UNIVERSITY OF SASKATCHEWAN

Channel estimation, data detection and carrier frequency offset estimation in
OFDM systems

by

Malihe Ahmadi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SASKATOON, SASKATCHEWAN

DECEMBER, 2007

To my beloved parents, Akram and Morteza, and my sister and brother, Marzieh an Mohammad

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering

University of Saskatchewan

Saskatoon, Saskatchewan (S7N 5A9)

Abstract

Orthogonal Frequency Division Multiplexing (OFDM) plays an important role in the implementation of high data rate communication. In this thesis, the problems of data detection and channel and carrier frequency offset estimation in OFDM systems are studied.

Multi-symbol non-coherent data detection is studied which performs data detection by processing multiple symbols without the knowledge of the channel impulse response (CIR).

For coherent data detection, the CIR needs to be estimated. Our objective in this thesis is to work on blind channel estimators which can extract the CIR using just one block of received OFDM data. A blind channel estimator for (Single Input Multi Output) SIMO OFDM systems is derived. The conditions under which the estimator is identifiable is studied and solutions to resolve the phase ambiguity of the proposed estimator are given.

A channel estimator for superimposed OFDM systems is proposed and its CRB is derived. The idea of simultaneous transmission of pilot and data symbols on each subcarrier, the so called superimposed technique, introduces the efficient use of bandwidth in OFDM context. Pilot symbols can be added to data symbols to enable CIR estimation without sacrificing the data rate.

Despite the many advantages of OFDM, it suffers from sensitivity to carrier frequency offset (CFO). CFO destroys the orthogonality between the subcarriers. Thus, it is necessary for the receiver to estimate and compensate for the frequency offset. Several high accuracy estimators are derived. These include CFO estimators, as well as a joint iterative channel/CFO

estimator/data detector for superimposed OFDM. The objective is to achieve CFO estimation with using just one OFDM block of received data and without the knowledge of CIR.

Acknowledgement

I would like to express my sincere gratitude to Dr. Saadat Mehr for his invaluable time and support and advice. Working under his supervision has been an honor for me. I am grateful to University of Saskatchewan, Electrical Engineering Department's faculty member, especially to Dr. Salt and Dr. Nguyen for making my time in school intriguing and memorable.

I wish to record my gratitude to my beloved parents, Akram and Morteza, and my lovely sister and brother, Marzieh and Mohamad, for their patience and continuous support. Special thank goes to my friends Maryam, Zohreh, Brian, Ben, Marie, Gillian and University of Saskatchewan Electrical Engineering Department's staff.

Acronyms

Acronyms	Definition
AWGN	Additive white Gaussian noise
BER	Bit error rate
BPSK	Binary phase shift keying
CFO	Carrier frequency offset
CGRV	Complex Gaussian random variable
CIR	Channel impulse response
CN	Complex normal
CP	Cyclic Prefix
CRB	Cramer-Rao bound
DFT	Discrete Fourier Transform
FDM	Frequency division multiplexing
FDMA	Frequency division multiplexing access
FFT	Fast Fourier Transform
FIM	Fisher information matrix
FIR	Finite impulse response
DPSK	Differential phase shift keying
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
ISI	Inter-symbol interference
LAN	Local area network

LS	Least squares
MIMO	Multi input multi output
ML	Maximum likelihood
MMSE	Minimum mean square error
MSE	Mean square error
OFDM	Orthogonal frequency division multiplexing
pdf	Probability density function
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
SD	Sphere decoding
SIMO	Single input multi output
SISO	Single input single output
SNR	Signal to noise ratio
V-BLAST	Vertical Bell laboratories layered space time architecture
VC	Virtual carrier
ZF	Zero forcing

List of Contents

x

List of Figures

List of Tables

Chapter One: Introduction

1.1 Background and motivation

The need for wireless connections with different systems and devices has been growing spectacularly in the past several years. In particular, there is a need to access high speed wireless internet to exchange media such as photos, music and video. Inter-symbol Interference (ISI) is a common problem in high speed wireless data transmission. Among the various transmission techniques to overcome ISI and meet high data rate demand, Orthogonal Frequency Division Multiplexing (OFDM) is a promising candidate.

Although the basic concept of OFDM was proposed almost fifty years ago, it was not widely utilized because its implementation required a large number of analog devices. The idea of using the Discrete Fourier Transform (DFT) for the implementation of the modulation and demodulation of OFDM signals has made OFDM technologically practical and commercially affordable. Currently, OFDM has been widely adopted and implemented in wired and wireless communication systems. It has been exploited for various wideband data communications, such as mobile radio FM channels, high bit rate digital subcarrier lines (HDSL, 1.6 Mbps) [47], asymmetric digital subcarrier lines (ADSL, up to 6 Mbps) [48], very high speed digital subcarrier lines (VHDSL, 100 MHz) and digital audio broadcasting (DAB) [49]. It is used in high data rate wireless local area network (WLAN) standards, such as HIPERLAN and IEEE 802.11a, providing data rates of up to 54Mbits/s and is being considered for the fourth generation (4G) mobile wireless systems and beyond [51].

OFDM is a multicarrier transmission technique which divides the available spectrum into many subcarriers. Each one is modulated by a low data rate stream. OFDM is similar to

Frequency Division Multiplexing Access (FDMA) in that multiple access is achieved by subdividing the available bandwidth into multiple channels, which are then allocated to users. However, OFDM uses the spectrum much more efficiently by spacing the channels much closer together. This is achieved by making all subcarriers orthogonal to one another, preventing interference between the closely spaced carriers.

In the concept of OFDM, there are some issues at the receiver to be considered. Data detection is one of those issues and it is the first problem to be studied in this thesis. Data detection algorithms can be categorized in two groups: Coherent and Non-coherent. A Non-coherent algorithm detects transmitted data without requiring the knowledge of the channel. In chapter two, we study a non-coherent multi symbol data detection algorithm for OFDM systems. When the data detection is coherent, the knowledge of channel impulse response (CIR) is needed, so the channel estimation is the second issue to be investigated. Channel estimators which don't use pilots (blind channel estimators) are of great interest. Our objective in this thesis is to develop blind channel estimators which use a very short amount of OFDM blocks of received data to extract the CIR. In chapter three, we study the problem of blind channel estimation for SIMO OFDM systems. Then, we develop a channel estimator for superimposed OFDM systems in chapter four. Both of our estimators use just one OFDM block of data so that they don't introduce large delay to the process of estimation. The third problem we worked on is the carrier frequency offset (CFO) estimation; which is necessary to compensate the non orthogonality between subcarriers and it is the subject of chapter five. Our objective for CFO estimation is to develop estimators which don't require the knowledge of CIR and those which require a very limited amount of OFDM blocks of received data to perform the estimation.

In the coherent data detection algorithms, the known (estimated) CIR is used for data detection. Thus, channel estimation is necessary ahead of data detection. The use of pilot tones for channel estimation constitutes a significant overhead or bandwidth loss, motivating the development of blind techniques for OFDM. Several blind channel estimators have been proposed by using statistical or deterministic properties of the transmitted and received signals. However, most of these blind estimators typically use averaging techniques over a large number of OFDM blocks (up to several thousands in some cases). These estimators thereby introduce a considerable latency into the overall system and have high complexity. Besides, because of the time varying nature of wireless channels, these averaging techniques are not feasible. We will develop a blind channel estimator for SIMO OFDM systems requiring only one OFDM block in chapter three. Thus, this estimator suits for communication over time varying channels.

The next channel estimator presented in chapter four is for superimposed OFDM systems. The idea of superimposed pilot and data was first proposed for analog communication and was later extended to digital signal carrier systems. The implementation of superimposed techniques in OFDM systems has attracted attention since they allow the simultaneous communication of data and pilots on the same subcarriers. Therefore they use the bandwidth more efficiently. Our estimator can estimate CIR using just one block of OFDM received data, thus it doesn't introduce large latency.

Although OFDM has many advantages over single carrier transmission, the desirable features of OFDM also come with some disadvantages. Being a multi carrier system, one major disadvantage of OFDM is its sensitivity to carrier frequency offset (CFO). Frequency offsets cause the loss of orthogonality among the subcarriers and result in intercarrier interference between subcarriers [52]. Hence the receiver needs to estimate and compensate the CFO before

taking the DFT on received samples. In Chapter 5, we will develop several CFO estimators for OFDM and superimposed OFDM systems.

Next, we will describe wireless channel models and OFDM principles.

1.2 Wireless Channel Model

The design of spectrally efficient wireless communication systems requires a good understanding of the physical properties of the wireless propagation environment. The following is a brief review of major concepts of wireless communication channels.

1.2.1 Multipath fading

When a signal is transmitted over a wireless channel, in most applications, there is no line of sight (LOS) path between the transmitter and receiver. Consequently, the receiver signal consists of multiple copies of the transmitted signal reflection, refraction and diffraction. At some times, the randomly distributed amplitudes, phases and arrival angles of these multipath copies may add destructively and result in weak received signal. At other times, they may add constructively, which result in a strong received signal. Physical factors in the propagation channel influence multipath fading. First, the time spread of the channel causes the transmitted signal to undergo either flat or frequency selective fading; second, the Doppler spread determines whether the signal undergoes a slow or fast fading.

1.2.2 Delay spread

The channel impulse response (CIR) can be written as

$$h(t) = \sum_{l=0}^{L-1} h_l(t)\delta(t - \tau_l(t)) \qquad (1.1)$$

where $h_l(t)$ is the complex gain of the *l*th path and $\tau_l(t)$ is the time delay for the *l*th path. The total number of paths is *L*. Usually; we assume that $h_l(t)$ and $\tau_l(t)$ are constant over one OFDM block. So, we can omit the variable t in the following. Then, the frequency response of the channel is given by

$$H(f)=\sum_{l=0}^{L-1}h_l e^{-j2\pi f\tau_l}\ .\tag{1.2}$$

The weighted average delay is given by

$$\bar{\tau}=\frac{\sum_{l=0}^{L-1}|h_l|^2\tau_l}{\sum_{l=0}^{L-1}|h_l|^2}\ .\tag{1.3}$$

The rms delay spread is defined as

$$\tau_{rms}^2=\overline{\tau^2}-\bar{\tau}^2\ .\tag{1.4}$$

where $\overline{\tau^2}=\dfrac{\sum_{l=0}^{L-1}|h_l|^2\tau_l^2}{\sum_{l=0}^{L-1}|h_l|^2}\ .$ (1.5)

     Analogous to the delay spread parameter in the time domain, the coherence bandwidth is used to characterize the channel in the frequency domain. Coherence bandwidth is a statistical measure of the range of frequencies over which the channel can be considered flat and there are different definitions for it. When the coherence bandwidth is defined as the bandwidth over which the frequency correlation is above 90%, the coherence bandwidth is approximately [1]

$$B_c \approx \frac{1}{50\tau_{rms}} . \qquad (1.6)$$

When the transmitted signal has its spectrum bandwidth greater than the coherence bandwidth of the channel, frequency components of the signal separated by more than B_c are subject to different amplitude gains and phase rotations. In such a case, the channel is called frequency selective. In other words, frequency selective fading is caused by multipath delays which approach or exceed symbol period of the transmitted symbol. In OFDM systems, since a high-speed data sequence is split into a number of parallel subcarriers, the symbol period on each subcarrier is usually much larger than the delay spread of the channel, so the fading on each subcarrier is always flat fading. On the other hand, the total bandwidth of all subcarriers is larger than the channel coherence bandwidth, which means different subcarrier suffer from different fading factors, when their difference is larger than B_c.

1.2.3 Doppler spread

Doppler spread gives us information about the time varying nature of the channel, which is measured by coherence time. The Motion of the receiver, transmitter or reflectors results in a Doppler shift of the frequency. If T_c is defined as the time over which the time correlation function of channel is above 0.5

$$T_c = \frac{9}{16\pi f_d} \qquad (1.7)$$

where $f_d = \frac{v f_c}{c}$, f_c is carrier frequency, v is the velocity of the transmitter relative to the receiver, c is speed of light ($3 \times 10^8 \, m/s$). The definition of coherent time implies that two received signals have a strong potential for amplitude correlation if they arrive within a time

duration less than T_c. If T_s denotes sampling period, then a signal will undergo fast fading if $T_s >$ T_c, and a signal will undergo slow fading if $T_s << T_c$.

1.2.4 Channel Model realization

A very effective and widely applied method to simulate a fading channel is based on the sum of sinusoids, which is known as the Jakes' Model [2]. It is a widely used method for simulating time-correlated Rayleigh fading waveforms. The model assumes that scattered signals are distributed uniformly around the mobile.

1.3   OFDM principles

The principles of orthogonal frequency division multiplexing (OFDM) modulation have been in existence for several decades. However, in recent years these techniques have quickly moved out of textbooks and research laboratories and into practice in modern communications systems. The techniques are employed in data delivery systems over the phone line, digital radio and television, and wireless networking systems. In what follows, we will review the principal of OFDM.

A single carrier system modulates information onto one carrier using frequency, phase, or amplitude adjustment of the carrier. For digital signals, the information is in the form of bits, or collections of bits called symbols, that are modulated onto the carrier. As higher bandwidths (data rates) are used, the duration of one bit or symbol of information becomes smaller. The system becomes more susceptible to loss of information from impulse noise, signal reflections and other impairments. These impairments can impede the ability to recover the information sent. In addition, as the bandwidth used by a single carrier system increases, the susceptibility to interference from other continuous signal sources becomes greater.

Frequency division multiplexing (FDM) extends the concept of single carrier modulation by using multiple subcarriers within the same single channel. The total data rate to be sent in the channel is divided between the various subcarriers. The data do not have to be divided evenly nor do they have to originate from the same information source. Advantages include using separate modulation/demodulation customized to a particular type of data, or sending out banks of dissimilar data that can be best sent using multiple, and possibly different, modulation schemes. FDM offers an advantage over single-carrier modulation in terms of narrowband frequency interference since this interference will only affect one of the frequency sub-bands. The other subcarriers will not be affected by the interference. Since each subcarrier has a lower information rate, the data symbol periods in a digital system will be longer, adding some additional immunity to impulse noise and reflections. FDM systems usually require a guard band between modulated subcarriers to prevent the spectrum of one subcarrier from interfering with another. These guard bands lower the system effective information rate when compared to a single carrier system with similar modulation. If the FDM system above had been able to use a set of subcarriers that were orthogonal to each other, a higher level of spectral efficiency could have been achieved. The guard bands that were necessary to allow individual demodulation of subcarriers in an FDM system would no longer be necessary. The use of orthogonal subcarriers would allow the subcarriers spectra to overlap, thus increasing the spectral efficiency. As long as orthogonality is maintained, it is still possible to recover the individual subcarriers signals despite their overlapping spectrums.

Recall from signals and systems theory that the sinusoids of the DFT form an orthogonal basis set, and a signal in the vector space of the DFT can be represented as a linear combination of the orthogonal sinusoids. One view of the DFT is that the transform essentially correlates its

input signal with each of the sinusoidal basis functions. If the input signal has some energy at a certain frequency, there will be a peak in the correlation of the input signal and the basis sinusoid that is at that corresponding frequency. This transform is used at the OFDM transmitter to map an input signal onto a set of orthogonal subcarriers, i.e., the orthogonal basis functions of the DFT. The orthogonal and uncorrelated nature of the subcarriers is exploited in OFDM with powerful results. Since the basis functions of the DFT are uncorrelated, the correlation performed in the DFT for a given subcarrier only sees energy for that corresponding subcarrier. The energy from other subcarriers does not contribute because it is uncorrelated. This separation of signal energy is the reason that the OFDM subcarriers spectrums can overlap without causing interference. Figure 1-1 shows the resultant frequency spectrum. In the frequency domain, the resulting sine function side lobes produce overlapping spectra. The individual peaks of sub-bands all line up with the zero crossings of the other sub-bands. This overlap of spectral energy does not interfere with the system's ability to recover the original signal.
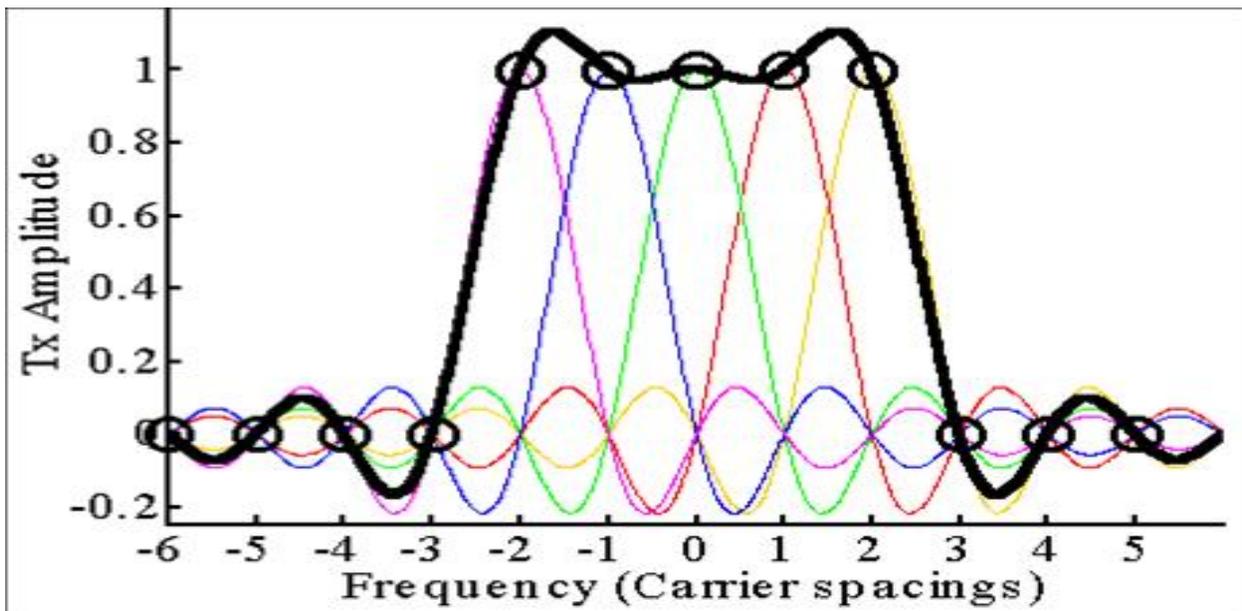
Figure 1-1: Overall spectrum of OFDM signal shown with five subcarriers

The idea behind the analog implementation of OFDM can be extended to the digital domain by using the Discrete Fourier Transform (DFT) and its counterpart, the Inverse Discrete Fourier Transform (IDFT). The IDFT is used to take in frequency-domain data and convert it to time-domain data. In order to perform that operation, the IDFT correlates the frequency-domain input data with its orthogonal basis functions, which are sinusoids at certain frequencies. This correlation is equivalent to mapping the input data onto the sinusoidal basis functions. In practice, OFDM systems are implemented using a combination of Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) blocks that are mathematically equivalent versions of the DFT and IDFT, respectively, but more efficient to implement.

An OFDM system treats the source symbols (e.g., the QPSK or QAM symbols that would be present in a single carrier system) at the transmitter as though they are in the frequency-domain. These symbols are used as the inputs to an IFFT block that brings the signal into the time domain. The IFFT takes in N symbols at a time where N is the number of subcarriers in the system. Each of these N input symbols has a symbol period of T_s seconds. Recall that the basis functions for an IFFT are N orthogonal sinusoids. These sinusoids each have a different frequency and the lowest frequency is DC. Each input symbol acts like a complex weight for the corresponding sinusoidal basis function. Since the input symbols are complex, the value of the symbol determines both the amplitude and phase of the sinusoid for that subcarrier. The IFFT output is the summation of all N sinusoids. Thus, the IFFT block provides a simple way to modulate data onto N orthogonal subcarriers. The block of N output samples from the IFFT make up a single OFDM block. The length of the OFDM block is NT_s where T_s is the IFFT input symbol period mentioned above.

After some additional processing, the time-domain signal that results from the IFFT is transmitted across the channel. At the receiver, an FFT block is used to process the received signal and bring it into the frequency domain. Ideally, the FFT output will be equal to the original symbols that were sent to the IFFT at the transmitter. When plotted in the complex plane, the FFT output samples will form a constellation, such as 16-QAM.

A major problem in most wireless systems is the presence of a multipath channel. In a multipath environment, the transmitted signal reflects off of several objects. As a result, multiple delayed versions of the transmitted signal arrive at the receiver. The multiple versions of the signal cause the received signal to be distorted. Inter-symbol Interference (ISI) is the problem caused by multipath characteristics of wireless channel. It is the result of interference amongst consequent OFDM subcarrier. The solution is to insert the guard interval at the beginning of each OFDM block.

Recall that in continuous time, a convolution in time is equivalent to a multiplication in the frequency domain. This property is true in discrete time only if the signals are of infinite length or if at least one of the signals is periodic over the range of the convolution. It is not practical to have an infinite length OFDM block; however, it is possible to make the OFDM block appears periodic. This periodic form is achieved by replacing the guard interval with something known as a *cyclic prefix* of length N_g samples. The cyclic prefix is a replica of the last N_g samples of the OFDM block. Since it contains redundant information, the cyclic prefix is discarded at the receiver. This step removes the effects of inter-symbol interference. Because of the way in which the cyclic prefix was formed, the cyclically-extended OFDM block now appears periodic when convolved with the channel. An important result is that the effect of the channel becomes multiplicative.

Denoting the block of symbols of length N by $\mathbf{X} = [X[0], X[1], ..., X[N-1]]^T$, and CIR length with L, we create an $N+L-1$ input block of (data symbol plus guard interval)

$$\mathbf{d} = [X[N-L+1], X[N-L+2], ..., X[N-1], X[0], X[1], ..., X[N-1]]^T \qquad (1.8)$$

i.e., we add a *prefix* of length *L-1* consisting of data symbols rotated cyclically. With this input to the channel and considering channel model as a finite impulse response channel, $h(n) = \sum_{l=0}^{L-1} h_l \delta(n-l)$, the output of the channel is

$$y_n = \sum_{l=0}^{L-1} h_l d_{n-l} + w_n. \qquad (1.9)$$

The ISI extends over the first *L-1* symbols and the receiver ignores it by considering only the output over the time interval $n \in [L, N+L-1]$. Due to the cyclic prefix, the output over this time interval (of length N) is

$$y_n = \sum_{l=0}^{L-1} h_l X[(n-L-l) \bmod N] + w_n. \qquad (1.10)$$

Denoting the output of length N by $\mathbf{y} = [y[L], ..., y[N+L-1]]^T$ and the channel by a vector of length N

$$\mathbf{h} = [h_0, h_1, ..., h_{L-1}, 0, ..., 0]^T \qquad (1.11)$$

(1.10) can be written as

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{X} + \mathbf{w} \qquad (1.12)$$

where $\mathbf{w} = [w[L], ..., w[N + L - 1]]^T$ denotes a vector of i.i.d. $CN\left(0, \sigma_w^2\right)$ random variables. The

notation $\otimes$ is used for *cyclic convolution* in (1.12). There are two definitions for DFT, in order

to make the DFT a unitary transfer, we scale it by $\dfrac{1}{\sqrt{N}}$. Thus, DFT of $\mathbf{X}$ is defined to be [19]

$$x_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n \exp(\frac{-j2\pi nk}{N}), \quad k = 0, ..., N-1. \tag{1.13}$$

Taking the DFT of both sides of (1.12) and using the identity

$$DFT(\mathbf{h} \otimes \mathbf{X})_n = \sqrt{N} DFT(\mathbf{h})_n \times DFT(\mathbf{X})_n, \quad n = 0, 1, ..., N-1, \tag{1.14}$$

we can write (1.12) as

$$Y_k = H_k X_k + W_k, \quad k = 0, 1, ..., N-1 \tag{1.15}$$

where we denote $W_0, ..., W_{N-1}$ as the N point DFT of the noise vector $\mathbf{w} = [w[1], ..., w[N]]^T$. The

vector $\mathbf{H} = [H[1], ..., H[N]]^T$ is defined as the DFT of the L tap channel $\mathbf{h}$, scaled by $\sqrt{N}$

$$H_k = \sum_{l=0}^{L-1} h_l \exp(\frac{-j2\pi lk}{N}), \quad k = 0, ..., N-1 \tag{1.16}$$

These operations are illustrated in Figure 1-2 which has the following interpretation: The

binary stream of data ($N \times \log_2 M$ bits where M is the size of constellation Q) are mapped into N

data symbols in constellation Q and form data symbols on N *tones* or *subcarriers*. Then a Serial

to Parallel converter makes a group of N subcarriers ready for IDFT. The data symbols on the

subcarriers are then converted (through the IDFT) to time domain. The procedure of introducing

the cyclic prefix (guard interval) before transmission allows for the removal of ISI. Then the

Parallel to Serial converts the group of data to a stream that passes through the channel. The

receiver ignores the part of the output signal containing the cyclic prefix (along with the ISI

terms) and converts the length N symbols back to the frequency domain through the DFT. The data symbols on the subcarriers are maintained to be orthogonal as they propagate through the channel and hence go through narrowband parallel channels. This interpretation justifies the name of OFDM for this communication scheme.
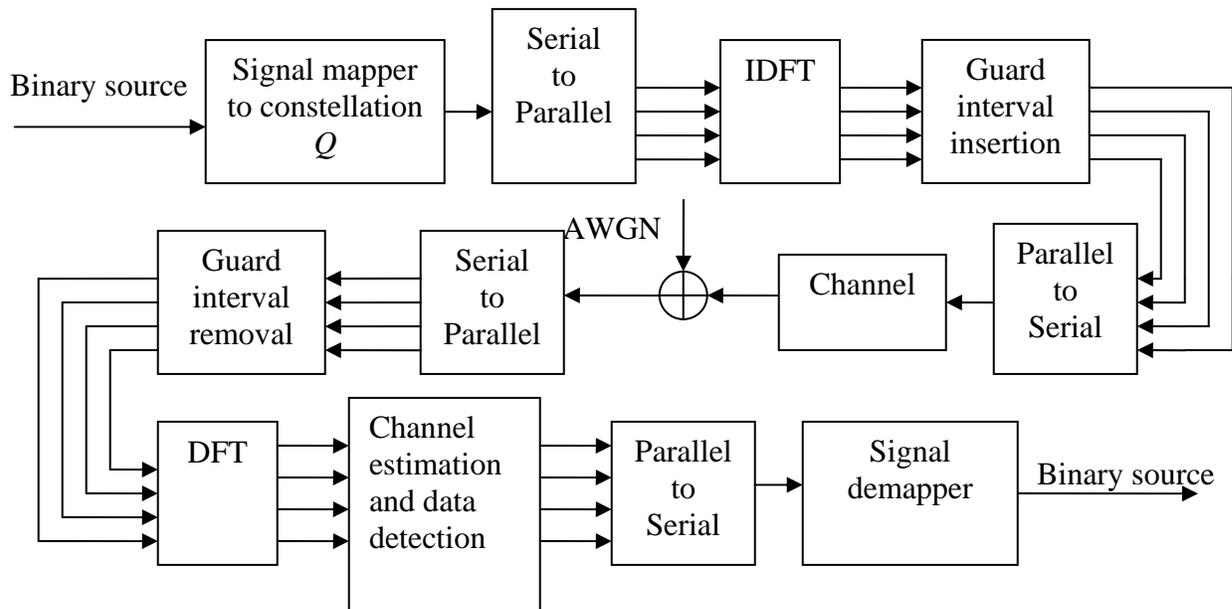
Figure 1-2: Baseband OFDM system model

In a digital communication system, the symbols that arrive at the receiver have been convolved with the time domain channel impulse response of length L samples. Thus, the effect of the channel is convolutional. In order to undo the effects of the channel, another convolution must be performed at the receiver using a time domain filter known as an equalizer. The length of the equalizer needs to be on the order of the length of the channel. The equalizer processes symbols in order to adapt its response in an attempt to remove the effects of the channel. Such an equalizer can be expensive to implement in hardware and often requires a large number of

symbols in order to adapt its response to a good setting. In OFDM, the time domain signal is still convolved with the channel response. However, the data will ultimately be transformed back into the frequency domain by the FFT in the receiver. Because of the periodic nature of the cyclically-extended OFDM symbol, this time domain convolution will result in the multiplication of the spectrum of the OFDM signal (i.e., the frequency domain constellation points) with the frequency response of the channel. The result is that each subcarrier symbol will be multiplied by a complex number equal to the channel's frequency response at that subcarrier frequency [(1.15)]. Each received subcarrier experiences a complex gain (amplitude and phase distortion) due to the channel. In order to undo these effects, a frequency domain equalizer is employed. Such an equalizer is much simpler than a time domain equalizer. The frequency domain equalizer consists of a single complex multiplication for each subcarrier. For the simple case of no noise, the ideal value of the equalizer's response is the inverse of the channel's frequency response. Thus the receiver needs to know (or estimate) the CIR so that it can detect data symbols. At the end, the stream of data symbols can be demapped to stream of binary data according to constellation Q.

To detect the transmitted data, coherent and non-coherent algorithms can be established. In general, coherent algorithms give better error rate performance [19]. In chapter two, we will study two non-coherent data detectors. For the case of coherent data detection, the channel impulse response (CIR) should be estimated. In chapter three and four we will study two channel estimation algorithms.

1.4 Detection Problem

The detection problem for a MIMO system (or a multicarrier system) can be formulated as

$$\mathbf{r} = \mathbf{Hx} + \mathbf{w} \tag{1.17}$$

where $\mathbf{x} \in Z^n$, $\mathbf{r}, \mathbf{w} \in R^m$ denote the system input, output and additive noise, and $\mathbf{H} \in R^{m \times n}$ represents the transfer matrix, and Z^n denotes the n-dimensional integer lattice and R^m denotes the m-dimensional real space. Generally, the noise terms $w_i, i = 1, ..., m$ are independent and identically distributed (i.i.d) zero mean Gaussian random variables with variance σ_w^2. If we assume that $n \geq m$ and $\mathbf{H}$ has full column rank and is perfectly known at the receiver, the optimal ML detector that minimize the average error probability is given by so called integer least square problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in Z^n} \left\| \mathbf{r} - \mathbf{Hx} \right\|^2 , \tag{1.18}$$

where $\| . \|$ denotes the Euclidean norm. Eq. (1.18) is known as the closest vector problem in the lattice theory. The lattice generated by a generation matrix $\mathbf{H}$ is [3]

$$\Delta(\mathbf{H}) = \{ \mathbf{Hx} : \mathbf{x} \in Z^n \} . \tag{1.19}$$

The columns of $\mathbf{H}$ are called basis vectors of Δ and the number n is called the dimension of Δ. Eq. (1.17) is a model for real signal. Often complex signal constellations such as quadrature amplitude modulation (QAM) are used. In this case, the resulting complex detection problem can be transformed into an equivalent real problem as

$$\tilde{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}} \in Q'^n} \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{H}} \tilde{\mathbf{x}} \right\|^2 \tag{1.20}$$

where

$$\tilde{\mathbf{r}} = \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix} , \quad \tilde{\mathbf{x}} = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} \tag{1.21}$$

and

$$\tilde{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix} \tag{1.22}$$

Q' denotes the real constellation after the transformation and $\Re\{.\}$, $\Im\{.\}$ denote the real and imaginary part of a function respectively. For example, a square QAM constellation is transformed to a pulse amplitude modulation (PAM) constellation in (1.20). Since (1.20) is similar to (1.18), we will concentrate on how to solve (1.18) in the following using two classic detection algorithms in communication theory. The vertical Bell Laboratories Layered Space-Time Architecture (V-BLAST) detection algorithm [4], which is suboptimal, and the optimal sphere decoding algorithm [5].

1.4.1 V-BLAST algorithm

The V-BLAST detection algorithm consists of nulling and interference cancellation. Nulling is performed by linearly weighing the received symbols to satisfy the zero forcing (ZF) or minimum mean square error (MMSE) performance criterion [4]. Denoting the *i*th column of $\mathbf{H}$ as $(\mathbf{H})_i$, the zero forcing nulling vector $\mathbf{w}_i$, $i = 1, ..., n$ is chosen such that

$$\mathbf{w}_i^T (\mathbf{H})_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{1.23}$$

For interference cancellation, the effect of already detected symbols can be subtracted from symbols yet to be detected. This improves the overall performance when the order of detection is chosen carefully. For example, denoting the received vector $\mathbf{r}$ by $\mathbf{r}_1$, if the nulling vector is $\mathbf{w}_1$, the first symbol is then detected by

$$\hat{x}_1 = \arg\min_{x \in Q} \left| x - \mathbf{w}_1^H \mathbf{r}_1 \right|^2 \tag{1.24}$$

The interference due to $\hat{x}_1$ on the other symbols can be subtracted by taking $\mathbf{r}_2 = \mathbf{r}_1 - \hat{x}_1 \mathbf{h}_1$.

