

**(SEMI) BLIND CHANNEL AND DATA
RECOVERY IN OFDM**

BY

AHMED ABDUL QUADEER

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

JUNE 2008


KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES


This thesis, written by **AHMED ABDUL QUADEER** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

Thesis Committee



Dr. Tareq Y. Al-Naffouri (Advisor)


Dr. Maan A. Kousa (Member)


Dr. M. Adnan Landolsi (Member)


Dr. Ali H. Muqaibel (Member)


Dr. Wail A. Mousa (Member)


Dr. Ibrahim O. Habiballah
Department Chairman

Dr. Mohammad S. Al-Homoud
Dean of Graduate Studies

Date

21 / 7 / 08



To My Parents

ACKNOWLEDGMENTS

I reserve most thanks and appreciation to ALLAH the Almighty for His countless blessings. It is only He who drags me out of every problem I face. I pray that this work and the growth I had at KFUPM will be used in His cause.

I would like to thank King Fahd University of Petroleum & Minerals for giving me the opportunity to study and research at such a prestigious university.

It was a true honor to work with my advisor, Dr. Tareq Y. Al-Naffouri. I would like to take this opportunity to express my sincere gratitude and appreciation to him for the endless hours of guidance, ideas and support throughout my KFUPM experience. I have learned immensely from his deep knowledge, discipline, and intuition. Good guidance is very rare but I can easily rank his guidance as the best. It would not be an overstatement to say that he was surely the best teacher I ever had. Halfway through the Thesis, he had to go to University of Southern California, U.S.A., as a Fulbright scholar. In spite of the geographical distance, he always remained in contact and helped me out through every problem I faced. I would like to thank him also for flying to KFUPM to attend my thesis defense.

I would like to thank the members of my thesis committee, Dr. Maan A. Kousa, Dr. Ali H. Muqaibel, Dr. M. Adnan Landolsi, and Dr. Wail A. Mousa for their assistance and valuable suggestions.

I would also like to thank Dr. Mohammad Moinuddin for his support and guidance throughout my stay at KFUPM, especially during the early days.

I was fortunate enough to find friends like Saqib, Zeeshan, and Salman here. It would be impossible for me to forget the time we spent together during courses, thesis, playing games, visiting places, and finding jobs. This work would not have been possible without their love, encouragement, motivation, and support.

It would not be an exaggeration to say that my time at KFUPM had been the best experience in my life. It was mainly due to Matam Al-Maiyaar which included Zeeshan, Saqib, Salman, Mazhar Bhai, Babar, Asif, and Akhlaq Bhai. All the good time we had together will remain a valuable part of my memory.

Finally, I would like to thank my parents whose blessings and good wishes have brought me to this stage. It is very difficult to express my gratitude in words for their invaluable love and support. I would also like to thank Tahreem Younus for her love, support, and motivation during my stay at KFUPM. The support and encouragement of my sisters was also priceless. Their indirect contributions to my life are too numerous to delineate here.

I would like to thank my laptop too for not betraying me even once during the course of this Thesis.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT (ENGLISH)	xi
ABSTRACT (ARABIC)	xiii
1 INTRODUCTION	1
1.1 The Need for Channel Estimation in a Wireless Environment	2
1.2 Techniques for Channel Estimation and Data Detection	5
1.2.1 Constraints Used in Channel Estimation and Data Detection . .	5
1.2.2 Approaches to Channel Estimation and Data Detection	7
1.3 Overview of Contributions	9
1.3.1 Cyclic Prefix Based Enhanced Data Recovery in OFDM	9
1.3.2 Utilizing Gaussian Assumption for Semiblind Channel Estimation	10
1.3.3 Performance of Forward-Backward Kalman Filter Based STBC MIMO OFDM Receiver over Spatially Correlated Channel	11
2 DATA-CENTERED BLIND ESTIMATION	12
2.1 Introduction	12
2.1.1 The Approach and Organization of the Chapter	13
2.2 Notation	15
2.3 System Overview	17

2.3.1	Circular Convolution (Subchannel)	20
2.3.2	Linear Convolution (Subchannel)	22
2.4	Maximum-Likelihood Estimation	23
2.5	ML Estimation in the Constant Modulus Case	26
2.6	Approximate Methods to Reduce Computational Complexity	28
2.6.1	Linearization Approach	28
2.6.2	ML Estimation at High SNR	30
2.6.3	Using Search Algorithms	33
2.6.4	Reduced Exhaustive Search Algorithm	34
2.6.5	Using Pilots and Frequency Correlation	36
2.6.6	Using Newton's Method	36
2.7	Enhanced Equalization Using CP	44
2.8	Simulation Results	45
2.8.1	Blind Data Detector	45
2.8.2	Enhanced Equalization Using CP	54
2.9	Conclusion	58
3	CHANNEL-CENTERED BLIND ESTIMATION	61
3.1	Introduction	61
3.1.1	The Approach and Organization of the Chapter	62
3.2	System Overview	62
3.3	Evaluating the Log Likelihood Function	63
3.3.1	Maximum Likelihood Estimation of the Channel IR	66
3.3.2	Plots of Likelihood Function vs Channel Taps	67
3.4	Finding Gradient of \mathcal{L} w.r.t. \mathbf{h}	67
3.4.1	Writing \mathbf{H} in Block Form	70
3.4.2	Evaluating Second Order Moment of Output \mathbf{Y}	71
3.4.3	Gradient Matrix of \mathcal{L} w.r.t \mathbf{h}	74
3.4.4	An Alternative Condition for Optimality	77
3.5	Semiblind Estimation Using Steepest Descent Algorithm	79

3.6	Computational Complexity	80
3.6.1	Calculating $\Sigma_{\mathbf{Y}}^{-1}$	82
3.6.2	Calculating $\det(\Sigma_{\mathbf{Y}})$	84
3.7	Simulation Results	85
3.7.1	BER vs SNR Comparison for BPSK Modulated Data	85
3.7.2	BER vs SNR Comparison for 16–QAM Modulated Data	86
3.7.3	Sensitivity to Number of Pilots	86
3.8	Conclusion	87
4	MIMO SEMI-BLIND CHANNEL ESTIMATION AND DATA DE- TECTION	89
4.1	Introduction	89
4.1.1	The Approach of this Chapter	90
4.1.2	Organization of the Chapter	91
4.1.3	Notation	91
4.2	System Overview	92
4.2.1	Transmitter	92
4.2.2	Channel Model	93
4.2.3	Receiver	97
4.3	I/O Equations for MIMO-OFDM	99
4.3.1	I/O Equations with Space Time Coding: Channel Estimation Version	100
4.3.2	I/O Equations with Space Time Coding: Data Detection Version	103
4.4	The EM-Based Forward-Backward Kalman	105
4.4.1	Channel Estimation–Known Input Case	105
4.4.2	Channel Estimation–Unknown Input Case	106
4.4.3	Initial Channel Estimation	107
4.4.4	Data Detection	108
4.4.5	Summary of the EM-Based FB Kalman	109
4.5	Two Extensions/Modifications of the FB Kalman Filter	111

4.5.1	Modification-I: Kalman- (Forward-Only) Based Estimation . . .	111
4.5.2	Modification-II: Helix Based FB Kalman	111
4.6	Simulation Results	112
4.6.1	Comparison of Kalman and FB Kalman over Spatially White and Spatially Correlated Channels	113
4.6.2	Comparison of Kalman, FB Kalman and Helix Based FB Kalman	113
4.6.3	Effect of Using Outer Code	115
4.6.4	Effect of Using Different Number of Pilots	115
4.7	Conclusion	116
APPENDIX A: Channel Model in the Presence of Spatial Correla- tion		118
APPENDIX B: Calculating the Moments of X		122
5	CONCLUSIONS AND FUTURE WORK	124
5.1	Concluding Remarks	124
5.2	Future Work	126
5.2.1	General Time Variant Case	126
5.2.2	Iterative Methods for Non-Constant Modulus Data	126
5.2.3	Improving the Performance of the Semiblind Channel Centered Receiver	127
REFERENCES		128
VITAE		142

LIST OF TABLES

4.1	Indices used in the analysis of MIMO semiblind channel estimation and data detection	91
-----	--	----

LIST OF FIGURES

2.1	BER vs SNR for BPSK-OFDM over a Rayleigh channel	46
2.2	BER vs SNR for 4QAM-OFDM over a Rayleigh channel	48
2.3	BER vs SNR for BPSK-OFDM over channel with zeros on FFT grid	48
2.4	BER vs SNR for BPSK-OFDM over constant channel	49
2.5	BER vs SNR for BPSK-OFDM with previous symbol assumed to be zero	50
2.6	Comparison of low complexity algorithms for BPSK-OFDM over a Rayleigh channel	51
2.7	Sensitivity of Reduced search algorithms to number of iterations for BPSK- OFDM over a Rayleigh channel	52
2.8	Comparison of Reduced search algorithms for BPSK-OFDM over a Rayleigh channel	53
2.9	Comparison of Newton's Method for 4QAM-OFDM with $N = 16$ and $L = 4$ over a Rayleigh channel	54
2.10	Comparison of Newton's Method for 4QAM-OFDM with $N = 64$ and $L = 16$ over a Rayleigh channel	55
2.11	Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over a Rayleigh channel	56
2.12	Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over channel with zeros on FFT grid	57
2.13	Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over a Rayleigh channel	58
2.14	Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over channel with zeros on FFT grid	59

3.1	Number of samples vs transmitted data (\mathbf{x}_i)	64
3.2	3D plot of likelihood function against channel taps with $\sigma_n^2 = 1$	68
3.3	Top view 3D plot of likelihood function against channel taps with $\sigma_n^2 = 1$	68
3.4	3D plot of likelihood function against channel taps with $\sigma_n^2 = 0.1$	69
3.5	Top view 3D plot of likelihood function against channel taps with $\sigma_n^2 = 0.1$	69
3.6	BER vs SNR comparison for BPSK modulated data	86
3.7	BER vs SNR comparison for 16-QAM modulated data	87
3.8	BER vs SNR comparison for BPSK modulated data with different number of pilots	88
4.1	OSTBC OFDM Transmitter	93
4.2	OSTBC OFDM Receiver	98
4.3	BER performance of Kalman and FB-Kalman using Soft data over spatially white and correlated channel models	114
4.4	BER performance of Kalman, FB Kalman (Cyclic) and Helix based FB Kalman over spatially correlated channel	114
4.5	Effect of using outer code on performance of FB Kalman (Cyclic) and FB Kalman (Helix) over spatially correlated channel	115
4.6	Sensitivity of FB Kalman (Cyclic) and FB Kalman (Helix) to different num- ber of pilots using an outer code	116
4.7	Sensitivity of FB Kalman (Cyclic) and FB Kalman (Helix) to different num- ber of pilots without using an outer code	117

THESIS ABSTRACT

NAME: Ahmed Abdul Quadeer
TITLE OF STUDY: (Semi) Blind Channel and Data Recovery in OFDM
MAJOR FIELD: Electrical Engineering
DATE OF DEGREE: JUNE 2008

OFDM modulation combines the advantages of high achievable rates and relatively easy implementation. In this Thesis, we focus on the receiver design for OFDM transmission over block fading channels. Blind and semiblind techniques are developed for channel estimation and data recovery. The techniques use as many a priori information about the communication problem as possible to reduce the training overhead and/or improve the channel estimation accuracy and bit error rate. The thesis develops two blind techniques, a data centered technique (i.e. the input data is the main unknown) and the other is channel centered (i.e. channel is the main unknown). In the data centered blind technique, we show how an OFDM symbol can be blindly detected using output symbol and associated cyclic prefix. The approach is based on the transformation of the OFDM channel into two parallel subchannels due to the presence of cyclic prefix at the input. In the channel centered blind approach,

the cyclic prefix and the Gaussian assumption on the transmitted data in an OFDM system are used for channel estimation and subsequent data detection. Finally, the thesis considers a more realistic problem in which we design a semiblind receiver for space time coded OFDM transmission over MIMO frequency selective, time-variant, and spatially correlated channels.

خلاصة الرسالة

الاسم الكامل	: احمد عبد القدير
عنوان الرسالة	: القناة الشبه اعمي و استعادة البيانات في نظام مضاعفة الانقسام الترددي المتعامد (OFDM)
التخصص	: هندسة الكهربائية
تاريخ الشهادة	: يونيو 2008

تضمن مضاعفة الانقسام الترددي المتعامد (OFDM) يجمع بين مزايا تحقيق معدلات عالية و سهولة التنفيذ نسبيا. في هذه الرسالة ، تم التركيز علي تصميم جهاز الاستقبال لظام ارسال (OFDM). تم تطوير طرق عمياء و شبه عمياء لتقدير القناة و استعادة البيانات. هذه الطريقة تستعمل اكبر عدد ممكن من المعلومات المسبقة عن مشكلة الاتصال لتقليل تكلفة التدريب و تطوير الدقة لتحسين دقة التقدير و معدل الخطاء (BER). تم ايضا تطوير طريقتان عمياء. الاولى تتمحور حول البيانات (أي بيانات المدخلات غير معروفة) والأخرى تركز على القناة (أي القناة غير معروفة). في الطريقة الاولى، نبين كيفية إكتشاف رمز بطريقة الاكتشاف الاعمي باستخدام رمز الخرج وما يرتبط بها من دوري الباءه (Cyclic prefix). ويستند هذا النهج الى تحويل القناة (OFDM) الي قناتين متوازيتين لوجود دوري الباءه (Cyclic prefix) في المدخل. في الطريقة الاخرى التي تركز على القناة، تم افتراض بادئة دوري (Cyclic prefix) و افتراض قوسي (Gaussian) علي البيانات المرسله لتقدير القناة والكشف عن البيانات اللاحقه. تم ايضا تقديم طريقة شبه اعمي للارسال علي قناة متعددة المدخلات و المخرجات (MIMO) متغيرة زمنيا و زات تردد انتقائي و ارتباطا مكانيا.

CHAPTER 1

INTRODUCTION

The motive of modern broadband wireless communication systems is to offer high data rate services. The main hindrance for such high data rate systems is multipath fading as it results in inter-symbol interference (ISI). It therefore becomes essential to use such modulation techniques that are robust to multipath fading. Multicarrier techniques especially Orthogonal Frequency Division Multiplexing (OFDM) has emerged as a modulation scheme that can achieve high data rate by efficiently handling multipath effects. The additional advantages of simple implementation and high spectral efficiency due to orthogonality contribute towards the increasing interest in OFDM. This is reflected by the many standards that considered and adopted OFDM, including those for digital audio and video broadcasting (DAB and DVB), WIMAX (Worldwide Interoperability for Microwave Access), high speed modems over digital subscriber lines, and local area wireless broadband standards such as the HIPER-LAN/2 and IEEE 802.11a, with data rates of up to 54 Mbps [1]. OFDM is also being considered for fourth-generation (4G) mobile wireless systems [2].

In order to achieve high data rate in OFDM, receivers must estimate the channel efficiently and subsequently the data. The receiver also needs to be of low complexity and should not require too much overhead. The problem becomes especially challenging in the wireless environment when the channel is time-variant. This Thesis is concerned with (semi) blind receivers for OFDM over block fading channels.

This introductory chapter sets the stage for the Thesis. It starts by discussing the need for channel estimation in OFDM systems in Section 1.1. The chapter then presents an overview of the various channel estimation techniques that have been proposed in literature. The chapter concludes by laying out the contributions of the subsequent chapters in Section 1.3 which also serve to outline the Thesis organization.

1.1 The Need for Channel Estimation in a Wireless Environment

In OFDM systems¹, a cyclic prefix (CP) is appended to the transmitted symbol. This allows OFDM to deal effectively with ISI by transforming the equalization problem into parallel single tap equalizers. This does not completely solve the problem in a wireless environment as the equalizer taps need to be estimated. These taps are usually time variant for a wireless channel. So it becomes essential for the OFDM receiver to estimate the channel continuously for proper data detection.

In the following, we summarize the major requirements in an OFDM receiver

¹While the remarks in this section apply to a general wireless channel, we concentrate here on OFDM systems.

design (channel estimation and data detection). The receiver needs to:

1. Deal with time variant channels

OFDM is a technology that is being increasingly employed in wireless systems. This means that the receiver needs to be able to deal with mobility, i.e. with time-variant channels. In doing so, the receiver needs to take care of the following constraints

Reduce training overhead: The easiest way to deal with time-variant channels is to send enough pilots. Since, the channel impulse response (IR) can be as long as the CP of the OFDM symbol, which is roughly one-fourth the OFDM symbol length [5], each symbol would waste one-fourth of the throughput in training. Thus, the OFDM receiver should employ more intelligent techniques for channel estimation that would avoid the need for excessive training and deal with time-variant channels.

Avoid any latency by relying on the current symbol only: Some techniques for channel estimation might deal with the lack of enough training by relying on past or future symbols to perform some averaging-based channel estimation as is the case with many blind-based estimation techniques. This inherently assumes that the channel remains constant over several OFDM symbols which might not be true in a wireless scenario. Even if the channel is correlated from one symbol to another [4], a filtering or smoothing approach to channel estimation requires excessive storage and results in undesirable latency.

Thus, the proper answer to time-variant channels is to use as much natural structure as possible in the current OFDM symbol. This includes 1) The cyclic prefix, 2) the finite alphabet constraint on the data, and 3) the channel finite delay spread and correlation, and rely as little as possible on smoothing or averaging techniques.

2. Reduce complexity and storage requirements

As pointed out above, the algorithm should bootstrap itself from the current symbol without need for storing past data and especially without having to rely on the future symbols. The bootstrap should not also come at the expense of increased complexity.

3. Deal with special channel conditions

In an OFDM setting, the receiver should be able to deal with some special channel conditions which include

CP length shorter than the length of channel impulse response: This is usually dealt with by using some channel impulse response shortening techniques.

Zeros on FFT grid of channel impulse response: The frequency domain element-by-element relationship in OFDM (see equation (2.14) in next chapter) is usually used for data detection. However, it is not unusual for channel's frequency response to be zero at some carrier i which makes it impossible to

detect the data resulting in an error floor in the BER curve. The receiver should deal with this abnormality too.

Time variation within the OFDM symbol leading to inter-carrier interference: For applications with high mobility, the receiver should be able to deal with channels that vary within the OFDM symbol which gives rise to inter-carrier interference. However, a prerequisite for solving this problem is the ability to design a receiver that can cope with the milder block-fading variation problem ².

1.2 Techniques for Channel Estimation and Data Detection

Channel estimation for OFDM systems has been an active area of research. There have been several algorithms proposed for channel estimation in literature. These algorithms can be classified according to the constraints that have been used in performing channel estimation (and data detection) or according to the estimation approach used.

1.2.1 Constraints Used in Channel Estimation and Data Detection

In literature, all algorithms for channel estimation use some inherent structure of the communication problem. This structure is produced by constraints on the data or the

²This Thesis focuses on the block fading model

channel. In the following, we categorize the research work done on channel estimation on the basis of the constraints used.

Data Constraints

Finite alphabet constraint: Data is usually drawn from a finite alphabet [4], [34], [35], [57], [58].

Code: Data is coded before being transmitted which introduces redundancy and helps in reducing probability of error [27], [28], [31], [33], [47]-[50], [56].

Transmit precoding: Precoding might be done on the data at the transmitter to assist channel estimation at the receiver such as cyclic prefix [4], [25], [30], [31], [40], [46], zero padding (silent guard bands) [8], [9], [10] and virtual carriers (the subcarriers that are set to zero without any information) [42], [43], [44], [63].

Pilots: Pilots i.e. training symbols for the receiver, have been extensively used for channel estimation in OFDM [11]-[23].

Channel Constraints

Finite delay spread: The length of channel impulse response is considered to be finite and known to the receiver.

Frequency correlation: It is assumed that some additional statistical information about the channel taps is known. This is usually captured by the frequency correlation in the frequency response of the channel taps [4], [12], [31], [51], [59].

Time correlation: As channels vary with time, they show some form of time correlation. In a wireless environment, it is introduced by the doppler effect [4], [10], [33], [52], [54].

1.2.2 Approaches to Channel Estimation and Data Detection

The algorithms used for channel estimation in OFDM can also be divided on the basis of approach used. These approaches can be divided into four main categories.

Training based Estimation

Pilots i.e. symbols which are known to the receiver are sent with the data symbols so that the channel can be estimated and hence the data at the receiver (see [11]-[23]). Use of training sequences decreases the system bandwidth efficiency [24] and they are suitable only if the channel is assumed to be time-invariant. But as the wireless channel is time-varying, it becomes essential to transmit pilots periodically to keep track of the varying channel. Thus this further decreases the channel throughput.

Blind Estimation

The above limitations in training based estimation techniques motivated interest in the spectrally efficient blind approach. Only natural constraints are used for estimation in blind algorithms. For example, cyclic prefix and the cyclostationarity introduced by it was used by [25], [26], [29], [30], and [46] while coding was also used along with cyclic prefix by [31]. Redundant and non-redundant linear precoding was exploited in [27], [28], [33], [47]-[50] for channel estimation. Virtual carriers have also been used

by [42]-[44] and constant modulus modulation was used by [45]. Receiver diversity was used in [36] while [37]-[41], [44] and the references therein developed a subspace approach using the second order statistics. The finite alphabet constraint on the data was explored by [34] and [35] and for reducing the computational complexity involved in it, adaptive techniques were explored by [32] and [33].

Semiblind Estimation

Semiblind techniques make use of both pilots and the natural constraints to efficiently estimate the channel. These methods use pilots to obtain an initial channel estimate and improve the estimate by using a variety of a priori information. Thus, in addition to the pilots, semiblind methods use the cyclic prefix [4], [31], [40], the finite alphabet constraint on the data as well as the frequency and time correlation of the channel [4], magnitude error in data [55], linear precoding [56], frequency correlation [12], [31], and [59], gaussian assumption on transmitted data [60], the first order statistics [61], subspace of the channel [62], receiver diversity and virtual carriers [63] for channel estimation and subsequent data detection. Semiblind adaptive approaches for channel estimation have also been exploited by [57] and [58] who in addition to pilots, utilized the finite alphabet nature of data and the second order statistics of the received signal, respectively.

Data-aided Estimation

The purpose of channel estimation is to use that estimate to detect data. The recovered data, in turn, can also be used to improve the channel estimate, thus giving

rise to an iterative technique for channel and data recovery. This idea is the basis of joint channel estimation and data detection. This iterative technique was used in a data-aided fashion by [39] or more rigorously by the expectation maximization (EM) approach [68]-[73].

1.3 Overview of Contributions

1.3.1 Cyclic Prefix Based Enhanced Data Recovery in OFDM

In Chapter 2, we develop a totally blind algorithm for data detection by using the outputs of the circular and linear channels (to be explained in Chapter 2) and we use exhaustive search to detect the data. By using exhaustive search, we obtain a bench mark against which other lower complexity versions of the algorithm can be compared. Moreover, we modify the algorithm to incorporate a priori information about the channel such as the finite delay spread and frequency correlation. We also investigate the ability of the algorithm to deal with zeros on the FFT grid. Since the blind algorithm uses both the linear and circular channel, it is robust to the presence of zeros on the FFT grid even as compared with the case when the channel is perfectly known at the receiver. In that chapter, we also propose approximate methods to reduce the computational complexity involved in the exhaustive search. As all standard-based OFDM systems involve some form of training, we have also studied the behavior of the blind receiver in the presence of pilots. A new method of enhanced equalization using cyclic prefix has also been proposed when the receiver has perfect or estimated

knowledge of the channel. Specifically, in this method, data is recovered using both the linear and circular subchannels as opposed to the conventional method which utilizes only circular subchannel.

1.3.2 Utilizing Gaussian Assumption for Semiblind Channel Estimation

In Chapter 3, we propose a receiver that is semiblind and emphasizes on channel estimation. Cyclic prefix and Gaussian assumption on the transmitted data in an OFDM system are used for channel estimation in this technique. The Gaussian assumption on input helps to easily evaluate the pdf of output and channel is estimated by maximizing the log likelihood function. We investigate the likelihood function by simulations to check whether it has a global maxima by plotting it against the channel taps. The results show that the likelihood function has more than one maxima. So, a semiblind approach is adopted. We derive the gradient of likelihood function with respect to the channel taps $\mathbf{h}_i(1)$, $\mathbf{h}_i(2)$, ..., $\mathbf{h}_i(L)$, and then use it in the Steepest Descent algorithm (initialized with a noisy channel estimate) to find the likelihood function maxima.

1.3.3 Performance of Forward-Backward Kalman Filter Based STBC MIMO OFDM Receiver over Spatially Correlated Channel

In Chapter 4, we solve a more practical problem where we design a receiver for channel estimation and data recovery for OFDM transmission over MIMO time-variant spatially correlated channels. The receiver uses all possible constraints on the channel (the finite delay spread, frequency and time correlation, and transmit and receive spatial correlation) and the data (the finite alphabet constraint, the cyclic prefix, pilots, and the orthogonal space time block coding (OSTBC)). Our approach is based on [96] which proposed a Kalman filter approach to channel estimation in MIMO OFDM. We have extended this approach to spatially correlated channels and explored the use of forward-backward Kalman filter with different implementations (cyclic and helix). We have also discussed the effect of using an outer code and using reduced number of pilots on the performance of the receiver.

CHAPTER 2

DATA-CENTERED BLIND ESTIMATION

2.1 Introduction

In this chapter, we show how the cyclic prefix can be used to enhance the performance of an OFDM receiver. In the first part of the chapter, we consider blind data detection for OFDM transmission over block fading channels. Specifically, we show how an OFDM symbol can be blindly detected using output symbol and associated cyclic prefix. Our approach relies on decomposing the OFDM channel into two subchannels (circular and linear) that share the same input and are characterized by the same channel parameters. This fact enables us to estimate the channel parameters from one subchannel and substitute the estimate into the other, thus obtaining a nonlinear relationship involving the input and output data only that can be searched for the maximum likelihood estimate of the input. This shows that OFDM systems are com-

pletely identifiable using output data only, irrespective of the channel zeros, as long as the channel delay spread is less than the length of the cyclic prefix. We also propose iterative methods to reduce the computational complexity involved in the maximum likelihood search of input. In the second part of the chapter, we show how the cyclic prefix can be used to enhance the operation of the channel equalizer when the channel is known at the receiver(perfectly or through training).

2.1.1 The Approach and Organization of the Chapter

This chapter presents the improvement in the performance of the OFDM receiver by using cyclic prefix (CP) whether operating in the blind, semiblind, training or perfectly known channel modes.

In the first part of the chapter, we perform channel identification and equalization from output data only (i.e. OFDM output symbol and associated cyclic prefix (CP)), without the need for a training sequence or a priori channel information. The advantage of our approach is three fold:

1. The method provides a blind estimate of the data from one output symbol without the need for training or averaging (contrary to the common practice in blind methods where averaging over several symbols is required). Thus, the method lends itself to block fading channels.
2. Data detection is done without any restriction on the channel (as long as the delay spread is shorter than the (CP)). In fact, data detection can be performed

even in the presence of zeros on the FFT grid.¹

3. The fact that we use two observations (the OFDM symbol and CP) to recover the input symbol enhances the diversity of the system as can be seen from simulations.

Our approach is based on the transformation of the OFDM channel into two parallel subchannels due to the presence of a cyclic prefix at the input (see Section 2.3). One is a circular subchannel that relates the input and output OFDM symbols and thus is free of any intersymbol interference (ISI) effects and is best described in the frequency domain (Section 2.3.1). The other one is a linear subchannel that carries the burden of ISI and that relates the input and output prefixes through linear convolution (Section 2.3.2). This subchannel is best studied in the time domain.

It can be shown that the two subchannels are characterized by the same set of parameters (or impulse response(IR)) and are driven by the same stream of data. They only differ in the way in which they operate on the data (i.e. linear vs circular convolution). This fact enables us in Section 2.4 to estimate the IR from one subchannel and eliminate its effect from the other, thus obtaining a nonlinear least squares relationship that involves the input and output data only. This relationship can in turn be optimized for the ML data estimate, something that can be achieved through exhaustive search (in the worst case scenario). The relationship takes a particularly simple form in constant modulus case (Section 2.5).

Exhaustive search is computationally very expensive. We thus suggest in Section

¹This comes contrary to the common belief that OFDM using CP cannot be equalized for channels with zeros on the FFT grid [1] and [3]

2.6, six approaches to reduce the computational complexity. The first approach is based on approximating the nonlinear least squares problem with a linear one. In the second approach, we consider the high SNR case and try to find a closed form solution of the nonlinear least squares problem. In the third approach, we use the Particle Swarm Optimization (PSO) [81], [82], [83], and the Genetic Algorithm (GA) [84], [85], to directly solve the nonlinear problem. The estimate obtained by the linear approximation approach can be used to start these search algorithms. In the fourth approach, we propose a reduced exhaustive search algorithm. We show in our fifth approach how the CP in addition to pilots and frequency correlation can be used to estimate the channel in a semiblind manner. The sixth approach also describes a semiblind algorithm in which we use Newton's method to estimate the data when it is initialized with an estimate using frequency correlation and less number of pilots.

In the second part of the chapter (Section 2.7), we show how the CP can be used to enhance the operation of the equalizer when the channel is perfectly known at the receiver or is obtained through training. Specifically, the CP observation enhances the BER performance especially when the channel exhibits zeros on the FFT grid.

To setup the stage, we introduce our notation in the following section.

2.2 Notation

We denote scalars with small-case letters, vectors with small-case boldface letters, and matrices with uppercase boldface letters. Calligraphic notation (e.g. \mathcal{X}) is reserved for vectors in the frequency domain. The individual entries of a vector like \mathbf{h} are

denoted by $h(l)$. A hat over a variable indicates an estimate of the variable (e.g., $\hat{\mathbf{h}}$ is an estimate of \mathbf{h}). When any of these variables become a function of time, the time index i appears as a subscript.

Now consider a length- N vector \mathbf{x}_i . We deal with three derivatives associated with this vector. The first two are obtained by partitioning \mathbf{x}_i into a lower (trailing) part $\underline{\mathbf{x}}_i$ (known as the cyclic prefix) and an upper vector $\tilde{\mathbf{x}}_i$ so that

$$\mathbf{x}_i = \begin{bmatrix} \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix}$$

The third derivative, $\overline{\mathbf{x}}_i$, is created by concatenating \mathbf{x}_i with a copy of CP i.e. $\underline{\mathbf{x}}_i$.

