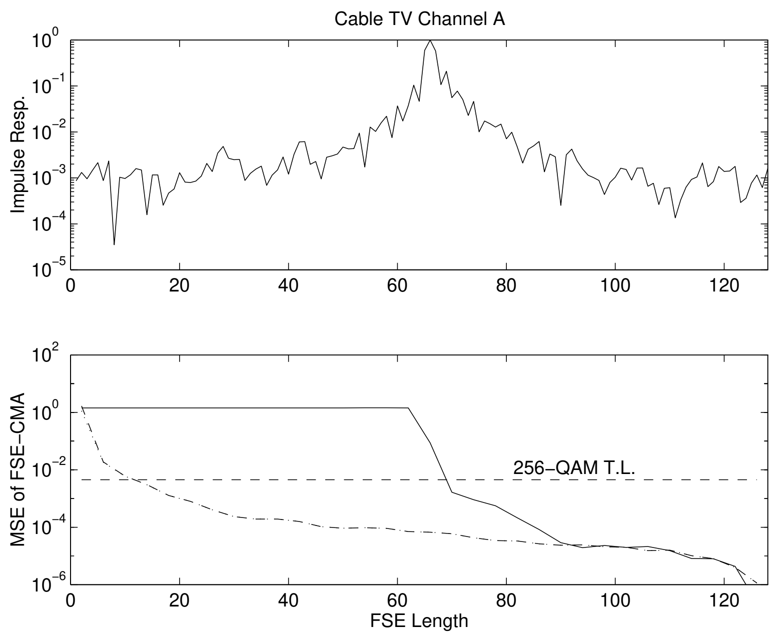


Figure A.3: Cable TV Channel A: MSE of CM Receiver versus FSE Length, 256-QAM surface, and that the CM minimum stays in a tight neighborhood of the Wiener solution. The "significant" portion of the channel for this signalling threshold and channel class seems to be those channel coefficients whose magnitudes are greater than approximately $2-5\%$ of the magnitude of the largest channel coefficient. For the database channels with 16-QAM signalling, we previously concluded that "significant" meant $15-20\%$ relative magnitudes.

The blind receiver uses the stochastic gradient descent implementation in (2.46) to transcend the CM error surface. With 256-QAM signalling, we expect severe misadjustment due to the multi-modulus source. We again use the approximation of excess MSE in [20] to approximate the MSE of CMA-FSE when undermodeled with a multi-modulus source as in (6.50). Figure A.4 is analogous to Figure 6.8, with stepsizes $\mu = 10^{-6}, 10^{-5}$, and 10^{-4}. Figure A.4 suggests a range of FSE

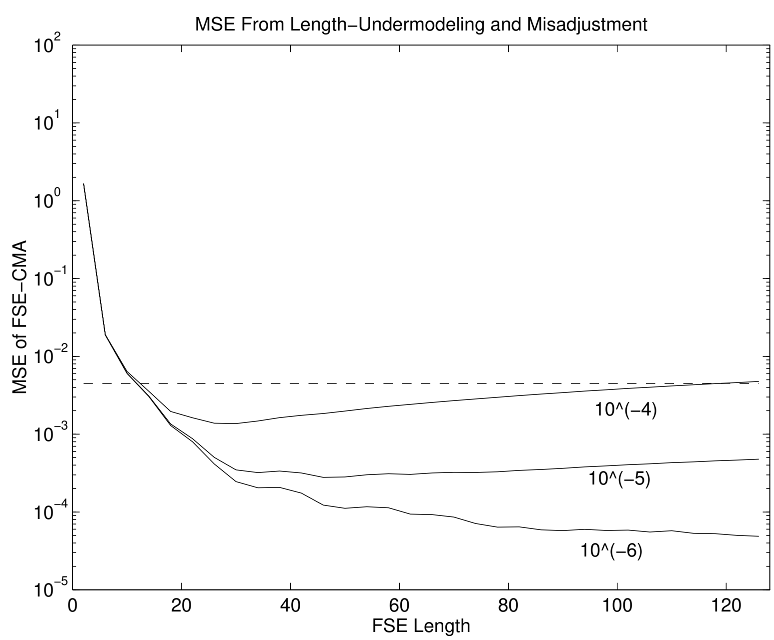


Figure A.4: Cable TV Channel A: MSE Due to Undermodeling and Misadjustment lengths and stepsizes which tradeoff between ISI removal and excess MSE, but still achieve the prescribed MSE Transfer Level. Based on these results, we choose an FSE length of $L_f = 32$, and a stepsize of $\mu = 5 \times 10^{-5}$.

