OPTIMAL TRAINING SEQUENCE DESIGN FOR MIMO-OFDM IN SPATIALLY CORRELATED FADING ENVIRONMENTS

Viet D. Luong

B.E. in Electronics and Telecommunications

Posts and Telecommunications Institute of Technology, Vietnam

A thesis submitted in fulfilment

of the requirements for the degree of

Master by Research

from

School of Electrical Engineering and Telecommunications

The University of New South Wales

Sydney, Australia

October 2009

Abstract

Multiple Input Multiple Output with Orthogonal Frequency Division Multiplexing (MIMO-OFDM) has been widely adopted as one of the most promising air interface solutions for future broadband wireless communication systems due to its high rate transmission capability and robustness against multipath fading. However, these MIMO-OFDM advantages cannot be achieved unless the channel state information (CSI) can be obtained accurately and promptly at the receiver to assist coherent detection of data symbols. Channel estimation and training sequence design are, therefore, still open challenges of great interest.

In this work, we investigate the Linear Minimum Mean Square Error (LMMSE) channel estimation and design nearly optimal training sequences for MIMO-OFDM systems in spatially correlated fading. We, first, review the LMMSE channel estimation model for MIMO-OFDM in spatially correlated fading channels. We, then, derive a tight theoretical lower bound of the channel estimation Mean Square Error (MSE). By exploiting the information on channel correlation matrices which is available at the transmitter, we design a practical and nearly optimal training sequence for MIMO-OFDM systems . The optimal transmit power allocation for training sequences is found using the Iterative Bisection Procedure (IBP). We also propose an approximate transmit power allocation algorithm which is computationally more efficient than the IBP while maintaining a similar MSE performance. The proposed training sequence design method is also applied to MIMO-OFDM systems where Cyclic Prefixing OFDM (CP-OFDM) is replaced by Zero Padding OFDM - OverLap-Add method (ZP-OFDM-OLA). The simulation results show that the performance of the proposed training sequence is superior to that of all existing training sequences and almost achieves the MSE theoretical lower bound.

Acknowledgements

I would like to take this opportunity to thank many people for their support, assistance and advice in completing this thesis. First of all, I would like to thank my supervisor, Associate Professor Tuan D. Hoang, for his generous support, enthusiastic guidance and insightful criticism during my postgraduate research program. I would like to express my gratitude to Associate Professor Robert Malaney, Dr. Kha H. Ha, Dr. Tran N. Nguyen and Professor Ha H. Nguyen for the informative research discussion and valuable advice. I am deeply grateful to Dr. Wei Zhang, Associate Professor Jinhong Yuan, Associate Professor Timothy Hesketh and Professor Chee Kwok for their valuable advice and enthusiastic support. I highly appreciate the anonymous thesis examiners for their hard work and helpful comments to improve the quality and presentation of the thesis.

I wish to extend my gratitude to my friends - Diep N. Nguyen, Hung Q. Ta, Vinh X. Nguyen, Ahmed Pasha, Faizan Javed, Kaveh Fanian, Boris Savkovic, Hamid Teimoori for their helpful support and sincere friendship. Many thanks go to all my friends in VDS group and colleagues in the BSL lab for sharing the joy and difficult moments with me.

I wish to express my deepest gratitude to my parents for their eternal love and unconditional support in all aspects of my life. Finally, I would like to devote this work to my dear wife and daughter. They are the greatest source of inspiration for me to complete this thesis.

To my parents, Honey and Bee

Author's Publications

The following publications have been produced by or in conjunction with the author during his Master by Research candidacy.

- Viet D. Luong, Tran N. Nguyen, and Tuan D. Hoang, "Optimal Training Sequence Design for MIMO-OFDM in Spatially Correlated Fading," in *2nd International Conference on Signal Processing and Telecommunication Systems, ICSPCS'2008*, (Gold Coast, Australia), pp 1-6, Dec 2008

- Tuan D. Hoang, Viet D. Luong, and Ha H. Nguyen, "Optimization of Training Sequences for Spatially Correlated MIMO-OFDM," in *ICASSP 2009*, (Taipei, Taiwan), pp 2681-2684, Apr 2009

- Tuan D. Hoang, Viet D. Luong, and Ha H. Nguyen, "Optimization of Training Sequences for Spatially Correlated MIMO-OFDM," in *IEEE Transactions on Wireless Communications*, to be submitted

Contents

List of Figures

Acronyms and Notations

Acronyms

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CP-OFDM	Cyclic Prefixing - OFDM
CSI	Channel State Information
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
FDM	Frequency Division Multiplexing
FFT	Fast Fourier Transform
IBI	InterBlock Interference
IBP	Iterative Bisection Procedure
IFFT	Inverse Fast Fourier Transform
i.i.d.	independent and identically distributed
ISI	InterSymbol Interference
KKT	Karush-Kuhn-Tucker
LMMSE	Linear Minimum Mean Square Error
LS	Least Square
MAP	Maximum A Posteriori
MIMO	Multiple Input Multiple Output
MIMO-OFDM	MIMO with OFDM
MLE	Maximum Likelihood Estimator

MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MVUE	Minimum Variance Unbiased Estimator
OFDM	Orthogonal Frequency Division Multiplexing
OLA	OverLap-Add method
OLS	OverLap-Save method
PDF	Probability Distribution Function
SDP	Semi-Definite Programming
SER	Symbol Error Rate
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
STBC	Space Time Block Coding
SVD	Singular Value Decomposition
ZP-OFDM-OLA	Zero Padding - OFDM - OLA

Notations

Symbol	**Description**
$\mathbf{X}$, $\mathbf{x}$	Matrix/column vector
$()^T$, $()^*$, $()^H$	Transpose, complex conjugate and Hermitian transpose
$\mathbf{C}$, $\mathbf{R}$	The complex/real number field
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{0}_{n\times m}$	$n \times m$ zero matrix
$\det(\mathbf{X})$	Determinant of matrix $\mathbf{X}$
$\otimes$	Kronecker product
vec	Vectorization of matrix X
$\mathrm{diag}\begin{pmatrix} d_0 & d_1 & \dots & d_N \end{pmatrix}$	A diagonal matrix in which $d_0, d_1, \dots, d_N$ are the elements on the main diagonal
$E[.]$	Statistical expectation
$\mathbf{dom} f$	Domain of function f
∇f	Gradient of function f
$\nabla^2 f$	Hessian of function f

Chapter 1

Introduction

1.1 Literature Review and Research Motivation

Multiple Input Multiple Output (MIMO) techniques promise to bring a revolution to wireless communication with more reliable communication link obtained through diversity [4] [38] [39] and higher data rate by spatial multiplexing [13] [14] [40] [45]. Unfortunately, the computational complexity of MIMO symbol detection in a frequency selective fading channel grows exponentially with the number of antennas, bandwidth and delay spread. It is because MIMO receivers must suppress both Multiple Access Interference (MAI) and Inter Symbol Interference (ISI) to adequately detect the signal. MIMO receiver complexity in a frequency selective fading channel, therefore, becomes prohibitive even for a small number of transmit and receive antennas [17]. Meanwhile, the basic idea of OFDM is to convert a frequency selective channel into a set of parallel flat fading channels. In addition, the implementation of OFDM is very computationally efficient thanks to the use of Fast Fourier Transform (FFT) algorithm. Therefore, the combination of MIMO and OFDM is expected to exploit the MIMO merits at a reasonably and practically computational complexity [8] [35] [37]. MIMO-OFDM is now adopted as one of the most promising air interface solutions for future broadband wireless communication systems [1] [2] [3].

The high data rate transmission capability of MIMO-OFDM can only be achieved if receiver can obtain the channel state information (CSI) accurately and promptly to assist coherent detection. This information, in practice, has to be calculated at receiver by channel estimation. There are two popular approaches to channel estimation: *decision-directed* and *pilot-symbol-aided*. In decision-directed channel estimation scheme, the received information symbols are used as reference to estimate the CSI. Meanwhile, in the pilot-symbol-aided channel estimation scheme, a known sequence (called a pilot or a training sequence) is transmitted and this sequence is used to estimate the channel at the receiver. Pilot-symbol-aided channel estimation methods have been proved to be more reliable and allow simpler channel estimation algorithms at receivers at the price of spending higher power and larger bandwidth for transmitting known training sequences [28].

Channel estimation for MIMO-OFDM systems has received great attention since 1990s. [7] is among the early works considering OFDM channel estimation using Least Square (LS) and MMSE estimation methods. In this work, Van de Beek et al. developed LS and MMSE estimators for OFDM systems in slow fading channels. To further reduce the complexity, the authors proposed a rank reduced method which only considered the channel taps with significant energy. In [25], Ye Li et al. derived an MMSE channel estimator which made full use of the time and frequency domain correlations of the frequency response of time-varying dispersive fading channels. The authors showed that their robust channel estimator could significantly improve the performance of OFDM systems in a rapid dispersive fading channel. Ye Li et al. extended their work in [26] to obtain a channel estimator for MIMO-OFDM systems. However, this method is highly complex due to the inversion of a large matrix as a result of a higher number of unknown channel coefficients. To resolve this problem, Ye Li proposed two techniques to reduce the complexity of channel parameter estimation in [27]: optimum training sequence design and simplified channel estimation method at the expense of a slight performance degradation. [5] is another attractive work on the topic of MIMO-OFDM channel estimation. In this paper, Barhumi et al. described an LS channel estimation scheme for MIMO-OFDM systems based on pilot tones. The authors derived optimal pilot sequences and optimal placement of the pilot tones with respect to the MSE of the LS channel estimate. An important result of this paper is to show that the optimal pilot sequences of this channel estimation method are equipowered, equispaced and phase shift orthogonal.

MIMO-OFDM channel estimators using LS or MMSE methods in [5] [27] only considered MIMO-OFDM uncorrelated fading channels. However, MIMO channels are often spatially correlated in practice. Channel estimation, thus, can be improved by exploiting the channel correlation. As statistics of MIMO channels change slowly with time [6], the correlation matrices can be estimated at the receiver and fed back to the transmitter. Therefore, it is assumed that information on the correlation matrices is available at both transmitter and receiver for channel estimation and training sequence design.

The topic of optimal training sequence design for MIMO-OFDM in spatially correlated fading channels has been considered in [46] in literature. This paper has successfully formulated an LMMSE channel estimation model for MIMO-OFDM in spatially correlated fading channels. However, many important issues of designing optimal training sequence have been left unsolved. Some of them are described below.

- Necessary conditions for optimal training sequences have been derived. However, the existence of a training sequence satisfying these conditions was not considered.

- The training sequence design in this paper only applies for some special cases of correlated fading channels. Therefore, the optimal training sequence design for the general case of correlated fading channels is still an open challenge.

- The transmit power allocation for training sequences only uses an asymptotic solution at very low and very high Signal to Noise Ratio (SNR). Thus, the MSE performance is suboptimal. It is expected that a more accurate transmit power allocation will help to improve the MSE performance.

- A critical issue of optimal training sequence design in correlated fading environments is the case when channel correlation matrices are singular due to very highly correlated fading. Actually, in this case, all the mathematical derivations which depend on the inverse of channel correlation matrices throughout this paper are invalid. Therefore, we need a different approach to consider optimal training sequence design for MIMO-OFDM in a more general case of correlated fading.

Our research aims to answer the above open research questions. The achievements that we have gained are summarized in the next section.

1.2 Thesis Contributions

This thesis focuses on several unsolved issues on optimal training sequence design for MIMO-OFDM systems in spatially correlated fading channels. The main contributions of the thesis are highlighted as follows.

- The conditions for MIMO-OFDM optimal training sequences in spatially correlated fading environments which have been discovered in [46] are reviewed in this thesis. However, we show that the existence of a training matrix cannot be guaranteed in general.
- A tight theoretical lower bound on the MSE of MIMO-OFDM channel estimation in spatially correlated fading environments is extracted in Section 3.2.
- If the channel correlation matrices are singular due to highly correlated fading, all the formulas depending on the inverse of the channel correlation matrices are invalid. We address this issue in Section 3.3 by considering the limit of the channel estimation MSE when the channel correlation matrices approach singularity. We derive a general expression of the MSE theoretical lower bound in (3.23). This general expression does not depend on the invertibility of the channel correlation matrix.
- We show that the optimal transmit power allocation for training sequences to achieve the MMSE performance of channel estimation is a convex optimization problem. The solution, therefore, can be found by resorting to the IBP [33].

- We propose an approximate algorithm for transmit power allocation in Section 3.5. This algorithm is computationally more efficient while maintaining a similar MSE performance in comparison with the above IBP.
- A practical and nearly optimal training sequence design is demonstrated in Section 3.6. The proposed training sequence shows to have a superior MSE performance when compared with all existing training sequences. Its MSE performance almost achieves the MSE theoretical lower bound as specified in (3.23). It is also showed that its SER performance is very close to the SER performance when assuming perfect CSI at receiver.
- The training sequence design for MIMO-OFDM in spatially correlated fading environments can also be applied to MIMO-OFDM systems in which CP-OFDM is replaced by ZP-OFDM-OLA. We prove that the MSE performance of MIMO-ZP-OFDM-OLA optimal training sequence is exactly the same as that for the MIMO-CP-OFDM case.

1.3 Thesis Organization

The remainder of the thesis is organized as follows.

In Chapter 2, we briefly review the literature on MIMO-OFDM and channel estimation. We carefully consider MIMO and OFDM techniques to understand why they can be combined together to provide an effective air interface solution for future broadband wireless communication systems. The reason why LMMSE estimation method is chosen for MIMO-OFDM channel estimation is explained. We then review the LMMSE channel estimation model for MIMO-OFDM systems with spatially correlated fading channels. Finally, we extend the LMMSE channel estimation scheme for MIMO-OFDM systems where CP-OFDM is replaced by ZP-OFDM-OLA.

In Chapter 3, the problem of optimal training sequence design for MIMO-OFDM in spatially correlated fading environments is comprehensively treated. We derive a tight theoretical lower bound on MIMO-OFDM channel estimation MSE. We then design a practical and nearly optimal training sequence for MIMO-OFDM in spatially correlated fading environments. The optimal training sequence design for MIMO-ZP-OFDM-OLA systems is also developed in a similar fashion to that of MIMO-CP-OFDM systems.

We present our conclusions and perspectives for future work in Chapter 4.

Chapter 2

MIMO-OFDM Channel Estimation

2.1 MIMO

2.1.1 MIMO Introduction

A MIMO system is defined as a point-to-point wireless communication link with multiple antennas at both the transmitter and receiver [15]. Figure 2.1 shows a typical MIMO system with M_t transmit and M_r receive antennas. Over the last few years, MIMO wireless communication has been being a very active research area because of its potential for achieving higher date rate and more reliable data transmission in comparison with traditional single antenna systems. The performance improvements of MIMO systems are due

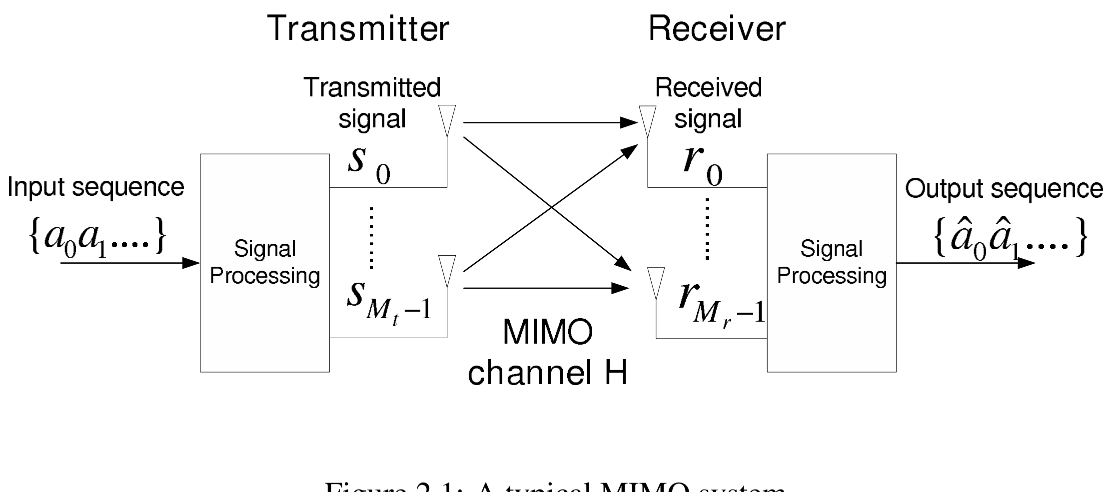


Figure 2.1: A typical MIMO system

to array gain, diversity gain and spatial multiplexing gain. Below is a brief review of the performance gains of MIMO communication scheme [8] [22] [35] [47] .

1. *Array gain*

   Array gain is the average increase in SNR at the receiver resulting from the coherent combining effect of multiple antennas at the receiver, transmitter or both. The transmit/receive array gain requires channel knowledge in the transmitter and receiver, respectively, and depends on the number of transmit and receive antennas.

2. *Diversity gain*

   Received signal strength in a wireless communication link fluctuates significantly in a random manner due to multipath fading. Diversity is a powerful technique to combat fading. Diversity techniques rely on transmitting the signal over multiple independent fading paths in time/frequency/space. Spatial (or antenna) diversity is preferable to time/frequency diversity as it does not incur an expenditure in transmission time or bandwidth. An $M_r \times M_t$ MIMO system can exploit the diversity capability of order up to $M_t.M_r$ with an appropriate space-time coding technique [4] [39] [38]

3. *Spatial Multiplexing Gain*

   It has been proved that MIMO communication offers an increase of the transmission rate (or capacity) in proportional with the number of transmit-receive antenna pairs (or $\min(M_t, M_r)$) for the same bandwidth and with no additional power expenditure [13] [14] [40] [45]. Such capacity gain can be achieved in a rich scattering propagation environment and can be realized by transmitting independent data streams over multiple antennas. The receiver exploits all the spatial degrees of freedom of the MIMO channel to decode the received signals, thereby realizing the capacity gain.

4. *Trade-off*

The price of using a MIMO system is the significantly increased complexity in transceiver signal processing and the difficulty of multiple antenna implementation.

In addition, it is important to know that it is impossible to exploit all the gains at the maximum capability simultaneously due to the limit of the available spatial degrees of freedom of MIMO channels.

2.1.2 MIMO Channel Model

A flat fading point-to-point wireless communication system employing M_t transmit and M_r receive antennas is shown in Figure 2.1. This system can be represented by the following discrete-time channel model

$$\begin{pmatrix} r_0 \\ \vdots \\ r_{M_r-1} \end{pmatrix} = \begin{pmatrix} h_{00} & \cdots & h_{0\,(M_t-1)} \\ \vdots & \ddots & \vdots \\ h_{(M_r-1)\,0} & \cdots & h_{(M_r-1)\,(M_t-1)} \end{pmatrix} \begin{pmatrix} s_0 \\ \vdots \\ s_{M_t-1} \end{pmatrix} + \begin{pmatrix} n_0 \\ \vdots \\ n_{M_r-1} \end{pmatrix}$$

or, simply,

$$\mathbf{r} = \mathbf{Hs} + \mathbf{n} \tag{2.1}$$

where

$\mathbf{s} = \begin{pmatrix} s_0 & s_1 & \dots & s_{M_t-1} \end{pmatrix}^T$ represents the M_t-dimensional transmitted signal vector.

$\mathbf{r} = \begin{pmatrix} r_0 & r_1 & \dots & r_{M_r-1} \end{pmatrix}^T$ represents the M_r-dimensional received signal vector.

$\mathbf{n} = \begin{pmatrix} n_0 & n_1 & \dots & n_{M_r-1} \end{pmatrix}^T$ is the M_r-dimensional Additive White Gaussian Noise (AWGN) noise vector $\mathscr{CN}(0, \sigma^2 \mathbf{I}_{M_r})$.

$\mathbf{H}$ is the $M_r \times M_t$ channel matrix whose elements h_{ij} represent the channel gain from j^{th} transmit antenna to i^{th} receive antenna.

2.1.3 Capacity of MIMO Channels

Capacity of Deterministic MIMO Channels

In this part, we consider the MIMO channel $\mathbf{H}$ to be deterministic.

The capacity of a Gaussian channel with transmit power constraint P is defined as [12]

$$C = \max_{f(x):E[X^2]\leq P} I(X;Y) \ bit/s/Hz \tag{2.2}$$

where X and Y are respectively transmitted and received symbol. The maximization is taken over all possible input distributions $f(x)$ under the transmit power constraint $E[X^2] \leq P$.

It is well-known that the capacity of a Gaussian channel is achieved when X is a circular symmetric complex Gaussian random variable $X \sim \mathscr{CN}(0,P)$.

The corresponding capacity of a deterministic MIMO channel is [40]

$$C = \max_{\mathbf{R}_{ss}} \log_2 \det\left(\mathbf{I}_{M_r} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H \right) bit/s/Hz \tag{2.3}$$

where the maximization is performed over all possible input covariance matrix satisfying the power constraint trace$\{\mathbf{R}_{ss}\} \leq P$.

Perfect Channel State Information at Transmitter and Receiver (CSIT and CSIR)
It is assumed here that the CSI is known at both transmitter and receiver. The SVD of the MIMO channel $\mathbf{H}$ is written as

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is the diagonal matrix of singular values of $\mathbf{H}$.

Accordingly, the transmitted signal can be precoded by multiplying with a transmit precoding matrix $\mathbf{V}$ and the received signal is multiplied by a receive shaping matrix $\mathbf{U}^H$ (see Figure 2.2).

The receiver output can be written as

$$\begin{aligned}
\mathbf{y} &= \mathbf{U}^H\mathbf{r} \\
&= \mathbf{U}^H\left(\mathbf{H}(\mathbf{V}\mathbf{s}) + \mathbf{n} \right) \\
&= \mathbf{U}^H\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H\mathbf{V}\mathbf{s} + \mathbf{U}^H\mathbf{n} \\
&= \mathbf{\Lambda}\mathbf{s} + \mathbf{U}^H\mathbf{n}
\end{aligned}$$

or, equivalently,

$$y_i = \lambda_i s_i + \tilde{n}_i \ , \ i = 1, \ldots, rank(\mathbf{H}) \tag{2.4}$$

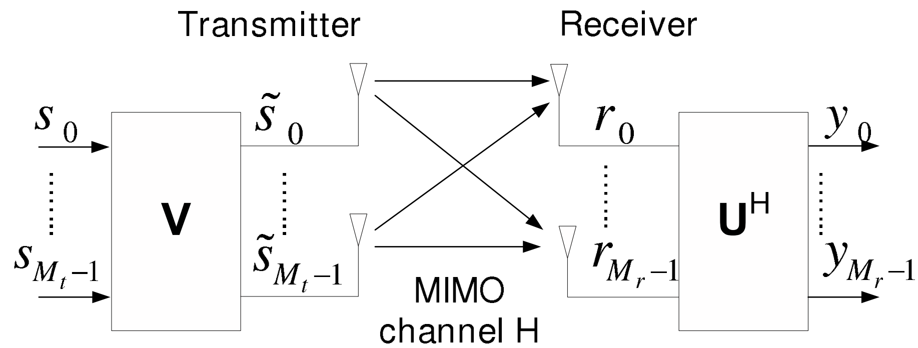


Figure 2.2: Transmit Precoding and Receive Shaping for MIMO channel

where λ_i is the i^{th} singular value of $\mathbf{H}$.

Equation (2.4) indicates that a MIMO channel can be transformed into parallel non-interfering Single Input Single Output (SISO) channels.

Following Equation (2.4), the MIMO channel capacity can be written as

$$C = \max_{P_i:\sum_i P_i \leq P} \sum_{i=1}^{rank(H)} \log_2(1+\lambda_i^2 \frac{P_i}{\sigma^2}) \qquad (2.5)$$

where P_i denotes the power allocated for the i^{th} channel of the parallel channels.

The optimal transmit power allocation to achieve the maximum data transmission rate can be found by resorting to the so-called water-filling principle [11] [12].