Thus, we have

$$\overline{\mathbf{x}}_i = \begin{bmatrix} \underline{\mathbf{x}}_i \\ \mathbf{x}_i \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{x}}_i \\ \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix} \quad (2.1)$$

In line with the above notation, a matrix \mathbf{Q} having N rows will have the natural partitioning

$$\mathbf{Q} = \begin{bmatrix} \tilde{\mathbf{Q}} \\ \underline{\mathbf{Q}} \end{bmatrix} \quad (2.2)$$

where the number of rows in $\tilde{\mathbf{Q}}$ and $\underline{\mathbf{Q}}$ are understood from the context and when it

is not clear, the number of rows will appear as a subscript. Thus, we write

$$\mathbf{Q} = \begin{bmatrix} \tilde{\mathbf{Q}}_{N-L} \\ \mathbf{Q}_L \end{bmatrix} \quad (2.3)$$

2.3 System Overview

In an OFDM system, data is transmitted in symbols \mathbf{x}_i of length N each. The symbol undergoes an IFFT operation to produce the time domain symbol \mathbf{x}_i , i.e.

$$\mathbf{x}_i = \sqrt{N}\mathbf{Q}\mathbf{x}_i \quad (2.4)$$

where \mathbf{Q} is the $N \times N$ IFFT matrix. When juxtaposed, these symbols result in the sequence $\{x_k\}$.² We assume a channel $\underline{\mathbf{h}}$ of maximum length $L + 1$. To avoid ISI caused by passing through the channel, a cyclic prefix (CP) $\underline{\mathbf{x}}_i$ (of length L) is appended to \mathbf{x}_i , resulting in super-symbol $\bar{\mathbf{x}}_i$ as defined in (2.1). The concatenation of these symbols produces the underlying sequence $\{\bar{x}_k\}$. When passed through the channel $\underline{\mathbf{h}}$, the sequence $\{\bar{x}_k\}$ produces the output sequence $\{\bar{y}_k\}$ i.e.

$$\bar{y}_k = \underline{h}_k * \bar{x}_k + \bar{n}_k \quad (2.5)$$

where \bar{n}_k is the additive white Gaussian noise and $*$ stands for linear convolution.

Motivated by the symbol structure of the input, it is convenient to partition the

²The time indices in the sequence \mathbf{x}_i and the underlying sequence $\{x_k\}$ are dummy variables. Nevertheless, we chose to index the two sequences differently to avoid any confusion that might arise from choosing identical indices.

output into length $N + L$ symbol as

$$\bar{\mathbf{y}}_i = \begin{bmatrix} \underline{\mathbf{y}}_i \\ \mathbf{y}_i \end{bmatrix}$$

This is a natural way to partition the output because the prefix $\underline{\mathbf{y}}_i$ actually absorbs all ISI that takes place between the adjacent symbols $\bar{\mathbf{x}}_{i-1}$ and $\bar{\mathbf{x}}_i$. Moreover, the remaining part \mathbf{y}_i of the symbol depends on the i th input OFDM symbol \mathbf{x}_i only. These facts can be seen from the input/output relationship

$$\begin{bmatrix} \mathbf{y}_{i-1} \\ \underline{\mathbf{y}}_i \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}} & \mathbf{O}_{N \times L} & \mathbf{O}_{N \times N} \\ \mathbf{O}_{L \times N} & \underline{\mathbf{H}}_U & \underline{\mathbf{H}}_L & \mathbf{O}_{L \times N} \\ \mathbf{O}_{N \times N} & \mathbf{O}_{N \times L} & \bar{\mathbf{H}} & \mathbf{O}_{N \times N} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_{i-1} \\ \tilde{\mathbf{x}}_{i-1} \\ \underline{\mathbf{x}}_{i-1} \\ \underline{\mathbf{x}}_i \\ \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{i-1} \\ \mathbf{n}_i \\ \mathbf{n}_i \end{bmatrix} \quad (2.6)$$

where \mathbf{n} is the output noise which we take to be white Gaussian. The matrices $\bar{\mathbf{H}}$, $\underline{\mathbf{H}}_L$, and $\underline{\mathbf{H}}_U$ are convolution (Toeplitz) matrices of proper sizes created from the vector $\underline{\mathbf{h}}$. Specifically, $\bar{\mathbf{H}}$ is the $N \times (N + L)$ matrix

$$\bar{\mathbf{H}} = \begin{bmatrix} \underline{h}(L) & \cdots & \underline{h}(1) & \underline{h}(0) & & \\ \vdots & \ddots & \cdots & \ddots & \ddots & \\ 0 & \cdots & \underline{h}(L) & \cdots & \underline{h}(1) & \underline{h}(0) \end{bmatrix} \quad (2.7)$$

The matrices $\underline{\mathbf{H}}_{\text{U}}$ and $\underline{\mathbf{H}}_{\text{L}}$ are square of size L .³

$$\underline{\mathbf{H}}_{\text{U}} = \begin{bmatrix} \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(1) \\ & \underline{h}(L) & \cdots & \underline{h}(2) \\ & & \ddots & \vdots \\ & & & \underline{h}(L) \end{bmatrix} \quad (2.8)$$

$$\underline{\mathbf{H}}_{\text{L}} = \begin{bmatrix} \underline{h}(0) & & & \\ \underline{h}(1) & \underline{h}(0) & & \\ \vdots & \ddots & \ddots & \\ \underline{h}(L-1) & \cdots & \underline{h}(1) & \underline{h}(0) \end{bmatrix} \quad (2.9)$$

Because of the redundancy in the input, the convolution in (2.6) can be decomposed into two distinct constituent convolution operations or subchannels [96]. This decomposition is essential for channel and data recovery, which is the center of attention in this chapter. In what follows, we shall describe each of these operations separately.

³The matrix $\underline{\mathbf{H}}_{\text{L}}$ ($\underline{\mathbf{H}}_{\text{U}}$) is lower (upper) triangular; this explains the superscript L (U).

2.3.1 Circular Convolution (Subchannel)

From (2.6), we can write

$$\mathbf{y}_i = \overline{\mathbf{H}} \begin{bmatrix} \underline{\mathbf{x}}_i \\ \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix} = \overline{\mathbf{H}} \overline{\mathbf{x}}_i + \mathbf{n}_i \quad (2.10)$$

This shows that \mathbf{y}_i is created solely from $\overline{\mathbf{x}}_i$ through convolution and hence is ISI-free.

Moreover, the existence of a cyclic prefix in $\overline{\mathbf{x}}_i$ allows us to rewrite (2.10) as

$$\mathbf{y}_i = \mathbf{H} \mathbf{x}_i + \mathbf{n}_i \quad (2.11)$$

where \mathbf{H} is the size- N circulant matrix.

$$\mathbf{H} = \begin{bmatrix} \underline{h}(0) & 0 & \cdots & 0 & \underline{h}(L) & \cdots & \underline{h}(1) \\ \underline{h}(1) & \underline{h}(0) & \cdots & 0 & 0 & \cdots & \underline{h}(2) \\ \vdots & \vdots & \ddots & \vdots & \cdots & \ddots & \vdots \\ \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(0) \end{bmatrix} \quad (2.12)$$

In other words, the cyclic prefix of $\overline{\mathbf{x}}_i$ renders the convolution in (2.11) cyclic, and we can write

$$\boxed{\mathbf{y}_i = \mathbf{h}_i \circledast \mathbf{x}_i + \mathbf{n}_i} \quad (2.13)$$

where \mathbf{h}_i is a length- N zero-padded version of $\underline{\mathbf{h}}_i$.

$$\mathbf{h}_i = \begin{bmatrix} \underline{\mathbf{h}}_i \\ \mathbf{O}_{(N-L-1) \times 1} \end{bmatrix}$$

In the frequency domain, the circular convolution (2.13) reduces to the element-by-element operation

$$\boxed{\mathbf{y}_i = \mathcal{H}_i \odot \mathcal{X}_i + \mathcal{N}_i} \quad (2.14)$$

where \odot stands for element-by-element multiplication and where \mathcal{H}_i , \mathcal{X}_i , \mathcal{N}_i , and \mathbf{y}_i , are the DFT's of \mathbf{h}_i , \mathbf{x}_i , \mathbf{n}_i , and \mathbf{y}_i respectively

$$\mathcal{H}_i = \mathbf{Q}^* \mathbf{h}_i, \quad \mathcal{X}_i = \frac{1}{\sqrt{N}} \mathbf{Q}^* \mathbf{x}_i, \quad \mathcal{N}_i = \frac{1}{\sqrt{N}} \mathbf{Q}^* \mathbf{n}_i, \quad \text{and} \quad \mathbf{y}_i = \frac{1}{\sqrt{N}} \mathbf{Q}^* \mathbf{y}_i \quad (2.15)$$

Since $\underline{\mathbf{h}}_i$ corresponds to the first $L + 1$ elements of \mathbf{h}_i , we can show that

$$\mathcal{H}_i = \mathbf{Q}_{L+1}^* \underline{\mathbf{h}}_i \quad \text{and} \quad \underline{\mathbf{h}}_i = \mathbf{Q}_{L+1} \mathcal{H}_i \quad (2.16)$$

where \mathbf{Q}_{L+1}^* consists of the first $L + 1$ columns of \mathbf{Q}^* and \mathbf{Q}_{L+1} consists of first $L + 1$ rows of \mathbf{Q} . This allows us to rewrite (2.14) as

$$\boxed{\mathbf{y}_i = \text{diag}(\mathcal{X}_i) \mathbf{Q}_{L+1}^* \underline{\mathbf{h}}_i + \mathcal{N}_i} \quad (2.17)$$

2.3.2 Linear Convolution (Subchannel)

From (2.5), we can also deduce that the cyclic prefixes at the input and output are related by linear convolution. Specifically, if we concatenate all cyclic prefixes at the input into a sequence $\{\underline{x}_k\}$ and the cyclic prefixes at the output into the corresponding sequence $\{\underline{y}_k\}$, then we can show that the two sequences are related by linear convolution [6]

$$\boxed{\underline{y}_k = \underline{h}_k * \underline{x}_k + \underline{n}_k} \quad (2.18)$$

From this we deduce that the cyclic prefix of OFDM symbol, \mathbf{y}_i , is related to the input cyclic prefixes $\underline{\mathbf{x}}_{i-1}$ and $\underline{\mathbf{x}}_i$ by

$$\boxed{\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_i \underline{\mathbf{h}}_i + \underline{\mathbf{n}}_i} \quad (2.19)$$

where $\underline{\mathbf{X}}_i$ is constructed from $\underline{\mathbf{x}}_{i-1}$ and $\underline{\mathbf{x}}_i$ according to

$$\underline{\mathbf{X}}_i = \underline{\mathbf{X}}_{U_{i-1}} + \underline{\mathbf{X}}_{L_i} \quad (2.20)$$

and where (compare with (2.8))

$$\underline{\mathbf{X}}_{\mathbf{U}_{i-1}} = \begin{bmatrix} 0 & \underline{x}_{i-1}(L-1) & \cdots & \underline{x}_{i-1}(0) \\ 0 & 0 & \cdots & \underline{x}_{i-1}(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \underline{x}_{i-1}(L-1) \end{bmatrix}, \quad (2.21)$$

$$\text{and } \underline{\mathbf{X}}_{\mathbf{L}_i} = \begin{bmatrix} \underline{x}_i(0) & 0 & \cdots & 0 \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \underline{x}_i(L-1) & \cdots & \underline{x}_i(0) & 0 \end{bmatrix} \quad (2.22)$$

This fact together with the FFT relationship (2.16) yields the time-frequency input/output equation

$$\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_i \mathbf{Q}_{L+1} \mathcal{H}_i + \underline{\mathbf{n}}_i \quad (2.23)$$

2.4 Maximum-Likelihood Estimation

Consider the frequency domain description of the circular subchannel (2.14)

$$\mathcal{Y}_i = \mathcal{H}_i \odot \mathcal{X}_i + \mathcal{N}_i$$

To obtain the maximum-likelihood (ML) estimate of \mathcal{H}_i , we assume that the sequence \mathcal{X}_i is deterministic and perform an element-by-element division of (2.14) by \mathcal{X}_i to get

$$\mathbf{D}_{\mathcal{X}}^{-1}\mathcal{Y}_i = \mathcal{H}_i + \mathbf{D}_{\mathcal{X}}^{-1}\mathcal{N}_i \quad (2.24)$$

where

$$\mathbf{D}_{\mathcal{X}} = \text{diag}(\mathcal{X}_i) \quad (2.25)$$

Equivalently, we can write (2.24) as

$$\mathbf{D}_{\mathcal{X}}^{-1}\mathcal{Y}_i = \mathcal{H}_i + \mathcal{N}'_i \quad (2.26)$$

where \mathcal{N}'_i is Gaussian distributed with zero mean and autocorrelation matrix

$$\mathbf{R}_{n'} = \sigma_n^2 \mathbf{D}_{\mathcal{X}}^{-1} \mathbf{D}_{\mathcal{X}}^{-*} = \sigma_n^2 |\mathbf{D}_{\mathcal{X}}|^{-2} \quad (2.27)$$

The maximum-likelihood estimate of \mathcal{H} can now be obtained by solving the system of equations (2.24) in the least-squares (LS) sense subject to the constraint

$$\tilde{\mathbf{Q}}_{N-L-1} \mathcal{H}_i \triangleq \tilde{\mathbf{Q}} \mathcal{H}_i = 0 \quad (2.28)$$

We can show that the ML estimate is given by [92]

$$\begin{aligned}\hat{\mathcal{H}}_i^{ML} &= \left[\mathbf{I} - \mathbf{R}_{n'} \tilde{\mathbf{Q}}^* \left(\tilde{\mathbf{Q}} \mathbf{R}_{n'} \tilde{\mathbf{Q}}^* \right)^{-1} \tilde{\mathbf{Q}} \right] \mathbf{D}_{\mathcal{X}}^{-1} \mathbf{y}_i \\ &= \left[\mathbf{I} - |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \left(\tilde{\mathbf{Q}} |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \right)^{-1} \tilde{\mathbf{Q}} \right] \mathbf{D}_{\mathcal{X}}^{-1} \mathbf{y}_i\end{aligned}\quad (2.29)$$

The ML estimate (2.29) was obtained solely from the circular convolution subchannel.

Upon replacing \mathcal{H}_i that appears in the time-frequency input/output equation (2.23)

(corresponding to the linear subchannel)

$$\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_i \mathbf{Q}_{L+1} \mathcal{H} + \underline{\mathbf{n}}_i$$

with its ML estimate (2.29), we obtain

$$\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_i \mathbf{Q}_{L+1} \left[\mathbf{I} - |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \left(\tilde{\mathbf{Q}} |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \right)^{-1} \tilde{\mathbf{Q}} \right] \mathbf{D}_{\mathcal{X}}^{-1} \mathbf{y}_i + \underline{\mathbf{n}}_i \quad (2.30)$$

This is an input/output relationship that does not depend on any channel information whatsoever. Since the data is assumed deterministic, maximum-likelihood estimation is the optimum way to detect it, i.e. we minimize

$$\boxed{\hat{\mathcal{X}}_i^{ML} = \arg \min_{\mathcal{X}_i} \left\| \underline{\mathbf{y}}_i - \underline{\mathbf{X}}_i \mathbf{Q}_{L+1} \left[\mathbf{I} - |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \left(\tilde{\mathbf{Q}} |\mathbf{D}_{\mathcal{X}}|^{-2} \tilde{\mathbf{Q}}^* \right)^{-1} \tilde{\mathbf{Q}} \right] \mathbf{D}_{\mathcal{X}}^{-1} \mathbf{y}_i \right\|^2} \quad (2.31)$$

This is a nonlinear least-squares problem in the data. In the worst case scenario, it can be solved by an exhaustive search over all possible sequences \mathcal{X}_i .

To gain more insight into this problem, we now treat the case of constant modulus

data which leads to more explicit results.

2.5 ML Estimation in the Constant Modulus Case

In the constant modulus case, we have

$$|\mathbf{D}_{\mathcal{X}}|^{-2} = \frac{1}{\mathcal{E}_X} \mathbf{I} \quad (2.32)$$

As a consequence, we can also write

$$\mathbf{D}_{\mathcal{X}}^{-1} = \frac{1}{\mathcal{E}_X} \mathbf{D}_{\mathcal{X}}^* \quad (2.33)$$

Thus, the ML estimate of \mathcal{H}_i (2.29) simplifies to

$$\hat{\mathcal{H}}_i^{\text{ML}} = \frac{1}{\mathcal{E}_X} \left[\mathbf{I} - \tilde{\mathbf{Q}}^* (\tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^*)^{-1} \tilde{\mathbf{Q}} \right] \mathbf{D}_{\mathcal{X}}^* \mathbf{y}_i \quad (2.34)$$

$$= \frac{1}{\mathcal{E}_X} \left[\mathbf{I} - \tilde{\mathbf{Q}}^* \tilde{\mathbf{Q}} \right] \mathbf{y}_i \odot \mathbf{x}_i^* \quad (2.35)$$

where in (2.35), we used the fact that $\tilde{\mathbf{Q}}$ is a left-inverse of $\tilde{\mathbf{Q}}^*$ - a consequence of the unitary nature of \mathbf{Q}

$$\mathbf{I} = \mathbf{Q} \mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_{L+1} \\ \tilde{\mathbf{Q}}_{N-L-1} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{L+1}^* & \tilde{\mathbf{Q}}_{N-L-1}^* \end{bmatrix} \quad (2.36)$$

Just as we did in the general case, we now replace the effect of \mathcal{H}_i in the linear convolution channel, as expressed in (2.23), by its ML estimate to get

$$\underline{\mathbf{y}}_i = \frac{1}{\mathcal{E}_X} \underline{\mathbf{X}}_i \underline{\mathbf{Q}}_{L+1} \left[\mathbf{I} - \tilde{\mathbf{Q}}^* \tilde{\mathbf{Q}} \right] \mathbf{y}_i \odot \mathbf{x}_i^* + \underline{\mathbf{n}}_i \quad (2.37)$$

$$= \frac{1}{\mathcal{E}_X} \underline{\mathbf{X}}_i \underline{\mathbf{Q}}_{L+1} \mathbf{y}_i \odot \mathbf{x}_i^* + \underline{\mathbf{n}}_i \quad (2.38)$$

where in going to (2.38), we used the fact that

$$\mathbf{Q}_{L+1} \tilde{\mathbf{Q}} = \mathbf{Q}_{L+1} \tilde{\mathbf{Q}}_{N-L-1} = 0$$

which can be deduced from (2.36). The ML estimate of \mathbf{x}_i is now obtained by performing the minimization

$$\hat{\mathbf{x}}_i^{\text{ML}} = \arg \min_{\mathcal{X}_i} \left\| \underline{\mathbf{y}}_i - \frac{1}{\mathcal{E}_X} \underline{\mathbf{X}}_i \underline{\mathbf{Q}}_{L+1} \mathbf{y}_i \odot \mathbf{x}_i^* \right\|^2 \quad (2.39)$$

Notice that the only unknowns in this minimization are $\underline{\mathbf{X}}_i$ and \mathbf{x}_i , i.e. the input data sequence. This minimization is nothing but a *nonlinear least-squares* problem in the data. In the worst case scenario, we can obtain the ML estimate through an exhaustive search.

2.6 Approximate Methods to Reduce Computational Complexity

The search for the optimal \mathbf{x}_i in (2.39) is computationally very complex. In the following, we describe six approaches to reduce this complexity:

2.6.1 Linearization Approach

One way to reduce the computational complexity is to transform the nonlinear into a linear least squares problem. To do so, note first that the $\underline{\mathbf{X}}_i$ involved in equation (2.39) is composed of an upper and lower triangle formed by the CP of previous (known) and current (unknown) symbol respectively as shown in equation (2.20), i.e.

$$\underline{\mathbf{X}}_i = \underline{\mathbf{X}}_{U_{i-1}} + \underline{\mathbf{X}}_{L_i}$$

Thus equation (2.39) can be rewritten as

$$\begin{aligned} \hat{\mathbf{x}}_i^{ML} &= \arg \min_{\mathcal{X}_i} \left\| \underline{\mathbf{y}}_i - \frac{1}{\mathcal{E}_X} (\underline{\mathbf{X}}_{U_{i-1}} + \underline{\mathbf{X}}_{L_i}) \mathbf{Q}_{L+1} \mathbf{D} \mathbf{y} \mathbf{x}_i^* \right\|^2 \\ &= \arg \min_{\mathcal{X}_i} \left\| \underline{\mathbf{y}}_i - (\underline{\mathbf{X}}_{U_{i-1}} + \underline{\mathbf{X}}_{L_i}) \mathbf{A} \mathbf{x}_i^* \right\|^2 \\ &= \arg \min_{\mathcal{X}_i} \left\| \underline{\mathbf{y}}_i - \mathbf{B} \mathbf{x}_i^* - \mathbf{C} \mathbf{x}_i^* \right\|^2 \end{aligned} \quad (2.40)$$

where

$$\mathbf{A} = \frac{1}{\mathcal{E}_X} \mathbf{Q}_{L+1} \mathbf{D} \mathbf{y}, \quad \mathbf{B} = \underline{\mathbf{X}}_{U_{i-1}} \mathbf{A}$$

and hence are completely known and where

$$\mathbf{C} = \underline{\mathbf{X}}_{L_i} \mathbf{A} \quad (2.41)$$

Thus, the elements of \mathbf{C} are linear in the input \mathbf{x}_i making $\mathbf{C}\mathbf{x}_i^*$ quadratic in \mathbf{x}_i . In fact, each element of $\mathbf{c} = \mathbf{C}\mathbf{x}_i^*$ can be written as

$$c(j) = \|\mathbf{x}_i\|_{\mathbf{W}_j}^2 \triangleq \mathbf{x}_i^* \mathbf{W}_j \mathbf{x}_i \quad (2.42)$$

for some weighted matrix \mathbf{W}_j that is independent from input \mathbf{x}_i . Thus, the nonlinear minimizing problem can be written as

$$\hat{\mathbf{x}}_i^{ML} = \arg \min_{\mathbf{x}_i} \left\| \underline{\mathbf{y}}_i - \mathbf{B}\mathbf{x}_i^* - \mathbf{c} \right\|^2 \quad (2.43)$$

The linear approximation is obtained by replacing the matrix \mathbf{W}_j by its diagonal, i.e.

$$\begin{aligned} c(j) &= \|\mathbf{x}_i\|_{\mathbf{W}_j}^2 \quad 1 \leq j \leq L \\ &\simeq \|\mathbf{x}_i\|_{\text{diag}(\mathbf{W}_j)}^2 \\ &= \mathcal{E}_X \text{trace}(\mathbf{W}_j) \\ &= z(j) \end{aligned} \quad (2.44)$$

where the third line follows from the fact that the elements of \mathbf{x}_i have constant modulus. The input dependent vector \mathbf{c} is thus replaced by the constant vector \mathbf{z} ,

and the objective function becomes linear in \mathcal{X}_i

$$\arg \min_{\mathcal{X}_i} \left\| \left(\underline{\mathbf{y}}_i - \mathbf{z} \right) - \mathbf{B} \mathcal{X}_i^* \right\|^2 \quad (2.45)$$

One way to solve equation (2.45) is by using least squares

$$\hat{\mathcal{X}}_i^* = (\mathbf{B}^* \mathbf{B} + \delta \mathbf{I})^{-1} \mathbf{B}^* \left(\underline{\mathbf{y}}_i - \mathbf{z} \right) \quad (2.46)$$

where δ is a small constant.⁴

We could refine the estimate obtained in (2.46) further by using the estimate $\hat{\mathcal{X}}_i^*$ to obtain the vector \mathbf{c} in (2.43) (as opposed to the vector \mathbf{z} which is obtained by approximating \mathbf{W}_j with its diagonal). We now solve the alternative least squares problem

$$\arg \min_{\mathcal{X}_i} \left\| \left(\underline{\mathbf{y}}_i - \mathbf{c} \right) - \mathbf{B} \mathcal{X}_i^* \right\|^2 \quad (2.47)$$

This procedure of refining the estimate \mathbf{c} and solving the least squares (2.47) could be repeated for a desired number of iterations.

2.6.2 ML Estimation at High SNR

In this section, we try to find a closed form solution of the non-linear problem (equation (2.39)) at high SNR by assuming noise to be zero. At high SNR, equation (2.43)

⁴The optimum choice for δ is \mathcal{E}_X as $\text{Cov}[\mathcal{X}_i] = \mathcal{E}_X \mathbf{I}$.

reduces to

$$\underline{\mathbf{y}}_i - \mathbf{B}\boldsymbol{\chi}_i^* - \mathbf{c} = 0 \quad (2.48)$$

This involves solving L equations but let us consider one equation which can be given as

$$\underline{y}_k - \boldsymbol{\chi}_i^H \mathbf{b}_k - \|\boldsymbol{\chi}_i\|_{\mathbf{W}_k}^2 = 0 \quad (2.49)$$

where \underline{y}_k is the k^{th} element of $\underline{\mathbf{y}}_i$ and \mathbf{b}_k is the k^{th} column of \mathbf{B}^T . Taking Hermitian transpose of both sides, we get

$$\underline{y}_k^* - \mathbf{b}_k^H \boldsymbol{\chi}_i - \|\boldsymbol{\chi}_i\|_{\mathbf{W}_k^H}^2 = 0 \quad (2.50)$$

Adding equations (2.49) and (2.50)

$$R_k = 2Re(\underline{y}_k) - \mathbf{b}_k^H \boldsymbol{\chi}_i - \boldsymbol{\chi}_i^H \mathbf{b}_k - \|\boldsymbol{\chi}_i\|_{(\mathbf{W}_k + \mathbf{W}_k^H)}^2 \quad (2.51)$$

Let α and β be such that

$$\alpha + \beta = 2Re(\underline{y}_k)$$

Then

$$R_k = \alpha - \mathbf{b}_k^H \mathcal{X}_i - \mathcal{X}_i^H \mathbf{b}_k + \|\mathcal{X}_i\|_{\left(\frac{\beta}{\mathcal{E}_{XN}} \mathbf{I}_N - \mathbf{W}_k - \mathbf{W}_k^H\right)}^2 \quad (2.52)$$

where we used the fact that

$$\beta = \|\mathcal{X}_i\|_{\frac{\beta}{\mathcal{E}_{XN}} \mathbf{I}_N}^2$$

By completing squares [89], we get

$$R_k = (\mathcal{X}_i - \mathbf{E}^{-1} \mathbf{b}_k)^H \mathbf{E} (\mathcal{X}_i - \mathbf{E}^{-1} \mathbf{b}_k) + \mathbf{b}_k^H \mathbf{E}^{-1} \mathbf{b}_k + \alpha \quad (2.53)$$

where

$$\mathbf{E} = \left(\frac{\beta}{\mathcal{E}_{XN}} \mathbf{I}_N - \mathbf{W}_k - \mathbf{W}_k^H \right)$$

To make equation (2.53) a perfect square, α should satisfy the following relation

$$\begin{aligned} \mathbf{b}_k^H \mathbf{E}^{-1} \mathbf{b}_k - \alpha &= 0 \\ \mathbf{b}_k^H \mathbf{E}^{-1} \mathbf{b}_k &= \alpha \\ \mathbf{b}_k^H \left(\frac{\beta}{\mathcal{E}_{XN}} \mathbf{I}_N - \mathbf{W}_k - \mathbf{W}_k^H \right)^{-1} \mathbf{b}_k &= 2\text{Re}(\underline{y}_k) - \beta \end{aligned} \quad (2.54)$$

The above equation can be solved for β as it is the only variable unknown. Thus equation (2.53) can now be given as

$$\begin{aligned} R_k &= (\mathcal{X}_i - \mathbf{E}^{-1}\mathbf{b}_k)^H \mathbf{E}(\mathcal{X}_i - \mathbf{E}^{-1}\mathbf{b}_k) \\ R_k &= \|(\mathcal{X}_i - \mathbf{E}^{-1}\mathbf{b}_k)\|^2 \end{aligned} \quad (2.55)$$

Now, we have L such quadratic terms and their sum is given by [89]

$$\sum_{k=1}^L R_k = (\mathcal{X}_i - \mathbf{m}_c)^T \Sigma_c^{-1} (\mathcal{X}_i - \mathbf{m}_c) \quad (2.56)$$

where

$$\begin{aligned} \Sigma_c^{-1} &= \mathbf{E}_1 + \mathbf{E}_2 + \cdots + \mathbf{E}_L \\ \text{and } \mathbf{m}_c &= \Sigma_c(\mathbf{b}_1 + \mathbf{b}_2 + \cdots + \mathbf{b}_L) \\ &= (\mathbf{E}_1 + \mathbf{E}_2 + \cdots + \mathbf{E}_L)^{-1}(\mathbf{b}_1 + \mathbf{b}_2 + \cdots + \mathbf{b}_L) \end{aligned}$$

Thus \mathbf{m}_c would be our required solution. The problem in this method lies in the fact that in finding β from equation (2.54), we again end up with exhaustive search as there will be $N + 1$ solutions.

2.6.3 Using Search Algorithms

We can use the search algorithms like Particle Swarm Optimization (PSO) [81], [82], [83], and the Genetic Algorithm (GA) [84], [85], to directly solve the nonlinear prob-

lem (equation (2.39)). PSO and GA are widely used algorithms to solve nonlinear problems. PSO and GA are motivated by the evolution of nature. Depending on the number of variables in the problem, a population of individuals is generated. The rule of survival of the fittest is used to manipulate the population by cooperation and competition within the individuals in case of PSO, and by using genetic operators like mutation, crossover and reproduction in case of GA. The best solution is selected from the generations.

The data estimated by using the linearization approach above can be used to initialize PSO or GA. This initialization, with close to optimal solution, will help to kick start them for better results.