Admittedly, thus far we have considered only the performance of the CM minimum corresponding to optimum system delay. To study the effects of system delay on performance of the CM minima locations, we evaluate the measure proposed in (7.24) based on our Taylor series expansion of the binary CM cost function. Figure A.5 is analogous to Figures 7.1-7.4 and is evaluated for a 32-tap FSE. The results agree with the conclusions in Chapter 7; the CM minima corresponding to better MSE performance stay closer to their corresponding Wiener solutions than those CM minima which correspond to system delay of poorer MSE performance. We choose an FSE initialization corresponding to a unity spike at the 16^{th} index,

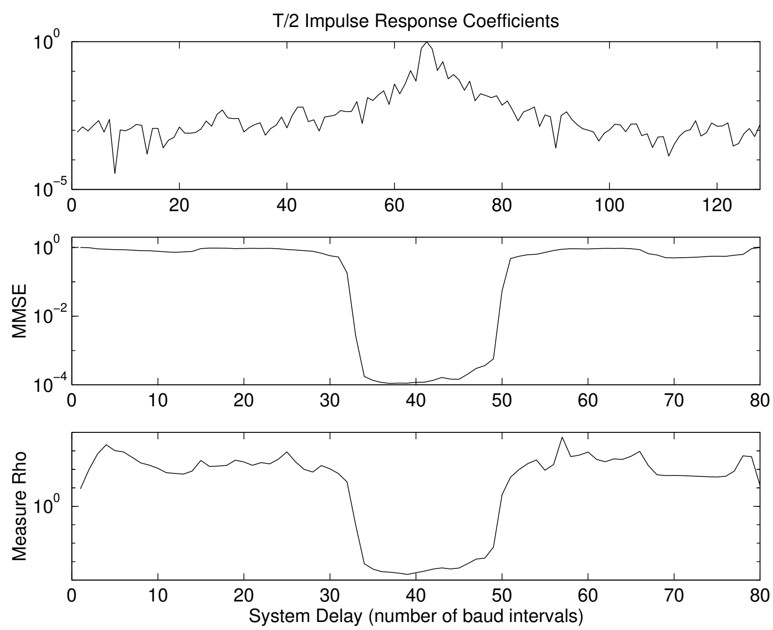


Figure A.5: Cable TV Channel A: MMSE and ρ versus System Delay and all other taps set to zero. The motivation for this initialization reflects the following reasoning. We estimate the cursor of the channel at the 66^{th} $T/2$-spaced position, so that the system delay in baud intervals at the start of adaptation is $\delta = \frac{66+16}{2} = 41$. Admittedly, there is no guarantee that the system delay will not wander during the adaptation process, but we at least start it in the trough described in Figure A.5.

Our design variables are now chosen based on our analysis: 32-tap FSE, $\mu = 5 \times 10^{-5}$, and a unity-spike initialization at 16^{th} position. A source sequence of 120,000 $T/2$-spaced samples drawn from a white 256-QAM alphabet is filtered through Cable TV Channel A, producing a received sequence for FSE-CMA. The MSE trajectory of our blind receiver is shown in Figure A.6, which demonstrates acceptable MSE reduction for transfer to a DD mode.

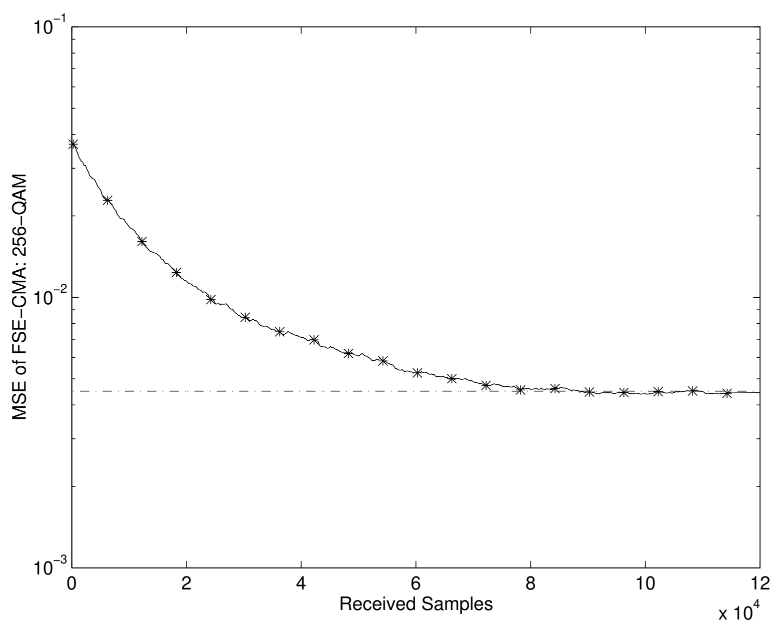


Figure A.6: Cable TV Channel A: MSE Trajectory of CMA-FSE

Appendix B

Polarity Equivalence of Algebraic Analysis

This appendix shows that for the binary algebraic analysis in Chapter 6 which considers $m_\delta = 1$, our result in (6.26) is valid also for the case with opposite polarity ($m_\delta = -1$).

We introduce the notation $\check{(\cdot)}$ to denote the system of "negative" polarity. For example, for the channel perturbation approach of §6.1.1,

$$\begin{aligned} \check{\mathbf{h}}_\downarrow &= \mathbf{C}_{\mathbf{m}\downarrow}\check{\mathbf{f}}_\mathbf{m} + \mathbf{C}_{\mathbf{p}\downarrow}\check{\mathbf{f}}_\mathbf{m} && \text{(B.1)} \\ &= \check{\mathbf{h}}_\mathbf{m} + \check{\mathbf{h}}_\mathbf{p} && \text{(B.2)} \end{aligned}$$