Assuming $\hat{x}_1 = x_1$ (i.e., the decision is correct), the next symbol x_2 is then detected using $\mathbf{w}_2$.

The detection process consists of n iterations. In the kth iteration, the signal with maximum post

detection SNR among the remaining $n\text{-}k\text{+}1$ symbols is detected, which is known to be the optimal

detection order. The post detection SNR for the kth detected symbol is given by

$$\rho_k = \frac{E\{|x_k|^2\}}{\sigma_n^2 \|\mathbf{w}_k\|^2}.$$

(1.25)

From (1.25), maximizing ρ_k is equivalent to minimizing $\|\mathbf{w}_k\|^2$. The whole algorithm is

described as follows

- Initialization:

$$\mathbf{r}_1 = \mathbf{r}$$

(1.26)

$$\mathbf{G}_1 = \mathbf{H}^+$$

(1.27)

$$k_1 = \arg\min_{j} \|(\mathbf{G}_1)_j\|^2$$

(1.28)

- Recursion: for $i = 1,...,n$

$$\mathbf{w}_{k_i} = (\mathbf{G}_i)_{k_i}$$

(1.29)

$$\hat{x}_{k_i} = \arg\min_{x \in Q} \left|x - \mathbf{w}_{k_i}^H \mathbf{r}_i\right|^2$$

(1.30)

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \hat{x}_{k_i} (\mathbf{H})_{k_i}$$

(1.31)

$$\mathbf{G}_{i+1} = \mathbf{H}_{\overline{k_i}}^+$$

(1.32)

$$k_{i+1} = \arg \min_{j \notin \{k_1,...,k_i\}} \left\| (\mathbf{G}_{i+1})_j \right\|^2 \qquad (1.33)$$

where $(\mathbf{G}_i)_j$ is the jth row of matrix $\mathbf{G}_i$ and $\mathbf{H}_{\bar{k}_i}$ is obtained by zeroing the k_1, ..., k_ith columns of $\mathbf{H}$. Assuming $\mathbf{\Pi}$ is the column permutation matrix obtained from optimum order, we apply $\mathbf{\Pi}$ to $\mathbf{H}$. Let the QR factorization of $\mathbf{G} = \mathbf{H}\mathbf{\Pi}$ be

$$\mathbf{G} = [\mathbf{Q}_1, \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \qquad (1.34)$$

where $\mathbf{R}$ is an $n \times n$ upper triangle matrix, $\mathbf{0}$ is the $(m-n) \times n$ all zero matrix, $\mathbf{Q}_1$ is an $m \times n$ unitary matrix and $\mathbf{Q}_2$ is an $m \times (m-n)$ unitary matrix. Eq. (1.17) is equivalent to

$$\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{v} \qquad (1.35)$$

where $\mathbf{y} = \mathbf{Q}_1^T \mathbf{r}$ and $\mathbf{v} = \mathbf{Q}_1^T \mathbf{w}$ is also an i.i.d complex Gaussian vector with mean zero and variance σ_w^2. The second description of V-BLAST algorithm is given by

- for $i = 1,....,n$

$$\hat{x}_i = \arg \min_{x \in Q} \left| y_i - r_{i,i} x \right|^2 \qquad (1.36)$$

$$\mathbf{y} = \mathbf{y} - (\mathbf{R})_i \hat{x}_i \qquad (1.37)$$

  end

where $r_{i,i}$ is the *(i,i)*th entry of $\mathbf{R}$.

1.4.2 Sphere Decoding (SD) algorithm

Another important algorithm for solving (1.18) is sphere decoding. SD only tests the lattice points lying inside a hypershpere. The popularity of the SD stems from the fact that the SD offers

ML decoding at lower complexity, as opposed to the exponential complexity incurred by the exhaustive search. In [6], it has been proven that for a wide range of SNR and lattice dimension, the expected complexity of SD is polynomial, often cubic in the lattice dimension. Following the formulation in (1.35), the lattice point $\mathbf{Rx}$ lies in a sphere of radius d, if and only if

$$\|\mathbf{y} - \mathbf{Rx}\|^2 \le d^2 . \tag{1.38}$$

Eq. (1.38) can be written as

$$\sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{n} r_{i,j} x_j \right|^2 \le d^2 \tag{1.39}$$

where $r_{i,j}$ denotes the *(i,j)*th entry of $\mathbf{R}$. The left hand side of the above inequality can be expanded as

$$(y_n - r_{n,n} x_n)^2 + (y_{n-1} - r_{n-1,n} x_n - r_{n-1,n-1} x_{n-1})^2 + \dots + \sum_{i=1}^{n} \left| y_1 - \sum_{j=1}^{n} r_{1,j} x_j \right|^2 \le d^2 \tag{1.40}$$

where the first term depends only on x_n, the second term on x_n, x_{n-1} and so forth. Therefore a necessary condition for $\mathbf{Rx}$ to lie inside the sphere is that $d^2 \ge (y_n - r_{n,n} x_n)^2$, which is equivalent to x_n belonging to the interval

$$\left\lceil \frac{-d + y_n}{r_{n,n}} \right\rceil \le x_n \le \left\lfloor \frac{-d + y_n}{r_{n,n}} \right\rfloor \tag{1.41}$$

where $\lceil . \rceil$ denotes the smallest integer greater than or equal to its argument and $\lfloor . \rfloor$ denotes the largest integer less than or equal to its argument.

For each candidate x_n satisfying the above bound, we define $d_{n-1}^2 = d^2 - (y_n - r_{n,n} x_n)^2$. We can get the following stronger necessary condition for x_{n-1}

$$(y_{n-1} - r_{n-1,n}x_n - r_{n-1,n-1}x_{n-1})^2 \le d_{n-1}^2 \tag{1.42}$$

which leads to the following bound

$$\left\lceil \frac{-d_{n-1} + y_{n-1} - r_{n-1,n}x_n}{r_{n-1,n-1}} \right\rceil \le x_{n-1} \le \left\lfloor \frac{d_{n-1} + y_{n-1} - r_{n-1,n}x_n}{r_{n-1,n-1}} \right\rfloor \tag{1.43}$$

The SD chooses a candidate for x_{n-1} from the above region. We continue in the same fashion for

x_{n-2} and so on. The bounds for x_k are

$$\left\lceil \frac{-d_k + y_k - \sum_{j=k+1}^{n} r_{k,j}x_j}{r_{k,k}} \right\rceil \le x_k \le \left\lfloor \frac{d_k + y_k - \sum_{j=k+1}^{n} r_{k,j}x_j}{r_{k,k}} \right\rfloor \tag{1.44}$$

where $d_k^2 = d_{k+1}^2 - (y_{k+1} - \sum_{j=k+1}^{n} r_{k+1,j}x_j)^2$. If there is no lattice point within the bounds for

x_n, the SD goes back to x_{k+1} and chooses another candidate value from the corresponding region

for x_{k+1}. If the SD reaches x_1, the SD finds a candidate lattice point $\mathbf{x}'$ within the hypersphere of

radius d. SD checks the value of $\|\mathbf{y} - \mathbf{Rx}'\|^2$. If this value is less than d, it updates the radius d

which means the search space is limited by new radius. The above process continues until no

further lattice points are found within the hypersphere. The lattice point that achieves the

smallest value of $\|\mathbf{y} - \mathbf{Rx}\|^2$ within the hypersphere is deemed as the ML solution. If no point in

the sphere is found, the sphere is empty and search fails. In this case, the initial search radius d

must be increased and the search is restarted with the new squared radius. In [6], the authors

analyzed the complexity based on the statistical property of the problem. They choose an initial

radius based on the statistics of the problem.

1.5 Literature review and Thesis Outline

In OFDM systems, data detection can be done either coherently or non-coherently. Non-coherent data detection, allows the receiver to process the received data entirely without the knowledge of the channel impulse response (CIR). Two multi-symbol non-coherent data detection algorithms will be introduced in Chapter two.

Alternatively, coherent data detection in OFDM systems requires the CIR. The use of pilot tones for channel estimation [30][53][54] constitutes a significant overhead or bandwidth loss, motivating the development of semi blind and blind techniques for OFDM. They use statistical or deterministic properties of the transmit and receive signals; properties such as cyclic prefix (CP) and pilot induced redundancy, cyclostationarity, finite alphabet and virtual carriers have been exploited in [55][56][57].

Joint estimation of CIR and data detection has been proposed in [66]. Several semi blind data detection and channel estimation techniques has been given in [60][61][62][64]. In [59], channel estimation improvement is done by using noise reduction. In Chapter three, we will develop a blind channel estimator for SIMO OFDM systems and discuss the identifiablity conditions.

The idea of superimposed training; i.e., simultaneous information and pilot transfer, was first used for analog communications and then was advocated for digital communication systems. Superimposed pilots have been exploited for the purpose of channel estimation [63]. In Chapter four, we first explain the idea of superimposed pilots in OFDM systems, and then we develop an approximately ML channel estimator for superimposed OFDM.

Despite its several advantages, OFDM suffers from sensitivity to synchronization errors. A ML carrier frequency offset (CFO) estimator has been proposed in [65]. In [67], a joint channel and CFO estimator has been developed. In Chapter five, we first formulate the CFO

problem in OFDM and then develop several CFO estimators. Conclusion and future work are given in Chapter 6.

Chapter Two: Multi symbol data detection in OFDM

This chapter develops two ML multi symbol data detectors for OFDM. The chapter organization is as follows. Section 2.1 develops the system model of a ML non-coherent multi symbol detector in OFDM and afterwards it discusses two efficient detection algorithms. In Section 2.2, the ML multi symbol differential detector is derived. Section 2.3 gives the numerical results and Section 2.4 concludes the chapter.

2.1 Non-coherent Maximum Likelihood Multi Symbol decoder

In the previous chapter, we have seen that a time-dispersive (frequency selective) channel in time domain is transferred into a flat (one tap) channel in frequency domain in an OFDM system. Therefore we have [Eq. (1.15)],

$$Y(k) = H(k)X(k) + W(k), \quad 0 \le k \le N-1 \tag{2.1}$$

where N is the DFT length, $X(k)$ is pre-DFT transmitted data on subcarrier k drawn from constellation Q, $Y(k)$ is the post-DFT received data on subcarrier k, $H(k)$ is the frequency response of channel for subcarrier k, and $W(k)$ is DFT of channel noise for subcarrier k, which are given by the following equations :

$$H(k) = \sum_{l=0}^{L-1} h_l e^{-j2\pi lk/N} \tag{2.2}$$

$$W(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} w_n e^{-j2\pi nk/N} \qquad (2.3)$$

where $h_l \in CN(0, \sigma_l^2)$, $l = 0, 1, ..., L-1$ are independent CIR coefficients and have been considered Rayleigh fading and $w_n \in CN(0, \sigma_w^2)$, $n = 0, 1, ..., N-1$ is AWGN.

In what follows we derive a data detector which combines several received data (multi symbol) to detect the transmitted data [7][8]. This detector doesn't need to know the CIR (non-coherent).

At the receiver, the multiple symbol decoder processes N' consequently received symbols

$$\mathbf{Y}_k = [Y((N'-1)k - (N'-1)), ..., Y((N'-1)k)]^T$$

to obtain ML estimate $\hat{\mathbf{X}}_k$ of the corresponding N' transmitted symbols

$$\mathbf{X}_k = [X((N'-1)k - (N'-1)), ..., X((N'-1)k)]^T.$$

The non-coherent detector at the receiver will try to detect the transmitted data with no knowledge of $\mathbf{H}_k = [H((N'-1)k - (N'-1)), ..., H((N'-1)k)]^T$.

The value of N' is referred to as the observation window size, and window size can be considered less than DFT length. With growing window size, the memory of the process is more completely taken into account so the performance of the decoder will be improved with increasing window size. In following, the non-coherent ML estimate of $\mathbf{X}_k$ based on the observation of $\mathbf{Y}_k$ will be formulated while the window size is equal to DFT length ($N'=N$).

Vectorizing equation (2.1), we have

$$\mathbf{Y}_k = \mathbf{X}_{k,D}\mathbf{H}_k + \mathbf{W}_k \qquad (2.4)$$

where

$$\mathbf{X}_{k,D} = diag(\mathbf{X}_k) = diag(X((N-1)k - (N-1)),...,X((N-1)k)),$$

$$\mathbf{H}_k = \mathbf{F}_L \mathbf{h}, \ \mathbf{h} = [h_0, h_1, ..., h_{L-1}]^T \in C^L,$$

$\mathbf{F}$ is the DFT matrix ($[\mathbf{F}]_{k,l} = \dfrac{1}{\sqrt{N}} e^{-j(2\pi/N)kl}$, $k,l \in 0, 1, ..., N-1$),

$\mathbf{F}_L$ is a $N \times L$ submatrix (first L columns) of DFT matrix $\mathbf{F}$ scaled by $\sqrt{N}$ and

$\mathbf{W}_k = [W((N-1)k - (N-1)),...,W((N-1)k)]^T$. Considering the model of channel and noise

explained in previous chapter (channel and noise are both vectors of CGRV), $\mathbf{Y}_k$ is a vector of

complex Gaussian random variables (CGRV) (since it is a linear combination of CGRVs) and

the pdf of $\mathbf{Y}_k$ given $\mathbf{X}_k$ can be written (for simplicity, we will discard subscript k):

$$f(\mathbf{Y}|\mathbf{X}) = \frac{1}{(\pi)^N \times \sqrt{\det(\mathbf{R_{YY}})}} \exp(-\mathbf{Y}^H \mathbf{R_{YY}}^{-1} \mathbf{Y}) \qquad (2.5)$$

where correlation matrix $\mathbf{R_{YY}}$ is:

$$\mathbf{R_{YY}} = E\{\mathbf{YY}^H\} = E\{(\mathbf{X}_D\mathbf{H} + \mathbf{W})(\mathbf{X}_D\mathbf{H} + \mathbf{W})^H\}. \qquad (2.6)$$

Since $w_n, n = 0, 1, ..., N-1$ are uncorrelated CGRVs and DFT is a unitary transformation

($\mathbf{FF}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}$), W_k, $k = 0, 1, ..., N-1$ are also uncorrelated CGRVs [19]. Therefore,

$E\{\mathbf{WW}^H\} = \sigma_w^2 \mathbf{I}$. Then,

$$\mathbf{R_{YY}} = E\{\mathbf{YY}^H\} = \mathbf{X}_D\mathbf{R_{HH}}\mathbf{X}_D^H + \sigma_w^2\mathbf{I} \qquad (2.7)$$

where

$$\mathbf{R_{HH}} = E\{\mathbf{HH}^H\} = \mathbf{F}_L E\{\mathbf{hh}^H\}\mathbf{F}_L^H \qquad (2.8)$$

and $\mathbf{R_h} = E\{\mathbf{hh}^H\} = diag(\sigma_0^2,...,\sigma_{L-1}^2).$ \qquad (2.9)

If we consider transmitted data are drawn from a unitary constellation, we will have

$$\mathbf{R_{YY}} = \mathbf{X}_D(\mathbf{R_{HH}} + \sigma_w^2\mathbf{I})\mathbf{X}_D^H \tag{2.10}$$

In this case, we can say $\det(\mathbf{R_{YY}}) = \det(\mathbf{X}_D)\det(\mathbf{R_{HH}} + \sigma_w^2\mathbf{I})\det(\mathbf{X}_D^H)$ is independent of transmitted data. Removing terms which are independent of transmitted data, maximizing ML estimation equation (2.5) is equivalent to minimizing $\mathbf{Y}^H\mathbf{R_{YY}}^{-1}\mathbf{Y}$. Therefore, the decision rule for data detection will be

$$\hat{\mathbf{X}}_D = \arg\min_{\mathbf{X}_D}\{\mathbf{Y}^H\mathbf{R_{YY}}^{-1}\mathbf{Y}\}. \tag{2.11}$$

Since $\mathbf{C} = \mathbf{R_{HH}} + \sigma_w^2\mathbf{I}$ is Hermitian and positive definite, we can apply Cholesky factorization [46] on its inverse, so that

$$\mathbf{C}^{-1} = \mathbf{LL}^H \tag{2.12}$$

where $\mathbf{L}$ is a lower triangular matrix. Replacing (2.10) and (2.12) into (2.11), we can rewrite decision rule as

$$\hat{\mathbf{X}}_D = \arg\min_{\mathbf{X}_D}\{\mathbf{Y}^H\mathbf{X}_D\mathbf{C}^{-1}\mathbf{X}_D^H\mathbf{Y}\} = \arg\min_{\mathbf{X}_D}\{\mathbf{Y}^H\mathbf{X}_D\mathbf{LL}^H\mathbf{X}_D^H\mathbf{Y}\}. \tag{2.13}$$

Defining $\mathbf{X} = [X((N-1)k-(N-1)), ..., X((N-1)k)]^T$

and $\mathbf{Y}_D = diag(\mathbf{Y}) = diag(Y((N-1)k-(N-1)), ..., Y((N-1)k))$, we can rewrite (2.13) as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{\mathbf{X}^T\mathbf{Y}_D^H\mathbf{LL}^H\mathbf{Y}_D\mathbf{X}^*\} = \arg\min_{\mathbf{X}}\{\|\mathbf{UX}^*\|^2\} \tag{2.14}$$

which is a quadratic form in $\mathbf{X}$, where $\mathbf{U} = \mathbf{L}^H\mathbf{Y}_D$.

Thus the maximum likelihood data detection can be regarded as integer shortest vector problem [9]. The Cholesky decomposition doesn't need to be repeated provided that channel statistics and SNR are constant. The brute-force solution which is to test M^{N-1} vector $\mathbf{X}$ (where M is the size of constellation Q) is computationally infeasible in the casa of large window size (N) because the complexity grows exponentially in N. Two efficient data detection algorithms, Sphere Decoding and V-BLAST solution, will be discussed in next section.

From equation (2.14), we can see that our data detector has phase ambiguity. In other words, if $\hat{\mathbf{X}}$ is a solution for (2.14), any rotation of $\hat{\mathbf{X}}$ ($\hat{\mathbf{X}}e^{j\theta}$, $0 \leq \theta \leq 2\pi$) is another solution as well. In order to resolve this ambiguity there are 2 solutions:

2.1.1 Differential encoding of transmitted data

At the transmitter, we can apply classical M-ary differential phase shift keying (DPSK) over adjacent subcarriers. $\log_2(M)$ binary data symbols are Gray mapped to M-ary data symbols $v(k)$ taken from MPSK signal constellation $Q \equiv \left\{ v = e^{j2\pi m/M} \mid m = 0,1,..., M-1 \right\}$. From $v(k)$, the transmitted symbol on subcarrier k is obtained via differential encoding $X(k) = v(k)X(k-1)$. At the receiver, from detected vector $\hat{\mathbf{X}}$, and via differential decoding $\hat{\mathbf{v}}_k$ of the *N-1* differential symbols $\mathbf{v}_k = \left[v((N-1)k-(N-2)),...,v((N-1)k)\right]^T$ can be obtained.

2.1.2 Insertion of embedded pilot symbols

At the transmitter at least one of the subcarriers should be embedded to a known *a priori* symbol (pilot) so that the receiver can solve the ambiguity based on it. In our simulation, we will investigate both solutions.

2.1.3 V-BLAST detection

A more detailed explanation of V-BLAST has been given in section 1.4.1. In following we will give the algorithm steps briefly.

The V-BLAST algorithm is based on detecting data symbols from the strongest to the weakest. This means it needs a permutation matrix $\mathbf{\Pi}$ to rearrange the columns of matrix $\mathbf{U}$ in Eq. (2.14). Afterwards, we can do a QR decomposition on permuted matrix, $\mathbf{QR} = \mathbf{M} = \mathbf{U\Pi}$, where $\mathbf{Q}$ is unitary and $\mathbf{R} = [r_{ij}]$ is upper triangular. Since $\mathbf{Q}$ is unitary, it will be ignored in norm Eq. (2.14). Permutation matrix $\mathbf{U}$ guarantees $\mathbf{R}$ has the property that $\min\limits_{1 \le i \le N} r_{ii}$ is maximized over all column permutations. For $k = N, N-1, ..., 1$, the algorithm chooses $\pi(k)$ such that

$$\pi(k) = \arg \min_{j \notin \{\pi(1),..., \pi(k-1)\}} \left\| \left(\mathbf{G}_k\right)_j \right\|^2 \tag{2.15}$$

where $\left(\mathbf{G}_k\right)_j$ is the jth row of $\mathbf{G}_k$, $\mathbf{G}_k$ is the pseudo inverse of $\mathbf{U}_k$, and $\mathbf{U}_k$ denotes the matrix obtained by zeroing columns $\pi(1),...,\pi(k-1)$ of $\mathbf{U}$. So (2.14) can be expressed as

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \left\| \mathbf{RX}^* \right\|^2 \right\}. \tag{2.16}$$

Because $\mathbf{R}$ is upper triangular matrix, the kth element of $\mathbf{X}$ is given by

$$\hat{X}_k = \arg \min_{X} \left(r_{kk} X_k^* + \sum_{i=k+1}^{N} r_{ki} X_i^* \right)^2 \tag{2.17}$$

where r_{ki} is the *(k,i)*th entry of matrix $\mathbf{R}$. Processing in the order $X_N,...,X_1$ and assuming correct previous detection, transmitted symbols can be detected.

The original V-BLAST ordering requires N matrix inversion. In [10], a new low complexity ordering was proposed. The ordering $\pi(1),...,\pi(N)$ is obtained by sorting $\left\|\left(\mathbf{U}^+\right)_j\right\|$ in an ascending order for $j = 1,...,N$ in this technique where $\mathbf{U}^+$ is the pseudo inverse of $\mathbf{U}$. Then QR decomposition will be done on the permuted matrix and the same steps will be followed to estimate transmitted symbols. The advantage of this method is that it requires only one matrix inversion and its complexity is lower that the original V-BLAST. Simulations in [10] show that degradation in performance is not significant. In our simulation, we will compare the performance and complexity of both ordering algorithms.

2.1.4 Sphere Decoding detection

More detailed explanation of SD has been given in section 1.4.2. In following we will give the algorithm steps briefly.

Eq. (2.14) for data detection is a discrete (integer) problem as data are drawn from a specific constellation. SD is an efficient method for solving an integer problem (finding the closest lattice point in N dimensions to a given point). SD searches the lattice inside a hyper sphere of radius "*r*" instead of searching the whole lattice:

$$\left\|\mathbf{U}\mathbf{X}^*\right\|^2 \leq r^2 \qquad\qquad (2.18)$$

Because $\mathbf{U}$ is upper triangular, the key idea is to generate boundary condition for $\hat{X}_l$, $1 \leq l \leq i+1$ based on the previously detected symbols $\hat{X}_i,...,\hat{X}_N$. To see this, let u_{ij} be the entry $\mathbf{U}$ in row *i* and column *j* where $1 \leq i, j \leq N$. By defining d_{i+1} like

$$d_{i+1}^2 = \sum_{l=i+1}^{N} \left| \sum_{j=l}^{N} u_{lj} X_j^* \right|^2 \qquad (2.19)$$

SD is the method to estimate the transmitted symbols that progressively satisfy:

$$d_i^2 = \left| u_{ii} X_j^* + \sum_{j=i+1}^{N} u_{ij} X_j^* \right|^2 + d_{i+1}^2 \le r^2 \qquad (2.20)$$

Radius r can be dynamically updated whenever one lattice point found.

Although SD reduces the complexity significantly comparing to brute-force search, still the complexity might be high for the case of big window size N.

It can be seen that the brute-force search is like searching in a tree with N Node and M branch in each Node (Figure 2-1). We can split this tree to several sub-trees of the length w ($w<N$), at the expense of degradation in performance (Figure 2-2). This way, we actually split search space to several smaller ones with lower complexity. Then we can use SD in each of sub-spaces fixing the detected data symbols from previous sub-spaces [11]. In other words, in each step it can be supposed that symbols detected using SD in previous steps are correct and can be used as the initial state for the next subspaces. Thus the complexity will decrease dramatically. The smaller the size of the sub-tree, the lower the complexity and the higher the degradation. In our simulation, we will compare the performance and complexity of both algorithms.
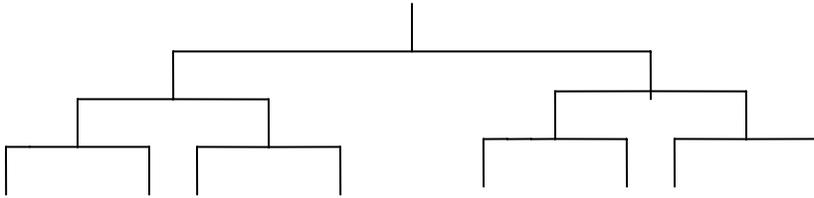
Figure 2-1: Search tree, *M=2*, *N=3*
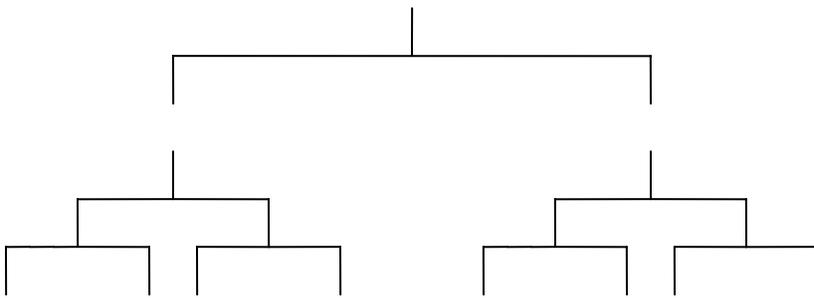
Figure 2-2: Splitting search tree, *M=2*, *N=3, w_1=1,w_2=2*

2.2 Multi Symbol Differential detection in OFDM

Conventional differential detection for OFDM suppose that CIR is constant for 2 consecutive OFDM blocks and differential encoding and decoding are considered for data on the same subcarriers but in adjacent blocks. In case of large DFT length and in wireless system, this assumption can't be true anymore.

In next section, we will show that when the CIR length is much smaller than DFT length, channel frequency response for consecutive subcarriers are highly correlated and can be considered constant so that the differential encoding-decoding can be done over adjacent subcarriers in one OFDM block. In section 2.2.2, we will look at the maximum likelihood detection of MPSK for differentially encoded OFDM symbols.

2.2.1 Channel frequency response autocorrelation

Using the channel model $h(t) = \sum_{l=0}^{L-1} h_l \delta(t - \tau_l)$, the channel frequency response on tone k

becomes

$$H(k) = \sum_{l=0}^{L-1} h_l e^{-j2\pi(k/N)\tau_l} . \tag{2.21}$$

Assuming τ_l's are independent, the correlation matrix for frequency response vector $\mathbf{H}$,

$\mathbf{R_{HH}} = E\{\mathbf{HH}^H\} = [r_{m,n}]$ can be expressed as

$$r_{m,n} = \int \ldots \int \prod_{k=0}^{L-1} f_{\tau_k}(\tau_k) \left[\sum_{i=0}^{L-1} r_{hh}(\tau_i) e^{-j2\pi\tau_i(m-n)/N} \right] d\tau_0 \ldots d\tau_{L-1} =$$
$$\sum_{i=0}^{L-1} \int f_{\tau_i}(\tau_i) r_{hh}(\tau_i) e^{-j2\pi\tau_i(m-n)/N} d\tau_i \tag{2.22}$$

where $r_{hh}(\tau)$ is the multipath power profile and $f_{\tau_k}(\tau_k)$ is the pdf of τ_k. For the case of

uniform delay profile

$$f_{\tau_k}(\tau_k) = \begin{cases} 1/L & \text{if } \tau_k \in [0, L-1] \\ 0 & \text{otherwise} \end{cases} \quad k = 0, 1 \ldots, L-1 \tag{2.23}$$

and exponential decaying power profile $r_{hh}(\tau) = Ce^{-\tau/\tau_{rms}}$, so by substituting in (2.22)

$$r_{m,n} = \frac{C}{L} \frac{1 - e^{-(L-1)((1/\tau_{rms}) + 2\pi j(m-n)/N)}}{(\dfrac{1}{\tau_{rms}} + j2\pi\dfrac{m-n}{N})} \tag{2.24}$$

and normalizing $r_{k,k}$ to unity gives us,

$$r_{m,n} = \frac{1 - e^{-L((1/\tau_{rms}) + 2\pi j(m-n)/N)}}{\tau_{rms}(1 - e^{-(L/\tau_{rms})})(\dfrac{1}{\tau_{rms}} + j2\pi \dfrac{m-n}{N})} . \tag{2.25}$$

Figure 2-3 shows the magnitude of $r_{m,n}$ for different values of *N* for a typical channel profile of

indoor office area as is modeled in [12].The rms delay spread of this channel is 35ns and the

maximum delay spread is 100ns, assuming sampling period of 20ns, $L = 5$, $\tau_{rms} \approx 2$. The figure

shows for small values of *m-n*, correlation is high and as *N* increases; the range of *m-n* with

highly correlated channel frequency response will increase as well. We will use this

characteristic to assume constant value of channel frequency response for adjacent subcarriers.

2.2.2 Maximum Likelihood Multiple symbol Differential detection in OFDM

Consider the transmission of MPSK signals for subcarriers in an OFDM block,

$v(k) \in Q$, $Q \equiv \left\{ v = e^{j2\pi m/M} \mid m = 0,1,...,M-1 \right\}$. From $v(k)$, the transmitted symbol on subcarrier

k is obtained via differential encoding $X(k) = v(k)X(k-1)$. The corresponding received signal

is then $Y(k) = H(k)X(k) + W(k)$, $0 \le k \le N-1$. Now consider a received sequence of length

P+1 and assume that $H(k)$ is independent of *k* over the length of this sequence, i.e., $H[k] = H$.

