國立交通大學

電機與控制工程學系

碩 士 論 文

正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測

On Frequency-Domain Equalization with Training-Based Channel Estimation for Orthogonal Space-Time Block Coded System via MIMO Frequency-Selective Fading Channels

研 究 生：楊傑智

指導教授：林清安　教授

中 華 民 國 九 十 六 年 八 月

正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測

On Frequency-Domain Equalization with Training-Based Channel Estimation for Orthogonal Space-Time Block Coded System via MIMO Frequency-Selective Fading Channels

研 究 生：楊傑智　　　Student：Chieh-Chih Yang

指導教授：林清安　　　Advisor：Ching-An Lin

國 立 交 通 大 學
電 機 與 控 制 工 程 學 系
碩 士 論 文

A Thesis

Submitted to Department of Electrical and Control Engineering

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Electrical and Control Engineering

August 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年八月

國 立 交 通 大 學

博碩士論文全文電子檔著作權授權書

（提供授權人裝訂於紙本論文書名頁之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學電機與控制工程系所 ______________組， 95 學年度第_____學期取得碩士學位之論文。

論文題目：正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測
指導教授：林清安

■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學與台灣聯合大學系統圖書館：基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學及台灣聯合大學系統圖書館得不限地域、時間與次數，以紙本、光碟或數位化等各種方法收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

論文全文上載網路公開之範圍及時間：

本校及台灣聯合大學系統區域網路	■ 立即公開
校外網際網路	■ 立即公開

■ 全文電子檔送交國家圖書館

授 權 人：楊傑智

親筆簽名：_________________________

中華民國　　　　年　　　　月　　　　日

國 立 交 通 大 學

博碩士紙本論文著作權授權書

（提供授權人裝訂於全文電子檔授權書之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學電機與控制工程系所 ＿＿＿＿＿＿＿＿組， 95 學年度第＿＿＿學期取得碩士學位之論文。

論文題目：正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測
指導教授：林清安

■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學，基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學圖書館得以紙本收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行閱覽或列印。

本論文為本人向經濟部智慧局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：＿＿＿＿＿＿＿＿＿＿＿＿＿＿，請將論文延至＿＿＿年＿＿＿月＿＿＿日再公開。

授 權 人：楊傑智

親筆簽名：＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

中華民國　　　　年　　　　月　　　　日

國家圖書館
博碩士論文電子檔案上網授權書

（提供授權人裝訂於紙本論文本校授權書之後）

ID:GT009412501

本授權書所授權之論文為授權人在國立交通大學電機與控制工程系所 95 學年度第___學期取得碩士學位之論文。

論文題目：正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測
指導教授：林清安

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

※ 讀者基於非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：楊傑智

親筆簽名：________________

民國　　年　　月　　日

國立交通大學

論文口試委員會審定書

本校 電機與控制工程 學系碩士班 楊傑智 君

所提論文 正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的擇頻通道估測

合於碩士資格標準、業經本委員會評審認可。

口試委員：

馬[illegible]	林源倍
吳卓諭	林清安

指導教授： 林清安 教授

系 主 任： 邱俊誠 教授

中 華 民 國 九十六 年 六 月 二十二 日

正交時空區塊碼之頻域等化及其基於訓練機制之多輸入多輸出的
擇頻通道估測

學生：楊傑智　　　　　　　　　　　　　　　　　　指導教授：林清安

國立交通大學電機與控制工程學系（研究所）碩士班

摘　要

針對正交時空區塊碼應用於多輸入多輸出之擇頻通道上的問題，吾人於本文中提出了一個嶄新且完整的推導方式；另著墨於此架構下所能達到的最大之多重增益－包括來自於天線以及多重路徑所貢獻者。藉由多次的模擬實驗結果發現，吾人可以得到一個於實際考量下為合理的假設，進而由此推導錯誤率分析。錯誤率分析的推導結果證實了吾人所提出之架構確實是有能力達到完整的多重增益。另一方面，藉由接收端的頻域等化機制以及所提出之傳輸架構，吾人可以進一步找到基於訓練機制時最佳的通道估測所需之相對應的訊號設計模式。本論文之結果為線性等化器能否達到完整的多重增益提供了一個正面的佐證。