is analogous to (6.7), and for the equalizer truncation approach of §6.1.2,

$$\begin{aligned} \check{\mathbf{h}}_\downarrow &= \mathbf{C}_\downarrow\check{\mathbf{f}}^\ddagger + \mathbf{C}_\downarrow\check{\check{\mathbf{f}}} && \text{(B.3)} \\ &= \check{\mathbf{h}}_\mathbf{m} + \check{\mathbf{h}}_\mathbf{p} && \text{(B.4)} \end{aligned}$$

is analogous to (6.12). Hence, $\check{m}_i \in \mathbf{C}_{\mathbf{m}\downarrow}\check{\mathbf{f}}_\mathbf{m}$ or $\check{m}_i \in \mathbf{C}_\downarrow\check{\mathbf{f}}^\ddagger$, and $\check{p}_i \in \mathbf{C}_{\mathbf{p}\downarrow}\check{\mathbf{f}}_\mathbf{m}$ or $\check{p}_i \in \mathbf{C}_\downarrow\check{\check{\mathbf{f}}}$, depending on the analysis approach (§6.1.1 or §6.1.2), and $\check{h}_i = \check{m}_i + \check{p}_i$.

We now derive a result analogous to (6.26), but assume $\check{m}_\delta = -1$, where the result in (6.26) assumes that $m_\delta = +1$. Consider the three terms of the CM cost function in (6.13) separately

$$\begin{aligned}
1 - 2\sum_{i=0}^{P} \check{h}_i^2 &= 1 - 2\sum_{i=0}^{P}(\check{m}_i + \check{p}_i)^2 && \text{(B.5)}\\
&= 1 - 2\sum_{i=0}^{P}(\check{m}_i^2 + 2\check{m}_i\check{p}_i + \check{p}_i^2) && \text{(B.6)}\\
&= 1 - 2(1 - 2\check{p}_\delta + \check{p}_\delta^2) - 2\sum_{i\neq\delta, i=0}^{P} \check{p}_i^2 && \text{(B.7)}\\
&= -1 + 2(2\check{p}_\delta - \check{p}_\delta^2) - 2\sum_{i\neq\delta, i=0}^{P} \check{p}_i^2 && \text{(B.8)}
\end{aligned}$$

since the “modeled” part of the channel achieves P.E. with $\check{m}_\delta = -1$ and $\check{m}_i = 0 \ \forall \ i \neq \delta$.

Similarly, express

$$\begin{aligned}
\sum_{i=0}^{P} \check{h}_i^4 &= \sum_{i=0}^{P}(\check{m}_i + \check{p}_i)^4 \\
&= \sum_{i=0}^{P} \check{m}_i^4 + 4\check{m}_i^3 p_i + 6\check{m}_i^2\check{p}_i^2 + 4\check{m}_i\check{p}_i^3 + \check{p}_i^4 && \text{(B.9)}\\
&= 1 + 6\check{p}_\delta^2 - 4\check{p}_\delta - 4\check{p}_\delta^3 + \sum_{i=0}^{P} \check{p}_i^4 && \text{(B.10)}
\end{aligned}$$

And for the double sum in (6.13)

$$\begin{aligned}
\sum_{i=0}^{P}\sum_{j=0, j\neq i}^{P} \check{h}_i^2\check{h}_j^2 &= \sum_{i=0}^{P}\sum_{j=0, j\neq i}^{P} (\check{m}_i + \check{p}_i)^2(\check{m}_j + \check{p}_j)^2 && \text{(B.11)}\\
&= \sum_{i=0}^{P}\sum_{j=0, j\neq i}^{P} \check{m}_i^2\check{p}_j^2 + \check{p}_i^2\check{m}_j^2 + \check{p}_i^2\check{p}_j^2 + 2\check{p}_i^2\check{m}_j\check{p}_j + 2\check{m}_i\check{p}_i\check{p}_j^2 && \text{(B.12)}\\
&= \sum_{j\neq\delta, j=0}^{P} p_j^2 + \sum_{i\neq\delta, i=0}^{P} \check{p}_i^2 + \sum_{i=0}^{P}\sum_{j\neq i, j=0}^{P} \check{p}_i^2\check{p}_j^2 && \text{(B.13)}\\
&+ 2\check{p}_\delta \sum_{i\neq\delta, i=0}^{P} \check{p}_i^2 + 2\check{p}_\delta \sum_{j\neq\delta, j=0}^{P} \check{p}_j^2 \\
&= 2(1 - 2\check{p}_\delta)\sum_{i=0, i\neq\delta}^{P} \check{p}_i^2 + \sum_{i=0}^{P}\sum_{j=0, j\neq i}^{P} \check{p}_i^2\check{p}_j^2 && \text{(B.14)}
\end{aligned}$$

Now collecting terms (B.8), (B.10), and (B.14), the CM cost incurred from undermodeled channel taps is expressed as

$$\begin{aligned} \Delta J_{CM|BPSK} &= \left[4||\check{\mathbf{h}}_{\mathbf{p}}||_2^2\right] - \check{p}_\delta \left[4\check{p}_\delta^2 + 12 \sum_{i\neq\delta, i=0}^{P} \check{p}_i^2\right] \\ &+ \left[\sum_{i=0}^{P} \check{p}_i^4 + 3 \sum_{i=0}^{P} \sum_{j=0, j\neq i}^{P} \check{p}_i^2 \check{p}_j^2\right] \end{aligned} \tag{B.15}$$

Observe that the (not-so-surprising) difference between the two polarities (between (B.15) and (6.26)) is that the cubic (odd power) contribution is negated. We next complete the proof by showing that $\check{p}_i = -p_i$ for both approaches in §6.1.