Now the received sequence $\mathbf{Y} = [Y(k), Y(k+1),..., Y(k+P)]^T$ is expressed as

$$\mathbf{Y} = \mathbf{X}H + \mathbf{W} \tag{2.26}$$

where $\mathbf{X} = [X(k), X(k+1),..., X(k+P)]^T$ and $\mathbf{W} = [W(k), W(k+1),..., W(k+P)]^T$.
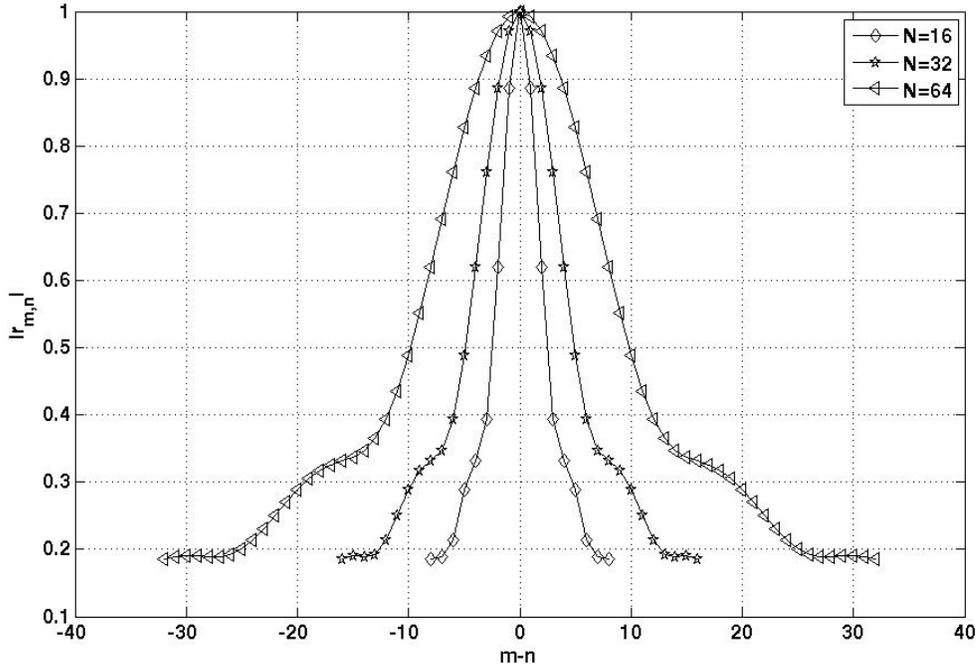
Figure 2-3: Amplitude of channel frequency response correlation matrix's elements for different values of N **when** $L = 5, \tau_{rms} = 2$

Following the same model for channel and noise as given in previous section [13],

$$f(\mathbf{Y} \mid \mathbf{X}, H) = \frac{1}{(\pi\sigma_w)^P} \exp(-\frac{\|\mathbf{Y} - \mathbf{X}H\|^2}{\sigma_w^2}). \tag{2.27}$$

To maximize this likelihood function, H should be selected such that

$$\hat{H} = \arg\min_H \|\mathbf{Y} - \mathbf{X}H\|^2 = \arg\min_H \sum_{p=0}^{P} |Y[k + p] - HX[k + p]|^2. \tag{2.28}$$

Since the constellation is unitary,

$$\hat{H} = \arg\min_{H} \sum_{p=0}^{P} |Y[k+p] - HX[k+p]|^2 =$$

$$\arg\min_{H} [-H \sum_{p=0}^{P} X[k+p]Y^*[k+p] - H^* \sum_{p=0}^{P} X^*[k+p]Y[k+p]] = \qquad (2.29)$$

$$\arg\min_{H} \left| \sum_{p=0}^{P} X^*[k+p]Y[k+p] - H \right|^2$$

Then $\hat{H} = \sum_{p=0}^{P} X^*[k+p]Y[k+p]$. Replacing $\hat{H}$ from (2.29) into (2.28), our cost function for

data detection is

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\| \mathbf{Y} - \mathbf{X}\hat{H} \right\|^2,$$

$$\left\| \mathbf{Y} - \mathbf{X}\mathbf{X}^H \mathbf{Y} \right\|^2 = \left\| (\mathbf{I} - \mathbf{X}\mathbf{X}^H)\mathbf{Y} \right\|^2 = \mathbf{Y}^H(\mathbf{I} - \mathbf{X}\mathbf{X}^H)(\mathbf{I} - \mathbf{X}\mathbf{X}^H)\mathbf{Y} = \mathbf{Y}^H\mathbf{Y} + (P-1)\mathbf{Y}^H\mathbf{X}\mathbf{X}^H\mathbf{Y}$$

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} \mathbf{Y}^H\mathbf{X}\mathbf{X}^H\mathbf{Y} = \arg\max_{\mathbf{X}} \left\| \mathbf{X}^H\mathbf{Y} \right\|^2$$

$$(2.30)$$

In (2.30), we have used the fact that $\mathbf{X}^H\mathbf{X} = (P+1)$.

Note that this decision rule has a phase ambiguity associated with it. Since the transmitted data has been differentially encoded, this ambiguity won't affect decoding of $\mathbf{v} = [v[k+1],...,v[k+P]]^T$. In [9], BER analysis for BPSK case has been done and it has been shown that the performance of multiple symbol differentially detected BPSK approaches that of ideal coherent detection BPSK with differential encoding in the limit as the observation interval P approaches infinity provided that channel response won't change during this interval. But in the developed model, because we can assume constant frequency response for limited number of subcarriers, we can not consider a very large observation interval P.

2.3 Simulation and results

Simulation results are given for the proposed data detectors. In simulation, an OFDM system has been developed over a frequency selective Rayleigh fading channel with 4 independent complex Gaussian taps with exponential power decay. The DFT length is considered to be 32 and data are from BPSK constellation. The length of CP is $L+1$ where L is the maximum length of CIR. CIR is constant over each block of OFDM data but can be varied from one block to another.

The SNR axis is in dB scale in our figures and explains the relation of the power in received signal to the power of noise. In the first simulation, differential coding has been used over data to solve the ambiguity. Figure 2-4 compares the performance of different data detectors.
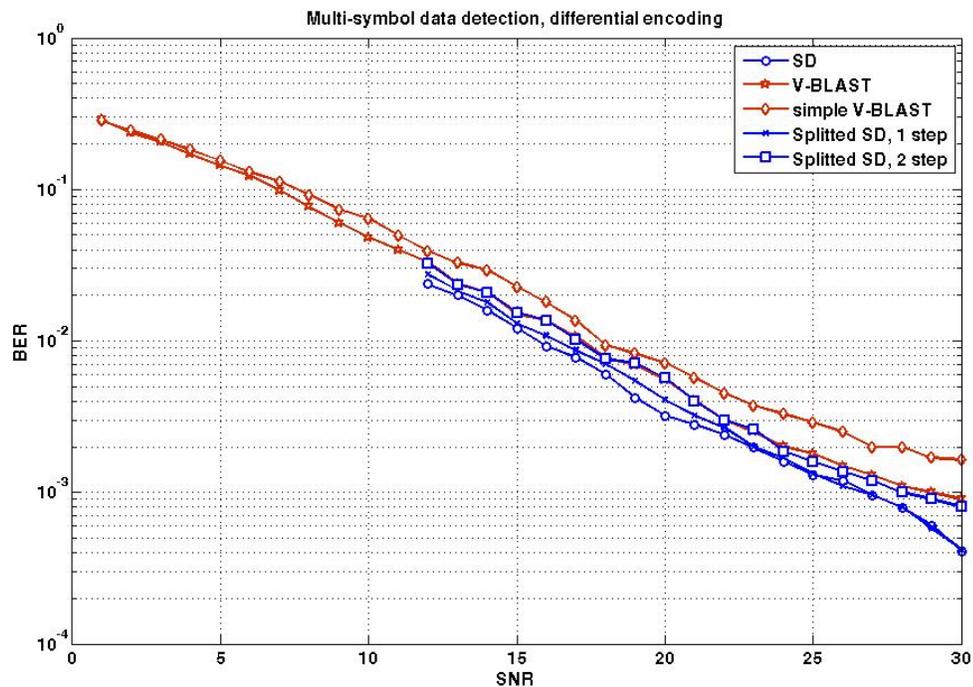
Figure 2-4: BER performance comparison of different multi symbol data detectors with differential coding for ambiguity solution

SD algorithm needs the initial radius to start the search with. We choose a relaxation approach to choose the initial radius. For OFDM symbol from MPSK, we relax (2.14) as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{\|\mathbf{U}\mathbf{X}^*\|^2\} = \arg\min_{\mathbf{X}}\{\mathbf{X}^T\mathbf{U}^H\mathbf{U}\mathbf{X}^*\}, \mathbf{X}^T\mathbf{X}^* = N \quad \text{where the vector} \quad \mathbf{X} \in C^N.$$ The

Lagrangian for this minimization problem is $l(\mathbf{X}, \lambda) = \mathbf{X}^T\mathbf{U}^H\mathbf{U}\mathbf{X}^* + \lambda(\mathbf{X}^T\mathbf{X}^* - N)$. The optimal λ here is the minimum eigenvalue of matrix $\mathbf{U}^H\mathbf{U}$ and $\tilde{\mathbf{X}}$ is the eigenvector corresponding to λ. We then quantize $\tilde{\mathbf{X}}$ into a point in Q^N as $\hat{\mathbf{X}}$. By substituting $\hat{\mathbf{X}}$ into (2.14), the initial radius is given by $r^2 = \hat{\mathbf{X}}^T\mathbf{U}^H\mathbf{U}\hat{\mathbf{X}}^*$. In Figure 2-4, the performance of SD when we split the search space to two subspace (*w=16*) has been shown. In high SNR, it performs very close to the original SD while in lower SNR the degradation can be up to 1 dB comparing to the original SD. Also, we simulated SD with two steps of splitting the search space (*w=8*).
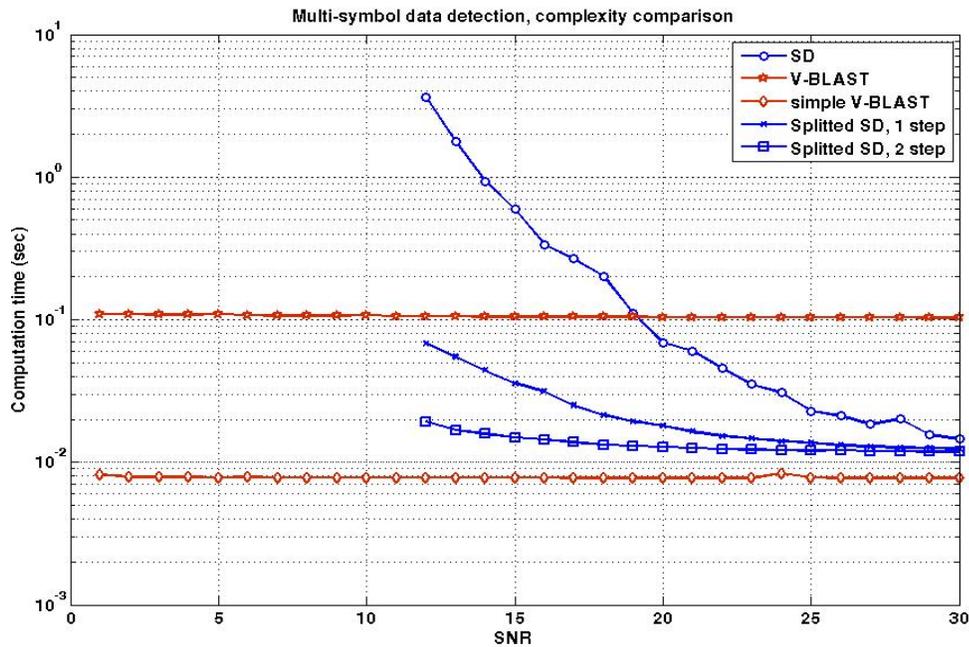
Figure 2-5: Complexity comparison between different algorithms proposed for multi symbol data detection

In low SNR, its performance is quite close to V-BLAST but in high SNR it works almost 1 dB better than V-BLAST and it shows almost 3dB degradation comparing to one step splitting SD the original SD. For V-BLAST algorithm, we have shown the BER performance of the original one and the comparison with the one in [10] (labelled simple V-BLAST). The simple V-BLAST works at most 2 dB worse than the original V-BLAST. Figure 2-5 compares the complexity (computation time) of different data detection algorithms. It shows there is a trade off between the complexity and accuracy. At the expense of degradation in BER, splitting SD algorithms suggest significant reduction in complexity (especially in lower SNR). As we can see, the complexity is not a function of SNR for V-BLAST algorithms.
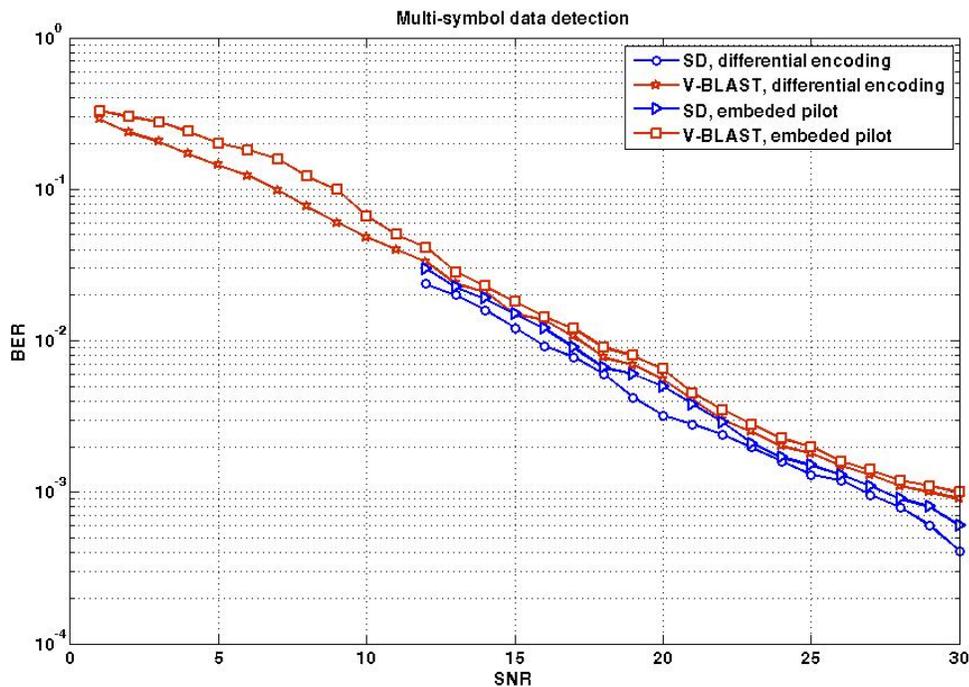
Figure 2-6: BER performance comparison of different algorithms and different ambiguity solutions

In low SNR regime, original SD has more complexity than VBLAST. The simplified V-BLAST has the minimum complexity which is significantly less that original V-BLAST algorithm. In fact the complexity of SD is a function of the size of constellation. For BPSK case, form Figure 2-5, we can see in high SNR the complexity of SD is even less than V-BLAST but for higher constellation size (like 4PSK in [8]), SD complexity can be always more than V-BLAST. Figure 2-6 compare the performance of 2 different ambiguity solutions: embedded pilot and differential coding. To make the comparison fair, there is one subcarrier dedicated to pilot in each OFDM block. No matter what solution has been applied to solve the ambiguity, SD outperforms V-BLAST in all SNR range. Simulation results in Figure 2-6 suggest that differential coding solution outperforms embedded pilot solution. The reason is that the error in ambiguity detection will not propagate in differential solution.
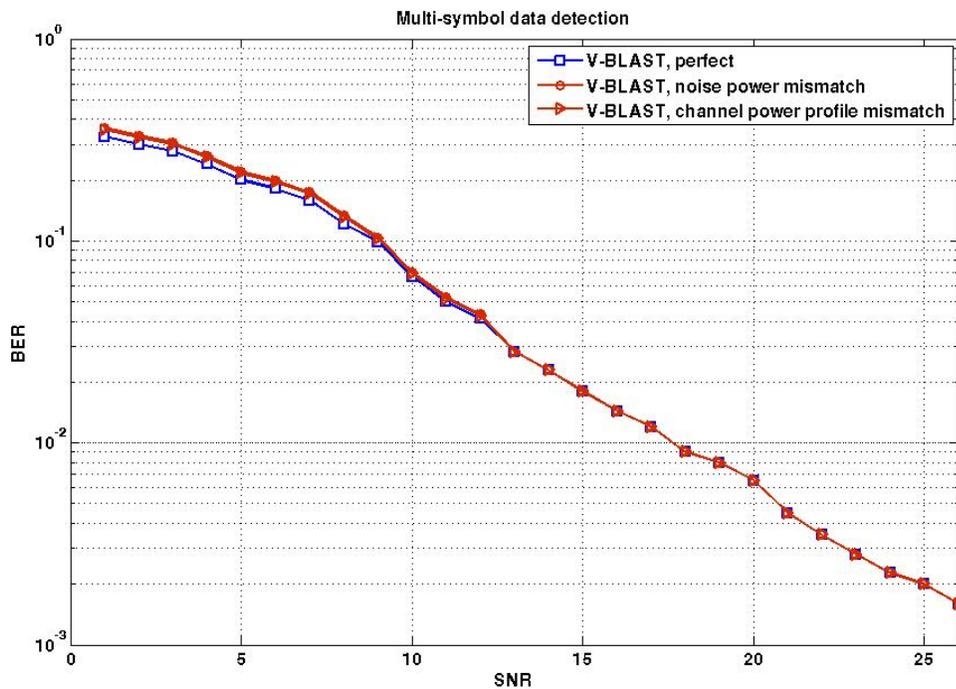
Figure 2-7: The effect of mismatch of channel power profile and noise power in receiver

As Eq. (2.10) shows, the proposed data detection algorithms require the receiver to know the channel power profile and noise power. However simulation results in Figure 2-7 show that the proposed data detectors are robust to mismatch in both power profile and noise power. We simulated the detector with perfect knowledge of channel power profile and noise power. In the other simulation the noise power has been fixed to 20 dB SNR value, and we can see the degradation in performance is insignificant. Since in low SNR the error caused from the mismatch will be concealed in noise, the receiver should pick a noise power value related to high SNR in case it is not aware of the real value of noise power. In the other simulation, receiver uses a unitary power profile instead of exponential decay to investigate the effect of channel power profile mismatch. Again, the data detector performs robust to this mismatch.
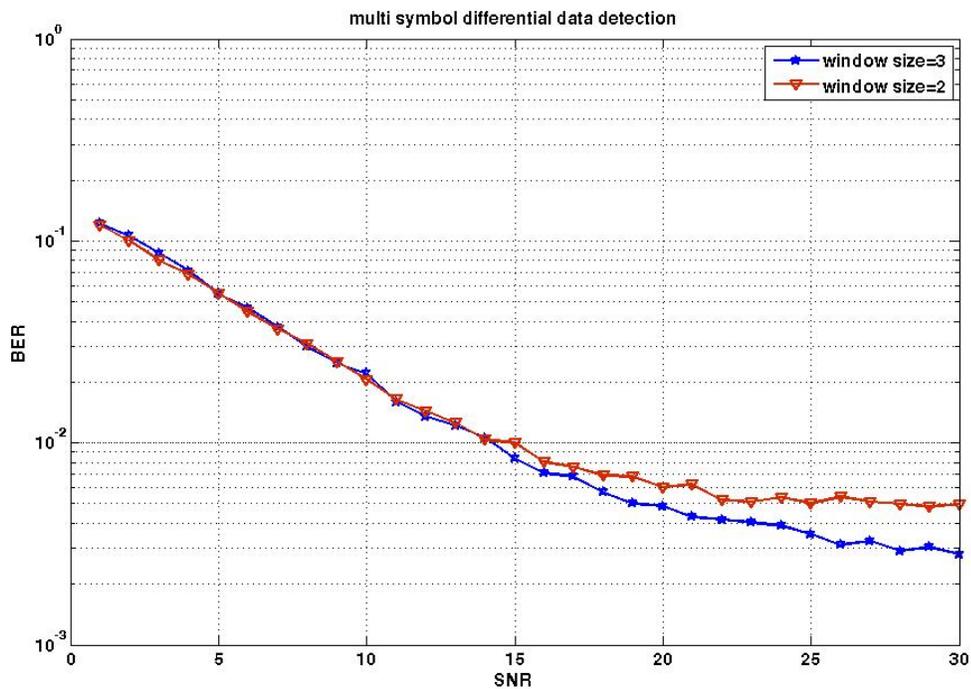
Figure 2-8: Multi symbol differential data detection BER performance with different window size

In Figure 2-8, we have compared the performance of multi symbol differential detector proposed in 2.2.2 for two different window sizes. Results show that multi symbol differential data detection over three consequent subcarriers outperforms the one over two. Although, as it has been mentioned before, we can't make the window size large, since the channel frequency response may not stay constant over a large window.

2.4 Conclusion

We have investigated two ML multi symbol data detector for OFDM systems. For the first data detector, we suggested two different solutions to resolve the phase ambiguity. As the cost function for this data detector is an integer quadratic, we have used V-BLAST and SD algorithm to solve it. We also used less complexity version of V-BLAST and SD algorithms. We investigated the performance and computation complexity of different algorithms. Because of the trade of between accuracy and complexity, based on the application somebody can pick a specific algorithm. Although the detection algorithm requires the receiver to know the channel power profile and noise power, simulation shows that it is robust to mismatch. We also derived a differential multi symbol detector and compared the performance of the detector for different window size.

Chapter Three: Blind channel estimation for an SIMO OFDM system

This chapter develops a channel estimator for a Single Input Multi Output (SIMO) OFDM system. The chapter organization is as follows. Section 3.1 develops the base band system model of an OFDM system and afterwards it discusses a SIMO OFDM model. In Section 3.2, we will derive the blind channel estimator for SIMO system. Section 3.3 gives identifiability conditions of estimator. Section 3.4 gives the numerical results and Section 3.5 concludes the chapter.

3.1 OFDM SISO base band model

In an OFDM system, the binary source data are mapped into symbols from a constellation Q and then grouped to blocks of length N. Each group is modulated by inverse discrete Fourier transform (IDFT) on N parallel subcarriers. The resulting time domain samples are

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j(2\pi kn/N)}, \quad n = 0, 1, ..., N-1 \tag{3.1}$$

where

$$X_k = \begin{cases} d_k & k \in I_d \\ p_k & k \in I_p \\ 0 & k \in I_v \end{cases} \tag{3.2}$$

I_d is the index set of data subcarriers with N_d elements, I_p is the index set of subcarriers reserved for pilot symbols with N_p elements, and I_v is the index set of virtual carriers with N_v elements and $N_d + N_p + N_v = N$. The symbols $X_k, k = 0, 1, ..., N-1$ are called OFDM input symbols.

The symbols after IDFT are denoted as $x_n, n = 0,1,...,N-1$. The term "OFDM block" denotes the entire IDFT output $\{x_0, x_1,...,x_{N-1}\}$. The input symbol duration is T_s and the OFDM block duration is NT_s. One transmitted OFDM block duration usually consists of a regular symbol interval and a guard interval (GI). Assuming that GI length is greater than or equal to the channel impulse response (CIR) length, Inter-symbol Interference (ISI) is completely eliminated and the orthogonality between the subcarriers can be maintained [14].

There are different alternatives for the GI. In Cyclic prefix OFDM (CP-OFDM), a copy of the last part of OFDM block is pre-appended to the transmitted block as the GI. In Zero padding OFDM (ZP-OFDM), zeros are pre-appended to the transmitted block as GI. In [14] it has been shown that ZP-OFDM can be converted to CP-OFDM at the receiver. It also shows that insertion of CP can convert linear convolution of data and CIR to circular convolution.

Time domain transmitted samples are appropriately pulse shaped to construct the time domain signal *x(t)* for transmission. Typically pilots $X_k, k \in I_p$ known *a priori* at the receiver, remain fixed from one OFDM block to the next one. Pilot arrangement can be such that $N_p <<$ *N*. Alternatively, entire OFDM block can be pilots and be transmitted periodically. In Semi Blind estimation and detection techniques, the former is studied.

It has been assumed that the composite CIR which includes transmit and receive pulse shaping and the physical channel response between the transmitter and receiver can be modeled as

$$h(\tau) = \sum_{l=0}^{L-1} h_l \delta(\tau - \tau_l) \tag{3.3}$$

where $h_l \sim CN(0, \sigma_l^2)$, τ_l is the delay of the lth tap and L is the total number of paths. The $[\sigma_0^2,...,\sigma_{L-1}^2]$ and $[\tau_0,...,\tau_{L-1}]$ constitute the power delay profile (PDP). The received signal after sampling is given by

$$y_n = \sum_{l=0}^{L-1} h_l x_{n-d_l} + w_n \tag{3.4}$$

where $w_n \sim CN(0, \sigma_w^2)$ is an additive white Gaussian noise (AWGN) and $d_l = \lfloor \tau_l / T_s \rfloor$ is the delay normalized by T_s. We assumed that d_l's are integers otherwise there is a leakage of energy to all the channel frequency taps [15][16]. We assume perfect synchronization and that channel remains constant during each OFDM block, but it might varies between OFDM blocks.

After removing the guard interval and taking the discrete Fourier transform (DFT) (N points) on the remaining samples at the receiver, the post-DFT received samples Y_k can be expressed as follows:

$$Y_k = H_k X_k + W_k, \quad 0 \leq k \leq N-1 \tag{3.5}$$

where $H_k = H(j2\pi k / N)$ is the complex channel frequency response at subcarrier k, $H(jw)$ is the Fourier transform of the CIR. W_k, $0 \leq k \leq N-1$ are the Fourier transform of w_n and are independent and identically distributed (i.i.d) complex Gaussian random variables (CGRV), each of which also has zero mean and variance σ_w^2. Assuming $\tau_l = lT_s$, we find $\mathbf{H} = \mathbf{F}_L \mathbf{h}$ where $\mathbf{H} = [H_0, H_1, ..., H_{N-1}]^T$, $\mathbf{h} = [h_0, h_1, ..., h_{L-1}]^T \in C^L$ is the CIR, $\mathbf{F}$ is the DFT matrix

$$[\mathbf{F}]_{k,l} = \frac{1}{\sqrt{N}} e^{-j(2\pi/N)kl}, k,l \in 0,1,...,N-1),\ \mathbf{F}_L \text{ is a } N{\times}L \text{ submatrix (first L columns) of}$$

DFT matrix $\mathbf{F}$ scaled by $\sqrt{N}$. We can vectorize (3.5) as

$$\mathbf{Y} = \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} \tag{3.6}$$

where $\mathbf{X}_D = diag(X_0, X_1,..., X_{N-1})$ is a diagonal matrix. Figure 1-2 shows the principal blocks in a baseband OFDM system.

3.2 Blind LS channel estimator for SIMO OFDM system

We will derive a lease square (LS) blind channel estimator for a single input multi output (SIMO) OFDM system [17]. It will basically take the advantage of the correlation between the received data at different receivers. It exploits the cross relation between each channel output pair, which is the basis of the approach in [18] and extends the idea to the OFDM systems. The channel estimator we will develop is not only for the case of spatial diversity [19]; it can be also used for an OFDM system with time diversity. We will concentrate on the former case though.

In a SIMO system, transmitted data are passed through different channels, which are supposed to be independent and uncorrelated, and are received at the different receivers. Suppose we have m_r receivers (Figure 3-1), and the maximum CIR length of the m_r channels is L. in other words, if we define the CIR length of ith channel by L_i, we know $\max(L_i) = L, 1 \le i \le m_r$.
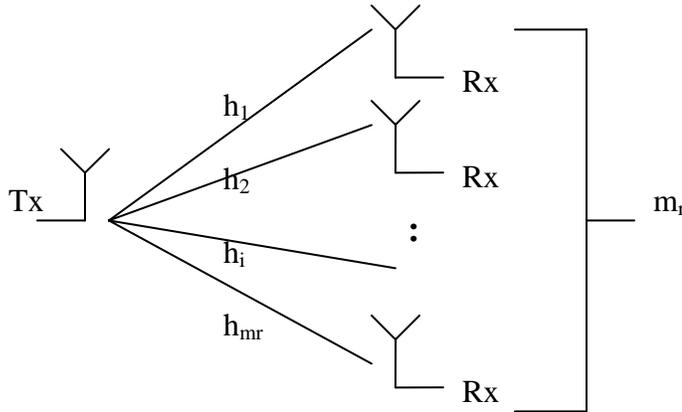
Figure 3-1: SIMO system with m_r receivers

The equation (3.6) can be written for each channel

$$\mathbf{Y}_m = \mathbf{X}_D \mathbf{F}_L \mathbf{h}_m + \mathbf{W}_m \qquad 1 \le m \le m_r \qquad (3.7)$$

where $\mathbf{h}_m$ is the CIR of channel m and $\mathbf{Y}_m$ is the post-DFT received data at the receiver m. In the noise free case, we have

$$Y_i(k) = X(k) H_i(k), \quad 1 \le i \le m_r \qquad (3.8)$$

where $X(k)$ is the transmitted symbol at the subcarrier k and $H_i(k)$ is the frequency response of the ith channel at the subcarrier k and $Y_i(k)$ is the post-DFT received data at the subcarrier k at the ith receiver. It is easy to show that in noise free case

$$Y_i(k) H_j(k) = Y_j(k) H_i(k) \quad i \ne j, 1 \le i, j \le m_r . \qquad (3.9)$$

Defining $\mathbf{Y}_{iD} = diag(\mathbf{Y}_i)$ and using equation (3.9), we can write

$$\mathbf{Y}_{iD} \mathbf{F}_L \mathbf{h}_j = \mathbf{Y}_{jD} \mathbf{F}_L \mathbf{h}_i \quad i \ne j, 1 \le i, j \le m_r , \qquad (3.10)$$

if we replace $\mathbf{Y}_{iD} \mathbf{F}_L$ with $\tilde{\mathbf{Y}}_i$, last equation can be rearranged to

$$\tilde{\mathbf{Y}}_i \mathbf{h}_j - \tilde{\mathbf{Y}}_j \mathbf{h}_i = 0 \qquad i \neq j, 1 \leq i, j \leq m_r . \qquad (3.11)$$

For a fixed value of i, if we write (3.11) for all possible values of j; considering $\mathbf{h} = \left[\mathbf{h}_1^T, ..., \mathbf{h}_{m_r}^T \right]^T$ and denoting

$$\tilde{\mathbf{Y}}^i = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{Y}}_{i+1} & -\tilde{\mathbf{Y}}_i & \mathbf{0} & & \mathbf{0} \\ & \vdots & & \vdots & & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{Y}}_{m_r} & \mathbf{0} & \cdots & & -\tilde{\mathbf{Y}}_i \\ \underbrace{\phantom{\mathbf{0} \cdots \mathbf{0}}}_{i-1 \quad blocks} & & & \underbrace{\phantom{\tilde{\mathbf{Y}}_{m_r} \quad \mathbf{0} \quad \cdots \quad -\tilde{\mathbf{Y}}_i}}_{m_r - i + 1 \quad blocks} \end{bmatrix} \qquad (3.12)$$

and when all received data are taken into consideration,

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{Y}}^1 \\ \vdots \\ \tilde{\mathbf{Y}}^{m_r - 1} \end{bmatrix} \qquad (3.13)$$

in the absence of noise, we construct the following equation :

$$\tilde{\mathbf{Y}} \mathbf{h} = \mathbf{0} . \qquad (3.14)$$

Hence the channel coefficient vector $\mathbf{h}$ can be identified by solving (3.14). In other words, $\mathbf{h}$ is in the null space of $\tilde{\mathbf{Y}}$.

As with any deterministic blind identification method, in order to ensure the identifiability, it is necessary to impose conditions upon the channels and the source data. In the following, we will show that the condition about the channel is that there is no common zero among all the channels, where a zero of the channel m is defined by $H_m(z_0) = \sum_{n=0}^{L-1} h_m(n) z_0^{-n} = 0$. The condition about the source data is that none of the data on different subcarriers are zero.

3.3 Identifiability of blind SIMO estimator

Matrix $\tilde{\mathbf{Y}}_i$ can also be expressed by

$$\tilde{\mathbf{Y}}_i = \mathbf{X}_D \mathbf{H}_{iD} \mathbf{F}_L \qquad\qquad 1 \le i \le m_r \qquad (3.15)$$

where $\mathbf{H}_{iD} = diag\left(H_{i,0}, H_{i,1}, ..., H_{i,N-1}\right)$.

Thus, if we consider $\tilde{\mathbf{H}}_i = \mathbf{H}_{iD}\mathbf{F}_L$, then

$$\tilde{\mathbf{Y}}^i = \hat{\mathbf{X}}_D \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{H}}_{i+1} & -\tilde{\mathbf{H}}_i & \mathbf{0} & \mathbf{0} \\ & \vdots & & \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{H}}_{m_r} & \mathbf{0} & \cdots & -\tilde{\mathbf{H}}_i \end{bmatrix} = \hat{\mathbf{X}}_D \tilde{\mathbf{H}}^i \qquad (3.16)$$

$$\underbrace{}_{i-1 \quad blocks} \qquad \underbrace{}_{m_r - i + 1 \quad blocks}$$

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{Y}}^1 \\ \vdots \\ \tilde{\mathbf{Y}}^{m_r - 1} \end{bmatrix} = \tilde{\mathbf{X}}_D \begin{bmatrix} \tilde{\mathbf{H}}^1 \\ \vdots \\ \tilde{\mathbf{H}}^{m_r - 1} \end{bmatrix} = \tilde{\mathbf{X}}_D \tilde{\mathbf{H}} \qquad (3.17)$$

where $\tilde{\mathbf{X}}_D = diag\,(\mathbf{X}_D, ..., \mathbf{X}_D) \in C^{(Nm_r(m_r-1)/2) \times (Nm_r(m_r-1)/2)}$.