CSIR only Acquiring perfect channel knowledge at transmitter is, in general, very difficult in practical systems. Thus, we assume that the CSI is known at the receiver only. In this case, the transmitter cannot optimize the transmit power allocation or input covariance as it has no knowledge of the channel. Then, it is reasonable to allocate transmit power equally between transmit antennas, i.e., $\mathbf{R}_{ss} = \frac{P}{M_t}\mathbf{I}_{M_t}$. Thus, from (2.3), the capacity of a deterministic MIMO channel with CSIR is

$$C = \log_2 \det\left(\mathbf{I}_{M_r} + \frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H \right) \; bit/s/Hz \qquad (2.6)$$

which may be decomposed as

$$C = \sum_{i=1}^{rank(H)} \log_2(1+\lambda_i^2 \frac{P}{M_t\sigma^2}).$$

Capacity of MIMO Fading Channels

We have so far discussed channel capacity when the MIMO channel **H** is deterministic. We now consider the case when **H** is a Rayleigh fading channel. We assume that the receiver has perfect knowledge of the channel while the transmitter has no knowledge of the channel.

Since the channel is random, the capacity corresponding to that channel is also random. We will consider two important notions of capacity - ergodic capacity and outage capacity - which respectively relate to the mean and tail behaviors of MIMO fading channel capacity.

Ergodic Capacity Ergodic capacity of a MIMO channel is defined as an ensemble average of the information rate over the whole distribution range of elements of the channel matrix **H**.

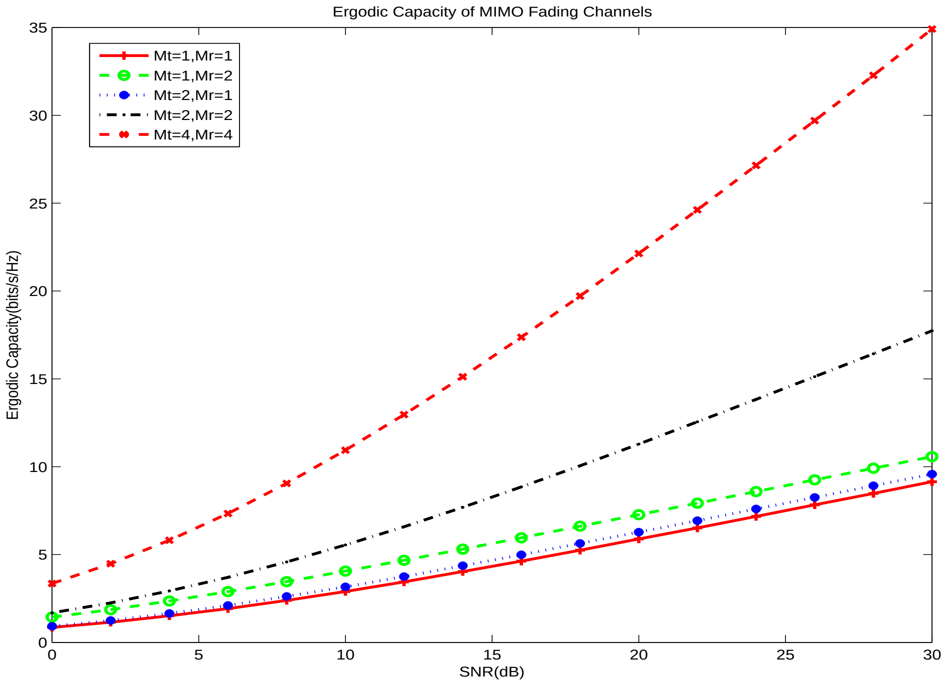
$$C_{ergodic} = E_{\mathbf{H}}\left[\ \log_2\ \det\left(\ \mathbf{I}_{M_r} + \frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H\ \right)\ \right]\ bit/s/Hz. \tag{2.7}$$

Ergodic capacity is often used as the capacity of fast fading channels where the channel conditions change very quickly in the fast fading environment. Therefore, the capacity of fast fading channels is calculated as the average capacity over a period of time.

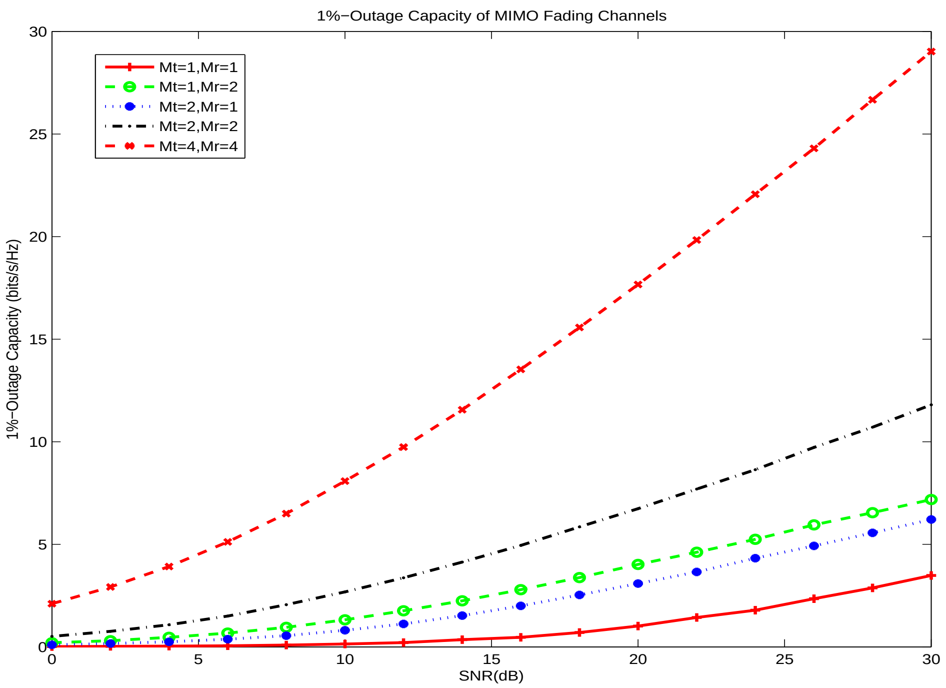
Outage Capacity As **H** is a Rayleigh fading channel, there is always a nonzero probability that the capacity of a channel realization is lower than an expected information rate. We define the $p\%$-outage capacity $C_{out,p}$ as the information rate that is guaranteed for $(100-p)\%$ of the channel realizations, i.e.,

$$P(C \leq C_{out,p}) = p\%. \tag{2.8}$$

Outage capacity is often referred as the capacity of slow fading channels where the channel gain is random but remains constant for a long time. Suppose that a transmitter sends data at a rate R bits/s/Hz. However, at the time when the channel is in deep fade and the channel capacity at that time does not exceed the rate R, then the decoding error probability cannot be made arbitrarily small irrespective of what channel code to be used in the transmitter. In this case, the communication system is said to be in outage.



(a) Ergodic Capacity



(b) 1%-Outage Capacity

Figure 2.3: Capacity of MIMO fading channel for various antenna configurations

Figure 2.3(a) and 2.3(b) respectively show the ergodic capacity and 1%-outage capacity of Rayleigh fading channel for several MIMO configurations as a function of SNR. Throughout this thesis, SNR is defined as the ratio between total transmit power P and total noise power σ^2. It is clear from these figures that the capacity of a MIMO channel increase by r bit/s/Hz $(r = \min(M_t, M_r))$ for every $3dB$ increase in SNR (for high SNR) as opposed to 1 bit/s/Hz in conventional SISO channels.

2.2 OFDM

2.2.1 OFDM Introduction

A high data rate wireless communication scheme is highly desired in many applications. However, when symbol rate increases, i.e., symbol duration reduces, the received signal will be suffered from more severe InterSymbol Interference (ISI) which is caused by the dispersive nature of wireless channels. Thus, a more complex equalization scheme is required. To facilitate the simplicity of equalization at receiver, OFDM modulation divides the entire bandwidth of a frequency selective wireless channel into many narrowband subchannels whose bandwidth is less than the coherence bandwidth of the wireless channel. Therefore, each subchannel experiences relatively flat fading and the ISI on each subchannel is negligible. Figure 2.4 shows this idea graphically.

Another advantage of using OFDM modulation is the high spectral efficiency due to the use of a set of orthogonal carrier frequencies. That means the spectra of the subcarriers are overlapped with each other but remain orthogonal as shown in Figure 2.5.

In addition, the data modulation and demodulation on the orthogonal subcarriers can be implemented by IFFT/FFT algorithm. The implementation of OFDM modulators and demodulators, therefore, benefits from the computational efficiency of IFFT/FFT.

To deal with channel delay spread, a guard interval is introduced between OFDM signal blocks. The guard interval, whose length should exceed the maximum excess delay, is in form of a cyclic prefix in standard CP-OFDM or in form of a zero-padding interval in ZP-OFDM-OLA. We will discuss about CP-OFDM in the next sections. ZP-OFDM-OLA

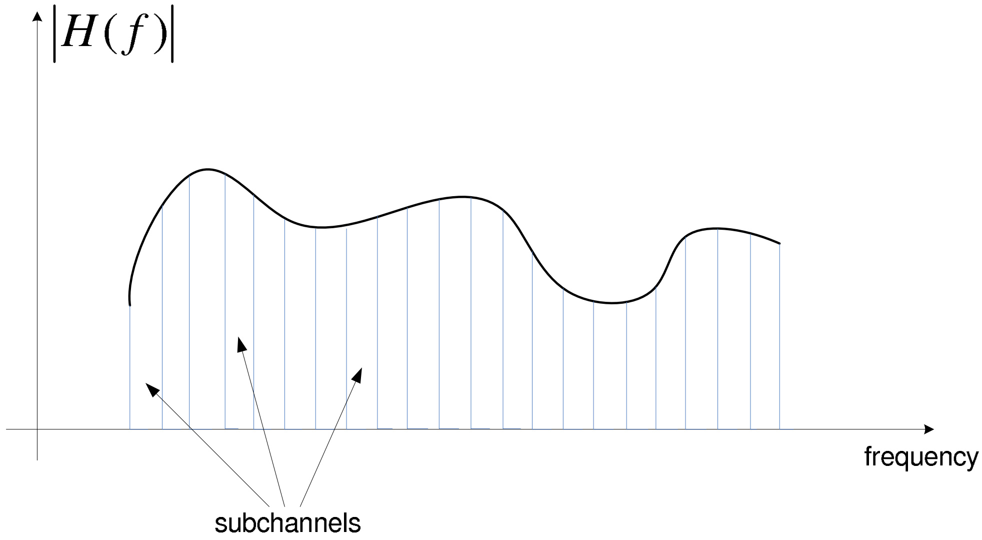
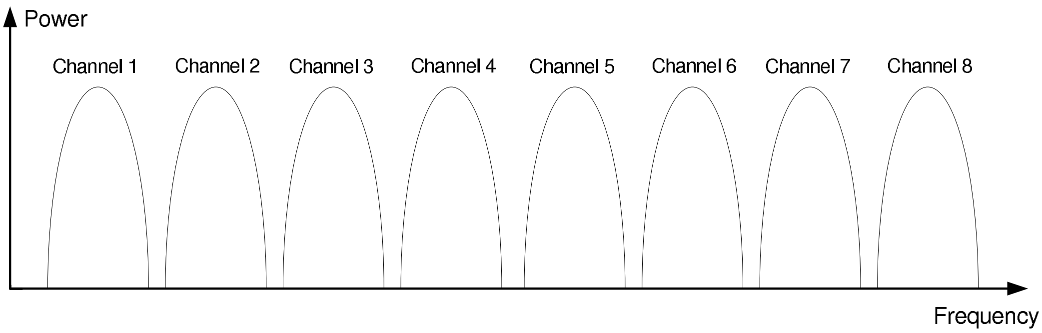
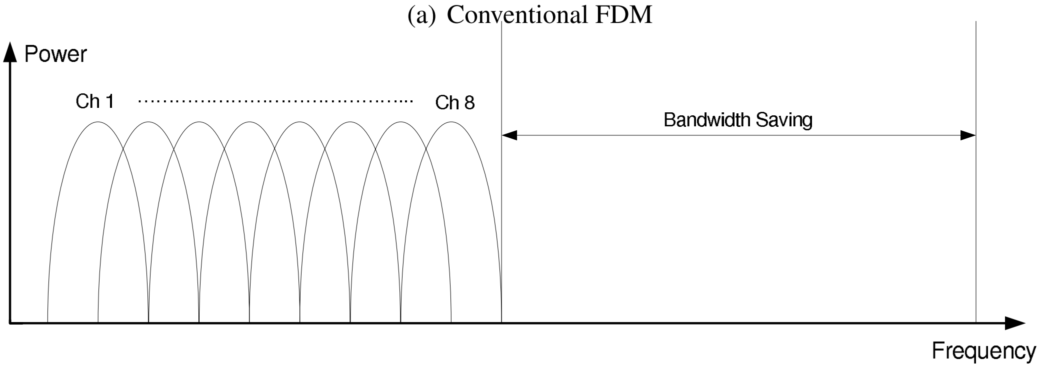


Figure 2.4: OFDM subchannels



(a) Conventional FDM



(b) OFDM

Figure 2.5: Comparison between conventional FDM and OFDM

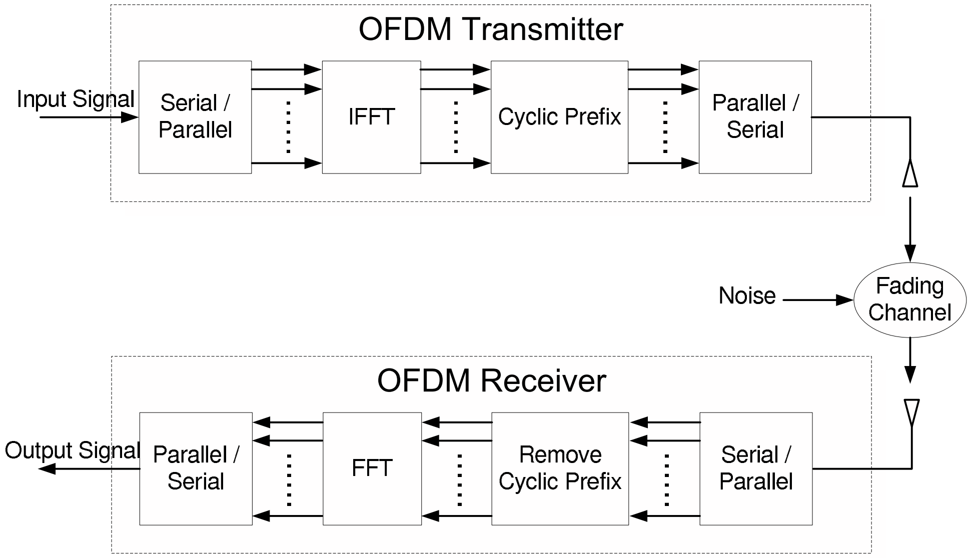


Figure 2.6: a typical OFDM system

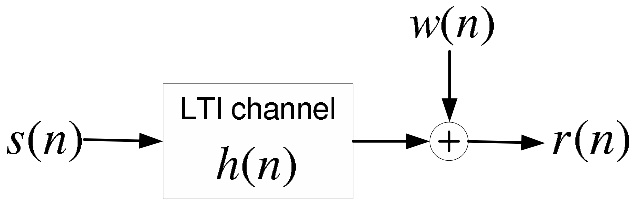


Figure 2.7: a discrete-time LTI channel

will be discussed in detail in Section 2.8.2.

A graphical view of a standard OFDM transmission system is provided in Figure 2.6. In the next part, we will elaborate on the mathematical details of OFDM systems [32] [44]
.

2.2.2 OFDM Mathematical View

Block Transmission

We consider a digital communication system with a Linear and Time Invariant (LTI) discrete-time baseband equivalent channel (Figure 2.7). Transmitted sequence $s(n)$ passes through an LTI discrete-time baseband equivalent channel with the finite impulse response

$h(l)$ of length L (the channel order L is determined by dividing the maximum path delay τ_{max} by the sampling period T_s).

The received baseband signal is expressed as

$$r(n) = \sum_{l=0}^{L-1} h(l)s(n-l) + w(n) \tag{2.9}$$

or,

$$r(n) = h(n) * s(n) + w(n)$$

where

$r(n)$ is the received signal at time index n,

$h(l)$ is the impulse response of the channel,

$w(n)$ is the AWGN noise at time index n.

The received baseband signal in (2.9) can be expressed in a different form as below

$$r(n) = h(0)s(n) + \sum_{l=1}^{L-1} h(l)s(n-l) + w(n) \tag{2.10}$$

where the first term is the desired received signal at time n and the second term plays as the ISI which distorts the desired received signal.

To mitigate such time-domain dispersive effect which gives rise to frequency selectivity, we divide the transmitted sequence into block of size P. P is chosen to be much longer than the channel order L.

Define the i^{th} transmitted block to be $\mathbf{s}(i) = [s(iP), s(iP+1), \ldots, s(iP+P-1)]^T$ and the i^{th} received block to be $\mathbf{r}(i) = [r(iP), r(iP+1), \ldots, r(iP+P-1)]^T$.

The linear convolution operation as in (2.9) can be expressed in a matrix-vector form as follows

$$\mathbf{r}(i) = \mathbf{H}\mathbf{s}(i) + \mathbf{H}_{IBI}\mathbf{s}(i-1) + \mathbf{w}(i) \tag{2.11}$$

where $\mathbf{w}(i) \sim \mathscr{CN}(0, \sigma^2 \mathbf{I}_P)$ is the corresponding $P \times 1$ AWGN noise vector.

$\mathbf{H}$ and $\mathbf{H}_{IBI}$ are respectively a $P \times P$ lower and upper triangular Toeplitz matrix which are determined as below

$$\mathbf{H} = \begin{pmatrix} h(0) & 0 & 0 & \dots & 0 \\ \vdots & h(0) & 0 & \dots & 0 \\ h(L-1) & \dots & \ddots & \dots & \vdots \\ \vdots & \ddots & \dots & \ddots & 0 \\ 0 & \dots & h(L-1) & \dots & h(0) \end{pmatrix}$$

and

$$\mathbf{H}_{IBI} = \begin{pmatrix} 0 & \dots & h(L-1) & \dots & h(1) \\ \vdots & \ddots & 0 & \ddots & \vdots \\ 0 & \dots & \ddots & \dots & h(L-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix}.$$

Due to the time-domain dispersive nature of the channel, InterBlock Interference (IBI) arises between successive blocks and renders $\mathbf{r}(i)$ dependent on both $\mathbf{s}(i)$ and $\mathbf{s}(i-1)$.

To obtain IBI-free received signal blocks, we must introduce a guard interval whose length is at least L into the transmitted blocks $\mathbf{s}(i)$. The guard interval of length L is added into the information-bearing $N \times 1$ vector $\bar{\mathbf{s}}(i)$ by pre-multiplying with a $P \times N$ precoding matrix $\mathbf{T}$, where $P = N + L$

$$\mathbf{s}(i) = \mathbf{T}\bar{\mathbf{s}}(i).$$

We can rewrite Equation (2.11) as

$$\mathbf{r}(i) = \mathbf{H}\mathbf{T}\bar{\mathbf{s}}(i) + \mathbf{H}_{IBI}\mathbf{T}\bar{\mathbf{s}}(i-1) + \mathbf{w}(i). \tag{2.12}$$

We observe that P transmitted symbols are now used to transfer only $N = P - L$ information-bearing symbols. The other L transmitted symbols are redundant symbols acting as a guard interval.

Two popular methods of filtering a long sequence on a block-by-block basis using FFT are the OverLap-Save method (OLS) and the OverLap-Add method (OLA) [36] [34]. Motivated by the OLS method, a cyclic prefix of length L, acting as a guard interval, is

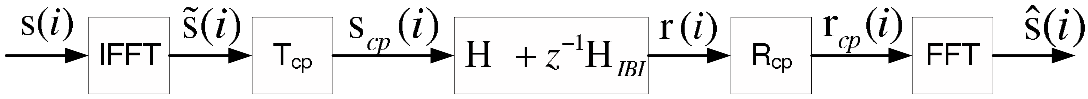


Figure 2.8: a mathematical view of CP-OFDM

added into the transmitted block. At receiver, the first L elements of the received blocks are simply discarded. This method is used in standard CP-OFDM system. Meanwhile, deriving from the OLA method, L trailing zeros is padded into the transmitted block $\mathbf{s}(i)$. This method is used in ZP-OFDM-OLA system. ZP-OFDM-OLA will be treated in detail in Section 2.8.2.

In the next section, we will look into mathematical details of CP-OFDM system.

Standard CP-OFDM

Figure 2.8 depicts the baseband discrete time block equivalent model of a standard CP-OFDM system. The i^{th} $N \times 1$ information block $\mathbf{s}(i)$ is first precoded by the IFFT matrix $\mathbf{F}_N^H$ to yield the time-domain block vector

$$\tilde{\mathbf{s}}(i) = \mathbf{F}_N^H \mathbf{s}(i). \tag{2.13}$$

Then a cyclic prefix of length L is added to the time-domain block vector by copying the last L elements of $\tilde{\mathbf{s}}(i)$ and placing onto the beginning of $\tilde{\mathbf{s}}(i)$ to form a $P \times 1$ transmitted block where $P = N + L$. This step is equivalent to pre-multiplying $\tilde{\mathbf{s}}(i)$ with a cyclic prefix matrix $\mathbf{T}_{cp}$ defined by

$$\mathbf{T}_{cp} = \begin{pmatrix} \mathbf{0}_{L \times N-L} & \mathbf{I}_L \\ \multicolumn{2}{c}{\mathbf{I}_N} \end{pmatrix}.$$

Thus, we have

$$\mathbf{s}_{cp}(i) = \mathbf{T}_{cp} \tilde{\mathbf{s}}(i). \tag{2.14}$$

The transmitted block $\mathbf{s}_{cp}(i)$ then passes through the LTI channel of order L. The received block is written in matrix-vector form as in (2.11)

$$\mathbf{r}(i) = \mathbf{H}\mathbf{s}_{cp}(i) + \mathbf{H}_{IBI}\mathbf{s}_{cp}(i-1) + \mathbf{w}(i)$$

At the receive side, the received block $\mathbf{r}(i)$ is pre-multiplied by a receive matrix $\mathbf{R}_{cp} = \begin{pmatrix} \mathbf{0}_{N\times L} & \mathbf{I}_N \end{pmatrix}$ in order to remove the first L entries of $\mathbf{r}(i)$, i.e., remove the IBI component. We have

$$\mathbf{R}_{cp}\mathbf{H}_{IBI} = \mathbf{0}.$$

Thus,

$$\begin{aligned} \mathbf{r}_{cp}(i) &= \mathbf{R}_{cp}\mathbf{r}(i) \\ &= \mathbf{R}_{cp}\mathbf{H}\mathbf{s}_{cp}(i) + \mathbf{R}_{cp}\mathbf{H}_{IBI}\mathbf{s}_{cp}(i-1) + \mathbf{R}_{cp}\mathbf{w}(i) \\ &= \mathbf{R}_{cp}\mathbf{H}\mathbf{s}_{cp}(i) + \mathbf{R}_{cp}\mathbf{w}(i). \end{aligned} \tag{2.15}$$

Finally, the IBI-free received block $\mathbf{r}_{cp}(i)$ is multiplied with the FFT matrix $\mathbf{F}_N$

$$\begin{aligned} \hat{\mathbf{s}}(i) &= \mathbf{F}_N\mathbf{r}_{cp}(i) \\ &= \mathbf{F}_N\mathbf{R}_{cp}\mathbf{H}\mathbf{T}_{cp}\mathbf{F}_N^H\mathbf{s}(i) + \mathbf{F}_N\mathbf{R}_{cp}\mathbf{w}(i) \\ &= \mathbf{F}_N\tilde{\mathbf{H}}\mathbf{F}_N^H\mathbf{s}(i) + \mathbf{F}_N\mathbf{R}_{cp}\mathbf{w}(i) \end{aligned} \tag{2.16}$$

where $\tilde{\mathbf{H}} = \mathbf{R}_{cp}\mathbf{H}\mathbf{T}_{cp}$ is an $N \times N$ circulant matrix with the first row $\begin{pmatrix} h(0) & 0 & \dots & 0 & h(L-1) & \dots & h(1) \end{pmatrix}$.

