Using the Cyclic Prefix for Blind Equalization in OFDM

Tareq Y. Al-Naffouri

Department of Electrical Engineering

KFUPM, Dhahran 31261

Saudi Arabia

e-mail: naffouri@kfupm.edu.sa

SABIC & FAST-TRACK Project

Project No.: SB080007

Final Report

Contents

Abstract

Orthogonal Frequency Division Multiplexing (OFDM) combines the advantages of high achievable rates and relatively easy implementation. However, for proper recovery of the input, the OFDM receiver needs accurate channel information. In this project, we show how the cyclic prefix can be used to enhance the performance of an OFDM receiver. In the first part of the project, we consider blind data detection for OFDM transmission over block fading channels. Specifically, we show how an OFDM symbol can be blindly detected using the 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 completely 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. In the second part, we propose six iterative methods to reduce the computational complexity involved in the maximum likelihood search of input. In the last part of the project, 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).

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 HIPERLAN/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 project is concerned with (semi) blind receivers for OFDM over block fading channels.

1.1 The Need for Channel Estimation in a Wireless Environment

In OFDM systems[1], 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.

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

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 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 (14) in the next section) 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].

[2]This project focuses on the block fading model

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 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], [30], [31], [53], [54].

Code: Data is coded before being transmitted which introduces redundancy and helps in reducing probability of error [23], [24], [27], [29], [43]-[46], [52].

Transmit precoding: Precoding might be done on the data at the transmitter to assist channel estimation at the receiver such as cyclic prefix [4], [21], [26], [27], [36], [42], zero padding (silent guard bands) [8], [9], [10] and virtual carriers (the subcarriers that are set to zero without any information) [38], [39], [40], [59].

Pilots: Pilots i.e. training symbols for the receiver, have been extensively used for channel estimation in OFDM [11]-[19].

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], [27], [47], [55].

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], [29], [48], [50].

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]-[19]). Use of training sequences decreases the system bandwidth efficiency [20] 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 [21], [22], [25], [26], and [42] while coding was also used along with cyclic prefix by [27]. Redundant and non-redundant linear precoding was exploited in [23], [24], [29], [43]-[46] for channel estimation. Virtual carriers have also been used by [38]-[40] and constant modulus modulation was used by [41]. Receiver diversity was used in [32] while [33]-[37], [40] and the references therein developed a subspace approach using the second order statistics. The finite

alphabet constraint on the data was explored by [30] and [31] and for reducing the computational complexity involved in it, adaptive techniques were explored by [28] and [29].

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], [27], [36], the finite alphabet constraint on the data as well as the frequency and time correlation of the channel [4], magnitude error in data [51], linear precoding [52], frequency correlation [12], [27], and [55], gaussian assumption on transmitted data [56], the first order statistics [57], subspace of the channel [58], receiver diversity and virtual carriers [59] for channel estimation and subsequent data detection. Semiblind adaptive approaches for channel estimation have also been exploited by [53] and [54] 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 [35] or more rigourously by the expectation maximization (EM) approach [64]-[69].

2 Project Objective and Motivation

In this project, we show how the cyclic prefix can be used to enhance the performance of an OFDM receiver. In the first part of the project, 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 completely 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. In the second part of the project, we propose iterative methods to reduce the computational complexity involved in the maximum likelihood search of input. In the last part, 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 Approach and Organization

This project 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 project, 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.[3]

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 3.1). 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 3.1.1). 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 3.1.2). 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 3.2 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 3.3).

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

Exhaustive search is computationally very expensive. We thus suggest in chapter 4, 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) [71], [72], [73], and the Genetic Algorithm (GA) [74], [75], 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 chapter 5, 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. The simulation results are presented in chapter 6 followed by the conclusions and outcomes of the project in chapter 7. To setup the stage, we introduce our notation in the following subsection.

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. $\boldsymbol{\mathcal{X}}$) is reserved for vectors in the frequency domain. The individual entries of a vector like $\boldsymbol{h}$ are denoted by $h(l)$. A hat over a variable indicates an estimate of the variable (e.g., $\hat{\boldsymbol{h}}$ is an estimate of $\boldsymbol{h}$). When any of these variables become a function of time, the time index i appears as a subscript.

Now consider a length-N vector $\boldsymbol{x}_i$. We deal with three derivatives associated with this vector. The first two are obtained by partitioning $\boldsymbol{x}_i$ into a lower (trailing) part $\underline{\boldsymbol{x}}_i$ (known as

the cyclic prefix) and an upper vector $\tilde{\boldsymbol{x}}_i$ so that

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

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

$$\overline{\boldsymbol{x}}_i = \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \boldsymbol{x}_i \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} \tag{1}$$

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

$$\boldsymbol{Q} = \begin{bmatrix} \tilde{\boldsymbol{Q}} \\ \underline{\boldsymbol{Q}} \end{bmatrix} \tag{2}$$

where the number of rows in $\tilde{\boldsymbol{Q}}$ and $\underline{\boldsymbol{Q}}$ are understood from the context and when it is not clear, the number of rows will appear as a subscript.Thus, we write

$$\boldsymbol{Q} = \begin{bmatrix} \tilde{\boldsymbol{Q}}_{N-L} \\ \underline{\boldsymbol{Q}}_{L} \end{bmatrix} \tag{3}$$

In the following chapter, we focus on blind channel and data detection in OFDM systems using cyclic prefix.

3 Blind Channel Estimation and Data Detection

We first present the essential elements of an OFDM communication system in the following section.

3.1 System Overview

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

$$\boldsymbol{x}_i = \sqrt{N}\boldsymbol{Q}\boldsymbol{\mathcal{X}}_i \tag{4}$$

where $\boldsymbol{Q}$ is the $N \times N$ IFFT matrix. When juxtaposed, these symbols result in the sequence $\{x_k\}$.[4] We assume a channel $\underline{\boldsymbol{h}}$ of maximum length $L+1$. To avoid ISI caused by passing the signal through the channel, a cyclic prefix (CP) $\underline{\boldsymbol{x}}_i$ (of length L) is appended to $\boldsymbol{x}_i$, resulting in the super-symbol $\overline{\boldsymbol{x}}_i$ as defined in (1). The concatenation of these symbols produces the underlying sequence $\{\overline{x}_k\}$. When passed through the channel $\underline{\boldsymbol{h}}$, the sequence $\{\overline{x}_k\}$ produces the output sequence $\{\overline{y}_k\}$ i.e.

$$\overline{y}_k = \underline{h}_k * \overline{x}_k + \overline{n}_k \tag{5}$$

where $\overline{n}_k$ 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 output into length $N+L$ symbol as

$$\overline{\boldsymbol{y}}_i = \begin{bmatrix} \underline{\boldsymbol{y}}_i \\ \boldsymbol{y}_i \end{bmatrix}$$

This is a natural way to partition the output because the prefix $\underline{\boldsymbol{y}}_i$ actually absorbs all ISI that takes place between the adjacent symbols $\overline{\boldsymbol{x}}_{i-1}$ and $\overline{\boldsymbol{x}}_i$. Moreover, the remaining part $\boldsymbol{y}_i$ of the

[4]The time indices in the sequence $\boldsymbol{x}_i$ 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.

symbol depends on the ith input OFDM symbol $\boldsymbol{x}_i$ only. These facts can be seen from the input/output relationship

$$\begin{bmatrix} \boldsymbol{y}_{i-1} \\ \underline{\boldsymbol{y}}_i \\ \boldsymbol{y}_i \end{bmatrix} = \begin{bmatrix} \overline{\boldsymbol{H}} & & \boldsymbol{O}_{N\times L} & \boldsymbol{O}_{N\times N} \\ \boldsymbol{O}_{L\times N} & \underline{\boldsymbol{H}}_U & \underline{\boldsymbol{H}}_L & \boldsymbol{O}_{L\times N} \\ \boldsymbol{O}_{N\times N} & \boldsymbol{O}_{N\times L} & \overline{\boldsymbol{H}} & \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_{i-1} \\ \tilde{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{n}_{i-1} \\ \underline{\boldsymbol{n}}_i \\ \boldsymbol{n}_i \end{bmatrix} \tag{6}$$

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

$$\overline{\boldsymbol{H}} = \begin{bmatrix} \underline{h}(L) & \cdots & \underline{h}(1) & \underline{h}(0) & & \\ \vdots & \ddots & \cdots & \ddots & \ddots & \\ 0 & \cdots & \underline{h}(L) & \cdots & \underline{h}(1) & \underline{h}(0) \end{bmatrix} \tag{7}$$

The matrices $\underline{\boldsymbol{H}}_{\mathsf{U}}$ and $\underline{\boldsymbol{H}}_{\mathsf{L}}$ are square matrices of size L and are defined by[5]

$$\underline{\boldsymbol{H}}_{\mathsf{U}} = \begin{bmatrix} \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(1) \\ & \underline{h}(L) & \cdots & \underline{h}(2) \\ & & \ddots & \vdots \\ & & & \underline{h}(L) \end{bmatrix} \tag{8}$$

$$\underline{\boldsymbol{H}}_{\mathsf{L}} = \begin{bmatrix} \underline{h}(0) & & & \\ \underline{h}(1) & \underline{h}(0) & & \\ \vdots & \ddots & \ddots & \\ \underline{h}(L-1) & \cdots & \underline{h}(1) & \underline{h}(0) \end{bmatrix} \tag{9}$$

[5]The matrix $\underline{\boldsymbol{H}}_{\mathsf{L}}$ ($\underline{\boldsymbol{H}}_{\mathsf{U}}$) is lower (upper) triangular; this explains the subscript L (U).