For the channel perturbation approach described in §6.1.1, we have that

$$\check{\mathbf{h}}_{\mathbf{m}} = -\mathbf{h}_{\mathbf{m}} \tag{B.16}$$

$$= -(\mathbf{C}_{\mathbf{m}\downarrow}\mathbf{f}_{\mathbf{m}}) \tag{B.17}$$

$$= \mathbf{C}_{\mathbf{m}\downarrow}(-\mathbf{f}_{\mathbf{m}}) \tag{B.18}$$

since $\check{m}_\delta = -m_\delta$. Also, by definition, $\check{\mathbf{h}}_{\mathbf{m}} = \mathbf{C}_{\mathbf{m}\downarrow}\check{\mathbf{f}}_{\mathbf{m}}$. This immediately implies that $\check{\mathbf{f}}_{\mathbf{m}} = -\mathbf{f}_{\mathbf{m}}$, which in turn implies that $\check{\mathbf{h}}_{\mathbf{p}} = -\mathbf{h}_{\mathbf{p}}$, or $\check{p}_i = -p_i$.

For the equalizer truncation approach of §6.1.2, we have that

$$\check{\mathbf{h}}_{\mathbf{m}} = -\mathbf{h}_{\mathbf{m}} \tag{B.19}$$

$$= -(\mathbf{C}_{\downarrow}\mathbf{f}^{\ddagger}) \tag{B.20}$$

$$= \mathbf{C}_{\downarrow}(-\mathbf{f}^{\ddagger}) \tag{B.21}$$

We also have, by definition, $\check{\mathbf{h}}_{\mathbf{m}} = \mathbf{C}_{\downarrow}\check{\mathbf{f}}^{\ddagger}$, which immediately implies that $\check{\mathbf{f}}^{\ddagger} = -\mathbf{f}^{\ddagger}$, and also $\check{\tilde{\mathbf{f}}} = -\tilde{\mathbf{f}}$. Hence, $\check{\mathbf{h}}_{\mathbf{p}} = -\mathbf{h}_{\mathbf{p}}$, or $\check{p}_i = -p_i$. □□□.

We conclude that the result derived in (6.26), which assumed a "positive" polarity, is also valid for a "negative" polarity. Presumably, the multi-modulus case is analogous.

Appendix C

Adjusting Push Rods on Iron Head Sportsters

Sportsters are nearly unique among V-twins in that they have four individual cam shafts to adjust four parameters (valves,) unlike todays Big Twins. Hence, Sportsters do not suffer from undermodeling and can use optimum valve-train angles, unlike a Big Twin which compromises with a single cam shaft. The result is usable horsepower.

The objective is the proper[1] adjustment of pushrods for early-model iron head Sportsters. By "early-model," we mean those pre-1971, magneto-ignition, 4-cammed relics with solid lifters. Though our technique may apply to some later iron head AMF years (and it does) it does not apply to the aluminum head, hydraulic lifters of the yuppie[2] years. Some think it mere speculation, but it can be shown that it is highly probable[3] that if Duane Allman had decided to take the time and *prop-*

[1] Proper does not mean what the manual says. Proper does mean the right way to do it.
[2] *But you can get those things to run.* –K. T.
[3] to something like one minus rare-events type probabilities

erly adjust the pushrods on that little hot-rod, peaches would conjure up entirely different images in 75% of the U. S. population today.

The motor must be cold. If it's not cold, go find something else to do. Though you can perform all of the necessary steps with the bike on its (lame) kickstand, it's easier to block the frame at the rear of the engine cradle so that the rear tire is (slightly) off the ground. Pull out both spark plugs–we do not need to build compression for this task. If you do not have a piece of Teflon in the proper shape, use a large flathead screwdriver to "pop off" the pushrod covers. Collapse the pushrod tube and slide it up towards the head so that the adjustable lifter is visible. Now use H. D. tool no. 114[4] to hold the pushrod tube covers from falling down.

If you screw this next step up, you will be re-building the heads (new valves, guides, etc.). We will discuss the process for the front cylinder. The process is repeated for the rear cylinder. The order does not matter. Turn the motor over[5] with the kickstart[6] so that both intake and exhaust cams have their lobes pointed down. Equivalently, the lifter is in it's lowest position. Equivalently, both intake and exhaust valves are closed. Equivalently, the motor is on a compression stroke. Using two open-end wrenches, loosen the lock nut on the adjustable lifter corresponding to the intake valve. Inspect it. If it is at all blown out, replace it. Otherwise, loosen the lifter so that the pushrod (with one end in the lifter and the other in the rocker arm) freely turns. Place the pushrod in the lifter/cam follower firmly and freely lengthen the adjuster until the pushrod firmly touches the rocker

[4]Tool No. 114 is a rubber band with a bent paper clip attached to it.