2.6.4 Reduced Exhaustive Search Algorithm

An alternative approach to blind data detection is to pursue an iterative data detection/channel estimation approach. Since the channel is of (maximum) length $L + 1$, we only need $L + 1$ data symbols to perform channel estimation. Given the N data symbols, which $L + 1$ symbols should we look for? To decide on this, consider the estimate (2.40) reproduced here for convenience

$$\begin{aligned}
\hat{\boldsymbol{x}}_i^{ML} &= \arg \min_{\boldsymbol{x}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{x}_i^* - \boldsymbol{C}\boldsymbol{x}_i^* \right\|^2 \\
&= \arg \min_{\boldsymbol{x}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{x}_i^* - \boldsymbol{c} \right\|^2 \\
&= \arg \min_{\boldsymbol{x}_i} \left\| \left(\underline{\boldsymbol{y}}_i - \boldsymbol{z} \right) - \boldsymbol{B}\boldsymbol{x}_i^* - \bar{\boldsymbol{c}} \right\|^2
\end{aligned} \tag{2.57}$$

where \mathbf{z} is defined in (2.44) and where

$$\bar{c}(j) = \|\mathbf{x}_i\|_{\overline{\mathbf{W}}_j}^2 \quad (2.58)$$

where $\overline{\mathbf{W}}_j$ is \mathbf{W}_j with all zero diagonal.

The $L + 1$ data symbols we should look for are the most significant symbols. This is determined by the symbols interacting with the largest coefficients in (2.57). Specifically, define

$$\mathbf{b} = \text{vec}(\mathbf{B}) \quad \text{and} \quad \overline{\mathbf{w}}_i = \text{vec}(\overline{\mathbf{W}}_i) \quad 1 \leq i \leq L \quad (2.59)$$

and find the largest element in these $L + 1$ vectors. This largest element interacts with at most two data symbols. Retain this one or two symbols and all coefficients in \mathbf{B} and $\overline{\mathbf{W}}_i$ that operate on this symbol (or symbols). Now, find the next largest element in (2.59) and determine the data symbols (mostly two) that interact with this coefficient. With this procedure, (2.57) is approximated with an optimization that looks for the most significant $L + 1$ data symbols by exhaustive search.

These $L + 1$ symbols of \mathbf{x}_i are then used to estimate the channel $\underline{\mathbf{h}}_i$ using the following input/output equation

$$\mathbf{y}_{L+1} = \text{diag}(\mathbf{x}_{L+1}) \mathbf{Q}_{L+1}^* \underline{\mathbf{h}}_i + \mathcal{N}_i \quad (2.60)$$

where \mathbf{x}_{L+1} are the $L + 1$ elements detected using the above method and \mathbf{y}_{L+1} are the $L + 1$ elements of \mathbf{y} corresponding to \mathbf{x}_{L+1} .

This complete procedure of data detection and channel estimation is then repeated for a desired number of iterations.

2.6.5 Using Pilots and Frequency Correlation

The channel estimate obtained in the previous subsection can be further improved in the presence of pilots and/or frequency correlation about the channel. Thus let \mathbf{R} be the channel correlation matrix and let

$$\mathbf{y}_p = \text{diag}(\mathbf{x}_p) \mathbf{Q}_{L+1}^* \underline{\mathbf{h}}_i + \mathcal{N}_p \quad (2.61)$$

be a subsystem of (2.17) corresponding to the pilot locations. Then (2.60) and (2.61) can be concatenated into a single system of equations and the channel $\underline{\mathbf{h}}_i$ can be obtained by solving the regularized least squares problem

$$\hat{\underline{\mathbf{h}}}_i = \arg \min_{\underline{\mathbf{h}}_i} \left\| \begin{bmatrix} \mathbf{y}_{L+1} \\ \mathbf{y}_p \end{bmatrix} - \begin{bmatrix} \text{diag}(\mathbf{x}_{L+1}) \\ \text{diag}(\mathbf{x}_p) \end{bmatrix} \mathbf{Q}_{L+1}^* \right\|_{\sigma_n^{-2} \mathbf{I}}^2 + \|\underline{\mathbf{h}}_i\|_{\mathbf{R}^{-1}}^2 \quad (2.62)$$

2.6.6 Using Newton's Method

Let

$$\mathbf{Z} = \left\| \underline{\mathbf{y}}_i - \mathbf{B} \mathbf{x}_i^* - \mathbf{C} \mathbf{x}_i^* \right\|^2$$

subject to $|\mathbf{x}_i|^2 = 1$ (2.63)

be the cost function to be minimized. If initial estimate of data $\boldsymbol{\mathcal{X}}_{-1}$ is available, then it can be refined by applying Newton's Method [92] given by

$$\boldsymbol{\mathcal{X}}_k = \boldsymbol{\mathcal{X}}_{k-1} - \mu [\nabla^2 \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})]^{-1} [\nabla \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})]^*, \quad k \geq 0 \quad (2.64)$$

where μ (called step size) is a small number, ∇ stands for gradient, and ∇^2 stands for hessian of cost function \boldsymbol{Z} subjected to the constant modulus constraint. The algorithm runs iteratively till a maximum number of iterations or a stopping criteria is reached. To implement Newton's method, we need to calculate gradient and hessian of the cost function.

Evaluating the Gradient

We evaluate the gradient of cost function and the constant modulus constraint on data separately.

Gradient of Cost Function The cost function \boldsymbol{Z} can be written as

$$\begin{aligned} \boldsymbol{Z} &= \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^* \right\|^2 \\ &= \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c} \right\|^2 \\ &= \|\boldsymbol{a}\|^2 \\ &= \boldsymbol{a}^H \boldsymbol{a} \end{aligned} \quad (2.65)$$

To find the gradient of \boldsymbol{Z} , we have to differentiate it with $\boldsymbol{\mathcal{X}}_i^*$ [86] (we consider $\boldsymbol{\mathcal{X}}_i^*$

a row vector while differentiation)

$$\begin{aligned}
\frac{\partial \mathbf{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \frac{\partial [\mathbf{a}^H \mathbf{a}]}{\partial \boldsymbol{\mathcal{X}}_i^*} \\
&= \left[\frac{\partial(\mathbf{a}^H \mathbf{a})}{\partial \mathbf{a}} \right] \left[\frac{\partial \mathbf{a}}{\partial \boldsymbol{\mathcal{X}}_i^*} \right] + \left[\frac{\partial(\mathbf{a}^H \mathbf{a})}{\partial \mathbf{a}^*} \right] \left[\frac{\partial \mathbf{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*} \right] \\
&= \mathbf{a}^H \left[\frac{\partial \mathbf{a}}{\partial \boldsymbol{\mathcal{X}}_i^*} \right] + \mathbf{a}^T \left[\frac{\partial \mathbf{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*} \right]
\end{aligned} \tag{2.66}$$

where the second line follows from the chain rule for complex matrices [86], given by

$$\frac{\partial \mathbf{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} = \left[\frac{\partial \mathbf{Z}}{\partial \mathbf{a}} \right] \left[\frac{\partial \mathbf{a}}{\partial \boldsymbol{\mathcal{X}}_i^*} \right] + \left[\frac{\partial \mathbf{Z}}{\partial \mathbf{a}^*} \right] \left[\frac{\partial \mathbf{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*} \right]$$

Now, as can be seen from equation (2.66), we need to find differential of \mathbf{a} and \mathbf{a}^*

w.r.t $\boldsymbol{\mathcal{X}}_i^*$

$$\begin{aligned}
\frac{\partial \mathbf{a}}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \frac{\partial [\underline{\mathbf{y}}_i - \mathbf{B} \boldsymbol{\mathcal{X}}_i^* - \mathbf{c}]}{\partial \boldsymbol{\mathcal{X}}_i^*} \\
&= \frac{\partial \underline{\mathbf{y}}_i}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial(\mathbf{B} \boldsymbol{\mathcal{X}}_i^*)}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i^*} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_1 \boldsymbol{\mathcal{X}}_i^* \\ \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_2 \boldsymbol{\mathcal{X}}_i^* \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_L \boldsymbol{\mathcal{X}}_i^* \end{bmatrix} \\
&= -\mathbf{B} - \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_1 \\ \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \mathbf{W}_L \end{bmatrix}
\end{aligned} \tag{2.67}$$

and

$$\begin{aligned}
\frac{\partial \mathbf{a}^*}{\partial \mathbf{x}_i^*} &= \frac{\partial [\mathbf{y}_i^* - \mathbf{B}^* \mathbf{x}_i - \mathbf{c}^*]}{\partial \mathbf{x}_i^*} \\
&= \frac{\partial \mathbf{y}_i^*}{\partial \mathbf{x}_i^*} - \frac{\partial (\mathbf{B}^* \mathbf{x}_i)}{\partial \mathbf{x}_i^*} - \frac{\partial}{\partial \mathbf{x}_i^*} \begin{bmatrix} \mathbf{x}_i^H \mathbf{W}_1^* \mathbf{x}_i \\ \mathbf{x}_i^H \mathbf{W}_2^* \mathbf{x}_i \\ \vdots \\ \mathbf{x}_i^H \mathbf{W}_L^* \mathbf{x}_i \end{bmatrix} \\
&= - \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1^H \\ \mathbf{x}_i^T \mathbf{W}_2^H \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L^H \end{bmatrix}
\end{aligned} \tag{2.68}$$

Substituting these values in equation (2.66), we get

$$\begin{aligned}
\frac{\partial \mathbf{Z}}{\partial \mathbf{x}_i^*} &= \mathbf{a}^H \left(-\mathbf{B} - \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1 \\ \mathbf{x}_i^T \mathbf{W}_2 \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L \end{bmatrix} \right) + \mathbf{a}^T \left(- \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1^H \\ \mathbf{x}_i^T \mathbf{W}_2^H \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L^H \end{bmatrix} \right) \\
&= -\mathbf{a}^H \mathbf{B} - \mathbf{a}^H \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1 \\ \mathbf{x}_i^T \mathbf{W}_2 \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L \end{bmatrix} - \mathbf{a}^T \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1^H \\ \mathbf{x}_i^T \mathbf{W}_2^H \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L^H \end{bmatrix}
\end{aligned} \tag{2.69}$$

which is a vector of size $1 \times N$.

Gradient of Constraint The constant modulus constraint on data can be written

as

$$\Psi = \|\mathcal{E}_X - \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_i \boldsymbol{\mathcal{X}}_i\|^2 \quad (2.70)$$

where \mathbf{E}_i is a $N \times N$ matrix given by

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{E}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \cdots, \quad \mathbf{E}_N = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Equation (2.70) can also be written as

$$\begin{aligned} \Psi &= \|\mathcal{E}_X - \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_i \boldsymbol{\mathcal{X}}_i\|^2 \\ &= \|\mathbf{b}\|^2 \\ &= \mathbf{b}^H \mathbf{b} \end{aligned} \quad (2.71)$$

which is similar to equation (2.65). So, the gradient of constant modulus constraint is calculated similar to the cost function and is given as

$$\frac{\partial \Psi}{\partial \mathbf{x}_i^*} = -\mathbf{b}^H \begin{bmatrix} \mathbf{x}_i^T \mathbf{E}_1^T \\ \mathbf{x}_i^T \mathbf{E}_2^T \\ \vdots \\ \mathbf{x}_i^T \mathbf{E}_N^T \end{bmatrix} - \mathbf{b}^T \begin{bmatrix} \mathbf{x}_i^T \mathbf{E}_1^* \\ \mathbf{x}_i^T \mathbf{E}_2^* \\ \vdots \\ \mathbf{x}_i^T \mathbf{E}_N^* \end{bmatrix} \quad (2.72)$$

which is a vector of size $1 \times N$.

So, the gradient of cost function subjected to the constraint is given by

$$\nabla(\mathbf{Z}) = \frac{\partial \mathbf{Z}}{\partial \mathbf{x}_i^*} + \frac{1}{\sigma_n^2} \frac{\partial \Psi}{\partial \mathbf{x}_i^*} \quad (2.73)$$

Evaluating the Hessian

Similar to gradient, we evaluate the hessian of cost function and constraint separately.

Hessian of Cost Function The Hessian of cost function is given by

$$\frac{\partial}{\partial \mathbf{x}_i} \left(\frac{\partial \mathbf{Z}}{\partial \mathbf{x}_i^*} \right)^T = \frac{\partial}{\partial \mathbf{x}_i} \left(-\mathbf{a}^H \mathbf{B} - \mathbf{a}^H \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1 \\ \mathbf{x}_i^T \mathbf{W}_2 \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L \end{bmatrix} - \mathbf{a}^T \begin{bmatrix} \mathbf{x}_i^T \mathbf{W}_1^H \\ \mathbf{x}_i^T \mathbf{W}_2^H \\ \vdots \\ \mathbf{x}_i^T \mathbf{W}_L^H \end{bmatrix} \right)^T$$

which can be expanded to get the following

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}_i} \left(\frac{\partial \mathbf{Z}}{\partial \boldsymbol{x}_i^*} \right)^T &= - \frac{\partial}{\partial \boldsymbol{x}_i} (\mathbf{B}^T \mathbf{a}^*) \\
&- \frac{\partial}{\partial \boldsymbol{x}_i} \left(\begin{bmatrix} \mathbf{w}_1^T \boldsymbol{x}_i & \cdots & \mathbf{w}_L^T \boldsymbol{x}_i \end{bmatrix} \mathbf{a}^* \right) \\
&- \frac{\partial}{\partial \boldsymbol{x}_i} \left(\begin{bmatrix} \mathbf{w}_1^* \boldsymbol{x}_i & \cdots & \mathbf{w}_L^* \boldsymbol{x}_i \end{bmatrix} \mathbf{a} \right) \quad (2.74)
\end{aligned}$$

Now, we apply chain rule of complex matrices (explained in the previous section) to each term of equation (2.74). The final result obtained is given by

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}_i} \left(\frac{\partial \mathbf{Z}}{\partial \boldsymbol{x}_i^*} \right)^T &= - \mathbf{B}^T \left(-\mathbf{B}^* - \begin{bmatrix} \boldsymbol{x}_i^H \mathbf{w}_1^* \\ \boldsymbol{x}_i^H \mathbf{w}_2^* \\ \vdots \\ \boldsymbol{x}_i^H \mathbf{w}_L^* \end{bmatrix} \right) \\
&- \begin{bmatrix} \mathbf{w}_1^T \boldsymbol{x}_i & \cdots & \mathbf{w}_L^T \boldsymbol{x}_i \end{bmatrix} \left(-\mathbf{B}^* - \begin{bmatrix} \boldsymbol{x}_i^H \mathbf{w}_1^* \\ \boldsymbol{x}_i^H \mathbf{w}_2^* \\ \vdots \\ \boldsymbol{x}_i^H \mathbf{w}_L^* \end{bmatrix} \right) \\
&- \begin{bmatrix} \mathbf{w}_1^* \boldsymbol{x}_i & \cdots & \mathbf{w}_L^* \boldsymbol{x}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_i^H \mathbf{w}_1^T \\ \boldsymbol{x}_i^H \mathbf{w}_2^T \\ \vdots \\ \boldsymbol{x}_i^H \mathbf{w}_L^T \end{bmatrix} \quad (2.75)
\end{aligned}$$

which is a matrix of size $N \times N$.

Hessian of Constraint The Hessian of constant modulus constraint on data is calculated in a similar fashion to cost function and the final result is given as under

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i} \left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*} \right)^T = & \begin{bmatrix} \mathbf{E}_1 \boldsymbol{\mathcal{X}}_i & \cdots & \mathbf{E}_N \boldsymbol{\mathcal{X}}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_1^H \\ \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_2^H \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_N^H \end{bmatrix} \\ & + \begin{bmatrix} \mathbf{E}_1^H \boldsymbol{\mathcal{X}}_i & \cdots & \mathbf{E}_N^H \boldsymbol{\mathcal{X}}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_1 \\ \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^H \mathbf{E}_N \end{bmatrix} \end{aligned} \quad (2.76)$$

which is also a matrix of size $N \times N$.

Thus, the Hessian of the cost function subjected to the constant modulus constraint on data is given by

$$\nabla^2(\mathbf{Z}) = \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i} \left(\frac{\partial \mathbf{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} \right)^T + \frac{1}{\sigma_n^2} \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i} \left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*} \right)^T \quad (2.77)$$

2.7 Enhanced Equalization Using CP

We have so far seen how the CP can be used for blind data detection. This can aid with reducing the training overhead and dealing with the presence of channel zeros on the FFT grid. However, can the CP be of value when the channel is known perfectly or an estimate of it is available at the receiver?

We show in this section how this is possible. To this end, consider the input/output equations of the circular and linear subchannels ((2.14) and (2.19)), reproduced here for convenience

$$\mathbf{y}_i = \mathcal{H}_i \odot \mathbf{x}_i + \mathcal{N}_i = \text{diag}(\mathcal{H}_i)\mathbf{x}_i + \mathcal{N}_i \quad (2.78)$$

$$\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_i \underline{\mathbf{h}}_i + \underline{\mathbf{n}}_i \quad (2.79)$$

Now, decompose $\underline{\mathbf{X}}_i$ as $(\underline{\mathbf{X}}_{U_{i-1}} + \underline{\mathbf{X}}_{L_i})$ and move the known part $\underline{\mathbf{X}}_{U_{i-1}} \underline{\mathbf{h}}_i$ to the left hand side to get

$$\underline{\mathbf{y}}_i - \underline{\mathbf{X}}_{U_{i-1}} \underline{\mathbf{h}}_i = \underline{\mathbf{X}}_{L_i} \underline{\mathbf{h}}_i + \underline{\mathbf{n}}_i \quad (2.80)$$

and exchange the roles of $\underline{\mathbf{X}}_{L_i} \underline{\mathbf{h}}_i$ as

$$\underline{\mathbf{X}}_{L_i} \underline{\mathbf{h}}_i = \underline{\mathbf{H}}_{\perp} \underline{\mathbf{x}}_i \quad (2.81)$$

$$= \underline{\mathbf{H}}_{\perp} \mathbf{Q}_{N-L+1} \mathbf{x}_i \quad (2.82)$$

where the second line follows from the fact that

$$\mathbf{x}_i = \mathbf{Q}\mathcal{X}_i$$

and that $\underline{\mathbf{x}}_i$ consists of the last $L + 1$ elements of \mathbf{x}_i . Thus (2.80) can be rewritten as

$$\underline{\mathbf{y}}_i - \underline{\mathbf{X}}_{U_{i-1}}\underline{\mathbf{h}}_i = \underline{\mathbf{H}}_L\mathbf{Q}_{N-L+1}\mathcal{X}_i + \underline{\mathbf{n}}_i \quad (2.83)$$

Combining (2.78) and (2.83) yields an $N + L$ system of equations in the unknown OFDM symbol \mathcal{X}_i

$$\begin{bmatrix} \mathbf{y}_i \\ \underline{\mathbf{y}}_i - \underline{\mathbf{X}}_{U_{i-1}}\underline{\mathbf{h}}_i \end{bmatrix} = \begin{bmatrix} \text{diag}(\mathcal{H}_i) \\ \underline{\mathbf{H}}_L\mathbf{Q}_{N-L+1} \end{bmatrix} \mathcal{X}_i + \begin{bmatrix} \mathcal{N}_i \\ \underline{\mathbf{n}}_i \end{bmatrix} \quad (2.84)$$

This system can be solved for \mathcal{X}_i using least squares.

2.8 Simulation Results

This section is divided into two parts. In the first part, simulation results of the blind data detector and the approximate methods are discussed while in the second part, we provide results for the enhanced equalization using CP.

2.8.1 Blind Data Detector

We consider an OFDM system with $N = 16$ and cyclic prefix of length $L = 4$. The OFDM symbol consists of BPSK or 4-QAM symbols. The channel IR consists of 5 iid

Rayleigh fading taps which remains constant over one OFDM symbol. We compare the BER performance of three methods: (i) Perfectly known channel, (ii) Channel estimated using $L + 1$ pilots and (iii) Blind based estimation using exhaustive search of equation (2.39). In all the following cases, we assume that we know the previous symbol unless mentioned otherwise.

BER vs SNR comparison for BPSK modulated data

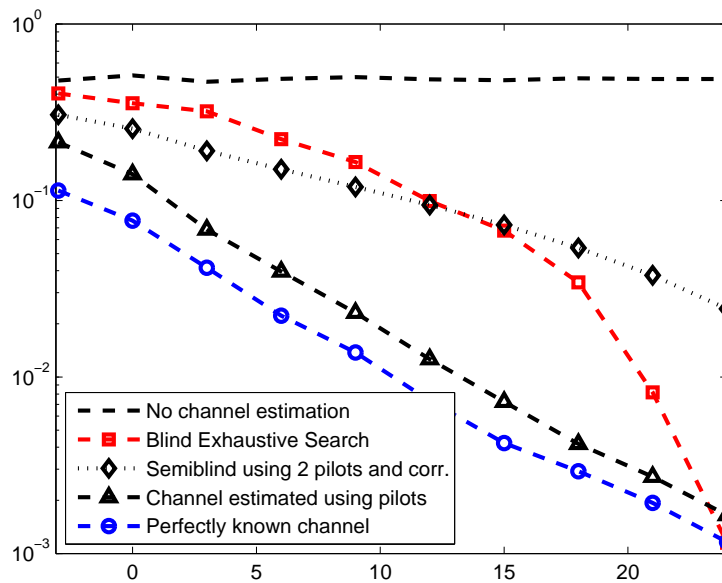


Figure 2.1: BER vs SNR for BPSK-OFDM over a Rayleigh channel

In Figure 2.1, we compare the three mentioned approaches, a semiblind least squares estimator using few pilots and frequency correlation, and an approach with no channel estimation, for BPSK modulated data over a Rayleigh fading channel. As expected, the best performance is achieved by the perfectly known channel, followed by that obtained by training based estimated channel. The approach with no channel

estimation appears to be flat at $\text{BER} = 0.5$ as the data is convolved with channel IR and detection is transformed to guessing without channel equalization. The semiblind approach performs better than the proposed blind approach at low SNR. Note however that in the high SNR region, the BER curve of blind based estimation exhibits steeper slope (higher diversity) which can be explained from the fact that the two subchannels (linear and circular) are used to detect the data in the blind case when only the linear subchannel is used in the pilot based and known channel cases. The presence of occasional nulls in the channel also makes the blind channel case better.

BER vs SNR comparison for 4-QAM modulated data

The same conclusion can be made for the 4-QAM input (see Figure 2.2). Here it can also be seen that the blind based estimation outperforms the pilot based and perfectly known channel cases at high SNR.

BER vs SNR comparison for BPSK modulated data with persistent channel nulls

In Figure 2.3, the three approaches are compared for BPSK modulated data when the channel IR has persistent zeros on the FFT grid. We note that at high SNR, the BER for perfectly known channel and that of the estimated channel reach an error floor. Our blind method does not suffer from this problem and thus blind case outperforms the other two cases when the channel has persistent nulls.

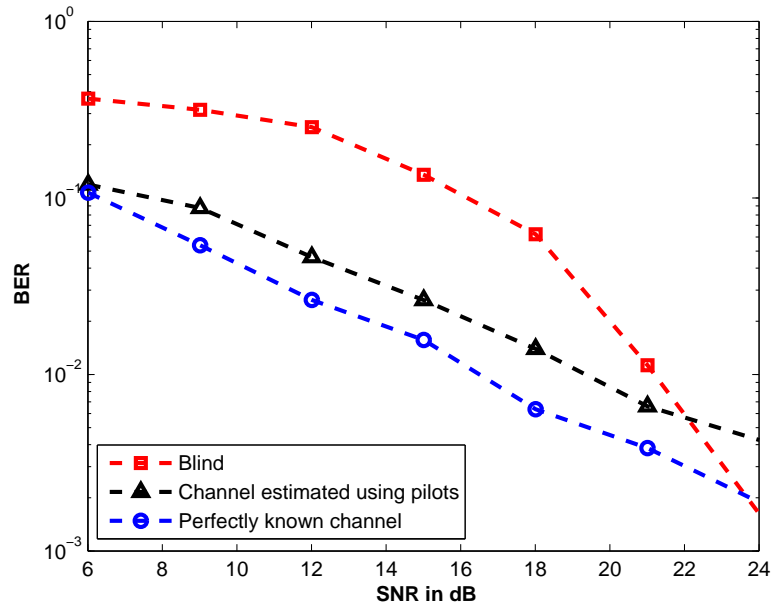


Figure 2.2: BER vs SNR for 4QAM-OFDM over a Rayleigh channel

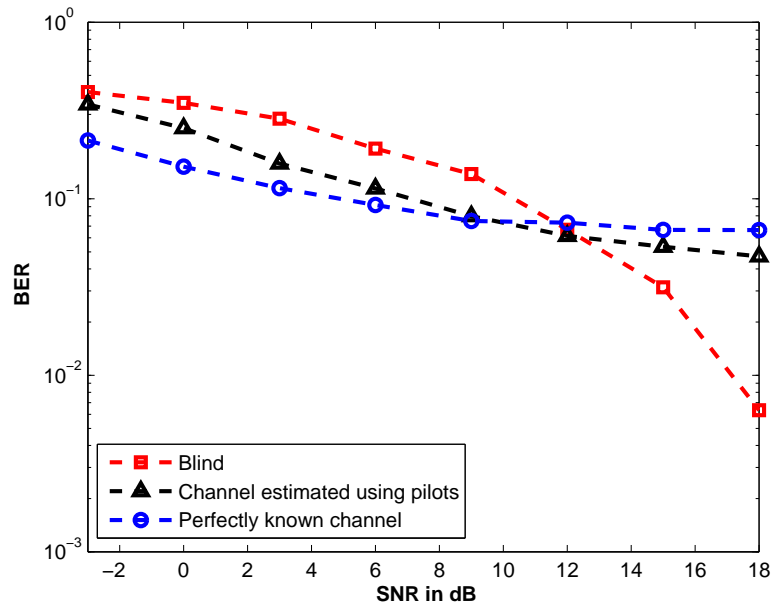


Figure 2.3: BER vs SNR for BPSK-OFDM over channel with zeros on FFT grid

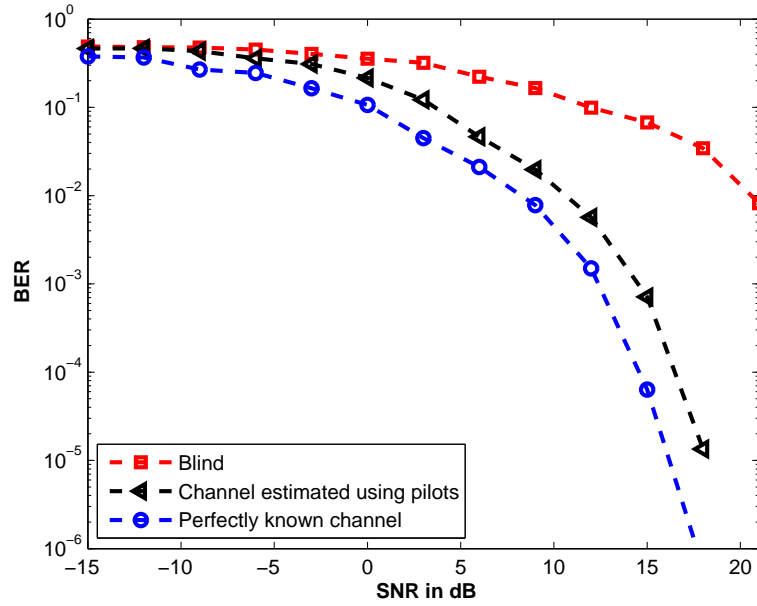


Figure 2.4: BER vs SNR for BPSK-OFDM over constant channel

BER vs SNR comparison for BPSK modulated data over a constant channel

To check whether occasional nulls are encountered in the Rayleigh channel, we compare the performance of the blind algorithm with the pilot based and perfectly known channel methods when the channel is constant with no zeros on FFT grid. Figure 2.4 indicates that when the channel is constant, the blind algorithm does not perform better than the pilot based or perfectly known channel methods even at high SNR. This proves that occasional nulls are present in the Rayleigh channel. The blind algorithm presented here is robust to channel nulls and thus outperforms even the perfectly known channel method in the practical scenario of random channel.

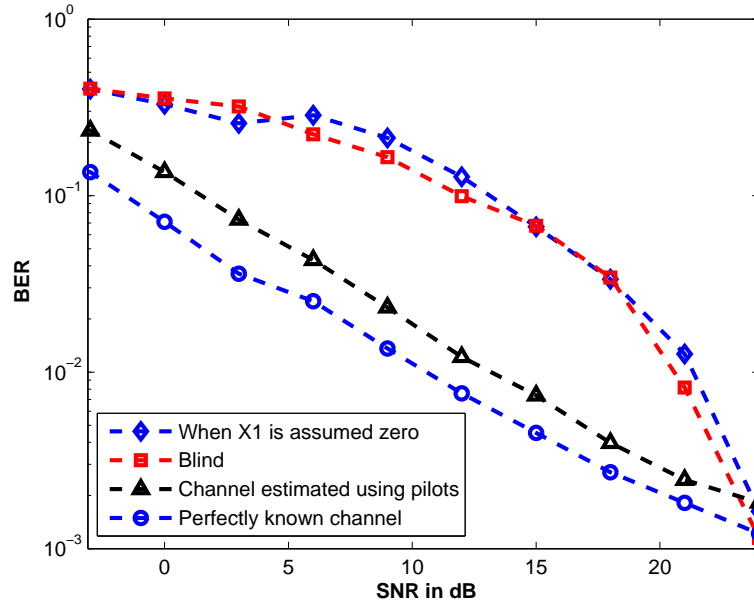


Figure 2.5: BER vs SNR for BPSK-OFDM with previous symbol assumed to be zero

BER vs SNR comparison for BPSK modulated data with previous symbol assumed to be zero

In Figure 2.5, contrary to the above cases, we assume the previous symbol to be zero and use the blind algorithm to detect the current symbol. The problem of sign ambiguity is faced in this scenario and we tackle it by assuming the first bit of current symbol to be known. This can be done by sending a pilot on the first bit of each symbol. Figure 2.5 clearly shows that the blind algorithm performs quite well in this scenario also.

Comparison of linearization approach and search algorithms

The low complexity algorithms proposed in Subsections 2.6.1 and 2.6.3 (linearization approach, PSO and GA) have been compared in Figure 2.6 for BPSK modulated data.