Equalization of CP-OFDM transmission relies on the following two well-known properties

- *Property 1 (Diagonalization of a circulant matrix):* an $N \times N$ circulant matrix $\tilde{\mathbf{H}}$ can be diagonalized by pre- and post- multiplication with N-point FFT and IFFT matrices, i.e.,

  $$\mathbf{F}_N\tilde{\mathbf{H}}\mathbf{F}_N^H = \mathbf{D}_H = \mathrm{diag}\begin{pmatrix} H_0 & H_1 & \dots & H_{N-1} \end{pmatrix} \tag{2.17}$$

  where

  $$H_k = \sum_{l=0}^{L-1} h(l)e^{\frac{-j2\pi kl}{N}}. \tag{2.18}$$

- *Property 2:* the statistics of a random noise vector is unchanged by a unitary transformation.

Consider a random noise vector $\mathbf{w} \sim \mathscr{CN}(\mathbf{0}, \mathbf{R}_w)$, then

$$E[\tilde{\mathbf{w}}] = E[\mathbf{F}_N \mathbf{w}] = \mathbf{0}$$

and

$$\begin{aligned} \mathbf{R}_{\tilde{w}} &= E[\mathbf{w}^H \mathbf{F}_N^H \mathbf{F}_N \mathbf{w}] \\ &= E[\mathbf{w}^H \mathbf{w}] \\ &= \mathbf{R}_w \end{aligned} \tag{2.19}$$

Therefore, Equation (2.16) can be rewritten as

$$\hat{\mathbf{s}}(i) = \operatorname{diag}\begin{pmatrix} H_0 & H_1 & \dots & H_{N-1} \end{pmatrix} \mathbf{s}(i) + \tilde{\mathbf{w}}(i) \tag{2.20}$$

where $\tilde{\mathbf{w}}(i) = \mathbf{F}_N \mathbf{R}_{cp} \mathbf{w}(i)$.

Without loss of generality, we can drop the block index i. Equation (2.20) is simplified to

$$\hat{s}_k = H_k s_k + \tilde{w}_k, k = 0, 1, \dots, N-1. \tag{2.21}$$

It is clear from (2.21) that the frequency selective channel is divided into parallel subchannels and received symbols on each subchannel can be decoded independently.

2.3 Space-Time Coding

Space Time Block Coding (STBC) is a very remarkable transmit diversity technique in MIMO communication. It starts with Alamouti's genuinely brilliant simple idea about a transmit diversity technique for wireless communication using orthogonal code and two transmit antennas [4]. A similar approach was used to generalise STBC for MIMO systems involving an arbitrary number of antennas in [38]. The extension to other transmit diversity techniques such as space time trellis coding [39], space frequency coding [10] and space time frequency coding [29] has also been well investigated. In this section, we will carefully examine Alamouti space time coding for a communication system using two transmit antennas and one receive antenna. A STBC scheme for a 4×4 MIMO

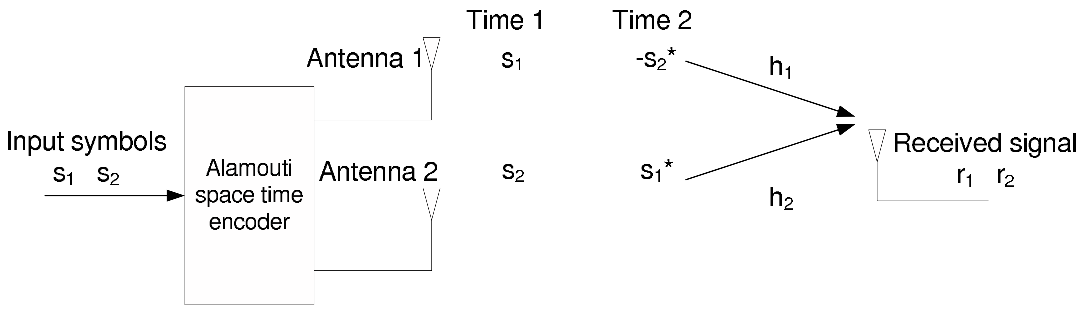


Figure 2.9: A block diagram of Alamouti space time encoder

communication system proposed in [38] is reviewed in Section 3.10. This STBC scheme will be used to examine the SER performance of the proposed training sequence design in Chapter 3. Readers who are interested in the topic of space time coding are recommended to look for further information in relevant textbooks [21] [43].

Alamouti code We consider a communication system with two transmit antennas and one receive antenna employing Alamouti code as in Figure 2.9.

First, the transmitter picks two symbols s_1 and s_2 from the input sequence. It then sends s_1 from antenna 1 and s_2 from antenna 2 at time 1. At time 2, it transmits $-s_2^*$ and s_1^* from Antenna 1 and 2, respectively. Accordingly, the transmitted codeword is

$$\mathbf{C} = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}.$$

It is important to note that the two columns of the square matrix $\mathbf{C}$ are orthogonal as their inner product is equal to zero,i.e.

$$s_1 s_2^* + (-s_2^* s_1) = 0.$$

Let us assume that the wireless channel is Rayleigh flat fading and the path gains from transmit antenna 1 and 2 to the receive antenna are respectively h_1 and h_2. We also assume that these gains are constant across two consecutive symbol transmission periods.

The received signals at time 1 and 2 can be expressed as:

$$\begin{aligned} r_1 &= h_1 s_1 + h_2 s_2 + n_1 \\ r_2 &= -h_1 s_2^* + h_2 s_1^* + n_2 \end{aligned}$$

where n_1 and n_2 represents AWGN noise at the receiver at time 1 and 2 respectively.

The received signals can be expressed in a vector-matrix form as follows.

$$\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$$

or

$$\mathbf{r} = \mathbf{Ch} + \mathbf{n}. \tag{2.22}$$

Take the negative complex conjugate of the second received signal, we have

$$\begin{pmatrix} r_1 \\ -r_2^* \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2^* \end{pmatrix}$$

or

$$\tilde{\mathbf{r}} = \mathbf{Hs} + \tilde{\mathbf{n}}. \tag{2.23}$$

We now assume that the channel gains h_1 and h_2 can be perfectly recovered at the receiver.

The decoding procedure can be performed by left-multiplying $\mathbf{H}^H$ as follows

$$\begin{aligned} \mathbf{y} &= \mathbf{H}^H \tilde{\mathbf{r}} \\ &= \mathbf{H}^H \mathbf{Hs} + \mathbf{H}^H \tilde{\mathbf{n}} \\ &= (|h_1|^2 + |h_2|^2)\mathbf{s} + \mathbf{w}. \end{aligned} \tag{2.24}$$

That is,

$$y_i = (|h_1|^2 + |h_2|^2) s_i + w_i, i = 1, 2. \tag{2.25}$$

We note that the statistics of the noise $\mathbf{w}$ in (2.24) is unchanged in comparison with the noise $\mathbf{n}$ in (2.22) due to a unitary transformation and conjugation operation.

It is clear from (2.25) that Alamouti space time coding provides the diversity order of 2 for a 2×1 Multiple Input Single Output (MISO) communication channel. It is straight forward to extend this scheme to the case of 2 transmit antennas and M_r receive antennas where the maximum diversity order is up to $2M_r$.

It is also very important to note that the Alamouti space time decoding procedure has low complexity as each symbol is decoded separately using only linear signal processing.

The performance of Alamouti code over a Rayleigh flat fading channel is provided in Figure 2.10. This figure shows the symbol error rate (SER) against SNR for wireless communication systems using QPSK constellation and different number of transmit and receive antennas. It is clear that the SER performance of communication systems using Alamouti code with two transmit antennas is much better than that of a SISO system.

2.4 MIMO-OFDM

2.4.1 Why MIMO and OFDM ?

MIMO techniques promise to bring a revolution to wireless communication with more reliable communication link obtained through diversity [4] [39] [38] and higher data rate by spatial multiplexing [40] [13] [14] [45]. However, a major challenge of MIMO system is the complexity of signal processing, especially at the receive side. Data transmission through a broadband wireless channel is expected to experience large delay spread and, thus, has to deal with frequency selectivity. Unfortunately, the computational complexity of MIMO symbol detection in a frequency selective fading channel grows exponentially with the bandwidth and delay spread [35] and becomes infeasible even for a small number of transmit/receive antennas. Hence, most MIMO signaling techniques developed so far are for frequency flat fading channel only.

Meanwhile, the basic idea of OFDM is to convert a frequency selective channel into a set of parallel flat fading channels. Moreover, the implementation of OFDM is very computationally efficient thanks to the use of IFFT/FFT algorithm.

Therefore, the combination of MIMO and OFDM is expected to exploit the MIMO

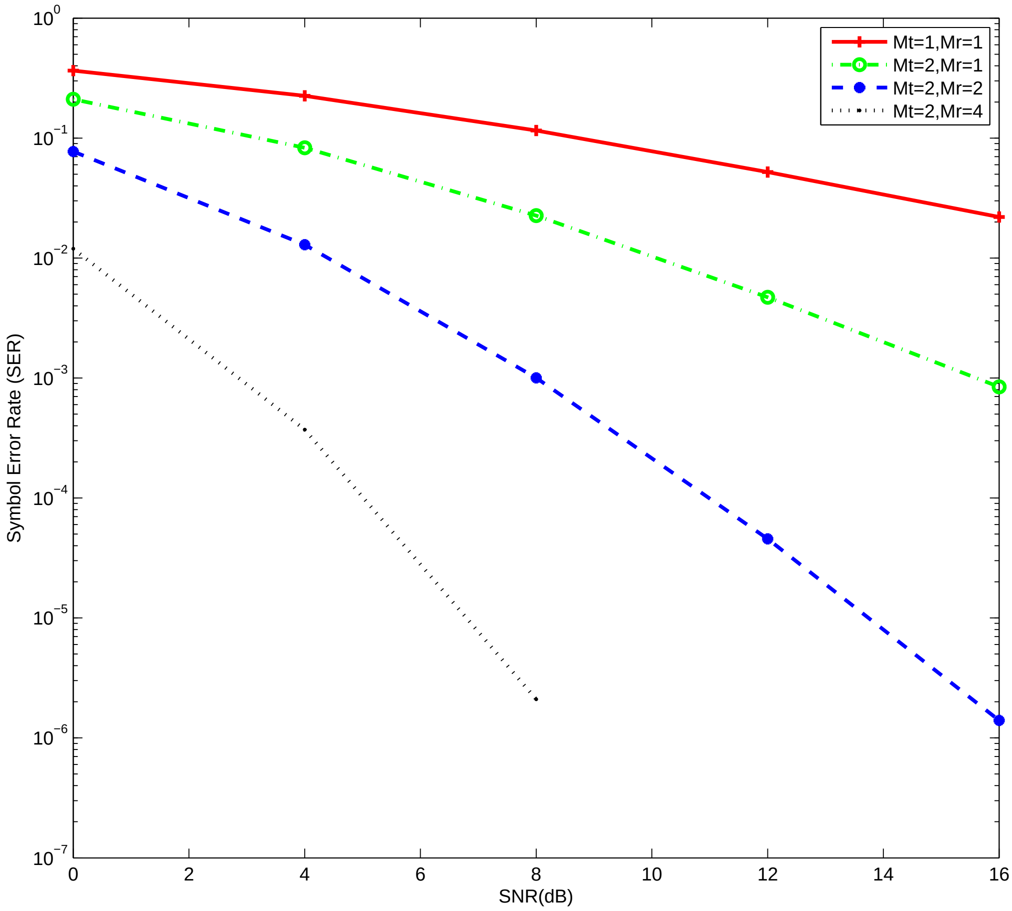


Figure 2.10: Performance of Alamouti space time code

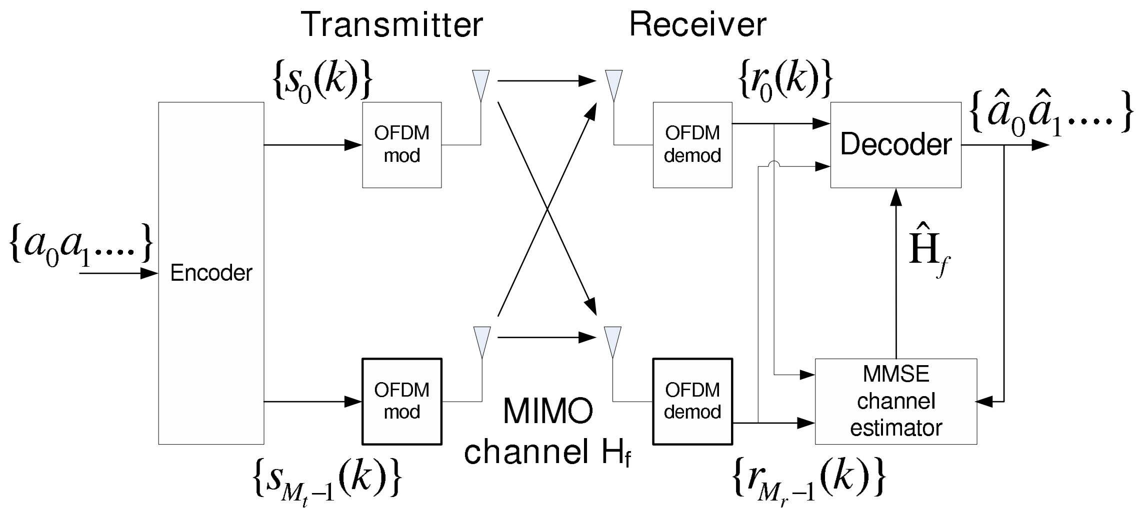


Figure 2.11: A typical MIMO-OFDM system model

merits at a reasonably and practically computational complexity. MIMO-OFDM is now recognized as one of the most promising air interface solutions for future broadband wireless communication systems.

2.4.2 MIMO-OFDM Signal Model

A typical MIMO-OFDM system is shown in Figure 2.11. The transmitted data sequence is first encoded into M_t sequences by an encoder such as a space–time encoder. Then the encoded sequence at each output of the encoder is divided into blocks of N symbols. Each block goes through an OFDM modulator to form an OFDM block. And each of these OFDM blocks is transmitted through a transmit antenna.

The MIMO-OFDM system has M_t transmit antennas and M_r receive antennas. N is the number of subchannels for each OFDM block. The OFDM cyclic prefix length is chosen to be longer than the channel order L to avoid Inter Block Interference (IBI).

The three-dimensional view of MIMO-OFDM transmitted signal is demonstrated in Figure 2.12.

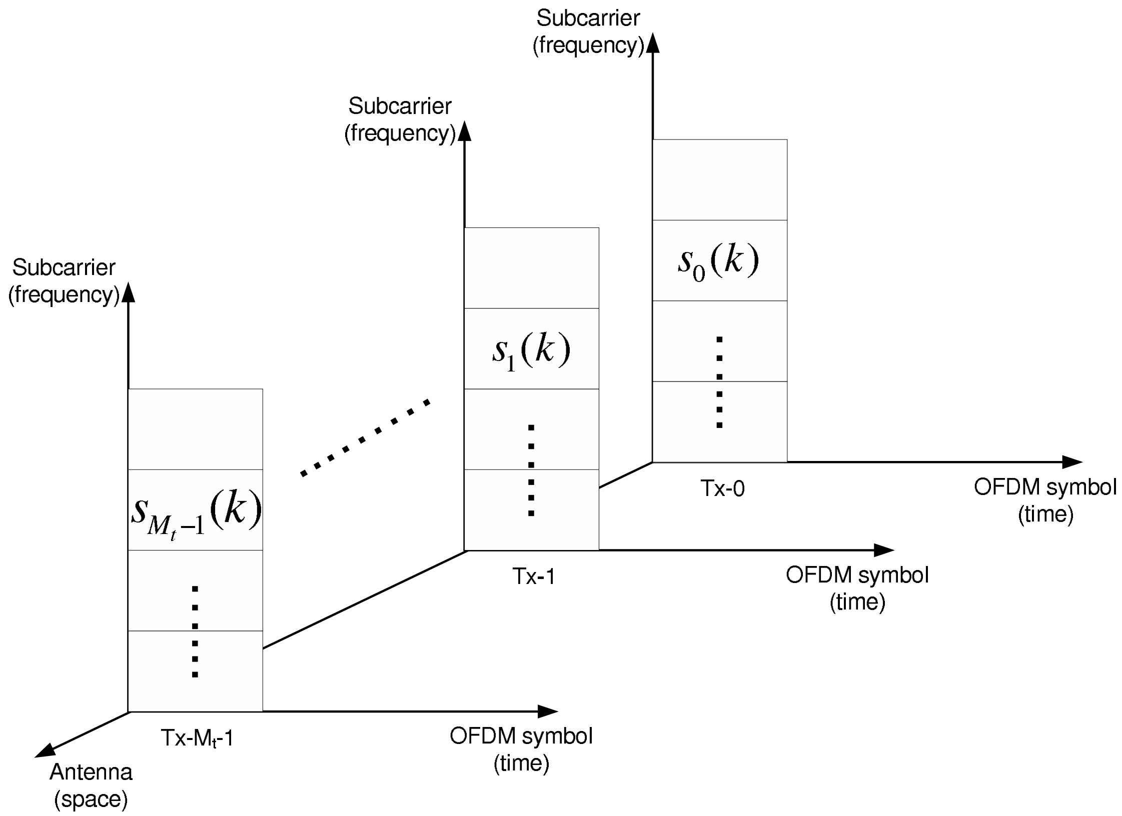


Figure 2.12: 3D view of MIMO-OFDM transmitted signal

For the k^{th} subchannel, received signal at the i^{th} receive antenna is

$$r_i(k) = \sum_{j=0}^{M_t-1} H_{ij}(k)s_j(k) + \tilde{n}_i(k)$$

or, equivalently,

$$r_i(k) = \mathbf{h}_i^T(k)\mathbf{s}(k) + \tilde{n}_i(k)$$

where $H_{ij}(k)$ is defined as in (2.18)

$$H_{ij}(k) = \sum_{l=0}^{L-1} h_{ij}(l)e^{\frac{-j2\pi kl}{N}}$$

and $h_{ij}(l), l = 0, \ldots, L-1$ is the channel impulse response from j^{th} transmit antenna to i^{th} receive antenna.

$\mathbf{h}_i^T(k) = \begin{pmatrix} H_{i0}(k) & H_{i1}(k) & \ldots & H_{i(M_t-1)}(k) \end{pmatrix}$ is a row vector.

Stacking signals at all receive antennas and normalize the noise terms to unit variance, we have

$$\mathbf{r}(k) = \sqrt{\frac{\rho}{M_t}}\mathbf{H}_f(k)\mathbf{s}(k) + \mathbf{n}(k) \tag{2.26}$$

where

ρ is the expected received SNR at each receive antenna.

$\mathbf{r}(k) = \begin{pmatrix} r_0(k) & r_1(k) & \ldots & r_{M_r-1}(k) \end{pmatrix}^T$ is the received signal vector.

$\mathbf{s}(k) = \begin{pmatrix} s_0(k) & s_1(k) & \ldots & s_{M_t-1}(k) \end{pmatrix}^T$ is the transmitted signal vector. Each entry of $\mathbf{s}(k)$ has unit mean square on average, i.e. $\sum_{k=0}^{N-1} \mathbf{s}(k)^H\mathbf{s}(k) = NM_t$.

$\mathbf{n}(k) = \begin{pmatrix} n_0(k) & n_1(k) & \ldots & n_{M_r-1}(k) \end{pmatrix}^T$ is the i.i.d complex Gaussian random noise vector with zero mean and unit variance $\mathcal{CN}(0, \mathbf{I}_{M_r})$.

$\mathbf{H}_f(k)$ is the MIMO channel frequency response matrix on the k^{th} subchannel

$$\mathbf{H}_f(k) = \sum_{l=0}^{L-1} \mathbf{H}_l e^{-j2\pi\frac{lk}{N}} \tag{2.27}$$

where $\mathbf{H}_l$ is the $M_r \times M_t$ channel matrix of the l^{th} tap, $l = 0, 1, \ldots, L-1$ with the coefficient $[\mathbf{H}_l]_{i,j}$ to be $h_{ij}(l)$ defined above.

2.5 Channel Estimation

2.5.1 Introduction to Channel Estimation

In MIMO-OFDM systems as well as other digital communication systems, CSI is needed for signal detection at the receiver. Its accuracy directly affects the overall performance of the system. In the previous sections, we had assumed perfect channel knowledge when analyzing the performance of MIMO systems. However, in practice, CSI has to be estimated.

There are two popular approaches to channel estimation: *decision-directed* and *pilot-symbol-aided*. Each of the two schemes has its own advantages and drawbacks. In the decision-directed channel estimation scheme, CSI is estimated by exploiting the statistical properties of the communication channel and the received information symbols. As no independent pilot signal is needed, the entire transmission session can be used to send information symbols. The decision directed channel estimation scheme is, therefore, ideally bandwidth efficient in theory. However, this channel estimation scheme has some critical drawbacks. Firstly, the computation based on second-order statistics of the channel and received signal is highly complicated. It is, therefore, impractical to implement decision directed channel estimation in real time communication systems. Secondly, decision-directed methods often rely on time averaging, thus are only suitable for slowly varying channel where the channel statistics do not change over a long period. It is well known that all decision directed channel estimation algorithms have very poor performance in fast fading communication channels. Consequently, decision-directed channel estimation methods have limited use in practice due to their high computational cost and restriction to some data and channel assumptions.

In pilot symbol aided channel estimation scheme, a pilot (or training) sequence, known to the receiver, is embedded into the transmitted signal frame and sent through the channel. At the receiver, CSI is estimated from the received signal and the known pilot sequence. Pilot symbol aided channel estimation algorithms, in general, have high accuracy and relatively low computational complexity. Therefore, this channel estimation scheme

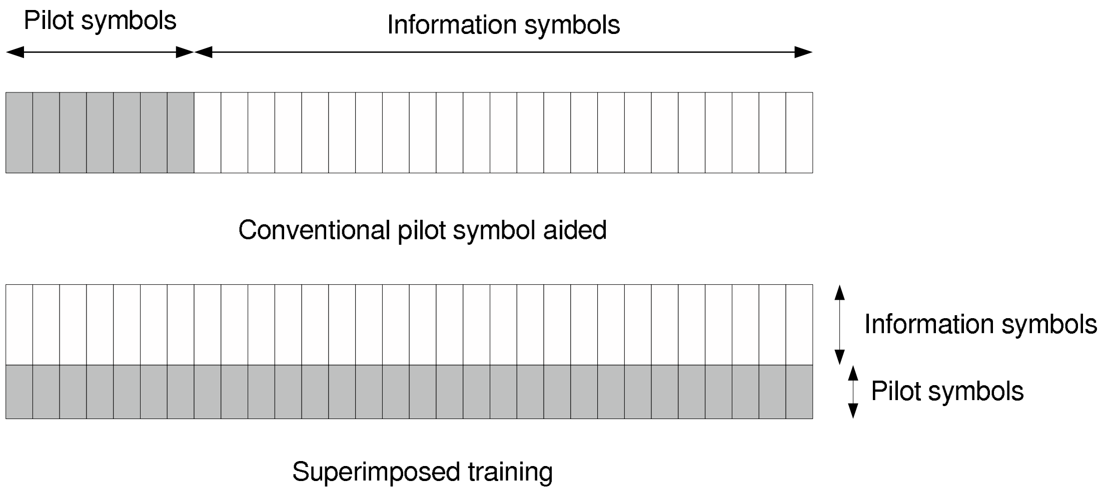


Figure 2.13: A schematic representation of channel estimation: conventional pilot symbol aided vs. superimposed training

is widely used in current wireless communication systems. A major drawback of pilot symbol aided channel estimation scheme is the reduced bandwidth efficiency due to the wasteful transmission of a known pilot sequence.