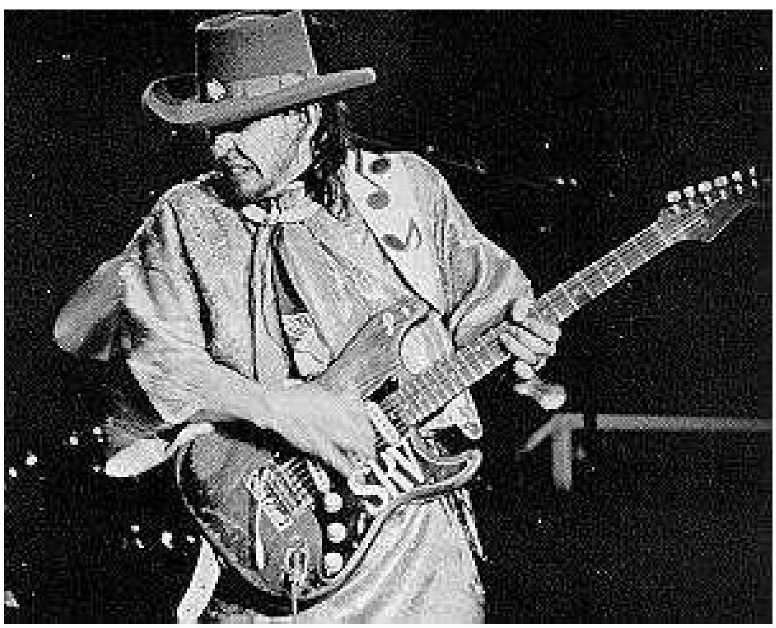
[5]Make sure the ignition is off.

[6]Those with solely an electric start can turn the motor over by placing the bike in 2^{nd} gear (trust me) and manually rotate the rear tire, thus rotating the flywheels through the tranny.

arm above. The objective is now the surgical[7] adjustment of the pushrod so that when the lock nut is tightened, the pushrod can *barely* be rotated with the grip of thumb, index and middle fingers. When this condition is satisfied, and the torque on the lock nut is 15-20 ft-lbs[8], repeat the process for the exhaust valve and then the rear head[9].

Turn the motor over through all four strokes and see that nothing seems weird. Then replace the pushrod covers and spark plugs. Get ready to start the bike, with the objective being to grab a handful of pushrod tube when the beast comes alive – when it starts, feel the pushrod tubes to make sure the pushrods are not being beaten to a pulp. If you have cams worth their weight in lead, you should feel a hearty vibration, but not a *rat-tat-tat* of looseness.

Burn rubber.



[7]Loss of horsepower measured at the rear wheel has been shown to be proportional to the fourth-order of the error from "optimum tightness."

[8]Do not over-tighten this poor little lock nut.

[9]Make sure the cams for the rear cylinder are in the correct position–pointed DOWN.

Bibliography

[1] K. Abed-Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, D. Slock, "Prediction Error Methods For Time-Domain Blind Identification Of Multichannel FIR Filters," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, pp. 1968-1971, May 1995.

[2] D. Adams, *Life, The Universe, And Everything*, London, England: Orion Books, 1982.

[3] S. T. Alexander, *Adaptive Signal Processing: Theory and Applications*, New York: Springer-Verlag, 1986.

[4] L. A. Baccalá, S. Roy, "A New Blind Time-Domain Channel Identification Method Based on Cyclostationarity," *IEEE Signal Processing Letters*, vol. 1, no. 6, pp. 89-91, 1994.

[5] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*, New York: John Wiley and Sons, 1966.

[6] S. Bellini, "Blind Equalization," *Alta Frequenza*, vol. LVII, no. 7, pp. 445-450, Sept. 1988.

[7] P. A. Bello, "Characterization of Randomly Time-Invariant Linear Channels," *IEEE Transactions on Communications Systems*, vol. CS-11, pp. 360-393, Dec. 1963.

[8] S. Benedetto, E. Biglieri, V. Castellani, *Digital Transmission Theory*, Englewood Cliffs, NJ: Prentice Hall, 1987.

[9] A. Benveniste, M. Goursat, G. Ruget, "Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385-399, June 1980.

[10] R. A. Casas, "Blind Adaptive Decision Feedback Equalization: A Class of *BAD* Channels," *M. S. Thesis*, Cornell University, Ithaca, NY, May 1996.

[11] R. A. Casas, F. Lopez de Victoria, I. Fijalkow, P. Schniter, T. Endres, C. R. Johnson, Jr., "On MMSE Fractionally Spaced Equalizer Design," *To appear in* 13^{th} *International Conference on Digital Signal Processing*, Santori, Greece, July 2-4 1997.

[12] Z. Ding, "Characteristics of Bandlimited Channels Unidentifiable from Second-Order Cyclosationary Statistics," *IEEE Signal Processing Letters*, vol. 3, no. 5, pp. 150-152, May 1996.

[13] Z. Ding and Y. Li, "On Channel Identification Based on Second-Order Cyclic Spectra," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1260-1264, May 1994.

[14] P. Duhamel, "Tutorial: Blind Equalization," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, May 1995.

[15] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr., and L. Tong, "On the Robustness of FIR Channel Identification from Fractionally-Spaced Received Signal Second-Order-Statistics," *IEEE Signal Processing Letters*, vol. 3, no. 5., pp. 153-155, May 1996.