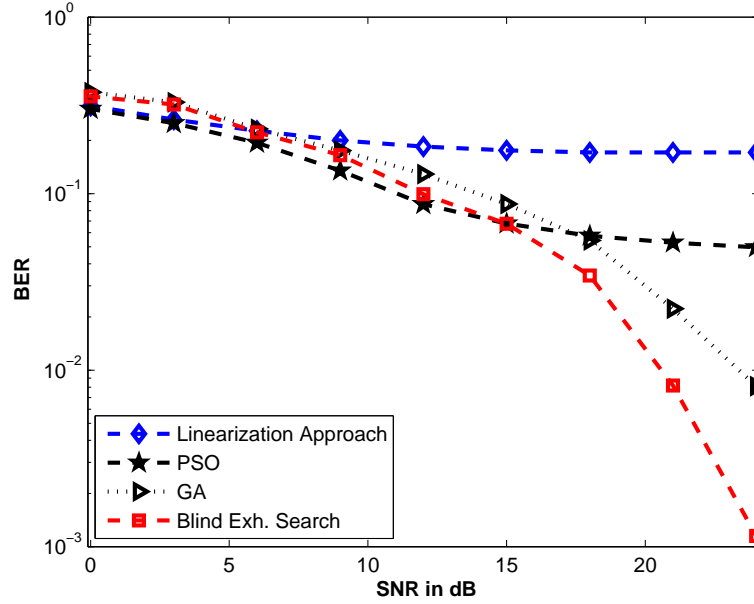


Figure 2.6: Comparison of low complexity algorithms for BPSK-OFDM over a Rayleigh channel

In the linearization approach, the estimate was refined for 1000 number of iterations. In GA, size of population used was 200 while the number of generations was fixed at 250. The algorithm was initialized by the estimate obtained from linearization approach. As the data is BPSK modulated, so minimization was performed subject to the constraint $|\mathcal{X}_i| = 1$. In PSO, the population size used was 300 and the particles were also initialized with the data estimated from linearization approach. We can observe in Figure 2.6 that PSO performs well at low SNR while GA performs quite close to the blind exhaustive search at low as well as high SNR.

Sensitivity of Reduced exhaustive search algorithm to number of iterations

The sensitivity of reduced $(L + 1)$ exhaustive search algorithms described in Subsections 2.6.4 and 2.6.5 to number of iterations is shown in Figure 2.7. Two pilots were

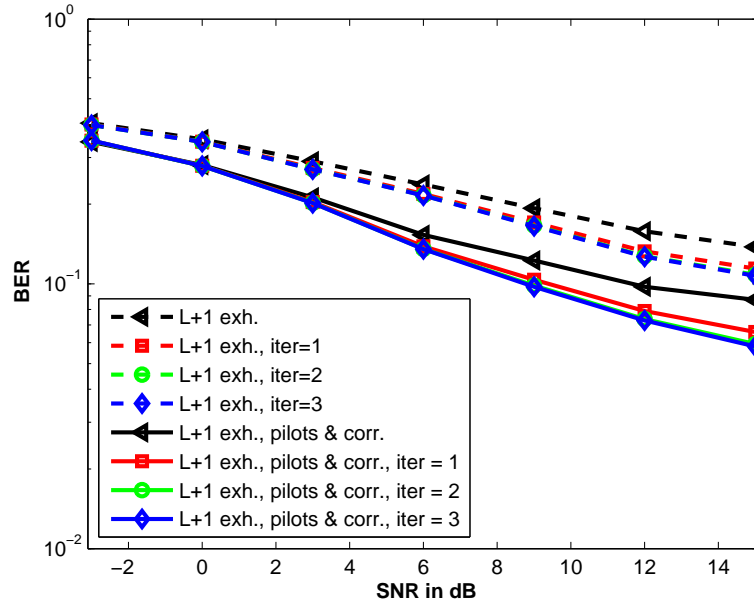


Figure 2.7: Sensitivity of Reduced search algorithms to number of iterations for BPSK-OFDM over a Rayleigh channel

used while the number of iterations were varied from 1 to 3. It can be observed that the first iteration yields substantial improvement in BER but iterating beyond that yields diminishing returns.

BER vs SNR Comparison of Reduced exhaustive search algorithms

In Figure 2.8, three variants of the reduced $(L + 1)$ exhaustive search algorithm (discussed in Subsections 2.6.4 and 2.6.5) have been compared with the N exhaustive search algorithm of equation (2.39). The three variants include: (i) $L + 1$ exhaustive search, (ii) $L + 1$ exhaustive search with pilots only, and (iii) $L + 1$ exhaustive search with both pilots and frequency correlation. The results are shown for two iterations and the number of pilots is also fixed at two. Figure 2.8 shows that $L + 1$ exhaustive search algorithm performs quite close to the N exhaustive search blind algorithm. We

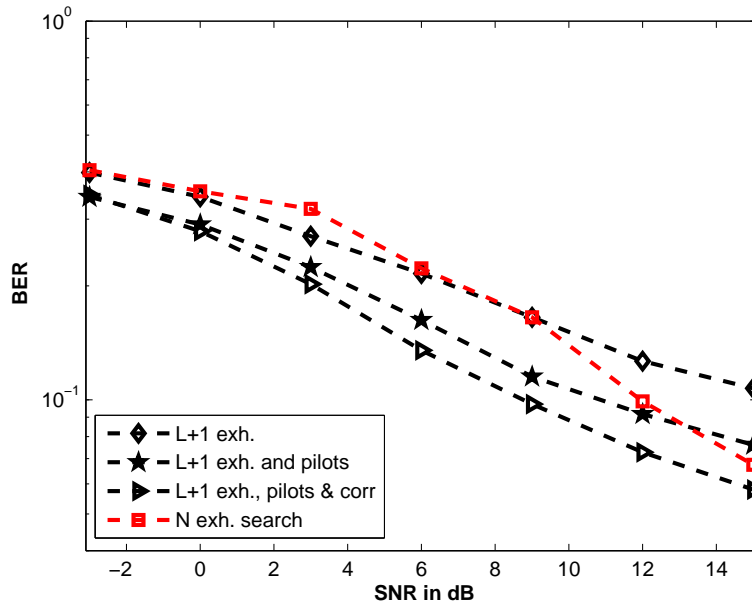


Figure 2.8: Comparison of Reduced search algorithms for BPSK-OFDM over a Rayleigh channel

can also notice that the reduced exhaustive search with pilots and frequency correlation performs better than the blind one as we are utilizing pilots and thus trading off with bandwidth efficiency.

BER vs SNR Comparison of Newton's Method

The Newton's method described in equation (2.64) was implemented using a step size of 0.5. The iterative algorithm was run till the difference between the value of current and previous cost function becomes less than 10^{-6} . Figure 2.9 shows the performance of Newton's method for 4-QAM with $N = 16$ and $L = 4$ when it is initialized by the estimate obtained by using 3 pilots and channel correlation. It can be seen that the 3 pilots based method reaches an error floor at high SNR while the Newton's method performs quite close to the blind exhaustive search. In Figure 2.10, the performance of

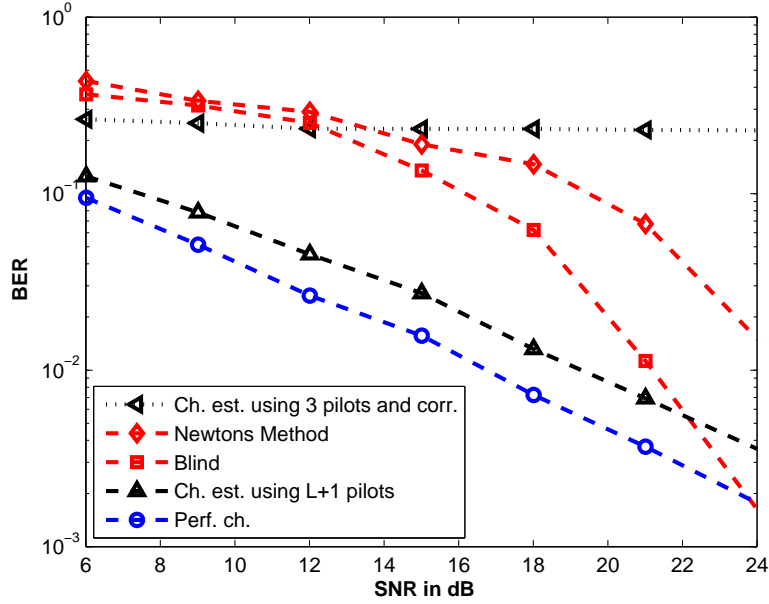


Figure 2.9: Comparison of Newton’s Method for 4QAM-OFDM with $N = 16$ and $L = 4$ over a Rayleigh channel

Newton’s method is compared with the $L + 1$ pilots case and perfectly known channel for a comparatively more realistic OFDM system using 4-QAM with $N = 64$ and $L = 16$. The Newton’s method is initialized with an estimate obtained by using 12 pilots and channel correlation. Figure 2.10 clearly indicates that Newton’s method performs quite well even for higher number of carriers.

2.8.2 Enhanced Equalization Using CP

We consider a realistic OFDM system with $N = 128$ and cyclic prefix of length $L = 32$. The channel IR consists of 33 iid Rayleigh fading taps and the OFDM symbol consists of BPSK or 16QAM modulated data. In this section, we assume that the receiver either knows the channel perfectly or estimates it using $L + 1$ pilots. We then compare the performance of the receiver in these two scenarios when (i)

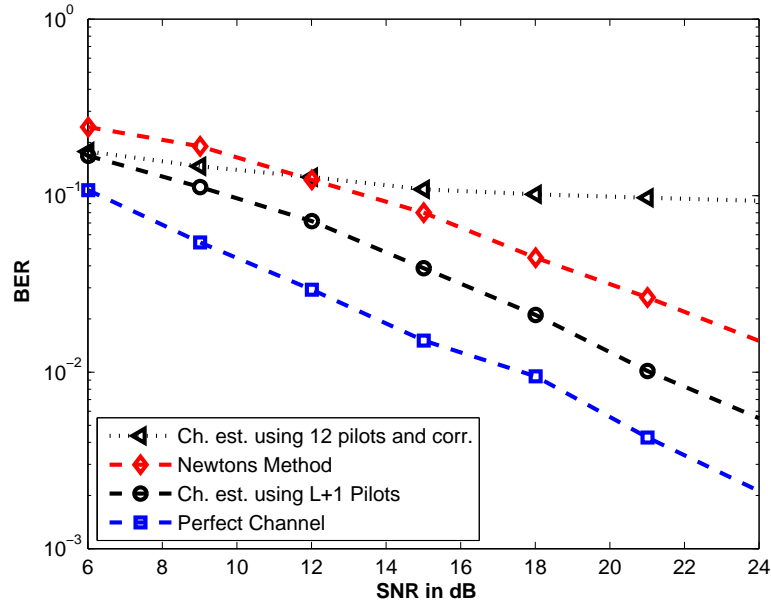


Figure 2.10: Comparison of Newton's Method for 4QAM-OFDM with $N = 64$ and $L = 16$ over a Rayleigh channel

data is detected using only the circular subchannel, (ii) data is detected using both the circular and linear subchannels with errors propagated (i.e. the error corrupted symbol detected in the current iteration is used as it is in the next iteration), and (iii) data is detected using both the channels with no errors propagated (i.e. we assume that we have detected the previous symbol perfectly in the next iterations).

BER vs SNR Comparison for BPSK-OFDM over a Rayleigh channel

In Figure 2.11, the performance of the receiver is compared for the above scenarios for BPSK-OFDM over a Rayleigh channel. It can be seen that the performance of the receiver improves when both subchannels are used for data recovery. The improvement is quite significant at high SNR. It can also be noticed that the case when errors are propagated does not perform well at low SNR but as the SNR increases, its

performance is improved and becomes equal to the case when no errors are propagated.

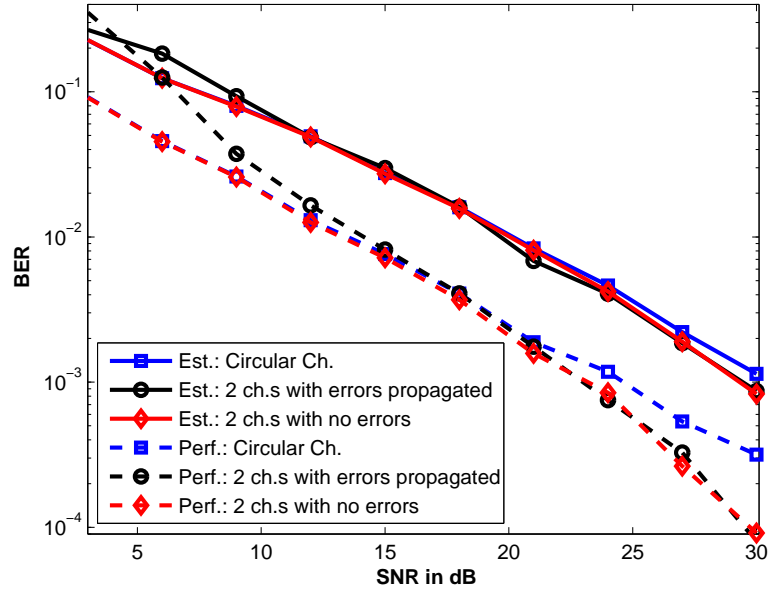


Figure 2.11: Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over a Rayleigh channel

BER vs SNR Comparison for BPSK-OFDM over channel with persistent nulls

In Figure 2.12, the performance of the receiver with enhanced equalization using CP is compared with the conventional one using only circular subchannel when channel IR has zeros on FFT grid. It can be observed that the case when data is estimated using only circular subchannel reaches an error floor as expected. No such error floor is observed if both the linear and circular subchannels are used for data detection. As noticed in Figure 2.11, the performance of the case when errors are propagated improves with increasing SNR.

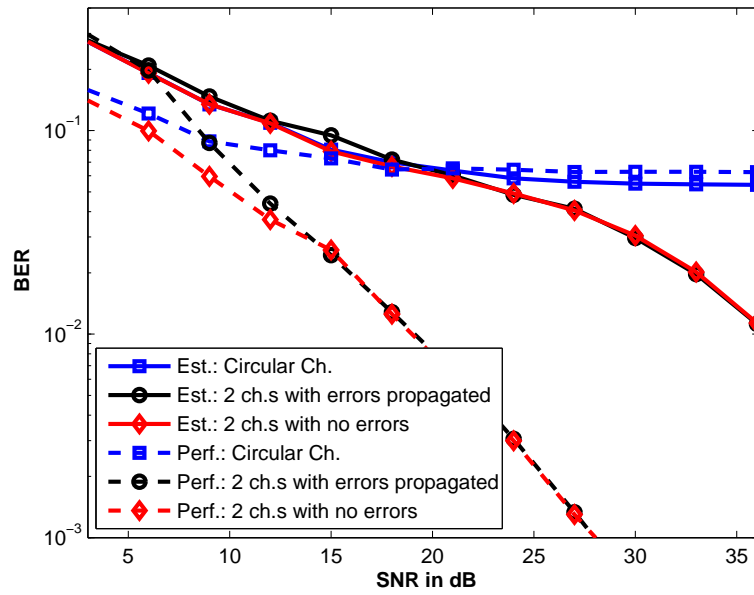


Figure 2.12: Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over channel with zeros on FFT grid

BER vs SNR Comparison for 16QAM-OFDM over a Rayleigh channel

Figure 2.13 shows the performance of the receiver with enhanced equalization using CP for 16QAM modulated (non-constant modulus) data over a Rayleigh channel. Similar to the BPSK modulated data case, the performance of the receiver when both subchannels are considered is better as compared to its performance when only circular subchannel is used for data recovery. The improvement is quite significant at high SNR.

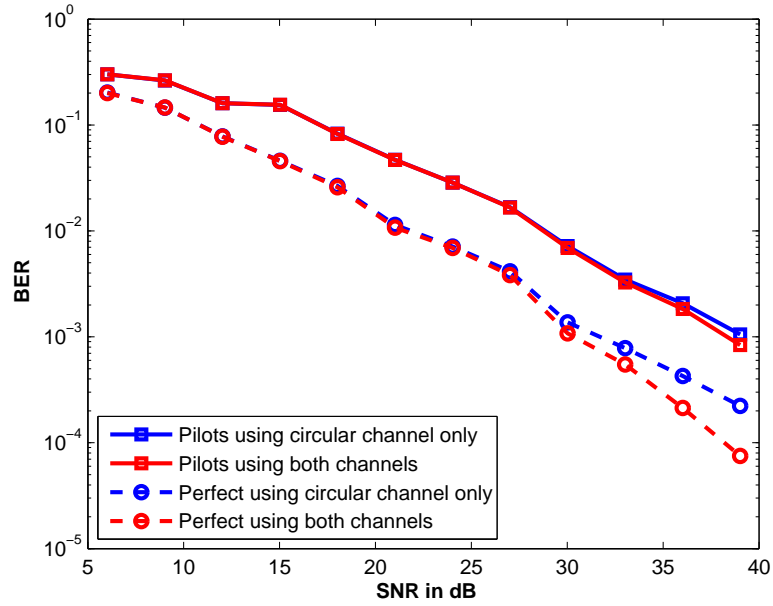


Figure 2.13: Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over a Rayleigh channel

BER vs SNR Comparison for 16QAM-OFDM over channel with persistent nulls

The performance of the receiver using the enhanced equalization using CP for 16QAM modulated data is shown in Figure 2.14. The case when only circular subchannel is used for data recovery flattens at high SNR while the case when both subchannel are used does not show any error floor.

2.9 Conclusion

In this chapter, we demonstrated how to perform blind ML data recovery in OFDM transmission. This is done using a single output OFDM symbol and associated CP. In particular, it was shown that the ML data estimate is the solution of an integer

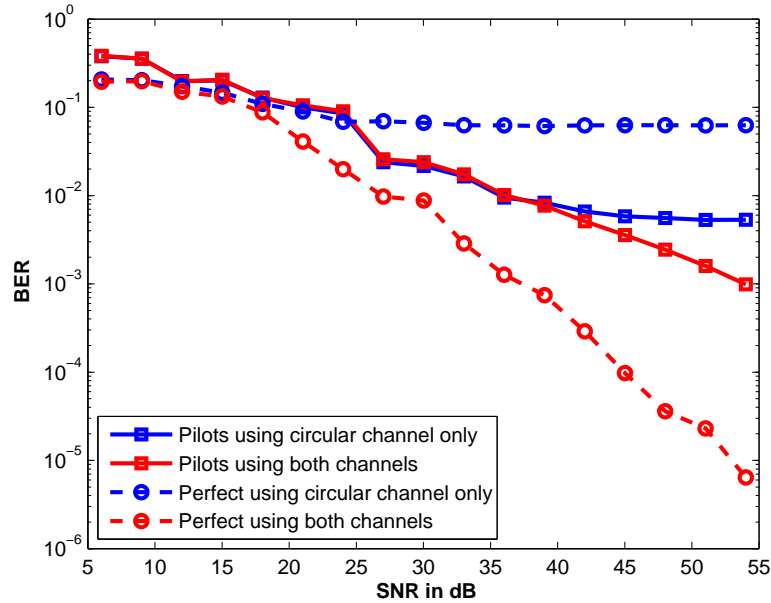


Figure 2.14: Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over channel with zeros on FFT grid

nonlinear-least squares problem which becomes simpler in the case of constant modulus data. We further proved that the data recovery is possible from output data only, irrespective of the channel zero locations and irrespective of the quality of the channel estimates or of its exact order.

We have also proposed approximate methods to reduce the exponential complexity entailed in the algorithm developed in the chapter. As is evident from the simulation results, these approximate methods perform quite close to the exhaustive search method especially at low SNR values. As all standard-based OFDM systems involve some form of training, we have also studied the behavior of the blind receiver in the presence of pilots and channel frequency correlation. It was found that Newton's method performs quite well at all values of SNR even for higher number of carriers.

A new method of enhanced equalization using CP was also proposed when the

receiver has perfect or estimated knowledge of channel. Specifically, in this method, data is recovered using both the linear and circular subchannels as opposed to the conventional method which utilized only circular subchannel. Simulation results proved that the proposed method performs better than the conventional one especially at high SNR values and when channel has zeros on FFT grid.

CHAPTER 3

CHANNEL-CENTERED BLIND ESTIMATION

3.1 Introduction

As opposed to the data-centric approach of Chapter 2, we adopt in this chapter a channel-centric approach for (semi) blind channel estimation and data detection in OFDM systems. In other words, we start by estimating channel and use it to estimate the data. The channel estimate is obtained by maximizing the log likelihood of the channel given the output data. Finding the likelihood function of a linear system can be very difficult. However, in the OFDM case, central limit arguments can be used to argue that the input is Gaussian [60]. Under the assumption that the noise is Gaussian, this makes the output Gaussian and allows us to easily write down the likelihood expression of the output. The likelihood function can then be maximized to obtain the ML estimate of the channel.

3.1.1 The Approach and Organization of the Chapter

We start by giving an overview of the system in Section 3.2. The log likelihood function is derived in Section 3.3 by using the Gaussian assumption on transmitted data. The channel can be estimated by maximizing the likelihood function. For blind estimation, the likelihood function should have global maxima when it is plotted against the channel taps. Experimental results of log likelihood function plot against channel taps are discussed in Section 3.3.2 which show that it is not unimodal. So channel cannot be estimated blindly and thus a semiblind approach is proposed by using steepest descent algorithm. This algorithm involves gradient of likelihood function with respect to channel taps. It is evaluated in Section 3.4 followed by the description of steepest descent algorithm in Section 3.5. The issue of computational complexity is discussed in Section 3.6 and simulation results are presented in Section 3.7.

3.2 System Overview

The OFDM system used here is similar to the previous chapter. Data is transmitted in symbols \mathbf{x}_i of length N each which undergo an IFFT operation to produce the time domain symbol, \mathbf{x}_i . A cyclic prefix of length L is appended to form the super-symbol $\bar{\mathbf{x}}_i$ as defined in equation (2.1) in the previous chapter. We assume an FIR channel of maximum length $L + 1$ given by

$$\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_L \end{bmatrix} \quad (3.1)$$

For reasons to be explained shortly, we will focus in this chapter on time domain signals. Here, the input/output relationship is given by

$$\begin{bmatrix} \underline{\mathbf{y}}_i \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} h_L & h_{L-1} & h_{L-2} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_L & h_{L-1} & \cdots & h_1 & h_0 & \cdots & 0 \\ & \ddots & & \ddots & \ddots & \ddots & & \\ 0 & 0 & 0 & \cdots & h_{L-1} & h_{L-2} & \cdots & h_0 \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_{i-1} \\ \underline{\mathbf{x}}_i \\ \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix} + \begin{bmatrix} \underline{\mathbf{n}}_i \\ \mathbf{n}_i \end{bmatrix}$$

or in matrix form

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \quad (3.2)$$

where \mathbf{n} is the output noise which we take to be white Gaussian. The matrices \mathbf{Y} , \mathbf{H} and \mathbf{X} are of size $(N + L) \times 1$, $(N + L) \times (N + 2L)$ and $(N + 2L) \times 1$, respectively.

3.3 Evaluating the Log Likelihood Function

To derive the likelihood function of a the output of a linear system, the input is assumed usually to be Gaussian (otherwise writing down the likelihood function can be very difficult). This is usually not true in a data communication system as the input is generated from a finite alphabet. Fortunately in an OFDM system, the time domain input can be assumed to be Gaussian by central limit theorem arguments.

Specifically, from the FFT relationship

$$\mathbf{x}_i = \sqrt{N}\mathbf{Q}\boldsymbol{\mathcal{X}}_i$$

we have the element by element relationship,

$$\mathbf{x}_i(1) = \sqrt{N}q_1\boldsymbol{\mathcal{X}}_i, \quad \mathbf{x}_i(2) = \sqrt{N}q_2\boldsymbol{\mathcal{X}}_i \dots \quad \mathbf{x}_i(N) = \sqrt{N}q_N\boldsymbol{\mathcal{X}}_i \quad (3.3)$$

where q_j are the rows of \mathbf{Q} . In other words, this shows that $\mathbf{x}_i(j)$ is a large (weighted) sum of iid random variables. The validity of this assumption is evident from the histogram plot shown in Figure 3.1 which describes the distribution of the transmitted data \mathbf{x}_i .

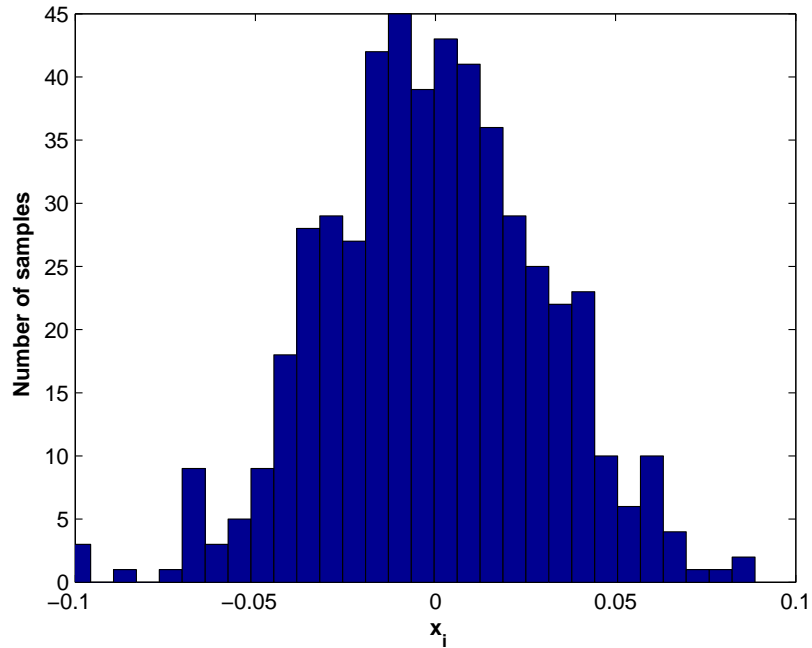


Figure 3.1: Number of samples vs transmitted data (\mathbf{x}_i)

Thus from this and the fact that noise is also Gaussian, we can conclude that

output \mathbf{Y} will also be Gaussian with pdf

$$\mathbf{Y} \sim \mathcal{N}(0, \Sigma_{\mathbf{Y}})$$

where $\Sigma_{\mathbf{Y}}$ is the second order moment of \mathbf{Y} which (from equation (3.2)) is given by

$$\Sigma_{\mathbf{Y}} = E[\mathbf{H}\mathbf{X}\mathbf{X}^*\mathbf{H}^*] + \sigma_n^2\mathbf{I} \quad (3.4)$$

$$= \mathbf{H}\Sigma_{\mathbf{X}}\mathbf{H}^* + \sigma_n^2\mathbf{I} \quad (3.5)$$

where $\Sigma_{\mathbf{X}}$ is a matrix of size $(N + 2L) \times (N + 2L)$ given by

$$\begin{aligned} \Sigma_{\mathbf{X}} &= \mathbf{E}[\mathbf{X}\mathbf{X}^H] \\ &= \mathbf{E} \begin{bmatrix} \underline{\mathbf{x}}_{i-1} \\ \underline{\mathbf{x}}_i \\ \tilde{\mathbf{x}}_i \\ \underline{\mathbf{x}}_i \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_{i-1}^H & \underline{\mathbf{x}}_i^H & \tilde{\mathbf{x}}_i^H & \underline{\mathbf{x}}_i^H \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_L & 0 & 0 & 0 \\ 0 & \mathbf{I}_L & 0 & \mathbf{I}_L \\ 0 & 0 & \mathbf{I}_{N-L} & 0 \\ 0 & \mathbf{I}_L & 0 & \mathbf{I}_L \end{bmatrix} \end{aligned} \quad (3.6)$$

The pdf of output \mathbf{Y} can thus be written as

$$P(\mathbf{Y}|\mathbf{h}) = \frac{1}{\det(\Sigma_{\mathbf{Y}})} \exp(-\mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y}) \quad (3.7)$$

So, the log likelihood function is given by

$$\mathcal{L}(\mathbf{Y}|\mathbf{h}) = -\ln \det (\boldsymbol{\Sigma}_{\mathbf{Y}}) - \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \quad (3.8)$$

3.3.1 Maximum Likelihood Estimation of the Channel IR

We can use the likelihood function derived above to obtain the ML estimate of the channel \mathbf{h} by maximizing it. i.e.

$$\begin{aligned} \hat{\mathbf{h}}_{\text{ML}} &= \max_{\mathbf{h}} \mathcal{L} \\ &= \max_{\mathbf{h}} -\ln \det (\boldsymbol{\Sigma}_{\mathbf{Y}}) - \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \end{aligned} \quad (3.9)$$

which depends only on the output data \mathbf{Y} and the channel \mathbf{h} (through the dependence of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ on \mathbf{h}).

This approach for channel estimation using the Gaussian input assumption is quite common in single carrier case, but has not been applied in the OFDM case. There are two disadvantages of applying it in the single carrier case [90], [91]:

- The method assumes that the input is Gaussian which is not the case in a single carrier system.
- Even if input is Gaussian, this method is usually phase blind i.e. it can only be used to identify minimum phase systems.

We avoid both of the problems in the OFDM case as the input is Gaussian by central limit theorem arguments and as the input is cyclostationary (due to the presence

of the cyclic prefix)[25].

Unfortunately, as we shall show next, the likelihood function is not unimodal (it could have local maxima) and so finding the global maxima might be challenging.

3.3.2 Plots of Likelihood Function vs Channel Taps

The likelihood function is plotted against the channel taps to investigate whether it has a global maxima. The input data is considered to be Gaussian of length $N = 64$ and a cyclic prefix of length $L = 2$ is used. Channel is considered to be an FIR of length $L + 1 = 3$ with first tap fixed at 1 to avoid sign ambiguity inherent in all blind techniques¹. In Figure 3.2 the log likelihood function is plotted against the remaining two channel taps h_1 and h_2 when $\sigma_n^2 = 1$. Figure 3.3 shows the top view of the same plot. Figures 3.4 and 3.5 show the plot of likelihood function and its top view respectively, when $\sigma_n^2 = 0.1$. These figures clearly indicate that the log likelihood function has multiple maxima.