Now the identification equation (3.14) can be rewritten as

$$\tilde{\mathbf{Y}}\,\mathbf{h} = \tilde{\mathbf{X}}_D \tilde{\mathbf{H}}\,\hat{\mathbf{h}} = \mathbf{0} \qquad (3.18)$$

where $\hat{\mathbf{h}}$ is the nontrivial solution to $\tilde{\mathbf{Y}}\,\mathbf{h} = \mathbf{0}$. If $X(k) \ne 0, k = 0,1,...,N-1$, then $\mathbf{X}_D, \tilde{\mathbf{X}}_D$ are non-singular and it follows

$$\tilde{\mathbf{H}}\,\hat{\mathbf{h}} = \mathbf{0} \qquad (3.19)$$

which means

$$\tilde{\mathbf{H}}_i \hat{\mathbf{h}}_j - \tilde{\mathbf{H}}_j \hat{\mathbf{h}}_i = 0 \qquad\qquad i \ne j, 1 \le i, j \le m_r. \qquad (3.20)$$

By expanding the last equation we will have

$$H_i(k)\hat{H}_j(k) = H_j(k)\hat{H}_i(k) \quad , \quad k = 0,1,...,N-1 \qquad (3.21)$$

which means $H_i(z)\hat{H}_j(z) = H_j(z)\hat{H}_i(z)$. Because there is no common zero among all the

channels, it means for any value of z_0, if $H_i(z_0) = 0$ then $\hat{H}_i(z_0) = 0$. So if $H_i(z)$ is of order

L, $\hat{H}_i(z)$ is also of order L, then $H_i(z) = \alpha\hat{H}_i(z)$, where α is a nonzero complex constant. It

yields that $\mathbf{h}_i = \alpha\hat{\mathbf{h}}_i$, $i = 1,...,m_r$. We can conclude:

The nontrivial solution $\hat{\mathbf{h}}$ to $\tilde{\mathbf{Y}}\hat{\mathbf{h}} = \mathbf{0}$ uniquely (up to a complex scalar constant) determines the

channel impulse response $\{\mathbf{h}_i\}_{i=1}^{m_r}$, if there is no common zero among all the channels and none of

the data on different subcarriers are zero.

In the presence of noise, when the received data are corrupted by noise, the channel

estimate $\hat{\mathbf{h}}$ is obtained by solving the LS problem of

$$\hat{\mathbf{h}} = \arg\min_{\|h\|^2 = 1} \left\|\tilde{\mathbf{Y}}\mathbf{h}\right\|^2 \qquad (3.22)$$

The above formula is a quadratic form in $\mathbf{h}$ and is the basis of our channel estimator. $\hat{\mathbf{h}}$ can be

computed as the eigenvector of corresponding to the minimum eigenvalue of [20].

Remarks:

▪   It is well known that blind techniques for channel identifications inherit the phase

ambiguity (channel can be identified uniquely up to a complex nonzero constant). In order to

solve phase ambiguity, two different algorithms will be used. They will be illustrated in simulation results section.

■   It has been mentioned that identifiability condition requires the transmitted data not to be zero on all subcarriers. Virtual carriers (VC) have been used in application (IEEE 802.11a [21] standards) for the purpose of mitigation of inter block interference (limiting the transmit spectrum) and frequency offset synchronization. It means that in presence of VC, just N_d+N_p subcarriers are carrying data and zero will be allocated to the rest of subcarriers. In that case, our estimator should use just N_d+N_p non-zero subcarriers to ensure identifiability and N should be replaced by N_d+N_p in our formulation.

■   Estimated   channel $\hat{\mathbf{h}}$ can   be   used   for   data   detection.   Defining

$$\hat{\mathbf{H}}_i = \left[\hat{H}_{i,0}, \hat{H}_{i,1}, ..., \hat{H}_{i,N-1}\right]^T, \qquad \hat{\mathbf{H}} = \begin{bmatrix} \hat{\mathbf{H}}_1^T \\ \vdots \\ \hat{\mathbf{H}}_{m_r}^T \end{bmatrix}, \qquad \mathbf{Y}_i = \left[Y_{i,0}, Y_{i,1}, ..., Y_{i,N-1}\right]^T, \qquad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1^T \\ \vdots \\ \mathbf{Y}_{m_r}^T \end{bmatrix} \quad \text{and}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1^T \\ \vdots \\ \mathbf{W}_{m_r}^T \end{bmatrix}, \text{ we have}$$

$$\mathbf{Y} = \hat{\mathbf{H}}\mathbf{X}_D + \mathbf{W} . \tag{3.23}$$

Therefore Maximum ratio combining (MRC) algorithm can be used to detect the transmitted data [19].

3.4 Simulation and results

Simulation results are given for the proposed channel estimator. In simulation, an SIMO OFDM system with 4 receive antennas ($m_r = 4$) has been developed and frequency selective static

channels have been considered. An OFDM system with 16 subcarriers and BPSK is simulated. The length of CP is *L+2* where *L* is the maximum length of CIRs and it is considered 4 in our simulation. CIR is constant over each block of OFDM data but can be varied from one block to another. The performance of equalizer with perfect knowledge of the CIR provides the benchmark. Figure 3-2 shows the mean square error (MSE) of channel estimation which is

defined as $\qquad MSE\ (dB) = 10\log_{10}(\frac{1}{R}\sum_{i=1}^{R}\left\|\hat{\mathbf{h}}_i - \mathbf{h}\right\|^2) \qquad$ (3.24)

where *R* is the number of Monte Carlo runs. In this simulation, we used static channels with channel coefficients shown in Table 3-1. There is no common zero among the channels so the identifiability condition for the channels is satisfied. Any sort of random coefficients (non- LOS scenario) can be used for CIR of different channels as long as they have no common zero among them.

	$h_1(n)$	$h_2(n)$	$h_3(n)$	$h_4(n)$
$n = 0$	-0.049+0.395i	0.443-0.0364i	-0.211-0.322i	0.417+0.030i
$n = 1$	0.482-0.569i	1	-0.199+0.918i	1
$n = 2$	-0.556+0.578i	0.921-0.194i	1	0.873+0.145i
$n = 3$	1	0.189-0.208i	-0.284-0.524i	0.285+0.309i
$n = 4$	-0.171+0.061i	-0.087-0.054i	0.136-0.19i	-0.049+0.16i

Table 3-1: CIR (channel coefficients) of 4 channels

In order to solve phase ambiguity of the blind estimator we used one embedded pilot. Using this pilot, the ambiguity is estimated and CIR is adjusted. In order to improve the channel estimation, we used the estimated channel to detect the data and detected data were iterated back to estimate the channel (Decision Directed algorithm). Simulation shows one step of iteration improves estimation by almost 10dB.

The second algorithm for solving the phase ambiguity is to modulate the adjacent subcarriers in an OFDM symbol differentially. This way, as long as data detection is of concern, the amount of rotation in channel estimation will not affect data detection. It worth mentioning that data detection is still coherent and the differential coding of data is just to combat the phase ambiguity. Figure 3-3 shows BER performance of SIMO OFDM system. In this figure the performance of two different phase ambiguity solutions have been compared with that of the benchmark. In order to make the comparison fair, just one pilot has been inserted in an OFDM symbol for the case of pilot embedded solution (so that both ambiguity solutions use the same amount of bandwidth).
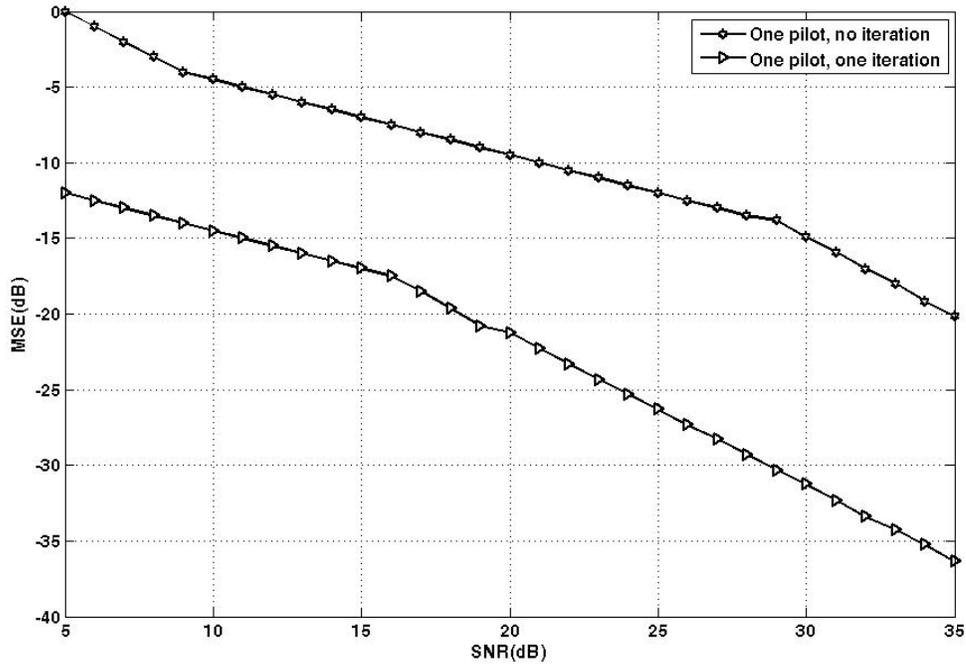
Figure 3-2: MSE of blind channel estimator

It can be seen that differential solution outperforms pilot solution in low SNR range. It is because in low SNR, just one pilot may not be able to compensate the phase ambiguity completely and the error will propagate through the OFDM symbols while in differential case, the error will not propagate through the whole OFDM symbols and is limited to the two consequent subcarriers. Figure 3-4 shows the BER performance of the estimator in the case that only 12 of 16 subcarriers have used for data and the rest have been allocated to VCs. It can be seen that in presence of VC, detection performance degrades because less amount of data are available to be used for channel estimation.
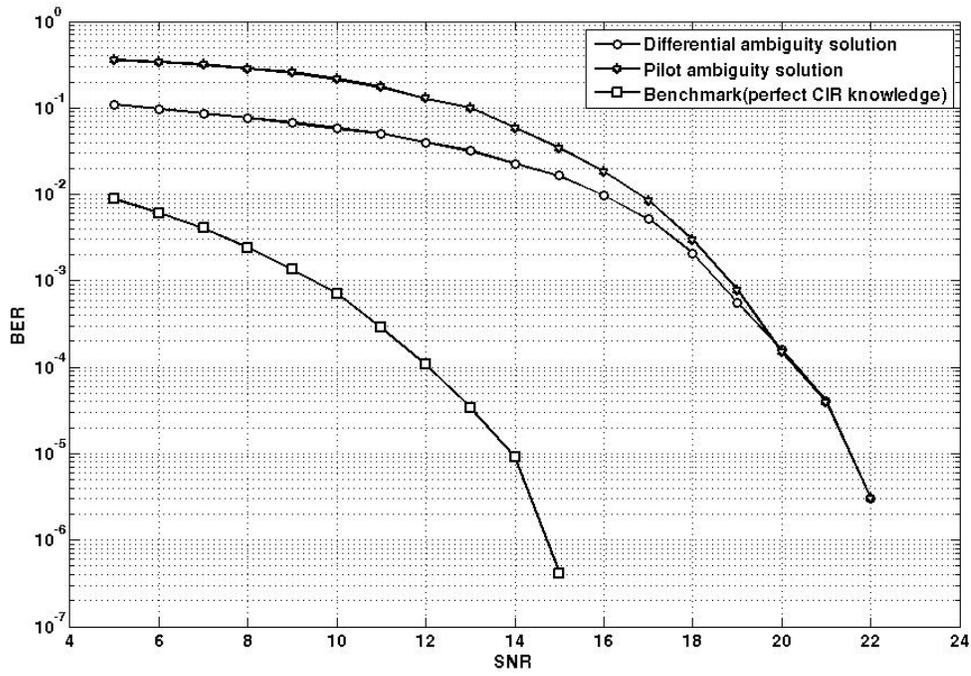
Figure 3-3: BER performance comparison of two different phase ambiguity solution with perfect channel knowledge
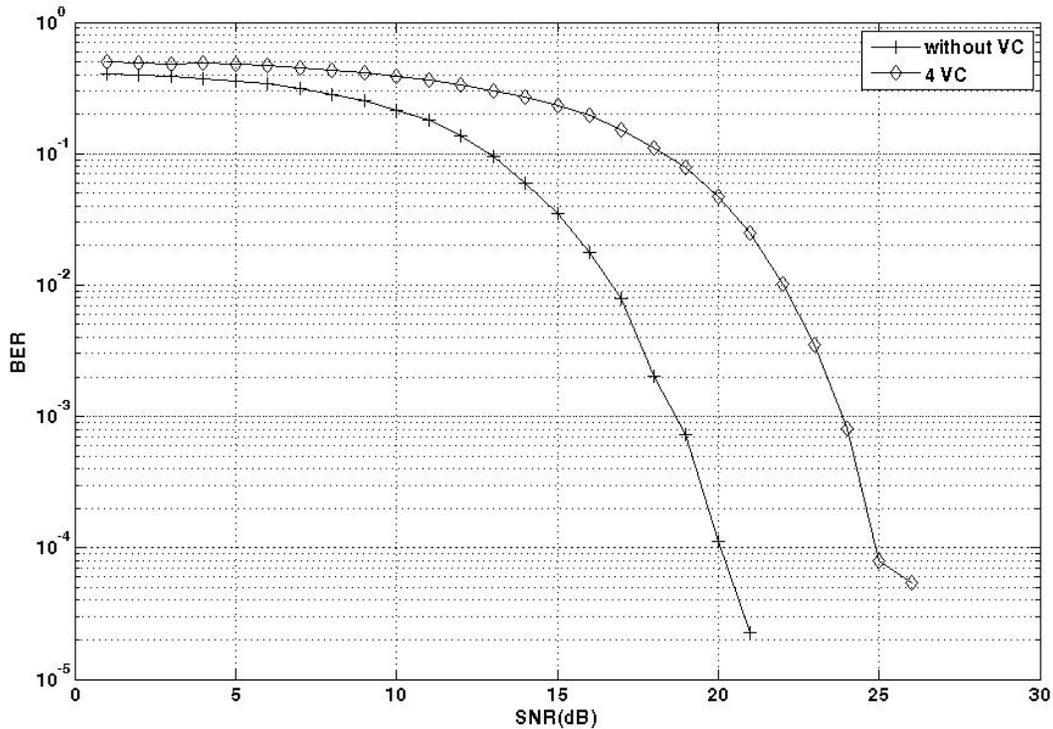
Figure 3-4: BER performance comparison of OFDM with VC and without VC

3.5 Conclusion

We have developed a blind channel estimation algorithm for SIMO OFDM systems. The algorithm exploits the relation between received data at different receivers. We then investigated the idetifiability condition of transmitted data and channel for the estimator. As the estimator is a blind one, it inherits the phase ambiguity and we used two different solutions to solve it. The performance of these two solutions has been compared in our simulations. Simulations also show that the iteration of estimated channel with data improves the performance of channel estimator.

Chapter Four: Superimposed OFDM and channel estimation

This chapter develops an approximately ML channel estimator and data detector for superimposed OFDM system. The chapter organization is as follows. Section 4.1 introduces the idea of superimposed data transmission. Section 4.2 develops the system model of a superimposed OFDM. In Section 4.3, the joint ML channel estimator and data detector are derived and discussed and the CRB of the estimator is presented in Section 4.4. Section 4.5 gives the numerical results and Section 4.6 concludes the chapter.

4.1 Introduction

The idea of superimposed training; i.e., simultaneous information transfer and channel sounding, was first described in [22], albeit for analog communications. It was advocated for digital communication systems by Farhang Boroujeny in [23] and more investigations followed in [24]– [27].The main advantage of superimposed pilot scheme is that the information symbols can be transmitted over all time-frequency slots, hence saving the bandwidth compared to time-multiplexed pilot scheme. In addition to this, in the OFDM context, none of the subcarriers need to be dedicated completely or partially for the pilots. In rapidly varying channels (in time or in frequency) superimposed pilots have an advantage in terms of improved channel tracking performance [28]. In [29], the potential of the superimposed pilot scheme for high data rate transmission has been demonstrated. Pilot-based channel estimation for OFDM has thus been widely studied [30], where pilots and data symbols are placed in separate subcarriers by periodic insertion of pilot symbols. The receiver estimates the channel at the pilot subcarriers first, and

these estimates are interpolated to estimate the channel at the data subcarriers. In mobile radio environments, the time-varying channel requires closely-spaced pilot symbols, resulting in a significant bandwidth loss. Semi-blind and blind equalization and channel estimation methods, however, need several OFDM blocks for channel estimation and exhibit both high complexity and phase ambiguities. On the other hand, pilot symbols can be added to data symbols to enable CIR estimation without sacrificing the data rate. In [31], a two-dimensional Wiener filter is employed to obtain the initial frequency domain channel estimate using second order statistics. In [32], periodic pilots are added to data symbols in time domain before transmission, and first order statics are exploited to identify the CIR. As adding pilots can increase the peak-to-average power ratio (PAPR), superimposed pilots must be carefully chosen to mitigate this problem. In [33], the data vector is distorted so that its discrete Fourier transform (DFT) at the pilot frequencies is zero, which cancels the performance degradation by the embedded unknown data. In [34], channel estimators and data detectors have been proposed for superimposed OFDM systems. The channel estimation is based on iterative ML and MMSE algorithms. The data and CIR estimates are then updated using a decision directed algorithm.

In this chapter, we derive an approximately ML channel estimator and data detector for superimposed training based OFDM. We propose channel estimator based on iterative maximum likelihood (ML). ML algorithm is used to obtain the initial estimate of the time domain CIR. The data and CIR estimate are then updated using a decision directed (DD) algorithm to improve the performance of channel estimator and data detector. Then we will derive the CRB of the approximately ML CIR estimator and optimize the superimposed pilots to attain the minimum MSE. We also derive the CRB for the CIR estimation of the joint iterative channel estimator and data detector.

4.2 Superimposed OFDM system model

We consider the discrete-time equivalent baseband model of an OFDM system over frequency-selective channels. Data are mapped into a finite constellation Q. We consider a generalized training strategy in which the transmitted symbol X_k at the kth subcarrier is a linear combination of a pilot symbol and a data symbol

$$X_k = \sqrt{\varphi_k}\, S_k + \sqrt{\phi_k}\, P_k, \quad k = 0, 1, ..., N-1 \qquad (4.1)$$

where $P_k \in Q$ is the known pilot, $S_k \in Q$ is a zero-mean randomly distributed data symbol, and both P_k and S_k have the unity average power. The coefficients ϕ_k and φ_k specify the power of the pilot and data symbols, respectively. The signal to pilot power ratio (SPR) for the kth subcarrier is defined as $SPR_k = \varphi_k / \phi_k$. The power $E_k = \varphi_k + \phi_k$ is the total power for the kth subcarrier and $E = \sum_{k=0}^{N-1} E_k$ is the total power for an OFDM block. If $\varphi_k = 0$, for $k \in I_p$, Eq. (4.1) reduces to the separated training scheme in [30], where I_p denotes the index set of N_p pilot subcarriers. Transmit symbols X_k's are modulated by an inverse DFT (IDFT), and the resulting time domain signal samples are

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j(2\pi kn/N)}, \quad n = 0, 1, ..., N-1. \qquad (4.2)$$

Note after IDFT, pilots and data symbols are superimposed in both time domain and frequency domain. A guard interval includes a cyclic prefix of $\{x(N-N_g+1), ..., x(N-1)\}$ where N_g is the number of samples in the guard interval. These samples are appropriately pulse shaped to construct the time domain signal $x(t)$ for transmission.

The composite response including transmit and receive nyquist pulse shaping response (provided that their length is less than *L)* and the physical channel response between the transmitter and the receiver may be modeled as

$$h(\tau) = \sum_{l=0}^{L-1} h_l \delta(\tau - \tau_l)$$
(4.3)

where $h_l \sim CN(0, \sigma_l^2)$, τ_l is the delay of the *l*th tap. Typically, it is modeled using $\tau_l = lT_s$, and this results in a finite impulse response filter with an effective length *L*. We consider that the channel taps h_l remain constant in each block so that inter-carrier interference (ICI) is negligible. Assuming perfect synchronization, the received signal after sampling can be represented as

$$y_n = \sum_{l=0}^{L-1} h_l x_{n-l} + w_n$$
(4.4)

where $w_n \sim CN(0, \sigma_w^2)$ is an additive white Gaussian noise (AWGN). After removing the guard interval and performing DFT demodulation, we can get

$$Y_k = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} y_k e^{j(-2\pi kn / N)} = H_k X_k + W_k , \quad 0 \le k \le N - 1$$
(4.5)

where $H_k = \sum_{l-0}^{L-1} h_l e^{-j2\pi lk / N}$ and $W_k = \frac{1}{\sqrt{N}} \sum_{n-0}^{N-1} w_n e^{-j2\pi nk / N}$ with zero mean and variance σ_w^2.

We define $\mathbf{H} = [H_0, H_1, ..., H_{N-1}]^T = \mathbf{F}_L \mathbf{h}$, where $\mathbf{h} = [h_0, h_1, ..., h_{L-1}]^T \in C^L$ is the CIR and $\mathbf{F}$ is the DFT matrix ($[\mathbf{F}]_{k,l} = \frac{1}{\sqrt{N}} e^{-j(2\pi / N)kl}, k,l \in 0,1,..., N-1$), $\mathbf{F}_L$ is a $N \times L$ submatrix (first L columns) of DFT matrix $\mathbf{F}$ scaled by $\sqrt{N}$. We can vectorize (4.5) as

$$\mathbf{Y} = \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} = (\mathbf{\Phi} \mathbf{P}_D + \mathbf{\Psi} \mathbf{S}_D) \mathbf{F}_L \mathbf{h} + \mathbf{W} \qquad (4.6)$$

or equivalently

$$\mathbf{Y} = \mathbf{H}_D \mathbf{X} + \mathbf{W} = \mathbf{H}_D (\mathbf{\Phi} \mathbf{P} + \mathbf{\Psi} \mathbf{S}) + \mathbf{W} \qquad (4.7)$$

where

$$\mathbf{X} = \begin{bmatrix} X_0, \ ..., \ X_{N-1} \end{bmatrix}^T, \ \mathbf{Y} = \begin{bmatrix} Y_0, \ ..., \ Y_{N-1} \end{bmatrix}^T, \ \mathbf{P} = \begin{bmatrix} P_0, \ ..., \ P_{N-1} \end{bmatrix}^T,$$

$$\mathbf{S} = \begin{bmatrix} S_0, \ ..., \ S_{N-1} \end{bmatrix}^T, \ \mathbf{\Phi} = diag \left\{ \sqrt{\phi_0}, \ ..., \ \sqrt{\phi_{N-1}} \right\},$$

$$\mathbf{\Psi} = diag \left\{ \sqrt{\varphi_0}, \ ..., \ \sqrt{\varphi_{N-1}} \right\}. \qquad (4.8)$$

4.3 Iterative approximately ML channel estimator

In what follows, we derive maximum likelihood estimation for CIR of a superimposed OFDM system. The information symbols are considered as unknown random variables. The ML channel estimator then can be used for decision directed algorithm to detect the information data.

Considering $S_k, k = 0, ..., N-1$ as zero-mean randomly distributed data symbols and noise $\mathbf{w}$ as a vector of complex Gaussian random variables and CIR $\mathbf{h}$ as a CGRV vector, $\mathbf{Y}$ is a CGRV vector so that we can write its pdf given $\mathbf{h}$ as

$$f(\mathbf{Y} \mid \mathbf{h}) = \frac{1}{(\pi)^N \times \sqrt{\det(\mathbf{R_{YY}})}} \exp(-(\mathbf{Y} - \mathbf{\mu})^H \mathbf{R_{YY}}^{-1} (\mathbf{Y} - \mathbf{\mu})) \qquad (4.9)$$

where the mean matrix $\mathbf{\mu}$ is:

$$\mathbf{\mu} = E\{\mathbf{Y} \mid \mathbf{h}\} = E\{\mathbf{\Phi} \mathbf{P}_D \mathbf{F}_L \mathbf{h} + \mathbf{\Psi} \mathbf{S}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} \mid \mathbf{h}\} = \mathbf{\Phi} \mathbf{P}_D \mathbf{F}_L \mathbf{h} \qquad (4.10)$$

and correlation matrix $\mathbf{R_{YY}}$ is

$$\mathbf{R}_{\mathbf{YY}} = E\{(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^H \mid \mathbf{h}\} = E\{(\boldsymbol{\Psi}\mathbf{S}_D\mathbf{F}_L\mathbf{h} + \mathbf{W})(\boldsymbol{\Psi}\mathbf{S}_D\mathbf{F}_L\mathbf{h} + \mathbf{W})^H \mid \mathbf{h}\} = \boldsymbol{\Psi}^2\mathbf{H}_{abs} + \sigma_w^2\mathbf{I}$$

$$(4.11)$$

where $\mathbf{H}_{abs} = diag\{|H_0|^2, ..., |H_{N-1}|^2\}$. Note that expectation in (4.11) is over information symbols. Now, we take an extra step of expectation over channel for correlation matrix $\mathbf{R}_{\mathbf{YY}}$ and we rewrite (4.9) approximately as

$$f(\mathbf{Y} \mid \mathbf{h}) \approx \frac{1}{(\pi)^N \times \sqrt{\det(\hat{\mathbf{R}}_{\mathbf{YY}})}} \exp(-(\mathbf{Y} - \boldsymbol{\mu})^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1}(\mathbf{Y} - \boldsymbol{\mu})) \qquad (4.12)$$

where $\hat{\mathbf{R}}_{\mathbf{YY}} = E_{\mathbf{h}}\{\mathbf{R}_{\mathbf{YY}}\} = E_{\mathbf{h}}\{\boldsymbol{\Psi}^2\mathbf{H}_{abs} + \sigma_w^2\mathbf{I}\}$. Defining $\alpha = \sum_{l=0}^{L-1}\sigma_l^2 / N$, we have

$$\hat{\mathbf{R}}_{\mathbf{YY}} = \alpha\boldsymbol{\Psi}^2 + \sigma_w^2\mathbf{I}. \qquad (4.13)$$

which is independent of $\mathbf{h}$.

If the CIR remains the same during K OFDM blocks, and if the delay is tolerable, the K consequent OFDM blocks can be combined for CIR estimation. Let $\mathbf{Y}_k$ denotes the kth received OFDM block, since the $\mathbf{Y}_k$'s for $k = 1, ..., K$ are independent (it is true because information symbols are independent of pilot symbols), the join pdf function of $(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K)$ approximately is

$$f(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \mathbf{h}) = \prod_{k=1}^{K} f(\mathbf{Y}_k \mid \mathbf{h}) \approx \frac{1}{\pi^{KN}\prod_{k=1}^{K}\sqrt{\det(\hat{\mathbf{R}}_{\mathbf{YY}})}} \exp(-\sum_{k=1}^{K}(\mathbf{Y}_k - \boldsymbol{\mu})^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1}(\mathbf{Y}_k - \boldsymbol{\mu}))$$

$$(4.14)$$

Removing terms which are independent of CIR, the log likelihood function is

$$\Lambda(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \mathbf{h}) = \ln f(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \mathbf{h}) = -\sum_{k=1}^{K}(\mathbf{Y}_k - \boldsymbol{\mu})^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1}(\mathbf{Y}_k - \boldsymbol{\mu}) \qquad (4.15)$$

Approximately maximum likelihood estimation $\hat{\mathbf{h}}$ for CIR is obtained by maximizing (4.15). Thus, we need to take the derivative of log likelihood function with respect to CIR and find its zero

$$\frac{\partial}{\partial \mathbf{h}}(\Lambda(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \hat{\mathbf{h}})) = \frac{\partial}{\partial \mathbf{h}^H}(\Lambda(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \hat{\mathbf{h}})) = 0 \tag{4.16}$$

From [35], we know

$$\frac{\partial}{\partial \mathbf{h}^H}(\Lambda(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K \mid \mathbf{h})) = (\mathbf{\Phi P}_D \mathbf{F}_L)^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1} \sum_{k=1}^{K} (\mathbf{Y}_k - \mathbf{\Phi P}_D \mathbf{F}_L \mathbf{h}). \tag{4.17}$$

Therefore the ML estimation is

$$(\mathbf{\Phi P}_D \mathbf{F}_L)^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1} \sum_{k=1}^{K} (\mathbf{Y}_k - \mathbf{\Phi P}_D \mathbf{F}_L \hat{\mathbf{h}}) = 0 \tag{4.18}$$

or equivalently,

$$\hat{\mathbf{h}} = ((\mathbf{\Phi P}_D \mathbf{F}_L)^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1} K \mathbf{\Phi P}_D \mathbf{F}_L)^{-1} (\mathbf{\Phi P}_D \mathbf{F}_L)^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1} (\sum_{k=1}^{K} \mathbf{Y}_k). \tag{4.19}$$

Next we use the estimated channel to form a joint channel estimator and data detector following decision directed technique.

The pdf of the received signal $\mathbf{Y}$ conditioned on $\mathbf{h}$ and $\mathbf{S}$ is

$$f(\mathbf{Y} \mid \mathbf{h}, \mathbf{S}) = \frac{1}{(\pi \sigma_w)^N} \exp \left\{ -\frac{1}{\sigma_w^2} \left\| \mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h} \right\|^2 \right\} \tag{4.20}$$

so the joint channel estimator and data detector is given by

$$\{\hat{\mathbf{S}}, \hat{\mathbf{h}}\} = \arg\min_{\mathbf{S} \in Q^N, \mathbf{h} \in C^L} \left\| \mathbf{Y} - (\mathbf{\Phi P}_D + \mathbf{\Psi S}_D) \mathbf{F}_L \mathbf{h} \right\|^2. \tag{4.21}$$

Starting from (4.19) as $\hat{\mathbf{h}}^0$, a decision directed (DD) technique can be used to improve the performance of both channel estimation and data detection. In the ith iteration, data symbols $\mathbf{S}$ can be estimated via

$$\hat{\mathbf{S}}^i = M_Q(\boldsymbol{\Psi}^{-1}\left[(\mathbf{H}_D^{i-1})^{-1}\mathbf{Y} - \boldsymbol{\Phi}\mathbf{P}_D\right]) \tag{4.22}$$

where $\mathbf{H}_D^{i-1} = diag\{\mathbf{F}_L \hat{\mathbf{h}}^{i-1}\}$, and $M_Q(.)$ quantize (.) to the nearest element in Q. The CIR estimation given $\hat{\mathbf{S}}^i$ follows as (LS solution to (4.21))

$$\hat{\mathbf{h}}^i = (\mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L)^{-1}\mathbf{F}_L^H \mathbf{X}_D^H \mathbf{Y} \tag{4.23}$$

where

$$\mathbf{X}_D = \boldsymbol{\Phi}\mathbf{P}_D + \boldsymbol{\Psi}\mathbf{S}_D^i \tag{4.24}$$

4.4 Cramer-Rao bound of the ML channel estimator

Cramer-Rao bound establishes a lower bound on the error covariance matrix of any unbiased estimator ($\hat{\boldsymbol{\theta}}$) of a parameter ($\boldsymbol{\theta}$) [35]:

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H\} \geq \mathbf{J}^{-1}(\boldsymbol{\theta}) \tag{4.25}$$

where $\mathbf{J}$ is called Fisher information matrix (FIM).

In order to evaluate the performance of our approximately ML estimator, we derive its CRB. We first derive the CRB of the initial approximately ML channel estimator given in (4.19) and then we derive the CRB of the joint channel estimator and data detector given in (4.23).

The FIM associated with a complex stochastic parameter vector $\boldsymbol{\theta}$ is defined as [36]

$$\mathbf{J}(\boldsymbol{\theta}) = E\{[\frac{\partial \ln f(\mathbf{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\mathbf{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H\} =$$
$$E\{E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H |\boldsymbol{\theta}\}\} + \quad (4.26)$$
$$E\{[\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H\} = E\{\mathbf{J}_1\} + \mathbf{J}_2$$

where $\boldsymbol{\theta} = \mathbf{h}$ for the channel estimator in (4.19).