Recently, there is a growing interest in a novel channel estimation scheme called *superimposed training*. Figure 2.13 shows the difference between the conventional pilot symbol aided and superimposed training channel estimation schemes. In superimposed training channel estimation scheme, the pilot signal is superimposed on the information signal. Using this approach, the entire transmitted frame can still be used to transmit information symbols. Therefore, superimposed training scheme is considered as more bandwidth efficient than the conventional pilot symbol aided channel estimation scheme. However, some useful power, which could have been allocated for transmitting information signal, is wasted in transmitting the superimposed training sequence. The current superimposed training channel estimation methods are, in general, still far from practical implementation due to their complexity in decoupling information/training signal at receiver and the orthogonal vulnerability of precoding and training matrices.

The research in this thesis uses the conventional pilot symbol aided channel estimation scheme.

2.5.2 Choosing an Estimator

The general problem of estimation theory is described as follows.

Given an observed data set $\mathbf{y} = \left\{ y[0], y[1], \ldots, y[N-1] \right\}$ which depends on an unknown parameter x, find an optimal estimator of x, denoted by $\hat{x}$.

The literature on estimation theory consists of many different estimation methods which are based on different optimal criteria. The choice of a good estimator for a particular application depends upon many considerations [23]

- The selection of a good data model is of primary concern. It should be complex enough to precisely describe the nature of the data but at the same time simple enough to allow an estimator to be practically implemented.

- Prior knowledge of the estimated parameters is also very important in choosing an appropriate estimator. It is a fundamental rule of estimation theory that the use of prior knowledge will result in a more accurate estimator.

- Dimensionality of the estimation problem should also be taken into account in the decision making process. In many practical cases, it is better to choose a suboptimal estimator which can be easily implemented rather than an optimal one which is too computationally expensive.

There are two popular approaches in estimation theory: classical estimation and Bayesian estimation.

In the classical approach, the parameter to be estimated is viewed as deterministic but unknown. Accordingly, this approach assumes no prior knowledge on the estimated parameter. The optimal estimator in classical approach is a Minimum Variance Unbiased Estimator (MVUE). An unbiased estimator means that it will yield the true value of the unknown parameter on average. And the optimality criterion of an MVUE is the minimum variance between the estimated and true value of the parameter. The problem is that an MVUE does not always exist. Even if an MVUE exists, there is no known procedure to find the MVUE in general. Maximum Likelihood Estimator (MLE) is a well-known alternative to MVUE. MLE is desirable in situations where an MVUE does not exist or does exist but cannot be found. Based on the maximum likelihood criterion, the performance of an MLE is asymptotically optimal for large enough data set. However, a closed form expression for an MLE problem does not always exist. A numerical approach has to be employed in that case but the computational complexity is very high and the convergence of the numerical algorithm is not guaranteed in general.

In the Bayesian approach, the parameter to be estimated is considered as a random variable whose particular realisation must be estimated. By this way, we can incorporate some prior statistical knowledge about the parameter into our estimator. Therefore, the Bayesian approach can improve the estimation accuracy if the prior knowledge is appropriate. On the contrary, if the prior knowledge is unreliable, Bayesian approach will result in biased estimators.

We define a Bayesian cost function $C(x,\hat{x})$ to be a nonnegative real value function of the true parameter x and the estimated parameter $\hat{x}$. A typical example of cost function is the quadratic error of the estimator $(x-\hat{x})^2$, which is used in MMSE estimator.

A Bayesian risk function is defined to be the average value of the cost function

$$\mathscr{R} = E[C(x,\hat{x})].$$

The minimisation of this criterion using different cost functions results in a variety of estimators. In the next section, we will elaborate on MMSE estimator, the most important and most commonly used class of Bayesian estimators.

2.5.3 MMSE Estimation

MMSE estimator Given an observed data set $\mathbf{y} = \left\{ y[0], y[1], \ldots, y[N-1] \right\}$ which depends on an unknown parameter x and assuming that the estimation parameter x to be a random variable with known PDF, find an estimator $\hat{x}$ of x in order to minimize the mean square error

$$\mathrm{MSE}(\hat{x}) = E[(x-\hat{x})^2]. \tag{2.28}$$

The optimal estimator $\hat{x}$ regarding the MMSE criterion is the mean of the posterior PDF $p[x/\mathbf{y})$

$$\hat{x} = E[x/\mathbf{y}]. \tag{2.29}$$

For a wireless fading channel, MMSE estimator is found to be the most accurate and practical one for the following reasons:

- Prior knowledge of the estimation parameters can be incorporated into the MMSE estimator, thus, leading to a more accurate estimation result.
- MMSE estimator always exists and is optimal on average with respect to the assumed prior Probability Distribution Function (PDF) of the estimation parameters.
- An MMSE estimator can be found easily under the jointly Gaussian distribution assumption.

2.5.4 LMMSE Estimation

It is, in general, difficult to express MMSE estimator in a closed form. And the implementation of MMSE estimator is often very computationally expensive as expression (2.29) usually involves multidimensional integration except under the jointly Gaussian distribution assumption.

For practical purpose, we can retain the MMSE criterion but constrain the estimator to be linear [23]. In other word, we can choose a linear estimator

$$\hat{x} = \sum_{n=0}^{N-1} a_n y[n] + a_N \tag{2.30}$$

in order to minimize the MSE (2.28)

This class of estimators are called LMMSE estimators.

If x and $\mathbf{y}$ are zero mean, the LMMSE estimator is found as [23]

$$\hat{x} = \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{y} \tag{2.31}$$

and

$$\text{MSE} = R_x - \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{yx} \tag{2.32}$$

where

R_x is the variance of x,

$\mathbf{R}_{xy}$ and $\mathbf{R}_{yx}$ are respectively the cross-covariance vector between x and $\mathbf{y}$ and vice versa. And we have $\mathbf{R}_{xy} = \mathbf{R}_{yx}^T$,

$\mathbf{R}_y$ is the $N \times N$ covariance matrix of $\mathbf{y}$.

2.5.5 LMMSE Vector Estimation

In a more general case, the vector LMMSE estimator is a straightforward extension of the scalar LMMSE estimator. It is summarized in the following theorem.

Bayesian Gauss-Markov Theorem [23]

If the data are described by Bayesian linear model form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.33}$$

where

$\mathbf{y}$ is an $N \times 1$ data vector,

$\mathbf{H}$ is a known $N \times p$ matrix,

$\mathbf{x}$ is a $p \times 1$ random vector whose realization is to be estimated and has zero mean and covariance matrix $\mathbf{R}_x$,

$\mathbf{n}$ is an $N \times 1$ random noise vector with zero mean and covariance matrix $\mathbf{R}_n$; $\mathbf{n}$ is uncorrelated with $\mathbf{x}$.

Then, the LMMSE estimator of $\mathbf{x}$ is

$$\hat{\mathbf{x}} = \mathbf{R}_x\mathbf{H}^H(\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{R}_n)^{-1}\mathbf{y} \tag{2.34}$$

and the MSE performance is

$$\begin{aligned} \mathrm{MSE} &= E\left[\ (\mathbf{x} - \hat{\mathbf{x}})^H(\mathbf{x} - \hat{\mathbf{x}})\ \right] \\ &= \mathrm{trace}\left\{\ \mathbf{R}_x - \mathbf{R}_x\mathbf{H}^H(\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{R}_n)^{-1}\mathbf{H}\mathbf{R}_x\ \right\}. \end{aligned} \tag{2.35}$$

If $\mathbf{R}_x$ and $\mathbf{R}_n$ are invertible, using the matrix inversion lemma (1.6), the LMMSE estimator of $\mathbf{x}$ can be re-written in a more convenient form

$$\hat{\mathbf{x}} = (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{y} \tag{2.36}$$

and the MSE performance is

$$\begin{aligned} \text{MSE} &= E\left[\ (\mathbf{x} - \hat{\mathbf{x}})^H (\mathbf{x} - \hat{\mathbf{x}}) \ \right] \\ &= \text{trace}\left\{ \ (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H})^{-1} \ \right\}. \end{aligned} \tag{2.37}$$

It can be seen that an LMMSE estimator is always determined in an explicit form and only depends on the first two moments of the PDFs. It is important to note that LMMSE estimator does not require Gaussian distribution assumption. The performance of an LMMSE estimator is identical to that of an MMSE estimator under the jointly Gaussian distribution assumption. Consequently, LMMSE is widely used for channel estimation in practice. In this thesis, we apply LMMSE estimator for MIMO-OFDM channel estimation.

2.6 MIMO Spatially Correlated Fading Channel Model

The simplest and most common probabilistic model for a MIMO fading channel is the i.i.d. Rayleigh fading model [42] where the entries of MIMO channel matrix $\mathbf{H}$ are independent, identically distributed and circular symmetric complex Gaussian. However, this fading model is only true in a richly scattered environment where there are a large number of statistically independent reflected and scattered transmission paths. Under this assumption, the mathematical calculations corresponding to MIMO techniques are significantly simplified. In practice, correlation does exist between different entries of the channel matrix as the antennas are often not spaced far enough or the wireless environment is not scattered sufficiently. Therefore, the i.i.d. Rayleigh fading channel model is not an accurate description of real-world MIMO fading channels.

In this section, we will consider a widely accepted MIMO spatially correlated fading channel model as presented in [9] [8]. This fading model represents a good balance between accurate simulating spatially correlated fading channels and being expressed in a simple analytical form.

A schematic representation of a MIMO spatially correlated fading channel is shown in figure 2.14. For simplicity, only relevant angles for transmit array are shown. Those of receive array are similar. We consider that each channel tap is corresponding to a significant scatterer cluster.

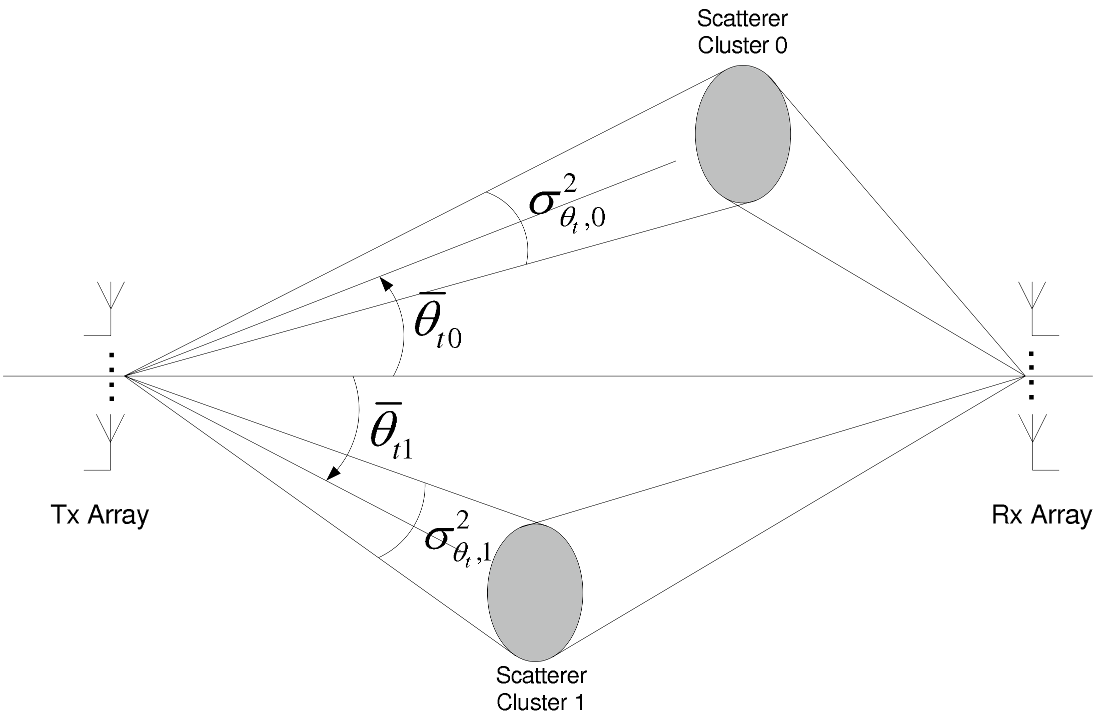


Figure 2.14: MIMO spatially correlated fading channel model composed of 2 clustered paths

It is assumed that the channel is a Rayleigh block fading channel. This means that the elements of the channel matrix $[\mathbf{H}_l]_{m,n}$ are (possibly correlated) circular symmetric zero-mean complex Gaussian random variables. In addition, the channel remains constant over a period of multiple of N signal symbols and then changes randomly to a new state which is independent of the previous ones. We also assume that spatially correlated fading occurs at both transmitter and receiver. Under these assumptions, according to [9], the

MIMO channel matrix can be decomposed to

$$\mathbf{H}_l = \mathbf{R}_{rl}^{1/2}\,\mathbf{H}_{wl}\,(\mathbf{R}_{tl}^{1/2})^T \tag{2.38}$$

where

$\mathbf{H}_{wl}$ is an $M_r \times M_t$ matrix whose elements are i.i.d complex Gaussian variables with zero mean and unit variance $\mathscr{C}\mathscr{N}(0,1)$,

$\mathbf{R}_{tl} = \mathbf{R}_{tl}^{1/2}\mathbf{R}_{tl}^{1/2}$ and $\mathbf{R}_{rl} = \mathbf{R}_{rl}^{1/2}\mathbf{R}_{rl}^{1/2}$ are the transmit and receive correlation matrix, respectively.

Elements of the correlation matrices $\mathbf{R}_{tl}$ and $\mathbf{R}_{rl}$ are given by

$$\begin{aligned} [\mathbf{R}_{tl}]_{m,n} &= \sqrt{\sigma_l^2}\,\xi\left(\ (n-m)\Delta_t, \bar{\theta}_{tl}, \sigma_{\theta_t,l}\ \right), \\ [\mathbf{R}_{rl}]_{m,n} &= \sqrt{\sigma_l^2}\,\xi\left(\ (n-m)\Delta_r, \bar{\theta}_{rl}, \sigma_{\theta_r,l}\ \right). \end{aligned} \tag{2.39}$$

where

$\Delta_t = \frac{d_t}{\lambda}$, $\Delta_r = \frac{d_r}{\lambda}$ are the relative transmit and receive antenna spacing respectively, where d_t and d_r stand for absolute antenna spacing and λ is the wavelength of the carrier. Here it is assumed that uniform linear antenna arrays are used at both the transmitter and receiver,

$\bar{\theta}_{tl}$ and $\bar{\theta}_{rl}$ are the mean angle of departure from the transmit array and the mean angle of arrival at the receive array respectively,

$\sigma_{\theta_t,l}^2$ and $\sigma_{\theta_r,l}^2$ are the cluster angle spread perceived by the transmitter and receiver respectively,

σ_l^2 is the l^{th}-path power delay profile.

Assuming real angles of departure from the transmitter and real angles of arrival at the receiver are Gaussian distributed around the mean angle $\bar{\theta}$, it is shown in [9] that for small cluster angle spread, we have

$$\xi\left(s\Delta, \bar{\theta}, \sigma_\theta\right) \approx e^{-j2\pi s\Delta cos(\bar{\theta})}\, e^{-\frac{1}{2}(2\pi s\Delta sin(\bar{\theta})\sigma_\theta)^2}. \tag{2.40}$$

It can be seen from the above formula that large antenna spacing and/or large cluster angle spread result in low spatial fading correlation and vice versa.

As statistics of MIMO channels change slowly with time [6], the correlation matrices can be estimated at the receiver and fed back to the transmitter. Therefore, it is assumed that information of the correlation matrices is available at both transmitter and receiver for channel estimation and training sequence design.

2.7 LMMSE Channel Estimation for MIMO-OFDM

We refer again to the MIMO-OFDM system model in Figure 2.11 and 2.12. The MIMO-CP-OFDM signal model provided in (2.26) and (2.27) is duplicated here

$$\mathbf{r}(k) = \sqrt{\frac{\rho}{M_t}}\mathbf{H}_f(k)\mathbf{s}(k) + \mathbf{n}(k)$$

where

$$\mathbf{H}_f(k) = \sum_{l=0}^{L-1} \mathbf{H}_l e^{-j2\pi\frac{lk}{N}}.$$

During the training period, the transmitted sequences $\mathbf{s}(k), k = 0, 1, \ldots, N-1$ are known to the receiver. Therefore, the channel matrices $\mathbf{H}_l,\ l = 0, 1, \ldots, L-1$ can be estimated from Equation (2.26). The LMMSE channel estimation model for MIMO-OFDM with correlated fading is presented in detail as below.

First, using the property of Kronecker product as in Section A.1.2: $\mathbf{H}_l\,\mathbf{s}(k) = (\mathbf{s}^T(k) \otimes \mathbf{I}_{M_r})\,\mathrm{vec}(\mathbf{H}_l)$, Equation (2.26) can be re-written as

$$\mathbf{r}(k) = \sqrt{\frac{\rho}{M_t}}\mathbf{M}(k)\mathbf{h} + \mathbf{n}(k) \tag{2.41}$$

where

$\mathbf{h} = \begin{pmatrix} \mathrm{vec}(\mathbf{H}_0)^T & \mathrm{vec}(\mathbf{H}_1)^T & \ldots & \mathrm{vec}(\mathbf{H}_{L-1})^T \end{pmatrix}^T \in \mathbf{C}^{LM_tM_r \times 1}$,

$\mathbf{M}(k) = \begin{pmatrix} \mathbf{M}_0(k), & \mathbf{M}_1(k), & \ldots, & \mathbf{M}_{L-1}(k) \end{pmatrix}$ with $\mathbf{M}_l(k) = e^{-j2\pi\frac{kl}{N}}\,\mathbf{s}^T(k) \otimes \mathbf{I}_{M_r}$.

Define

$\mathbf{r} = \begin{pmatrix} \mathbf{r}(0)^T & \mathbf{r}(1)^T & \ldots & \mathbf{r}(N-1)^T \end{pmatrix}^T \in \mathbf{C}^{NM_r \times 1}$,

$\mathbf{n} = \begin{pmatrix} \mathbf{n}(0)^T & \mathbf{n}(1)^T & \ldots & \mathbf{n}(N-1)^T \end{pmatrix}^T \in \mathbf{C}^{NM_r \times 1}$,

$\mathbf{M} = \begin{pmatrix} \mathbf{M}(0)^T & \mathbf{M}(1)^T & \ldots & \mathbf{M}(N-1)^T \end{pmatrix}^T \in \mathbf{C}^{NM_r \times LM_tM_r}$.

Stacking received signal vectors of all OFDM subchannels, Equation (2.41) can be expressed as

$$\mathbf{r} = \sqrt{\frac{\rho}{M_t}}\mathbf{M}\mathbf{h} + \mathbf{n}. \tag{2.42}$$

This equation represents a Bayesian linear model as in (2.33). Therefore, we can apply the LMMSE channel estimation for MIMO-OFDM with spatially correlated fading channels.

Now, we have to find the correlation matrix of $\mathbf{h}$. Applying the Kronecker product property: $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{X})$ into Equation (2.38), we have

$$\mathrm{vec}(\mathbf{H}_l) = (\mathbf{R}_{tl}^{1/2} \otimes \mathbf{R}_{rl}^{1/2})\,\mathrm{vec}(\mathbf{H}_{wl}).$$

Note that $\mathbf{H}_{wl},\ l = 0,\ldots,L-1$ are $M_r \times M_t$ matrices whose elements are i.i.d complex Gaussian variables $\mathscr{CN}(0,1)$. Thus, the correlation matrix of $\mathbf{h}$ in (2.42) is given by

$$\begin{aligned}\mathbf{R}_h &= E\{\mathbf{h}\mathbf{h}^H\}\\ &= \begin{pmatrix} \mathbf{R}_{t0}\otimes\mathbf{R}_{r0} & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \mathbf{R}_{t(L-1)}\otimes\mathbf{R}_{r(L-1)}\end{pmatrix}\end{aligned} \tag{2.43}$$

where $\mathbf{R}_{tl}$ and $\mathbf{R}_{rl}$ are determined as in (2.39).

According to (2.34), the LMMSE estimator of the channel vector $\mathbf{h}$ in (2.42) is determined by

$$\hat{\mathbf{h}} = \sqrt{\frac{\rho}{M_t}}\mathbf{R}_h\mathbf{M}^H(\frac{\rho}{M_t}\mathbf{M}\mathbf{R}_h\mathbf{M}^H + \mathbf{I}_{NM_r})^{-1}\mathbf{r}. \tag{2.44}$$

If $\mathbf{R}_h$ is invertible, we can rewrite $\hat{\mathbf{h}}$ as follows

$$\hat{\mathbf{h}} = \sqrt{\frac{\rho}{M_t}}(\mathbf{R}_h^{-1} + \frac{\rho}{M_t}\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{r} \tag{2.45}$$

and the MSE is

$$\begin{aligned}\mathrm{MSE} &= E\left\{\ (\hat{\mathbf{h}}-\mathbf{h})^H(\hat{\mathbf{h}}-\mathbf{h})\ \right\}\\ &= \mathrm{trace}\left\{\ (\mathbf{R}_h^{-1} + \tfrac{\rho}{M_t}\mathbf{M}^H\mathbf{M})^{-1}\ \right\}.\end{aligned} \tag{2.46}$$

It is important to note that $\mathbf{R}_h$ can approach singularity if the channel fading is highly correlated, i.e. small antenna spacing and/or small cluster angle spread. We will also consider this case in Section 3.3 of the next chapter.

2.8 MIMO-ZP-OFDM-OLA and LMMSE Channel Estimation

A dual of CP-OFDM is ZP-OFDM-OLA. While CP-OFDM cyclic prefixing idea derives from the fast convolution algorithm based on the *OverLap-Save* (OLS) method, ZP-OFDM-OLA is based on the *OverLap-Add* (OLA) method [36] [34] [44] [32]. ZP-OFDM-OLA has equivalent computational complexity as CP-OFDM and is an alternative solution for CP-OFDM in practical wireless systems. We will elaborate on the MIMO-ZP-OFDM signal model and its LMMSE channel estimation in spatially correlated fading environments in the following sections.

2.8.1 ZP-OFDM-OLA Signal Model

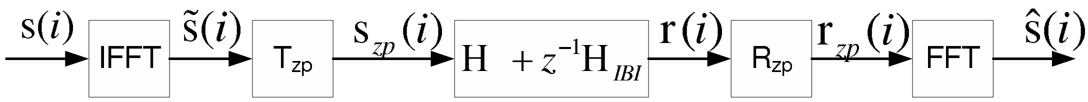


Figure 2.15: A mathematical view of ZP-OFDM-OLA

Figure 2.15 depicts the baseband discrete time block equivalent model of a ZP-OFDM-OLA system. The major difference to CP-OFDM is that the guard interval is L trailing zeros instead of a cyclic prefix. The trailing zeros are inserted by multiplication with a zero-padded matrix $\mathbf{T}_{zp}$ at the transmit side and removed by multiplication with a receive matrix $\mathbf{R}_{zp}$ at the receive side. $\mathbf{T}_{zp}$ and $\mathbf{R}_{zp}$ are defined as follows

$$\mathbf{T}_{zp} = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{0}_{L\times N} \end{pmatrix} \quad \text{and} \quad \mathbf{R}_{zp} = \begin{pmatrix} \mathbf{I}_N & \begin{matrix} \mathbf{I}_L \\ \mathbf{0}_{N-L\times L} \end{matrix} \end{pmatrix}. \tag{2.47}$$

According to the signal flow as shown in Figure 2.15, the frequency domain received signal block $\hat{\mathbf{s}}(i)$ is written as

$$\begin{aligned}
\hat{\mathbf{s}}(i) &= \mathbf{F}_N \mathbf{R}_{zp} \left(\mathbf{H}\mathbf{T}_{zp}\mathbf{F}_N^H \mathbf{s}(i) + \mathbf{H}_{IBI}\mathbf{T}_{zp}\mathbf{F}_N^H \mathbf{s}(i-1) + \mathbf{w}(i) \right) && (2.48)\\
&= \mathbf{F}_N \mathbf{R}_{zp}\mathbf{H}\mathbf{T}_{zp}\mathbf{F}_N^H \mathbf{s}(i) + \mathbf{F}_N \mathbf{R}_{zp}\mathbf{w}(i) && (2.49)\\
&= \mathbf{F}_N \tilde{\mathbf{H}}\mathbf{F}_N^H \mathbf{s}(i) + \mathbf{F}_N \bar{\mathbf{w}}(i) && (2.50)\\
&= \mathrm{diag}\left(H_0 \quad H_1 \quad \ldots \quad H_{N-1} \right) \mathbf{s}(i) + \tilde{\mathbf{w}}(i). && (2.51)
\end{aligned}$$

The first simplifying step from (2.48) to (2.49) results from the property that $\mathbf{R}_{zp}\mathbf{H}_{IBI} = \mathbf{0}$. In the second step, we have $\mathbf{R}_{zp}\mathbf{H}\mathbf{T}_{zp} = \tilde{\mathbf{H}}$, which is an $N \times N$ circulant matrix with the first row $\left(h(0) \quad 0 \quad \ldots \quad 0 \quad h(L-1) \quad \ldots \quad h(1) \right)$. Therefore, the first term in (2.51) is achieved according to (2.17).