This shows that a completely blind approach for channel estimation would be challenging. So, we pursue a semiblind approach in which we use Steepest Descent algorithm to estimate the channel.

3.4 Finding Gradient of \mathcal{L} w.r.t. h

As it is difficult to obtain the global maximum of the likelihood function, we will pursue a semiblind approach here where we use a few pilots to obtain an initial estimate of

¹A channel with only two effective taps is chosen so that we can plot the likelihood function against them in three dimensions.

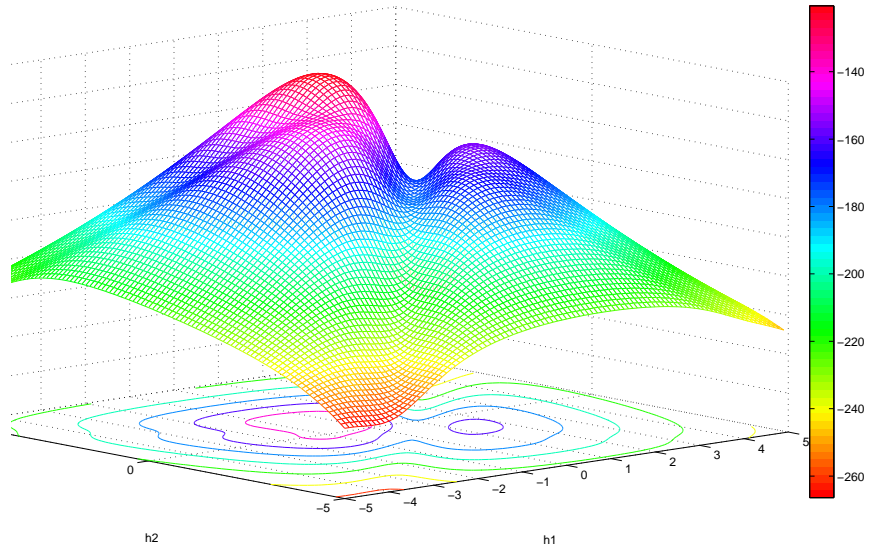


Figure 3.2: 3D plot of likelihood function against channel taps with $\sigma_n^2 = 1$

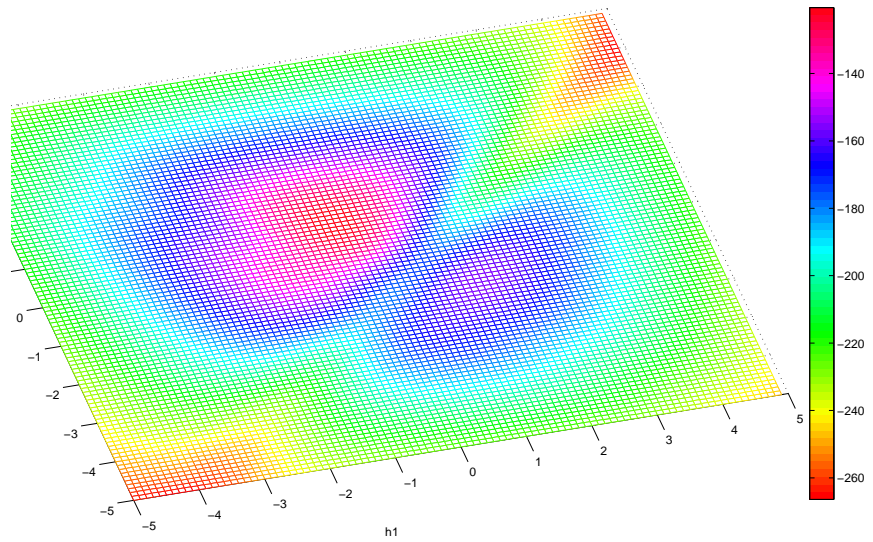


Figure 3.3: Top view 3D plot of likelihood function against channel taps with $\sigma_n^2 = 1$

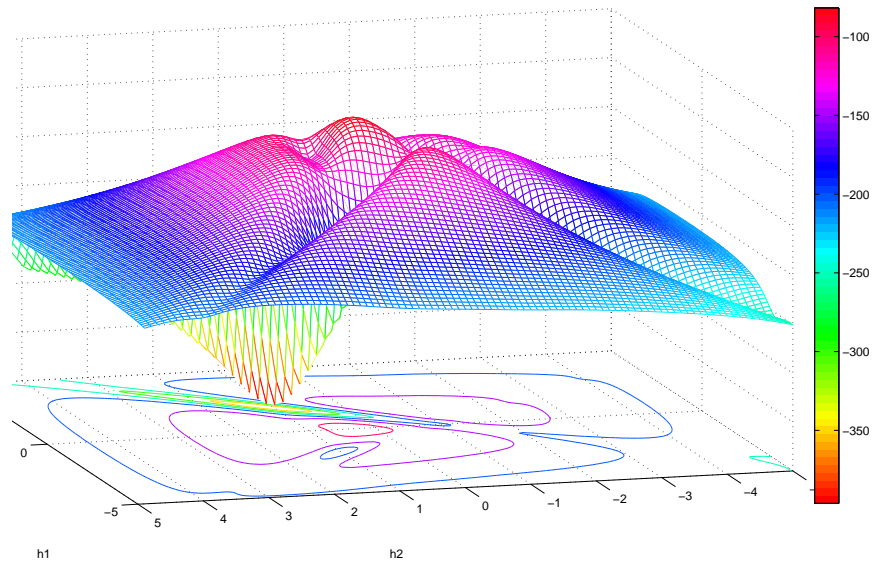


Figure 3.4: 3D plot of likelihood function against channel taps with $\sigma_n^2 = 0.1$

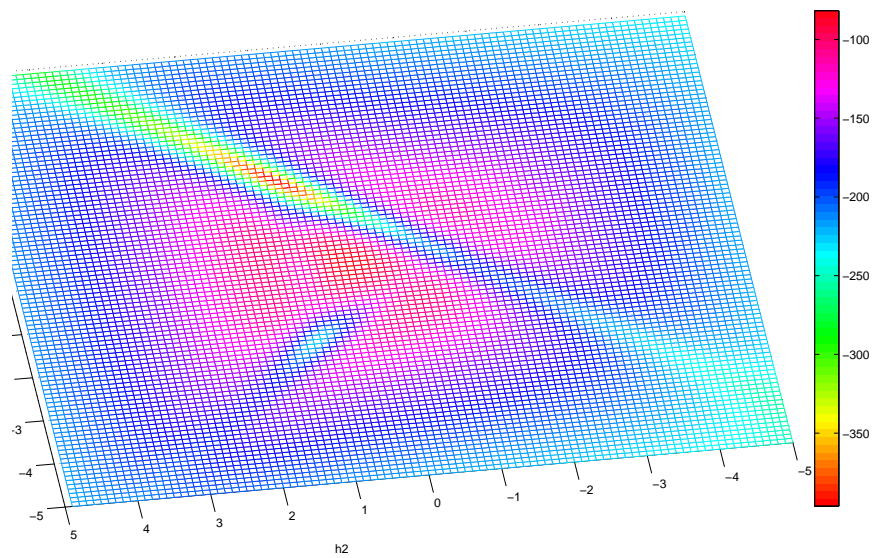


Figure 3.5: Top view 3D plot of likelihood function against channel taps with $\sigma_n^2 = 0.1$

the channel and subsequently improve the channel estimate using the steepest descent algorithm. To this end, we need to evaluate the gradient of the likelihood function with respect to the channel. We start by representing the channel convolution matrix \mathbf{H} in block form.

3.4.1 Writing \mathbf{H} in Block Form

While our approach works in the general case, we assume here that L (the CP length) divides N (OFDM symbol length) as this allows us to represent the matrices involved in block form. We emphasize however that our approach is valid in general. Now define,

$$\mathbf{B} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \cdots & 0 \\ \vdots & & \ddots & \\ h_{L-1} & h_{L-2} & \cdots & h_0 \end{bmatrix} \quad (3.10)$$

$$\mathbf{C} = \begin{bmatrix} h_L & h_{L-1} & \cdots & h_1 \\ 0 & h_L & \cdots & h_2 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & h_L \end{bmatrix} \quad (3.11)$$

then we can write the convolution matrix in the following block form

$$\mathbf{H} = \begin{bmatrix} \mathbf{C} & \mathbf{B} & & & & \\ & \mathbf{C} & \mathbf{B} & & & \\ & & \ddots & \ddots & & \\ & & & & \mathbf{C} & \mathbf{B} \end{bmatrix} \quad (3.12)$$

3.4.2 Evaluating Second Order Moment of Output \mathbf{Y}

As the log likelihood function (equation (3.8)) involves second order moment of output \mathbf{Y} , we want to evaluate it in terms of channel \mathbf{h} or specifically in terms of \mathbf{B} and \mathbf{C} .

The output autocorrelation matrix $\Sigma_{\mathbf{Y}}$ can be decomposed as

$$\Sigma_{\mathbf{Y}} = \mathbf{G}\mathbf{G}^H + \sigma_n^2 \mathbf{I} \quad (3.13)$$

where

$$\mathbf{G}\mathbf{G}^H = \begin{bmatrix} \mathbf{C}\mathbf{C}^H + \mathbf{B}\mathbf{B}^H & \mathbf{B}\mathbf{C}^H & \mathbf{O} & \dots & \mathbf{O} & \mathbf{B}\mathbf{B}^H \\ \mathbf{C}\mathbf{B}^H & \mathbf{C}\mathbf{C}^H + \mathbf{B}\mathbf{B}^H & \mathbf{B}\mathbf{C}^H & \dots & \mathbf{O} & \mathbf{C}\mathbf{B}^H \\ \mathbf{O} & \mathbf{C}\mathbf{B}^H & \mathbf{C}\mathbf{C}^H + \mathbf{B}\mathbf{B}^H & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & & \mathbf{C}\mathbf{C}^H + \mathbf{B}\mathbf{B}^H & \mathbf{B}\mathbf{C}^H \\ \mathbf{B}\mathbf{B}^H & \mathbf{B}\mathbf{C}^H & \mathbf{O} & & \mathbf{C}\mathbf{B}^H & \mathbf{C}\mathbf{C}^H + \mathbf{B}\mathbf{B}^H \end{bmatrix}$$

with

$$\mathbf{G} = \begin{bmatrix} \mathbf{C} & \mathbf{B} & 0 & \cdots & 0 \\ 0 & \mathbf{C} & \mathbf{B} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{B} \\ 0 & \mathbf{B} & 0 & \cdots & \mathbf{C} \end{bmatrix}$$

The factor matrix \mathbf{G} has the following properties

1. It is a square matrix of size $N + L$.
2. It is full rank if and only if $h_L \neq 0$

Structure of \mathbf{G} in terms of channel parameters

Remember that we need to differentiate the likelihood function with respect to the channel IR. Now, the likelihood function is a function of $\Sigma_{\mathbf{Y}}$ which is a function of \mathbf{G} (see equation (3.13)). The matrix \mathbf{G} is itself a linear function of the channel IR. Specifically, we can write \mathbf{G} as a linear combination of $L + 1$ constant matrices $\mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_L$ i.e.

$$\mathbf{G} = \sum_{i=0}^L h_i \mathbf{F}_i \tag{3.14}$$

The matrix \mathbf{F}_i is an indicator matrix, i.e. it indicates the entries of \mathbf{G} that depend on h_i . The matrices \mathbf{F}_i have the following properties

1. It is easy to deduce the matrix \mathbf{F}_i from the structure of \mathbf{G} .
2. \mathbf{F}_i is square matrix just like \mathbf{G} .
3. The entries of \mathbf{F}_i are zeros or ones. In fact, each \mathbf{F}_i contains exactly $N + L$ ones (the same as the dimension of the output vector \mathbf{Y}).
4. It is easy to see that the \mathbf{F}'_i s are linearly independent.
5. The \mathbf{F}'_i s commute but they are not symmetric.

We can thus write

$$\mathbf{G}^T = \sum_{i=0}^L h_i \mathbf{F}_i^T \quad (3.15)$$

or

$$\begin{aligned} \text{vec}(\mathbf{G}^T) &= \sum_{i=0}^L h_i \text{vec}(\mathbf{F}_i^T) \\ &= \begin{bmatrix} \text{vec}(\mathbf{F}_0^T) & \text{vec}(\mathbf{F}_1^T) & \cdots & \text{vec}(\mathbf{F}_L^T) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_L \end{bmatrix} \\ &= \mathcal{F} \mathbf{h}^T \end{aligned} \quad (3.16)$$

where the vec operation transforms a matrix \mathbf{G} into a long column vector consisting of the concatenation of the columns of \mathbf{G} . Thus,

$$\begin{aligned}\frac{\partial \mathbf{G}}{\partial \mathbf{h}} &\stackrel{\Delta}{=} \frac{\partial \text{vec}(\mathbf{G}^T)}{\partial \mathbf{h}} \\ &= \mathcal{F}^T\end{aligned}\tag{3.17}$$

We will use this relation in evaluating gradient of \mathcal{L} w.r.t \mathbf{h} in the following section.

3.4.3 Gradient Matrix of \mathcal{L} w.r.t \mathbf{h}

We would like to find the gradient of \mathcal{L} w.r.t \mathbf{h} . By the chain rule, we can write

$$\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = \frac{\partial \mathcal{L}}{\partial \mathbf{G}} \frac{\partial \mathbf{G}}{\partial \mathbf{h}}\tag{3.18}$$

In carrying out the differentiation $\frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}}$, $\Sigma_{\mathbf{Y}}$ is treated as a general matrix. Thus, despite the fact that $\Sigma_{\mathbf{Y}}$ is symmetric and positive definite, we ignore this fact in obtaining $\frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}}$. All properties of $\Sigma_{\mathbf{Y}}$ are captured in its relation to \mathbf{G} and in the relation of the latter to \mathbf{h} .

We have already evaluated $\frac{\partial \mathbf{G}}{\partial \mathbf{h}}$. Lets now evaluate $\frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}}$. We can show that [87]

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}} &= -\frac{\partial}{\partial \Sigma_{\mathbf{Y}}} (\ln \det(\Sigma_{\mathbf{Y}}) + \mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y}) \\ &= -\text{vec}(\Sigma_{\mathbf{Y}}^{-T}) - \frac{\partial}{\partial \Sigma_{\mathbf{Y}}} \text{tr}(\mathbf{Y} \mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-1}) \\ \frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}} &= -\text{vec}(\Sigma_{\mathbf{Y}}^{-T} - \Sigma_{\mathbf{Y}}^{-T} \mathbf{Y} \mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-T})\end{aligned}\tag{3.19}$$

Similarly in carrying out the differentiation $\frac{\partial \Sigma_{\mathbf{Y}}}{\partial \mathbf{G}}$, we ignore the sparse structure of \mathbf{G} . The sparse structure is captured in the relation of \mathbf{G} to the channel parameters \mathbf{h} .

$$\begin{aligned}
\frac{\partial \Sigma_{\mathbf{Y}}}{\partial \mathbf{G}} &= \frac{\partial}{\partial \mathbf{G}} (\mathbf{G}\mathbf{G}^T + \sigma_n^2 \mathbf{I}) \\
&= \frac{\partial \mathbf{G}}{\partial \mathbf{G}} (\mathbf{I}_s \otimes \mathbf{G}^T) + \frac{\partial \mathbf{G}^T}{\partial \mathbf{G}} (\mathbf{G}^T \otimes \mathbf{I}_s) \\
\frac{\partial \Sigma_{\mathbf{Y}}}{\partial \mathbf{G}} &= (\mathbf{I} \otimes \mathbf{G}^T) + \mathbf{K}_{s,m}(\mathbf{G}^T \otimes \mathbf{I}_s)
\end{aligned} \tag{3.20}$$

where the second line is obtained by the product rule, and \otimes and $\mathbf{K}_{s,m}$ stand for Kronecker product and Commutation matrix respectively [87]. Combining the results (3.19) and (3.20) yields

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{G}} &= - ((\mathbf{I}_s \otimes \mathbf{G}^T) + \mathbf{K}_{s,m}(\mathbf{G}^T \otimes \mathbf{I}_s)) \text{vec}(\Sigma_{\mathbf{Y}}^{-T} - \Sigma_{\mathbf{Y}}^{-T} \mathbf{Y}\mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-T}) \\
&= - \text{vec} [\mathbf{G}^T \Sigma_{\mathbf{Y}}^{-T} - \mathbf{G}^T \Sigma_{\mathbf{Y}}^{-T} \mathbf{Y}\mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-T}] \\
&\quad - \mathbf{K}_{s,m} \text{vec} [\Sigma_{\mathbf{Y}}^{-T} \mathbf{G} - \Sigma_{\mathbf{Y}}^{-T} \mathbf{Y}\mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-T} \mathbf{G}] \\
\frac{\partial \mathcal{L}}{\partial \mathbf{G}} &= - 2 \text{vec} [\mathbf{G}^T \Sigma_{\mathbf{Y}}^{-1} - \mathbf{G}^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y}\mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-1}]
\end{aligned} \tag{3.21}$$

where in the last line, we used the property that

$$\mathbf{K}_{s,m} \text{vec}(\mathbf{A}^T) = \text{vec}(\mathbf{A})$$

So, equation (3.18) can now be written as

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{h}} &= \frac{\partial \mathbf{G}}{\partial \mathbf{h}} \frac{\partial \mathcal{L}}{\partial \mathbf{G}} \\
&= -2 \mathcal{F}^T \text{vec} [\mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}] \\
&= -2 \begin{bmatrix} \text{vec}(\mathbf{F}_0^T)^T \\ \text{vec}(\mathbf{F}_1^T)^T \\ \vdots \\ \text{vec}(\mathbf{F}_L^T)^T \end{bmatrix} \text{vec} [\mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}]
\end{aligned}$$

or

$$\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = -2 \begin{bmatrix} \text{tr}(\mathbf{F}_0 \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{F}_0 \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}) \\ \text{tr}(\mathbf{F}_1 \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{F}_1 \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}) \\ \vdots \\ \text{tr}(\mathbf{F}_L \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{F}_L \mathbf{G}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}) \end{bmatrix} \quad (3.22)$$

which is our required gradient of size $(L + 1) \times 1$. This is the gradient that we will use in the steepest descent algorithm further ahead. Thus a necessary condition for optimality is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = 0 \quad (3.23)$$

This gradient can be used to estimate the channel using the steepest descent algorithm. Before we do so however, we digress in the next subsection to show how this gradient provides an alternative form for the optimality condition.

3.4.4 An Alternative Condition for Optimality

The blind algorithm can also be used to estimate the noise standard deviation. Lets compute the first derivative of the likelihood function with respect to the noise standard deviation.

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial \sigma_n} &= \frac{\partial \Sigma_{\mathbf{Y}}}{\partial \sigma_n} \frac{\partial \mathcal{L}}{\partial \Sigma_{\mathbf{Y}}} \\
 &= -2\sigma_n \text{vec}(\mathbf{I})^T \text{vec}(\Sigma_{\mathbf{Y}}^{-T} - \Sigma_{\mathbf{Y}}^{-T} \mathbf{Y} \mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-T}) \\
 \frac{\partial \mathcal{L}}{\partial \sigma_n} &= -2\sigma_n \text{tr}(\Sigma_{\mathbf{Y}}^{-1} - \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y} \mathbf{Y}^T \Sigma_{\mathbf{Y}}^{-1}) \tag{3.24}
 \end{aligned}$$

The optimality condition $\frac{\partial \mathcal{L}}{\partial \sigma_n} = 0$ together with the condition $\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = 0$ allows us to represent our optimization problem in a different form. Specifically, start from the optimality conditions

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial \mathbf{h}} &= 0 \\
 \frac{\partial \mathcal{L}}{\partial \sigma_n} &= 0 \tag{3.25}
 \end{aligned}$$

and write the first one in an element by element form while performing the multiplication indicated

$$\begin{aligned}
\text{tr}(\mathbf{F}_0 \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} - \mathbf{F}_0 \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) &= 0 && \times \mathbf{h}_0 \\
\text{tr}(\mathbf{F}_1 \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} - \mathbf{F}_1 \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) &= 0 && \times \mathbf{h}_1 \\
&\vdots && \vdots \\
\text{tr}(\mathbf{F}_L \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} - \mathbf{F}_L \mathbf{G}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) &= 0 && \times \mathbf{h}_L \\
\text{tr}(\boldsymbol{\Sigma}_Y^{-1} - \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) &= 0 && \times \sigma_n^2
\end{aligned}$$

Adding the above equations yields

$$\begin{aligned}
\text{tr} \left[\left\{ \left(\sum_{i=0}^L h_i \mathbf{F}_i \right) \mathbf{G}^T + \sigma_n^2 \mathbf{I} \right\} \left\{ \boldsymbol{\Sigma}_Y^{-1} - \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \right\} \right] &= 0 \\
\text{tr} \left[(\mathbf{G} \mathbf{G}^T + \sigma_n^2 \mathbf{I}) (\boldsymbol{\Sigma}_Y^{-1} - \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) \right] &= 0 \\
\text{tr} \left[\boldsymbol{\Sigma}_Y (\boldsymbol{\Sigma}_Y^{-1} - \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1}) \right] &= 0 \\
\text{tr} \left[\mathbf{I} - \mathbf{Y} \mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \right] &= 0
\end{aligned}$$

The necessary conditions (3.25) imply that

$$\boxed{\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} = L + N} \tag{3.26}$$

Substituting this condition back into our original maximization problem (equation (3.9)) yields the following new form for the objective function.

$$\begin{aligned} \hat{\mathbf{h}}_{\text{ML}} &= \min_{\mathbf{h}} \ln \det(\boldsymbol{\Sigma}_{\mathbf{Y}}) \\ \text{subject to} & \\ \mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y} &= L + N \end{aligned} \tag{3.27}$$

So the optimization problem can be represented in any of the two formulations: the original one of (3.9) and the one represented by (3.27). Neither of these formulations are helpful unfortunately as we are not able to solve them for the global optimum value of \mathbf{h} . Instead, as we explained before, we will estimate the channel using the steepest descent algorithm.

3.5 Semiblind Estimation Using Steepest Descent Algorithm

As we cannot obtain an explicit solution to the problem at hand and come up with a convex formulation to pursue a blind approach, we adopt a semiblind technique by using the gradient to obtain a locally optimum solution. We can do this using Steepest Descent algorithm initialized with a rough channel estimate obtained by using a few pilots and channel correlation, i.e.

$$\mathbf{h}_{(k+1)}^T = \mathbf{h}_{(k)}^T - \mu \frac{\partial \mathcal{L}}{\partial \mathbf{h}_{(k)}} \tag{3.28}$$

where μ is called step size which is a small scalar value and $\mathbf{h}_{(k)}$ represents the estimate of channel \mathbf{h} at k^{th} iteration. The algorithm continues to iterate until a maximum number of iterations is reached or until a stopping threshold is crossed. Before we demonstrate the performance of the algorithm using simulations, we touch upon the issue of reducing computational complexity.

3.6 Computational Complexity

Regardless of how we solve the ML problem, the solution involves some heavy computations. We aim here to present some methods to simplify these computations. Specifically, in calculating the gradient and the likelihood function, two matrix operations are involved

1. $\Sigma_{\mathbf{Y}}^{-1}$
2. $\det(\Sigma_{\mathbf{Y}})$

Although $\det(\Sigma_{\mathbf{Y}})$ does not appear in the gradient but we still might need it to calculate the log likelihood function (equation (3.8)) to determine the stopping criteria of the steepest descent algorithm.

In simplifying the matrix calculations, we will rely on block matrix calculations. $\Sigma_{\mathbf{Y}}$ is given by the equation (3.13), reproduced here for convenience

$$\Sigma_{\mathbf{Y}} = \mathbf{G}\mathbf{G}^H + \sigma_n^2\mathbf{I} \quad (3.29)$$

where \mathbf{G} is a square matrix of size $(N+L) \times (N+L)$. We can write \mathbf{G} in the following block form

$$\mathbf{G} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{O} & \mathbf{A} \end{bmatrix} \quad (3.30)$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{B} & \mathbf{O} & \dots & \mathbf{O} \end{bmatrix}$$

and

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{B} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{C} & \mathbf{B} & \dots & \mathbf{O} \\ \vdots & \ddots & \ddots & \dots & \vdots \\ \mathbf{B} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{C} \end{bmatrix} \quad (3.31)$$

and where the matrices \mathbf{B} and \mathbf{C} were defined in equations (3.10) and (3.11). It is easy to see that the matrix \mathbf{A} is circulant so it is diagonalizable by an FFT matrix.

So, $\Sigma_{\mathbf{Y}}$ can now be written as

$$\begin{aligned}
\Sigma_{\mathbf{Y}} &= \mathbf{G}\mathbf{G}^H + \sigma_n^2 \mathbf{I} \\
&= \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{O} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{C}^H & \mathbf{O} \\ \mathbf{D}^H & \mathbf{A}^H \end{bmatrix} + \sigma_n^2 \mathbf{I} \\
&= \begin{bmatrix} \mathbf{C}\mathbf{C}^H + \mathbf{D}\mathbf{D}^H + \sigma_n^2 \mathbf{I} & \mathbf{D}\mathbf{A}^H \\ \mathbf{A}\mathbf{D}^H & \mathbf{A}\mathbf{A}^H + \sigma_n^2 \mathbf{I} \end{bmatrix} \tag{3.32}
\end{aligned}$$

In the following sections, we use the above structure of $\Sigma_{\mathbf{Y}}$ to calculate its inverse and determinant.

3.6.1 Calculating $\Sigma_{\mathbf{Y}}^{-1}$

Now to calculate $\Sigma_{\mathbf{Y}}^{-1}$, we use the following block inversion formula [88] (page 30, formula (2)),

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\gamma} & \boldsymbol{\xi} \end{bmatrix}^{-1} &= \\
&\begin{bmatrix} (\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})^{-1} & -(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})^{-1}\boldsymbol{\beta}\boldsymbol{\xi}^{-1} \\ -\boldsymbol{\xi}^{-1}\boldsymbol{\gamma}(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})^{-1} & \boldsymbol{\xi}^{-1} + \boldsymbol{\xi}^{-1}\boldsymbol{\gamma}(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})^{-1}\boldsymbol{\beta}\boldsymbol{\xi}^{-1} \end{bmatrix}
\end{aligned}$$

which is valid provided the inverses involved are valid.

There are two inverses that need to be calculated here

1. $\boldsymbol{\xi}^{-1}$
2. $(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})^{-1}$

Given the structure of $\Sigma_{\mathbf{Y}}$ in equation (3.32), this transforms to calculating

$$(\mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I})^{-1} \quad (N \times N) \quad (3.33)$$

and

$$(\mathbf{C}\mathbf{C}^H + \mathbf{D}\mathbf{D}^H + \sigma_n^2\mathbf{I} - \mathbf{D}\mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I})^{-1}\mathbf{A}\mathbf{D}^H)^{-1} \quad (L \times L) \quad (3.34)$$

Now \mathbf{A} is circulant (equation (3.31)), so we can write it as

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^H \quad (3.35)$$

where \mathbf{Q} is the FFT matrix and Λ is the FFT of the first row of \mathbf{A} (which is the FFT of $[h_L \ h_{L-1} \ \dots \ h_1 \ h_0 \ 0 \ \dots \ 0]$). It is easy to see then that the inverse in (3.33) can be computed as follows

$$(\mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I})^{-1} = \mathbf{Q}(\Lambda + \sigma_n^2\mathbf{I})^{-1}\mathbf{Q}^H \quad (3.36)$$

$$= \mathbf{Q} \begin{bmatrix} \frac{1}{|\lambda_1|^2 + \sigma_n^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{|\lambda_N|^2 + \sigma_n^2} \end{bmatrix} \mathbf{Q}^H \quad (3.37)$$

Now, (3.34) is an $L \times L$ inverse, so it is easy to calculate. However, it involves an

$N \times N$ matrix inversion, namely $\mathbf{A}(\mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I})^{-1}\mathbf{A}^H$ which can be evaluated as

$$\mathbf{A}(\mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I})^{-1}\mathbf{A}^H = \mathbf{Q} \begin{bmatrix} \frac{|\lambda_1|^2}{|\lambda_1|^2 + \sigma_n^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{|\lambda_N|^2}{|\lambda_N|^2 + \sigma_n^2} \end{bmatrix} \mathbf{Q}^H \quad (3.38)$$

So, evaluating inverse of $(N + L) \times (N + L)$ matrix $\Sigma_{\mathbf{Y}}$ reduces to calculating an FFT (to find Λ) and to calculating the $L \times L$ inverse of (3.34).

3.6.2 Calculating $\det(\Sigma_{\mathbf{Y}})$

In order to calculate determinant of $\Sigma_{\mathbf{Y}}$, we use [88] (page 50, formula (6))

$$\det \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\gamma} & \boldsymbol{\xi} \end{bmatrix} = \det(\boldsymbol{\xi}) \det(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma}) \quad (3.39)$$

Now, from (3.32),

$$\boldsymbol{\xi} = \mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I} \quad (3.40)$$

$$\Rightarrow \det(\boldsymbol{\xi}) = \det(\mathbf{Q}(\Lambda + \sigma_n^2\mathbf{I})\mathbf{Q}^H) \quad (3.41)$$

$$= \det(\Lambda + \sigma_n^2\mathbf{I}) \quad (3.42)$$

We already know how to calculate $\boldsymbol{\xi}^{-1}$ from equation (3.37), so $\det(\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\xi}^{-1}\boldsymbol{\gamma})$ involves calculating the determinant of an $L \times L$ matrix.

So to summarize, the inverse and the determinant of the $(N + L) \times (N + L)$ matrix

$\Sigma_{\mathbf{Y}}$ reduces to calculating

1. FFT (to find Λ)
2. Inverse and determinant of an $L \times L$ matrix.

3.7 Simulation Results

We consider an OFDM system with $N = 64$ and cyclic prefix of length $L = 8$. The OFDM symbol consists of BPSK or 16-QAM symbols. The channel IR consists of 9 iid Rayleigh fading taps. The proposed semiblind algorithm was run for 20 iterations in all cases. We compare the BER performance of four methods: (i) Perfectly known channel, (ii) Channel estimated using $L+1$ pilots, (iii) Channel estimated using semi-blind method proposed in this chapter using the steepest descent algorithm, and (iv) Channel estimated using low number of pilots (either 6 or 8) and channel correlation which is how we initialize the steepest descent algorithm.