[16] T. J. Endres, S. D. Halford, C. R. Johnson, Jr., G. B. Giannakis, "Blind Adaptive Channel Equalization Using Fractionally-Spaced Receivers: A Comparison Study," *Proc. Conference on Information Sciences and Systems*, Princeton, N.J., March 1996.

[17] T. J. Endres, S. D. Halford, C. R. Johnson, Jr., G. B. Giannakis, "Simulated Comparisons of Blind Equalization Algorithms for Cold Start-Up Applications," *International Journal of Adaptive Control & Signal Processing*, Submitted July 1996.

[18] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr, M. Green, "On the Robustness of the Fractionally-Spaced Constant Modulus Criterion to Channel Order Undermodeling: Part I," *To appear in Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, April 16-18, 1997.

[19] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr, M. Green, "On the Robustness of the Fractionally-Spaced Constant Modulus Criterion to Channel Order Undermodeling: Part II," *To appear in Proc. International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, April 20-24, 1997.

[20] I. Fijalkow, C. E. Manlove, C. R. Johnson, Jr., "Adaptive Fractionally Spaced Blind CMA Equalization," *IEEE Transactions on Signal Processing*, Submitted January 1995.

[21] I. Fijalkow, J. R. Treichler, and C. R. Johnson, Jr., "Fractionally Spaced Blind Equalization: Loss of Channel Disparity," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, pp. 1988-1991, May 1995.

[22] I. Fijalkow, "Multichannel Equalization Lower Bound: A Function of Channel Noise and Disparity," *Proc. Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, pp. 344-347, June 24-26, 1996.

[23] I. Fijalkow, A. Touzni, J. R. Treichler, "Fractionally Spaced Equalization Using CMA: Robustness to Channel Noise and Lack of Disparity," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 56-66, Jan. 1997.

[24] G. J. Foschini, "Equalizing Without Altering or Detecting Data," *AT&T Technical Journal*, vol. 64, no. 8, pp. 1885-1911, Oct. 1985.

[25] F. M. Gardner, *Phaselock Techniques*, New York: John Wiley and Sons, Inc., Second Edition, 1979.

[26] W. A. Gardner, "A New Method of Channel Identification," *IEEE Transactions on Communications*, vol. 39, no. 6, pp. 813-817, June 1991.

[27] D. Gesbert, P. Duhamel, S. Mayrargue, "Blind Multichannel Adaptive MMSE Equalization with Controlled Delay," *Proc. Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, pp. 172-175, June 24-26, 1996.

[28] G. Giannakis, S. Halford, "Blind Fractionally-Spaced Equalization of Noisy FIR Channels: Optimal and Adaptive Solutions," *IEEE Transactions on Signal Processing*, Submitted April 28, 1995.

[29] R. D. Gitlin, J. H. Hayes, and S. B. Weinstein, *Data Communications Principles*, New York: Plenum Press, 1992.

[30] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867-1875, Oct. 1980.

[31] R. Gooch, B.Daellanbach, "Prevention of Interference Capture in a Blind (CMA-Based) Adaptive Receive Filter," *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 898-902, Oct. 1989.

[32] R. Gooch, B. Daellanbach, R. Tsui, "Wideband PCM Channel Characterization Study: Final Report (Revision 2)," *Applied Signal Technology*, Technical Report FR-020-87R2, May 1988.

[33] J. B. Gómez del Moral, E. Biglieri, "Blind Identification of Digital Communication Channels with Correlated Noise," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3154-3156, Dec. 1996.

[34] S. D. Halford, "Blind Channel Equalization for Wireless Communications," *Ph.D. Dissertation*, University of Virginia, Charlottesville, VA, January 1997.

[35] B. Hartley, T. O. Hawkes, *Rings, Modules and Linear Algebra*, London, England: Chapman and Hall, 1994.

[36] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, second edition, 1991.

[37] S. Haykin, *Blind Deconvolution*, Englewood Cliffs, NJ: Prentice Hall, 1994.

[38] C. Heegard, "Constellation Shaping for the Gaussian Channel," *Proc. Information Symposium on Information Theory*, Whistler, British Columbia, Sept. 1995.

[39] R. A. Horn, C. R. Johnson, *Matrix Analysis*, New York, NY: Cambridge Press, 1985.

[40] N. K. Jablon, "Joint Blind Equalization, Carrier Recovery, and Timing Recovery for Higher-Order QAM Signal Constellations," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1383-1398, June 1992.

[41] C. R. Johnson, Jr., P. B. Schniter, T. J. Endres, R. D. Brown, R. A. Casas, D. R. Brown, and C. U. Berg, "Blind Equalization Using the Constant Modulus Criterion: A Review," *Proccedings of the IEEE*, Submitted April 1997 to special issue on blind identification and equalization.

[42] C. R. Johnson, Jr., H. J. Lee, J. P. LeBlanc, T. J. Endres, et al., "On Fractionally-Spaced Equalizer Design for Digital Microwave Radio Channels," *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 290-294, Oct. 1995.

[43] C. R. Johnson, Jr., B. D. O. Anderson, "Godard Blind Equalizer Error Surface Characteristics: White, Zero-Mean, Binary Case," *International Journal of Adaptive Control & Signal Processing*, vol. 9, pp. 301-324, July-August 1995.