We now consider the noise term. To simplify the notation, we will omit the block index i. Similar to the noise term in (2.11), $\mathbf{w} \sim \mathscr{CN}(0, \sigma^2 \mathbf{I}_P)$ is a $P \times 1$ AWGN noise vector. Then, we have

$$\begin{aligned}
\bar{\mathbf{w}} &= \mathbf{R}_{zp}\mathbf{w}\\
&= \left(w_0 + w_N, \quad \ldots, \quad w_{L-1} + w_{N+L-1}, \quad w_L, \quad \ldots, \quad w_{N+L-1} \right)^T. && (2.52)
\end{aligned}$$

Accordingly, $\bar{\mathbf{w}}$ is a random vector with zero mean and covariance matrix

$$\begin{aligned}
\mathbf{R}_{\bar{w}} &= E[\bar{\mathbf{w}}\bar{\mathbf{w}}^H]\\
&= \begin{pmatrix} 2\sigma^2 \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{N-L} \end{pmatrix}. && (2.53)
\end{aligned}$$

The noise term $\tilde{\mathbf{w}}(i)$ in (2.51) has the same distribution as $\bar{\mathbf{w}}$ as a unitary transformation does not change the statistics of a random noise as specified in (2.19).

From the above analysis, we can draw the following conclusions about ZP-OFDM-OLA

- ZP-OFDM-OLA can be decoded in the same way as CP-OFDM since the received signal in (2.51) has a similar form as the received signal in (2.20)

- The spectral efficiency of ZP-OFDM-OLA is the same as that of CP-OFDM if the length of zero-padding interval is equal to the length of cyclic prefix.

- The advantage of ZP-OFDM-OLA in comparison with CP-OFDM is to save transmit power. While CP-OFDM has to transmit the cyclic prefix as a guard interval, ZP-OFDM-OLA does not spend power for zero-padding interval. In other words, the power efficiency rate of ZP-OFDM-OLA is 1 in comparison with $\frac{N}{N+L}$ in CP-OFDM.

- However, in ZP-OFDM-OLA systems, colored noise is incurred at the receiver. This may be problematic for the receiver.

2.8.2 LMMSE Channel Estimation for MIMO-ZP-OFDM-OLA in Spatially Correlated Fading Environments

If ZP-OFDM-OLA replaces CP-OFDM in MIMO-OFDM systems, the ”frequency domain” received signal is expressed in a similar form as (2.26)

$$\mathbf{r}(k) = \sqrt{\frac{\rho}{M_t}}\mathbf{H}_f(k)\mathbf{s}(k) + \tilde{\mathbf{n}}(k). \tag{2.54}$$

All the terms are defined exactly the same as in (2.26) except the following two points.

First, as the power efficiency rate of ZP-OFDM-OLA is higher than that of CP-OFDM, the total transmit power constraint of the MIMO-ZP-OFDM-OLA system is $\sum_{k=0}^{N-1}\mathbf{s}(k)^H\mathbf{s}(k) = (N+L)M_t$ as opposed to NM_t in the standard MIMO-OFDM system.

Second, the noise term $\tilde{\mathbf{n}}(k)$ is colored. We have

$$\tilde{\mathbf{n}}(k) = \begin{cases} \mathbf{n}(k) + \mathbf{n}(N+k) & \text{if } 0 \le k \le L-1 \\ \mathbf{n}(k) & \text{if } L \le k \le N-1 \end{cases} \tag{2.55}$$

where $\mathbf{n}(k)$ is the i.i.d complex Gaussian random noise vector with zero mean and unit variance $\mathscr{CN}(0, \mathbf{I}_{M_r})$.

We define

$$\tilde{\mathbf{n}} = \begin{pmatrix} \tilde{\mathbf{n}}(0)^T & \tilde{\mathbf{n}}(1)^T & \ldots & \tilde{\mathbf{n}}(N-1)^T \end{pmatrix}^T \in \mathbf{C}^{NM_r \times 1}. \tag{2.56}$$

Then, the covariance of $\tilde{\mathbf{n}}$ is expressed as

$$\begin{aligned} \mathbf{R}_{\tilde{n}} &= E\{\tilde{\mathbf{n}}^H\tilde{\mathbf{n}}\} \\ &= \operatorname{diag}\begin{pmatrix} E\{\tilde{\mathbf{n}}(0)^H\tilde{\mathbf{n}}(0)\} & \ldots & E\{\tilde{\mathbf{n}}(N-1)^H\tilde{\mathbf{n}}(N-1)\} \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
\text{For } 0 \le k \le L-1: \quad E\{\tilde{\mathbf{n}}(k)^H \tilde{\mathbf{n}}(k)\} &= 2\mathbf{I}_{M_r} \\
\text{For } L \le k \le N-1: \quad E\{\tilde{\mathbf{n}}(k)^H \tilde{\mathbf{n}}(k)\} &= \mathbf{I}_{M_r}
\end{aligned}
$$

Accordingly, the covariance of $\tilde{\mathbf{n}}$ is determined as

$$
\mathbf{R}_{\tilde{n}} = \begin{pmatrix} 2\mathbf{I}_{LM_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(N-L)M_r} \end{pmatrix}. \tag{2.57}
$$

Using similar mathematical manipulation as in Section 2.7, the MIMO-ZP-OFDM-OLA received signal can be rewritten as

$$
\mathbf{r} = \sqrt{\frac{\rho}{M_t}}\mathbf{M}\mathbf{h} + \tilde{\mathbf{n}} \tag{2.58}
$$

where all the terms are determined exactly in the same manner as in (2.42) except $\tilde{\mathbf{n}}$ defined as in (2.56).

According to (2.36) and (2.37), the LMMSE channel estimation for the MIMO-ZP-OFDM-OLA system in spatially correlated fading environments is determined by

$$
\hat{\mathbf{h}} = \sqrt{\frac{\rho}{M_t}}(\mathbf{R}_h^{-1} + \frac{\rho}{M_t}\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M})^{-1}\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{r} \tag{2.59}
$$

and the channel estimation MSE is

$$
\begin{aligned}
\text{MSE} &= E\left[\ (\hat{\mathbf{h}} - \mathbf{h})^H(\hat{\mathbf{h}} - \mathbf{h})\ \right] \\
&= \text{trace}\left\{\ (\mathbf{R}_h^{-1} + \tfrac{\rho}{M_t}\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M})^{-1}\ \right\}.
\end{aligned} \tag{2.60}
$$

Different to the MIMO-CP-OFDM case, the LMMSE channel estimation for MIMO-ZP-OFDM-OLA systems in spatially correlated fading environments is affected by colored noise. This can be seen from the existence of the covariance noise term $\mathbf{R}_{\tilde{n}}$ in (2.59) and (2.60). The optimal training sequence design for MIMO-ZP-OFDM-OLA systems which will be presented in Section 3.7 will have to address the affect of colored noise onto the MSE performance of training sequences.

Chapter 3

Optimal Training Sequence Design for MIMO-OFDM in Spatially Correlated Fading Environments

The objective of optimal training sequence design for MIMO-OFDM in spatially correlated fading environments is to find training sequences $\mathbf{s}(k),\ k = 0, 1, \ldots, N-1$, or equivalently a training symbol matrix $\mathbf{S} = \begin{pmatrix} \mathbf{s}(0) & \mathbf{s}(1) & \ldots & \mathbf{s}(N-1) \end{pmatrix}^T \in \mathbf{C}^{N \times M_t}$, to minimize the MSE of channel estimation under the transmit power constraint. It is equivalent to the following mathematical optimization problem

$$\begin{aligned} \min_{\mathbf{S} \in \mathbf{C}^{N \times M_t}} \quad & \mathrm{trace}\{(\mathbf{R}_h^{-1} + \tfrac{\rho}{M_t}\mathbf{M}^H\mathbf{M})^{-1}\} \\ \text{subject to} \quad & \mathrm{trace}\{\mathbf{S}^H\mathbf{S}\} = NM_t. \end{aligned} \tag{3.1}$$

3.1 Conditions for Optimal Training Sequences

As the transmit and receive correlation matrices $\mathbf{R}_{tl}$ and $\mathbf{R}_{rl}$ are Hermitian Toeplitz positive-definite matrices, they are unitarily diagonalizable.

$$\mathbf{R}_{tl} = \mathbf{U}_l \mathbf{\Lambda}_{tl} \mathbf{U}_l^H \quad \text{and} \quad \mathbf{R}_{rl} = \mathbf{V}_l \mathbf{\Lambda}_{rl} \mathbf{V}_l^H \tag{3.2}$$

where

$\mathbf{U}_l$ and $\mathbf{V}_l$ are respectively $M_t \times M_t$ and $M_r \times M_r$ unitary matrices,

$\mathbf{\Lambda}_{tl}$ and $\mathbf{\Lambda}_{rl}$ are diagonal matrices with the diagonal elements to be the eigenvalues of $\mathbf{R}_{tl}$ and $\mathbf{R}_{rl}$ respectively.

According to the property of Kronecker product in Section A.1.2, we have

$$\mathbf{R}_{tl} \otimes \mathbf{R}_{rl} = (\mathbf{U}_l \otimes \mathbf{V}_l)(\mathbf{\Lambda}_{tl} \otimes \mathbf{\Lambda}_{rl})(\mathbf{U}_l \otimes \mathbf{V}_l)^H. \tag{3.3}$$

Applying (3.3) into (2.43), the eigen-decomposition of $\mathbf{R}_h$ is written as

$$\mathbf{R}_h = \mathbf{U}_h \mathbf{\Lambda}_h \mathbf{U}_h^H \tag{3.4}$$

where

$$\begin{aligned} \mathbf{U}_h &= \operatorname{diag}\Big(\ \mathbf{U}_0 \otimes \mathbf{V}_0, \quad \ldots, \quad \mathbf{U}_{L-1} \otimes \mathbf{V}_{L-1} \ \Big), \\ \mathbf{\Lambda}_h &= \operatorname{diag}\Big(\ \mathbf{\Lambda}_{t0} \otimes \mathbf{\Lambda}_{r0}, \quad \ldots, \quad \mathbf{\Lambda}_{t(L-1)} \otimes \mathbf{\Lambda}_{r(L-1)} \ \Big). \end{aligned}$$

Hence, the MSE of channel estimation in Equation (2.46) can be re-written as

$$\mathrm{MSE} = \operatorname{trace}\Big\{ \ (\mathbf{\Lambda}_h^{-1} + \tfrac{\rho}{M_t} \mathbf{U}_h^H \mathbf{M}^H \mathbf{M} \mathbf{U}_h)^{-1} \ \Big\}. \tag{3.5}$$

Lemma [18]: From the inverse matrix formulae (p.14 of [48]), for a positive definite matrix $\mathbf{X} = [x_{ij}] \in \mathbf{C}^{n \times n}$, the following inequality holds true

$$\operatorname{trace}\{\mathbf{X}^{-1}\} \ \geq \ \sum_{i=1}^{n} x_{ii}^{-1}. \tag{3.6}$$

The equality occurs if and only if $\mathbf{X}$ is a diagonal matrix.

From the above lemma, it can be seen that the MSE in (3.5) will be minimized if $\mathbf{U}_h^H \mathbf{M}^H \mathbf{M} \mathbf{U}_h$ is a diagonal matrix [24] [41] [46] .

For simplification, we can partition matrix $\mathbf{M}\mathbf{U}_h$ as follows

$$\mathbf{M}\mathbf{U}_h = \begin{pmatrix} \mathbf{A}_0(0) & \ldots & \mathbf{A}_{L-1}(0) \\ \vdots & \ddots & \vdots \\ \mathbf{A}_0(N-1) & \ldots & \mathbf{A}_{L-1}(N-1) \end{pmatrix} \tag{3.7}$$

where

$$\mathbf{A}_l(k) = \mathbf{M}_l(k)\,(\mathbf{U}_l \otimes \mathbf{V}_l) = e^{-j2\pi\frac{kl}{N}} \left(\ \mathbf{s}^T(k)\,\mathbf{U}_l \ \right) \otimes \mathbf{V}_l. \tag{3.8}$$

Define $\mathbf{A}_l$ as a block column of matrix $\mathbf{MU}_h$, i.e.

$$\mathbf{A}_l = \left(\ \mathbf{A}_l(0)^T \quad \ldots \quad \mathbf{A}_l(N-1)^T \ \right)^T. \tag{3.9}$$

$\mathbf{A}_l$ can also be expressed as

$$\mathbf{A}_l = (\mathbf{D}_l\,\mathbf{S}\,\mathbf{U}_l) \otimes \mathbf{V}_l \tag{3.10}$$

where

$$\mathbf{D}_l = \mathrm{diag}\left(\ e^{-j2\pi\frac{0.l}{N}}, \quad e^{-j2\pi\frac{1.l}{N}}, \quad \ldots, \quad e^{-j2\pi\frac{(N-1).l}{N}} \ \right). \tag{3.11}$$

$\mathbf{U}_h^H\mathbf{M}^H\mathbf{MU}_h$ is a diagonal matrix if and only if the columns of matrix $\mathbf{MU}_h$ are orthogonal. This means that the following two conditions must be satisfied

$$\mathbf{A}_l^H\mathbf{A}_l = (\mathbf{U}_l^H\mathbf{S}^H\mathbf{SU}_l) \otimes \mathbf{I}_{M_r} \text{ must be diagonal}$$

or, equivalently,

$$\mathbf{U}_l^H\mathbf{S}^H\mathbf{SU}_l \text{ must be diagonal} \tag{3.12}$$

For $l' \neq l$

$$\mathbf{A}_{l'}^H\mathbf{A}_l = (\mathbf{U}_{l'}^H\mathbf{S}^H\mathbf{D}_{l'}^H\mathbf{D}_l\,\mathbf{S}\,\mathbf{U}_l) \otimes (\mathbf{V}_{l'}^H\mathbf{V}_l) = \mathbf{0}_{M_rM_t}$$

which is equivalent to

$$\mathbf{S}^H\mathbf{D}_{l-l'}\mathbf{S} = \mathbf{0}_{M_t}. \tag{3.13}$$

In general, (3.12) and (3.13) are a set of nonlinear equations where the number of the equations excesses the number of the unknowns. Therefore, the existence of a training matrix $\mathbf{S}$ satisfying the diagonality of $\mathbf{U}_h^H\mathbf{M}^H\mathbf{MU}_h$ cannot be guaranteed in general.

3.2 MSE Lower Bound of MIMO-OFDM Channel Estimation

In this section, we will find a theoretical lower bound on the MSE of channel estimation under the assumption of diagonality of the matrix $\mathbf{U}_h^H\mathbf{M}^H\mathbf{MU}_h$.

Under this assumption, condition (3.12) is equivalent to

$$(\mathbf{SU}_l)^H(\mathbf{SU}_l) = \mathrm{diag}\left(\alpha_0, \quad \alpha_1, \quad \ldots, \quad \alpha_{M_t-1} \right) \tag{3.14}$$

where $\alpha_i \geq 0,\ i = 0, \ldots, M_t - 1$ and $\sum_{i=0}^{M_t-1} \alpha_i = NM_t$ to be the total power constraint of the training sequence.

Therefore, a lower bound on the MSE in (3.5) can be written as

$$\begin{aligned}
\underline{\mathrm{MSE}} &= \sum_{l=0}^{L-1} \mathrm{trace}\left\{ (\mathbf{\Lambda}_{tl}^{-1} \otimes \mathbf{\Lambda}_{rl}^{-1} + \tfrac{\rho}{M_t}\mathbf{A}_l^H \mathbf{A}_l)^{-1} \right\} \\
&= \sum_{l=0}^{L-1} \mathrm{trace}\left\{ [\mathbf{\Lambda}_{tl}^{-1} \otimes \mathbf{\Lambda}_{rl}^{-1} + \tfrac{\rho}{M_t}\mathrm{diag}\left(\alpha_0, \quad \alpha_1, \quad \ldots, \quad \alpha_{M_t-1} \right) \otimes \mathbf{I}_{M_r}]^{-1} \right\} \\
&= \sum_{l=0}^{L-1} \sum_{i=0}^{M_t-1} \sum_{j=0}^{M_r-1} \left(\frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})} + \frac{\rho}{M_t}\alpha_i \right)^{-1}
\end{aligned} \tag{3.15}$$

where $\lambda_i(\mathbf{X})$ denotes the i^{th} eigenvalue of matrix $\mathbf{X}$.

Define

$$\lambda_{ik} = \frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})}, \quad \begin{array}{c} i = 0, \ldots, M_t - 1 \\ k = 0, \ldots, LM_r - 1 \end{array}. \tag{3.16}$$

We have

$$\underline{\mathrm{MSE}} = \sum_{i=0}^{M_t-1} \sum_{k=0}^{LM_r-1} \frac{1}{a\alpha_i + \lambda_{ik}} \tag{3.17}$$

where $a = \frac{\rho}{M_t}$.

This lower bound on the MSE is attainable only when both conditions (3.12) and (3.13) are satisfied.

3.3 Limit of the MSE when Channel Correlation Matrix Approaches Singularity

Up to now, we have only considered the case when the channel correlation matrix $\mathbf{R}_h$ is invertible. Accordingly, the MSE expression in (2.44) can be written as (2.45) using the matrix inversion lemma. However, it is the fact that in very highly correlated fading environments, $\mathbf{R}_h$ approaches singularity and the matrix inversion lemma cannot be applied in (2.44). Therefore, all the previous results do not hold. To resolve this problem, we will consider the limit of channel estimation MSE when $\mathbf{R}_h$ is singular.

Let

$$\hat{\mathbf{R}}_h = \mathbf{R}_h + \varepsilon \mathbf{I} \tag{3.18}$$

where ε is a positive scalar.

We observe that $\hat{\mathbf{R}}_h$ is non-singular, and

$$\mathbf{R}_h = \lim_{\varepsilon \to 0} \hat{\mathbf{R}}_h.$$

From (2.46), we have

$$\text{MSE} = \lim_{\varepsilon \to 0} \text{trace}\left\{ (\hat{\mathbf{R}}_h^{-1} + \tfrac{\rho}{M_t}\mathbf{M}^H\mathbf{M})^{-1} \right\}.$$

Similar to (3.4), an eigen-decomposition of $\hat{\mathbf{R}}_h$ is written as

$$\hat{\mathbf{R}}_h = \mathbf{U}_h(\mathbf{\Lambda}_h + \varepsilon \mathbf{I})\mathbf{U}_h^H. \tag{3.19}$$

Hence, equivalent to (3.5), we have

$$\text{MSE} = \lim_{\varepsilon \to 0} \text{trace}\left\{ ((\mathbf{\Lambda}_h + \varepsilon \mathbf{I})^{-1} + \tfrac{\rho}{M_t}\mathbf{U}_h^H\mathbf{M}^H\mathbf{M}\mathbf{U}_h)^{-1} \right\}. \tag{3.20}$$

Based on the derivation as in Section 3.1 and Section 3.2, the lower bound of MSE when $\mathbf{U}_h^H\mathbf{M}^H\mathbf{M}\mathbf{U}_h$ is a diagonal matrix is written as

$$\begin{aligned}
\underline{\text{MSE}} &= \lim_{\varepsilon \to 0} \sum_{l=0}^{L-1} \text{trace}\left\{ \left((\mathbf{\Lambda}_{tl} \otimes \mathbf{\Lambda}_{rl} + \varepsilon \mathbf{I}_{M_t M_r})^{-1} + \tfrac{\rho}{M_t}\mathbf{A}_l^H\mathbf{A}_l \right)^{-1} \right\} \\
&= \lim_{\varepsilon \to 0} \sum_{l=0}^{L-1} \text{trace}\left\{ \left(\begin{array}{c} (\mathbf{\Lambda}_{tl} \otimes \mathbf{\Lambda}_{rl} + \varepsilon \mathbf{I}_{M_t M_r})^{-1} + \\ \tfrac{\rho}{M_t}\text{diag}\left(\alpha_0, \ \alpha_1, \ \ldots, \ \alpha_{M_t-1} \right) \otimes \mathbf{I}_{M_r} \end{array} \right)^{-1} \right\} \\
&= \lim_{\varepsilon \to 0} \sum_{l=0}^{L-1} \sum_{i=0}^{M_t-1} \sum_{j=0}^{M_r-1} \left(\frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl}) + \varepsilon} + \tfrac{\rho}{M_t}\alpha_i \right)^{-1}.
\end{aligned} \tag{3.21}$$

Let K_i be the number of non-zero eigenvalue products $\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})$ for different l and j.

Define

$$\lambda_{ik} = \frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})}, \quad \begin{array}{l} i = 0, \ldots, M_t - 1 \\ k = 0, \ldots, K_i - 1 \end{array} \tag{3.22}$$

where $\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl}) > 0$.

Thus, limit of MSE when $\varepsilon \to 0$ can be written as

$$\underline{\mathrm{MSE}} = \sum_{i=0}^{M_t-1}\sum_{k=0}^{K_i-1}\frac{1}{a\alpha_i+\lambda_{ik}} \tag{3.23}$$

where $a = \frac{\rho}{M_t}$

When $\mathbf{R}_h$ is invertible, all $\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})$ are non-zero and K_i is equal to LM_r.

We now have to find the optimal transmit power allocation $\alpha_{i,\ i=0,\dots,M_t-1}$ in order to minimize the MSE lower bound in (3.23). This is the general form of channel estimation MSE lower bound regardless of the invertibility of the channel correlation matrix $\mathbf{R}_h$.

3.4 Finding Optimal Transmit Power Allocation by the IBP

The optimization problem (3.1) is converted to the following optimization problem

$$\begin{array}{cc} \min\limits_{\alpha_{i,\ i=0,1,\dots,M_t-1}} & \underline{\mathrm{MSE}} = \sum_{i=0}^{M_t-1}\sum_{k=0}^{K_i-1}\frac{1}{a\alpha_i+\lambda_{ik}} \\ \text{subject to:} & \sum_{i=0}^{M_t-1}\alpha_i = NM_t \\ & \alpha_i \geq 0,\ i=0,\dots,M_t-1. \end{array} \tag{3.24}$$

From the Gradient and Hessian of the objective function $\underline{\mathrm{MSE}}(\alpha_{i,\ i=0,1,\dots,M_t-1})$ in the optimization problem (3.24) we can easily see that it is both convex and decreasing with all $\alpha_i,\ i=0,\dots,M_t-1$.