Due to the redundancy in the input, the convolution in (6) can be decomposed into two distinct constituent convolution operations or subchannels [86]. 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.

3.1.1 Circular Convolution (Subchannel)

From (6), we can write

$$\boldsymbol{y}_i = \overline{\boldsymbol{H}} \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} = \overline{\boldsymbol{H}}\, \overline{\boldsymbol{x}}_i + \boldsymbol{n}_i \tag{10}$$

This shows that $\boldsymbol{y}_i$ is created solely from $\overline{\boldsymbol{x}}_i$ through convolution and hence is ISI-free. Moreover, the existence of a cyclic prefix in $\overline{\boldsymbol{x}}_i$ allows us to rewrite (10) as

$$\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i + \boldsymbol{n}_i \tag{11}$$

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

$$\boldsymbol{H} = \begin{bmatrix} \underline{h}(0) & 0 & \cdots & 0 & \underline{h}(L) & \cdots & \underline{h}(1) \\ \underline{h}(1) & \underline{h}(0) & \cdots & 0 & 0 & \cdots & \underline{h}(2) \\ \vdots & \vdots & \ddots & \vdots & \cdots & \ddots & \vdots \\ \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(0) \end{bmatrix} \tag{12}$$

In other words, the cyclic prefix of $\overline{\boldsymbol{x}}_i$ renders the convolution in (11) cyclic, and we can write

$$\boxed{\boldsymbol{y}_i = \boldsymbol{h}_i \circledast \boldsymbol{x}_i + \boldsymbol{n}_i} \tag{13}$$

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

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

In the frequency domain, the circular convolution (13) reduces to the element-by-element operation

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i} \tag{14}$$

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

$$\boldsymbol{\mathcal{H}}_i = \boldsymbol{Q}^* \boldsymbol{h}_i, \quad \boldsymbol{\mathcal{X}}_i = \frac{1}{\sqrt{N}} \boldsymbol{Q}^* \boldsymbol{x}_i, \quad \boldsymbol{\mathcal{N}}_i = \frac{1}{\sqrt{N}} \boldsymbol{Q}^* \boldsymbol{n}_i, \text{ and } \boldsymbol{\mathcal{Y}}_i = \frac{1}{\sqrt{N}} \boldsymbol{Q}^* \boldsymbol{y}_i \tag{15}$$

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

$$\boldsymbol{\mathcal{H}}_i = \boldsymbol{Q}^*_{L+1} \underline{\boldsymbol{h}}_i \quad \text{and} \quad \underline{\boldsymbol{h}}_i = \boldsymbol{Q}_{L+1} \boldsymbol{\mathcal{H}}_i \tag{16}$$

where $\boldsymbol{Q}^*_{L+1}$ consists of the first $L+1$ columns of $\boldsymbol{Q}^*$ and $\boldsymbol{Q}_{L+1}$ consists of first $L+1$ rows of $\boldsymbol{Q}$. This allows us to rewrite (14) as

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i) \boldsymbol{Q}^*_{L+1} \underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i} \tag{17}$$

3.1.2 Linear Convolution (Subchannel)

From (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]

$$\boxed{\underline{y}_k = \underline{h}_k * \underline{x}_k + \underline{n}_i} \tag{18}$$

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

$$\boxed{\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{n}}_i} \tag{19}$$

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

$$\underline{\boldsymbol{X}}_i = \underline{\boldsymbol{X}}_{Ui-1} + \underline{\boldsymbol{X}}_{Li} \tag{20}$$

and where (compare with (8))

$$\underline{\boldsymbol{X}}_{\mathsf{U}i-1} = \begin{bmatrix} 0 & \underline{x}_{i-1}(L-1) & \cdots & \underline{x}_{i-1}(0) \\ 0 & 0 & \cdots & \underline{x}_{i-1}(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \underline{x}_{i-1}(L-1) \end{bmatrix}, \tag{21}$$

$$\text{and } \underline{\boldsymbol{X}}_{\mathsf{L}i} = \begin{bmatrix} \underline{x}_i(0) & 0 & \cdots & 0 \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \underline{x}_i(L-1) & \cdots & \underline{x}_i(0) & 0 \end{bmatrix} \tag{22}$$

This fact together with the FFT relationship (16) yields the time-frequency input/output equation

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \boldsymbol{\mathcal{H}}_i + \underline{\boldsymbol{n}}_i \tag{23}$$

3.2 Maximum-Likelihood Estimation

Consider the frequency domain description of the circular subchannel (14)

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

To obtain the maximum-likelihood (ML) estimate of $\boldsymbol{\mathcal{H}}_i$, we assume that the sequence $\boldsymbol{\mathcal{X}}_i$ is deterministic and perform an element-by-element division of (14) by $\boldsymbol{\mathcal{X}}_i$ to get

$$\boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i + \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{N}}_i \tag{24}$$

where

$$\boldsymbol{D}_{\mathcal{X}} = \text{diag}(\boldsymbol{\mathcal{X}}_i) \tag{25}$$

Equivalently, we can write (24) as

$$\boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i + \boldsymbol{\mathcal{N}}_i^{'} \tag{26}$$

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

$$\boldsymbol{R}_{n'} = \sigma_n^2 \boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{D}_{\mathcal{X}}^{-*} = \sigma_n^2 |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tag{27}$$

The maximum-likelihood estimate of $\boldsymbol{\mathcal{H}}$ can now be obtained by solving the system of equations (24) in the least-squares (LS) sense subject to the constraint

$$\tilde{\boldsymbol{Q}}_{N-L-1}\boldsymbol{\mathcal{H}}_i \triangleq \tilde{\boldsymbol{Q}}\boldsymbol{\mathcal{H}}_i = 0 \tag{28}$$

We can show that the ML estimate is given by [82]

$$\begin{aligned} \hat{\boldsymbol{\mathcal{H}}}_i^{ML} &= \left[\boldsymbol{I} - \boldsymbol{R}_{n'}\tilde{\boldsymbol{Q}}^*\left(\tilde{\boldsymbol{Q}}\boldsymbol{R}_{n'}\tilde{\boldsymbol{Q}}^*\right)^{-1}\tilde{\boldsymbol{Q}}\right]\boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{\mathcal{Y}}_i \\ &= \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\left(\tilde{\boldsymbol{Q}}|\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\right)^{-1}\tilde{\boldsymbol{Q}}\right]\boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{\mathcal{Y}}_i \end{aligned} \tag{29}$$

The ML estimate (29) was obtained solely from the circular convolution subchannel. Upon replacing $\boldsymbol{\mathcal{H}}_i$ that appears in the time-frequency input/output equation (23) (corresponding to the linear subchannel)

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

with its ML estimate (29), we obtain

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i\boldsymbol{Q}_{L+1}\left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\left(\tilde{\boldsymbol{Q}}|\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\right)^{-1}\tilde{\boldsymbol{Q}}\right]\boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{\mathcal{Y}}_i + \underline{\boldsymbol{n}}_i \tag{30}$$

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

$$\boxed{\hat{\boldsymbol{\mathcal{X}}}_i^{ML} = \arg\min_{\mathcal{X}_i}\left\|\underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_i\boldsymbol{Q}_{L+1}\left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\left(\tilde{\boldsymbol{Q}}|\boldsymbol{D}_{\mathcal{X}}|^{-2}\tilde{\boldsymbol{Q}}^*\right)^{-1}\tilde{\boldsymbol{Q}}\right]\boldsymbol{D}_{\mathcal{X}}^{-1}\boldsymbol{\mathcal{Y}}_i\right\|^2} \tag{31}$$

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 $\boldsymbol{\mathcal{X}}_i$.

To gain more insight into this problem, we now treat the case of constant modulus data which leads to more explicit results.

3.3 Maximum-Likelihood Estimation in the Constant Modulus Case

In the constant modulus case, we have

$$|\boldsymbol{D}_{\mathcal{X}}|^{-2} = \frac{1}{\mathcal{E}_X}\boldsymbol{I} \tag{32}$$

As a consequence, we can also write

$$\boldsymbol{D}_{\mathcal{X}}^{-1} = \frac{1}{\mathcal{E}_X}\boldsymbol{D}_{\mathcal{X}}^{*} \tag{33}$$

Thus, the ML estimate of $\boldsymbol{\mathcal{H}}_i$ (29) simplifies to

$$\hat{\boldsymbol{\mathcal{H}}}_i^{\mathsf{ML}} = \frac{1}{\mathcal{E}_X}\left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^{*}\left(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}^{*}\right)^{-1}\tilde{\boldsymbol{Q}}\right]\boldsymbol{D}_{\mathcal{X}^{*}}\boldsymbol{\mathcal{Y}}_i \tag{34}$$

$$= \frac{1}{\mathcal{E}_X}\left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^{*}\tilde{\boldsymbol{Q}}\right]\boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i^{*} \tag{35}$$

where in (35), we used the fact that $\tilde{\boldsymbol{Q}}$ is a left-inverse of $\tilde{\boldsymbol{Q}}^{*}$ - a consequence of the unitary nature of $\boldsymbol{Q}$

$$\boldsymbol{I} = \boldsymbol{Q}\boldsymbol{Q}^{*} = \begin{bmatrix} \boldsymbol{Q}_{L+1} \\ \tilde{\boldsymbol{Q}}_{N-L-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{Q}_{L+1}^{*} & \tilde{\boldsymbol{Q}}_{N-L-1}^{*} \end{bmatrix} \tag{36}$$