3.7.1 BER vs SNR Comparison for BPSK Modulated Data

In Figure 3.6, the proposed semiblind algorithm is compared with the above mentioned methods when the input data is BPSK modulated. In this case, the algorithm is initialized with an estimate obtained by using 6 pilots and channel correlation. The step size μ used in this case was 7.5×10^{-4} . This figure clearly indicates the favorable performance of the proposed method especially at high SNR. Thus, while the pilot based estimate flattens, the BER of the steepest descent algorithm continues to decrease with SNR.

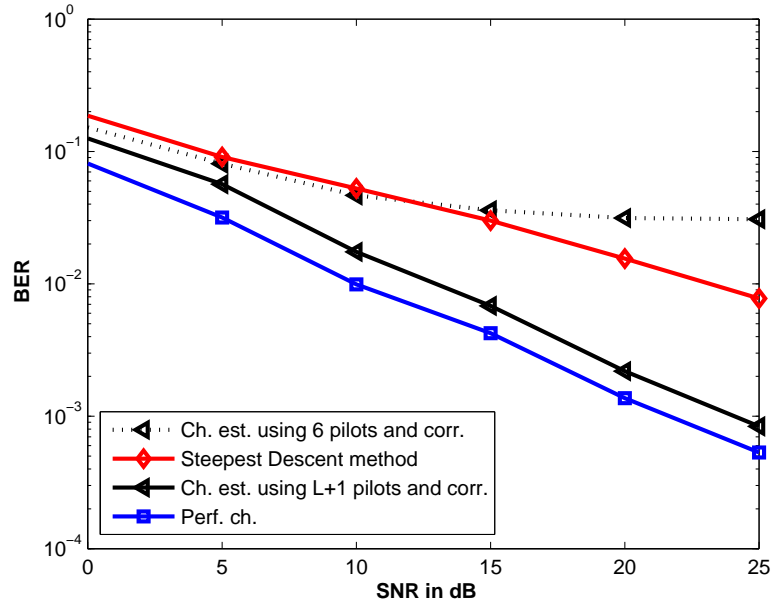


Figure 3.6: BER vs SNR comparison for BPSK modulated data

3.7.2 BER vs SNR Comparison for 16-QAM Modulated Data

In Figure 3.7, the performance of the proposed semi blind algorithm is compared with the remaining three methods for the case of 16-QAM modulated data. Similar to the BPSK case, the algorithm is again initialized with an estimate obtained by using 6 pilots and channel correlation. The step size μ used in this case was 2.5×10^{-4} . Our conclusion is similar to the conclusion in the BPSK, namely, the steepest descent method decreases with SNR while the pilot based estimate remains flat for high SNR.

3.7.3 Sensitivity to Number of Pilots

The sensitivity of the proposed algorithm to number of pilots used is studied in Figure 3.8. Specifically, we compare the performance of the semiblind algorithm for BPSK modulated data when it is initialized by an estimate using 6 or 8 pilots. The step size

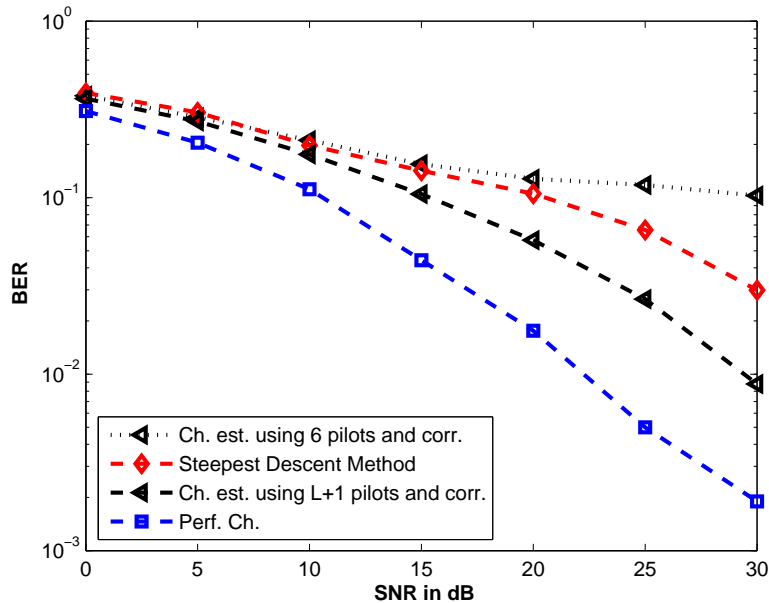


Figure 3.7: BER vs SNR comparison for 16-QAM modulated data

μ used was fixed at 7.5×10^{-4} . As expected, the algorithm performs well when it is initialized with a better estimate.

3.8 Conclusion

In this chapter, we presented a semiblind channel-centered approach for channel estimation and data recovery in OFDM transmission. It was argued in this part that the transmitted data in OFDM is Gaussian. Thus the output is also Gaussian and its pdf can be evaluated easily. The channel can then be estimated by maximizing the likelihood function given the output pdf. The experimental results unfortunately demonstrated that the log likelihood function does not have a unique maxima when it is plotted against the channel taps. Thus to pursue a completely blind approach

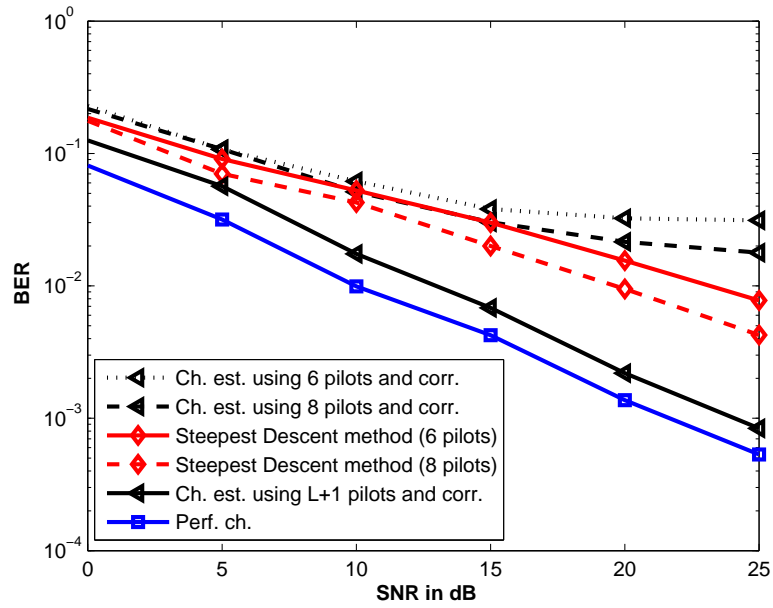


Figure 3.8: BER vs SNR comparison for BPSK modulated data with different number of pilots

might be challenging. Therefore, a semiblind approach was adopted using the steepest descent algorithm initialized by a rough estimate of channel obtained by using a few of pilots and channel correlation. Methods to reduce computational complexity involved in evaluating the gradient and the log likelihood function were also suggested.

CHAPTER 4

MIMO SEMI-BLIND CHANNEL ESTIMATION AND DATA DETECTION

4.1 Introduction

This chapter considers receiver design for orthogonal space time block coded¹ MIMO OFDM transmission over frequency selective time-variant spatially correlated channels. The receiver employs the expectation-maximization (EM) algorithm for joint channel and data recovery. It makes collective use of the data and channel constraints that characterize the communication problem. The data constraints include pilots, the cyclic prefix, the finite alphabet constraint, and space-time block coding. The channel

¹We concentrate on space-frequency codes in this work. We note however that a very similar approach can be used for space-time block codes (STBC). Given the similarity between the two approaches and the fact that the abbreviation STBC is more familiar, we continue to use this abbreviation to refer to our frequency-space codes.

constraints include the finite delay spread and frequency and time correlation as well as transmit and receive correlation. The channel estimation part of the receiver boils down to an EM-based forward-backward Kalman filter. A different implementation of the forward-backward Kalman filter has also been presented. To avoid the latency and storage associated with smoothing, we introduce a forward-only Kalman that can perform channel (and data) recovery with no latency. Simulations have been performed to compare the performance of the different implementations of the Kalman filter and investigate the sensitivity of the receiver to an outer code and different number of pilots.

4.1.1 The Approach of this Chapter

This chapter considers receiver design for OSTBC-OFDM transmission over frequency selective, time-variant, and spatially correlated channels. Our approach is based on [96] which proposed a Kalman filter approach to channel estimation in MIMO OFDM transmission. We have extended this approach to spatially correlated channels and explored the use of forward-backward Kalman filter with different implementations.

The receiver makes a collective use of the structure of the communication problem (i.e. the constraints on the data and on the channel). The data constraints include the finite alphabet constraint, the cyclic prefix, pilots, and the OSTBC. In addition, the receiver uses the following constraints on the channel: the finite delay spread, frequency and time correlation, and spatial correlation.

In spite of the many dimensions we deal with (see Table 4.1), we maintain the

Table 4.1: Indices used in the analysis of MIMO semiblind channel estimation and data detection

Variable	Index employed
Channel tap	$p = 0, 1, \dots, P$
Frequency tone	$n = 1, 2, \dots, N$
uncoded OFDM symbol (prior to ST coding)	$n_u = 1, 2, \dots, N_u$
Transmit antenna	$t_x = 1, 2, \dots, T_x$
Receive antenna	$r_x = 1, 2, \dots, R_x$
coded OFDM symbol i.e. part of a ST block	$n_c = 1, 2, \dots, N_c$
Space-time block	$t = 1, 2, \dots$
Sample time	$m = 1, 2, \dots$

transparency of the presentation.

4.1.2 Organization of the Chapter

The chapter is organized as follows. After introducing our notation, we give an overview of the transceiver in Section 4.2. Section 4.3 then presents the input/output equations for MIMO-OFDM with ST coding (the equations are needed for channel and data recovery). Channel estimation using the FB-Kalman is described in Section 4.4 at the end of which we summarize the transceiver algorithm. Section 4.5 presents two extensions/modifications of the algorithm and Section 4.6 presents our simulations. We conclude the chapter in Section 4.7.

4.1.3 Notation

Proper choice of notation is essential for clarity and consistency. One challenge in choosing notation is the many dimensions we deal with in this chapter including

sample time, (frequency) tone, (channel) tap, (OFDM) symbol index, and space.

We continue to use the notation of Chapters 2 and 3. Additionally in this chapter, we use $*$ to denote conjugate transpose, \otimes to denote Kronecker product, \mathbf{I}_N to denote the size $N \times N$ identity matrix, and $\mathbf{0}_{M \times N}$ to denote the all zero $M \times N$ matrix. Given a sequence of vectors $\mathbf{h}_{r_x}^{t_x}$ for $r_x = 1 \cdots R_x$ and $t_x = 1 \cdots T_x$, we define the following stack variables

$$\mathbf{h}_{r_x} = \begin{bmatrix} \mathbf{h}_{r_x}^1 \\ \vdots \\ \mathbf{h}_{r_x}^{T_x} \end{bmatrix} \quad \text{and} \quad \mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{R_x} \end{bmatrix} \quad (4.1)$$

The notation $\text{vec}(\mathbf{X})$ is column vector consisting of the concatenation of all column vectors of \mathbf{X} while the operation $\text{diag}(\mathbf{x})$ transforms the vector \mathbf{x} into a matrix with diagonal \mathbf{x} .

4.2 System Overview

In this section, we give an overview of the communications system: transmitter, channel, and receiver [96].

4.2.1 Transmitter

A block diagram of the transmitter is shown in Figure 4.1. The bit sequence to be transmitted passes through a convolutional encoder that serves as an outer code for the system. The coded output then passes through a random interleaver which rearranges the order of the bits according to a random permutation. The interleaved bit sequence

is mapped to QAM (or any modulation scheme for that matter) symbols using gray coding and the QAM symbols are in turn mapped to the OFDM symbols with some tones reserved for the pilot symbols. The STBC encoder uses the OFDM symbols to construct the ST block by mapping the various OFDM symbols to a specific antenna and specific time slot depending on the ST code used. Each antenna performs an IFFT operation on the OFDM symbols to produce the time-domain OFDM symbols and adds a cyclic prefix to each prior to transmission.

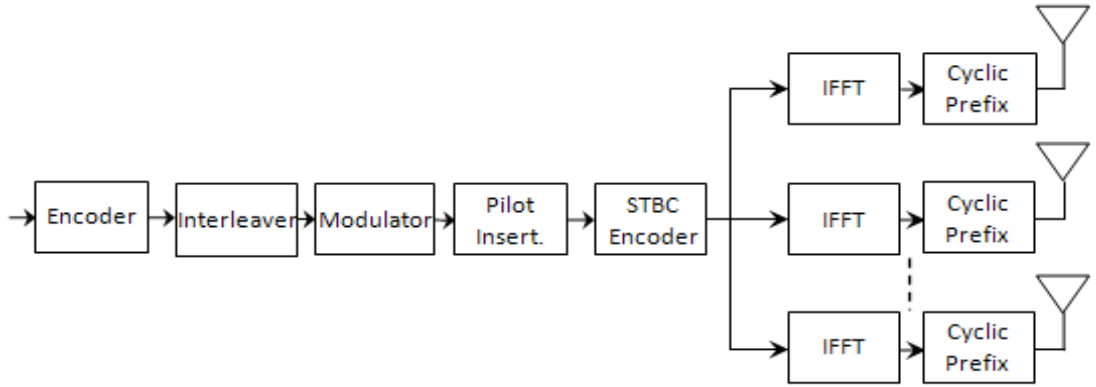


Figure 4.1: OSTBC OFDM Transmitter

4.2.2 Channel Model

We consider a time-variant and frequency selective MIMO channel. For a general MIMO system, the input/output time-domain relationship is described by

$$\mathbf{y}(m) = \sum_{p=0}^P \mathbf{H}(p) \mathbf{x}(m - p)$$

where $\mathbf{H}(p)$ is the $R_x \times T_x$ MIMO impulse response at tap p and where m represents the sample time index. The taps $\mathbf{H}(p)$ usually incorporate the effect of the transmit filter and the effects of the transmit and receive correlation making $\mathbf{H}(p)$ correlated across space and tap. Here it is assumed for simplicity that $\mathbf{H}(p)$ is iid for all p (In Appendix A, the general case where the channel exhibits transmit and receive correlation is described which was derived in [96]). We will also assume that the tap $\mathbf{H}(p)$ remains constant over a single ST block (and hence over the constituent OFDM symbols) and changes from the current block ($\mathbf{H}_t(p)$) to the next ($\mathbf{H}_{t+1}(p)$) according to the dynamical equation²

$$\mathbf{H}_{t+1}(p) = \alpha(p)\mathbf{H}_t(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}\mathbf{U}_t(p) \quad (4.2)$$

Here, $\mathbf{U}_t(p)$ is an iid matrix with entries that are $\mathcal{N}(0, 1)$ and $\alpha(p)$ is related to the Doppler frequency $f_D(p)$ by $\alpha(p) = J_0(2\pi f_D T(p))$, where T is the time duration of one ST block. The variable β in (4.2) corresponds to the exponent of the channel decay profile while the factor $\sqrt{(1 - \alpha^2(p))e^{-\beta p}}$ ensures that each link maintains the exponential decay profile ($e^{-\beta p}$) for all time.

Using this dynamical model, we can obtain the state-space model for the impulse response $\mathbf{h}_{r_x}^{t_x}$ between transmit antenna t_x and receive antenna r_x . From (4.2), we can write

$$h_{r_{xt+1}}^{t_x}(p) = \alpha(p)h_{r_{xt}}^{t_x}(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}u_{r_{xt}}^{t_x}(p) \quad (4.3)$$

²We shall at times suppress the time dependence for notational convenience.

By stacking (4.3) over the taps $p = 0, 1, \dots, P$, we obtain the dynamical model

$$\mathbf{h}_{r_x t+1}^{t_x} = \mathbf{F} \mathbf{h}_{r_x}^{t_x} + \mathbf{G} \mathbf{u}_{r_x}^{t_x} \quad (4.4)$$

where

$$\mathbf{F} = \begin{bmatrix} \alpha(0) & & \\ & \ddots & \\ & & \alpha(P) \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \sqrt{1 - \alpha^2(0)} & & \\ & \ddots & \\ & & \sqrt{(1 - \alpha^2(P))e^{-\beta P}} \end{bmatrix}$$

By further stacking (4.4) over all transmit and receive antennas (refer to our stacking notation in (4.1)), we obtain

$$\boxed{\mathbf{h}_{t+1} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \mathbf{h}_t + (\mathbf{I}_{T_x R_x} \otimes \mathbf{G}) \mathbf{u}_t} \quad (4.5)$$

where \mathbf{h}_{t+1} , \mathbf{h}_t , and \mathbf{u}_t , are vectors of size $T_x R_x (P + 1) \times 1$. The dynamical equation (4.5) shows explicit dependence on the space-time index t (so $t=1$ for the 1st space-time symbol which consists of two OFDM symbols in the Alamouti case, $t=2$ for the 2nd space-time symbol, and so on).

For complete characterization of the dynamical model, we need to specify the covariance of \mathbf{u}_t , and also the covariance of the channel at the first time instant. It is

easy to show that

$$\begin{aligned}
E[\mathbf{u}_t \mathbf{u}_t^*] &= \mathbf{I}_{R_x} \otimes E[\mathbf{u}_{r_x} \mathbf{u}_{r_x}^*] \\
&= \mathbf{I}_{R_x} \otimes (\mathbf{I}_{T_x} \otimes E[u_{r_x}^{t_x} u_{r_x}^{t_x*}]) \\
&= \mathbf{I}_{R_x} \otimes \mathbf{I}_{T_x} \otimes \mathbf{I}_{P+1} = \mathbf{I}_{T_x R_x (P+1)}
\end{aligned}$$

We can similarly show that the channel covariance at the first time instant is given by

$$E[\mathbf{h}_0 \mathbf{h}_0^*] = \mathbf{I}_{T_x R_x} \otimes \mathbf{G} \mathbf{G}^*$$

The covariance information is important for employing the Kalman filter which is used for channel estimation. We finally note that while (4.2) and (4.5) are equivalent, the latter model is in vector form and hence lends itself more to the Kalman filter operations.

A note about time variation:

One drawback of the approach in this chapter is the block fading model that we adopt. For it is more realistic to assume that the channel continuously varies with time. There are 3 justifications for using the block fading model:

1. While it is more realistic to assume that the channel continuously varies over the OFDM symbol, the model we assume is more valid than the block fading model that is widely used in literature. For in the block fading model, it is usually assumed that the channel remains constant over any one symbol and

varies *independently* from one symbol to another. Here, we account for the time-correlation across symbols.

2. The purpose of this chapter is to design an algorithm that makes a collective use of the underlying structure of the communication problem to lower the training overhead required in the time-variant case. Solving the general time-variant case is a future research problem that builds upon the findings in this chapter.
3. One could still envision applications where the channel is constant over a ST block, but varies substantially from one symbol to the next. Consider for example a multiuser application in which the wireless channel is time shared. Imagine also that the channel is very slowly time-variant but the duty cycle is very large. In that case, the channel that each user experiences during his transmission burst is very slow, but from one burst to another, the channel would change substantially due to the long duty cycle. This situation would also make sense in random access scenarios.

4.2.3 Receiver

This chapter is concerned with designing a receiver for the system described above. For completeness and as an allusion to the developments further ahead, Figure 4.2 shows a block diagram of the proposed receiver. As we shall show, the receiver's core operation is based on the EM algorithm which performs joint channel and data recovery:

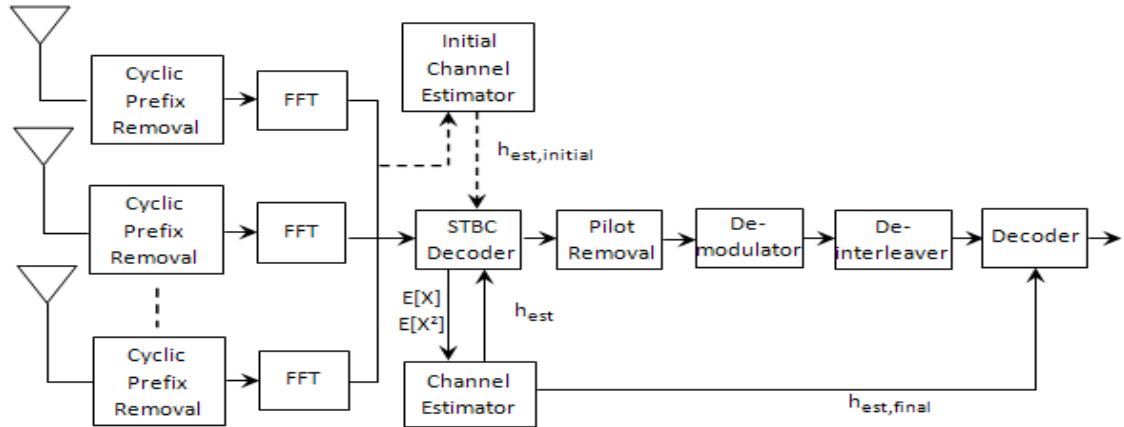


Figure 4.2: OSTBC OFDM Receiver

STBC Decoder/Data Detector (Estimation Step)

The STBC decoder/data detector calculates the conditional first and second moments of the transmitted data (soft estimate) to be used by the channel estimator.

Channel Estimator (Maximization Step)

Pilots are used to initialize channel estimation. The channel estimator then uses the soft data estimates together with the data and channel constraints to improve the channel estimate. These two processes (channel estimation and data detection) go on iteratively until a stopping criterion is satisfied.

In what follows, we describe the main results which were derived in [96]. We start by explaining the input/output equations for MIMO-OFDM system.

4.3 I/O Equations for MIMO-OFDM

As pointed out in Subsection 4.2.3, the receiver performs two operations, channel estimation and data detection. As such, we need to derive two forms of the (I/O) equations: one that lends itself to *channel estimation* (i.e. treats the channel impulse response as the unknown) and a dual version that lends itself to *data detection* (i.e. treats the input in its uncoded form as the unknown). To this end, Let \mathbf{x}_{t_x} be the OFDM symbol transmitted through antenna t_x which first undergoes an IDFT $\mathbf{x}_{t_x} = 1/N\mathbf{Q}\mathbf{x}_{t_x}$ where \mathbf{Q} is the $N \times N$ IDFT matrix. The system then appends a cyclic prefix before transmission. At the receiver end, the receiver strips the cyclic prefix to obtain the time domain symbol $\mathbf{y}_{r_x}^{t_x}$. The I/O equation of the OFDM system between transmit antenna t_x and receive antenna r_x is best described in the frequency domain

$$\mathbf{y}_{r_x}^{t_x} = \text{diag}(\mathbf{x}_{t_x}) \mathbf{Q}_{P+1}^* \mathbf{h}_{r_x}^{t_x} + \mathcal{N}_{r_x} \quad (4.6)$$

where $\mathbf{y}_{r_x}^{t_x}$, \mathbf{x}_{t_x} , $\mathcal{H}_{r_x}^{t_x}$, and $\mathcal{N}_{r_x}^{t_x}$ are the (length- N) DFT's of $\mathbf{y}_{r_x}^{t_x}$, \mathbf{x}_{t_x} , $\mathbf{h}_{r_x}^{t_x}$, \mathbf{n}_{r_x} , respectively, and where (4.6) follows from the fact that

$$\mathcal{H}_{r_x}^{t_x} = \mathbf{Q}^* \begin{bmatrix} \mathbf{h}_{r_x}^{t_x} \\ \mathbf{O}_{(N-P-1) \times 1} \end{bmatrix} = \mathbf{Q}_{P+1}^* \mathbf{h}_{r_x}^{t_x} \quad (4.7)$$

Here \mathbf{Q}_{P+1} represents the first $P + 1$ rows of \mathbf{Q} . By superposition and using the stacking notation (4.1), we can express the I/O equation at receive antenna r_x as

$$\mathbf{y}_{r_x} = [\text{diag}(\boldsymbol{\chi}_1) \cdots \text{diag}(\boldsymbol{\chi}_{T_x})] (\mathbf{I}_{T_x} \otimes \mathbf{Q}_{P+1}^*) \mathbf{h}_{r_x} + \mathcal{N}_{r_x} \quad (4.8)$$

4.3.1 I/O Equations with Space Time Coding: Channel Estimation Version

Consider a set of N_u uncoded OFDM symbols $\{\boldsymbol{\mathcal{S}}(1), \dots, \boldsymbol{\mathcal{S}}(N_u)\}$ which we would like to transmit over T_x antennas and N_c time slots. Following [95], we can perform ST coding using the set of $T_x \times N_c$ matrices $\{\mathbf{A}(1), \mathbf{B}(1), \dots, \mathbf{A}(N_u), \mathbf{B}(N_u)\}$ which characterizes the ST code. We can now show that the OFDM symbol transmitted from antenna t_x at time n_c is given by

$$\boldsymbol{\chi}_{t_x}(n_c) = \sum_{n_u=1}^{N_u} a_{t_x, n_c}(n_u) \text{Re} \boldsymbol{\mathcal{S}}(n_u) + j b_{t_x, n_c}(n_u) \text{Im} \boldsymbol{\mathcal{S}}(n_u) \quad (4.9)$$

where $a_{t_x, n_c}(n_u)$ is the (t_x, n_c) element of $\mathbf{A}(n_u)$ and $b_{t_x, n_c}(n_u)$ is the (t_x, n_c) element of $\mathbf{B}(n_u)$. Thus, in the presence of ST coding, (4.8) reads

$$\mathbf{y}_{r_x}(n_c) = [\text{diag}(\boldsymbol{\chi}_1(n_c)) \cdots \text{diag}(\boldsymbol{\chi}_{T_x}(n_c))] (\mathbf{I}_{T_x} \otimes \mathbf{Q}_{P+1}^*) \mathbf{h}_{r_x} + \mathcal{N}_{r_x}(n_c)$$

This represents the I/O equation at antenna r_x at OFDM symbol n_c of a ST block.

Collecting this equation for all such symbols yields

$$\mathbf{y}_{r_x} = \mathbf{X}\mathbf{h}_{r_x} + \mathcal{N}_{r_x} \quad (4.10)$$

where

$$\mathbf{y}_{r_x} = \begin{bmatrix} \mathbf{y}_{r_x}(1) \\ \vdots \\ \mathbf{y}_{r_x}(N_c) \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \text{diag}(\boldsymbol{\mathcal{X}}_1(1)) \cdots \text{diag}(\boldsymbol{\mathcal{X}}_{T_x}(1)) \\ \text{diag}(\boldsymbol{\mathcal{X}}_1(2)) \cdots \text{diag}(\boldsymbol{\mathcal{X}}_{T_x}(2)) \\ \vdots \\ \text{diag}(\boldsymbol{\mathcal{X}}_1(N_c)) \cdots \text{diag}(\boldsymbol{\mathcal{X}}_{T_x}(N_c)) \end{bmatrix} \quad (4.11)$$

Now, by further collecting this relationship over all receive antennas, we obtain

$$\boxed{\mathbf{y}_t = (\mathbf{I}_{R_x} \otimes \mathbf{X}_t)\mathbf{h}_t + \mathcal{N}_t} \quad (4.12)$$

This equation captures the I/O relationship at *all* frequency bins, for *all* input and output antennas, and for *all* OFDM symbols of the t^{th} ST block. We can also construct a similar I/O relationship that incorporates (4.12) as well as the effect of the cyclic prefix observation.

$$\boxed{\bar{\mathbf{y}}_t = (\mathbf{I}_{R_x} \otimes \bar{\mathbf{X}}_t)\mathbf{h}_t + \bar{\mathcal{N}}_t} \quad (4.13)$$

where at a particular antenna r_x ,

$$\bar{\mathbf{y}}_{r_x} = \begin{bmatrix} \mathbf{y}_{r_x} \\ \underline{\mathbf{y}}_{r_x} \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{X}} = \begin{bmatrix} \bar{\mathbf{X}}_1 & \bar{\mathbf{X}}_2 & \cdots & \bar{\mathbf{X}}_{T_x} \end{bmatrix} \quad (4.14)$$

with

$$\bar{\mathbf{X}}_{t_x} = \begin{bmatrix} \bar{\mathbf{X}}_{t_x}(1) \\ \bar{\mathbf{X}}_{t_x}(2) \\ \vdots \\ \bar{\mathbf{X}}_{t_x}(N_c) \end{bmatrix}$$

and

$$\bar{\mathbf{X}}_{t_x}(n_c) = \begin{bmatrix} \mathbf{X}_{t_x}(n_c) \\ \underline{\mathbf{X}}_{t_x}(n_c) \end{bmatrix} = \begin{bmatrix} \text{diag}(\boldsymbol{\mathcal{X}}_{t_x}(n_c))\mathbf{Q}_{P+1}^* \\ \underline{\mathbf{X}}_{t_x}(n_c) \end{bmatrix}$$

The matrix $\underline{\mathbf{X}}_{t_x}$ is constructed using the cyclic prefix of current and previous symbol.

For a particular transmitter, with i representing the current symbol and $i - 1$ the previous one, $\underline{\mathbf{X}}$ is given by

$$\underline{\mathbf{X}}_i = \begin{bmatrix} \underline{x}_i(0) & \underline{x}_{i-1}(P-1) & \cdots & \underline{x}_{i-1}(0) \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & \underline{x}_{i-1}(1) \\ \vdots & \ddots & \ddots & \vdots \\ \underline{x}_i(P-1) & \cdots & \underline{x}_i(0) & \underline{x}_{i-1}(P-1) \end{bmatrix} \quad (4.15)$$

To perform initial channel estimation, we select those equations where the pilots are present. Let I_p denote the index set of the pilots bins. Then, the pilot/output equation takes the form

$$\boxed{\mathcal{Y}_{t_{I_p}} = (\mathbf{I}_{R_x} \otimes \mathbf{X}_{t_{I_p}})\mathbf{h}_t + \mathcal{N}_{t_{I_p}}} \quad (4.16)$$

As can be seen, the equations (4.12), and (4.16) are quite similar and using them in a Kalman filter context will be similar as well.