The Lagragian associated with the optimization problem (3.24) is [11]

$$L(\alpha_i,\mu,\mu_i,\ i=0,\dots,M_t-1) = \underline{\mathrm{MSE}} - \sum_{i=0}^{M_t-1}(\alpha_i\mu_i) + \mu\left(\sum_{i=0}^{M_t-1}\alpha_i - NM_t\right) \tag{3.25}$$

(for Lagrange multipliers $\mu \geq 0$, $\mu_i \geq 0,\ i=0,\dots,M_t-1$).

The optimal value of $\alpha_{i,\ i=0,\dots,M_t-1}$, μ and $\mu_{i\ i=0,\dots,M_t-1}$ have to satisfy the KKT conditions [11]

$$\begin{cases} \sum_{i=0}^{M_t-1} \alpha_i = NM_t \\ \alpha_i \geq 0,\ i = 0,\ldots,M_t-1 \\ \mu_i \geq 0,\ i = 0,\ldots,M_t-1 \\ \mu_i \alpha_i = 0,\ i = 0,\ldots,M_t-1 \\ \frac{\partial L}{\partial \alpha_i} = \sum_{k=0}^{K_i-1} \frac{-a}{(a\alpha_i+\lambda_{ik})^2} - \mu_i + \mu = 0,\ i = 0,\ldots,M_t-1 \end{cases} \tag{3.26}$$

Note that μ_i acts as a slack variable in the last equation, so it can be eliminated.

The KKT conditions, therefore, can be re-written as

$$\begin{cases} \sum_{i=0}^{M_t-1} \alpha_i = NM_t \\ \alpha_i \geq 0,\ i = 0,\ldots,M_t-1 \\ \alpha_i(\mu - \sum_{k=0}^{K_i-1} \frac{a}{(a\alpha_i+\lambda_{ik})^2}) = 0,\ i = 0,\ldots,M_t-1 \\ \sum_{k=0}^{K_i-1} \frac{a}{(a\alpha_i+\lambda_{ik})^2} \leq \mu,\ i = 0,\ldots,M_t-1 \end{cases} \tag{3.27}$$

It can be seen that this optimization problem belongs to the class of water-filling optimization problems. The optimal solution of (3.24), thus, can be expressed as

$$\alpha_{i\,opt} = \alpha_i^+(\mu) := \max(0, \alpha_i^*) \quad , i = 0,\ldots,M_t-1 \tag{3.28}$$

where α_i^* is the solution of the following equation

$$f(\alpha_i) = \sum_{k=0}^{K_i-1} \frac{a}{(a\alpha_i+\lambda_{ik})^2} = \mu \tag{3.29}$$

and the value of $\alpha_i^+(\mu),\ i = 0,\ldots,M_t-1$ as functions of the scalar Lagrange multiplier μ must satisfy the total transmit power constraint: $\sum_{i=0}^{M_t-1} \alpha_i^+(\mu) = NM_t$.

Function $f(\alpha_i)$ in equation (3.29) is a decreasing function over the range $\alpha_i \in [0, NM_t]$. Therefore, if μ is known, equation (3.29) can be solved by the IBP [33] as below.

Iterative Bisection Procedure (IBP): Solve an equation: $f(x) = \alpha, x \in [\underline{x}, \overline{x}]$ where $f(x)$ is a decreasing function.

- If $f(\underline{x}) < \alpha$ or $f(\overline{x}) > \alpha$, then there is no solution for $x \in [\underline{x}, \overline{x}]$

- Otherwise, set $x = \frac{\underline{x}+\overline{x}}{2}$. If $f(x) > \alpha$, reset $\underline{x} = x$
  If $f(x) < \alpha$, reset $\overline{x} = x$

- Repeat until $|f(x) - \alpha| < \varepsilon$ (where ε is the tolerance of the solution deviation).

Observe that the objective function $\underline{\text{MSE}}$ in (3.24) has the following lower bound and upper bound

$$\sum_{i=0}^{M_t-1} \frac{K_i}{a\alpha_i + \lambda_{i\,max}} \leq \underline{\text{MSE}} \leq \sum_{i=0}^{M_t-1} \frac{K_i}{a\alpha_i + \lambda_{i\,min}} \tag{3.30}$$

where $\lambda_{i\,max} = \max_k \lambda_{ik}$ and $\lambda_{i\,min} = \min_k \lambda_{ik}$.

Therefore, we can find the value of μ in the range $[\underline{\mu}, \overline{\mu}]$, where $\underline{\mu}$ and $\overline{\mu}$ are respectively the Lagrange multipliers corresponding to the equality constraint in the following two simple optimization problems.

$$\begin{aligned} \min_{\alpha_{i,\ i=0,1,\ldots,M_t-1}} \quad & \textstyle\sum_{i=0}^{M_t-1} \frac{K_i}{a\alpha_i+\lambda_{i\,max}} \\ \text{subject to:} \quad & \textstyle\sum_{i=0}^{M_t-1} \alpha_i = NM_t \\ & \alpha_i \geq 0,\ i = 0, \ldots, M_t - 1. \end{aligned} \tag{3.31}$$

and

$$\begin{aligned} \min_{\alpha_{i,\ i=0,1,\ldots,M_t-1}} \quad & \textstyle\sum_{i=0}^{M_t-1} \frac{K_i}{a\alpha_i+\lambda_{i\,min}} \\ \text{subject to:} \quad & \textstyle\sum_{i=0}^{M_t-1} \alpha_i = NM_t \\ & \alpha_i \geq 0,\ i = 0, \ldots, M_t - 1. \end{aligned} \tag{3.32}$$

These are two simple water-filling optimization problems and can be easily solved.

Figure 3.1 shows the flow diagram to find the optimal value of $\alpha_{i,\ i=0,\ldots,M_t-1}$ where ε_1 and ε_2 are respectively the tolerance of the difference between $\underline{\mu}$ and $\overline{\mu}$, and the tolerance of the difference to the total allocated power constraint.

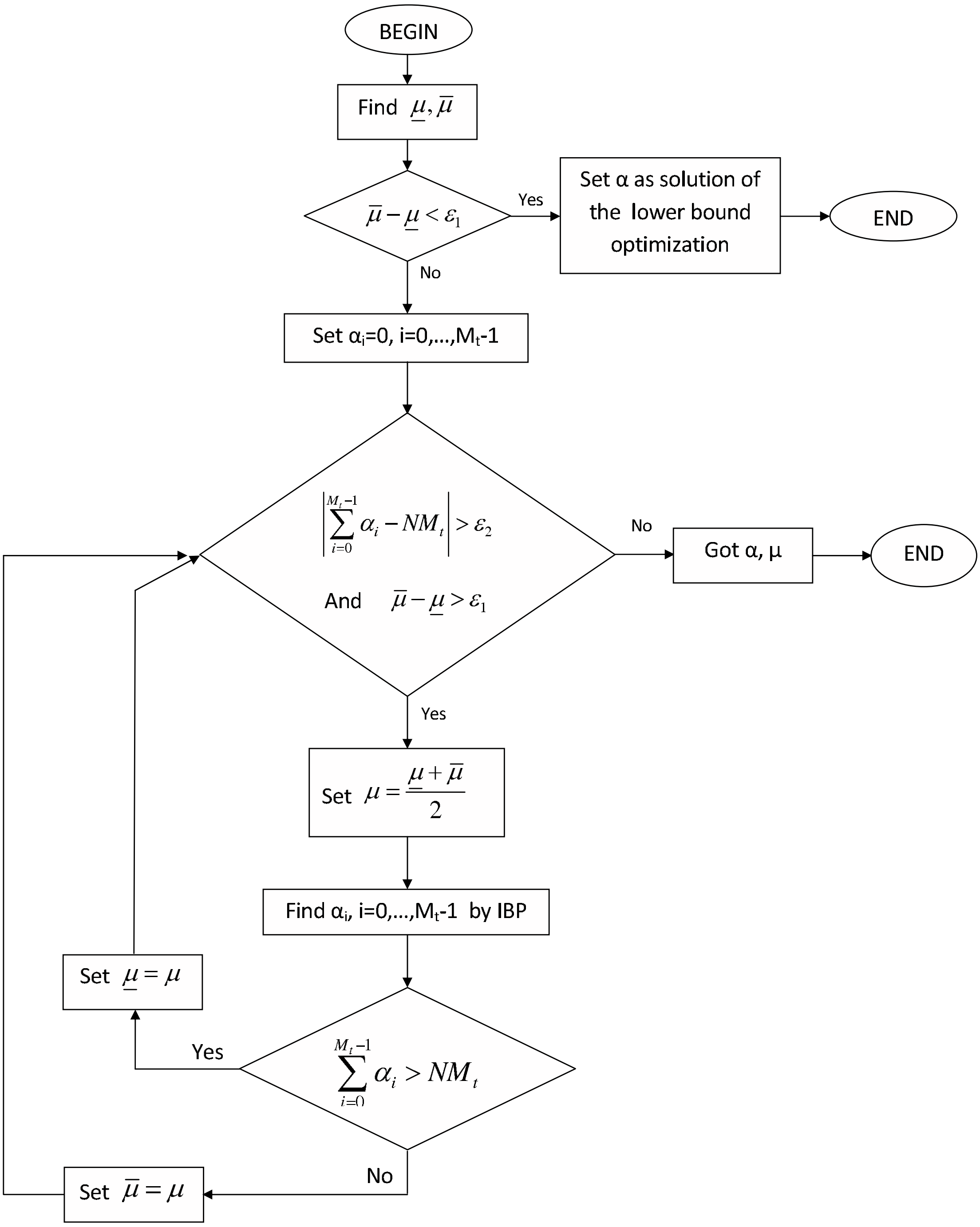


Figure 3.1: Flow Diagram for Optimal Power Allocation Algorithm

3.5 Approximate Optimal Transmit Power Allocation Algorithm

The optimal solution of the optimization problem (3.24) is guaranteed with the IBP described in the previous section. However, the computational complexity of such algorithm is quite high. It involves two water-filling problems and a computationally intensive loop of IBP. This gives us a motivation to find a more efficient algorithm.

From (3.15), we define a function

$$f_i(\alpha_i) = \sum_{j=0}^{M_r-1}\sum_{l=0}^{L-1}\left(\frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})} + \frac{\rho}{M_t}\alpha_i \right)^{-1}. \tag{3.33}$$

Since $(x^{-1}+b)^{-1}$ is a concave function of scalar variable $x>0$, we can find an upper bound for $f_i(\alpha_i)$ in (3.33) as follows

$$f_i(\alpha_i) \le LM_r(\gamma_i^{-1} + \frac{\rho}{M_t}\alpha_i)^{-1} \tag{3.34}$$

where

$$\gamma_i = \frac{1}{LM_r}\sum_{j=0}^{M_r-1}\sum_{l=0}^{L-1}\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl}). \tag{3.35}$$

We now consider the following optimization problem

$$\begin{aligned} \min_{\alpha_i,\ i=0,1,\ldots,M_t-1} \quad & \textstyle\sum_{i=0}^{M_t-1} \frac{LM_r}{a\alpha_i+\gamma_i^{-1}} \\ \text{subject to:} \quad & \textstyle\sum_{i=0}^{M_t-1}\alpha_i = NM_t \\ & \alpha_i \ge 0,\ i=0,\ldots,M_t-1. \end{aligned} \tag{3.36}$$

which is the minimization of an upper bound of the objective function in (3.24). This is only a simple water-filling optimization problem. Therefore, the iterative solution can be easily implemented.

We will see in the simulation results that the MSE performance of this approximate solution of transmit power allocation is almost the same as that of the optimal solution in the previous part.

3.6 Nearly-Optimal Training Sequence Design

As mentioned in Section 3.1, there is no guarantee for the existence of an optimal training sequence. However, a class of nearly-optimal training sequences can be designed only according to the correlation information of the strongest delay tap. The correlation matrix of this tap is denoted as **R** and its eigen-matrix corresponding to the eigen-decomposition is **U**.

From condition (3.12), we have

$$(\mathbf{SU})^H(\mathbf{SU}) = \mathrm{diag}\begin{pmatrix} \alpha_0, & \alpha_1, & \ldots, & \alpha_{M_t-1} \end{pmatrix}. \tag{3.37}$$

Thus, a nearly-optimal training symbol matrix **S** can be designed as follows

$$\begin{aligned} \mathbf{SU} &= (\sqrt{\alpha_0}\mathbf{q}_0, \ldots, \sqrt{\alpha_{M_t-1}}\mathbf{q}_{M_t-1}) \\ \Leftrightarrow \mathbf{S} &= (\sqrt{\alpha_0}\mathbf{q}_0, \ldots, \sqrt{\alpha_{M_t-1}}\mathbf{q}_{M_t-1})\mathbf{U}^H. \end{aligned} \tag{3.38}$$

where

The values of $\alpha_i,\ i = 0, \ldots, M_t - 1$ are chosen using the IBP or the approximate algorithm as in the previous sections,

$\mathbf{q}_0, \ldots, \mathbf{q}_{M_t-1}$ are a set of N-element orthonormal vectors.

To satisfy condition (3.13), we can choose $\mathbf{q}_i$ according to [27] as follows

Let $\mathbf{q}_0 = \begin{pmatrix} q_0(0) & \ldots & q_0(N-1) \end{pmatrix}^T$ be any training sequence that is good for timing and frequency synchronization and $|q_0(k)| = \frac{1}{\sqrt{N}}$, for $k = 0, 1, \ldots, N-1$

and $\mathbf{q}_i = \begin{pmatrix} q_0(0)W_N^{-K_0 i 0} & \ldots & q_0(N-1)W_N^{-K_0 i(N-1)} \end{pmatrix}^T$, $i = 0, \ldots, M_t - 1$ where $W_N = e^{-j\frac{2\pi}{N}}$ and $K_0 = \lfloor \frac{N}{M_t} \rfloor$ to be the largest integer smaller than or equal to $\frac{N}{M_t}$.

3.7 Extension to MIMO-ZP-OFDM-OLA Optimal Training Sequence Design

As ZP-OFDM is considered as an alternative to CP-OFDM in practical wireless systems, it is worthwhile to have further discussion on the topic of optimal training sequence design for MIMO-ZP-OFDM. Due to the similar implementation of the two OFDM systems, the optimal training sequence design for MIMO-ZP-OFDM will follow the same manner as having been discussed throughout this chapter.

3.7.1 Conditions for MIMO-ZP-OFDM-OLA Optimal Training Sequences

The objective of optimal training sequence design for MIMO-ZP-OFDM-OLA in spatially correlated fading environments is to find training sequences to minimize the MSE of channel estimation under the transmit power constraint. It is equivalent to the following mathematical optimization problem

$$\begin{array}{ll} \min\limits_{\mathbf{S}\in\mathbf{C}^{N\times M_t}} & \text{trace}\left\{ (\mathbf{R}_h^{-1}+\frac{\rho}{M_t}\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M})^{-1} \right\} \\ \text{subject to} & \text{trace}\{\mathbf{S}^H\mathbf{S}\}=(N+L)M_t. \end{array} \tag{3.39}$$

Using the same mathematical manipulations as in Section 3.1, the optimization problem (3.39) can be re-written as

$$\begin{array}{ll} \min\limits_{\mathbf{S}\in\mathbf{C}^{N\times M_t}} & \text{trace}\left\{ (\mathbf{\Lambda}_h^{-1}+\frac{\rho}{M_t}\mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h)^{-1} \right\} \\ \text{subject to} & \text{trace}\{\mathbf{S}^H\mathbf{S}\}=(N+L)M_t. \end{array} \tag{3.40}$$

The objective function in (3.40) will be minimized if $\mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h$ is a diagonal matrix [24] [46] [41].

To simplify the above condition, we can partition matrix $\mathbf{M}\mathbf{U}_h$ as follows

$$\mathbf{M}\mathbf{U}_h=\begin{pmatrix} \mathbf{A}_0(0) & \dots & \mathbf{A}_{L-1}(0) \\ \vdots & \ddots & \vdots \\ \mathbf{A}_0(N-1) & \dots & \mathbf{A}_{L-1}(N-1) \end{pmatrix}=\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} \tag{3.41}$$

where

$$\mathbf{A}_l(k)=\mathbf{M}_l(k)\,(\mathbf{U}_l\otimes\mathbf{V}_l)=e^{-j2\pi\frac{kl}{N}}(\mathbf{s}^T(k)\,\mathbf{U}_l)\otimes\mathbf{V}_l,$$

$$\mathbf{B}_1\in\mathbf{C}^{LM_r\times LM_tM_r} \quad \text{and} \quad \mathbf{B}_2\in\mathbf{C}^{(N-L)M_r\times LM_tM_r}.$$

Accordingly, we have

$$\begin{aligned} \mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h &= \begin{pmatrix} \mathbf{B}_1^H & \mathbf{B}_2^H \end{pmatrix}\begin{pmatrix} \frac{1}{2}\mathbf{I}_{LM_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(N-L)M_r} \end{pmatrix}\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} \\ &= \frac{1}{2}\mathbf{B}_1^H\mathbf{B}_1+\mathbf{B}_2^H\mathbf{B}_2. \end{aligned} \tag{3.42}$$

The training symbol matrix $\mathbf{S}$ can be partitioned as follows

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{s}(0) & \mathbf{s}(1) & \dots & \mathbf{s}(N-1) \end{pmatrix}^T \in \mathbf{C}^{N \times M_t} \tag{3.43}$$

where $\mathbf{S}_1 \in \mathbf{C}^{L \times M_t}$ and $\mathbf{S}_2 \in \mathbf{C}^{(N-L) \times M_t}$.

We define

$$\mathbf{D}_l = \begin{pmatrix} \mathbf{D}_{1l} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{2l} \end{pmatrix} = \mathrm{diag}\begin{pmatrix} e^{-j2\pi\frac{0.l}{N}}, & e^{-j2\pi\frac{1.l}{N}}, & \dots, & e^{-j2\pi\frac{(N-1).l}{N}} \end{pmatrix} \in \mathbf{C}^{N \times N} \tag{3.44}$$

where $\mathbf{D}_{1l} \in \mathbf{C}^{L \times L}$ and $\mathbf{D}_{2l} \in \mathbf{C}^{(N-L) \times (N-L)}$.

The l^{th} block column of $\mathbf{B}_1$ and $\mathbf{B}_2$ can be respectively represented as follows

$$\begin{aligned} (\mathbf{B}_1)_l &= (\mathbf{D}_{1l}\mathbf{S}_1\mathbf{U}_l) \otimes \mathbf{V}_l, \\ (\mathbf{B}_2)_l &= (\mathbf{D}_{2l}\mathbf{S}_2\mathbf{U}_l) \otimes \mathbf{V}_l. \end{aligned}$$

Accordingly, we have

$$\begin{aligned} (\mathbf{B}_1)_l^H(\mathbf{B}_1)_l &= (\mathbf{U}_l^H\mathbf{S}_1^H\mathbf{S}_1\mathbf{U}_l) \otimes \mathbf{I}_{M_r}, \\ (\mathbf{B}_2)_l^H(\mathbf{B}_2)_l &= (\mathbf{U}_l^H\mathbf{S}_2^H\mathbf{S}_2\mathbf{U}_l) \otimes \mathbf{I}_{M_r}. \end{aligned}$$

The diagonal structure of $\frac{1}{2}\mathbf{B}_1^H\mathbf{B}_1 + \mathbf{B}_2^H\mathbf{B}_2$ is equivalent to

$$\mathbf{U}_l^H(\frac{1}{2}\mathbf{S}_1^H\mathbf{S}_1 + \mathbf{S}_2^H\mathbf{S}_2)\mathbf{U}_l \qquad \text{must be diagonal} \tag{3.45}$$

$$\frac{1}{2}\mathbf{S}_1^H\mathbf{D}_{1\,l-l'}\mathbf{S}_1 + \mathbf{S}_2^H\mathbf{D}_{2\,l-l'}\mathbf{S}_2 = \mathbf{0}_{M_t} \tag{3.46}$$

We define a modified training symbol matrix as follows

$$\tilde{\mathbf{S}} = \begin{pmatrix} \frac{1}{\sqrt{2}}\mathbf{S}_1 \\ \mathbf{S}_2 \end{pmatrix}. \tag{3.47}$$

Conditions (3.45) and (3.46) are respectively equivalent to

$$\mathbf{U}_l^H\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\mathbf{U}_l \text{ must be diagonal} \tag{3.48}$$

and for $l' \neq l$

$$\tilde{\mathbf{S}}^H\mathbf{D}_{l-l'}\tilde{\mathbf{S}} = \mathbf{0}_{M_t} \tag{3.49}$$

(3.48) and (3.49) are the necessary and sufficient conditions for the diagonality of $\mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h$. We observe that the conditions (3.48) – (3.49) are exactly the same as conditions (3.12) – (3.13) in the MIMO-CP-OFDM case which was presented in Section 3.1.

3.7.2 Lower Bound of the MSE of MIMO-ZP-OFDM-OLA Channel Estimation

In this section, we will find a theoretical lower bound of the MSE of channel estimation when assuming the diagonality of the matrix $\mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h$. Under this assumption, condition (3.48) is equivalent to

$$(\tilde{\mathbf{S}}\mathbf{U}_l)^H(\tilde{\mathbf{S}}\mathbf{U}_l) = \mathrm{diag}\left(\begin{array}{cccc}\alpha_0, & \alpha_1, & \ldots, & \alpha_{M_t-1}\end{array}\right) \tag{3.50}$$

where $\alpha_i \geq 0,\ i = 0,\ldots,M_t-1$ and

$$\begin{aligned}\sum_{i=0}^{M_t-1}\alpha_i &= \mathrm{trace}\{\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\}\\ &= \frac{N}{N+L}\mathrm{trace}\{\mathbf{S}^H\mathbf{S}\}\\ &= NM_t.\end{aligned} \tag{3.51}$$

Therefore, a lower bound of the objective function in (3.40) can be written as

$$\begin{aligned}\underline{\mathrm{MSE}} &= \mathrm{trace}\left\{\ (\mathbf{\Lambda}_h^{-1}+\tfrac{\rho}{M_t}\mathbf{U}_h^H\mathbf{M}^H\mathbf{R}_{\tilde{n}}^{-1}\mathbf{M}\mathbf{U}_h)^{-1}\ \right\}\\ &= \mathrm{trace}\left\{\ \left[\ \mathbf{\Lambda}_h^{-1}+\tfrac{\rho}{M_t}(\tfrac{1}{2}\mathbf{B}_1^H\mathbf{B}_1+\mathbf{B}_2^H\mathbf{B}_2)\ \right]^{-1}\ \right\}\\ &= \sum_{l=0}^{L-1}\mathrm{trace}\left\{\ \left[\ \mathbf{\Lambda}_{tl}^{-1}\otimes\mathbf{\Lambda}_{rl}^{-1}+\tfrac{\rho}{M_t}\mathbf{U}_l^H(\tfrac{1}{2}\mathbf{S}_1^H\mathbf{S}_1+\mathbf{S}_2^H\mathbf{S}_2)\mathbf{U}_l\otimes\mathbf{I}_{M_r}\ \right]^{-1}\ \right\}\\ &= \sum_{l=0}^{L-1}\mathrm{trace}\left\{\ \left(\ \mathbf{\Lambda}_{tl}^{-1}\otimes\mathbf{\Lambda}_{rl}^{-1}+\tfrac{\rho}{M_t}\mathbf{U}_l^H\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\mathbf{U}_l\otimes\mathbf{I}_{M_r}\ \right)^{-1}\ \right\}\\ &= \sum_{l=0}^{L-1}\mathrm{trace}\left\{\ \left[\ \mathbf{\Lambda}_{tl}^{-1}\otimes\mathbf{\Lambda}_{rl}^{-1}+\tfrac{\rho}{M_t}\mathrm{diag}\left(\begin{array}{cccc}\alpha_0, & \alpha_1, & \ldots, & \alpha_{M_t-1}\end{array}\right)\otimes\mathbf{I}_{M_r}\ \right]^{-1}\ \right\}\\ &= \sum_{l=0}^{L-1}\sum_{i=0}^{M_t-1}\sum_{j=0}^{M_r-1}\left(\ \frac{1}{\lambda_i(\mathbf{R}_{tl})\lambda_j(\mathbf{R}_{rl})}+\tfrac{\rho}{M_t}\alpha_i\ \right)^{-1}.\end{aligned}$$

The optimization problem (3.39), thus, can be converted to

$$\begin{array}{cl} \min\limits_{\alpha_i,\ i=0,1,\ldots,M_t-1} & \sum_{i=0}^{M_t-1}\sum_{k=0}^{K_i-1}\frac{1}{a\alpha_i+\lambda_{ik}} \\ \text{subject to:} & \sum_{i=0}^{M_t-1}\alpha_i = NM_t \\ & \alpha_i \geq 0,\ i=0,\ldots,M_t-1 \end{array} \tag{3.52}$$

where the parameters K_i, λ_{ik} and a are defined the same as in Section 3.3.