Just as we did in the general case, we now replace the effect of $\boldsymbol{\mathcal{H}}_i$ in the linear convolution channel, as expressed in (23), by its ML estimate to get

$$\underline{\boldsymbol{y}}_i = \frac{1}{\mathcal{E}_X}\underline{\boldsymbol{X}}_i\underline{\boldsymbol{Q}}_{L+1}\left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^{*}\tilde{\boldsymbol{Q}}\right]\boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i^{*} + \underline{\boldsymbol{n}}_i \tag{37}$$

$$= \frac{1}{\mathcal{E}_X}\underline{\boldsymbol{X}}_i\underline{\boldsymbol{Q}}_{L+1}\boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i^{*} + \underline{\boldsymbol{n}}_i \tag{38}$$

where in going to (38), we used the fact that

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

which can be deduced from (36). The ML estimate of $\boldsymbol{\mathcal{X}}_i$ is now obtained by performing the minimization

$$\boxed{\hat{\boldsymbol{\mathcal{X}}}_i^{\mathsf{ML}} = \arg\min_{\mathcal{X}_i}\left\|\underline{\boldsymbol{y}}_i - \tfrac{1}{\mathcal{E}_X}\underline{\boldsymbol{X}}_i\underline{\boldsymbol{Q}}_{L+1}\boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i^{*}\right\|^2} \tag{39}$$

Notice that the only unknowns in this minimization are $\underline{\boldsymbol{X}}_i$ and $\boldsymbol{\mathcal{X}}_i$, 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.

4 Approximate Methods to Reduce Computational Complexity

The search for the optimal $\boldsymbol{\mathcal{X}}_i$ in (39) is computationally very complex. In the following, we describe six approaches to reduce this complexity:

4.1 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 $\underline{\boldsymbol{X}}_i$ involved in equation (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 (20), i.e.

$$\underline{\boldsymbol{X}}_i = \underline{\boldsymbol{X}}_{\mathsf{U}i-1} + \underline{\boldsymbol{X}}_{\mathsf{L}i}$$

Thus equation (39) can be rewritten as

$$\begin{aligned}
\hat{\boldsymbol{\mathcal{X}}}_i^{ML} &= \arg\min_{\mathcal{X}_i} \left\| \underline{\boldsymbol{y}}_i - \frac{1}{\mathcal{E}_X} \left(\underline{\boldsymbol{X}}_{\mathsf{U}i-1} + \underline{\boldsymbol{X}}_{\mathsf{L}i} \right) \boldsymbol{Q}_{L+1} \boldsymbol{D}_{\mathcal{Y}} \boldsymbol{\mathcal{X}}_i^* \right\|^2 \\
&= \arg\min_{\mathcal{X}_i} \left\| \underline{\boldsymbol{y}}_i - \left(\underline{\boldsymbol{X}}_{\mathsf{U}i-1} + \underline{\boldsymbol{X}}_{\mathsf{L}i} \right) \boldsymbol{A} \boldsymbol{\mathcal{X}}_i^* \right\|^2 \\
&= \arg\min_{\mathcal{X}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B} \boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C} \boldsymbol{\mathcal{X}}_i^* \right\|^2
\end{aligned} \tag{40}$$

where

$$\boldsymbol{A} = \frac{1}{\mathcal{E}_X} \boldsymbol{Q}_{L+1} \boldsymbol{D}_{\mathcal{Y}}, \qquad \boldsymbol{B} = \underline{\boldsymbol{X}}_{\mathsf{U}i-1} \boldsymbol{A}$$

and hence are completely known and where

$$\boldsymbol{C} = \underline{\boldsymbol{X}}_{\mathsf{L}i} \boldsymbol{A} \tag{41}$$

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

$$c(j) = \|\boldsymbol{\mathcal{X}}_i\|^2_{\boldsymbol{W}_j} \triangleq \boldsymbol{\mathcal{X}}_i^* \boldsymbol{W}_j \boldsymbol{\mathcal{X}}_i \tag{42}$$

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

$$\hat{\boldsymbol{\mathcal{X}}}_i^{ML} = \arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c} \right\|^2 \tag{43}$$

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

$$\begin{aligned} c(j) &= \|\boldsymbol{\mathcal{X}}_i\|^2_{\boldsymbol{W}_j} \qquad 1 \le j \le L \\ &\simeq \|\boldsymbol{\mathcal{X}}_i\|^2_{\mathrm{diag}(\boldsymbol{W}_j)} \\ &= \mathcal{E}_X \,\mathrm{trace}\,(\boldsymbol{W}_j) \\ &= z(j) \end{aligned} \tag{44}$$

where the third line follows from the fact that the elements of $\boldsymbol{\mathcal{X}}_i$ have constant modulus. The input dependent vector $\boldsymbol{c}$ is thus replaced by the constant vector $\boldsymbol{z}$, and the objective function becomes linear in $\boldsymbol{\mathcal{X}}_i$

$$\arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \left(\underline{\boldsymbol{y}}_i - \boldsymbol{z} \right) - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* \right\|^2 \tag{45}$$

One way to solve equation (45) is by using least squares

$$\hat{\boldsymbol{\mathcal{X}}}_i^* = (\boldsymbol{B}^*\boldsymbol{B} + \delta \boldsymbol{I})^{-1} \boldsymbol{B}^* \left(\underline{\boldsymbol{y}}_i - \boldsymbol{z} \right) \tag{46}$$

where δ is a small constant.[6]

We could refine the estimate obtained in (46) further by using the estimate $\hat{\boldsymbol{\mathcal{X}}}_i^*$ to obtain the vector $\boldsymbol{c}$ in (43) (as opposed to the vector $\boldsymbol{z}$ which is obtained by approximating $\boldsymbol{W}_j$ with its diagonal). We now solve the alternative least squares problem

$$\arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \left(\underline{\boldsymbol{y}}_i - \boldsymbol{c} \right) - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* \right\|^2 \tag{47}$$

This procedure of refining the estimate $\boldsymbol{c}$ and solving the least squares (47) could be repeated for a desired number of iterations.

[6] The optimum choice for δ is $\mathcal{E}_X$ as $\mathrm{Cov}\,[\boldsymbol{\mathcal{X}}_i] = \mathcal{E}_X \boldsymbol{I}$.

4.2 ML Estimation at High SNR

In this section, we try to find a closed form solution of the non-linear problem (equation (39)) at high SNR by assuming noise to be zero. At high SNR, equation (43) reduces to

$$\underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c} = 0 \tag{48}$$

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

$$\underline{y}_k - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{b}_k - \|\boldsymbol{\mathcal{X}}_i\|^2_{\boldsymbol{W}_k} = 0 \tag{49}$$

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

$$\underline{y}_k{}^* - \boldsymbol{b}_k^H \boldsymbol{\mathcal{X}}_i - \|\boldsymbol{\mathcal{X}}_i\|^2_{\boldsymbol{W}_k^H} = 0 \tag{50}$$

Adding equations (49) and (50)

$$R_k = 2Re(\underline{y}_k) - \boldsymbol{b}_k^H \boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{b}_k - \|\boldsymbol{\mathcal{X}}_i\|^2_{(\boldsymbol{W}_k + \boldsymbol{W}_k^H)} \tag{51}$$

Let α and β be such that

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

Then

$$R_k = \alpha - \boldsymbol{b}_k^H \boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{b}_k + \|\boldsymbol{\mathcal{X}}_i\|^2_{(\frac{\beta}{\mathcal{E}_X N}\boldsymbol{I}_N - \boldsymbol{W}_k - \boldsymbol{W}_k^H)} \tag{52}$$

where we used the fact that

$$\beta = \|\boldsymbol{\mathcal{X}}_i\|^2_{\frac{\beta}{\mathcal{E}_X N}\boldsymbol{I}_N}$$

By completing squares [79], we get

$$R_k = (\boldsymbol{\mathcal{X}}_i - \boldsymbol{E}^{-1}\boldsymbol{b}_k)^H \boldsymbol{E}(\boldsymbol{\mathcal{X}}_i - \boldsymbol{E}^{-1}\boldsymbol{b}_k) + \boldsymbol{b}_k^H \boldsymbol{E}^{-1}\boldsymbol{b}_k + \alpha \tag{53}$$

where

$$\boldsymbol{E} = (\frac{\beta}{\mathcal{E}_X N}\boldsymbol{I}_N - \boldsymbol{W}_k - \boldsymbol{W}_k^H)$$

To make equation (53) a perfect square, α should satisfy the following relation

$$\begin{aligned} \boldsymbol{b}_k^H \boldsymbol{E}^{-1}\boldsymbol{b}_k - \alpha &= 0 \\ \boldsymbol{b}_k^H \boldsymbol{E}^{-1}\boldsymbol{b}_k &= \alpha \\ \boldsymbol{b}_k^H (\frac{\beta}{\mathcal{E}_X N}\boldsymbol{I}_N - \boldsymbol{W}_k - \boldsymbol{W}_k^H)^{-1}\boldsymbol{b}_k &= 2Re(\underline{y}_k) - \beta \end{aligned} \tag{54}$$

The above equation can be solved for β as it is the only variable unknown. Thus equation (53) can now be given as

$$\begin{aligned} R_k &= (\boldsymbol{\mathcal{X}}_i - \boldsymbol{E}^{-1}\boldsymbol{b}_k)^H \boldsymbol{E}(\boldsymbol{\mathcal{X}}_i - \boldsymbol{E}^{-1}\boldsymbol{b}_k) \\ R_k &= \left\|(\boldsymbol{\mathcal{X}}_i - \boldsymbol{E}^{-1}\boldsymbol{b}_k)\right\|^2 \end{aligned} \tag{55}$$

Now, we have L such quadratic terms and their sum is given by [79]

$$\sum_{k=1}^{L} R_k = (\boldsymbol{\mathcal{X}}_i - \boldsymbol{m}_c)\Sigma_c^{-1}(\boldsymbol{\mathcal{X}}_i - \boldsymbol{m}_c) \tag{56}$$

where

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

Thus $\boldsymbol{m}_c$ would be our required solution. The problem in this method lies in the fact that in finding β from equation (54), we again end up with exhaustive search as there will be $N+1$ solutions.