Using [35, p. 237, eq. (6.134)] yields

$$\mathbf{J}_1(i, j) = \frac{K}{2}tr[\mathbf{R_{YY}}^{-1}\frac{\partial \mathbf{R_{YY}}}{\partial \theta_i^*}\mathbf{R_{YY}}^{-1}\frac{\partial \mathbf{R_{YY}}}{\partial \theta_j^*}] + K\frac{\partial \boldsymbol{\mu}^H}{\partial \theta_i^*}\mathbf{R_{YY}}^{-1}(\frac{\partial \boldsymbol{\mu}^H}{\partial \theta_j^*})^H \quad (4.27)$$

Again we replace $\mathbf{R_{YY}}^{-1}$ with $\hat{\mathbf{R}}_{\mathbf{YY}}^{-1}$ so that the derivative of $\mathbf{R_{YY}}^{-1}$ with respect to $\mathbf{h}$ is zero

and the expression of (4.27) can be simplified to

$$\mathbf{J}_1 \approx K(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L)^H \hat{\mathbf{R}}_{\mathbf{YY}}^{-1}(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L) \quad (4.28)$$

Considering that CIR is a zero mean CGRV we will have

$$\mathbf{J}_2 = E\{[\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H\} = \mathbf{R_h}^{-1}, \quad (4.29)$$

therefore

$$\mathbf{J}(\mathbf{h}) \approx E_{\mathbf{h}}\{K(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L)^H(\alpha\boldsymbol{\Psi}^2 + \sigma_w^2\mathbf{I})^{-1}(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L)\} + \mathbf{R_h}^{-1} = K(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L)^H(\alpha\boldsymbol{\Psi}^2 + \sigma_w^2\mathbf{I})^{-1}(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L) + \mathbf{R_h}^{-1}$$
$$. \quad (4.30)$$

The approximation of the CRB for the MSE of approximately ML channel estimation is then

given by

$$CRB_{\mathbf{h}} \approx trace\{(K(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L)^H(\alpha\boldsymbol{\Psi}^2 + \sigma_w^2\mathbf{I})^{-1}(\boldsymbol{\Phi}\mathbf{P}_D\mathbf{F}_L) + \mathbf{R_h}^{-1})^{-1}\} \quad (4.31)$$

Since the trace in (4.27) is always positive, the actual CRB will always be tighter than the approximated CRB given in (4.31). However the difference is not negligible in realistic situations.

The place of pilots and power distribution of data and pilots can be optimized to minimize the CRB subject to the power constraint $\sum_{k=0}^{N-1} \varphi_k = D$ and $\sum_{k \in I_p} \phi_k = P$ where I_p is the index of N_p subcarriers with superimposed pilot ($\phi_k \neq 0$) and P and D are the total power on pilots and data symbols respectively. The problem of optimal pilot design becomes

$$(I_p, \varphi_k, \phi_k) = \underset{\sum_{k=0}^{N-1} \varphi_k = D, \sum_{k \in I_p} \phi_k = P}{\arg \min} \quad trace\{((\mathbf{\Phi P}_D \mathbf{F}_L)^H (\alpha \mathbf{\Psi}^2 + \sigma_w^2 \mathbf{I})^{-1} (\mathbf{\Phi P}_D \mathbf{F}_L) + \mathbf{R_h}^{-1})^{-1}\}.$$

$$(4.32)$$

From [37], the lower bound on the MSE of channel estimation is attained if and only if $\mathbf{A} = (\mathbf{\Phi P}_D \mathbf{F}_L)^H (\alpha \mathbf{\Psi}^2 + \sigma_w^2 \mathbf{I})^{-1} (\mathbf{\Phi P}_D \mathbf{F}_L)$ is diagonal. The (r,s)th ($0 \leq r, s \leq L-1$) entry of $\mathbf{A}$ can be written as

$$[\mathbf{A}]_{r,s} = \sum_{k \in I_p} e^{j \frac{2\pi(r-s)k}{N}} \frac{\phi_k^2 |P_k|^2}{\alpha \varphi_k^2 + \sigma_w^2} \qquad (4.33)$$

therefore we require

$$\sum_{k \in I_p} e^{j \frac{2\pi(r-s)k}{N}} \frac{\phi_k^2 |P_k|^2}{\alpha \varphi_k^2 + \sigma_w^2} = 0, \quad r - s \neq 0. \qquad (4.34)$$

Eq. (4.34) is satisfied if the following conditions are satisfied

$$z = N / N_p \in Z \quad I_p = \{z_0 + k'z, \; k' = 0, 1, ..., N_p - 1\}, z_0 \in \{0, 1, ..., z - 1\} \qquad (4.35)$$

$$\frac{\phi_k^2 |P_k|^2}{\alpha \varphi_k^2 + \sigma_w^2} = cons. \quad k \in I_p . \qquad (4.36)$$

In case we want to design the power allocation independent of SNR, condition (4.36) becomes

$$\phi_k |P_k| = cons., \; \varphi_k = cons \quad k \in I_p . \qquad (4.37)$$

Also we should consider the amount of pilots (N_p) to follow

$$L - 1 < N_p \qquad (4.38)$$

to make sure that equation (4.34) is always true.

Conditions (4.34) and (4.37) and (4.38) mean that pilots must be equispaced and the power of signal and pilot should be the same for all subcarriers in I_p and we need at least L superimposed pilots. These conditions lead us to the initial channel estimation with minimum MSE.

Next we derive the CRB for the joint channel estimator and data detector given in (4.21). Starting from (4.26), where $\boldsymbol{\theta} = [\mathbf{h}^T, \mathbf{S}^T]^T$ for the channel estimator in (4.21), we have

$$\mathbf{J}_1 = E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H | \boldsymbol{\theta}\} =$$

$$\begin{Bmatrix} E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H | \boldsymbol{\theta}\} & E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{S}^*}]^H | \boldsymbol{\theta}\} \\ E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{S}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H | \boldsymbol{\theta}\} & E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{S}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{S}^*}]^H | \boldsymbol{\theta}\} \end{Bmatrix} = \begin{Bmatrix} \mathbf{J}_{1,1} & \mathbf{J}_{1,2} \\ \mathbf{J}_{1,3} & \mathbf{J}_{1,4} \end{Bmatrix}$$

$$(4.39)$$

and

$$\mathbf{J}_2 = E\{[\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}]^H\} = E\left\{\begin{matrix} [\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H & [\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{S}^*}]^H \\ [\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{S}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H & [\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{S}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{S}^*}]^H \end{matrix}\right\}$$

$$= \begin{Bmatrix} \mathbf{J}_{2,1} & \mathbf{J}_{2,2} \\ \mathbf{J}_{2,3} & \mathbf{J}_{2,4} \end{Bmatrix}$$

, $\qquad\qquad$ (4.40)

thus the CRB is defined as

$$CRB_{\mathbf{h}} = trace\,(\tilde{\mathbf{J}}^{-1}), \quad \tilde{\mathbf{J}} = E\{\mathbf{J}_{1,1}\} + \mathbf{J}_{2,1}\,. \qquad (4.41)$$

From (4.20) and [35] we know

$$\mathbf{J}_{1,1} = E\{[\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H \,|\,\boldsymbol{\theta}\} = \frac{1}{\sigma_w^2}(\mathbf{X}_D\mathbf{F}_L)^H(\mathbf{X}_D\mathbf{F}_L)\,. \qquad (4.42)$$

Since the information data is independent of CIR,

$$f(\boldsymbol{\theta}) = f(\mathbf{h},\mathbf{S}) = f(\mathbf{h})f(\mathbf{S}) \qquad (4.43)$$

therefore

$$\mathbf{J}_{2,1} = E\{[\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\boldsymbol{\theta})}{\partial \mathbf{h}^*}]^H\} = E\{[\frac{\partial \ln f(\mathbf{h})}{\partial \mathbf{h}^*}][\frac{\partial \ln f(\mathbf{h})}{\partial \mathbf{h}^*}]^H\} = \mathbf{R}_{\mathbf{h}}^{-1}\,. \qquad (4.44)$$

Therefore we have

$$\tilde{\mathbf{J}} = \frac{1}{\sigma_w^2}\mathbf{F}_L^H E\{\mathbf{X}_D^H\mathbf{X}_D\}\mathbf{F}_L + \mathbf{R}_{\mathbf{h}}^{-1} = \frac{1}{\sigma_w^2}\mathbf{F}_L^H(\mathbf{\Psi}^2 + \mathbf{\Phi}^2)\mathbf{F}_L + \mathbf{R}_{\mathbf{h}}^{-1} \qquad (4.45)$$

and

$$CRB_{\mathbf{h}} = trace((\frac{1}{\sigma_w^2}\mathbf{F}_L^H(\mathbf{\Psi}^2 + \mathbf{\Phi}^2)\mathbf{F}_L + \mathbf{R}_{\mathbf{h}}^{-1})^{-1}) \qquad (4.46)$$

4.5 Simulation results

Simulation results are given for the proposed channel estimator. In simulation, a superimposed OFDM system with 32 subcarriers has been developed. A COST 207 6-ary channel model with power profile $[0.189, 0.379, 0.239, 0.095, 0.061, 0.037]$ is considered. Each path is an independently generated complex Gaussian random process. CIR is constant over K consequent OFDM blocks but can vary from one K blocks to another. Both information data and pilot data have been drawn form BPSK constellation. We compare the performance of 8 equispaced superimposed pilots with that of 16 equispaced superimposed pilots and with different number of iterations. The total power at each subcarrier is 1 and the superimposed pilot subcarriers have 0.7 power. The notation $i=n$ denotes the performance in nth iteration. Ideal detectors, assuming the availability of perfect CIR knowledge, are used as benchmarks.

Figure 4-1 shows the mean square error (MSE) of channel estimation which is defined as

$$MSE = (\frac{1}{R}\sum_{i=1}^{R}\left\|\hat{\mathbf{h}}_i - \mathbf{h}\right\|^2)$$
(4.47)

where R is the number of Monte Carlo runs. In first simulation $K=1$ and 8 out of 32 subcarriers have superimposed pilots. Superimposed pilots are equispaced and equipower. It also includes the CRB of the initial estimation (Eq. (4.31)) and the iterated estimation (Eq. (4.46)). As it can be seen, in high SNR, one step of iteration improves the MSE of estimator significantly (almost 10 dB) but the 4th iteration doesn't do more than 3rd iteration. In high SNR, estimators with 3 or 4 iterations get quite close to the CRB of iterated estimation.
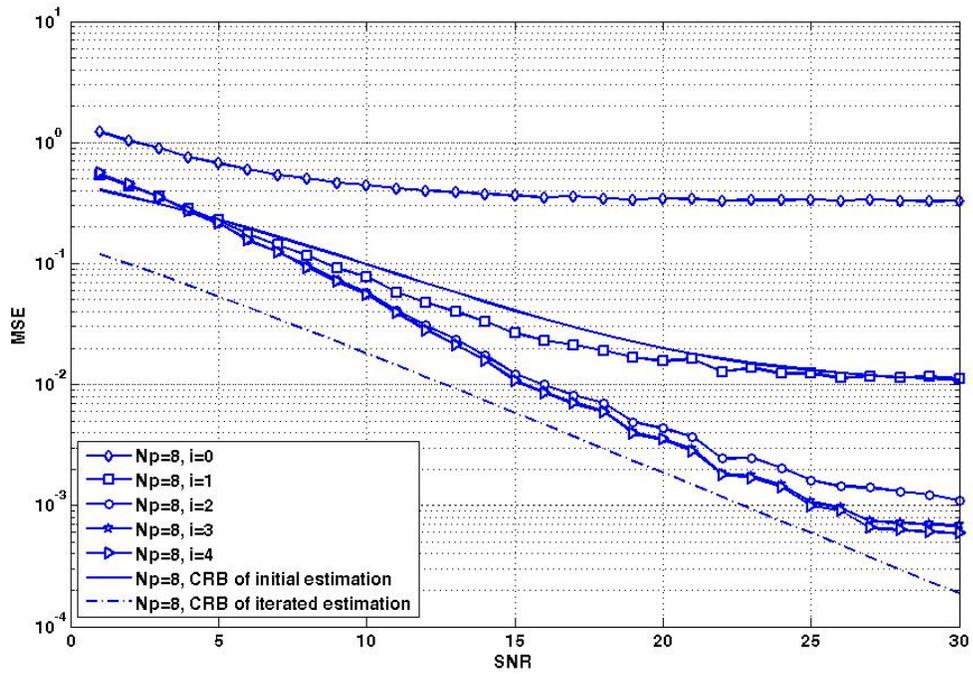
Figure 4-1: MSE of proposed channel estimator with different number of iteration and CRB of estimator vs. SNR
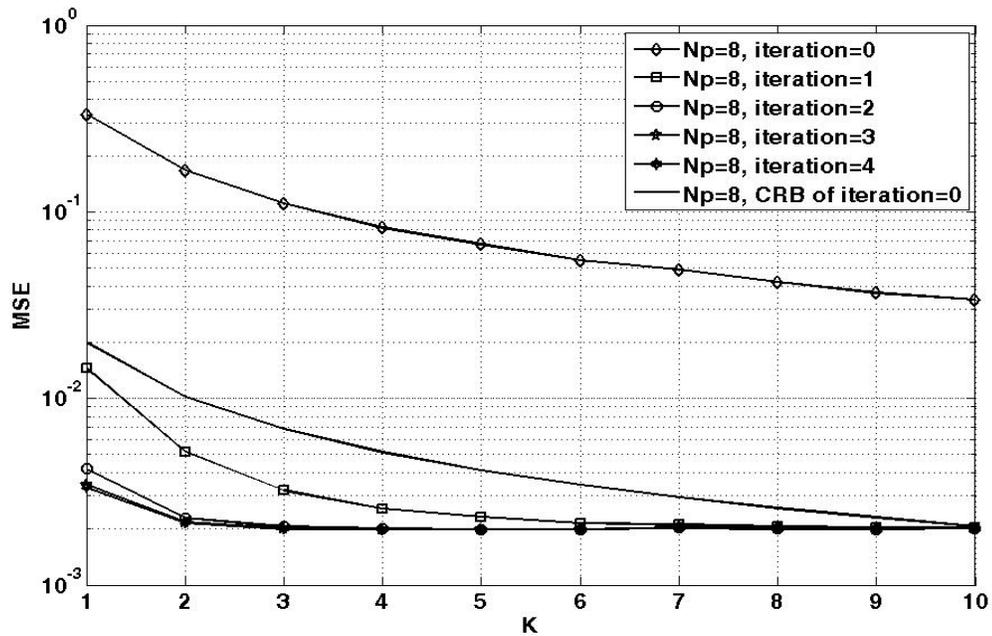
Figure 4-2: MSE of channel estimator vs. *K*

Next we investigate the effect of number of OFDM blocks included in estimation (K) on MSE of our estimator. Figure 4-2 shows that increase of K from 1 to 10 improve MSE by almost 9 and 10 dB for initial estimation CRB and channel estimator with no iteration respectively while the improvement is about 1 dB provided that 3 or more iterations are done. This means for the case of time varying channel, when we can't assume channel is unchanged for big value of K, iteration should be done to get good estimation of channel. In this simulation, we fixed the value of SNR to 20dB.

Since the proposed channel estimator needs to know the noise power (Eq. (4.19)), we investigate the robustness of our estimator to mismatch. Figure 4-3, shows the MSE of channel estimator with zero and one iteration and compares them with the MSE for the case with a mismatch. The mismatched value of noise power is the noise power related to SNR equal to 20 dB. As it can be seen both of the estimators perform robust to this mismatch provided that the approximated noise power is related to a high SNR.
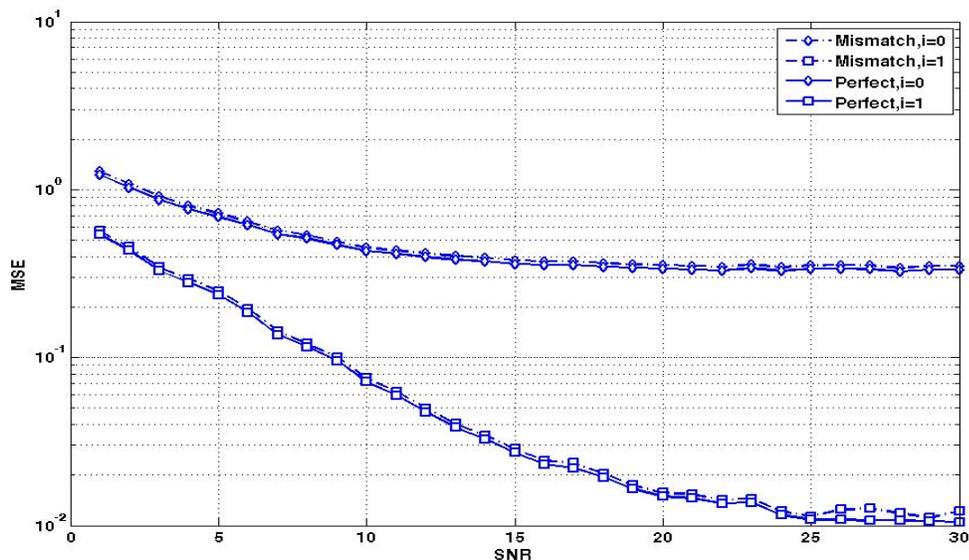
Figure 4-3: Comparison of MSE of channel estimator with perfect knowledge of noise power and mismatch

Next, we would like to see the effect of number of superimposed pilots on MSE of proposed estimators. Figure 4-4 compares the performance of CRB and estimator with different number of iterations between the system with 8 superimposed pilots and 16 pilots. As it can be seen CRB of system with 16 pilots is almost 2 dB better than the one with 8 pilots for the case of zero iteration while the difference increase as the number of iterations increase. It suggests that performing iteration can save the transmitted power. In other words, instead of increasing the number of pilots for better estimation, we can use decision directed algorithm and iteration. MSE of the estimator with 16 superimposed pilots with 3 or more iteration has less than 1 dB difference with the CRB in high SNR. This figure also includes the CRB of the initial estimation for 8 and 16 superimposed pilots. Not that the CRB of iterated estimator is not a function of the number of pilots, instead it is a function of the total power on each subcarrier.
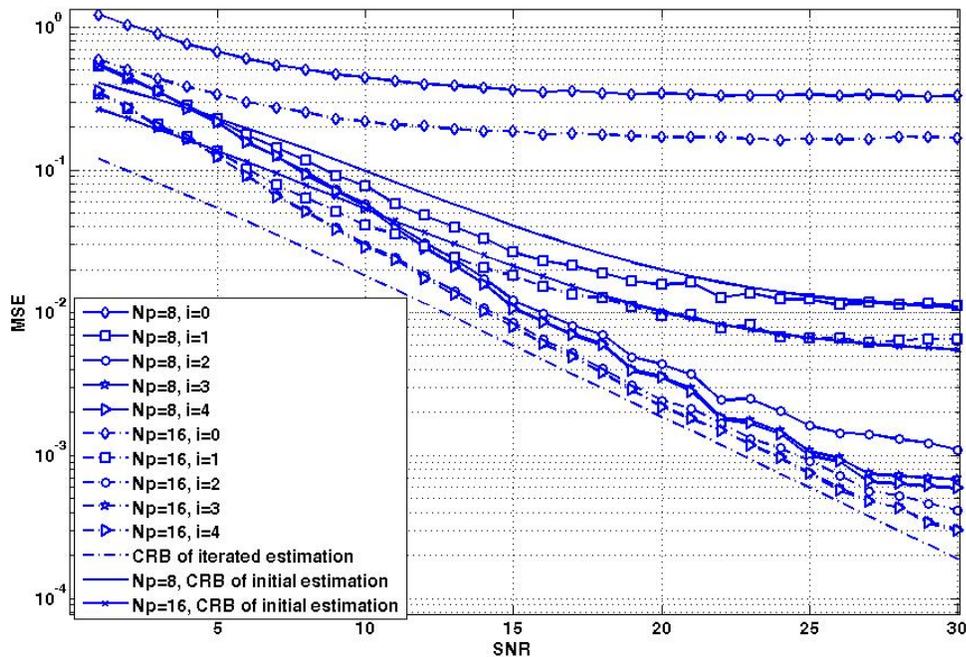
Figure 4-4: Comparison of MSE of channel estimator with different numbers of superimposed pilots

The last simulation compares the BER performance of our channel estimator and data detector for different number of iteration and 8 and 16 superimposed pilots. The estimated channel has been used to detect the information symbols. The performance of equalizer with perfect knowledge of the CIR provides the benchmark. Figure 4-5 shows that in high SNR the difference between zero and one iteration is almost 6 dB while it decreases to 4 dB going from 1st iteration to the 2nd one. Eventually the improvement caused by increasing the number of iterations from 3 to 4 is insignificant. Data detector with perfect knowledge of CIR performs just 3 dB better than the one with 16 superimposed pilots and more than 3 iterations.
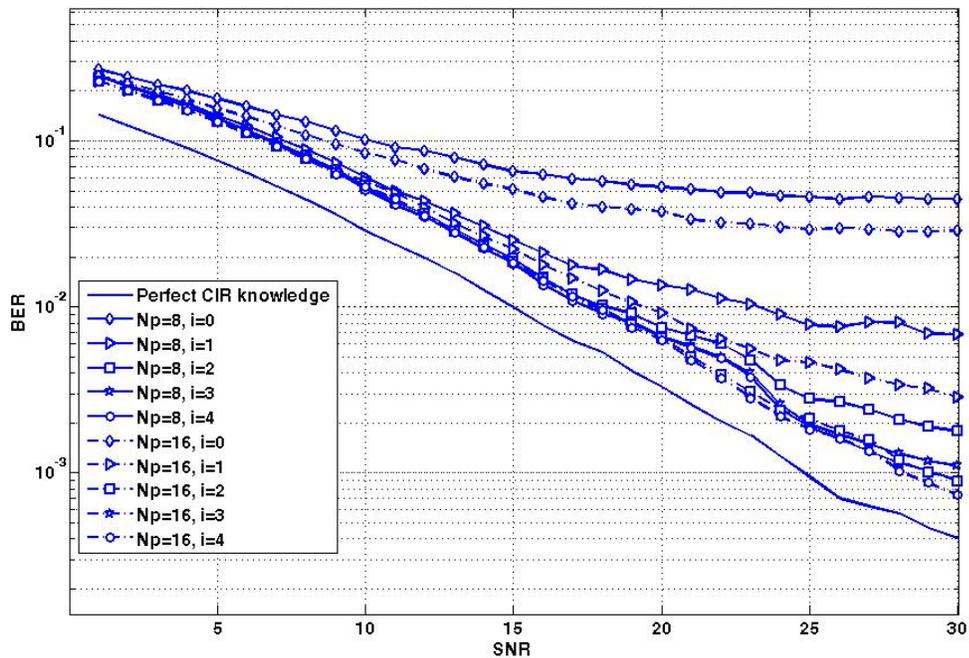
Figure 4-5: BER performance of estimated channel used for data detection

4.6 Conclusion

We have proposed an iterative ML channel estimator and data detector for superimposed OFDM system. Since we have approximated the correlation matrix in derivative of our ML estimator, it shows an error floor in high SNR. To improve the MSE of the channel estimator we used a decision directed algorithm to ensure good performance. We have derived the CRB for our approximately ML channel estimator and optimized the allocation and the power distribution of superimposed pilots such that the MSE (CRB) is minimized. We also derived the CRB of our iterative channel estimator. We investigated the effect of the amount of superimposed pilots and also number of OFDM blocks and number of iterations on our estimator's performance in our simulations.

Chapter Five: Carrier Frequency Offset in OFDM

This chapter presents several algorithms for carrier frequency offset (CFO) estimation in OFDM. Section 5.1 introduces the CFO issue in OFDM and addresses some of the present estimation algorithms and formulates CFO in OFDM. Section 5.2 and 5.3 and 5.4 and 5.5 present several algorithms for CFO estimation and Section 5.6 introduces a joint channel and CFO estimator. Section 5.7 gives the numerical results and Section 5.8 concludes the chapter.

5.1 Introduction

5.1.1 Literature review

OFDM introduces spectral efficiency (no guard bands are needed between adjacent frequency channels). More importantly, its implementation simplicity compared to traditional time domain modulation methods in channels with severe inter-symbol interference (ISI), is huge. OFDM does have its drawbacks relative to time domain modulation, most significantly its extreme sensitivity to time varying multiplicative effects such as fast fading, Doppler shifts, and oscillator jitter. The latter two effects lead to a mismatch between the carrier frequencies of the received signal and the local oscillator, so that a frequency offset is created.

OFDM provides an efficient way to combat multipath fading by dividing one high bit rate data stream into multiple low bit rate streams for simultaneous transmission on multiple subcarriers. However, as a multicarrier transmission technique, OFDM is more susceptible to the carrier frequency offset (CFO) than single carrier systems [38]. A carrier offset at the receiver can destroy the mutual orthogonality between subcarriers and thus introduce inter-channel interference (ICI) and cause severe degradation in system performance [38]. Consequently,

accurate estimation and compensation of CFO is necessary at the receiver before OFDM demodulation.

CFO estimation algorithms can be divided into two categories: data aided and blind schemes. Data aided schemes exploit the training sequence known to the receiver while blind schemes utilize the redundant information in the received sequence such as virtual carrier (VC). Blind algorithms are more efficient in the view of bandwidth usage of the system. In [39], a non-blind scheme was proposed where repeated symbols were used for estimation. The algorithm in [40] exploits cyclic prefix (CP) preceding the OFDM symbols for CFO estimation, thus reducing the need for pilots, but mostly developed for flat fading channels. Another solution to the estimation problem was proposed in [43] which take the advantage of virtual carriers. A ML estimator was proposed in [41] and it was shown that it is equivalent to the MUSIC-like algorithm in [42][43].

In what follows, we will develop several CFO estimators and derive their CRB and use this bound to optimize the parameters existed in the estimators.

5.1.2 Background

In the presence of CFO, every sample at the receiver is modulated with a complex exponential which is a function of sample index and frequency offset. In other words, CFO causes a rotation in received samples. Let ε_0 denotes the normalized CFO, which is the actual CFO divided by the subcarrier spacing. In the presence of carrier offset, we have

$$y(n) = e^{\frac{j2\pi\varepsilon_0 n}{N}} \sum_{l=0}^{L-1} h_l x(n-l) + w(n), \quad -\infty \leq n \leq \infty \qquad (5.1)$$

where $y(n)$, $x(n)$ and $w(n)$ are the nth received sample, transmitted sample and noise sample

respectively and h_l is the CIR for lth tap . Thus, in an OFDM system, from (5.1), after

discarding CP, the pre-DFT data can be rewritten as

$$y_m(n) = e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(n+N_g)}{N}} \frac{1}{\sqrt{N}} \sum_{l=0}^{L-1} h_l \sum_{k=0}^{N-1} X_m(k) e^{-\frac{j2\pi k(n-l)}{N}} + w_m(n) =$$

$$e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(n+N_g)}{N}} \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_m(k)(\sum_{l=0}^{L-1} h_l e^{\frac{j2\pi kl}{N}}) e^{-\frac{j2\pi kn}{N}} + w_m(n) = \qquad (5.2)$$

$$e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(n+N_g)}{N}} \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_m(k) H_k e^{-\frac{j2\pi kn}{N}} + w_m(n), \quad n = 0,1,....,N-1$$

where m is the block interval index.

Now, if we group the pre-DFT data in blocks of length N, we will have

$$\mathbf{y}_m = e^{j2\pi\varepsilon_0(N_g + m(N+N_g))/N} \mathbf{V}\mathbf{F}^H \mathbf{H}_D \mathbf{X}_m + \mathbf{w}_m \qquad (5.3)$$

where $\mathbf{V} = diag\ (1, e^{j2\pi\varepsilon_0/N},..., e^{j2\pi\varepsilon_0(N-1)/N})$. The definition of other matrixes is the same as

per previous chapters.

5.2 Frequency Offset Estimation by exploiting correlation of CP carrying received data

In this section, we will develop cyclic prefix based CFO estimator for OFDM symbols over

multipath fading channels. The key is to use the fact that CP contains the repeated samples which

introduces a special correlation structure on the received samples. The correlation between each

received signal sample over the CP interval and its corresponding sample at the end of the

OFDM block will be used to estimate CFO.

The nth OFDM symbol during the mth block interval is denoted as $x_m(n)$ and the received signal is denoted as $y_m(n)$. At the border between two OFDM blocks $(-N_g \leq n < 0)$, the received signal samples can be written as

$$y_m(n) = e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(N_g+n)}{N}} (\sum_{l=0}^{L-1} h_l x_m(n-l)U(n-l+N_g) +$$

$$\sum_{l=0}^{L-1} h_l x_{m-1}(N+n-l+N_g)U(l-n+N_g)) + w_m(n) \tag{5.4}$$

where $U(.)$ is the step function. The correlation between the samples of received data over the beginning and end of OFDM block can be given by

$$E\left\{y_m(-k)y_m^*(N-k)\right\} = \begin{cases} \sigma_h^2 e^{-j2\pi\varepsilon_0} & 0 < k \leq N_g - (L-1) \\ (\sum_{l=0}^{L-1} \sigma_l^2 U(N_g-k-l))e^{-j2\pi\varepsilon_0} & N_g - (L-1) < k \leq N_g \end{cases} \tag{5.5}$$

where $\sigma_h^2 = \sum_{l=0}^{L-1} \sigma_l^2$. Note that the expectation is taken with respect to both data and channel.

From the Eq. (5.5), it can be seen that the angle of $E\left\{y_m(-k)y_m^*(N-k)\right\}$ is a function of CFO and therefore it can be considered as an estimator for CFO. Note that this estimator doesn't need to have any knowledge of SNR and channel power delay profile but the length of CIR.

In practical implementation, expectation is replaced with averaging over M consecutive values. In other words,

$$E\left\{y_m(-k)y_m^*(N-k)\right\} \approx \frac{1}{M}\sum_{m=1}^{M} y_m(-k)y_m^*(N-k) \tag{5.6}$$

while we assumed that CFO is invariant within M OFDM symbols.

After estimation and compensation of CFO, channel power delay profile and noise power can be estimated using the algorithm in [44].

Looking back at Eq. (5.4), for $(L-1) - N_g < n < 0$, the received signal sample over the CP interval is just a function of the current block of transmitted OFDM symbols provided that $(L-1) < N_g$. In other words,

$$y_m(n) = e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(N_g+n)}{N}} (\sum_{l=0}^{L-1} h_l x_m(n-l)) + w_m(n). \qquad (5.7)$$

We know

$$y_m(N+n) = e^{\frac{j2\pi\varepsilon_0 m(N+N_g)}{N}} e^{\frac{j2\pi\varepsilon_0(N_g+n+N)}{N}} (\sum_{l=0}^{L-1} h_l x_m(n-l)) + w_m(N+n), \qquad (5.8)$$

therefore in the noise free case (or the high SNR regime),

$$\frac{y_m(N+n)}{y_m(n)} = e^{-j2\pi\varepsilon_0} \qquad (5.9)$$

This equation shows when the GI length (N_g) is longer than CIR length ($L-1$), the CFO can be estimated by taking the difference of the angle of the received data sample over the CP interval and its corresponding sample at the end of OFDM block.

5.3 Superimposed training aided Carrier Frequency Offset Estimation by exploiting cross correlation of consequent OFDM blocks

In this section, we will propose a method which utilizes superimposed training data for frequency offset estimation. Our estimator doesn't require knowing the CIR [50].

Let $\mathbf{S}_i = [S_i(0), S_i(1),..., S_i(N-1)]^T$ denote the ith block of information data to be transmitted which is assumed to consist of independent identically distributed (i.i.d) random variables with zero mean ($S_i(k) \in Q$, $k = 0, ..., N-1$ are zero-mean randomly distributed data symbols and have the unity average power) and covariance matrix $\mathbf{R}_{\mathbf{S}_i \mathbf{S}_j}$

$$\mathbf{R}_{\mathbf{S}_i \mathbf{S}_j} = E\left\{\mathbf{S}_i \mathbf{S}_j^H\right\} = \begin{cases} \mathbf{0}_N & i \neq j \\ \mathbf{I}_N & i = j \end{cases} \tag{5.10}$$

Let $\mathbf{P} = [P(0), P(1),..., P(N-1)]^T$ represent the pilot vector which is independent of the information data and is superimposed onto it. If we assume the same pilot vector for different OFDM blocks, the OFDM block to be transmitted can be written as

$$\mathbf{X}_i = [X_i(0), X_i(1),..., X_i(N-1)]^T = \mathbf{\Phi P} + \mathbf{\Psi S}_i \tag{5.11}$$

where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are given in (4.8). From (5.10) and our assumptions above, we can write the covariance matrix $\mathbf{R}_{\mathbf{X}_i \mathbf{X}_j}$ as

$$\mathbf{R}_{\mathbf{X}_i \mathbf{X}_j} = E\left\{\mathbf{X}_i \mathbf{X}_j^H\right\} = \begin{cases} \mathbf{\Phi}^2 & i \neq j \\ \mathbf{\Phi}^2 + \mathbf{\Psi}^2 & i = j \end{cases} \tag{5.12}$$

After taking the N-point IFFT, the resulting signal can be expressed by $\mathbf{x}_i = \mathbf{F}^H \mathbf{X}_i$. One transmitted OFDM block period usually consists of a regular symbol interval and a guard interval (CP). Assuming that the CP length (N_g) is greater than or equal to the CIR length (L), ISI is completely eliminated. The last N_g samples of OFDM block $\mathbf{x}_i$ are duplicated and appended to the beginning of the OFDM block to form CP, and then the OFDM symbols are transmitted serially on the channel. Composite CIR modeled has been considered the same as (4.3).