On Frequency-Domain Equalization with Training-Based Channel Estimation for Orthogonal Space-Time Block Coded System via MIMO Frequency-Selective Fading Channels

Student：Chieh-Chih Yang Advisor：Ching-An Lin

Department（Institute）of Electrical and Control Engineering
National Chiao Tung University

Abstract

We propose an instructive derivation for the generalized block-level orthogonal space-time block encoder, capable of achieving full spatial diversity via frequency-selective fading environment provided that channel order is known. Instead of dealing with special case and then extending the results intuitively, we provide an alternative by starting with the general signal model with multiple transmit and multiple receive antennas, from which a general form of block-level orthogonality is established. In particular, transmit diversity with more than two transmit antennas can be achieved without compromise by means of frequency-domain equalization, in contrast to the QO-STBC-based approach. Pairwise error probability analysis is derived, under certain assumption which is numerically supported by simulation results, for analytical verifications of our claim on full diversity, inclusive of transmit-receive diversity and the multipath one. Moreover, the encoder structure enables us to generalize a training-based channel estimation technique, originally proposed for flat-fading scenario, to the frequency-selective fading scenario. Surprisingly we even obtain similar optimality criteria for optimal training block design which in our case, the signal block are fixed as OSTBC-based and the design derivation reduces to derive optimal power constraint over the training blocks. The optimality criteria for the training blocks are easy to satisfy when randomness of signal constellation is not a concern. Simulation results validate our discussion of the behaviors of the least-squares and linear MMSE channel estimates.

Acknowledgement

I would like thank my instructor, Prof. Ching-An Lin, for his teaching, enlightenment and support over the past two years. The thesis would never be complete without his dedication to the revisions and his provision of ingenuity. The valuable joint advice from Prof. Yuan-Pei Lin, Prof. Wing-Kin Ma and Dr. Jwo-Yu Wu serves no less than the above efforts. Especially, issues regarding simulation accuracy and rigorousness of proofs are greatly improved with their assistances. Moreover, the generous helps from the seniors of our laboratory and my friends see me through many difficulties, and I am cordially grateful to all of them. Above all, I would like thank my family for their encouragement.

Contents

摘要 .... **i**

Abstract .... **ii**

Acknowledgement .... **iii**

List of Figures .... **vii**

Chap. 1 Introduction **1**

1.1 Motivation and Background 1

1.2 Thesis Overview 3

1.3 Notations 4

Chap. 2 System Model **5**

2.1 System Configuration 5

2.1.1 BGOSTBC Encoder 5

2.1.2 BGOSTBC Encoder (with CP Insertions) 7

2.2 Block Transmission and IBI-free Model upon CP Removal 8

2.2.1 MIMO Frequency-selective Fading Channel and Noise Model 8

2.2.2 Block Transmission via MIMO Frequency-Selective Fading Channels 8

Chap. 3 Frequency-Domain Equalization 11

3.1 Overall Frequency-Domain Received Signal Model 11

3.1.1 Space-Time Combining Using Signal DFT's 12

3.1.2 Frequency-Domain Equalization with Perfect CSI 13

Chap. 4 Pairwise Error Probability Analysis 17

4.1 PEP Analysis for Suboptimal Detection Problem 17

Chap. 5 Training-Based Channel Estimation 28

5.1 Least-Squares Channel Estimate 28

5.1.1 Optimality Criterion for LS Estimate under a Power Constraint 31

5.1.2 Selection of training signal blocks for optimal LS channel estimate 33

5.2 Linear Minimum-Mean-Square Error Estimate 34

5.2.1 Optimality Criterion for Linear MMSE Estimate under a Power Constraint 36

Chap. 6 Simulation Results 40

6.1 Assumptions for Numerical Simulation Runs 40

6.2 Result Overview and Discussion 40

6.2.1 Arbitrary Training: NMSE Performance 41

6.2.2 Optimal/Orthogonal Training: NMSE Performance 41

Chap. 7 Conclusion 42

Appendix 43

A1. Proof of Block-Level Orthogonality part I 43

A2. Proof of block-level orthogonality part II 44

A3. OSTBC construction 45

References 53

List of Figures

Fig. 1 Transmission scheme based on BGOSTBC encoder and the associated FDE based on training-based channel estimation. **48**