4.3 Using Search Algorithms

We can use the search algorithms like Particle Swarm Optimization (PSO) [71], [72], [73], and the Genetic Algorithm (GA) [74], [75], to directly solve the nonlinear problem (equation (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.

4.4 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 (40) reproduced here for convenience

$$\begin{aligned} \hat{\boldsymbol{\mathcal{X}}}_i^{ML} &= \arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^* \right\|^2 \\ &= \arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c} \right\|^2 \\ &= \arg\min_{\boldsymbol{\mathcal{X}}_i} \left\| \left(\underline{\boldsymbol{y}}_i - \boldsymbol{z} \right) - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \overline{\boldsymbol{c}} \right\|^2 \end{aligned} \tag{57}$$

where $\boldsymbol{z}$ is defined in (44) and where

$$\overline{c}(j) = \|\boldsymbol{\mathcal{X}}_i\|^2_{\overline{\boldsymbol{W}}_j} \tag{58}$$

where $\overline{\boldsymbol{W}}_j$ is $\boldsymbol{W}_j$ 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 (57). Specifically, define

$$\boldsymbol{b} = \mathrm{vec}(\boldsymbol{B}) \;\; and \;\; \overline{\boldsymbol{w}}_i = \mathrm{vec}(\overline{\boldsymbol{W}}_i) \qquad 1 \le i \le L \tag{59}$$

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 $\boldsymbol{B}$ and $\overline{\boldsymbol{W}}_i$ that operate on this symbol (or symbols). Now, find the next largest element in (59) and determine the data symbols (mostly two) that interact with this coefficient. With this procedure, (57) is approximated with an optimization that looks for the most significant $L+1$ data symbols by exhaustive search.

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

$$\boldsymbol{\mathcal{Y}}_{L+1} = \mathrm{diag}(\boldsymbol{\mathcal{X}}_{L+1})\boldsymbol{Q}^*_{L+1}\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i \tag{60}$$

where $\boldsymbol{\mathcal{X}}_{L+1}$ are the $L+1$ elements detected using the above method and $\boldsymbol{\mathcal{Y}}_{L+1}$ are the $L+1$ elements of $\boldsymbol{\mathcal{Y}}$ corresponding to $\boldsymbol{\mathcal{X}}_{L+1}$.

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

4.5 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 $\boldsymbol{R}$ be the channel correlation matrix and let

$$\boldsymbol{\mathcal{Y}}_p = \mathrm{diag}(\boldsymbol{\mathcal{X}}_p)\boldsymbol{Q}^*_{L+1}\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_p \tag{61}$$

be a subsystem of (17) corresponding to the pilot locations. Then (60) and (61) can be concatenated into a single system of equations and the channel $\underline{\boldsymbol{h}}_i$ can be obtained by solving the regularized least squares problem

$$\hat{\underline{\boldsymbol{h}}}_i = \arg\min_{\underline{\boldsymbol{h}}_i} \left\| \begin{bmatrix} \boldsymbol{\mathcal{Y}}_{L+1} \\ \boldsymbol{\mathcal{Y}}_p \end{bmatrix} - \begin{bmatrix} \mathrm{diag}(\boldsymbol{\mathcal{X}}_{L+1}) \\ \mathrm{diag}(\boldsymbol{\mathcal{X}}_p) \end{bmatrix} \boldsymbol{Q}^*_{L+1} \right\|^2_{\sigma_n^{-2}\boldsymbol{I}} + \|\underline{\boldsymbol{h}}_i\|^2_{\boldsymbol{R}^{-1}} \tag{62}$$

4.6 Using Newton's Method

Let

$$\boldsymbol{Z} = \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^* \right\|^2$$
$$\text{subject to} \quad |\boldsymbol{\mathcal{X}}_i|^2 = 1 \tag{63}$$

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 [82] given by

$$\boldsymbol{\mathcal{X}}_k = \boldsymbol{\mathcal{X}}_{k-1} - \mu \left[\nabla^2 \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})\right]^{-1} \left[\nabla \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})\right]^*, \quad k \geq 0 \tag{64}$$

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.

4.6.1 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

$$\begin{aligned} \boldsymbol{Z} &= \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^* \right\|^2 \\ &= \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c} \right\|^2 \\ &= \|\boldsymbol{a}\|^2 \\ &= \boldsymbol{a}^H \boldsymbol{a} \end{aligned} \tag{65}$$

To find the gradient of $\boldsymbol{Z}$, we have to differentiate it with $\boldsymbol{\mathcal{X}}_i^*$ [76] (we consider $\boldsymbol{\mathcal{X}}_i^*$ a row

vector while differentiation)

$$\begin{aligned}\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \frac{\partial \left[\boldsymbol{a}^H \boldsymbol{a}\right]}{\partial \boldsymbol{\mathcal{X}}_i^*} \\ &= \left[\frac{\partial (\boldsymbol{a}^H \boldsymbol{a})}{\partial \boldsymbol{a}}\right]\left[\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right] + \left[\frac{\partial (\boldsymbol{a}^H \boldsymbol{a})}{\partial \boldsymbol{a}^*}\right]\left[\frac{\partial \boldsymbol{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*}\right] \\ &= \boldsymbol{a}^H \left[\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right] + \boldsymbol{a}^T \left[\frac{\partial \boldsymbol{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*}\right] \end{aligned} \tag{66}$$

where the second line follows from the chain rule for complex matrices [76], given by

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

Now, as can be seen from equation (66), we need to find differential of $\boldsymbol{a}$ and $\boldsymbol{a}^*$ w.r.t $\boldsymbol{\mathcal{X}}_i^*$

$$\begin{aligned}\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \frac{\partial \left[\underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c}\right]}{\partial \boldsymbol{\mathcal{X}}_i^*} \\ &= \frac{\partial \underline{\boldsymbol{y}}_i}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial (\boldsymbol{B}\boldsymbol{\mathcal{X}}_i^*)}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i^*}\begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1 \boldsymbol{\mathcal{X}}_i^* \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2 \boldsymbol{\mathcal{X}}_i^* \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L \boldsymbol{\mathcal{X}}_i^* \end{bmatrix} \\ &= -\boldsymbol{B} - \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1 \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L \end{bmatrix} \end{aligned} \tag{67}$$

and

$$\begin{aligned}
\frac{\partial \boldsymbol{a}^*}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \frac{\partial \left[\underline{\boldsymbol{y}_i}^* - \boldsymbol{B}^* \boldsymbol{\mathcal{X}}_i - \boldsymbol{c}^*\right]}{\partial \boldsymbol{\mathcal{X}}_i^*} \\
&= \frac{\partial \underline{\boldsymbol{y}_i^*}}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial (\boldsymbol{B}^* \boldsymbol{\mathcal{X}}_i)}{\partial \boldsymbol{\mathcal{X}}_i^*} - \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i^*} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^H \boldsymbol{W}_1^* \boldsymbol{\mathcal{X}}_i \\ \boldsymbol{\mathcal{X}}_i^H \boldsymbol{W}_2^* \boldsymbol{\mathcal{X}}_i \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^H \boldsymbol{W}_L^* \boldsymbol{\mathcal{X}}_i \end{bmatrix} \\
&= -\begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1^H \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2^H \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L^H \end{bmatrix}
\end{aligned} \tag{68}$$

Substituting these values in equation (66), we get

$$\begin{aligned}
\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} &= \boldsymbol{a}^H \left(-\boldsymbol{B} - \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1 \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L \end{bmatrix} \right) + \boldsymbol{a}^T \left(- \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1^H \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2^H \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L^H \end{bmatrix} \right) \\
&= -\boldsymbol{a}^H \boldsymbol{B} - \boldsymbol{a}^H \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1 \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L \end{bmatrix} - \boldsymbol{a}^T \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_1^H \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_2^H \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{W}_L^H \end{bmatrix}
\end{aligned} \tag{69}$$

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

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

$$\Psi = \left\| \mathcal{E}_X - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{E}_i \boldsymbol{\mathcal{X}}_i \right\|^2 \tag{70}$$

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

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

Equation (70) can also be written as

$$\begin{aligned} \Psi &= \left\| \mathcal{E}_X - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{E}_i \boldsymbol{\mathcal{X}}_i \right\|^2 \\ &= \|\boldsymbol{b}\|^2 \\ &= \boldsymbol{b}^H \boldsymbol{b} \end{aligned} \tag{71}$$

which is similar to equation (65). So, the gradient of constant modulus constraint is calculated similar to the cost function and is given as

$$\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*} = -\boldsymbol{b}^H \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_1^T \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_2^T \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_N^T \end{bmatrix} - \boldsymbol{b}^T \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_1^* \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_2^* \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^T \boldsymbol{E}_N^* \end{bmatrix} \tag{72}$$