4.3.2 I/O Equations with Space Time Coding: Data Detection Version

Signal detection in ST-coded OFDM is done on a tone-by-tone basis (i.e., as in SISO OFDM), except that the tones are collected for the whole ST block (i.e., for R_x receive antennas and over N_c time slots). From (4.6), we can construct the following I/O equation *at any tone n* belonging to the OFDM symbol n_c

$$\mathcal{Y}_{r_x}(n_c) = \begin{bmatrix} \mathcal{H}_{r_x}^1 & \cdots & \mathcal{H}_{r_x}^{T_x} \end{bmatrix} \begin{bmatrix} \mathcal{X}_1(n_c) \\ \vdots \\ \mathcal{X}_{T_x}(n_c) \end{bmatrix} + \mathcal{N}_{r_x}(n_c) \quad (4.17)$$

We suppress the dependence on n for notational convenience. Collecting this relationship for all receive antennas yields

$$\begin{bmatrix} \mathcal{Y}_1(n_c) \\ \vdots \\ \mathcal{Y}_{R_x}(n_c) \end{bmatrix} = \begin{bmatrix} \mathcal{H}_1^1 & \cdots & \mathcal{H}_1^{T_x} \\ \vdots & \cdots & \vdots \\ \mathcal{H}_{R_x}^1 & \cdots & \mathcal{H}_{R_x}^{T_x} \end{bmatrix} \begin{bmatrix} \mathcal{X}_1(n_c) \\ \vdots \\ \mathcal{X}_{T_x}(n_c) \end{bmatrix} + \begin{bmatrix} \mathcal{N}_1(n_c) \\ \vdots \\ \mathcal{N}_{R_x}(n_c) \end{bmatrix}$$

Or, more succinctly,

$$\mathbf{y}(n_c) = \mathbf{H}\mathbf{x}(n_c) + \mathbf{N}(n_c)$$

By further concatenating this relationship for $n_c = 1, \dots, N_c$, we can show that the following relationship holds (see [95])

$$\mathbf{y} = \mathbf{C} \begin{bmatrix} \text{Re } \mathbf{s} \\ \text{Im } \mathbf{s} \end{bmatrix} + \mathcal{N} \quad (4.18)$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(N_c) \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} \mathbf{s}(1) \\ \vdots \\ \mathbf{s}(N_u) \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_a & \mathbf{C}_b \end{bmatrix}$$

with

$$\mathbf{C}_a = \begin{bmatrix} \text{vec}(\mathbf{H}\mathbf{A}(1)) & \cdots & \text{vec}(\mathbf{H}\mathbf{A}(N_u)) \end{bmatrix},$$

$$\text{and } \mathbf{C}_b = \begin{bmatrix} \text{vec}(\mathbf{H}\mathbf{B}(1)) & \cdots & \text{vec}(\mathbf{H}\mathbf{B}(N_u)) \end{bmatrix}$$

We finally note that the STBC code is orthogonal if and only if the matrix \mathbf{C} satisfies [95]

$$\operatorname{Re}[\mathbf{C}^* \mathbf{C}] = \|\mathcal{H}\|^2 \mathbf{I}_{2N_u} \quad \forall \mathcal{H} \quad (4.19)$$

This property is essential to perform data detection. We stress that the relationships (4.17) through (4.19) apply at a particular tone n and that this dependence has been omitted for notational convenience.

4.4 The EM-Based Forward-Backward Kalman

Consider the OFDM system of this chapter, essentially described by the state-space model

$$\mathbf{h}_{t+1} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \mathbf{h}_t + (\mathbf{I}_{T_x R_x} \otimes \mathbf{G}) \mathbf{u}_t \quad (4.20)$$

$$\mathbf{y}_t = (\mathbf{I}_{R_x} \otimes \mathbf{X}_t) \mathbf{h}_t + \mathcal{N}_t \quad (4.21)$$

with $\mathbf{h}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Pi})$ and $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_u)$. The MMSE estimate of the channel sequence given the input and output sequences \mathbf{X}_0^T and \mathbf{y}_0^T is obtained by the forward-backward (FB) Kalman filter. We consider the following two cases.

4.4.1 Channel Estimation–Known Input Case

Consider the state-space model (4.20)–(4.21). Given the input and output sequences \mathbf{X}_0^T and \mathbf{y}_0^T , the MAP (or equivalently MMSE) estimate of \mathbf{h}_0^T is obtained by applying

the following (forward-backward Kalman) filter to the state-space model (4.20)–(4.21) [96]

Forward run: For $i = 1, \dots, T$, calculate

$$\mathbf{R}_{e,t} = \sigma_n^2 \mathbf{I}_{T_x R_x N} + (\mathbf{I}_{R_x} \otimes \mathbf{X}_t) \mathbf{P}_{t|t-1} (\mathbf{I}_{R_x} \otimes \mathbf{X}_t^*) \quad \mathbf{P}_{0|-1} = \mathbf{\Pi}_0 \quad (4.22)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} (\mathbf{I}_{R_x} \otimes \mathbf{X}_t^*) \mathbf{R}_{e,t}^{-1} \quad (4.23)$$

$$\hat{\mathbf{h}}_{t|t} = (\mathbf{I}_{T_x R_x (P+1)} - \mathbf{K}_t (\mathbf{I}_{R_x} \otimes \mathbf{X}_t)) \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t \mathbf{y}_t, \quad (4.24)$$

$$\hat{\mathbf{h}}_{t+1|t} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \hat{\mathbf{h}}_{t|t}, \quad \mathbf{h}_{0|-1} = \mathbf{0} \quad (4.25)$$

$$\mathbf{P}_{t+1|t} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) (\mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{R}_{e,t} \mathbf{K}_t^*) (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}^*) + \mathbf{G} \mathbf{R}_u \mathbf{G}^* \quad (4.26)$$

Backward run: Starting from $\boldsymbol{\lambda}_{T+1|T} = \mathbf{0}$ and for $t = T, T-1, \dots, 0$, calculate

$$\begin{aligned} \boldsymbol{\lambda}_{t|T} = & (\mathbf{I}_{P+N} - (\mathbf{I}_{R_x} \otimes \mathbf{X}_t^*) \mathbf{K}_t^*) (\mathbf{I} \otimes \mathbf{F}^*) \boldsymbol{\lambda}_{t+1|T} \\ & + (\mathbf{I} \otimes \mathbf{X}_t) \mathbf{R}_{e,t}^{-1} (\mathbf{y}_t - (\mathbf{I} \otimes \mathbf{X}_t) \hat{\mathbf{h}}_{t|t-1}) \end{aligned} \quad (4.27)$$

$$\hat{\mathbf{h}}_{t|T} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{P}_{t|t-1} \boldsymbol{\lambda}_{t|T} \quad (4.28)$$

The desired estimate is $\hat{\mathbf{h}}_{t|T}$.

4.4.2 Channel Estimation–Unknown Input Case

Consider the state-space model (4.20)–(4.21) and assume that the receiver does not have access to the transmitted data \mathbf{X}_0^T . The channel estimate at the j^{th} iteration

$\mathbf{h}_0^{T(j)}$ of the EM algorithm is obtained by applying the forward-backward Kalman (4.22)–(4.28) to the following state-space model [96]

$$\mathbf{h}_{t+1} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \mathbf{h}_t + (\mathbf{I}_{T_x R_x} \otimes \mathbf{G}) \mathbf{u}_t \quad (4.29)$$

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{0}_{T_x R_x (P+1) \times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{R_x} \otimes E[\mathbf{X}_t] \\ \mathbf{I}_{R_x} \otimes \text{Cov}[\mathbf{X}_t^*]^{1/2} \end{bmatrix} \mathbf{h}_t + \begin{bmatrix} \mathcal{N}_t \\ \underline{\mathbf{n}}_t \end{bmatrix} \quad (4.30)$$

where $\underline{\mathbf{n}}_t$ is virtual noise that is not physically present and that is independent of the physical noise \mathcal{N}_t .

To fully implement the EM algorithm, we need to initialize the algorithm and calculate the first and second moments of the input, which we do next.

4.4.3 Initial Channel Estimation

We obtain the initial channel estimate from the pilot/output equation (4.16) together with the dynamical channel model (4.5). Specifically, we do this by applying the FB Kalman to the following state-space model

$$\mathbf{h}_{t+1} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \mathbf{h}_t + (\mathbf{I}_{T_x R_x} \otimes \mathbf{G}) \mathbf{u}_t \quad (4.31)$$

$$\mathbf{y}_{t_{I_p}} = (\mathbf{I}_{R_x} \otimes \mathbf{X}_{t_{I_p}}) \mathbf{h}_t + \mathcal{N}_{t_{I_p}} \quad (4.32)$$

i.e., by applying the FB Kalman filter (4.22)–(4.28) with the following substitution

$$\mathbf{y}_t \longrightarrow \mathbf{y}_{t_{I_p}}, \quad \mathbf{X}_t \longrightarrow \mathbf{X}_{t_{I_p}}, \quad \text{and} \quad \mathbf{I}_{T_x R_x N} \longrightarrow \mathbf{I}_{T_x R_x |I_p|}$$

4.4.4 Data Detection

To detect the data, we use the data detection version of the I/O equation (4.18). Upon multiplying both sides by \mathbf{C}^* and taking the real part, we obtain

$$\tilde{\mathcal{Y}} = \|\mathcal{H}\|^2 \begin{bmatrix} \text{Re } \mathcal{S} \\ \text{Im } \mathcal{S} \end{bmatrix} + \tilde{\mathcal{N}} \quad (4.33)$$

where $\tilde{\mathcal{Y}}$ and $\tilde{\mathcal{N}}$ are $2N_u \times 1$ vectors defined by $\tilde{\mathcal{Y}} = \text{Re } \mathbf{C}^* \mathcal{Y}$ and $\tilde{\mathcal{N}} = \text{Re } \mathbf{C}^* \mathcal{N}$. Since \mathbf{C} is orthogonal, the noise $\tilde{\mathcal{N}}$ remains white, and the input can be detected on an element-by-element basis. We will now demonstrate how to detect the elements of $\text{Re } \mathcal{S}$ (the imaginary part can be treated similarly). So let $\mathcal{R} = \{r_1, \dots, r_{|\mathcal{R}|}\}$ denote the alphabet set from which the elements of $\text{Re } \mathcal{S}$ take their values. We can evaluate the conditional pdf $f(\text{Re } \mathcal{S}(n_u) | \tilde{\mathcal{Y}}(n_u))$ by applying Bayes rule on it which yields

$$\begin{aligned} f(r_i | \tilde{\mathcal{Y}}(n_u), \mathcal{H}) &= \frac{f(r_i, \tilde{\mathcal{Y}}(n_u) | \mathcal{H})}{f(\tilde{\mathcal{Y}}(n_u) | \mathcal{H})} \\ &= \frac{f(r_i, \tilde{\mathcal{Y}}(n_u) | \mathcal{H})}{\sum_{i=1}^{|\mathcal{R}|} f(\tilde{\mathcal{Y}}(n_u), r_i | \mathcal{H})} \\ &= \frac{f(\tilde{\mathcal{Y}}(n_u) | r_i, \mathcal{H}) f(r_i | \mathcal{H})}{\sum_{i=1}^{|\mathcal{R}|} f(\tilde{\mathcal{Y}}(n_u) | r_i, \mathcal{H}) f(r_i | \mathcal{H})} \\ f(r_i | \tilde{\mathcal{Y}}(n_u)) &= \frac{e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|^2 r_i|^2}{2\sigma_n^2}}}{\sum_{i=1}^{|\mathcal{R}|} e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|^2 r_i|^2}{2\sigma_n^2}}} \end{aligned} \quad (4.34)$$

We can use this pdf to calculate conditional expectation of $\text{Re } \mathcal{S}(n_u)$ and its second

moment given the output $\tilde{\mathcal{Y}}(n_u)$

$$E[\text{Re } \mathcal{S}(n_u) | \tilde{\mathcal{Y}}(n_u)] = \frac{\sum_{i=1}^{|\mathcal{R}|} r_i e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|_{r_i}^2}{2\sigma_n^2}}}{\sum_{i=1}^{|\mathcal{R}|} e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|_{r_i}^2}{2\sigma_n^2}}} \quad (4.35)$$

$$E[|\text{Re } \mathcal{S}(n_u)|^2 | \tilde{\mathcal{Y}}(n_u)] = \frac{\sum_{i=1}^{|\mathcal{R}|} r_i^2 e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|_{r_i}^2}{2\sigma_n^2}}}{\sum_{i=1}^{|\mathcal{R}|} e^{-\frac{|\tilde{\mathcal{Y}}(n_u) - \|\mathcal{H}\|_{r_i}^2}{2\sigma_n^2}}} \quad (4.36)$$

We can similarly calculate the two moments of the imaginary part. Now equation (4.35)–(4.36), just like (4.17)–(4.19), apply at a certain frequency tone n . So collecting (4.35) for all tones ($n = 1, \dots, N$) produces the two moments of the uncoded OFDM symbols. Specifically, we can calculate

$$E[\text{Re } \mathcal{S}(n_u)], \quad E[\text{Im } \mathcal{S}(n_u)], \quad E[\text{diag}(\text{Re } \mathcal{S}(n_u))^2], \quad \text{and} \quad E[\text{diag}(\text{Im } \mathcal{S}(n_u))^2] \quad (4.37)$$

We show in Appendix B that these moments are enough to characterize the first and second moments $E[\mathbf{X}]$ and $E[\mathbf{X}^* \mathbf{X}]$, which are needed for channel estimation.

4.4.5 Summary of the EM-Based FB Kalman

We now have all the elements for the iterative receiver for channel and data recovery, and for ease of reference, we summarize the receiver algorithm in the following. Given a sequence of input and output symbols \mathbf{X}_0^T and \mathbf{Y}_0^T , perform the following operations:

1. Calculate the initial channel estimate $\mathbf{h}_0^T(0)$ by applying the FB Kalman filter to the state-space model (4.31)–(4.32), i.e. by applying (4.22)–(4.28) with the

following substitutions

$$\mathbf{y}_t \longrightarrow \mathbf{y}_{t_{I_p}}, \quad \mathbf{X}_t \longrightarrow \mathbf{X}_{t_{I_p}}, \quad \text{and} \quad \mathbf{I}_{T_x R_x N} \longrightarrow \mathbf{I}_{T_x R_x |I_p|}$$

2. Iterate between the expectation and maximization steps for $j = 1, \dots, N_{\text{iter}}$:

(a) **Expectation:**

- Compute the first two moments of the uncoded OFDM symbols

$\mathcal{S}(1), \mathcal{S}(2), \dots, \mathcal{S}(n_u)$, given the output \mathbf{y}_0^T and the most recent estimate of the channel, $\mathbf{h}_0^T(j-1)$.

- Use these moments to calculate the moments of \mathbf{X} through the relationships (4.9) and (4.35).

(b) **Maximization:** Obtain the channel estimate $\mathbf{h}_0^T(j)$ by employing the FB

Kalman to the state-space model (4.29)–(4.30), i.e. by applying (4.22)–

(4.28) with the following substitutions

$$\begin{aligned} \mathbf{I}_{R_x} \otimes \mathbf{X}_t &\longrightarrow \begin{bmatrix} \mathbf{I}_{R_x} \otimes E[\mathbf{X}_t] \\ \mathbf{I}_{R_x} \otimes \text{Cov}[\mathbf{X}_t^*]^{1/2} \end{bmatrix}, \\ \mathbf{y}_t &\longrightarrow \begin{bmatrix} \mathbf{y}_t \\ \mathbf{0}_{T_x R_x (P+1) \times 1} \end{bmatrix}, \quad \text{and} \\ \mathbf{I}_{T_x R_x N} &\longrightarrow \mathbf{I}_{T_x R_x (N+P+1)} \end{aligned}$$

The algorithm can be stopped when the maximum number of iterations N_{iter} is reached

or when the difference between two consecutive estimates $\|\mathbf{h}_0^T(j) - \mathbf{h}_0^T(j-1)\|^2$ is

below a certain threshold.

4.5 Two Extensions/Modifications of the FB Kalman Filter

In the following, we present two modification of the FB Kalman filter described above.

4.5.1 Modification-I: Kalman- (Forward-Only) Based Estimation

One disadvantage of the FB Kalman (summarized in Subsection 4.4.5 above) is the storage and latency involved. The algorithm needs to wait for all $T + 1$ symbols before it can execute the backward run and hence obtain the channel estimate. One way around this is to reduce the window size T . Alternatively, we can run the filter in the forward direction only (i.e., run (4.22)–(4.26)) for both the initial estimation and the EM iteration. The algorithm then collapses to the Kalman-based filter proposed in [64] where the data and channel are recovered within one ST symbol.

4.5.2 Modification-II: Helix Based FB Kalman

The FB Kalman explained in Subsection 4.4.5 involves a forward run (in which channel and data are estimated using EM algorithm from one symbol and passed on to the next) followed by a backward run and this whole cycle is repeated for a certain number of iterations. As the iterations in this case resemble a cyclic process, so this Kalman

filter can be called Cyclic FB Kalman. One modification of this can be the case when in the forward run, the EM algorithm is applied for a certain number of iterations on one symbol so that the channel estimate is refined as much as possible before being passed on to the next symbol. In the end, a backward run can be applied once. As in this case the iterations resemble a helix, so it is called a Helix based Kalman filter.

4.6 Simulation Results

The transmitter and receiver illustrated in Figures 4.1 and 4.2 were implemented. The outer encoder is a rate 1/2 convolutional encoder and the coded bits are mapped to 16-QAM symbols using gray coding. We use the OSTBC commonly known as the Alamouti code with number of time slots $N_s = 2$ and number of transmitters $T_x = 2$ [94].

We use spatially correlated MIMO channel model (described in Appendix A) with transmit and receive correlation matrices

$$\mathbf{T}(p) = \begin{bmatrix} 1 & \zeta \\ \zeta & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{R}(p) = I$$

where $\zeta = 0.8$. Other parameters used are $\alpha = 0.8$, $\beta = 0.2$, and $P = 16$.

Packets are transmitted at each SNR value until a minimum number of errors occur. Each packet consists of 12 OFDM symbols transmitted over six ST blocks. Each OFDM symbol consists of 64 frequency tones and a cyclic prefix of length 16. The first ST block comprises of 16 pilots while the number of pilots in subsequent

blocks is fixed at twelve or mentioned otherwise. Two EM iterations were used in all cases.

In the following, we compare the performance of different implementation of Kalman filter and study the effect of using outer code and different number of pilots on the performance of the receiver.

4.6.1 Comparison of Kalman and FB Kalman over Spatially White and Spatially Correlated Channels

The performance of Kalman and the FB-Kalman is compared in Figure 4.3 over the two channel models i.e. spatially correlated and spatially white channel model. In both cases, soft estimate of data is used. The number of pilots used in this case was 12. It can be observed from Figure 4.3 that both Kalman and FB-Kalman perform closer to the perfect channel in the case of spatially correlated channel. So, spatial correlation (practical scenario) makes the performance of both Kalman and FB-Kalman better as compared to the case of spatially white channel model.

4.6.2 Comparison of Kalman, FB Kalman and Helix Based FB Kalman

Figure 4.4 compares the performance of the different implementations of the Kalman filter (Forward Only Kalman, Cyclic FB Kalman and Helix based FB Kalman) over spatially correlated channel. It can be seen that the helix based FB Kalman filter outperforms the other two implementations.

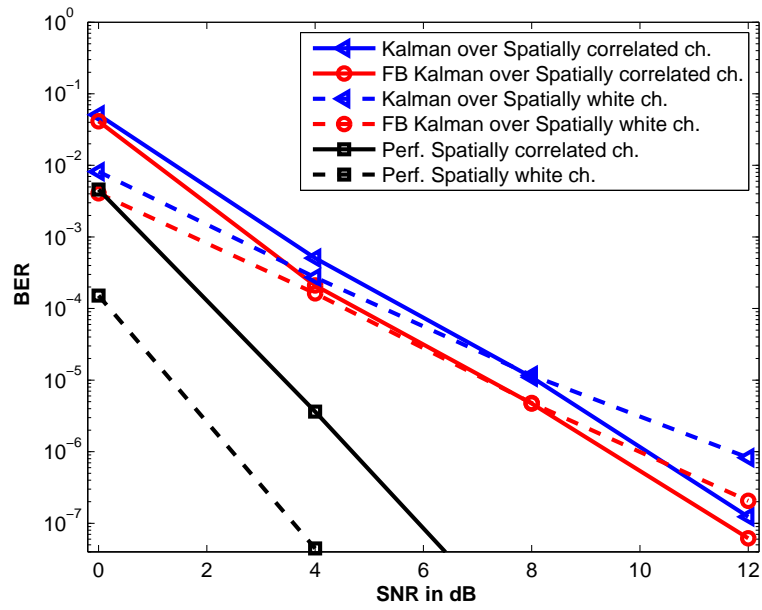


Figure 4.3: BER performance of Kalman and FB-Kalman using Soft data over spatially white and correlated channel models

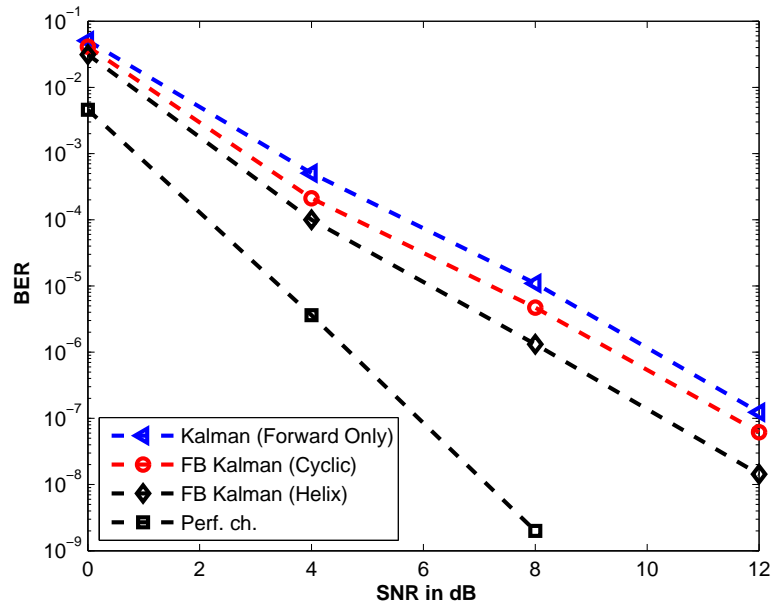


Figure 4.4: BER performance of Kalman, FB Kalman (Cyclic) and Helix based FB Kalman over spatially correlated channel

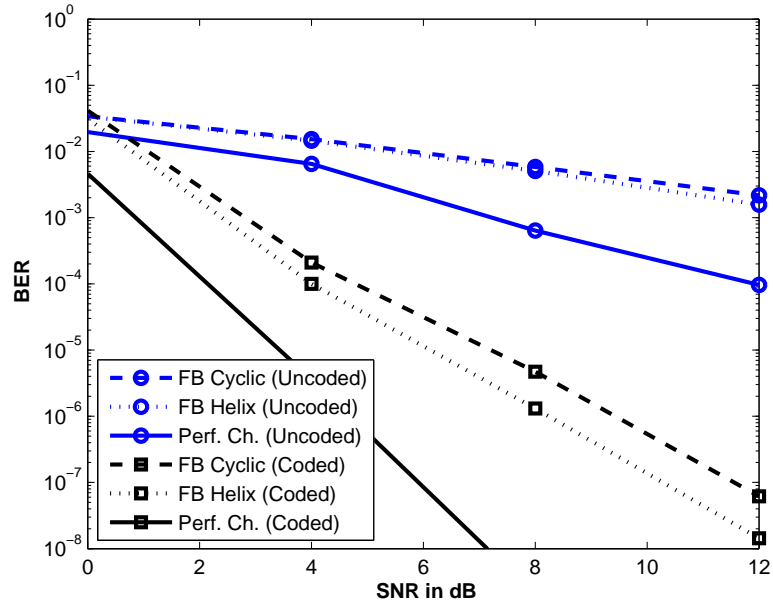


Figure 4.5: Effect of using outer code on performance of FB Kalman (Cyclic) and FB Kalman (Helix) over spatially correlated channel

4.6.3 Effect of Using Outer Code

In Figure 4.5, the effect of using outer code on the performance of the receiver employing the two implementations of FB Kalman over spatially correlated channel is presented. It can be observed that by using outer code, the performance of the receiver is improved quite significantly.

4.6.4 Effect of Using Different Number of Pilots

The sensitivity of the algorithm (using Cyclic FB Kalman and Helix based FB Kalman) to different number of pilots (six and twelve in this case) has been presented in Figure 4.6. Clearly, the twelve pilots case outperforms the six pilots case. The same conclusion can be drawn from Figure 4.7 in which the performance of the two imple-

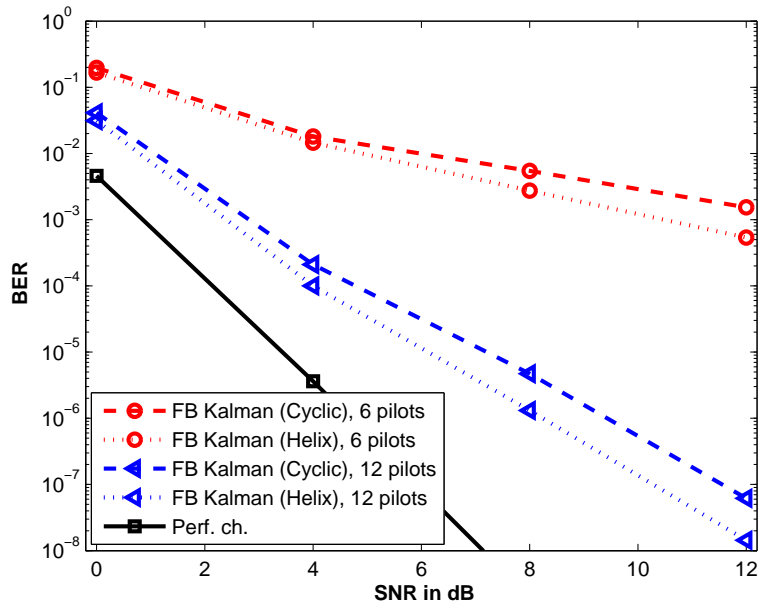


Figure 4.6: Sensitivity of FB Kalman (Cyclic) and FB Kalman (Helix) to different number of pilots using an outer code

mentations of FB Kalman filter has been compared without using the outer code.

4.7 Conclusion

In this chapter we considered a semiblind design for space time block coded MIMO-OFDM transmission over time-variant spatially correlated channels. This is an extension of the work done by [96] which proposed a Kalman filter approach to channel estimation in spatially white MIMO-OFDM systems. All possible constraints on the channel (the finite delay spread, frequency, time, and spatial correlation) and the data (the finite alphabet constraint, the cyclic prefix, pilots and the orthogonal space time block coding) were utilized. The channel estimation part boils down to EM based forward backward Kalman filter. A relaxed (forward-only) version of the algorithm

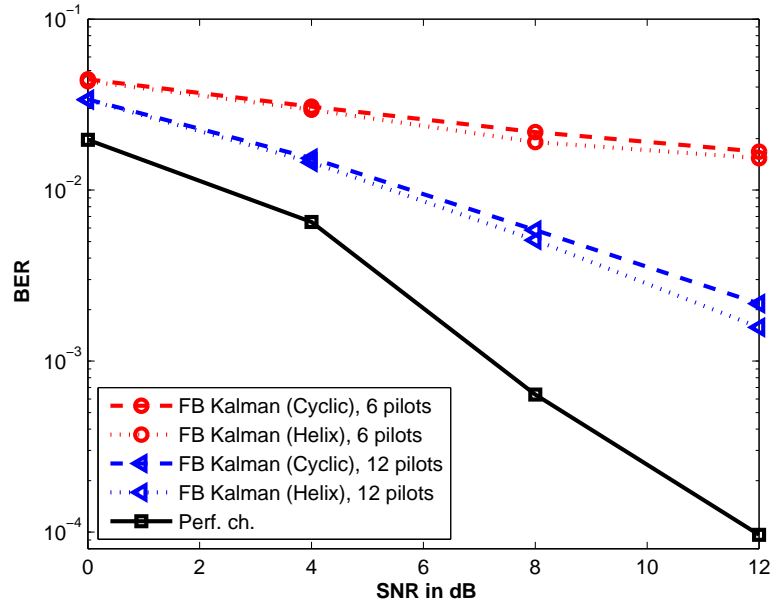


Figure 4.7: Sensitivity of FB Kalman (Cyclic) and FB Kalman (Helix) to different number of pilots without using an outer code

that is able to perform recovery with no latency and hence avoid the delay and storage shortcomings of the FB-Kalman was also considered. A different implementation of FB Kalman called Helix based FB Kalman filter was proposed and compared with the (forward-only) Kalman and FB Kalman (cyclic). Simulation results show the improvement in the performance of the receiver when it employs helix based FB Kalman filter. The performance of the receiver in the presence of an outer code and reduced number of pilots was also presented.

APPENDIX A: Channel Model in the Presence of Spatial Correlation

In what follows, the spatially white channel model described in Section 4.2.2 is presented for the practical scenario of spatial correlation between the MIMO channels. It was derived in [96] but its effect on the performance of the receiver was not studied.

We start by presenting the transmit correlation case, and then generalize the results to deal with the general (transmit and receive) correlation case. In the transmit correlation case, $\mathbf{H}(p)$, the MIMO impulse response at tap p , is given by

$$\mathbf{H}(p) = \mathbf{W}(p)\mathbf{T}^{1/2}(p) \quad (\text{A-1})$$

where $\mathbf{T}^{1/2}(p)$ is the transmit correlation matrix (of size T_x) at tap p and where $\mathbf{W}(p)$ consists of iid elements. The matrix $\mathbf{W}(p)$ remains constant over a single ST block and varies from one ST block to the next according to³

$$\mathbf{W}_{t+1}(p) = \alpha(p)\mathbf{W}_t(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}\mathbf{U}_t(p) \quad (\text{A-2})$$

where $\alpha(p)$, β , and $\mathbf{U}_t(p)$ are as defined in Subsection 4.2.2.