This optimization problem is exactly the same as the optimization problem in (3.24) for MIMO-CP-OFDM case. Therefore, the optimal solution and the approximate solution can be found by the same algorithms as in Section 3.4 and 3.5 respectively.

3.7.3 Nearly-Optimal Training Sequence Design

Similar to the training sequence design specified in Section 3.6, a class of nearly-optimal training sequences can be designed only according to the correlation information of the strongest delay tap. The correlation matrix of this tap is denoted as $\mathbf{R}$ and its eigen-matrix corresponding to the eigen-decomposition is $\mathbf{U}$.

From condition (3.48), we have

$$(\tilde{\mathbf{S}}\mathbf{U})^H(\tilde{\mathbf{S}}\mathbf{U}) = \mathrm{diag}\begin{pmatrix} \alpha_0, & \alpha_1, & \ldots, & \alpha_{M_t-1} \end{pmatrix}. \tag{3.53}$$

Thus, we must have

$$\begin{aligned} \tilde{\mathbf{S}}\mathbf{U} &= (\sqrt{\alpha_0}\mathbf{q}_0,\ldots,\sqrt{\alpha_{M_t-1}}\mathbf{q}_{M_t-1}) \\ \Leftrightarrow \tilde{\mathbf{S}} &= (\sqrt{\alpha_0}\mathbf{q}_0,\ldots,\sqrt{\alpha_{M_t-1}}\mathbf{q}_{M_t-1})\mathbf{U}^H \end{aligned} \tag{3.54}$$

where

The values of $\alpha_i,\ i=0,\ldots,M_t-1$ are chosen using the IBP as in the previous section.

$\mathbf{q}_0,\ldots,\mathbf{q}_{M_t-1}$ are a set of N-element orthonormal vectors which are specified as in Section 3.6.

The modified training symbol matrix $\tilde{\mathbf{S}}$ in (3.54) has the same value as the optimal training symbol matrix $\mathbf{S}$ in (3.38). This means the optimal training symbol matrix $\mathbf{S}$ for MIMO-ZP-OFDM-OLA systems as defined in (3.43) has to increase power in the first

L training symbols. However, as the optimization problem (3.52) is exactly the same as the optimization problem in (3.24) for MIMO-CP-OFDM case, the MSE performance of the nearly optimal training sequence for MIMO-ZP-OFDM-OLA channel estimation is identical to that of MIMO-CP-OFDM case. Therefore, the simulation results presented in Section 3.8 are also valid for the MIMO-ZP-OFDM-OLA training sequence design.

3.8 Simulation Results

3.8.1 MSE Performance

In this section, we present the performance of the proposed training sequence in comparison with that of the existing ones and the MSE lower bound through simulation results.

The simulated MIMO-OFDM system has 4 transmit antennas ($M_t = 4$) and 4 receive antennas ($M_r = 4$). 32 OFDM subchannels ($N = 32$) are used and the cyclic prefix is long enough to avoid Inter Block Interference. Assuming that the channel has 5 delay taps ($L = 5$) with

power delay profile $\sigma^2 = \begin{pmatrix} 0.3 & 0.2 & 0.2 & 0.15 & 0.15 \end{pmatrix}$,

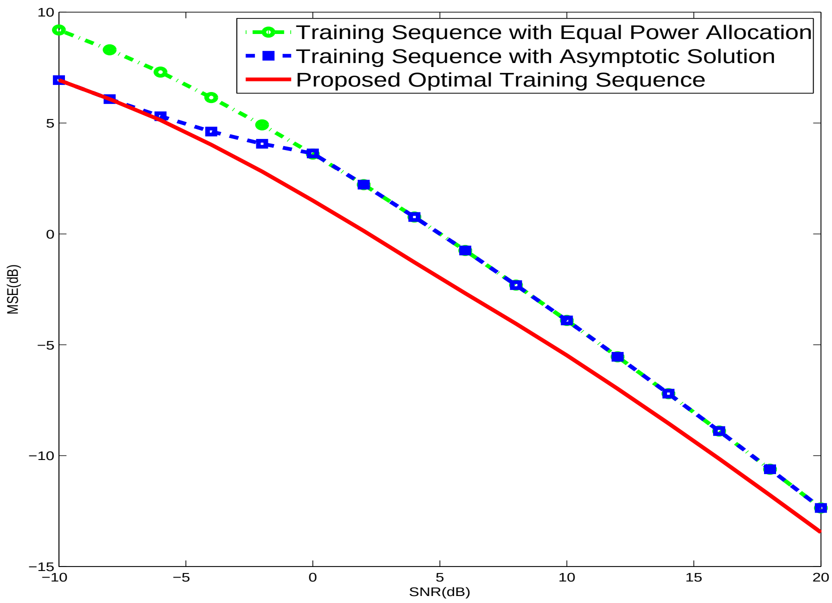
the mean angles of departure $\bar{\theta}_t = \begin{pmatrix} 13^o & 16^o & 20^o & 24^o & 27^o \end{pmatrix}$,

the mean angles of arrival $\bar{\theta}_r = \begin{pmatrix} 290^o & 300^o & 315^o & 320^o & 335^o \end{pmatrix}$.

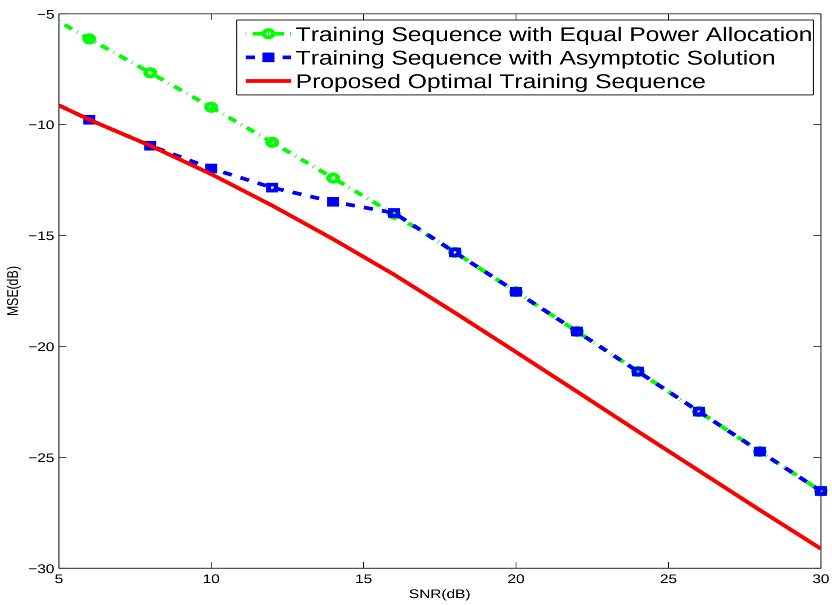
The angle spread for all taps at both transmitter and receiver are the same and equal to 8.6^o. The relative transmit and receive antenna spacing are both equal to 1.

With respect to the MSE of channel estimation, figure 3.2 demonstrates the performance of three different training sequences: training sequence with equal power allocation, training sequence with asymptotic solution as in [46] and the proposed training sequence with optimal transmit power allocation as in Section 3.4. It verifies that the training sequence with asymptotic solution as described in [46] is only optimal at very low and very high SNR. The training sequence which is proposed in Section 3.6 provides a better MSE over the whole range of SNR.

Figure 3.3 compares the MSE performance of the proposed training sequence with the MSE theoretical lower bound as specified in Equation (3.23). It can be seen that the gap

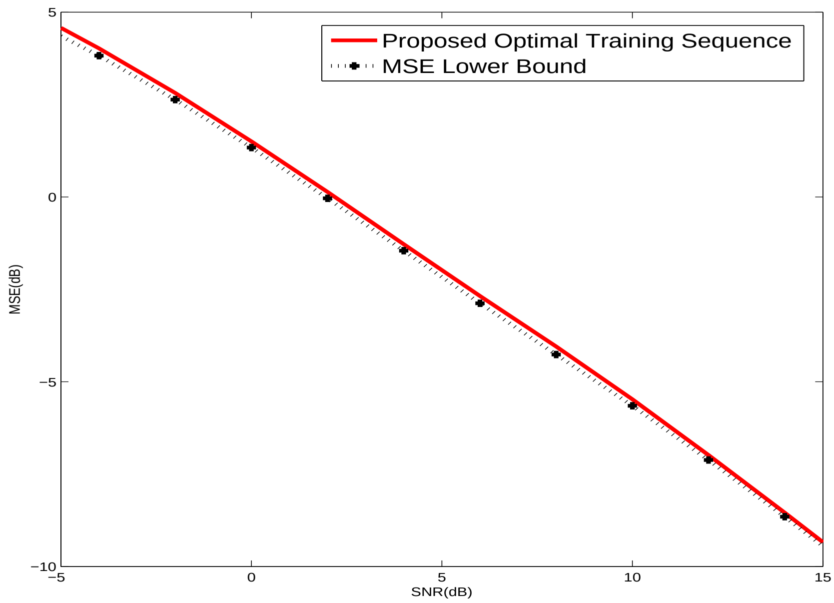


(a) $\sigma_\theta = 8.6^o$

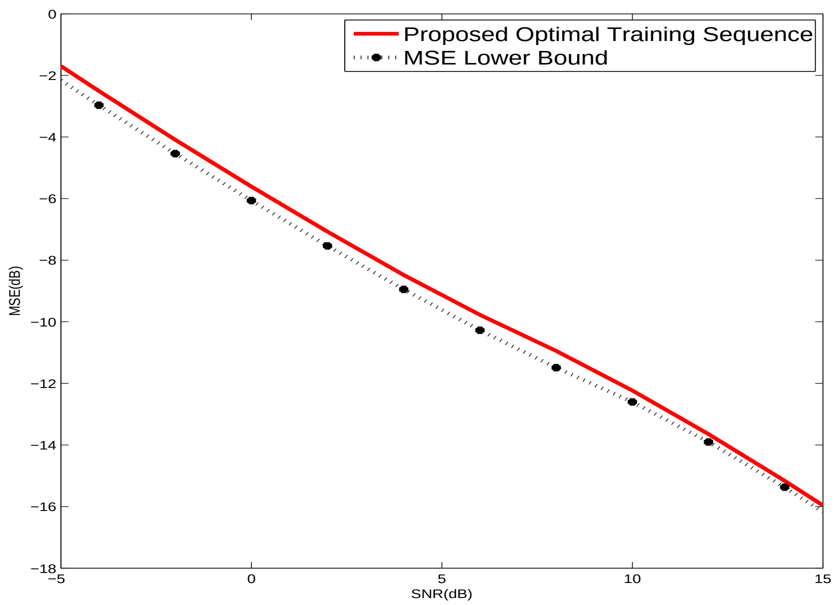


(b) $\sigma_\theta = 1^o$

Figure 3.2: The proposed training sequence vs. existing training sequences

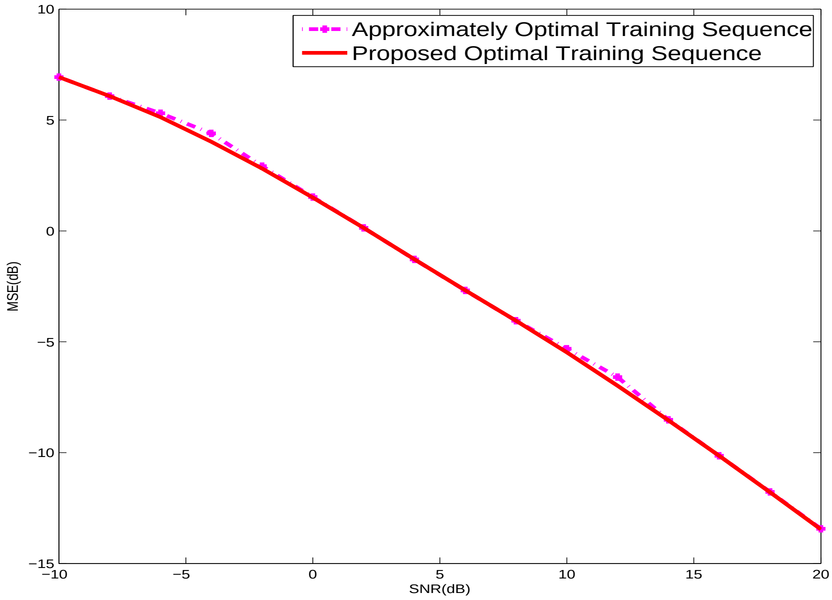


(a) $\sigma_\theta = 8.6^o$

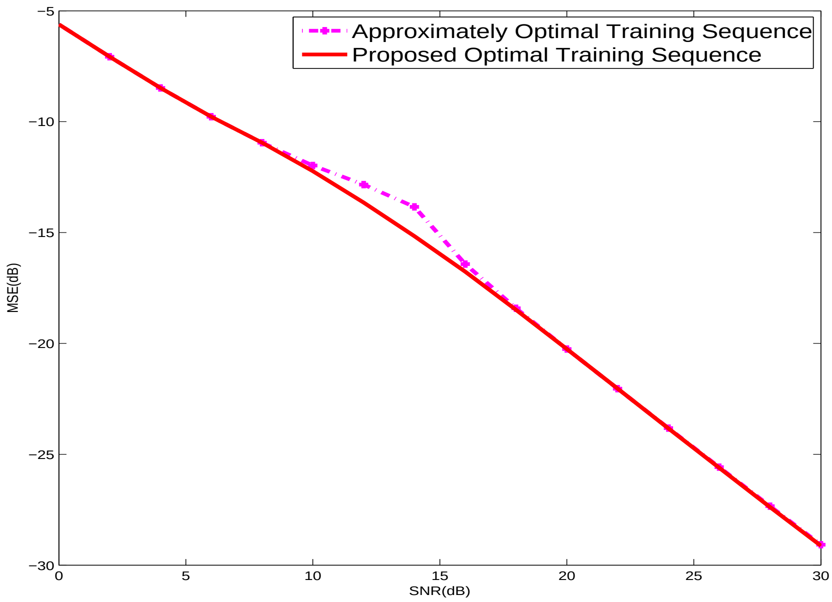


(b) $\sigma_\theta = 1^o$

Figure 3.3: MSE performance of the proposed training sequence vs. theoretical lower bound



(a) $\sigma_\theta = 8.6^o$



(b) $\sigma_\theta = 1^o$

Figure 3.4: The approximate training sequence vs. the proposed training sequence

between the two curves is very narrow. The MSE performance of a genuinely optimal training sequence will lie between these two curves. Therefore, we can conclude that the MSE performance of the nearly-optimal training sequence as in Section 3.6 is optimal.

It can be seen from Figure 3.4 that the MSE performance of the proposed training sequence using the approximate algorithm as in Section 3.5 is almost identical to that of the proposed training sequence with optimal transmit power allocation. Therefore, it is expected that the training sequence using the approximately optimal transmit power allocation algorithm can be considered as a strong candidate for practical MIMO-OFDM wireless systems due to its simple implementation and optimal MSE performance.

More significantly, Figure 3.2(b), 3.3(b) and 3.4(b) show that the proposed training sequence works well even in the case $\mathbf{R}_h$ is rank deficient due to very high spatially correlated fading ($\sigma_\theta = 1^o$).

3.8.2 SER Performance

In this section, we set up a simulated scenario of MIMO-OFDM communication system to examine how much the proposed training sequence improve the SER in comparison with other training sequences.

The MIMO-OFDM system configuration and the MIMO spatially correlated fading channel model are chosen exactly the same as those specified in Section 3.8.1. The angle spread for all taps at both transmitter and receiver are the same and equal to 8.6^o. The transmission system uses QPSK constellation without any channel coding scheme. A rate $3/4$ STBC scheme as specified in Appendix 3.10 is employed.

The SNR range to be examined is from 0dB to 12dB. Within this SNR range, the MSE performance of the training sequence with equal power allocation and that of the training sequence with asymptotic solution as in [46] are exactly the same. Therefore, the SER performances of MIMO-OFDM systems using these two training sequences are identical. Similarly, the SER performances of MIMO-OFDM systems using the proposed training sequence with optimal and approximate optimal power allocation are equal. To simplify the graph, figure 3.5 only demonstrates the SER performance of the MIMO-OFDM system with three different ways to obtain CSI: training sequence with equal power allocation, proposed optimal training sequence and perfect CSI at receiver.

It is observed that the SER performance when using the proposed optimal training sequence is very close to that when assuming perfect CSI at receiver. The proposed optimal training sequence provides 0.2dB power gain in SER performance in comparison with the training sequence using equal power allocation. This benefit is achieved at a slight complexity increase in training power allocation. It is important to note again that the approximate transmit power allocation algorithm as proposed in Section 3.5 is only a simple water-filling optimisation algorithm. Its iterative solution, therefore, can be easily implemented in practice.

3.9 Chapter Summary

In this chapter, we have comprehensively addressed the topic of optimal training sequence design for MIMO-OFDM in spatially correlated fading environments.

Section 3.1 reviewed the conditions for optimal training sequences which have been identified in [46]. At the end of this section, it was shown that the existence of a training matrix satisfying these conditions is not guaranteed. Consequently, we can only say that there is a lower bound for channel estimation MSE when assuming these conditions are satisfied. This lower bound on the channel estimation MSE is derived in Section 3.2.

Section 3.3 considered the limit of the MSE of channel estimation when the channel correlation matrix $\mathbf{R}_h$ approaches singularity due to very high correlated fading. A general expression of the MSE lower bound was given in (3.23) irrespective of the invertibility of the channel correlation matrix.

The optimal transmit power allocation which minimizes the MSE lower bound was presented in Section 3.4 and an approximate algorithm was proposed in Section 3.5.

A nearly optimal training sequence design using proposed power transmit allocation algorithms was presented in Section 3.6.

In Section 3.7, we also extended the optimal training sequence design to MIMO-

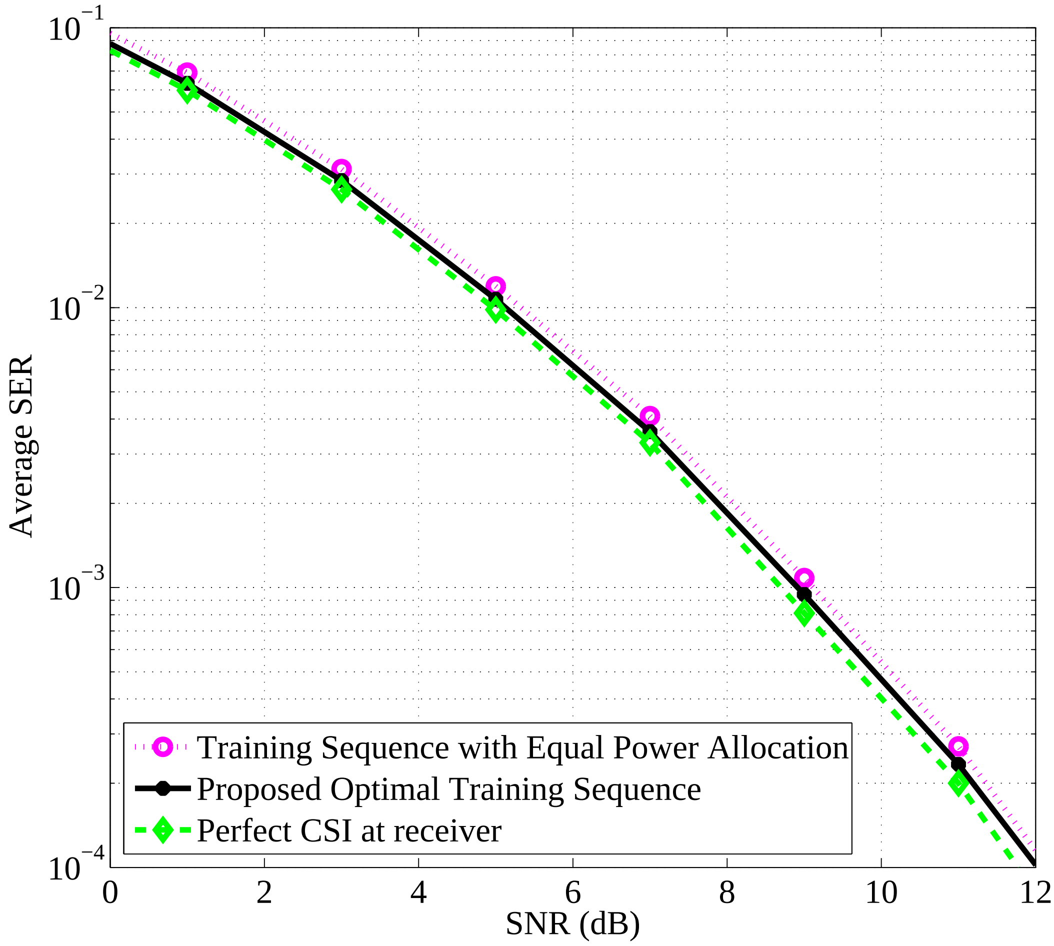


Figure 3.5: SER performance of the MIMO-OFDM system using the proposed training sequence

OFDM systems where standard CP-OFDM is replaced by ZP-OFDM-OLA. It is concluded that the optimal training sequence design for MIMO-ZP-OFDM-OLA systems is similar to that of MIMO-CP-OFDM systems and the performance of the LMMSE channel estimation is exactly the same.

The superior performance of the proposed training sequence in comparison to that of the existing training sequences was presented through the simulation results in Section 3.8.

3.10 Appendix: a STBC scheme for 4×4 MIMO communication system

In this appendix, we consider a STBC scheme for 4×4 MIMO communication system as presented in [38]

3.10.1 The STBC transmission model

We consider a wireless communication system with M_t transmit antennas and M_r receive antennas. At each time slot t, symbols $c_t^i, i = 1, 2, \ldots, M_t$ are simultaneously transmitted from the M_t transmit antennas. The channel is assumed to be a flat fading channel and the path gain from transmit antenna i to receive antenna j is defined to be α_{ij}. The wireless channel is assumed to be quasi-static so that the path gains are constant over a frame of length l and vary from one frame to another.

At time t, the received signal at antenna j is given by

$$r_t^j = \sum_{i=1}^{M_t} \alpha_{ij} c_t^i + n_t^j \tag{3.55}$$

where n_t^j is the AWGN noise presented at antenna j at time t.

Assuming perfect CSI is available, the receiver has to computes the decision metric

$$\sum_{t=1}^{l} \sum_{j=1}^{M_r} \left| \; r_t^j - \textstyle\sum_{i=1}^{M_t} \alpha_{ij} c_t^i \; \right|^2 \tag{3.56}$$

over all code words $c_1^1 c_1^2 \dots c_1^{M_t} c_2^1 c_2^2 \dots c_2^{M_t} \dots c_l^1 c_l^2 \dots c_l^{M_t}$ and decides in favour of the code word that minimizes the above sum.

3.10.2 Encoding algorithm

We consider a STBC scheme with the code rate of $3/4$, which means transmitting 3 information symbols over a frame of 4 time slots.