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

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

$$\nabla(\boldsymbol{Z}) = \frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} + \frac{1}{\sigma_n^2} \frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*} \tag{73}$$

4.6.2 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

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

which can be expanded to get the following

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T = &- \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\boldsymbol{B}^T\boldsymbol{a}^*\right)\\ &- \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\begin{bmatrix}\boldsymbol{W}_1^T\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{W}_L^T\boldsymbol{\mathcal{X}}_i\end{bmatrix}\boldsymbol{a}^*\right)\\ &- \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\begin{bmatrix}\boldsymbol{W}_1^*\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{W}_L^*\boldsymbol{\mathcal{X}}_i\end{bmatrix}\boldsymbol{a}\right)\end{aligned} \tag{74}$$

Now, we apply chain rule of complex matrices (explained in the previous section) to each term of equation (74). The final result obtained is given by

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T = &- \boldsymbol{B}^T\left(-\boldsymbol{B}^* - \begin{bmatrix}\boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_1^*\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_2^*\\ \vdots\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_L^*\end{bmatrix}\right)\\ &- \begin{bmatrix}\boldsymbol{W}_1^T\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{W}_L^T\boldsymbol{\mathcal{X}}_i\end{bmatrix}\left(-\boldsymbol{B}^* - \begin{bmatrix}\boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_1^*\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_2^*\\ \vdots\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_L^*\end{bmatrix}\right)\\ &- \begin{bmatrix}\boldsymbol{W}_1^*\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{W}_L^*\boldsymbol{\mathcal{X}}_i\end{bmatrix}\begin{bmatrix}\boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_1^T\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_2^T\\ \vdots\\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{W}_L^T\end{bmatrix}\end{aligned} \tag{75}$$

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

$$\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T = \begin{bmatrix} \boldsymbol{E}_1\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{E}_N\boldsymbol{\mathcal{X}}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_1^H \\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_2^H \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_N^H \end{bmatrix} + \begin{bmatrix} \boldsymbol{E}_1^H\boldsymbol{\mathcal{X}}_i & \cdots & \boldsymbol{E}_N^H\boldsymbol{\mathcal{X}}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_1 \\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_N \end{bmatrix} \tag{76}$$

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

$$\nabla^2(\boldsymbol{Z}) = \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T + \frac{1}{\sigma_n^2}\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T \tag{77}$$

5 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 ((14) and (19)), reproduced here for convenience

$$\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i = \mathrm{diag}(\boldsymbol{\mathcal{H}}_i)\boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i \tag{78}$$

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{n}}_i \tag{79}$$

Now, decompose $\underline{\boldsymbol{X}}_i$ as $\left(\underline{\boldsymbol{X}}_{Ui-1} + \underline{\boldsymbol{X}}_{Li}\right)$ and move the known part $\underline{\boldsymbol{X}}_{Ui-1}\underline{\boldsymbol{h}}_i$ to the left hand side to get

$$\underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_{Ui-1}\underline{\boldsymbol{h}}_i = \underline{\boldsymbol{X}}_{Li}\underline{\boldsymbol{h}}_i + \underline{\boldsymbol{n}}_i \tag{80}$$

and exchange the roles of $\underline{\boldsymbol{X}}_{Li}\underline{\boldsymbol{h}}_i$ as

$$\underline{\boldsymbol{X}}_{Li}\underline{\boldsymbol{h}}_i = \underline{\boldsymbol{H}}_{\mathsf{L}}\underline{\boldsymbol{x}}_i \tag{81}$$

$$= \underline{\boldsymbol{H}}_{\mathsf{L}}\boldsymbol{Q}_{N-L+1}\boldsymbol{\mathcal{X}}_i \tag{82}$$

where the second line follows from the fact that

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

and that $\underline{\boldsymbol{x}}_i$ consists of the last $L+1$ elements of $\boldsymbol{x}_i$. Thus (80) can be rewritten as

$$\underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_{Ui-1}\underline{\boldsymbol{h}}_i = \underline{\boldsymbol{H}}_{\mathsf{L}}\boldsymbol{Q}_{N-L+1}\boldsymbol{\mathcal{X}}_i + \underline{\boldsymbol{n}}_i \tag{83}$$

Combining (78) and (83) yields an $N+L$ system of equations in the unknown OFDM symbol $\boldsymbol{\mathcal{X}}_i$

$$\begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_{Ui-1}\underline{\boldsymbol{h}}_i \end{bmatrix} = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\mathcal{H}}_i) \\ \underline{\boldsymbol{H}}_{\mathsf{L}}\boldsymbol{Q}_{N-L+1} \end{bmatrix} \boldsymbol{\mathcal{X}}_i + \begin{bmatrix} \boldsymbol{\mathcal{N}}_i \\ \underline{\boldsymbol{n}}_i \end{bmatrix} \tag{84}$$

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

6 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.

6.1 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 (39). In all the following cases, we assume that we know the previous symbol unless mentioned otherwise.

6.1.1 Bench Marking

To bench mark our blind algorithm, we compared it with the blind algorithm proposed by Muquet et al. in [36], a method in which channel is estimated using $L+1$ pilots, and perfectly known channel case. As opposed to our algorithm in which estimation is done in a single block, the subspace method in [36] requires large memory to collect sufficiently enough data blocks to render the covariance matrix full rank. It also suffers from sign ambiguity and cannot be implemented without transmitting a pilot in the first symbol of the block. Moreover, it fails in the presence of nulls on subcarriers while our algorithm is robust to channel nulls.

In Figure 1, the performance of both algorithms is compared for BPSK modulated data over a Rayleigh channel. To make the covariance matrix full rank, 50 blocks of data were considered to implement the subspace algorithm while only a single block of data was used for channel estimation in our algorithm. As expected, the best performance is achieved by the perfectly known channel, followed by that obtained by training based estimated channel. It can