At the receiver, the CP will be discarded from the received OFDM block. In the presence of carrier offset, the received signal is modulated by a residual carrier $e^{j2\pi\varepsilon_0 n/N}$. After removing CP, the ith N-point received OFDM symbol block can be expressed by

$$\mathbf{y}_i = e^{j2\pi\varepsilon_0(N_g+i(N+N_g))/N}\mathbf{VF}^H\mathbf{H}_D\mathbf{X}_i + \mathbf{w}_i \qquad (5.13)$$

where $\mathbf{H}_D = diag\{H(0), H(1),..., H(N-1)\}$ and $\mathbf{w}_i$ is a white Complex Gaussian noise vector, with zero mean and variance of $\sigma_{w,i}^2$.

If the channel is static, and CFO can be assumed the same for two consequent OFDM blocks

$$\mathbf{y}_{i+1} = \gamma e^{j2\pi\varepsilon_0(N+N_g)/N}\mathbf{VF}^H\mathbf{H}_D\mathbf{X}_{i+1} + \mathbf{w}_{i+1} \qquad (5.14)$$

where $\gamma = e^{j2\pi\varepsilon_0(N_g+i(N+N_g))/N}$. So now,

$$\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}} = E\{\mathbf{y}_i\mathbf{y}_{i+1}^H\} = E\{e^{-j2\pi\varepsilon_0(N+N_g)/N}\mathbf{VF}^H\mathbf{H}_D\mathbf{X}_i\mathbf{X}_{i+1}^H\mathbf{H}_D^H\mathbf{FV}^H\} + E\{\mathbf{w}_i\mathbf{w}_{i+1}^H\}. (5.15)$$

We know

$$E\{H(s)H^*(s)\} = E\{\mathbf{F}_L(j)\mathbf{h}\mathbf{h}^H\mathbf{F}_L^H(j)\} = \frac{1}{N}\sum_{i=0}^{L-1}\sigma_i^2 \qquad . \qquad (5.16)$$

Then taking the average both over OFDM symbols and channel and considering the fact that

$$[\mathbf{H}_D\mathbf{X}_i\mathbf{X}_{i+1}^H\mathbf{H}_D^H]_{r,s} = H(r)X_i(r)H^*(s)X_{i+1}^*(s) =$$
$$H(r)H^*(s)(\sqrt{\varphi(r)}S_i(r)+\sqrt{\phi(r)}P(r))(\sqrt{\varphi(s)}S_{i+1}(s)+\sqrt{\phi(s)}P(s))$$
$$E\{[\mathbf{H}_D\mathbf{X}_i\mathbf{X}_{i+1}^H\mathbf{H}_D^H]_{r,s}\} = E\{H(r)H^*(s)\}(E\{\sqrt{\varphi(r)}S_i(r)\sqrt{\varphi(s)}S_{i+1}(s)\}+E\{\sqrt{\phi(r)}P(r)\sqrt{\phi(s)}P(s)\}) =$$
$$E\{H(r)H^*(s)\}E\{\sqrt{\phi(r)}P(r)\sqrt{\phi(s)}P(s)\}\delta(r-s)$$

$$(5.17)$$

and defining $\alpha = \sum_{l=0}^{L-1}\sigma_l^2/N$ and using (5.16) and (5.17), we will have

$$\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}} = \alpha\, e^{-j2\pi\varepsilon_0(N+N_g)/N}\mathbf{VF}^H\mathbf{\Phi}^2\mathbf{FV}^H. \tag{5.18}$$

Since DFT is a unitary transform, the diagonal elements of $\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}}$ can be written as

$$\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}}(k,k) = \frac{\alpha}{N} e^{-j2\pi\varepsilon_0(N+N_g)/N}\sum_{j=0}^{N-1}\phi_j, \quad k=1,....,N \tag{5.19}$$

(5.19) suggests using the angle of the diagonal elements of $\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}}$ to estimate the normalized CFO ε_0,

$$Phase(\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}}(k,k)) = -2\pi\varepsilon_0(N+N_g)/N, \quad k=1,....,N \tag{5.20}$$

so (5.20) is the basic of CFO estimator.

Now we want to use the same procedure for slowly time varying channels. Actually we want to know how slow the channel should change if we want to use the same equations. Eq. (5.13) can be written equivalently as

$$\mathbf{y}_i = e^{j2\pi\varepsilon_0(N_g+i(N+N_g))/N}\mathbf{VF}^H\mathbf{X}_{D,i}\mathbf{F}_L\mathbf{h}_i + \mathbf{w}_i \tag{5.21}$$

where $\mathbf{X}_{D,i} = \mathbf{\Phi P}_D + \mathbf{\Psi S}_{D,i}$ and $\mathbf{h}_i = [h_i(0),\, h_i(1),\, ...,\, h_i(L-1)]^T \in C^L$.

Considering two consequent OFDM blocks, we next derive the cross correlation as:

$$\mathbf{R}_{\mathbf{y}_i\mathbf{y}_{i+1}} = E\{\mathbf{y}_i\mathbf{y}_{i+1}^H\} = E\{e^{-j2\pi\varepsilon_0(N+N_g)/N}\mathbf{VF}^H\mathbf{X}_{D,i}\mathbf{F}_L\mathbf{h}_i\mathbf{h}_{i+1}^H\mathbf{F}_L^H\mathbf{X}_{D,(i+1)}^H\mathbf{FV}^H\} + E\{\mathbf{w}_i\mathbf{w}_{i+1}^H\} \tag{5.22}$$

Taking the average both over OFDM symbols and channel and defining $\mathbf{R}_{\mathbf{h}_i\mathbf{h}_{i+1}} = E\{\mathbf{h}_i\mathbf{h}_{i+1}^H\}$, we have

$$\mathbf{R}_{\mathbf{y}_i \mathbf{y}_{i+1}} = e^{-j2\pi\varepsilon_0(N+N_g)/N} E\left\{ \mathbf{V}\mathbf{F}^H (\mathbf{\Phi}\mathbf{P}_D + \mathbf{\Psi}\mathbf{S}_{Di})\mathbf{F}_L \mathbf{R}_{\mathbf{h}_i \mathbf{h}_{i+1}} \mathbf{F}_L^H (\mathbf{\Phi}\mathbf{P}_D + \mathbf{\Psi}\mathbf{S}_{D,(i+1)})^H \mathbf{F}\mathbf{V}^H \right\}$$

(5.23)

This equality follows from the fact that data and channel are statistically independent.

Assuming that channel taps are uncorrelated, and from Jakes' model [2], we have

$$E\{h_{l_1}(t+\Delta t)h_{l_2}^*(t)\} = \sigma_{l_1}^2 r_h(\Delta t)\delta(l_1 - l_2) = \sigma_{l_1}^2 J_0(2\pi f_d \Delta t)\delta(l_1 - l_2)$$

(5.24)

where $J_0(.)$ denotes the zeroth order Bessel function of first kind and f_d is Doppler frequency

in hertz. If we use the Taylor series expansion of $J_0(2\pi f_d \Delta t) = r_0 + r_1 \times (\Delta t) + r_2 \times (\Delta t)^2 + ...$,

since the correlation function is an even function, $r_{2k+1} = 0$, $k = 0,1,...$. The zeroth order Bessel

function of the first order can be expanded as $J_0(2\pi x) \approx 1 - (\pi x)^2$, so when $\Delta t = NT_s$

$$E\{h_{l_k}(t+\Delta t)h_{l_k}^*(t)\} = \sigma_{l_k}^2 J_0(2\pi f_d NT_s) \approx \sigma_{l_k}^2 (1 - (\pi f_d NT_s)^2)$$

(5.25)

where $f_d NT_s$ is the normalized Doppler frequency. When $\pi f_d NT_s < 0.1$ or $f_d NT_s < 0.03$, the

second term can be neglected and then we have

$$\mathbf{R}_{\mathbf{h}_i \mathbf{h}_{i+1}} = diag\{\sigma_0^2,...,\sigma_{L-1}^2\}.$$

Replacing in (5.23), we have

$$\mathbf{R}_{\mathbf{y}_i \mathbf{y}_{i+1}} = e^{-j2\pi\varepsilon_0(N+N_g)/N} \mathbf{V}\mathbf{F}^H \mathbf{\Phi}\mathbf{P}_D \mathbf{F}_L \mathbf{R}_{\mathbf{h}_i \mathbf{h}_{i+1}} \mathbf{F}_L^H \mathbf{P}_D^H \mathbf{\Phi}\mathbf{F}\mathbf{V}^H.$$

(5.26)

Again because matrix $\mathbf{F}^H \mathbf{\Phi}\mathbf{P}_D \mathbf{F}_L \mathbf{R}_{\mathbf{h}_i \mathbf{h}_{i+1}} \mathbf{F}_L^H \mathbf{P}_D^H \mathbf{\Phi}\mathbf{F}$ is Hermitian and $\mathbf{V}$ is diagonal, the diagonal

elements of matrix $\mathbf{V}\mathbf{F}^H \mathbf{\Phi}\mathbf{P}_D \mathbf{F}_L \mathbf{R}_{\mathbf{h}_i \mathbf{h}_{i+1}} \mathbf{F}_L^H \mathbf{P}_D^H \mathbf{\Phi}\mathbf{F}\mathbf{V}^H$ are real and

$$Phase(\mathbf{R}_{\mathbf{y}_i \mathbf{y}_{i+1}}(k,k)) = -2\pi\varepsilon_0(N+N_g)/N, \quad k = 1,...,N.$$

(5.27)

This means as long as $f_d N T_s < 0.03$, our estimator can be used for a time varying channel as well.

5.4 ML Frequency Offset Estimation

In this section, we will develop ML estimation of CFO considering that CSI is unknown for the receiver for the case of superiompsed and nonsuperimosped OFDM symbols.

To use the ML principal, we need to derive the autocorrelation of the received vector as function of tranmistted data, channel and frequency offest. Given the CFO (ε_0), the received symbol vector $\mathbf{y}$,

$$\mathbf{y} = \mathbf{V}\mathbf{F}^H \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{w} \tag{5.28}$$

is approximately Gussian wih zero mean (because $\mathbf{h}$ is CGRV with zero mean and covariance vector $\mathbf{R_{hh}}$) and covariance (autocorrelation) vector $\mathbf{R_{yy}}$, so its distribution is written

$$f(\mathbf{y} \mid \varepsilon_0) = \frac{1}{\pi^N \det(\mathbf{R_{yy}})} \exp(-\mathbf{y}^H \mathbf{R_{yy}^{-1}} \mathbf{y}) \tag{5.29}$$

where $\mathbf{R_{yy}} = E\{\mathbf{y}\mathbf{y}^H \mid \varepsilon_0\}$ (the expectation is over both OFDM symbol and channel). We now consider two different OFDM symbol configuration :

5.4.1 Non-superimposed combination of pilot and data

Defining I_d as the index set of data subcarriers with N_d elements , I_p as the index set of subcarriers reserved for pilot symbols with N_p elements, and I_v as the index set of virtual subcarriers with N_v

elements ($N = N_p + N_d + N_v$), the OFDM symbol's elements selected from a unitary constellation are

$$\mathbf{X}_D(k,k) = \begin{cases} d_k & k \in I_d \\ p_k & k \in I_p \\ 0 & k \in I_v \end{cases}. \tag{5.30}$$

Then

$$\mathbf{R}_{\mathbf{yy}} = \mathbf{V}\mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{D,p}\mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{D,p}^H)\mathbf{F}\mathbf{V}^H + \sigma_w^2 \mathbf{I} \tag{5.31}$$

where $\alpha = \sum_{l=0}^{L-1} \sigma_l^2 / N$ and

$$\mathbf{\Lambda}_d = diag[\lambda_0, \lambda_1, ..., \lambda_{N-1}], \lambda_k = \begin{cases} 1 & k \in I_d \\ 0 & otherwise \end{cases} \tag{5.32}$$

$$\mathbf{X}_{Dp} = diag[\lambda_0, \lambda_1, ..., \lambda_{N-1}], \lambda_k = \begin{cases} p_k & k \in I_p \\ 0 & otherwise \end{cases}. \tag{5.33}$$

In Eq. (5.31), we used the fact that channel and data are independent and the same procedure as Eq. (5.17) has been followed.

$\mathbf{R}_{\mathbf{yy}}$ can be written equivalently as

$$\mathbf{R}_{\mathbf{yy}} = \mathbf{V}\mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp}\mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H + \sigma_w^2 \mathbf{I})\mathbf{F}\mathbf{V}^H \tag{5.34}$$

so we have

$$\det(\mathbf{R}_{\mathbf{yy}}) = \det(\mathbf{V})\det(\mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp}\mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H + \sigma_w^2 \mathbf{I})\mathbf{F})\det(\mathbf{V}^H) =$$
$$\det(\mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp}\mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H + \sigma_w^2 \mathbf{I})\mathbf{F}) \tag{5.35}$$

which shows $\det(\mathbf{R}_{yy})$ is independent of CFO. We drop the terms in (5.29) that are independent of CFO and derive the log likelihood function as

$$\Lambda(\mathbf{y} \mid \varepsilon_0) = -\mathbf{y}^H \mathbf{V} \mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp} \mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H + \sigma_n^2 \mathbf{I})^{-1} \mathbf{F} \mathbf{V}^H \mathbf{y} =$$
$$-\mathbf{y}^H \mathbf{V} \mathbf{G}^{-1} \mathbf{V}^H \mathbf{y} = -\boldsymbol{\beta}^T \mathbf{y}_D^* \mathbf{G}^{-1} \mathbf{y}_D^T \boldsymbol{\beta}^*$$

(5.36)

where $\mathbf{G} = \mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp} \mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H + \sigma_w^2 \mathbf{I})^{-1} \mathbf{F}$, $\boldsymbol{\beta} = [1, e^{j2\pi\varepsilon_0 / N}, ..., e^{j2\pi\varepsilon_0 (N-1)/N}]^T$ and

$$\mathbf{y}_D = diag[y_0, ..., y_{N-1}].$$

(5.37)

So now maximizing the log likelihood function is equivalent to

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} g(\varepsilon_0)$$

(5.38)

where $g(\varepsilon_0) = \mathbf{y}^H \mathbf{V} \mathbf{G}^{-1} \mathbf{V}^H \mathbf{y}$.

If the CFO is the same during K OFDM symbols, and if the delay is tolerable, the K consequent OFDM symbols can be combined for frequency offset estimation. Let $\mathbf{y}_k$ denotes kth received OFDM symbol. Since $\mathbf{y}_k$'s for $k = 1, ..., K$ are independent, the joint pdf function of $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_K)$ is

$$f(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_K \mid \varepsilon_0) = \prod_{k=1}^{K} f(\mathbf{y}_k \mid \varepsilon_0) = \frac{1}{\pi^{KN} \prod_{k=1}^{K} \det(\mathbf{R}_{yy})} \exp(-\sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{V} \mathbf{G}^{-1} \mathbf{V}^H \mathbf{y}_k).$$

(5.39)

So the cost function (5.38) will be

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{V} \mathbf{G}^{-1} \mathbf{V}^H \mathbf{y}_k = \arg\min_{\varepsilon_0} \sum_{k=1}^{K} \boldsymbol{\beta}^T \mathbf{y}_{kD}^* \mathbf{G}^{-1} \mathbf{y}_{kD}^T \boldsymbol{\beta}^* = \arg\min_{\varepsilon_0} \boldsymbol{\beta}^T \mathbf{B} \boldsymbol{\beta}^*$$

(5.40)

where $\mathbf{B} = \sum_{k=1}^{K} \mathbf{y}_{kD}^* \mathbf{G}^{-1} \mathbf{y}_{kD}^T$.

Since the receiver knows the pilots and their positions, provided that the receiver knows $\mathbf{R_h}$

and σ_w^2, matrix $\mathbf{G}^{-1}$ can be pre-computed. The cost function (5.40), can be written as

$$g(\varepsilon_0) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} b_{i,k} e^{j2\pi(i-k)\varepsilon_0/N} \tag{5.41}$$

which is a polynomial of CFO, where $b_{i,k}$ is the (i,k)th entry of $\mathbf{B}$. Since $\mathbf{B}$ is Hermitian, the

cost function can be written as

$$g(\varepsilon_0) = 2\Re\{\sum_{i=0}^{N-1} a_i z^i\} \tag{5.42}$$

$$\text{where } z = e^{-j2\pi\varepsilon_0/N} \text{ and } a_i = \begin{cases} \sum_{j=0}^{N-i-1} b_{j,j+i} & i \neq 0 \\ \frac{1}{2}\sum_{j=0}^{N-1} b_{j,j} & i = 0 \end{cases}. \tag{5.43}$$

We add $(m-1)N$ zeros to the end of sequence $[a_0, a_1, ..., a_{N-1}]$ and perform the mN point DFT,

which yields

$$A(k) = \frac{1}{mN}\sum_{i=0}^{N-1} a_i e^{-j\frac{2\pi i}{N}\frac{k}{m}}, \quad k = 0, 1, ..., mN - 1. \tag{5.44}$$

Let the index of the minimum $\Re(A(k))$ denote $\hat{k}$, therefore the ε_0 that minimizes (5.42) can be

approximated as $\varepsilon_0 = \hat{k}/m$. The value of m should be chosen in according to the amount of

complexity we can afford and accuracy we desire. The larger the m, the better the estimation, but

the higher the complexity.

Remarks

- *No pilot case($N_p =0$)*

In case that there is no pilot in the transmitted symbol ($N_p =0$), the matrix $\mathbf{G}$ can be written as

$\mathbf{G} = \mathbf{F}^H (\alpha \mathbf{\Lambda}_d + \sigma_w^2 \mathbf{I})^{-1} \mathbf{F}$. Thus the cost function can be written as

$$g(\varepsilon_0) = \mathbf{y}^H \mathbf{V} \mathbf{F}^H \mathbf{B} \mathbf{F} \mathbf{V}^H \mathbf{y} \qquad (5.45)$$

where $\mathbf{B} = (\alpha \mathbf{\Lambda}_d + \sigma_w^2 \mathbf{I})^{-1} = diag[b_0, b_1, ..., b_{N-1}]$ and

$$b_k = \begin{cases} \dfrac{1}{\alpha + \sigma_w^2} & k \notin I_v \\[3mm] \dfrac{1}{\sigma_w^2} & k \in I_v \end{cases} . \qquad (5.46)$$

In high SNR $\dfrac{1}{\alpha + \sigma_w^2} << \dfrac{1}{\sigma_w^2}$, hence we can accept $b_k = 0, \quad k \notin I_v$. Denoting the kth column of

the IDFT matrix $\mathbf{F}^H$ with $\mathbf{w}_k$, the cost function (5.45) becomes

$$g(\varepsilon_0) = \mathbf{y}^H \mathbf{V}[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{N-1}] \mathbf{B}[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{N-1}]^H \mathbf{V}^H \mathbf{y} = \frac{1}{\sigma_w^2} \sum_{k \in I_v} \mathbf{y}^H \mathbf{V} \mathbf{w}_k \mathbf{w}_k^H \mathbf{V}^H \mathbf{y}. \quad (5.47)$$

Then the CFO can be estimated as

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \sum_{k \in I_v} \mathbf{y}^H \mathbf{V} \mathbf{w}_k \mathbf{w}_k^H \mathbf{V}^H \mathbf{y} = \arg\min_{\varepsilon_0} \sum_{k \in I_v} \left\| \mathbf{y}^H \mathbf{V} \mathbf{w}_k \right\|^2 = \arg\min_{\varepsilon_0} \sum_{k \in I_v} \mathbf{w}_k^H \mathbf{V}^H \mathbf{y} \mathbf{y}^H \mathbf{V} \mathbf{w}_k$$

$$(5.48)$$

which is the same as the cost function given in [43]. Therefore this CFO estimator in absence of

pilots and in high SNR is equivalent to the one in [43].

- *At least one pilot case($N_p > 0$)*

When pilots exist, $\mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H$ is a circulant matrix of rank L. If $N_p \le L$, $\mathbf{X}_{Dp} \mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H$ has

N_p nonzero columns and rows, the N_p columns are independent. Since $\alpha \mathbf{\Lambda}_d$ is a diagonal matrix,

$\mathbf{A} = \alpha \mathbf{\Lambda}_d + \mathbf{X}_{Dp} \mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H$ has rank N_{p+} N_d. Let the singular value decomposition (SVD) of

$\mathbf{A}$ be denoted as $\mathbf{A} = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$, where the first N_{p+} N_d diagonal elements of $\mathbf{\Lambda}$ are nonzero. Then

$g(\varepsilon_0)$ can be rewritten as

$$g(\varepsilon_0) = \mathbf{y}^H \mathbf{V} \mathbf{E}^H \mathbf{C} \mathbf{E} \mathbf{V}^H \mathbf{y} \qquad (5.49)$$

where $\mathbf{E} = \mathbf{U}\mathbf{F}$ and $\mathbf{C} = (\mathbf{\Lambda} + \sigma_w^2 \mathbf{I})^{-1} = diag\{c_0, c_1, ..., c_{N-1}\}$ and

$$c_k = \begin{cases} \dfrac{1}{\lambda_{k,k} + \sigma_w^2} & k = 0, ..., N_d + N_p - 1 \\[3mm] \dfrac{1}{\sigma_w^2} & otherwise \end{cases} \qquad (5.50)$$

Again in high SNR regime, $\dfrac{1}{\lambda_{k,k} + \sigma_w^2} \ll \dfrac{1}{\sigma_w^2}$. Let $\mathbf{e}_k$ denotes the kth column of matrix $\mathbf{E}^H$.

Eq. (5.49) becomes

$$g(\varepsilon_0) = \frac{1}{\sigma_w^2} \sum_{k=N_d+N_p}^{N-1} \mathbf{y}^H \mathbf{V} \mathbf{e}_k \mathbf{e}_k^H \mathbf{V}^H \mathbf{y} . \qquad (5.51)$$

Since there are $N - N_p - N_d = N_v$ terms in (5.51), it acts as a subspace based frequency offset

estimator with N_v virtual carriers. Thus when $N_p \le L$, the pilots can not improve the

performance of the frequency offset estimator in high SNR. However, improvement is possible

in low SNR. When $N_p > L$, $\mathbf{X}_{Dp} \mathbf{F}_L \mathbf{R_h} \mathbf{F}_L^H \mathbf{X}_{Dp}^H$ is of rank L and $\mathbf{A}$ has rank $L + N_d$. Therefore

there are $N - L - N_d = N_v + N_p - L$ terms in cost function $g(\varepsilon_0)$ and the CFO estimator is

the same as estimator in [43] with $N_v + N_p - L$ virtual carriers. It means with taking the

advantages of pilots when $N_p > L$, this estimator can outperforms the one in [43] in high SNR

regime.

It can be seen in (5.42) that the cost function $g(\varepsilon_0)$ is periodic with the period N, which

means that the range of frequency offset estimator is wider than and not limited to the half of the

frequency separation between adjacent subcarriers. It means this estimator doesn't divide the

frequency offset into an integer part and a fraction part.

Next we will show this estimator is unbiased, and then will find the Cramer-Rao bound

for it. [45] shows that the expectation of the estimate in high SNR is approximated as

$$E\{\hat{\varepsilon}_0\} = \varepsilon_0 - \frac{E\{g'(\varepsilon_0)\}}{E\{g''(\varepsilon_0)\}} \tag{5.52}$$

where $g'(\varepsilon_0)$ and $g''(\varepsilon_0)$ are the first and second derivatives of $g(\varepsilon_0)$. From [46], we know the

derivative of a quadratic form $\mathbf{x}^H \mathbf{C} \mathbf{x}$ provided that $\mathbf{C}$ is Hermitian can be written as

$$\partial(\mathbf{x}^H \mathbf{C} \mathbf{x}) = 2\Re(\mathbf{x}^H \mathbf{C} \, \partial \mathbf{x}) \tag{5.53}$$

therefore

$$\frac{\partial g(\varepsilon_0)}{\partial \varepsilon_0} = 2\Re\{\mathbf{y}^H \mathbf{V} \mathbf{G}^{-1} \frac{\partial(\mathbf{V}^H \mathbf{y})}{\partial \varepsilon_0}\} \tag{5.54}$$

and since $\dfrac{\partial(\mathbf{V}^H \mathbf{y})}{\partial \varepsilon_0} = (-j\dfrac{2\pi}{N})\mathbf{M}\mathbf{V}^H \mathbf{y}$ where $\mathbf{M} = diag\{0,1,...,N-1\}$

$$\frac{\partial g(\varepsilon_0)}{\partial \varepsilon_0} = 2\Re\{\mathbf{y}^H \mathbf{V} \mathbf{G}^{-1}(-j\frac{2\pi}{N})\mathbf{M}\mathbf{V}^H \mathbf{y}\} = \mathbf{y}^H \mathbf{V} \mathbf{G}^{-1}(-j\frac{2\pi}{N})\mathbf{M}\mathbf{V}^H \mathbf{y} +$$

$$(\mathbf{y}^H \mathbf{V} \mathbf{G}^{-1}(-j\frac{2\pi}{N})\mathbf{M}\mathbf{V}^H \mathbf{y})^H = j\frac{2\pi}{N}\mathbf{y}^H \mathbf{V}(\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{V}^H \mathbf{y}$$

$$(5.55)$$

if $\mathbf{x}$ is a complex Gaussian vector with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$, and $\mathbf{A}$ is a matrix,

we have $E\{\mathbf{x}^H \mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{S}) + \mathbf{m}^H \mathbf{A}\mathbf{m}$ [46] so

$$E\{g'(\varepsilon_0)\} = E\{j\frac{2\pi}{N}\mathbf{y}^H \mathbf{V}(\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{V}^H \mathbf{y}\} = j\frac{2\pi}{N}tr(\mathbf{V}(\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{V}^H \mathbf{R}_y) =$$

$$j\frac{2\pi}{N}tr(\mathbf{V}(\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{G}\mathbf{V}^H) = tr((\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{G}) = tr(\mathbf{M}) - tr(\mathbf{G}^{-1}\mathbf{M}\mathbf{G}) = 0$$

$$(5.56)$$

The last equalities come from the trace property $tr(\mathbf{B}\mathbf{C}) = tr(\mathbf{C}\mathbf{B})$.

From (5.52) and (5.56), we have $E\{\hat{\varepsilon}_0\} = \varepsilon_0$, which means the estimator is unbiased.

For any unbiased estimation $\hat{\varepsilon}_0$, from the Cramer-Rao lower bound theorem we know

$$\text{var}\{\hat{\varepsilon}_0\} \geq CRB \qquad\qquad (5.57)$$

where CRB for a scalar parameter is defined as

$$CRB = -\frac{1}{E\{\frac{\partial^2 \Lambda(\mathbf{y}, \varepsilon_0)}{\partial \varepsilon_0^2}\}} = \frac{1}{E\{g''(\varepsilon_0)\}} \qquad\qquad (5.58)$$

Now we will use the first derivative of cost function to derive the second derivative. We have

$$\frac{\partial g(\varepsilon_0)}{\partial \varepsilon_0} = j\frac{2\pi}{N}\mathbf{y}^H \mathbf{V}\mathbf{L}\mathbf{V}^H \mathbf{y}, \text{ where}$$

$$\mathbf{L} = (\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}). \qquad\qquad (5.59)$$

Since $\mathbf{L}^H = -\mathbf{L}$, we have

$$g''(\varepsilon_0) = \frac{\partial g'(\varepsilon_0)}{\partial \varepsilon_0} = j\frac{2\pi}{N} 2\Im(\mathbf{y}^H \mathbf{VL}\frac{\partial(\mathbf{V}^H\mathbf{y})}{\partial \varepsilon_0}) = j\frac{2\pi}{N} 2\Im(\mathbf{y}^H \mathbf{VL}(-j\frac{2\pi}{N})\mathbf{MV}^H\mathbf{y}) =$$

$$j\frac{2\pi}{N}(\mathbf{y}^H \mathbf{VL}(-j\frac{2\pi}{N})\mathbf{MV}^H\mathbf{y} - (\mathbf{y}^H \mathbf{VL}(-j\frac{2\pi}{N})\mathbf{MV}^H\mathbf{y})^H = \qquad (5.60)$$

$$(\frac{2\pi}{N})^2(\mathbf{y}^H \mathbf{VLMV}^H\mathbf{y} - \mathbf{y}^H \mathbf{VMLV}^H\mathbf{y}) = (\frac{2\pi}{N})^2\mathbf{y}^H \mathbf{V}(\mathbf{LM} - \mathbf{ML})\mathbf{V}^H\mathbf{y}$$

and eventually using (5.59), we will have

$$g''(\varepsilon_0) = (\frac{2\pi}{N})^2\mathbf{y}^H \mathbf{V}(2\mathbf{MG}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2)\mathbf{V}^H\mathbf{y}. \qquad (5.61)$$

Therefore,

$$E\{g''(\varepsilon_0)\} = E\{(\frac{2\pi}{N})^2\mathbf{y}^H \mathbf{V}(2\mathbf{MG}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2)\mathbf{V}^H\mathbf{y}\} =$$

$$(\frac{2\pi}{N})^2 tr(\mathbf{V}(2\mathbf{MG}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2)\mathbf{V}^H\mathbf{R}_y) = (\frac{2\pi}{N})^2 tr(\mathbf{V}(2\mathbf{MG}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2)\mathbf{GV}^H)$$

$$= tr((2\mathbf{MG}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2)\mathbf{G}) = \frac{8\pi^2}{N^2} tr(\mathbf{MG}^{-1}\mathbf{MG} - \mathbf{M}^2)$$

$$(5.62)$$

so

$$\mathrm{var}\{\hat{\varepsilon}_0\} \geq CRB = \frac{1}{E\{g''(\varepsilon_0)\}} = \frac{N^2}{8\pi^2} \frac{1}{tr(\mathbf{MG}^{-1}\mathbf{MG} - \mathbf{M}^2)}. \qquad (5.63)$$

When there are neither pilots nor virtual carriers, the cost function (5.38) can be written as

$$g(\varepsilon_0) = (\alpha + \sigma_w^2)^{-1}\mathbf{y}^H \mathbf{VF}^H \mathbf{FV}^H\mathbf{y} = (\alpha + \sigma_w^2)^{-1}\mathbf{y}^H\mathbf{y},$$ which is independents of CFO.

Therefore, in order to have an estimator of CFO, we need to either have embedded pilots or virtual carriers.

5.4.2 Superimposed data and pilot

When data and pilot are superimposed ($\mathbf{X}_D = \mathbf{\Phi}\mathbf{P}_D + \mathbf{\Psi}\mathbf{S}_D$), we have

$$\mathbf{R_{yy}} = \mathbf{V}\mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi})\mathbf{F}\mathbf{V}^H + \sigma_w^2\mathbf{I} \tag{5.64}$$

or equivalently

$$\mathbf{R_{yy}} = \mathbf{V}\mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})\mathbf{F}\mathbf{V}^H. \tag{5.65}$$

Again since

$$\det(\mathbf{R_{yy}}) = \det(\mathbf{V})\det(\mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})\mathbf{F})\det(\mathbf{V}^H) =$$
$$\det(\mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})\mathbf{F}) \tag{5.66}$$

is independent of CFO, the log likelihood function is

$$\Lambda(\mathbf{y}\,|\,\varepsilon_0) = -\mathbf{y}^H\mathbf{V}\mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})^{-1}\mathbf{F}\mathbf{V}^H\mathbf{y} =$$
$$-\mathbf{y}^H\mathbf{V}\mathbf{G}^{-1}\mathbf{V}^H\mathbf{y} = -\boldsymbol{\beta}^T\mathbf{y}_D^*\mathbf{G}^{-1}\mathbf{y}_D^T\boldsymbol{\beta}^* \tag{5.67}$$

where $\mathbf{G} = \mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})^{-1}\mathbf{F}$ and the other vectors have the same definition as the previous section.