The transmitted codeword matrix of this STBC scheme is as follows

$$\mathscr{H}_4 = \begin{pmatrix} s_1 & s_2 & \frac{s_3}{\sqrt{2}} & \frac{s_3}{\sqrt{2}} \\ -s_2^* & s_1^* & \frac{s_3}{\sqrt{2}} & -\frac{s_3}{\sqrt{2}} \\ \frac{s_3^*}{\sqrt{2}} & \frac{s_3^*}{\sqrt{2}} & \frac{-s_1-s_1^*+s_2-s_2^*}{2} & \frac{-s_2-s_2^*+s_1-s_1^*}{2} \\ \frac{s_3^*}{\sqrt{2}} & -\frac{s_3^*}{\sqrt{2}} & \frac{s_2+s_2^*+s_1-s_1^*}{2} & -\frac{s_1+s_1^*+s_2-s_2^*}{2} \end{pmatrix}. \tag{3.57}$$

3.10.3 Decoding algorithm

Maximum likelihood decoding of this STBC scheme can be achieved by minimizing the decision metric for different information symbols independently.

More specifically, the decoder minimises the decision metric

$$\left| \left[\sum_{j=1}^{M_r} \left(r_1^j \alpha_{1j}^* + (r_2^j)^* \alpha_{2j} + \frac{(r_4^j - r_3^j)(\alpha_{3j}^* - \alpha_{4j}^*)}{2} - \frac{(r_3^j + r_4^j)^* (\alpha_{3j} + \alpha_{4j})}{2} \right) \right] - s_1 \right|^2 + \left(-1 + \sum_{j=1}^{M_r} \sum_{i=1}^{M_t} |\alpha_{ij}|^2 \right) |s_1|^2$$

for decoding s_1,

minimises the decision metric

$$\left| \left[\sum_{j=1}^{M_r} \left(r_1^j \alpha_{2j}^* - (r_2^j)^* \alpha_{1j} + \frac{(r_4^j + r_3^j)(\alpha_{3j}^* - \alpha_{4j}^*)}{2} + \frac{(-r_3^j + r_4^j)^* (\alpha_{3j} + \alpha_{4j})}{2} \right) \right] - s_2 \right|^2 + \left(-1 + \sum_{j=1}^{M_r} \sum_{i=1}^{M_t} |\alpha_{ij}|^2 \right) |s_2|^2$$

for decoding s_2,

minimises the decision metric

$$\left| \left[\sum_{j=1}^{M_r} \left(\frac{(r_1^j + r_2^j)\alpha_{3j}^*}{\sqrt{2}} + \frac{(r_1^j - r_2^j)\alpha_{4j}^*}{\sqrt{2}} + \frac{(r_3^j)^* (\alpha_{1j} + \alpha_{2j})}{\sqrt{2}} + \frac{(r_4^j)^* (\alpha_{1j} - \alpha_{2j})}{\sqrt{2}} \right) \right] - s_3 \right|^2 + \left(-1 + \sum_{j=1}^{M_r} \sum_{i=1}^{M_t} |\alpha_{ij}|^2 \right) |s_3|^2$$

for decoding s_3.

Chapter 4

Conclusion and Future Work

4.1 Thesis Conclusion

This thesis has focused on the optimal training sequence design for MIMO-OFDM systems in spatially correlated fading channels. Some important original contributions of this work are summarized as follows.

- The conditions for MIMO-OFDM optimal training sequences in spatially correlated fading environments discovered in [46] have been reviewed in this thesis. However, we have discovered that the existence of a training matrix cannot be guaranteed in general.

- A tight theoretical lower bound on the MSE of MIMO-OFDM channel estimation in spatially correlated fading environments has been extracted in Section 3.2

- We have considered the limit of the channel estimation MSE when the channel correlation matrices approach singularity in Section 3.3. We, then, have derived a general expression of the MSE lower bound in (3.23). This general expression does not depend on the invertibility of the channel correlation matrix.

- The optimal transmit power allocation for training sequences to achieve the MMSE performance of channel estimation has been found by resorting to the IBP [33]

- We have proposed an approximate algorithm for transmit power allocation in Section 3.5. This algorithm is more computationally efficient while maintaining a similar MSE performance in comparison with the above optimal solution.

- A practical and nearly optimal training sequence design has been demonstrated in Section 3.6. The proposed training sequence shows to have a superior MSE performance when compared with all existing training sequences. Its MSE performance almost achieves the MSE theoretical lower bound. It is also confirmed that its SER performance is very close to the SER performance when assuming perfect CSI at receiver.

- The training sequence design for MIMO-OFDM in spatially correlated fading environments can also be applied to MIMO-OFDM systems in which CP-OFDM is replaced by ZP-OFDM-OLA. This idea has been presented in Section 3.7.

4.2 Future Work

Although this thesis has addressed some fundamental problems of MIMO-OFDM optimal training sequence design, it will not be the last-mile research work on this topic. We are still considering some unsolved problems for our future research.

1. The training sequence design in this thesis has shown to be practically implemented and has superior MSE performance when comparing with that of all existing training sequences. The further step is to test this training sequence in various practical correlated fading models to verify its superiority.

2. Although the training sequence design proposed in this thesis demonstrates nearly optimal channel estimation MSE performance, the exact optimal training sequence in a general case of correlated fading channels has not been found yet. We emphasize that this optimal solution is not necessary to satisfy the conditions as reported in [46]. It is optimal in a sense that it has the best MSE performance. This

optimal solution can be found by formulating the mathematical optimization problem (3.1) into an equivalent Semi-Definite Programming (SDP) optimization problem [18]. This SDP problem can then be solved by some SDP softwares such as YALMIP [30]. However, the computational complexity of the current SDP softwares is still very high and impractical for wireless communication applications. Consequently, a more computationally efficient algorithm for finding the MIMO-OFDM optimal training sequence is still an open challenge.

3. In this thesis, information on the MIMO channel correlation matrices is assumed to be available at both transmitter and receiver for channel estimation and training sequence design. This is because the statistics of MIMO channels often change slowly with time. The receiver, therefore, has enough time to estimate the MIMO channel correlation matrices and feed back to the transmitter. However, the estimation of MIMO channel correlation matrices cannot achieve perfect accuracy in practice. This inaccurate information certainly has negative effect on the performance of the proposed training sequence which is presented in Chapter 3. This effect should be carefully considered and quantitatively evaluated when implementing the training sequence in practice.

Appendix A

Mathematical Background

In this appendix, we introduce some mathematical concepts which are necessary to understand the mathematical derivations in the thesis. Anyone who is interested in the topic of MIMO-OFDM is strongly recommended to carefully review Matrix Analysis in advance. Meanwhile, Convex Optimization is a very important class of mathematical optimization problems. It arises in a variety of applications in practice and can be solved numerically very efficiently. In this thesis, convex optimization is used to find the optimal transmit power allocation for MIMO-OFDM training sequences.

All the proofs will be omitted to keep the mathematical review brief. Interested readers can easily find the detailed treatment of these topics in the referenced standard textbooks [19] [31] [20] [16] [48] [11].

A.1 Matrix Analysis

A.1.1 Singular Value Decomposition (SVD)

Definition: For each matrix $\mathbf{A} \in \mathbf{C}^{m\times n}$ of rank r, there are unitary matrices $\mathbf{U} \in \mathbf{C}^{m\times m}$, $\mathbf{V} \in \mathbf{C}^{n\times n}$ and a diagonal matrix $\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ such that

$$\mathbf{A} = \mathbf{U} \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{m\times n} \mathbf{V}^H \tag{1.1}$$

The σ_is are called nonzero *singular values* of $\mathbf{A}$. The values of σ_is are the positive square roots of the eigenvalues of $\mathbf{A}^H\mathbf{A}$. When $r < p = \min\{m,n\}$, $\mathbf{A}$ is said to have $p-r$ additional zero singular values.

The above factorization is called SVD of matrix $\mathbf{A}$, and the columns of $\mathbf{U}$ and $\mathbf{V}$ are called left-hand and right-hand singular vectors for $\mathbf{A}$, respectively.

A.1.2 Kronecker Product

Definition: The *Kronecker product* of a matrix $\mathbf{A} \in \mathbf{C}^{m\times n}$ and a matrix $\mathbf{B} \in \mathbf{C}^{p\times q}$, is denoted by $\mathbf{A}\otimes\mathbf{B}$ and is defined to be the following block matrix

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix} \in \mathbf{C}^{mp\times nq} \tag{1.2}$$

Notice that $\mathbf{A}\otimes\mathbf{B} \neq \mathbf{B}\otimes\mathbf{A}$ in general.

Definition: With each matrix $\mathbf{A} \in \mathbf{C}^{m\times n}$, an associated vector vec $\mathbf{A} \in \mathbf{C}^{mn\times 1}$ is defined as

$$\text{vec } \mathbf{A} = [a_{11},\dots,a_{m1},a_{12},\dots,a_{m2},\dots,a_{1n},\dots,a_{mn}]^T \tag{1.3}$$

Some basic properties of the Kronecker product include: $(\alpha\mathbf{A})\otimes\mathbf{B} = \alpha(\mathbf{A}\otimes\mathbf{B})$

$(\mathbf{A}\otimes\mathbf{B})^H = \mathbf{A}^H\otimes\mathbf{B}^H$

$(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$

$(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C}+\mathbf{B}\otimes\mathbf{C}$

$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$

If $\mathbf{A} \in \mathbf{C}^{m\times m}$ and $\mathbf{B} \in \mathbf{C}^{n\times n}$ are nonsingular, then so is $\mathbf{A}\otimes\mathbf{B}$, and $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$

vec $\mathbf{AXB} = (\mathbf{B}^T\otimes\mathbf{A})$vec $\mathbf{X}$

vec $\mathbf{AX} = (\mathbf{I}\otimes\mathbf{A})$vec $\mathbf{X}$

A.1.3 Matrix Inversion Formulae

Let $\mathbf{A}$ be an invertible matrix partitioned as follows

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are also invertible matrices.

The blockwise inverse of matrix $\mathbf{A}$ is provided by the following *matrix inversion formulae*

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\boldsymbol{\Delta}_1^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\boldsymbol{\Delta}_1^{-1} \\ -\boldsymbol{\Delta}_1^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \boldsymbol{\Delta}_1^{-1} \end{pmatrix} \tag{1.4}$$

and

$$\mathbf{A}^{-1} = \begin{pmatrix} \boldsymbol{\Delta}_2^{-1} & -\boldsymbol{\Delta}_2^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\boldsymbol{\Delta}_2^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\boldsymbol{\Delta}_2^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{pmatrix} \tag{1.5}$$

where

$\boldsymbol{\Delta}_1 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ is the *Schur complement* of $\mathbf{A}_{11}$

$\boldsymbol{\Delta}_2 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ is the *Schur complement* of $\mathbf{A}_{22}$

Equating equation (1.4) and (1.5), we have the following important identity, which is called the *matrix inversion lemma*

$$(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\boldsymbol{\Delta}_1^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \tag{1.6}$$

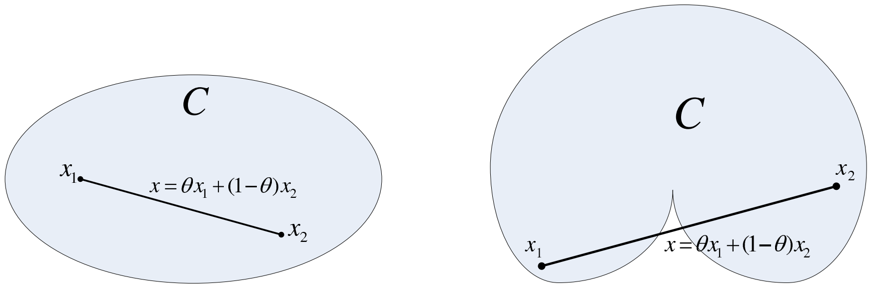
A.2 Convex Optimization

A.2.1 Convex Sets

Definition: A set C is *convex* if the line segment between any two points in C lies in C, i.e., if for any $x_1, x_2 \in C$ and any θ with $0 \leq \theta \leq 1$, we have

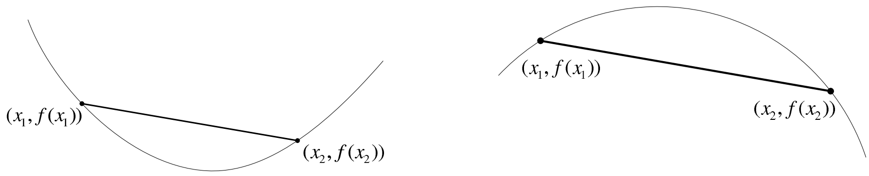
$$\theta x_1 + (1-\theta)x_2 \in C \tag{1.7}$$

Figure A.1 illustrates a simple example of convex and nonconvex set.



(a) Convex Set (b) Nonconvex Set

Figure A.1: An example of convex and nonconvex set



(a) Convex Function (b) Concave Function

Figure A.2: An example of convex and concave function

A.2.2 Convex Functions

Definition: A function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is *convex* if the domain of function f, denoted by $\mathbf{dom} f$, is a convex set and if for all $x_1, x_2 \in \mathbf{dom} f$, and θ with $0 \leq \theta \leq 1$, we have

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2) \tag{1.8}$$

Geometrically, this inequality means that the line segment between $(x_1, f(x_1))$ and $(x_2, f(x_2))$, which is the chord from x_1 to x_2, lies above the graph of f (figure A.2).

We say f is *concave* if $-f$ is convex.

A.2.3 Necessary and Sufficient Conditions for Convex Functions

First-order conditions: Suppose f is differentiable (i.e., its gradient ∇f exists at each point in $\mathbf{dom} f$). Then f is convex if and only if $\mathbf{dom} f$ is convex and

$$f(x) \geq f(x_1) + \nabla f(x_1)^T (x - x_1) \tag{1.9}$$

hold for all $x_1, x \in \mathbf{dom} f$

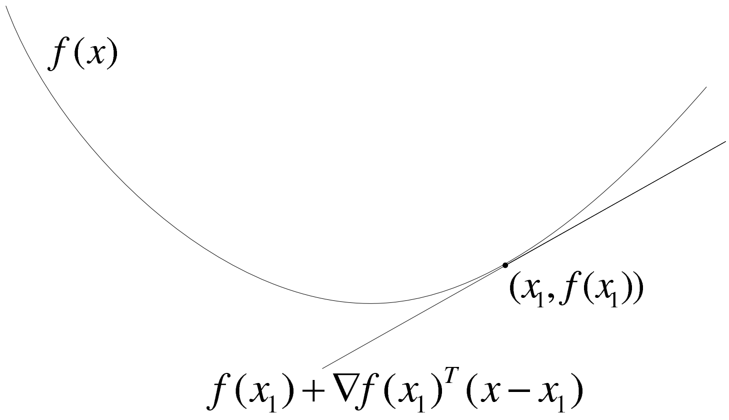


Figure A.3: If f is convex and differentiable, then $f(x) \geq f(x_1) + \nabla f(x_1)^T(x - x_1)$ for all $x_1, x \in \textbf{dom}f$

Second-order conditions: We now assume that f is twice differentiable, that is, its Hessian $\nabla^2 f$ exists at each point in $\textbf{dom}f$. Then f is convex if and only if $\textbf{dom}f$ is convex and its Hessian is positive semidefinite: for all $x \in \textbf{dom}f$, which is

$$\nabla^2 f(x) \succeq 0 \tag{1.10}$$

For a function on $\mathbf{R}$, this reduces to the simple condition $f''(x) \geq 0$

A.2.4 Optimization Problems

Definition: A *mathematical optimization problem*, or just *optimization problem*, has the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, p \end{aligned} \tag{1.11}$$

where

The vector $\mathbf{x} = (x_1, \ldots, x_1)$ is the *optimization variable* of the problem

The function $f(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$ is the *objective function*

The functions $g_i(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$ are called the *inequality constraint functions* and the inequalities $g_i(\mathbf{x}) \leq 0$ are called *inequality constraints*

The functions $h_j(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$ are called the *equality constraint functions* and the equalities $h_j(\mathbf{x}) \leq 0$ are called *equality constraints*

The set of points for which the objective and all constraint functions are defined

$$\mathscr{D} = \mathbf{dom}\, f \cap \bigcap_{i=1}^{m} \mathbf{dom}\, g_i \cap \bigcap_{j=1}^{p} \mathbf{dom}\, h_j$$

is called the *domain* of the optimization problem (1.11).

A point $\mathbf{x} \in \mathscr{D}$ is *feasible* if it satisfies the constraints $g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m$ and $h_j(\mathbf{x}) = 0,\ j = 1, \ldots, p$. The problem (1.11) is said to be *feasible* if there exists at least one feasible point, and *infeasible* otherwise. The set of all feasible points is called the *feasible set*.

A vector $\mathbf{x}^*$ is called an optimal solution of the problem (1.11), if it has the smallest objective value among all vectors that satisfy the constraints, i.e., for any $\mathbf{x}$ with $g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m$ and $h_j(\mathbf{x}) = 0,\ j = 1, \ldots, p$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$

A.2.5 Convex Optimization

Definition: A *convex optimization problem* has a form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, p \end{aligned} \tag{1.12}$$

where all the objective and constraint functions are convex. And the equality constraint functions must be affine, i.e., $h_j(\mathbf{x}) = \mathbf{a}_j^T \mathbf{x} - \mathbf{b}_j$

Notice that the feasible set of a convex optimization problem is convex, since it is the intersection of convex sets $\mathscr{D} = \mathbf{dom}\, f \cap \bigcap_{i=1}^{m} \mathbf{dom}\, g_i$

A.2.6 The Lagragian

Consider an optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, p \end{aligned}$$

with variable $\mathbf{x} \in \mathbf{R}^n$. Assuming that domain of the optimization problem $\mathscr{D} = \mathbf{dom}\, f \cap \bigcap_{i=1}^{m} \mathbf{dom}\, g_i \cap \bigcap_{j=1}^{p} \mathbf{dom}\, h_j$ is nonempty.

Definition: *Lagragian* $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ associated with the above optimization problem as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=0}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=0}^{p} \nu_j h_j(\mathbf{x}) \tag{1.13}$$

with $\mathbf{dom}\, L = \mathscr{D} \cap \mathbf{R}^m \cap \mathbf{R}^p$

λ_i is referred as the *Lagrange multiplier* associated with the i^{th} inequality constraint $g_i(\mathbf{x}) \le 0$

ν_j is referred as the *Lagrange multiplier* associated with the j^{th} equality constraint $h_j(\mathbf{x}) = 0$

A.2.7 Karush-Kuhn-Tucker (KKT) Optimality Conditions for Convex Problems

Consider a convex optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, p \end{aligned}$$

where $f, g_1, \ldots, g_m$ are differentiable and convex functions.

Necessary and sufficient conditions for $\mathbf{x}^*$ to be the optimal solution for the above convex optimization problem is that there exist $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$, $\boldsymbol{\nu}^*$ satisfy the following KKT conditions

$$
\begin{cases}
g_i(\mathbf{x}^*) \leq 0,\ i = 1, \ldots, m \\
\mathbf{a}_j^T \mathbf{x}^* - \mathbf{b}_j = 0,\ j = 1, \ldots, p \\
\lambda_i^* \geq 0, i = 1, \ldots, m \\
\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m \\
f(\mathbf{x}^*) + \sum_{i=0}^{m} \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=0}^{p} \nu_j^* h_j(\mathbf{x}^*) = 0
\end{cases}
\tag{1.14}
$$

Bibliography

[1] "Part 11: Wireless LAN medium access control (MAC) and physical layer specifications," *IEEE Standard 802.11n-2009*.

[2] "Part 16: Air interface for broadband wireless access systems," *IEEE Standard 802.16-2009*.

[3] "Physical layer - general description," *3GPP Technical Specification 25.201, Release 9*.

[4] S. M. Alamouti, "A simple transmit diversity technique for wireless," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, Oct 1998.

[5] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1615–1624, Jun 2003.

[6] G. Barriac and U. Madhow, "Space time communication for OFDM with implicit channel feedback," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3111–3129, Dec 2004.

[7] J. V. D. Beek and et al, "On channel estimation in ofdm systems," *45th IEEE Vehicular Technology Conference*, vol. 2, pp. 815–819, Jul 1995.

[8] H. Bölcskei, "MIMO-OFDM wireless systems: Basics, perspectives, and challenges," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 31–37, Aug 2006.

[9] H. Bolcskei, *Principles of MIMO-OFDM wireless systems*, 2004. [Online]. Available: http://www.nari.ee.ethz.ch/commth/pubs/p/crc03

[10] H. Bolskei and A. Paulraj, "Space-frequency coded broadband OFDM systems," *Wireless Communications and Networking Conference*, vol. 1, pp. 1–6, September 2000.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience; 2nd edition, 2006.

[13] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, Oct 1996.

[14] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications - Springer*, vol. 6, pp. 311–335, March 1998.

[15] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[16] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.

[17] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm I. expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2806–2818, August 2005.

[18] T. Hoang, V. Luong, and H. Nguyen, "Optimization of training sequences for spatially correlated MIMO-OFDM," *ICASSP 2009, to appear*.

[19] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[20] ——, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[21] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge University Press, 2005.

[22] M. Jankiraman, *Space-Time Codes and MIMO Systems*. Artech House Publishers, 2004.

[23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall PTR, 1993.

[24] J. H. Kotecha and A. M. Sayeed, “Transmit signal design for optimal estimation,” *IEEE Transactions on Signal Processing*, vol. 52, pp. 546–557, Feb 2004.

[25] Y. Li, L. J. Cimini, and N. R. Sollegberger, “Robust channel estimation for OFDM systems with rapid dispersive fading channels,” *IEEE Transactions on Communications*, vol. 46, pp. 902 – 915, July 1998.

[26] Y. Li, N. Seshadri, and S. Ariyavisitakul, “Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels,” *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 461 – 471, March 1999.

[27] Y. G. Li, “Simplified channel estimation for OFDM systems with multiple transmit antennas,” *IEEE Transactions on Wireless Communications*, vol. 1, pp. 67–75, Jan 2002.

[28] Y. G. Li and G. L. Stuber, *Orthogonal Frequency Division Multiplexing for Wireless Communications*. Springer, 2006.

[29] Z. Liu, Y. Xin, and G. Giannakis, *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2465–2476, October 2002.

[30] J. Lofberg, “YALMIP: a toolbox for modeling and optimization in MATLAB,” *2004 IEEE Interntional Symposium on Computer Aided Control System Design*, pp. 284–289, Sep 2004.

[31] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.

[32] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?" *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 2136– 2148, Dec 2002.

[33] V. Nguyen, T. D. Hoang, H. H. Nguyen, and N. N. Tran, "Optimal superimposed training design for spatially correlated fading MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, Aug 2008.

[34] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Prentice Hall, 1999.

[35] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bölcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb 2004.

[36] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*. Prentice Hall, 2006.

[37] G. Stuber, J. Barry, S. Mclaughlin, Y. Li, and et al, "Broadband MIMO-OFDM wireless communications," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 271–294, Feb 2004.

[38] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, Jul 1999.

[39] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744 – 765, March 1998.

[40] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999.

[41] N. N. Tran, H. D. Tuan, and H. H. Nguyen, "Superimposed training design for spatially correlated MIMO-OFDM systems," *IEEE Globecom 2008, to appear*.

[42] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[43] B. Vucetic and J. Yuan, *Space-Time Coding*. Wiley, 2003.

[44] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 29–48, May 2000.

[45] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering channel," *Proc. Int. Symp. Signals, Systems and Electronics (ISSE 1998)*, pp. 295 – 300, Sep 1998.

[46] H. Zhang, Y. Li, A. Reid, and J. Terry, "Optimum training symbol design for MIMO OFDM in correlated fading channels," *IEEE Trans. Wireless Comm*, vol. 5, no. 9, pp. 2343–2347, Sep 2006.

[47] W. Zhang, X. G. Xia, and K. B. Letaief, "Space-time/frequency coding for MIMO-OFDM in next generation broadband wireless systems," *IEEE Wireless Communications Magazine*, vol. 14, no. 3, pp. 32–43, Jun 2007.

[48] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Prentice Hall, 1997.