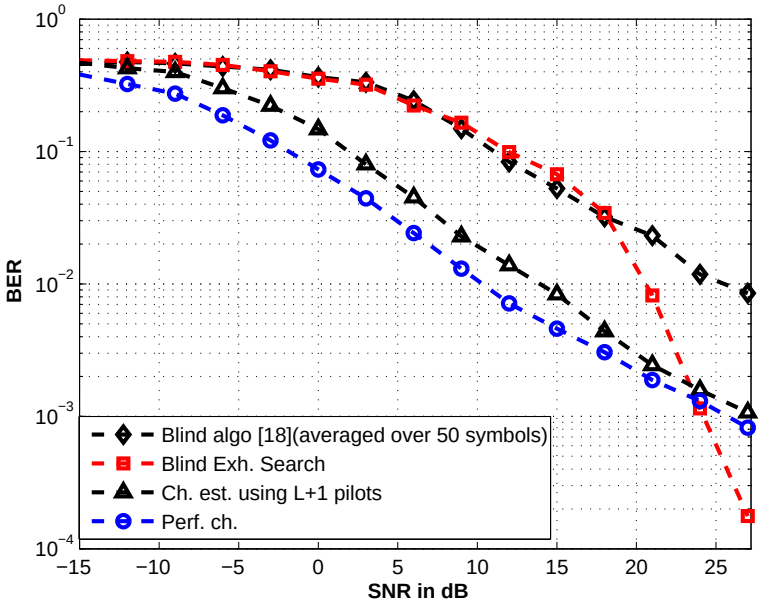

Figure 1: BER vs SNR comparison for BPSK-OFDM over a Rayleigh channel

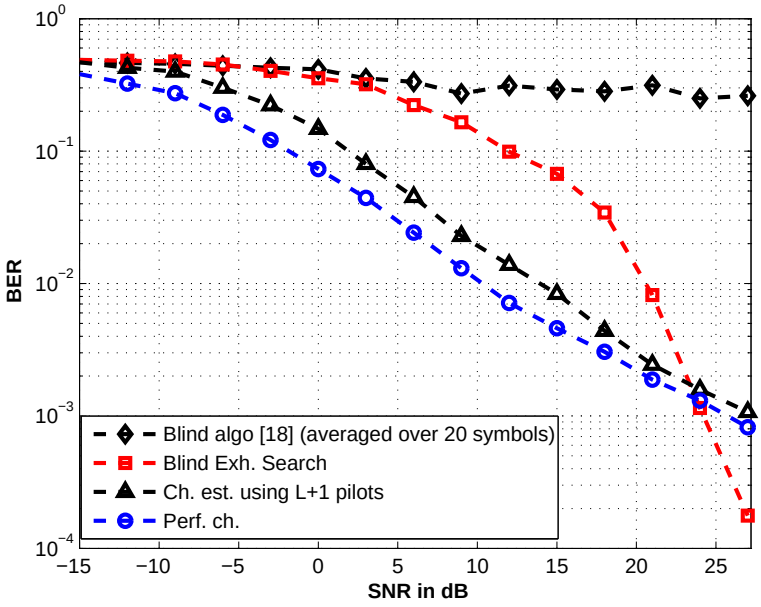

Figure 2: BER vs SNR comparison for BPSK-OFDM over a Rayleigh channel

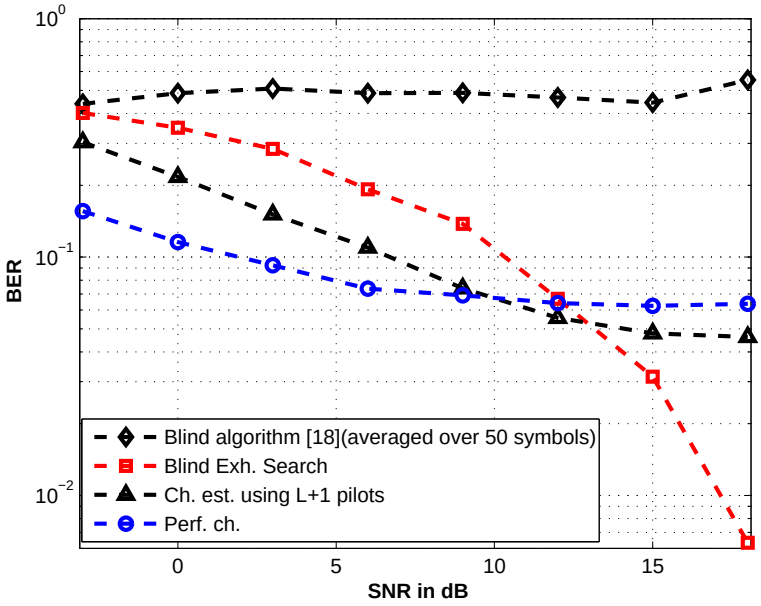

Figure 3: BER vs SNR comparison for BPSK-OFDM over a Rayleigh channel with persistent nulls

be observed that both the blind algorithms perform close to each other at low SNR but our algorithm outperforms the subspace algorithm at high SNR. It should also be noted that in the high SNR region, the BER curve of our blind algorithm 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 our case when only the circular subchannel is used in the pilot based and known channel cases. An alternative way to see this is to note that a Rayleigh fading channel will occasionally hit a (near) zero on the FFT grid resulting in a loss of the corresponding BPSK symbol. Our blind algorithm does not suffer from this and thus demonstrates improved performance in higher SNR.

In Figure 2, it can be seen that the performance of subspace algorithm becomes worse when only 20 data blocks are used as the covariance matrix is not full rank. Figure 3 compares the performance of the algorithms when the channel has persistent nulls. The number of blocks used to implement the subspace algorithm is fixed at 50. Our algorithm is robust to channel

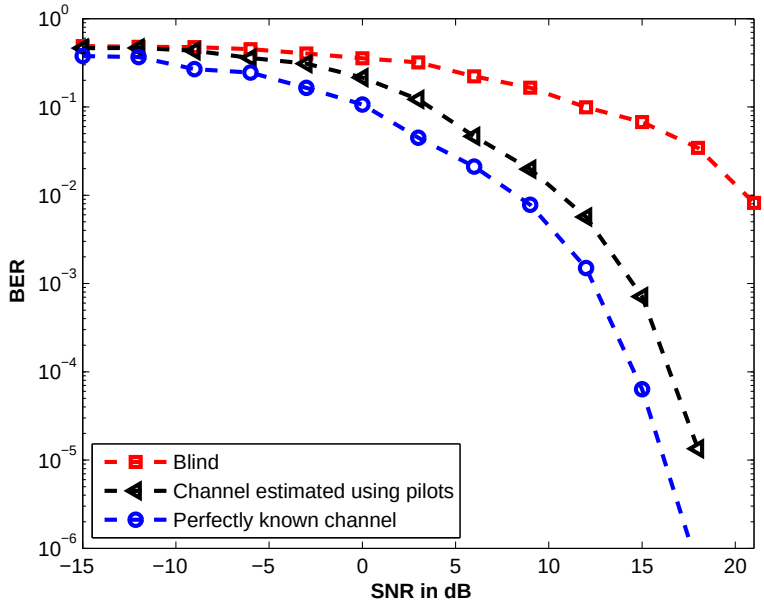

Figure 4: BER vs SNR for BPSK-OFDM over constant channel

nulls and thus in this case, it easily outperforms the subspace one. It should also be noted that at high SNR, the BER for perfectly known channel and that of the estimated channel reach an error floor but our algorithm does not suffer from it and outperforms them too.

6.1.2 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 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.

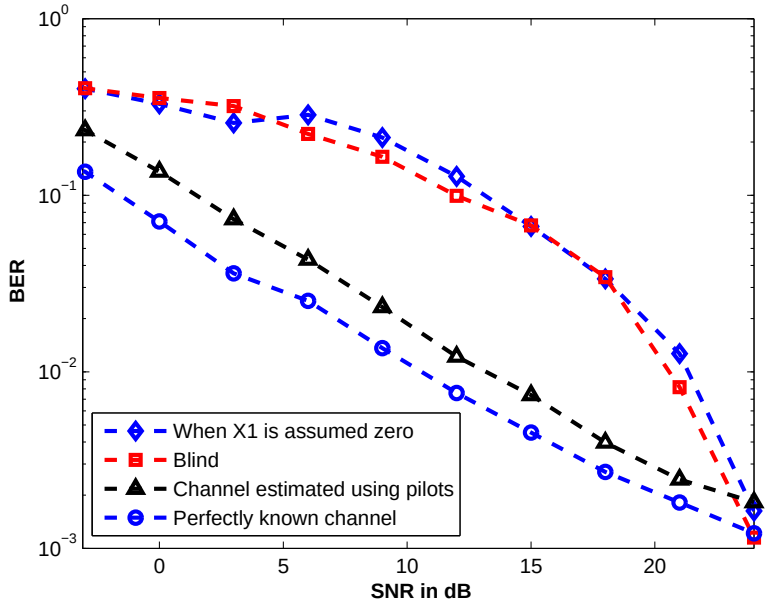

Figure 5: BER vs SNR for BPSK-OFDM with previous symbol assumed to be zero

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

In Figure 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 5 clearly shows that the blind algorithm performs quite well in this scenario also.

6.1.4 BER vs SNR comparison for 4-QAM modulated data

The same conclusion can be made for the 4-QAM input (see Figure 6). Here it can also be seen that the blind based estimation outperforms the pilot based and perfectly known channel cases at high SNR.

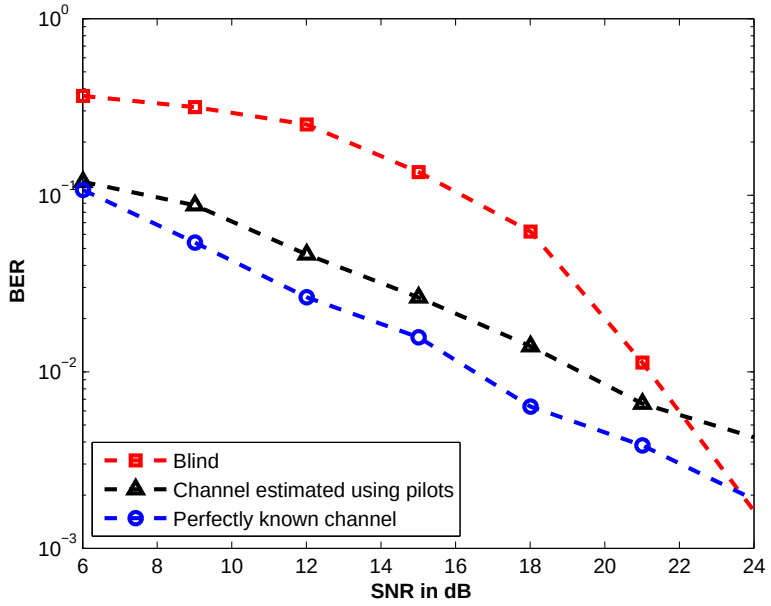

Figure 6: BER vs SNR for 4QAM-OFDM over a Rayleigh channel

6.1.5 Comparison of linearization approach and search algorithms

The low complexity algorithms proposed in Subsections 4.1 and 4.3 (linearization approach, PSO and GA) have been compared in Figure 7 for BPSK modulated data. 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 $|\boldsymbol{\mathcal{X}}_i| = 1$. 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 7 that PSO performs well at low SNR while GA performs quite close to the blind exhaustive search at low as well as high SNR.

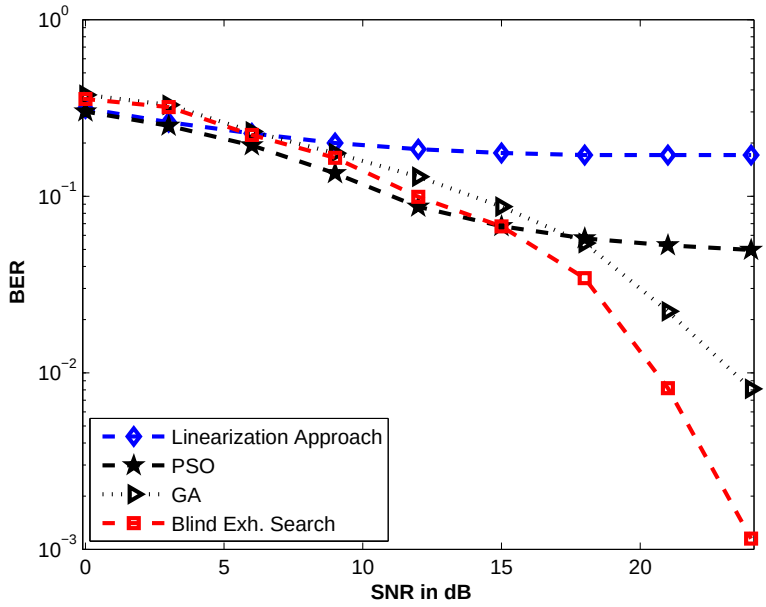

Figure 7: Comparison of low complexity algorithms for BPSK-OFDM over a Rayleigh channel

6.1.6 Sensitivity of Reduced exhaustive search algorithm to number of iterations

The sensitivity of reduced $(L+1)$ exhaustive search algorithms described in Subsections 4.4 and 4.5 to number of iterations is shown in Figure 8. Two pilots were 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.