Fig. 2 A BGOSTBC encoder employing K=4, M_t=3 configuration. **48**

Fig. 3-1 Condition numbers generated for 10^6 independent simulation runs under $(N,M_r,M_t,L,K)=(10,1,3,1,4)$ **49**

Fig. 3-2 Condition numbers generated for 10^6 independent simulation runs under $(N,M_r,M_t,L,K)=(16,1,3,1,4)$ **49**

Fig. 3-3 Condition numbers generated for 10^6 independent simulation runs under $(N,M_r,M_t,L,K)=(16,1,3,2,4)$ **50**

Fig. 3-4 Condition numbers generated for 10^6 independent simulation runs under $(N,M_r,M_t,L,K)=(10,2,6,1,8)$ **50**

Fig. 3-5 Condition numbers generated for 10^6 independent simulation runs under $(N,M_r,M_t,L,K)=(10,2,6,3,8)$ **51**

Fig. 3 Compare NMSE of LS and linear MMSE estimates, with arbitrary training. 51

Fig. 4 Compare NMSE of arbitrary training and optimal/orthogonal training. **52**

Fig. 5 Demonstration of full multipath diversity at high SNR region, averaged over 100 independent runs. **52**

Chapter 1

Introduction

1.1 Motivation and Background

Orthogonal space-time block code (OSTBC) was introduced in [1] for combating channel fading by exploiting diversities in multiple-input-multiple-output (MIMO) antenna configuration. By virtue of orthogonal design, the symbol decoding relies on only linear processing with relatively lower computational complexity than its trellis counterpart, provided that the channel state information is available. Hence both blind channel estimation and joint symbol detection using OSTBC architecture are well studied in flat-fading scenario.

Various transmission schemes and algorithms based on suitably manipulated orthogonality conditions have been proposed for flat-fading environment[19],[20]. In [19], joint signal detection for a flat-fading MIMO communication under OSTBC scheme was studied, where the exponent, embedded with OSTBC structure, of the closed-form log-likelihood was to be minimized in the maximum likelihood (ML) sense. A non-polynomial time (NP-hard) ML exhaustive search was cleverly reformulated and relaxed as a convex optimization problem which had been well studied theoretically and numerically. The relaxation is known as semidefinite relaxation (SDR) with corresponding numerical algorithm tailored in [21]. The resultant SDR-ML proposed by Ma *et al.* performs substantially better than the cyclic ML method[22]. It is noteworthy that the reformulation itself exploited the orthogonality of OSTBC, by which a near optimum signal detection with relatively lower computational cost than the optimal ML/sphere decoding was achieved. In [20], a closed-form channel estimation technique based on a variant of the generic orthogonality of OSTBC was developed. This variant was first derived by Alex B. Gershman *et al.* in [23], where real parts and imaginary parts of the received signal model were separated with discretion. The closed-form channel estimation, unlike the other approaches such as subspace method, suffers from sign ambiguity only due to its special real-valued formulation of signal models. Several sufficient antenna configurations for unique channel estimate were studied numerically. The uniqueness of the channel estimate was determined by the discrepancy between the eigenvectors of a certain objective which was formed by exploiting the orthogonality. For those configurations where uniqueness failed, i.e., algebraic multiplicity of eigenvectors greater than 1, the authors proposed a diagonal precoder to alienate the eigenvectors. Benefited from the orthogonality, this can be surprisingly easily done by assigning the power weighting coefficients with sufficient discrepancy in an ad-hoc manner.