Again if the CFO is the same during *K* OFDM symbols, and if the delay is tolerable, the *K* consequent OFDM symbols can be combined for frequency offset estimation. The same sort of equation are true in superimposed case provided that $\mathbf{G} = \mathbf{F}^H(\alpha\mathbf{\Psi}^2 + \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R_h}\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi} + \sigma_w^2\mathbf{I})^{-1}\mathbf{F}$. Equations (5.40) to (5.44) are applicable to the superimposed case with the new matrix $\mathbf{G}$.

5.5 Superimposed training aided Carrier Frequency Offset Estimation by exploiting the correlation between different received data samples

The correlation between adjacent subcarriers can be exploits in superimposed OFDM for the purpose of CFO estimation. We will investigate this idea in this section.

For the pre-DFT data at the receiver, we have

$$
y(n) = e^{\frac{j2\pi\varepsilon_0 n}{N}} \sum_{l=0}^{L-1} h_l(n)x(n-l) + w(n) = \frac{1}{\sqrt{N}} e^{\frac{j2\pi\varepsilon_0 n}{N}} \sum_{l=0}^{L-1} h_l(n) \sum_{k=0}^{N-1} X_k e^{-\frac{j2\pi k(n-l)}{N}} + w(n).
$$

$$(5.68)$$

The correlation between $y(n_1)$ and $y(n_2)$, using Eq. (5.68) and (5.24), can be written as

$$
r_{yy}(n_1, n_2) = E\{y(n_1)y^*(n_2)\} = \frac{1}{N} e^{\frac{j2\pi\varepsilon_0 (n_1-n_2)}{N}} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} E\{X_{k_1} X_{k_2}^*\} e^{j2\pi k_1 n_1 / N} e^{-j2\pi k_2 n_2 / N}
$$

$$
\times r_h((n_1-n_2)T_s) \sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi(k_1-k_2)l/N} + \sigma_w^2 \delta(n_1-n_2)
$$

$$(5.69)$$

Assuming uncorrelated data symbols

$$
r_{yy}(n_1, n_2) = E\{y(n_1)y^*(n_2)\} = \frac{1}{N} e^{\frac{j2\pi\varepsilon_0 (n_1-n_2)}{N}} \sum_{k=0}^{N-1} E\{|X_k|^2\} e^{\frac{j2\pi k(n_1-n_2)}{N}}
$$

$$
\times r_h((n_1-n_2)T_s) \sum_{l=0}^{L-1} \sigma_l^2 + \sigma_w^2 \delta(n_1-n_2)
$$

$$. (5.70)$$

In the case that data is from a unitary constellation so that $E\{|X_k|^2\}$ is the same for all the subcarriers, $\sum_{k=0}^{N-1} E\{|X_k|^2\} e^{\frac{j2\pi k(n_1-n_2)}{N}} = 0, n_1 \neq n_2$; which means the correlation between subcarriers can't be used for this CFO estimation. Still unitary constellations can be used for the

superimposed case provided that the total power for each subcarrier is not the same for different subcarriers. It means that if both data and pilot are drawn from a unitary constellation, pilots shouldn't be superimposed on all subcarriers. Assuming that the aforementioned condition is true, for the superimposed OFDM data

$$E\{|X_k|^2\} = \varphi_k + \phi_k, \quad k = 0, 1, ..., N-1 \tag{5.71}$$

and from (5.24), we know

$$r_h((n_1 - n_2)T_s) = J_0(2\pi f_d T_s(n_1 - n_2)) \approx 1 - (\pi f_d T_s(n_1 - n_2))^2 \tag{5.72}$$

so when $\pi f_d N T_s < 0.1$, $r_h((n_1 - n_2)T_s) \approx 1$ and we have

$$r_{yy}(n_1, n_2) = E\{y(n_1)y^*(n_2)\} = \alpha e^{\frac{j2\pi\varepsilon_0(n_1-n_2)}{N}} \sum_{k=0}^{N-1}(\varphi_k + \phi_k)e^{\frac{j2\pi k(n_1-n_2)}{N}} + \sigma_w^2\delta(n_1 - n_2)$$

$$\tag{5.73}$$

Therefore the autocorrelation matrix $\mathbf{R_{yy}}$ of the received signal $\mathbf{y}$, where $\mathbf{y} = [y(0), y(1), ..., y(N-1)]^T$, can be written as

$$\mathbf{R_{yy}} = \alpha\mathbf{V}\mathbf{F}^H\mathbf{\Omega}\mathbf{F}\mathbf{V}^H + \sigma_w^2\mathbf{I} = \mathbf{V}(\alpha\mathbf{F}^H\mathbf{\Omega}\mathbf{F} + \sigma_w^2\mathbf{I})\mathbf{V}^H \tag{5.74}$$

where $\mathbf{\Omega} = diag(\Omega_0, ..., \Omega_{N-1})$, $\Omega_k = \varphi_k + \phi_k$

The pdf of $\mathbf{y}$ conditional on the carrier offset is therefore

$$f(\mathbf{y}|\varepsilon_0) = \frac{1}{\pi^N \det(\mathbf{R_{yy}})}\exp(-\mathbf{y}^H\mathbf{R_{yy}}^{-1}\mathbf{y}) \tag{5.75}$$

and as we know $\det(\mathbf{R_{yy}}) = \det(\mathbf{V})\det(\alpha\mathbf{F}^H\mathbf{\Omega F} + \sigma_w^2\mathbf{I})\det(\mathbf{V})$ which is independent of carrier

offset. Ignoring terms that are independent of carrier offset, we find that maximizing the log

likelihood function is equivalent to minimizing the following cost function

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \mathbf{y}^H \mathbf{R_{yy}}^{-1}\mathbf{y} = \arg\min_{\varepsilon_0} \mathbf{y}^H \mathbf{V}(\alpha\mathbf{F}^H\mathbf{\Omega F} + \sigma_w^2\mathbf{I})^{-1}\mathbf{V}^H\mathbf{y} . \quad (5.76)$$

Solving (5.76) is equivalent to solving

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \mathbf{\beta}^T \mathbf{y}_D^*(\alpha\mathbf{F}^H\mathbf{\Omega F} + \sigma_w^2\mathbf{I})^{-1}\mathbf{y}_D^T\mathbf{\beta}^* \quad (5.77)$$

where $\mathbf{\beta} = [1, e^{j2\pi\varepsilon_0/N}, ..., e^{j2\pi\varepsilon_0(N-1)/N}]^T$ and $\mathbf{y}_D = diag[y_0,..., y_{N-1}]$.

For the case that $\pi f_d NT_s < 0.1$, the cost function remains the same which means the estimator

(5.77) is robust for normalized Doppler frequencies less than 0.03.

Again if the CFO can be supposed to be the same for K different OFDM symbols, assuming that

these symbols are independent, the joint pdf is the multiplication of the individual pdfs so that

the cost function can be written as

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{R_{yy}}^{-1}\mathbf{y}_k = \arg\min_{\varepsilon_0} \sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{V}(\alpha\mathbf{F}^H\mathbf{\Omega F} + \sigma_w^2\mathbf{I})^{-1}\mathbf{V}^H\mathbf{y}_k \quad (5.78)$$

Or equivalently

$$\hat{\varepsilon}_0 = \arg\min_{\varepsilon_0} \mathbf{\beta}^T (\sum_{k=1}^{K} (\mathbf{y}_k^*)_D(\alpha\mathbf{F}^H\mathbf{\Omega F} + \sigma_w^2\mathbf{I})^{-1}(\mathbf{y}_k^T)_D)\mathbf{\beta}^* = \arg\min_{\varepsilon_0} \mathbf{\beta}^T \mathbf{B}\mathbf{\beta}^* . \quad (5.79)$$

Again we can write the cost function as a polynomial

$$\mathbf{\beta}^T \mathbf{B}\mathbf{\beta}^* = g(\varepsilon_0) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} b_{i,k} e^{j2\pi(i-k)\varepsilon_0/N} \quad (5.80)$$

where $b_{i,k}$ is the *(i,k)*th entry of $\mathbf{B}$. Following the same equations as (5.42) to (5.44), the CFO can be estimated by an FFT based algorithm.

In order to find the CRB, we need to compute the second derivative of the cost function. Following the same procedure as in (5.54) to (5.63), we have

$$\text{var}\{\hat{\varepsilon}_0\} \ge CRB = \frac{1}{E\{g''(\varepsilon_0)\}} \rightarrow \text{var}\{\hat{\varepsilon}_0\} \ge \frac{N^2}{8\pi^2} \frac{1}{Ktr(\mathbf{MG}^{-1}\mathbf{MG} - \mathbf{M}^2)} \qquad (5.81)$$

where $\mathbf{G} = \alpha \mathbf{F}^H \mathbf{\Omega} \mathbf{F} + \sigma_w^2 \mathbf{I}$.

Note that if all the subcarriers have the same amount of power we will have $\mathbf{G} = \beta \mathbf{I}$, where β is a real number, and therefore $tr(\mathbf{MG}^{-1}\mathbf{MG} - \mathbf{M}^2) = 0$. It means the variance of our estimator will be infinity and it is the same as the fact that the cost function will be independent of CFO in this case.

The place of pilots and power distribution of data and pilot can be optimized to minimize the CRB subject to the power constraint $\sum_{k=0}^{N-1} \varphi_k = D$ and $\sum_{k \in I_p} \phi_k = P$ where I_p is the index of N_p subcarriers with superimposed pilot ($\phi_k \ne 0$) and P and D are the total power on pilots and data symbols respectively. From (5.81), minimizing CRB is equivalent to maximizing $tr(\mathbf{MG}^{-1}\mathbf{MG} - \mathbf{M}^2)$ and consequently, it is equivalent to maximize $tr(\mathbf{MG}^{-1}\mathbf{MG})$

$$(I_p, \varphi_k, \phi_k) = \underset{\substack{\sum_{k=0}^{N-1} \varphi_k = D, \sum_{k \in I_p} \phi_k = P}}{\arg\max} \quad tr(\mathbf{MG}^{-1}\mathbf{MG}). \qquad (5.82)$$

We can rewrite matrix $\mathbf{G}$ as $\mathbf{G} = \mathbf{F}^H (\alpha \mathbf{\Omega} + \sigma_w^2 \mathbf{I}) \mathbf{F}$ and so $\mathbf{G}^{-1} = \mathbf{F}^H (\alpha \mathbf{\Omega} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{F}$. Therefore we have

$$g_{r,s} = [\mathbf{G}]_{r,s} = \sum_{k=0}^{N-1} (\Omega_k + \sigma_w^2) e^{j\frac{2\pi(r-s)k}{N}} = g^*_{s,r} \tag{5.83}$$

$$f_{r,s} = [\mathbf{G}^{-1}]_{r,s} = \sum_{k=0}^{N-1} (\Omega_k + \sigma_w^2)^{-1} e^{j\frac{2\pi(r-s)k}{N}} = f^*_{s,r} . \tag{5.84}$$

We defined $\mathbf{M} = diag\{a_0, ..., a_{N-1}\}$, so

$$tr(\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G}) = \sum_{r=0}^{N-1}\sum_{s=0}^{N-1} a_r a_s g_{r,s} f_{s,r} = 2\Re\{\sum_r \sum_{s \geq r} a_r a_s g_{r,s} f_{s,r}\} \tag{5.85}$$

Note that

$$[\mathbf{G}]_{r,s} = \begin{cases} \sum_{k=0}^{N-1}(\Omega_k + \sigma_w^2) = N\sigma_w^2 + D + P, & r = s \\ \sum_{k \in I_p} \phi_k e^{j\frac{2\pi(r-s)k}{N}} + \sum_{k=0}^{N-1} \varphi_k e^{j\frac{2\pi(r-s)k}{N}}, & r \neq s \end{cases} \tag{5.86}$$

which means the elements of sum when $r = s$ are positive.

Therefore to maximize the trace, we need

$$\Re\{g_{r,s} f_{s,r}\} \geq 0, \quad 0 \leq r, s \leq N-1, \quad s > r . \tag{5.87}$$

If

$$\varphi_k = cons., \quad k \notin I_p \text{ and } \varphi_k + \phi_k = cons. \quad k \in I_p \tag{5.88}$$

$$z = N / N_p \in Z \quad I_p = \{z_0 + k'z, k' = 0, 1, ..., N_p - 1\}, z_0 \in \{0, 1, ..., z-1\} \tag{5.89}$$

we have

$$\sum_{k \in I_p} \phi_k e^{j\frac{2\pi(r-s)k}{N}} + \sum_{k=0}^{N-1} \varphi_k e^{j\frac{2\pi(r-s)k}{N}} \geq 0, \quad 0 \leq r, s \leq N-1 \tag{5.90}$$

so the condition (5.87) is satisfied.

(5.88) means that the total power on the subcarriers with superimposed pilots should be equal and also the power on subcarriers without pilot should be equal. (5.89) means that the superimposed pilots should be equispace. In the separate training scheme, (5.88) and (5.89) agree with the equipower and equispace conditions. Note that (5.88) doesn't assign the power allocated to pilots but the total power allocated to each subcarrier.

5.6 Iterative Joint channel and CFO estimation and data detection for superimposed training aided OFDM systems

Starting form (5.28), since the noise vector is i.i.d Gaussian, the ML joint estimators of the CSI and CFO and transmitted symbols are given by

$$(\hat{\mathbf{h}}, \hat{\mathbf{S}}_D, \hat{\varepsilon}_0) = \underset{\mathbf{h} \in C^L, \mathbf{S}_D \in Q^N, \varepsilon_0 \in R}{\arg \min} \left\| \mathbf{y} - \mathbf{V}\mathbf{F}^H \mathbf{X}_D \mathbf{F}_L \mathbf{h} \right\|^2 \qquad (5.91)$$

The minimization in (5.91) is a complex LS problem for $\hat{\mathbf{h}}$ and $\hat{\varepsilon}_0$ and an integer LS problem for $\hat{\mathbf{S}}_D$. Given $\hat{\varepsilon}_0$ and $\hat{\mathbf{S}}_D$ (we assume that $\hat{\varepsilon}_0 = \varepsilon_0$ and $\hat{\mathbf{S}}_D = \mathbf{S}_D$), the channel response that minimizes (5.91) is given by the LS estimate

$$\hat{\mathbf{h}} = \left(\mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L \right)^{-1} \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{F}\mathbf{V}^H \mathbf{y} \qquad (5.92)$$

Substituting (5.92) into (5.91), we obtain

$$(\hat{\mathbf{S}}_D, \hat{\varepsilon}_0) = \underset{\mathbf{S}_D \in Q^N, \varepsilon_0 \in R}{\arg \min} \left\| \mathbf{y} - \mathbf{V}\mathbf{F}^H \mathbf{X}_D \mathbf{F}_L \hat{\mathbf{h}} \right\|^2 \qquad (5.93)$$

Denoting $\mathbf{A} = \mathbf{y} - \mathbf{V}\mathbf{F}^H \mathbf{X}_D \mathbf{F}_L \hat{\mathbf{h}}$, we have

$$A = y - VF^H X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H FV^H y) =$$
$$\left[I - VF^H X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H FV^H \right] y =$$
$$VF^H \left[I - X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H \right] FV^H y$$

(5.94)

Then

$$A^H A = \left\| y - VF^H X_D F_L \hat{h} \right\|_2^2 = y^H VF^H \left[I - X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H \right] \times$$
$$\left[I - X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H \right] FV^H y = y^H VF^H \left[I - X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H \right] \times$$
$$FV^H y$$

(5.95)

So the decision function for data and CFO detection will be,

$$(\hat{S}_D, \hat{\varepsilon}_0) = \underset{S_D \in Q^N, \varepsilon_0 \in R}{\arg\min} \left\| y - VF^H X_D F_L \hat{h} \right\|^2 =$$
$$\underset{S_D \in Q^N, \varepsilon_0 \in R}{\arg\min} y^H VF^H \left[I - X_D F_L \left((X_D F_L)^H X_D F_L \right)^{-1} F_L^H X_D^H \right] \times FV^H y$$

(5.96)

From this decision rule, we can use the iteration algorithm as follow

▪ Initialization:

$$X_D^0 = \Phi P_D$$

(5.97)

▪ Recursion: for $i = 1,...,I$

$$\hat{\varepsilon}_0^i = \underset{\varepsilon_0 \in R}{\arg\min} \, y^H VF^H \left[I - X_D^{i-1} F_L \left((X_D^{i-1} F_L)^H X_D^{i-1} F_L \right)^{-1} F_L^H X_D^{i-1^H} \right] FV^H y$$

(5.98)

$$\mathbf{V}^i = diag\ (1, e^{j2\pi\hat{\varepsilon}_0^i/N}, ..., e^{j2\pi\hat{\varepsilon}_0^i(N-1)/N}) \qquad (5.99)$$

$$\hat{\mathbf{h}}^i = \left(\mathbf{F}_L^H (\mathbf{X}_D^{i-1})^H \mathbf{X}_D^{i-1}\mathbf{F}_L\right)^{-1}\mathbf{F}_L^H (\mathbf{X}_D^{i-1})^H \mathbf{F}(\mathbf{V}^i)^H \mathbf{y} \qquad (5.100)$$

$$\hat{\mathbf{H}}_D^i = diag\{\mathbf{F}_L\hat{\mathbf{h}}^i\} \qquad (5.101)$$

$$\mathbf{Y}^i = \mathbf{F}(\mathbf{V}^i)^H \mathbf{y} \qquad (5.102)$$

$$\hat{\mathbf{S}}^i = M_Q(\boldsymbol{\Psi}^{-1}\left[(\mathbf{H}_D^i)^{-1}\mathbf{Y}^i - \boldsymbol{\Phi}\mathbf{P}_D\right]) \qquad (5.103)$$

$$\hat{\mathbf{X}}_D^i = \boldsymbol{\Phi}\mathbf{P}_D + \boldsymbol{\Psi}\hat{\mathbf{S}}_D^i \qquad (5.104)$$

5.7 Simulation and Results

Simulation results are given for the proposed CFO estimators. In first simulation, we compared the performance of the CFO estimators given in section 5.2. An OFDM system with 32 subcarriers has been developed and frequency selective channel with exponential power decaying profile have been considered. The length of CP is *L+3* where *L* is the maximum length of CIRs and it is considered 4 in our simulation. CFO is assumed constant over *25* blocks of consequent OFDM symbols [*M=25* in (5.6)]. Data symbols have been drawn form BPSK constellation.
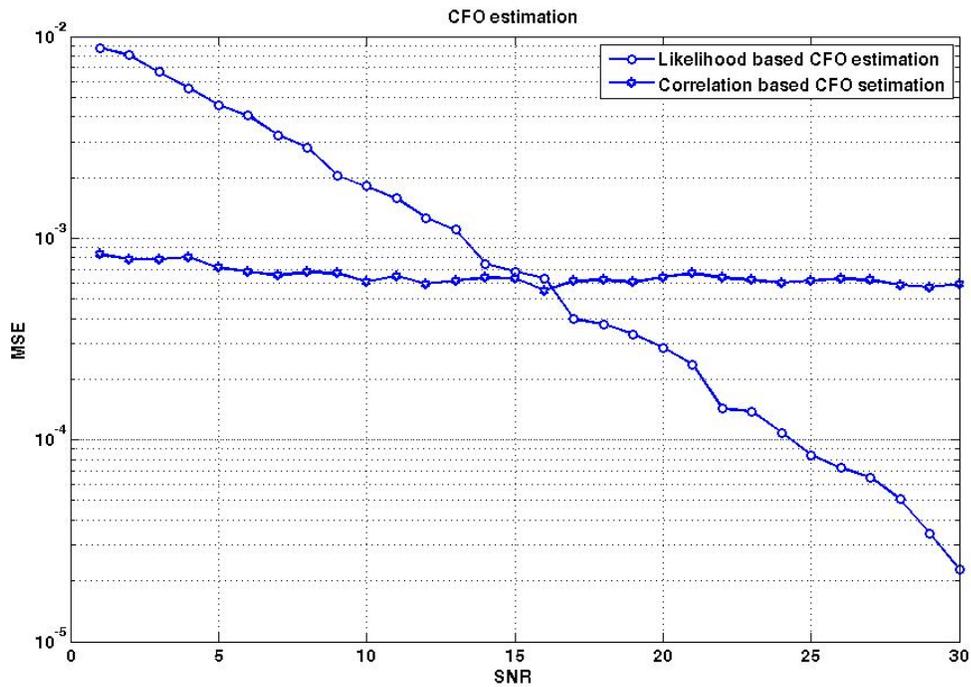
Figure 5-1: MSE comparison of CFO estimators presented in (5.5) and (5.9)

Figure 5-1 shows the mean square error (MSE) of CFO estimation which is defined as

$$MSE = (\frac{1}{M}\sum_{i=1}^{M}(\hat{\varepsilon}_0^i - \varepsilon_0)^2) \qquad (5.105)$$

where M is the number of Monte Carlo runs and $\varepsilon_0 = 0.15$. As it can be seen, correlation based CFO estimator [Eq. (5.5)] shows a good performance over the whole range of SNR. At the expense of insertion of extra 3 CP, the likelihood base CFO estimator [Eq. (5.9)] shows even less MSE in high SNR. Note that estimator in (5.9) needs just one OFDM symbol block to work.
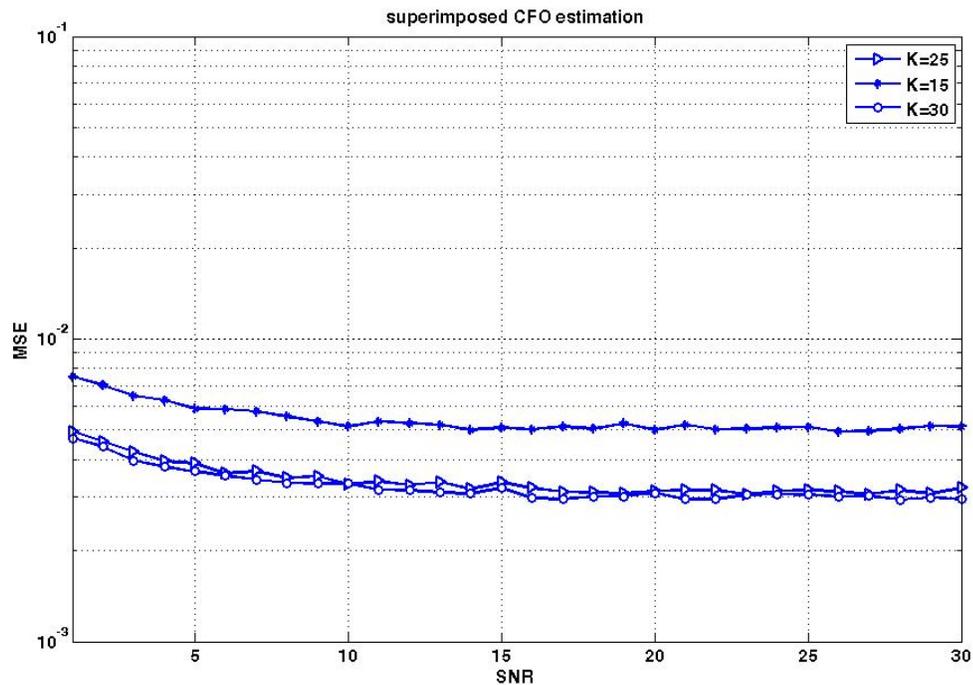
Figure 5-2: MSE vs. SNR for the CFO estimator in superimposed OFDM presented in (5.20) for different values of *K*

Figure 5-2 shows MSE of the estimator presented in (5.20) as a function SNR for three different values of *K*. A superimposed OFDM system with 64 subcarriers and $\varepsilon_0 = 0.15$ has been considered in next four simulation results. 16 out of 64 subcarriers have superimposed pilots for the first one. Superimposed pilots are equispaced and equipower. As it can be seen, increase of *K* form 15 to 25 improves the MSE of estimator about 2 dB but further increase doesn't do more. Figure 5-3 shows the effect of number of superimposed pilots on accuracy of estimator. As it can be increase form 8 pilots to 16 pilots makes the MSE 4 dB less. Figure 5-4 represents the MSE as a function of *K*. Simulation results in the figure suggests that a fair amount of *K (K=25)* should be considered for estimation but increase in *K* after that doesn't improve the MSE significantly.
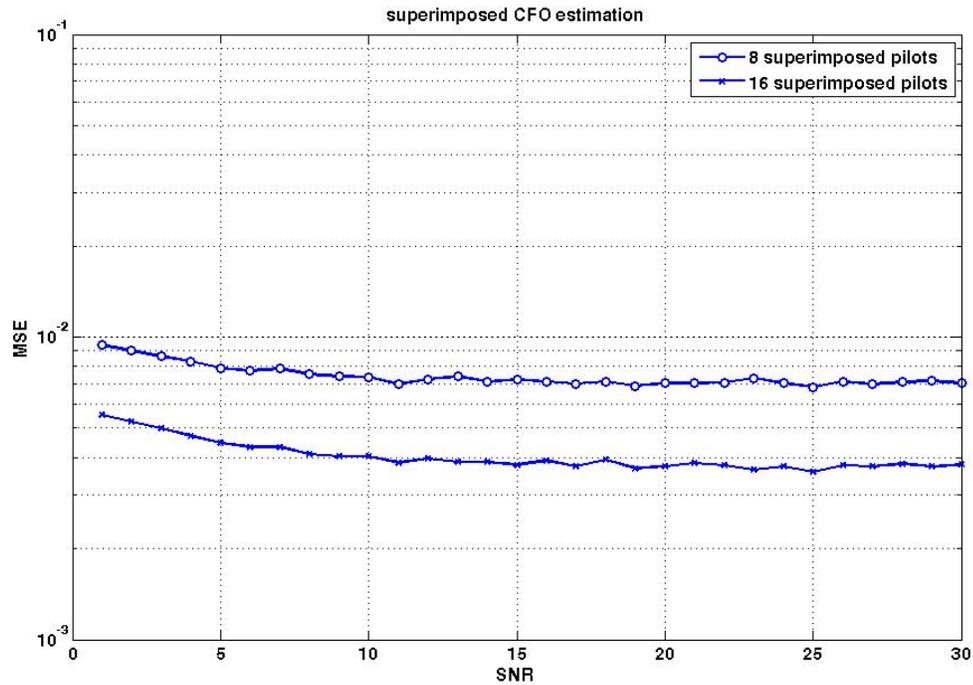
Figure 5-3: MSE comparison of CFO estimator presented in (5.20) for two different amount of superimposed pilots

Figure 5-5 shows the MSE of the estimator in (5.20) for two different channels. One of the channel is static (f_d=0) and for the other one $f_d NT_s = 0.0288 < 0.03$ which means it meets the requirement given in (5.25). Simulation results shows that the degradation in performance is insignificant (about 3 dB) as long as $f_d NT_s < 0.03$.
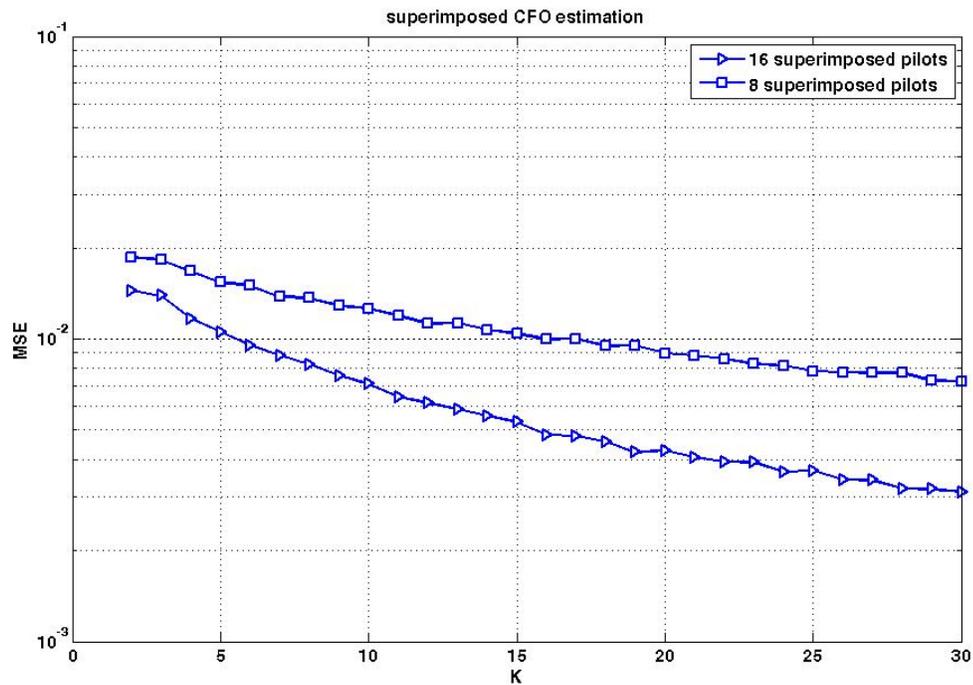
Figure 5-4: MSE vs. *K* for two different amount of superimposed pilots for CFO estimator in (5.20)

Next simulation is done for the ML CFO estimator presented in Section 5.4. An OFDM system with 64 subcarriers and $\varepsilon_0 = 0.25$ has been implemented. It has 16 VCs and channel's model is the same as previous simulations. Figure 5-6 compares the MSE of CFO estimator (denoted by ML) and CRB for different number of pilots. CRB of estimator with 4 pilots is almost 11 dB better than the one with no pilots. This difference decreases to 5 dB when the number of pilots changes from 4 to 8 and reduces to 3 dB when number of pilots changes from 8 to 16. The true MSE of ML estimator approaches the CRB in high SNR for every number of pilots. Figure 5-7 shows the effect of the number of the blocks (*K*) on the MSE of estimator with 4 pilots and 16 VCS. The use of *K=8* blocks yields a 2 dB gain over the use of *K=4* blocks and a 7 dB gain over the use of *K=2* blocks at SNR = 25 dB. Note that the CRB for (5.39) using *K*
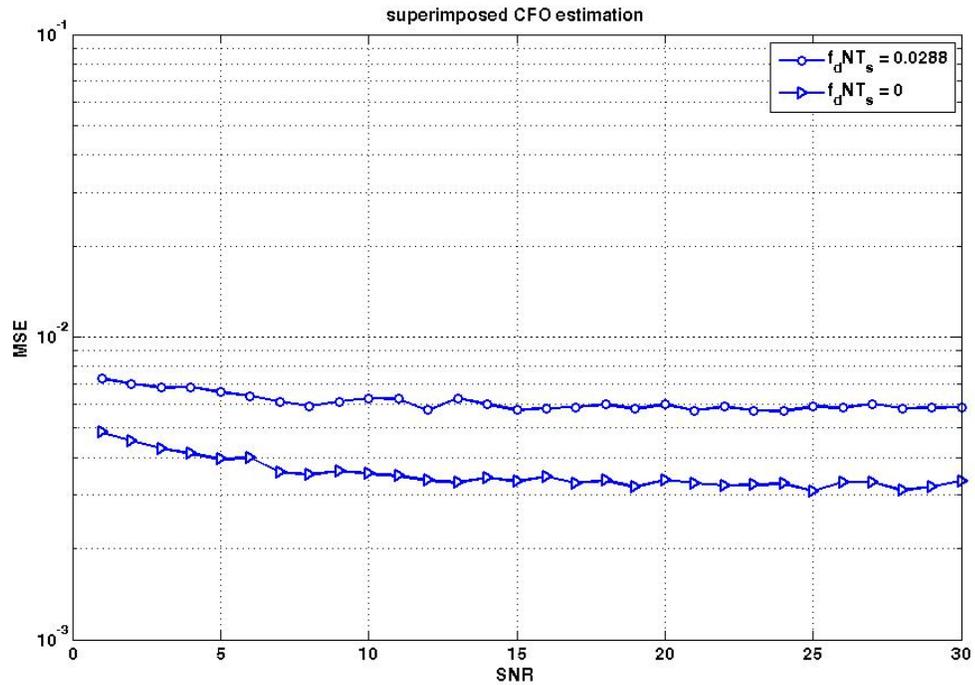
Figure 5-5: MSE of CFO estimator for two channels with different Doppler shift