6.1.7 BER vs SNR Comparison of Reduced exhaustive search algorithms

In Figure 9, three variants of the reduced $(L+1)$ exhaustive search algorithm (discussed in Subsections 4.4 and 4.5) have been compared with the N exhaustive search algorithm of equation (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 9 shows that $L+1$ exhaustive search algorithm performs quite close to the N exhaustive search blind

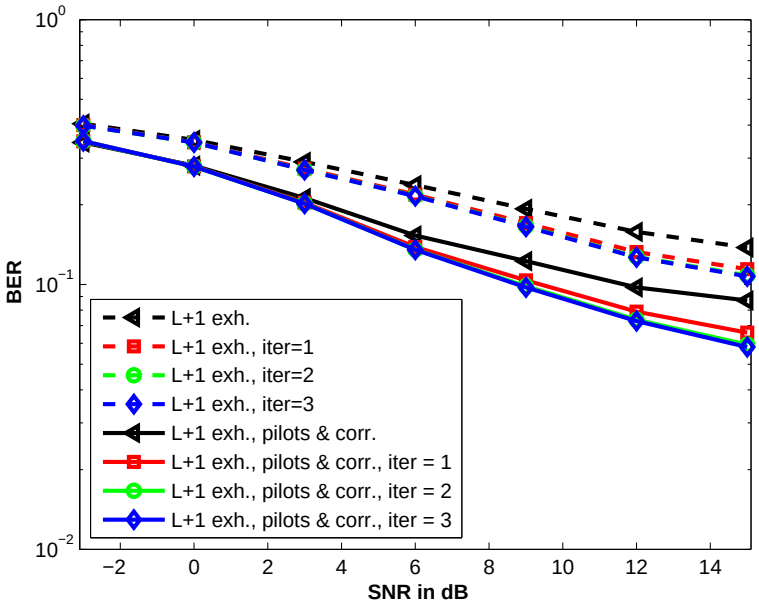

Figure 8: Sensitivity of Reduced search algorithms to number of iterations for BPSK-OFDM over a Rayleigh channel

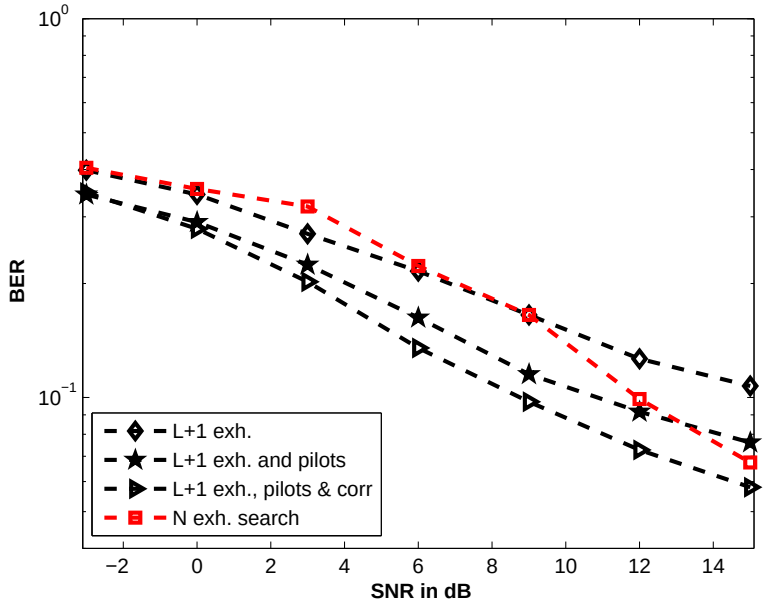

Figure 9: Comparison of Reduced search algorithms for BPSK-OFDM over a Rayleigh channel

algorithm. We 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.

6.1.8 BER vs SNR Comparison of Newton's Method

The Newton's method described in equation (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 10 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 11, the performance of 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

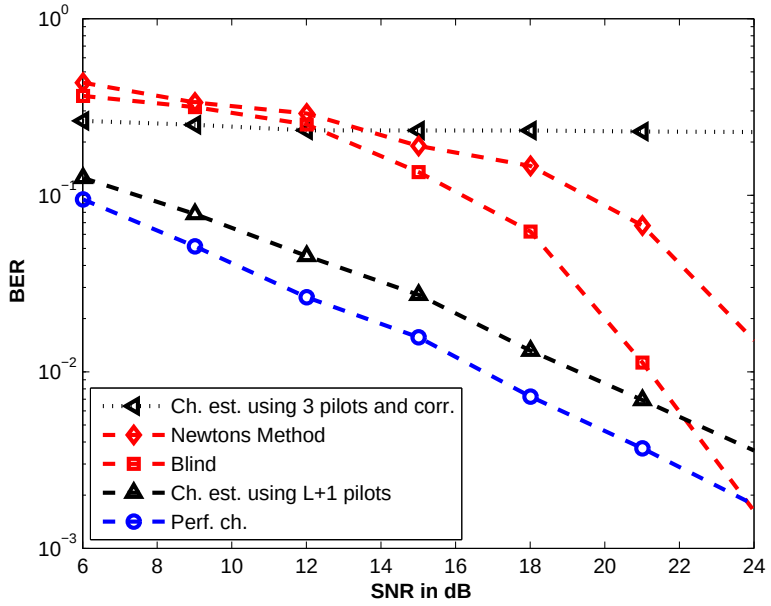

Figure 10: Comparison of Newton's Method for 4QAM-OFDM with $N = 16$ and $L = 4$ over a Rayleigh channel

$N = 64$ and $L = 16$. The Newton's method is initialized with an estimate obtained by using 12 pilots and channel correlation. Figure 11 clearly indicates that Newton's method performs quite well even for higher number of carriers.

6.2 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) 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

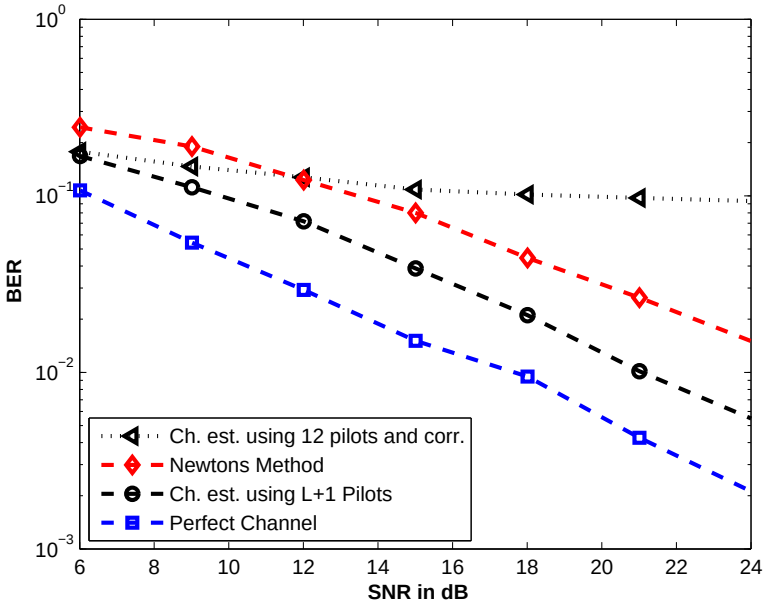

Figure 11: Comparison of Newton's Method for 4QAM-OFDM with $N = 64$ and $L = 16$ over a Rayleigh channel

(*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).

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

In Figure 12, 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.

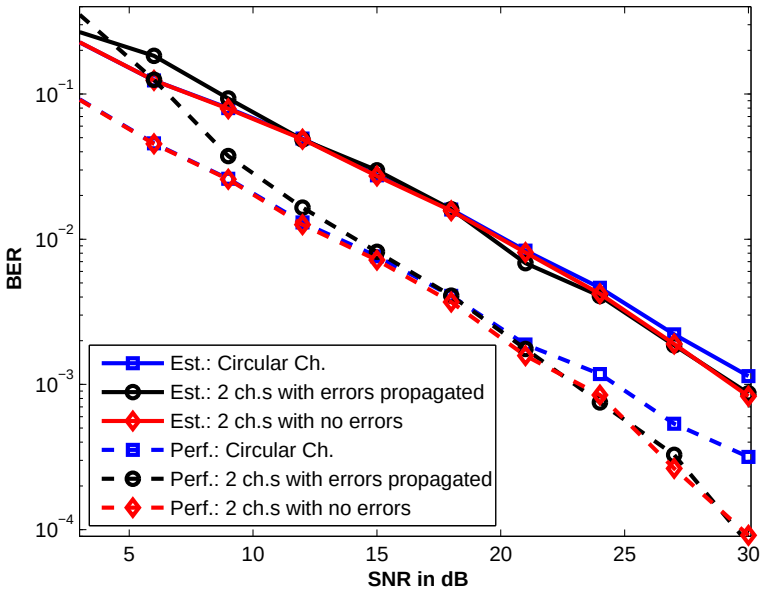

Figure 12: Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over a Rayleigh channel

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

In Figure 13, 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 12, the performance of the case when errors are propagated improves with increasing SNR.