However, it remains challenging for the above mentioned methods to be effectively extended to frequency-selective fading environment which is a more practical consideration. By effective extension we mean the algorithm on which the estimation or decoding is based should enjoy either full or partial diversity due to orthogonal nature of the coding itself. To apply the OSTBC structure to frequency- selective case without compromising the code orthogonality, E. Lindskog and A. Paulraj [7] cleverly combine the cyclic-prefix (CP) mechanism with time-reversal operation on symbol blocks, bringing the space-time concept to "block-level". It was then incorporated with Alamouti scheme[8] and known as time-reversed Alamouti-like (TR-Alamouti) scheme[9]. The TR-Alamouti scheme is unique in that it enjoys full 2-fold transmit diversity and nearly full transmission rate, when neglecting the CP overheads, at the same time. A general block-level orthogonal space-time block code was first proposed by Z. Liu *et al.*[14] for consideration regarding preserving orthogonality over frequency-selective channels with more than two transmit antennas, where the generalized complex orthogonal design (GOSTBC) [1] was adopted with ZP assistance for mitigating channel distortion. On the other hand, there have been extensive studies based on the 2-fold diversity scheme with frequency-domain equalization (FDE), such as [11],[12] and [13]. As for a general FDE scheme having more than 2 transmit antennas, a compromising method based on quasi-orthogonal STBC (QO-STBC) [10] was proposed in [2], where the nearly full rate was preserved at the cost of achieving only partial spatial diversity. Achieving full spatial diversity, particularly based on FDE and cast into the general structure in [14], was reported in [26]. However, the derivation in [26] is based on the 2-fold structure as a start and then generalized to multiple-antenna scenario *intuitively* by introducing the block-level concept in [14]. It is as instructive as important to build up the signal model with multiple transmit and multiple receive antennas, from which a general form of block-level orthogonality will be established.

Aside from the derivation issue, we will counterbalance the deduction in [26] that the "CP-only" scheme cannot exploit full multipath diversity, by giving a PEP analysis. Although a premise which is numerically supported by simulations must be met for justification of our PEP analysis, it does shade some light on extending the discussions in [17] to a more general scheme.

Also, optimal training design for MIMO communications in either flat-fading or frequency-selective fading environments is an important topic in practice. A pilot symbol-aided linear MMSE-based training scheme with optimal/orthogonal training is considered in [15]. In [16], the discussion in [15] was generalized to take advantage of space-time diversity. Nonetheless, none of which particularly considered the OSTBC class, and hence the optimal designs were quite involved. It will be shown that with the structure of training blocks fixed as OSTBC, the optimal training design can be simplified to optimal power allocation design.

1.2 Thesis Overview

In this thesis, we introduce a generalized FDE technique based on structure of [14] and by extending the training-based channel estimation approach in [3] we arrive at similar design criteria for optimal/orthogonal training as those obtained in [3] which considers flat-fading scenario. We propose an extended block-level OSTBC scheme capable of achieving full spatial and multipath diversities over frequency-selective fading channels when more than two transmit antennas involved by using FDE. What differs from [26] is that instead of starting from the 2-fold special case, we provide a new and instructive derivation based on general multiple-antenna signal model. The proposed scheme has nearly 1/2 symbol rate when discarding CP overheads. Pairwise error probability (PEP) analysis for demonstrating full spatial and full multipath diversities will be given. It will also be shown that since the signal model resembles that in flat-fading scenario, the optimal training designs such as those developed by M. Biguesh *et al.*[3] can be generalized to the frequency-selective fading scenario in a straightforward manner. The extended training-based channel estimation is optimal in least-squares (LS) sense in frequency-domain under a given power constraint, provided that a power criterion is satisfied. Also a time-domain linear MMSE channel estimation technique can be developed in a similar fashion as [3]. Adopting the Alamouti scheme for the construction of the proposed encoder with proper changes in dimension and scalar factor makes the 2-fold diversity scheme a special case.

In the next chapter, we will propose a transmission scheme and outline that entire system model. Post-processing including the subsequent FDE for data transmission and the training-based channel estimation for training mode will be introduced in chapter 3 and 5, respectively. Also, the optimality conditions for achieving lowest possible NMSE will be derived accordingly. In chapter 4, we will derive our PEP analysis. Simulation results of NMSE vs. SNR for training-based channel estimation are discussed in chapter 6.