[44] C. R. Johnson, Jr., J. P. LeBlanc, V. Krishnamurthy, "Godard Blind Equalizer Misbehavior With Correlated Sources," *Journal Marocain d'Automatique, d'Informatique, et de Traitment du Signal*, vol. 2, pp. 1-39, June 1993.

[45] E. I. Jury, *Inners and Stability of Dynamic Systems*, Malabar, FA: R. E. Krieger Publishing Company, second edition, 1982.

[46] T. Kailath, *Linear Systems*, Englewood Cliffs, NJ: Prentice Hall, 1980.

[47] R. A. Kennedy, "Operational Aspects of Decision Feedback Equalizers," *Ph.D. Dissertation*, The Australian National University, Canberra ACT, Australia, Dec. 1988.

[48] S. Lambotharan, J. Chambers, C. R. Johnson, Jr., "Attraction of Saddles and Slow Convergence in CMA Adaptation," *Submitted to Elsevier Science*, February 12, 1997.

[49] M. G. Larimore, S. L. Wood, J. R. Treichler, "Performance Costs for Theoretical Minimal-Length Equalizers," *To appear in Proc. International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, April 20-24, 1997.

[50] R. Laroia, N. Favardin, S. A. Tretter, "On Optimal Shaping of Multidimensional Constellations," *IEEE Transactions on Information Theory*, vol. IT-40, pp. 1044-1056, July 1994.

[51] J. P. LeBlanc, "Effects of Source Distributions and Correlation on Fractionally Spaced Blind Constant Modulus Algorithm Equalizers," *Ph.D. Dissertation*, Cornell University, Ithaca, NY, August 1995.

[52] J. P. LeBlanc, S. W. McLaughlin, "Nonequiprobable Constellation Shaping and Blind Constant Modulus Algorithm Equalization," *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, March 1996.

[53] J. P. LeBlanc, I. Fijalkow, B. Huber, C. R. Johnson, Jr., "Fractionally Spaced CMA Equalizers Under Periodic and Correlated Inputs," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, vol. 2, pp. 1041-1044, May 1995.

[54] J. P. LeBlanc, K. Dogancay, R. A. Kennedy, C. R. Johnson, Jr., "Effects of Input Data Correlation on the Convergence of Blind Adaptive Equalizers," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Adelaide, Aus., vol. 3, pp. 313-316, April 1994.

[55] E. A. Lee, D. G. Messerschmitt, *Digital Communication*, Boston, MA: Kluwer Academic Publishers, 1994.

[56] Y. Li, Z. Ding, "Global Convergence of Fractionally Spaced Godard (CMA) Adaptive Equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 818-826, April 1996.

[57] Y. Li, K. J. R. Liu, Z. Ding, "Length- and Cost-Dependent Local Minima of Unconstrained Blind Channel Equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2726-2735, Nov. 1996.

[58] Y. Li, K. J. R. Liu, "Static and Dynamic Convergence Behavior of Adaptive Blind Equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2736-2745, Nov. 1996.

[59] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, New York, NY: Dover Publications, Inc., 1973.

[60] R. W. Lucky, "Techniques for Adaptive Equalization of Digital Communication Systems," *Bell Systems Technical Journal*, vol. 45, no. 2, pp. 255-286, Feb. 1966.

[61] R. W. Lucky, H. R. Rudin, "An Automatic Equalizer for General-Purpose Communications Channels," *Bell Systems Technical Journal*, pp. 2179-2208, Nov. 1967.

[62] E. Moulines, P. Duhamel, J. Cardoso, S. Mayrargue, "Subspace Methods for Blind Identification of Multichannel FIR Filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-525, Feb. 1995.

[63] J. G. Proakis, *Digital Communications*, New York, NY: McGraw Hill, second edition, 1989.

[64] S. U. H. Qureshi, "Adjustment of the Position of the Reference Tap of an Adaptive Equalizer," *IEEE Transactions on Communications*, vol. COM-21, no. 9, pp. 1046-1052, 1973.

[65] S. U. H. Qureshi, "Adaptive Equalization," *Proccedings of the IEEE*, vol. 73, no. 9, pp. 1349-1387, Sept. 1985.

[66] Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulated Systems," *IEEE Transactions on Communications*, vol. COM-23, no. 6, pp. 679-682, June 1975.

[67] P. Schniter, R. Casas, F. Lopez de Victoria, "CMA-FSE Behavioral Examples," *Technical Report: Cornell University Blind Equalization Research Group*, Ithaca, NY, July 1996.

[68] J. J. Shynk, R. P. Gooch, G. Krishnamurthy, and C. K. Chan, "A Comparative Performance Study of Several Blind Equalization Algorithms," *Proc. The International Society for Optical Engineering(Adaptive Signal Processing)*, San Diego, CA, pp. 102-117, July 1991.

[69] M. K. Simon, S. M. Hinedi, W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Englewood Cliffs, NJ: Prentice Hall, 1995.

[70] D. T. M. Slock, "Blind Fractionally-Spaced Equalization, Perfect Reconstruction Filter Banks and Multichannel Linear Prediction," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Adelaide, Aus., pp. 585-588, April 19-22, 1994.

[71] D. T. M. Slock and C. B. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, pp. 1964-1967, May 1995.

[72] G. Strang, *Linear Algebra and Its Applications*, Orlando, FA: Harcourt Brace Jovanovich, second edition, 1988.