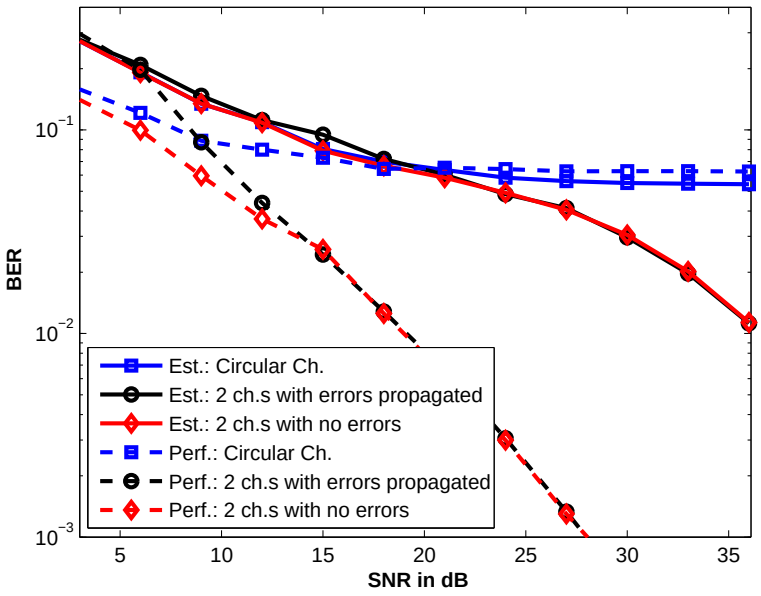

Figure 13: Comparison of perfect and pilot based estimation with enhanced equalization using CP for BPSK-OFDM over channel with zeros on FFT grid

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

Figure 14 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 modu-

lated 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.

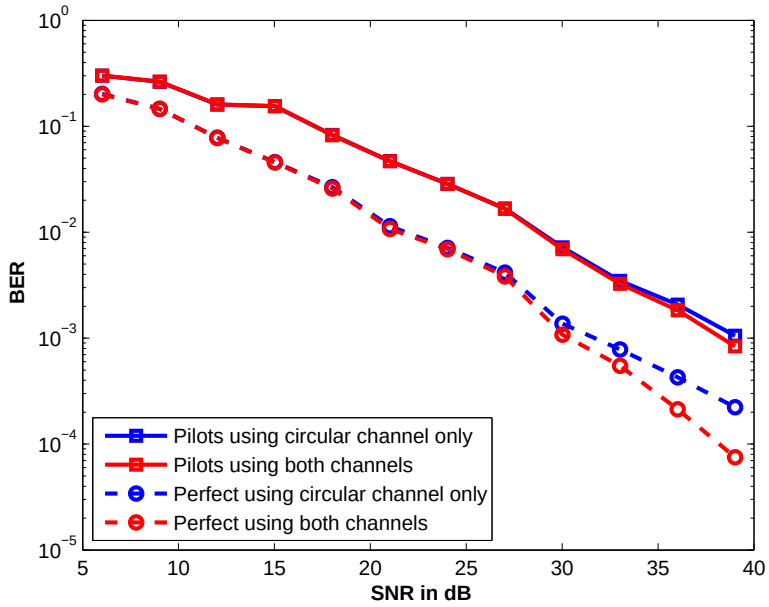

Figure 14: Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over a Rayleigh channel

6.2.4 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 15. 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.

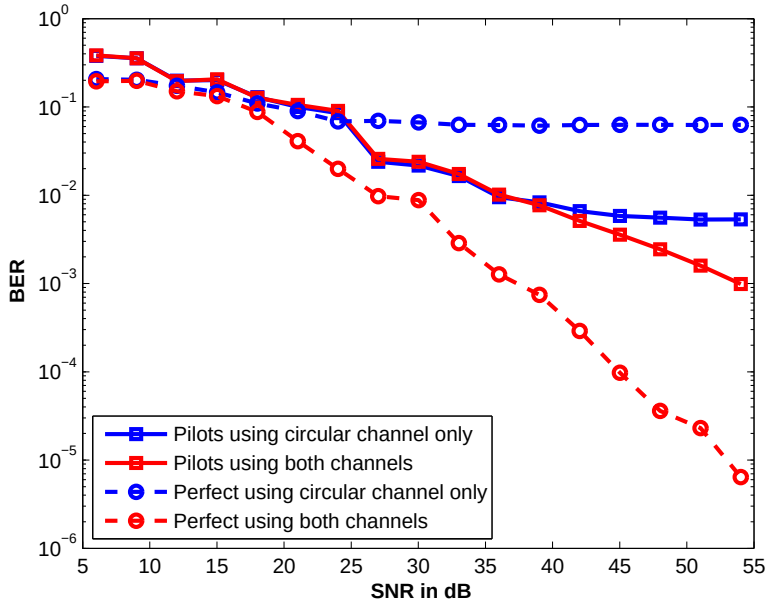

Figure 15: Comparison of perfect and pilot based estimation with enhanced equalization using CP for 16QAM-OFDM over channel with zeros on FFT grid

7 Conclusions, Recommendations, Outcomes, and Publications

7.1 Conclusions

In this project, 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 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 project. 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.

7.2 Recommendations

7.2.1 General Time Variant Case

This project deals with block fading channels i.e. the channel is assumed to be constant during the transmission of one block. 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 project. 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 project to perform channel estimation, ICI cancelation, and data detection.

7.2.2 Iterative Methods for Non-Constant Modulus Data

In Section 2, the blind algorithm presented is valid for constant as well as non-constant modulus data. The emphasis in this project 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

project. 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.

7.3 Summary of the Outcomes of the Project

The project resulted in the following outcomes:

1. The project studied the blind channel estimation and data detection in OFDM systems utilizing the cyclic prefix and finite alphabet constraint on the data.

2. The project showed the data can be recovered in OFDM systems blindly by using only a single OFDM block.

3. As two OFDM symbols are used, the blind algorithm developed in the project does not suffer from sign ambiguity that is inherent in all the blind algorithms present in the literature.

4. The blind algorithm presented in the project was compared with a state-of-art blind channel estimation algorithm [9]. The developed algorithm outperformed it easily.

5. As the blind algorithm presented in the project entails exponential complexity, iterative methods to reduce this complexity were also studied.

6. Out of the approximate methods studied in the project, the Genetic algorithm and Newton's method showed quite good performance.

7. The project also showed the improvement in performance when the cyclic prefix information was used in data detection with channel state information available (either through training or in the perfect case).

7.4 Conference Papers that Resulted from the Project

Here is a summary of the conference papers that resulted from this work.

1. T. Y. Al-Naffouri and A. A. Quadeer, "Blind maximum-likelihood data recovery in OFDM," IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP), vol. 1, pp. 2829-2832, Apr. 2008.

2. A. A. Quadeer, T. Y. Al-Naffouri and M. Shadaydeh, "Iterative blind data detection in constant modulus OFDM systems," European Signal Processing Conference (EUSIPCO), Aug. 2008.

7.5 Journal Paper that Resulted from the Project

The project also resulted in the following journal paper.

T. Y. Al-Naffouri and A. A. Quadeer, "Cyclic prefix based enhanced data recovery in OFDM," accepted in IEEE Transactions on Signal Processing.

7.6 Patent that Resulted from the Project

A patent was also filed from the work done in this project.

T. Y. Al-Naffouri and A. A. Quadeer, *Cylic prefix-based enhanced data recovery method*, Serial no. 12/385,076, US patent pending.

7.7 Master Thesis Related to the Project

This project also contributed to the following Master thesis

A. A. Quadeer, "(Semi) blind channel and data recovery in OFDM," Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, June 2008.

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 multicarrier 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] 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.

[16] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Vehicular Tech. Conf.*, vol. 2, pp. 1131−1135, May 1999.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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

[24] 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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] Y. Liang, H. Luo, and J. Huang, "Redundant precoding assisted blind channel estimation for OFDM systems," *Int. Conf. Signal Proc.*, vol. 3, 2006.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] F. Yang and W. Ser, "Adaptive semi-blind channel estimation for OFDM systems," *IEEE Veh. Technol. Conf.*, vol. 3, pp. 1773−1776, May 2004.

[55] 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.

[56] 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.

[57] W. Yang, Y. Cai, and Y. Xun, "Semi-blind Channel estimation for OFDM Systems," *IEEE Vehicular Tech. Conf.*, vol. 1, pp. 226−230, 2006.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space time wireless communications*. Cambridge University Press, 2003.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE Int. Conf. on Neural Networks*, vol. 4, pp. 1942−1948, Nov. 1995.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] A. Hjorungnes, 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.

[77] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons, Inc., 1999.

[78] H. Lutkepohl, *Handbook of Matrices.* John Wiley and Sons, Inc., 1996.

[79] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook.* Version 2006.

[80] 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.

[81] 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.

[82] Ali H. Sayed, *Fundamentals of Adaptive Filtering.* John Wiley and Sons, Inc., 2003.

[83] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation.* Prentice Hall, 2000.

[84] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451−1458, Oct. 1998.

[85] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[86] T. Y. Al-Naffouri, Adaptive Algorithms for Wireless Channel Estimation: Transient Analysis and Semi-Blind Design, Stanford University, Jan. 2005.