Just as we did in Subsection 4.2.2, we would like to construct a recursion for the tap $h_{r_x}^{t_x}(p)$ and subsequently scale it up for the SISO and MIMO cases. Now since $h_{r_x}^{t_x}(p)$ is the (r_x, t_x) element of $\mathbf{H}(p)$, we deduce from (A-1) that it is the inner product

³We suppress the time dependence at times for notational convenience.

of the r_x row of $\mathbf{W}(p)$ and the t_x column of $\mathbf{T}^{1/2}$, i.e.

$$h_{r_x}^{t_x}(p) = \mathbf{w}_{r_x}(p)\mathbf{t}^{t_x}(p) \quad (\text{A-3})$$

Moreover, from (A-2), we have the following recursion for $\mathbf{w}_{r_x}(p)$

$$\mathbf{w}_{r_x,t+1}(p) = \alpha(p)\mathbf{w}_{r_x,t}(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}\mathbf{u}_{r_x,t}(p)$$

Post-multiplying both sides by $\mathbf{t}^{t_x}(p)$ yields

$$\mathbf{w}_{r_x,t+1}(p)\mathbf{t}^{t_x}(p) = \alpha(p)\mathbf{w}_{r_x,t}(p)\mathbf{t}^{t_x}(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}\mathbf{u}_{r_x,t}(p)\mathbf{t}^{t_x}(p)$$

This means that $h_{r_x}^{t_x}(p)$ satisfies the dynamical equation

$$h_{r_x,t+1}^{t_x}(p) = \alpha(p)h_{r_x,t}^{t_x}(p) + \sqrt{(1 - \alpha^2(p))e^{-\beta p}}ut_{r_x,t}^{t_x}(p) \quad (\text{A-4})$$

where $ut_{r_x}^{t_x}$ is defined by

$$ut_{r_x}^{t_x}(p) = \mathbf{u}_{r_x}(p)\mathbf{t}^{t_x}(p)$$

Concatenating (A-4) for $p = 1, 2, \dots, P$ yields a dynamic equation for the impulse response

$$\mathbf{h}_{r_x}^{t_x} = \begin{bmatrix} h_{r_x}^{t_x}(0) \\ \vdots \\ h_{r_x}^{t_x}(P) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{r_x}(0)\mathbf{t}^{t_x}(0) \\ \vdots \\ \mathbf{w}_{r_x}(P)\mathbf{t}^{t_x}(P) \end{bmatrix}$$

which is the same as the dynamic equation (see (4.4)) for the spatially uncorrelated

case

$$\mathbf{h}_{r_x, t+1}^{t_x} = \mathbf{F}\mathbf{h}_{r_x, t}^{t_x} + \mathbf{G}\mathbf{u}_{r_x, t}^{t_x} \quad (\text{A-5})$$

The only difference from the uncorrelated case is that $\mathbf{u}_{r_x}^{t_x}$ is no more white. Rather, we have

$$\begin{aligned} E[\mathbf{u}_{r_x}^{t_x} \mathbf{u}_{r_x}^{t_x*}] &\triangleq E \begin{bmatrix} \mathbf{u}_{r_x}(0)\mathbf{t}^{t_x}(0) \\ \mathbf{u}_{r_x}(1)\mathbf{t}^{t_x}(1) \\ \vdots \\ \mathbf{u}_{r_x}(P)\mathbf{t}^{t_x}(P) \end{bmatrix} \\ &\quad \begin{bmatrix} \mathbf{t}^{t_x*}(0)\mathbf{u}_{r_x}^*(0) & \mathbf{t}^{t_x*}(1)\mathbf{u}_{r_x}^*(1) & \cdots & \mathbf{t}^{t_x*}(P)\mathbf{u}_{r_x}^*(P) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{t}\mathbf{t}_{t_x}^{t_x}(0) & & & \\ & \mathbf{t}\mathbf{t}_{t_x}^{t_x}(1) & & \\ & & \ddots & \\ & & & \mathbf{t}\mathbf{t}_{t_x}^{t_x}(P) \end{bmatrix} \triangleq \text{diag}(\mathbf{t}\mathbf{t}_{t_x}^{t_x}) \end{aligned}$$

where

$$\mathbf{t}\mathbf{t}_{r_x}^{t_x} = \begin{bmatrix} \mathbf{t}^{r_x*}(0)\mathbf{t}^{t_x}(0) \\ \mathbf{t}^{r_x*}(1)\mathbf{t}^{t_x}(1) \\ \vdots \\ \mathbf{t}^{r_x*}(P)\mathbf{t}^{t_x}(P) \end{bmatrix} = \begin{bmatrix} \mathbf{t}_{r_x}(0)\mathbf{t}^{t_x}(0) \\ \mathbf{t}_{r_x}(1)\mathbf{t}^{t_x}(1) \\ \vdots \\ \mathbf{t}_{r_x}(P)\mathbf{t}^{t_x}(P) \end{bmatrix}$$

and where the second line follows from the fact that $\mathbf{t}^{r_x*}(p) = \mathbf{t}_{t_x}(p)$ since $\mathbf{T}^{1/2}(p)$ is

conjugate symmetric. In general, we can show that

$$E[\mathbf{u}\mathbf{t}_{r_x}\mathbf{u}\mathbf{t}_{r'_x}^*] = \begin{bmatrix} \text{diag}(\mathbf{t}\mathbf{t}_1^1) & \text{diag}(\mathbf{t}\mathbf{t}_1^2) & \cdots & \text{diag}(\mathbf{t}\mathbf{t}_1^{T_x}) \\ \text{diag}(\mathbf{t}\mathbf{t}_2^1) & \text{diag}(\mathbf{t}\mathbf{t}_2^2) & \cdots & \text{diag}(\mathbf{t}\mathbf{t}_2^{T_x}) \\ \vdots & \vdots & \cdots & \vdots \\ \text{diag}(\mathbf{t}\mathbf{t}_{T_x}^1) & \text{diag}(\mathbf{t}\mathbf{t}_{T_x}^2) & \cdots & \text{diag}(\mathbf{t}\mathbf{t}_{T_x}^{T_x}) \end{bmatrix}$$

for $r_x = r'_x$ and is zero otherwise. Alternatively, we can write this as

$$E[\mathbf{u}\mathbf{t}_{r_x}\mathbf{u}\mathbf{t}_{r'_x}^*] = \begin{cases} \sum_{p=0}^P \mathbf{T}(p) \otimes (\underline{\mathbf{I}}^p \mathbf{B} \bar{\mathbf{I}}^p) & \text{for } r_x = r'_x \\ \mathbf{O} & \text{otherwise} \end{cases}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \underline{\mathbf{I}} = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & \ddots & & \\ & & \ddots & 0 & \\ & & & & 1 & 0 \end{bmatrix},$$

$$\text{and } \bar{\mathbf{I}} = \begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & 1 & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix},$$

Collecting (A-5) for all transmit and receive antennas yields

$$\mathbf{h}_{t+1} = (\mathbf{I}_{T_x R_x} \otimes \mathbf{F}) \mathbf{h}_t + (\mathbf{I}_{T_x R_x} \otimes \mathbf{G}) \mathbf{u} \mathbf{t}_t \quad (\text{A-6})$$

where

$$\begin{aligned} E[\mathbf{u} \mathbf{u}^*] &= \mathbf{I}_{R_x} \otimes E[\mathbf{u} \mathbf{t}_{r_x} \mathbf{u} \mathbf{t}_{r_x}^*] \\ &= \sum_{p=0}^P \mathbf{I}_{R_x} \otimes \mathbf{T}(p) \otimes (\underline{\mathbf{I}}^p \mathbf{B} \bar{\mathbf{I}}^p) \end{aligned} \quad (\text{A-7})$$

When the channel exhibits both transmit and receive correlation, the IR \mathbf{h} continues to satisfy the dynamical equation (A-6) except that the correlation of the innovation \mathbf{u} is now given by

$$E[\mathbf{u} \mathbf{u}^*] = \sum_{p=0}^P \mathbf{R}(p) \otimes \mathbf{T}(p) \otimes (\underline{\mathbf{I}}^p \mathbf{B} \bar{\mathbf{I}}^p)$$

APPENDIX B: Calculating the Moments of \mathbf{X}

In this appendix, we demonstrate that the four moments (4.37) of the uncoded OFDM symbol $\mathcal{S}(n_u)$ are enough to calculate the first two moments of \mathbf{X} , $E[\mathbf{X}]$ and $E[\mathbf{X}^* \mathbf{X}]$. Since \mathbf{X} depends linearly on $\text{Re } \mathcal{S}(n_u)$ and $\text{Im } \mathcal{S}(n_u)$ (see (4.9) and (4.11)), it is straight forward to calculate the mean of \mathbf{X} starting from the means of $\text{Re } \mathcal{S}(n_u)$ and $\text{Im } \mathcal{S}(n_u)$. Now from (4.11), we note that evaluating $E[\mathbf{X}^* \mathbf{X}]$ boils down to evaluating the cross

correlation $E[\text{diag}(\boldsymbol{\mathcal{X}}_i^*(n_c))\text{diag}(\boldsymbol{\mathcal{X}}_j(n'_c))]$ Recall also that

$$\boldsymbol{\mathcal{X}}_{t_x}(n_c) = \sum_{n_u=1}^{N_u} a_{t_x, n_c}(n_u) \text{Re } \boldsymbol{\mathcal{S}}(n_u) + j b_{t_x, n_c}(n_u) \text{Im } \boldsymbol{\mathcal{S}}(n_u) \quad (\text{B-1})$$

This means that calculating the cross expectation boils down to calculating the cross correlation of $\text{Re } \boldsymbol{\mathcal{S}}(n_u)$, $\text{Im } \boldsymbol{\mathcal{S}}(n_u)$, $\text{Re } \boldsymbol{\mathcal{S}}(n'_u)$, and $\text{Im } \boldsymbol{\mathcal{S}}(n'_u)$ for $n_u, n'_u = 1, \dots, N_u$. It is easy to see that these variables are independent for $n_u \neq n'_u$. Moreover, since the noise in (4.33) is white, one can also see that $\text{Re } \boldsymbol{\mathcal{S}}(n_u)$ and $\text{Im } \boldsymbol{\mathcal{S}}(n_u)$ are independent. As a result, we can completely characterize the cross correlation $E[\text{diag}(\boldsymbol{\mathcal{X}}_i^*(n_c))\text{diag}(\boldsymbol{\mathcal{X}}_j(n'_c))]$ and hence the expectations $E[\boldsymbol{\mathcal{X}}]$ and $E[\boldsymbol{\mathcal{X}}^* \boldsymbol{\mathcal{X}}]$ starting from the first and second moments of (4.37).

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Concluding Remarks

This Thesis has considered blind and semiblind algorithms for channel estimation and data recovery in OFDM transmission. The first part of the Thesis presented a blind algorithm with data-centered approach. The blind algorithm boils down to an integer nonlinear-least squares problem which becomes simpler in the case of constant modulus data. It was proved that the data can be recovered in an OFDM system from output data only, irrespective of the channel zero locations and the quality of channel estimate. To reduce the high computational complexity involved in the blind algorithm, several iterative methods were proposed. It was found that using Newton's method (initialized with a rough estimate of data) performed better than all other iterative methods proposed. It was also shown that how the CP can be used to improve

the performance of the OFDM receiver whether operating in blind, semiblind, training or perfectly known channel modes. Specifically, in this method, data is recovered using both the linear and circular subchannels as opposed to the conventional method which utilized only circular subchannel. This method performs well, especially for the cases when channel has zeros on FFT grid.

The second part of Thesis presented a semiblind channel-centered approach. It was proved in this part that the transmitted data in OFDM can be assumed Gaussian. As the noise is also considered Gaussian, this makes the output also Gaussian and its pdf is easily evaluated. The channel is estimated by maximizing the log likelihood function given the output pdf. The experimental results unfortunately demonstrated that the log likelihood function is not unimodal when it is plotted against the channel taps. Therefore, a semiblind approach was adopted using the steepest descent algorithm initialized by a rough estimate of channel obtained by using a few pilots and channel correlation. Methods to reduce computational complexity involved in evaluating the gradient and log likelihood function were also presented.

The last part of Thesis considered a semiblind design for space time block coded MIMO-OFDM transmission over time-variant spatially correlated channels. All possible constraints on the channel (the finite delay spread, frequency, time, and spatial correlation) and the data (the finite alphabet constraint, the cyclic prefix, pilots and the orthogonal space time block coding) were utilized. The channel estimation part boils down to EM based forward backward (FB) Kalman filter. Two modification/extensions of the FB Kalman filter were presented which include a relaxed

(forward-only) version of the algorithm (that is able to perform recovery with no latency) and a helix based FB Kalman filter. The performance of the receiver by employing the three different implementations of the Kalman filter was presented. Simulation results showed that the helix based FB Kalman filter outperforms the other two Kalman filters. The performance of the receiver in the presence of an outer code and by using reduced number of pilots was also studied.

5.2 Future Work

5.2.1 General Time Variant Case

This Thesis deals with block fading channels i.e. the channel is assumed to be constant during the transmission of one block. By block, we mean an OFDM symbol for SISO transmission and a space-time block for MIMO transmission. It is more realistic to assume that the channel continuously varies with time which is a future research problem which can be build upon the findings in this Thesis. Specifically, we will assume that the channel varies within the OFDM symbol (resulting in intercarrier interference (ICI)) and use the various constraints used in this Thesis to perform channel estimation, ICI cancelation, and data detection.

5.2.2 Iterative Methods for Non-Constant Modulus Data

In Chapter 2, the blind algorithm presented is valid for constant as well as non-constant modulus data. The emphasis in the chapter was on constant modulus data

as things simplified in this case. The approximate methods proposed to reduce the computational complexity involved in the blind algorithm are also applicable only to constant modulus data. Iterative methods for non-constant modulus data can also be developed similar to the methods presented in this Thesis. Of particular importance is the Newton's method that was developed and showed very good performance when initialized using pilots. It is easy to evaluate the gradient and the hessian for the non-constant modulus case in a similar manner and use it for semiblind data detection.

5.2.3 Improving the Performance of the Semiblind Channel Centered Receiver

In Chapter 3, a semiblind technique was presented for the channel estimation in OFDM transmission. Steepest descent algorithm initialized with a noisy estimate of channel was used. The gradient (of the log likelihood function with respect to the channel) involved in the steepest descent algorithm was derived. One way to enhance the performance of the algorithm is to evaluate the hessian of the log likelihood function with respect to the channel and use Newton's method [92]. Another way to improve the algorithm is to devise some techniques (e.g. Genetic algorithm) to find the global maximum of the likelihood function.

REFERENCES

- [1] Z. Wang and G. B. Giannakis, “Wireless multicarrier communications: Where Fourier meets Shannon,” *IEEE Signal Proc. Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
- [2] G. Stuber, J. Barry, S. McLaughlin, Y. Li, M. Ingram, and T. Pratt, “Broadband MIMO-OFDM wireless communications,” *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.
- [3] E. Chen, R. Tao, and X. Zhao, “Channel Equalization for OFDM System Based on the BP Neural Network,” *Int. Conf. Signal Proc.*, vol. 3, Nov. 2006.
- [4] T. Y. Al-Naffouri, “An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM,” *IEEE Trans. Signal Proc.*, vol. 55, no. 7, pp. 3924–3930, Jul. 2007.
- [5] N. Wang and S. D. Blostein, “Adaptive zero-padding OFDM over frequency-selective multipath channels,” *EURASIP Journal on App. Signal Proc.*, pp. 1478–1488, Oct. 2004.

- [6] X. Wang and R. Liu, "Adaptive channel estimation using cyclic prefix in multi-carrier modulation system," *IEEE Commun. Lett.*, vol. 3, no. 10, pp. 291–293, Oct. 1999.
- [7] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.
- [8] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Proc. Lett.*, vol. 4, no. 6, pp. 184–187, Jun. 1997.
- [9] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136–2148, Dec. 2002.
- [10] B. Muquet, M. de Courville, G. B. Giannakis, Z. Wang, and P. Duhamel, "Reduced complexity equalizers for zero-padded OFDM transmissions," *IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. 5, pp. 2973–2976, Jun. 2000.
- [11] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," *IEEE Veh. Technol. Conf.*, vol. 2, pp. 815–819, Jul. 1995.
- [12] O. Edfors, M. Sandell, J. van de Beek, K. S. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Signal Proc.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

- [13] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [14] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Proc.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.
- [15] M. Biguesh and A. B. Gershman, "MIMO channel estimation: optimal training and tradeoffs between estimation techniques," *IEEE Int. Conf. Commun.*, vol. 5, pp. 2658–2662, Jun. 2004.
- [16] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electr.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.
- [17] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Vehicular Tech. Conf.*, vol. 2, pp. 1131–1135, May 1999.
- [18] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.
- [19] M. Shin, H. Lee, and C. Lee, "Enhanced channel estimation technique for MIMO–OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 261–265, Jan. 2004.

- [20] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
- [21] I. Barhumi, G. Leus, and M. Moonen, "Optimal design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Proc.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [22] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Vehicular Tech. Conf.*, vol. 3, pp. 1639–1643, May 1997.
- [23] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.
- [24] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
- [25] R. W. Heath and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Proc.*, vol. 47, no. 3, pp. 848–856, Mar. 1999.
- [26] B. Muquet, M. de Courville, P. Duhamel, and V. Buzenac, "A subspace based blind and semi-blind channel identification method for OFDM systems," *IEEE Workshop on Signal Proc. Advances in Wireless Comm.*, pp. 170–173, May 1999.

- [27] A. P. Petropulu, R. Zhang, and R. Lin, “Blind OFDM channel estimation through simple linear precoding,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 647–655, Mar. 2004
- [28] Z. Wang and G. B. Giannakis, “Linear precoded or coded OFDM against wireless channel fades?,” *IEEE Workshop Sig. Proc. Adv. Wireless Commun.*, pp. 266–270, Mar. 2001
- [29] W. Bai, C. He, L. Jiang, and S. China, “Blind channel estimation in MIMO-OFDM Systems,” *IEEE GLOBECOM*, vol. 1, pp. 317–321, Nov. 2002.
- [30] H. Bölcskei, R. W. Heath, and A. J. Paulraj, “Blind channel identification and equalization in OFDM-based multi-antenna systems,” *IEEE Trans. Signal Proc.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.
- [31] G. A. Al-Rawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, “Exploiting error-control coding and cyclic-prefix in channel estimation for coded OFDM systems,” *IEEE Commun. Lett.*, vol. 7, no. 8, pp. 388–390, Aug. 2003.
- [32] X. G. Doukopoulos and G.V. Moustakides, “Blind adaptive channel estimation in OFDM systems,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1716–1725, Jul. 2006.
- [33] R. Zhang and W. Chen, “A mixture Kalman filter approach for blind OFDM channel estimation,” *Asilomar Conf. on Signals, Syst., and Computers*, vol. 1, pp. 350–354, Nov. 2004.

- [34] T. Petermann, S. Vogeler, K. Kammeyer, and D. Boss, "Blind turbo channel estimation in OFDM receivers," *Asilomar Conf. on Signals, Syst., and Computers*, vol. 2, pp. 1489–1493, Nov. 2001.
- [35] G. B. Giannakis and C. Tepedelenlioglu, "Direct blind equalizers of multiple FIR channels: a deterministic approach," *IEEE Trans. Signal Proc.*, vol. 47, no. 1, pp. 62–74, January 1999.
- [36] H. Wang, Y. Lin, and B. Chen, "Data-efficient blind OFDM channel estimation using receiver diversity," *IEEE Trans. Signal Proc.*, vol. 51, no. 10, pp. 2613–2623, Oct. 2003.
- [37] D. Xu and L. Yang, "Subspace-based blind channel estimation for STBC-OFDM," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 4, pp. IV – IV, May 2006.
- [38] S. Roy and C. Li, "A subspace blind channel estimation method for OFDM systems without cyclic prefix," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 572–579, Oct. 2002.
- [39] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. Signal Proc.*, vol. 50, no. 5, pp. 1215–1228, May 2002.
- [40] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Proc.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

- [41] L. Chengyang and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141–150, Jan. 2003.
- [42] C. Shin and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems using virtual carriers," *IEEE GLOBECOM*, vol. 4, pp. 2465–2469, Dec. 2004.
- [43] Y. Song, S. Roy, and L. A. Akers, "Joint blind estimation of channel and data symbols in OFDM," *IEEE Vehicular Tech. Conf.*, vol. 1, pp. 46–50, May 2000.
- [44] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM Systems," *IEEE Trans. Vehicular Tech.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.
- [45] Z. Liu, G. B. Giannakis, A. Scaglione, and S. Barbarossa, "Decoding and equalization of unknown multipath channels based on block precoding and transmit-antenna diversity," *Asilomar Conf. on Signals, Syst., and Computers*, vol. 2, pp. 1557–1561, Oct. 1999.
- [46] T. Kim and I. Eo, "Reliable blind channel estimation scheme based on cross-correlated cyclic prefix for OFDM system," *Int. Conf. Adv. Commun. Technol.*, vol. 1, Feb. 2006.
- [47] F. Gao and A. Nallanathan, "Blind channel estimation for OFDM systems via a general non-redundant precoding," *IEEE Int. Conf. Commun.*, vol. 10, pp. 4612–4617, Jun. 2006.

- [48] F. Gao and A. Nallanathan, “Blind channel estimation for MIMO OFDM systems via nonredundant linear precoding,” *IEEE Trans. Signal Proc.*, vol. 55, no. 2, pp. 784–789, Jan. 2007.
- [49] R. Lin and A. P. Petropulu, “Linear precoding assisted blind channel estimation for OFDM systems,” *IEEE Trans. Vehicular Tech.*, vol. 54, no. 3, pp. 983–995, May 2005.
- [50] Y. Liang, H. Luo, and J. Huang, “Redundant precoding assisted blind channel estimation for OFDM systems,” *Int. Conf. Signal Proc.*, vol. 3, 2006.
- [51] M. Chang and Y. T. Su, “Blind and semiblind detections of OFDM signals in fading channels,” *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 744–754, May 2004.
- [52] T. Y. Al-Naffouri, O. Awoniyi, O. Oteri, and A. Paulraj, “Receiver design for MIMO-OFDM transmission over time variant channels,” *IEEE GLOBECOM*, vol. 4, pp. 2487–2492, Dec. 2004.
- [53] B. Muquet and M. de Courville, “Blind and semi-blind channel identification methods using second order statistics for OFDM systems,” *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 5, pp. 2745–2748, Mar. 1999.
- [54] M. C. Necker and G. L. Stuber, “Totally blind channel estimation for OFDM on fast varying mobile radio channels,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1514–1525, Sep. 2004.

- [55] K. Y. Ho and S. H. Leung, "A generalized semi-blind channel estimation for pilot-aided OFDM systems," *IEEE Int. Symp. Circuits and Systems*, vol. 6, pp. 6086–6089, May 2005.
- [56] S. Zhou, B. Muquet, and G. B. Giannakis, "Semi-blind channel estimation for block precoded space-time OFDM transmissions," *Proceedings IEEE Signal Proc. Workshop on Statistical Signal Proc.*, pp. 381–384, Aug. 2001.
- [57] T. Y. Al-Naffouri, D. Toumpakaris, A. Bahai and A. Paulraj, "An adaptive semi-blind algorithm for channel identification in OFDM," *Asilomar Conf. on Signals, Syst., and Computers*, vol. 2, pp. 921–925, Nov. 2001.
- [58] F. Yang and W. Ser, "Adaptive semi-blind channel estimation for OFDM systems," *IEEE Veh. Technol. Conf.*, vol. 3, pp. 1773–1776, May 2004.
- [59] T. Cui and C. Tellambura, "Joint data detection and channel estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 54, no.4, pp. 670–679, Apr. 2006.
- [60] T. Cui and C. Tellambura, "Semi-blind channel estimation and data detection for OFDM systems over frequency-selective fading channels," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 3, pp. iii/597–iii/600, Mar. 2005.
- [61] W. Yang, Y. Cai, and Y. Xun, "Semi-blind Channel estimation for OFDM Systems," *IEEE Vehicular Tech. Conf.*, vol. 1, pp. 226–230, 2006.
- [62] B. Yang, K. B. Letaief, R. S. Cheng, and C. Zhigang, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modelling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

- [63] W. Kunji, Z. Jianhua, L. Chaojun, and H. Chen, "Semi-blind OFDM channel estimation using receiver diversity in the presence of virtual carriers," *Int. Conf. Commun. and Networking*, pp. 1–4, Oct. 2006.
- [64] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj, "An EM-based OFDM receiver for time-variant channels," *IEEE Globecom*, vol. 1, pp. 589–593, Nov. 2002.
- [65] T. Y. Al-Naffouri, "Receiver design for MIMO OFDM transmission over time variant channels," *IEEE Workshop on Signal Proc. Advances in Wireless Comm.*, pp. 1–6, Jun. 2007.
- [66] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space time wireless communications*. Cambridge University Press, 2003.
- [67] Z. Shengli and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1402–1414, Aug. 2001.
- [68] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj, "Semi-blind channel identification and equalization in OFDM: an expectation-maximization approach," *IEEE Vehicular Tech. Conf.*, vol. 1, pp. 13–17, Sep. 2002.
- [69] B. Lu, X. Wang, and Y. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 213–225, Apr. 2002.

- [70] Y. Li, C. N. Georghiades, and G. Huang, "Iterative maximum-likelihood sequence estimation for space-time coded systems," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 948–951, Jun. 2001.
- [71] C. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Proc.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.
- [72] Y. Xie and C. N. Georghiades, "Two EM-type channel estimation algorithms for OFDM with transmitter diversity," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 106–115, Jan. 2003.
- [73] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.
- [74] G. Stüber, *Principles of Mobile Communication*. Kluwer Academic, 2001.
- [75] R.A. Iltis, "Joint estimation of PN code delay and multipath using the extended Kalman filter," *IEEE Trans. Commun.*, vol. 38, no. 10, pp. 1677–1685, Oct. 1990.
- [76] C. Komninakis, C. Fragouli, A. H. Sayed, and R. Wesel, "Multi-input -input multi-output fading channel tracking and equalization using estimation," *IEEE Trans. Signal Proc.*, vol. 50, no. 5, pp. 5, pp. 1065–1076, May 2002.

- [77] I. Kang, M. P. Fitz, and S. B. Gelfand, "Blind estimation of multipath channel parameters: a modal analysis approach," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1140–1150, Aug. 1999.
- [78] M.C. Vanderveen, A. -J. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Proc.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.
- [79] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 2, pp. 1093–1096, May 1996.
- [80] T. Y. Al-Naffouri and A. Paulraj, "A forward-backward Kalman for the estimation of time-variant channels in OFDM," *IEEE Workshop on Signal Proc. Advances in Wireless Comm.*, pp. 670–674, Jun. 2005.
- [81] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE Int. Conf. on Neural Networks*, vol. 4, pp. 1942–1948, Nov. 1995.
- [82] M. Clerc and J. Kennedy, "The particle swarm: Explosion, stability, and convergence in a multi-dimensional complex space," *IEEE Trans. Evol. Comput.*, no. 1, vol. 6, pp. 58–73, Feb. 2002.
- [83] R. Mendes, J. Kennedy, and J. Neves, "The fully Informed Particle Swarm: Simpler May be Better," *IEEE Trans. Evol. Comput.*, vol 8, no.3, pp. 204–210, Jun. 2004.

- [84] K. C. Sharman and G. D. McClurkin, “Genetic algorithms for maximum likelihood parameter estimation,” *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 4, pp. 2716–2719, May 1989.
- [85] S. Chen and Y. Wu, “Maximum likelihood joint channel and data estimation using genetic algorithms,” *IEEE Trans. Signal Proc.*, vol. 46, no. 5, pp. 1469–1473, May 1998.
- [86] A. Hjørungnes, D. Gesbert, and D. P. Palomar, “Unified theory of complex-valued matrix differentiation,” *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 3, pp. III-345–III-348, Apr. 2007.
- [87] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons, Inc., 1999.
- [88] H. Lutkepohl, *Handbook of Matrices*. John Wiley and Sons, Inc., 1996.
- [89] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Version 2006.
- [90] A. Benveniste, M. Goursat, and G. Ruget, “Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications,” *IEEE Trans. on Automatic Control*, vol. 25, no. 3, pp. 385–399, Jun. 1980.
- [91] G. B. Giannakis and J. M. Mendel, “Identification of nonminimum phase systems using higher order statistics,” *IEEE Trans. on Acoust., Speech, and Signal Proc.*, vol. 37, no. 3, pp. 360–377, Mar. 1989.

- [92] Ali H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley and Sons, Inc., 2003.
- [93] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [94] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [95] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.
- [96] T. Y. Al-Naffouri, *Adaptive Algorithms for Wireless Channel Estimation: Transient Analysis and Semi-Blind Design*, Stanford University, Jan. 2005.

Vitae

- Ahmed Abdul Quadeer.
- Born in Jeddah, Saudi Arabia on November 6th, 1983.
- Received Bachelor of Engineering (BE.) degree in Electronics from NED University of Engineering & Technology, Karachi, Pakistan in January 2006.
- Joined King Fahd University of Petroleum & Minerals in September 2006.
- Email: aquadeer@kfupm.edu.sa

List of Publications

- T. Y. Al-Naffouri and **A. Quadeer**, “A Forward-Backward Kalman Filter based STBC MIMO OFDM Receiver,” *submitted to EURASIP Signal Processing Advances for Wireless Communications*.
- T. Y. Al-Naffouri and **A. Quadeer**, “Cyclic Prefix based Enhanced Data Recovery in OFDM,” *submitted to IEEE Transactions on Signal Processing*.
- T. Y. Al-Naffouri and **A. Quadeer**, “Blind Maximum Likelihood Data Recovery in OFDM,” *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, Apr. 2008.
- **A. Quadeer**, T. Y. Al-Naffouri and M. Shadaydeh, “Iterative Blind Data Detection in Constant Modulus OFDM Systems,” *European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.