[73] R. S. Strichartz, *The Way of Analysis*, Boston, MA: Jones and Bartlett Publishers, 1995.

[74] R. Swaminathan, J. K. Tugnait, "On Improving the Convergence of Constant Modulus Algorithm Adaptive Filters," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, pp. 340-343, April 1993.

[75] L. Tong, G. Xu, and T. Kailath, "A New Approach to Blind Identification and Equalization of Multipath Channels," *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 856-860, Nov. 1991.

[76] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization of Multipath Channels," *Proc. International Conference on Communications*, Chicago, IL, pp. 1513-1517, June 1992.

[77] L. Tong, G. Xu, and T. Kailath, "Fast Blind Equalization Via Antenna Arrays," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Mineapolis, MN, vol. 4, pp. 272-275, April 1993.

[78] L. Tong, G. Xu, and T. Kailath, "Blind Channel Identification Based on Second-Order Statistics: A Time Domain Approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340-349, March 1994.

[79] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind Channel Identification Based on Second-Order Statistics: A Frequency Domain Approach," *IEEE Transactions on Information Theory*, vol. 41, pp. 329-334, Jan. 1995.

[80] L. Tong, "Identifiability of Minimal, Stable, Rational, Causal Systems Using Second-Order Output Cyclic Spectra," *IEEE Transactions on Automatic Control*, vol. 41, pp. 329-334, Jan. 1995.

[81] L. Tong, H. H. Zeng, "Channel-Surfing Re-Initialization for the Constant Modulus Algorithm," *IEEE Signal Processing Letters*, vol. 4, no. 3, pp. 85-87, March 1997.

[82] A. Touzni, I. Fijalkow, J. R. Treichler, "Robustness of Fractionally-Spaced Equalization by CMA to Lack of Channel Disparity and Noise," *Proc. Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, pp. 144-147, June 1996.

[83] A. Touzni, I. Fijalkow, J. R. Treichler, "Fractionally-Spaced CMA Under Channel Noise," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, pp. 2674-2677, May 1996.

[84] A. Touzni, I. Fijalkow, "Does Fractionally-Spaced CMA Converge Faster Than LMS?," *Proc. European Signal and Image Processing Conference*, Trieste, Italy, pp. 1227-1230, Sept. 1996.

[85] J. R. Treichler, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally-Spaced Equalizers: How Long Should They Really Be?," *Signal Processing Magazine*, vol. 13, no. 3, pp. 65-81, May 1996.

[86] J. R. Treichler, "Private correspondence," *Cornell University*, Ithaca, NY, Feb. 1995.

[87] J. R. Treichler, C. R. Johnson, Jr., M. G. Larimore, *Theory and Design of Adaptive Filters*, New York: John Wiley and Sons, 1987.

[88] J.R. Treichler and B. G. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 2, pp. 459-472, April 1983.

[89] J. K. Tugnait, "On Blind Identifiability of Multipath Channels Using Fractional Sampling and 2^{nd}-Order Cyclostationarity Statistics," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 308-311, 1995.

[90] J. K. Tugnait, "On Fractionally Spaced Blind Adaptive Equalization Under Symbol Timing Offsets Using Godard and Related Equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1817-1821, July 1996.

[91] G. Ungerboeck, "Fractional Tap-Spacing Equalizer and Consequences for Clock Recovery in Data Modems," *IEEE Transactions on Communications*, vol. COM-24, no. 8, pp. 856-864, Aug. 1976.

[92] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[93] A. J. Viterbi, *Principles of Coherent Communication*, New York: McGraw-Hill, 1966.

[94] B. Widrow, J. McCool, and M. Ball, "The Complex LMS Algorithm," *Proccedings of the IEEE*, vol. 63, no. 4, pp. 719-720, April 1975.

[95] B. Widrow, J. M. McCool, M. G. Larimore, C. R. Johnson, Jr., "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *Proccedings of the IEEE*, vol. 64, no. 8, pp. 1151-1162, Aug. 1976.

[96] G. Xu, H. Liu, L. Tong, T. Kailath, "A Least-Squares Approach to Blind Channel Identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982-2993, Dec. 1995.

[97] H. H. Zeng and L. Tong, "Mean-Squared Error Performance of Constant Modulus Receiver for Singular Channels," *To appear in proc. International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, April 1997.

[98] H. H. Zeng, L. Tong, "On the Performance of CMA in the Presence of Noise," *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, March 1996.

[99] H. H. Zeng, L. Tong, C. R. Johnson, Jr., "Relationships Between the Constant Modulus and Wiener Receivers," *submitted to IEEE Transactions on Information Theory*, August 1996.

[100] H. H. Zeng, L. Tong, C. R. Johnson, Jr., "Behavior of Fractionally-Spaced Constant Modulus Algorithm: Mean Square Error, Robustness and Local Minima," *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1996.

[101] S. Zeng, H. H. Zeng, L. Tong, "Blind Equalization via Multiobjective Optimization," *Proc. Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, pp. 160-163, June 1996.

[102] S. Zeng, H. H. Zeng, L. Tong, "A Blind Channel Estimator Using the Constant Modulus Criterion with Subspace Constraints," *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, March 1996.