1.3 Notations

The following notations are adopted throughout the thesis: $\mathbf{P}^*$, $\mathbf{P}^T$ and $\mathbf{P}^H$ denote conjugate, transpose and conjugate-transpose of matrix $\mathbf{P}$, respectively. $\mathbf{A}\otimes\mathbf{B}$ stands for the kronecker product of matrix $\mathbf{A}$ and $\mathbf{B}$. Let Re$\{\mathbf{P}\}$ and Im$\{\mathbf{P}\}$ stand for the real and imaginary parts of matrix $\mathbf{P}$, respectively. Let Tr$\{\mathbf{P}\}$ and vec($\mathbf{P}$) denote the trace and the vectorization of matrix $\mathbf{P}$, respectively. For $\mathbf{y}\in\mathbb{C}^{\mathrm{N}}$, $Diag(\mathbf{y})\in\mathbb{C}^{\mathrm{N\times N}}$ stands for an N by N diagonal matrix with $\mathbf{y}$ on its main diagonal. For $\mathbf{A}\in\mathbb{C}^{\mathrm{N\times N}}$, $Diag(\mathbf{A})\in\mathbb{C}^{\mathrm{N\times N}}$ stands for the vector whose $i^{\underline{th}}$ entry is the $i^{\underline{th}}$ diagonal entry of A. For a matrix $\mathbf{A}\in\mathbb{C}^{\mathrm{N\times M}}$, $[\mathbf{A}]_{ij}$ denotes the entry at the $i^{\underline{th}}$ row-and-$j^{\underline{th}}$ column position of $\mathbf{A}$. $\mathbf{P}(i{:}j,m{:}n)$ denotes an extracted submatrix consists of from $i^{\underline{th}}$ to $j^{\underline{th}}$ rows and from $m^{\underline{th}}$ to $n^{\underline{th}}$ columns of matrix $\mathbf{P}$. $\mathbf{P}(:,m{:}n)$ indicates that all rows ranging from $m^{\underline{th}}$ to $n^{\underline{th}}$ columns are referred. Similarly define $\mathbf{P}(i{:}j,:)$. The symbol $\boldsymbol{\mathcal{F}}$ is preserved for N by N normalized discrete-time Fourier transform (DFT) matrix, with the $(m,n)^{\underline{th}}$ entry of $\boldsymbol{\mathcal{F}}$ being $[\boldsymbol{\mathcal{F}}]_{mn}=\frac{1}{\sqrt{N}}e^{-j\frac{2\pi(m-1)(n-1)}{N}}$, $1\le m,n\le N$.

Chapter 2

System Model

In this chapter, we will introduce a transmission scheme based on a block-level extension of OSTBC. As a preliminary, we will also review some basic properties of orthogonal design of space-time block codes, which are to be employed for introducing block-level generalized orthogonal space-time block code (BGOSTBC) encoder as depicted in Fig. 2. The mechanism of CP insertion and CP removal for combating inter-block interference (IBI) will be reviewed at the bottom of this chapter.

2.1 System Configuration

As depicted in Fig. 1, the overall system consists of a BGOSTBC encoder followed by CP insertions at the transmitting end while the receiving end comprises CP removal, DFT and the subsequent channel estimation plus FDE. Let M_t and M_r denote the number of transmit antennas and receive antennas, respectively. Assume that the channel order is L (L+1 taps) for all the subchannels and known a priori. We assume the CP length is exactly L. Let N be the block length and $N \geq L+1$. The information symbol blocks to be transmitted are accumulated over K blocks, each one of which will be sent by a certain transmit antenna during a specific time epoch with CP insertion. The block ordering is set up according to the proposed BGOSTBC encoder which occupies $2K$ time epochs for transmitting KN symbols. Each time epoch lasts for N+L symbol periods, where the redundant L symbol duration accounts for the CP insertion. At the receiving end, the received signal blocks over the entire $2K$ time epochs are buffered after CP removals. Then the equivalent IBI-free received blocks are Fourier transformed. With the outputs at the DFT system block available, we can either acquire channel estimation in training mode or perform FDE on the transmitted information symbols.

2.1.1 BGOSTBC Encoder

In this subsection, the referred time epoch lasts for N symbol periods only since no CP insertion involved. Now suppose we have collected K symbol blocks, $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_K$. Let $\boldsymbol{\mathcal{C}} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_K\end{bmatrix}$, where $\mathbf{c}_k \in \mathbb{C}^{\mathrm{N}}$ denotes the $k^{\underline{th}}$ information symbol block to be transmitted, whose $n^{\underline{th}}$ symbol is denoted by $c_k(n)$, $0 \le n \le N-1$, $1 \le k \le K$. The transmission scheme of the BGOSTBC encoder can be regarded as a forward transmission mode followed by a reversed transmission mode, each of them occupying exactly K epochs. In the latter mode each incoming signal block is time-reversed and conjugated prior to transmission, while in the former mode transmission is carried out with blocks unmodified.