OFDM blocks can be readily obtained as *1/K* times of (5.63). The gap between the MSE and the CRB decreases with the increase of *K*. This gap is almost 1 dB when *K=8* while it is about 2 dB when *K=4* at SNR=25 dB.
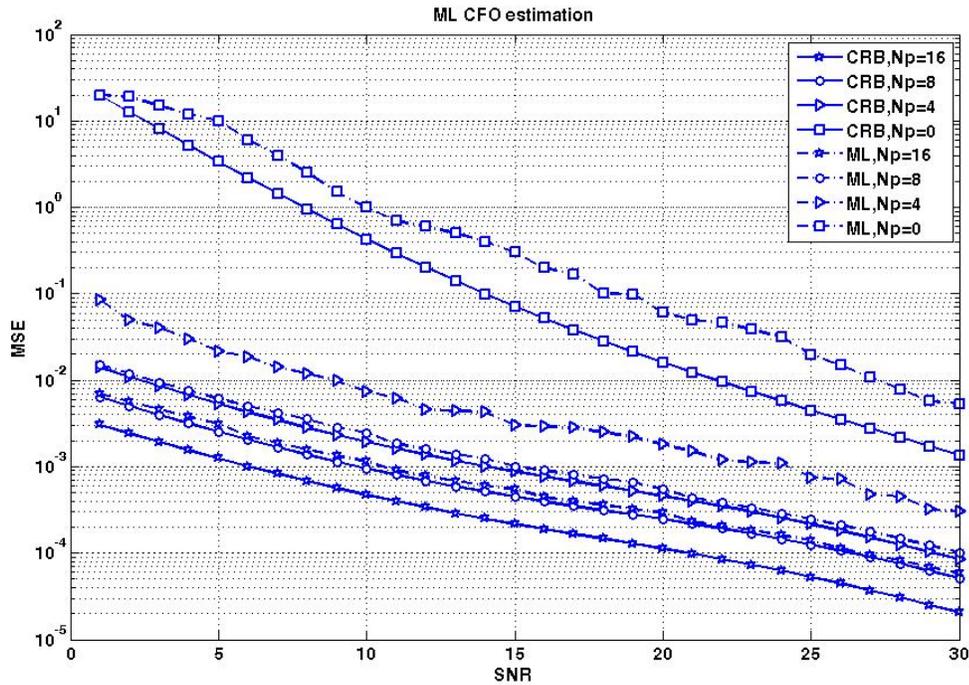
Figure 5-6: Comparison of MSE of ML CFO estimator for different pilot numbers

Figure 5-8 compares the MSE performance of CFO estimators presented in Sections 5.4.1 and 5.4.2 and 5.5. Different CRBs and MSEs are marked by their section number. Note that the estimator in Section 5.4.2 doesn't need an embedded subcarrier for pilot since the pilots are superimposed. The total amount of subcarriers is 64 in this simulation. When there is no VCs, for 4 pilots, CRB of 5.4.1 and 5.4.2 are almost the same but by the increase of pilots to 8, the estimator in 5.4.1 outperforms the other one in very high SNR by almost 1 dB. The true MSE (noted by ML) approaches the CRB and the difference is less for more number of pilots. Results show that the performance of ML CFO estimator presented in 5.5 is quite poor comparing to the ones presented in 5.4.1 and 5.4.2 (even with 32 pilots) but its advantage is its lower complexity and independency from channel power profile.
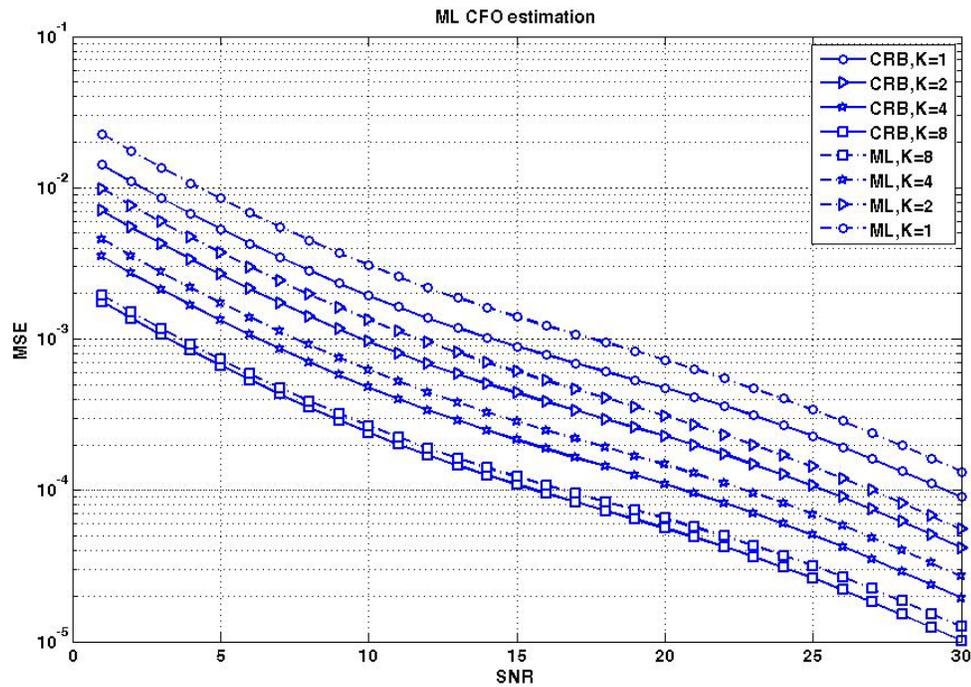
Figure 5-7: Comparison of MSE of ML CFO estimator in (5.39) for different values of *K*

Figure 5-9 investigates the effect of number of OFDM block on MSE of CFO estimator in 5.5. . It show the use of *K=4* blocks improves the MSE about 10 dB over the use of *K=1* blocks. Since the performance of this estimator comparing to the other ones is poor, it sounds that using more than one block of OFDM symbols is necessary for obtaining a fair amount of MSE.
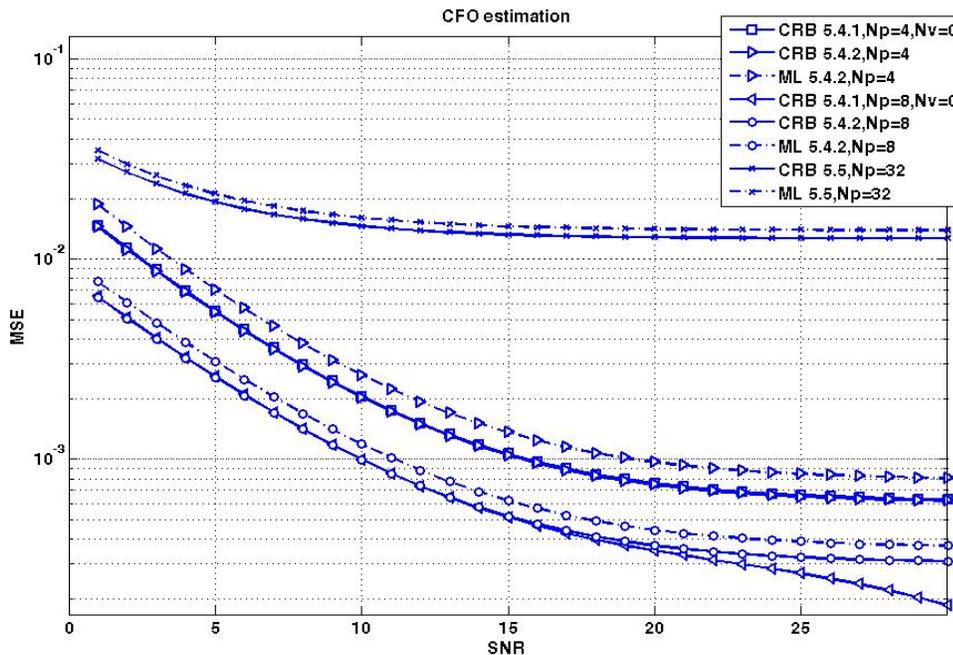
Figure 5-8: comparison of MSE of the CFO estimators presented in 5.4.1 and 5.4.2 and 5.5
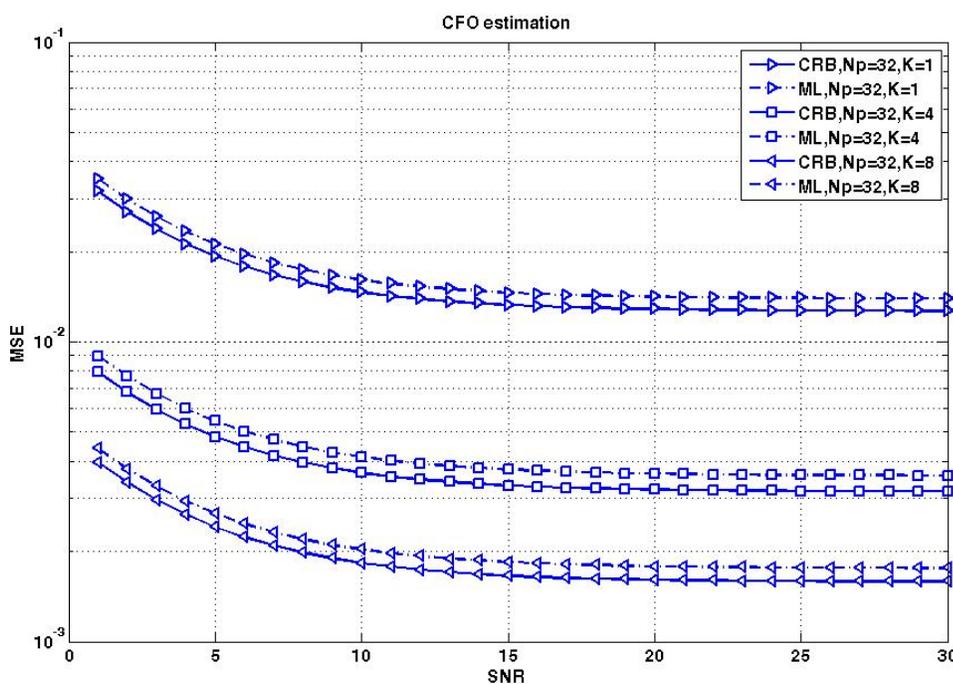
Figure 5-9: MSE of CFO estimator presented in 5.5 for different values of *K*
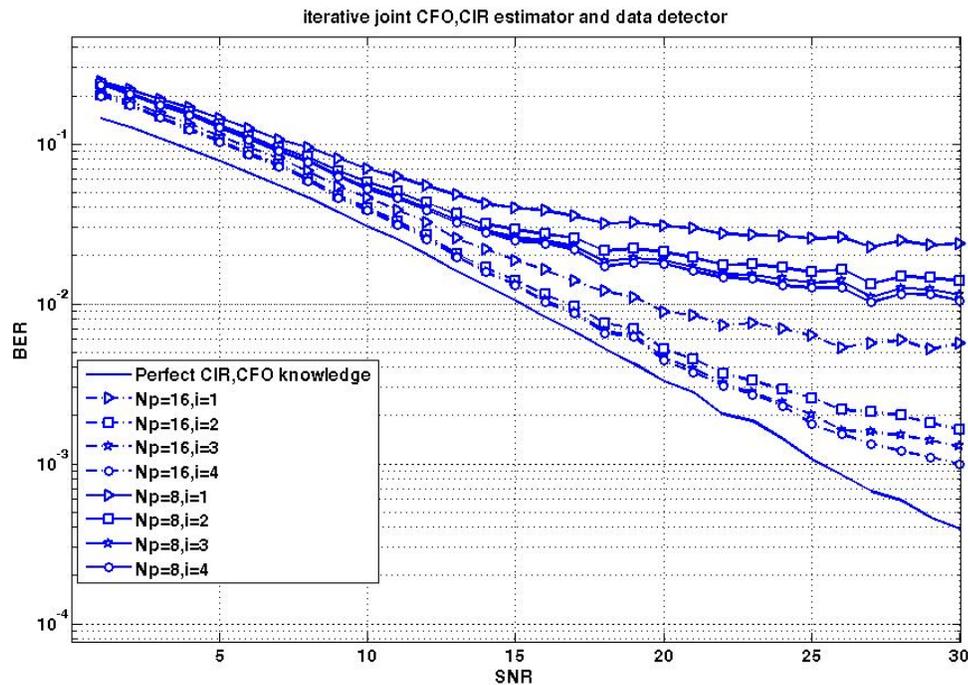
Figure 5-10: BER of iterative joint estimator and detector for different numbers of superimposed pilots and iterations

Figure 5-10 shows the BER performance of proposed joint iterative CFO and CIR estimator and data detector for a superimposed OFDM system with 64 subcarriers. A COST 207 6-ary channel model with power profile $[0.189, 0.379, 0.239, 0.095, 0.061, 0.037]$ is considered. Each path is an independently generated complex Gaussian random process. Both information data and pilot data have been drawn form BPSK constellation. We compare the performance of 8 equispaced superimposed pilots with that of 16 equispaced superimposed pilots and with different number of iterations. The total power at each subcarrier is 1 and the superimposed pilot subcarriers have 0.7 power. The notation $i=n$ denotes the performance in nth iteration. Ideal detectors, assuming the availability of perfect CIR knowledge and CFO, are used as benchmarks. The data detector with 4 iterations performs almost 2 dB and 4 dB better than the one with 1

iteration in high SNR for the case of 8 pilots and 16 pilots respectively. The detector with 8 superimposed pilots with 3 iterations performs quite the same as the one with 8 superimposed pilots and 4 iterations. Data detector with 16 pilots and 4 iterations approaches the benchmark in high SNR. Figure 5-11 and Figure 5-12 show the MSE of CIR estimator and CFO estimator respectively. More than 3 iterations don't improve the MSE of CIR significantly. The MSE of the CFO estimator improves almost 14 dB from 2nd iteration to the 3rd one for the system with 16 pilots in SNR=25 dB but extra iteration after that makes not more than 1 dB improvement.
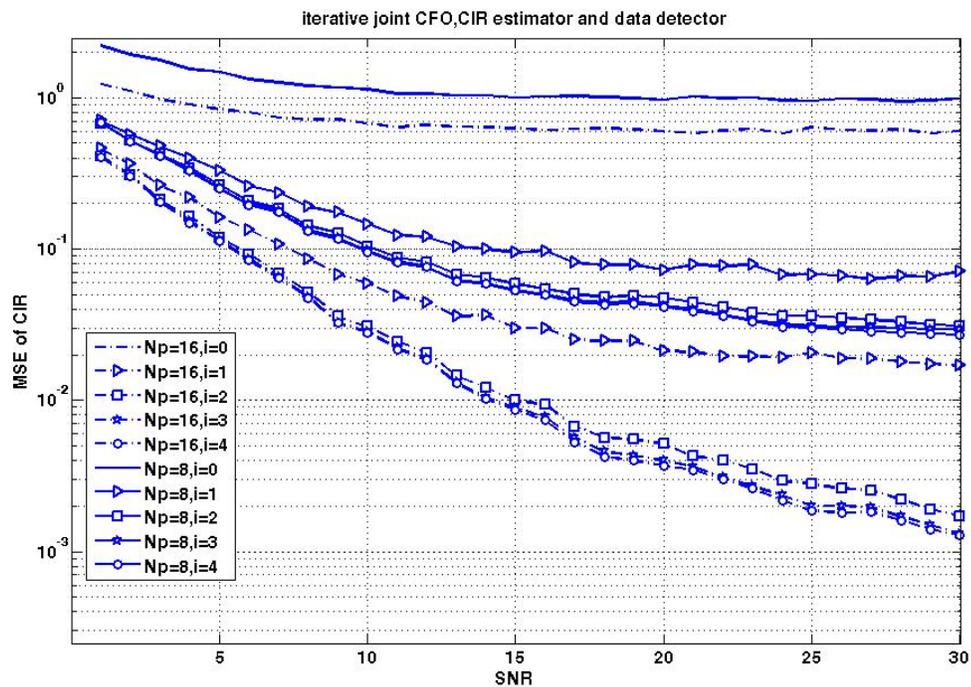
Figure 5-11: MSE of CIR of iterative joint estimator and detector for different numbers of superimposed pilots and iterations
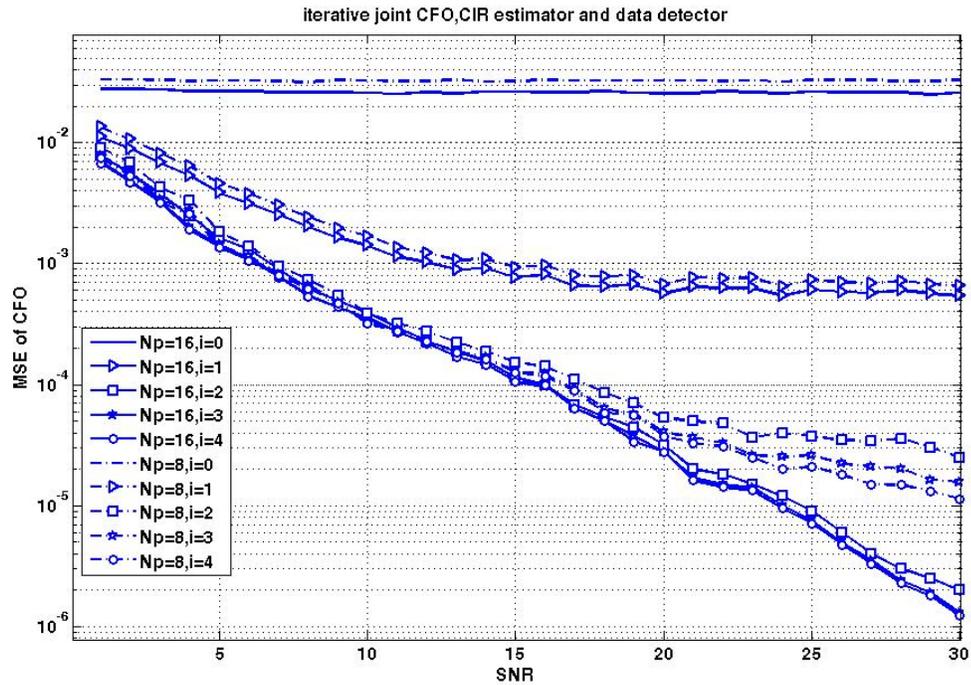
Figure 5-12: MSE of CFO of iterative joint estimator and detector for different numbers of superimposed pilots and iterations

5.8 Conclusion

We have developed several carrier frequency offset estimators for OFDM and superimposed OFDM systems. We have derived the CRB for our ML CFO estimators and in the case of superimposed OFDM; we optimized the allocation and the power distribution of superimposed pilots such that the MSE (CRB) is minimized. We investigated the effect of the amount of superimposed pilots and also number of OFDM blocks on our estimator performance in our simulation. We also compared the performance of different proposed CFO estimators. We developed a joint iterative CFO and channel estimator and data detector for superimposed OFDM systems. We compared the performance of this joint estimator for different number of iterations and pilots. Simulations show good performance of this joint estimator and detector.

Chapter Six: Conclusion

Both 4th generation and beyond 3rd generation cellular networks are characterized by their ability to provide high data rates over wireless links. Multiple antennas and orthogonal frequency division multiplexing (OFDM) are emerging as key technologies for high data rate communication. OFDM has high spectral efficiency, and it is robust to frequency selective fading. Moreover, it permits one-tap equalization. In this thesis, we have studied the data detection, CFO estimation, and channel estimation, which is critical in coherent detection.

In Chapter 2, we developed a ML multi-symbol non-coherent data detector. We used two different algorithms to solve the phase ambiguity problem. Then, we compared the performance of two efficient detection algorithms: V-BLAST and SD. Also, we studied a multi-symbol differential data detection algorithm.

Pilot based channel estimation entails a significant bandwidth loss, motivating blind methods. In Chapter 3, we introduced a semi-blind channel estimator for SIMO OFDM systems. We pointed out the phase ambiguity, which is common to all blind techniques. Then, we discussed identifiability conditions of the estimator.

The idea of superimposed data and pilot transmission has received attention in digital communication systems. In OFDM, it can enable CIR estimation without sacrificing the data rate. In Chapter 4, we developed a joint iterative channel estimator and data detector for a superimposed OFDM system. We also derived the CRB of our estimator and optimized pilot placement in order to minimize the MSE of our estimator.

While OFDM comes with so many advantages, its sensivity to carrier offset is an issue. CFO destroys the orthogonality between subcarriers and introduces inter-carrier interference. In

Chapter 5, we investigated several techniques to estimate CFO. We derived CRB for ML estimators. We also developed a joint channel and CFO estimator and data detector for superimposed OFDM systems.

The study in this thesis might be continued by research to solve the following problems:

- Our channel estimator and data detectors may be extended to MIMO OFDM systems. Also, CFO estimators can be developed over MIMO systems.

- The joint channel and CFO estimator and data detector is an open problem that hasn't been investigated vastly. It would be interesting to derive such a joint estimator.

- Superimposed transmission is a relatively new idea in digital communication. We explored its benefit in channel and CFO estimation. However, there is a good potential for improvements on the use of the superimposed techniques. It is worth developing new algorithms for detectors and estimators in superimposed OFDM systems.

REFERENCES

[1] William C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw Hill, 1989.

[2] William C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1974.

[3] J. H. Conway, and N. J. Sloane, *Sphere packing, Lattice and Groups*, 3rd ed. New York, Springer-Verlag, 1998.

[4] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics letters*, vol. 35, no. 1, pp. 14-15, Jan. 1999.

[5] E. Viterbo, and J. Bouros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639-1642, Jul. 1999.

[6] B. Hassibi, and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806-2818, Aug. 2005.

[7] Lampe, L., Schober R., Pauli V., and Windpassinger C., **"**Multiple-symbol differential sphere decoding," *IEEE Trans. on Communicatins*, vol. 53, no. 12, pp. 1981-1985, Dec. 2005.

[8] M. Ahmadi, and A. Saadat Mehr, "Multi Symbol Differnetial Detection:V-BLAST vs. Sphere Decoding," *Canadian Conference on Electrical and Computer Engineering, CCECE*, 22-26 April 2007, pp. 333 – 335.

[9] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201-2214, Aug. 2002.

[10] Wong Kwan Wai, Chi-Ying Tsui, and Cheng, R.S., "A low complexity architecture of the V-BLAST system," *IEEE Wireless Commun. And Networking Conf.*, 2000. vol. 1, 23-28 Sept. 2000, pp. 310 – 314.

[11] Tao Cui, and Tellambura, C., "Generalized feedback detection for MIMO systems," *IEEE Global Telecommunications Conf.*, vol. 5, 28 Nov.-2 Dec. 2005.

[12] H. Ochiai, and H. Imai, "MDPSK-OFDM with highly power efficient block codes for frequency selective fading channels," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 74-82, Jan. 2000.

[13] Divsalar D., and Simon M.K., "Multiple-symbol differential detection of MPSK," *IEEE Trans. on Communicatins*, vol. 38, no. 3, March 1990, pp. 300 – 308.

[14] Zhendao Wang, and Giannakis G.B., "Wireless multicarrier communications," *IEEE signal processing magazine*, vol. 17, no. 3, May 2000, pp. 29 – 48.

[15] van de Beek J.-J., Edfors O., Sandell M., Wilson S.K., and Borjesson P.O., "On channel estimation in OFDM systems," *Veh. Technol. Conf.*, vol. 2, 25-28 July 1995, pp. 815 – 819.

[16] Ye Li, Seshadri N., and Ariyavisitakul S., "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, March 1999, pp. 461 – 471.

[17] M. Ahmadi, and A. Saadat Mehr, "Blind Channel Identification and Data Detection for SIMO OFDM Systems," *Canadian Conference on Electrical and Computer Engineering*, CCECE, 22-26 April 2007 pp. 56-58.

[18] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2892-2993, Dec. 1995.

[19] David Tse, snd Pramod Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Sep. 2004.

[20] R. A. Horn, and C. R. Johnson*, Matrix Analysis*. Cambridge, U.K.:Cambridge Univ. Press, 1999.

[21] IEEE 802.11a, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High speed physical layer in the 5 GHz band," *IEEE standard*, 1999.

[22] C. E. Kastenholz, and W. P. Birkemeier, "A simultaneous information transfer and channel sounding modulation technique for wide-band channels," *IEEE Trans. on Communication Technology*, pp. 162-165, June 1965.

[23] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *IEE Electronics Letters*, vol. 31, no. 13, pp. 1044-1046, June 1995.

[24] G. T. Zhou, M. Viberg, and T. Mckelvey, "Superimposed periodic pilots for blind channel estimation," *Proc. 35th Asilomar Conference on Signals, Systems, and Computers*, pp. 653-657, Pacific Grove, CA, Nov. 2001.

[25] N. Chen, and G. T. Zhou, "A superimposed periodic pilot scheme for semi-blind channel estimation of OFDM systems," *Proc. 10th IEEE DSP Workshop*, pp. 362- 365, Pine Mountain, GA, Oct. 2002.

[26] J. K. Tugnait, and W. Luo, "On channel estimation using superimposed training and first-order statistics," *Proc. IEEE Int. Conf. Acoust, Speech, Signal Processing*, pp. 624-627, Hong Kong, China, Apr. 2003.

[27] G. T. Zhou, and N. Chen, "Superimposed training for doubly selective channels," *Proc. IEEE Statistical Signal Processing Workshop*, pp. 73-76, St. Louis, MO, Sept. 2003.

[28] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 5, pp. 1403 – 1418, 2004.

[29] S. Balasubramanian, B. Farhang-Boroujeny, and V.J. Mathews, "Pilot embedding for channel estimation and tracking in OFDM systems," in *Proc. of IEEE Globecom*, Nov. 2004, vol. 2, pp. 1244 – 1248.

[30] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[31] C. K. Ho, B. Farhang-Boroujeny, and F. Chin, "Added pilot semi-blind channel estimation scheme for OFDM in fading channels," in *Proc. Of Globecom*, vol. 5, Nov. 2001, pp. 3075 – 3079.

[32] N. Chen, and G. T. Zhou, "A superimposed periodic pilot scheme for semi-blind channel estimation of OFDM systems," in *Proc. of IEEE 10th Digital Signal Processing Workshop*, Oct. 2002, pp. 362 – 365.

[33] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Lett.*, vol. 12, no. 3, pp. 226 – 229, Mar. 2005.

[34] Tao Cui, and Tellambura C.," Superimposed Pilot Symbols for Channel Estimation in OFDM Systems," *IEEE Globecom Conf.*, vol. 4, 28 Nov.-2 Dec. 2005 pp. 2229-2233.

[35] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, New York: Addison-Wesley Publishing Co., 1990.

[36] M. Dong, and L. Tong, "optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055-3069, Dec. 2002.

[37] S. Ohno, and G. B. Giannakis, "Capacity maximizing MMSE optimal pilots for wireless OFDM over frequency selective block," *IEEE Trans. Info. Theory*, vol. 50, no. 9, pp. 2138-2145, Sep. 2004.

[38] T. Pollet, and M. Moeneclaey, "The effect of carrier frequency offset on the performance of band limited single carrier and ofdm signals," in *Proc. of IEEE Globecom96*, London, U.K., 1996, pp. 719-723.

[39] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, Oct. 1994, pp. 2908 – 2914.

[40] J. J. van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1800–1805, Jul. 1997.

[41] Biao Chen, "Maximum likelihood estimation of OFDM carrier frequency offset," *IEEE Signal Processing letters*, vol. 9, no. 4, April 2002, pp. 123 – 126.

[42] M. Ghogho, and A. Swami, G. B. Giannakis, "Optimizing null-subcarrier selection for CFO estimation in OFDM over frequency-selective fading channels," in *Proc. GLOBECOM*, Nov. 2001.

[43] Tureli, U., and Hui Liu, "A high efficiency carrier estimator for OFDM communications," *IEEE Commun. Letters*, vol. 2, no. 4, pp. 104-106, April 1998.

[44] Tao Cui , and Tellambura C., "Cyclic-prefix based noise variance and power delay profile estimation for OFDM systems,**"** *2005 IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, 2005, PACRIM, 24-26 Aug. 2005, pp. 518 – 521.

[45] M. H. Meyrs, and L. Franks, "Joint carrier phase and symbol timing recovery for PAM systems," *IEEE Trans. Commun*., vol. 28, no. 8, pp. 1121-1129, Aug. 1980.

[46] M. Brookes, "Matrix reference manual", [online] available:

http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html

[47] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Select. Areas Commun*., vol. 9, pp. 895-908, Aug. 1991.

[48] American National Standards Institute (ANSI), "Asymmetric digital subscriber line (ADSL) metallic interface," *Draft American National Standards for Telecommunications*, Dec. 1995.

[49] European Telecommunications Standards Institute (ETSI), "Radio broadcasting systems, digital video broadcasting (DVB), framing structure, channel coding and modulation for digital terrestrial television," *European Telecommunication Standard ETS 300 744*, 1st edition, March 1997.

[50] M. Ahmadi, and A. Saadat Mehr, "Superimposed training aided Carrier Frequency Estimation in OFDM systems," *IEEE EIT conference 2007*, May 17-20 2007.

[51] J. A. C. Ringham, "Multi Carrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.* ,vol. 28, no. 5.

[52] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM to carrier frequency offset and Wiener phase noise," *IEEE Trans. on Commun.*, vol. 43, no. 2/3/4, ,Feb./Mar./Apr. 1995, pp. 191-193.

[53] O. Edfors, M. Sandell, J. J. Van de Beek, S. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931-939, Jul. 1998.

[54] Y. Li, J. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902-915, July 1998.

[55] B. Muguet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1699-1712, July 2002.

[56] R. Health, and G. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Processing*, vol. 47, no. 3, pp. 848-856, Mar. 1999.

[57] C. Li, 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.

[58] Frenger P.K., Arne N., and Svensson B., "Decision-directed coherent detection in multicarrier systems on Rayleigh fading channels," *IEEE Trans. on Vehicular Technology,* vol. 48, no. 2, March 1999, pp. 490 – 498.

[59] Zamiri-Jafarian H., Omidi M.J., and Pasupathy S., "Improved channel estimation using noise reduction for OFDM systems," *IEEE Semiannual Vehicular Technology Conf., VTC 2003-Spring*, vol. 2, 22-25 April 2003, pp. 1308 – 1312.

[60] de Courville M., Duhamel P., Madec P., and Palicot, J., "Blind equalization of OFDM systems based on the minimization of a quadratic criterion," *IEEE International Conference on Record, Converging Technologies for Communications, Tomorrow's Applications*, vol. 3, 23-27 June 1996, pp. 1318 – 1322.

[61] Tao Cui, and Tellambura C., "Semi-blind equalization for OFDM systems over fast fading channels," *IEEE International Conference on Communications, ICC 2005*, vol. 2, 16-20 May 2005, pp. 1137 – 1141.

[62] Tao Cui, and Tellambura C., "Semi-blind channel estimation and data detection for OFDM systems over frequency-selective fading channels," *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '05).*

[63] Tao Cui, and Tellambura C., "Pilot symbols for channel estimation in OFDM systems," *IEEE Global Telecommun. Conf., GLOBECOM '05*, vol. 4, 28 Nov.-2 Dec. 2005, pp.

[64] Tao Cui, and Tellambura C., " Semi-blind channel estimation and data detection for OFDM systems over frequency-selective fading channels," *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, *(ICASSP '05)*, vol. 3, 18-23 March 2005, pp. iii/597 - iii/600.

[65] Tao Cui, and Tellambura, C., "Maximum-likelihood carrier frequency offset estimation for OFDM systems over frequency-selective fading channels," *IEEE International Conference on Communications,ICC 2005*, vol. 4, 16-20 May 2005 pp. 2506 – 2510.

[66] Tao Cui, and Tellambura C., "Joint data detection and channel estimation for OFDM systems," *IEEE Transactions on Commun.*, vol. 54, no. 4, April 2006 pp. 670 – 679.

[67] Darryl Dexu Lin, Pacheco R.A., Teng Joon Lim, and Hatzinakos D., "Optimal OFDM channel estimation with carrier frequency offset and phase noise," *IEEE Wireless Commun. and Networking Conf., WCNC 2006*, vol. 2, pp. 1050 – 1055.