Now, let us define the encoder output. The encoder has single serial input and parallel M_t outputs. Let $\mathbf{s}_m^{(p)} \in \mathbb{C}^{\mathrm{N}}$ denote the signal block transmitted from the $m^{\underline{th}}$ encoder output path during the $p^{\underline{th}}$ time epoch. Collecting encoder output over $2K$ time epochs and across M_t encoder output paths yields a matrix of dimension $2NK$ by M_t, whose $((p\text{-}1)N+1{:}pN,m)^{\underline{th}}$ block-entry is $\mathbf{s}_m^{(p)}$ by definition. The so obtained matrix is essentially a block-level extension of generalized complex orthogonal design[1]. See also [14]. Let us define $\mathbf{G}_B$ as the overall output at the BGOSTBC encoder stacking across $2K$ epochs and herein list some properties of it:

$$\mathbf{G}_B \triangleq \sum_{k=1}^{K}\left[\mathbf{X}_{A_k} \otimes \mathbf{c}_k + \mathbf{X}_{B_k} \otimes \mathbf{d}_k^*\right] \in \mathbb{C}^{2\mathrm{NK}\times \mathrm{M_t}}, \tag{2.1}$$

where

$$\begin{cases}\mathbf{d}_k = \begin{bmatrix}d_k(0) & d_k(1) & \cdots & d_k(N-1)\end{bmatrix}^T. \\ d_k(n) = c_k((\text{-}n)_N),\ 0 \le n \le N-1. \\ \mathbf{X}_{A_k},\ \mathbf{X}_{B_k} \in \mathbb{R}^{2\mathrm{K}\times\mathrm{M_t}},\ 1 \le k \le K. \\ \mathbf{X}_{A_k}^T\mathbf{X}_{A_l} = \begin{cases}\mathbf{I}_{M_t},\ 1 \le k = l \le K \\ -\mathbf{X}_{A_l}^T\mathbf{X}_{A_k},\ 1 \le k \ne l \le K\end{cases}. \\ \mathbf{X}_{B_k}^T\mathbf{X}_{B_l} = \begin{cases}\mathbf{I}_{M_t},\ 1 \le k = l \le K \\ -\mathbf{X}_{B_l}^T\mathbf{X}_{B_k},\ 1 \le k \ne l \le K\end{cases}. \\ \mathbf{X}_{A_k}^T\mathbf{X}_{B_l} = \mathbf{0}_{M_t\times M_t},\ 1 \le k,l \le K.\end{cases}$$

Specifically

$$\mathbf{X}_{A_k} = \begin{bmatrix} \boldsymbol{\mathcal{G}}_{A_k} \\ \hline \mathbf{0}_{K\times M_t} \end{bmatrix} \text{ and } \mathbf{X}_{B_k} = \begin{pmatrix} \mathbf{0}_{K\times K} & \mathbf{I}_K \\ \mathbf{I}_K & \mathbf{0}_{K\times K} \end{pmatrix} \mathbf{X}_{A_k}, \tag{2.2}$$

where $\boldsymbol{\mathcal{G}}_{A_k} \in \mathbb{C}^{\mathrm{M_t}\times\mathrm{K}}$ are generic constituent matrices of certain OSTBC design of dimensions depending of choices of K and M_t. See also Appendix A3 and [1] for a detailed description about the construction of constituent matrices of OSTBC. The time-reversal, $d_k(n) = c_k((\text{-}n)_N)$, is carried out via the modulo-N operation defined as $(\text{-}n)_N = N - n$ for $1 \le n \le N-1$ and $(\text{-}n)_N = 0$ for $n = 0$. Note that $\mathbf{X}_{A_k}, \mathbf{X}_{B_k} \in \mathbb{R}^{2\mathrm{K}\times\mathrm{M_t}}$, $1 \le k \le K$, are constituent matrices of a certain GOSTBC determined by choice of (K,M_t). $\mathbf{X}_{A_k}$ and $\mathbf{X}_{B_k}$ are non-overlapping matrices consisting of only ones and zeros up to sign changes, and are responsible for designating space-time ordering of the forward transmission mode and the reversed one, respectively, i.e., $\mathbf{s}_m^{(p)}$ takes $\mathbf{c}_k$ if $\left|\left[\mathbf{X}_{A_k}\right]_{pm}\right|$ is one, for $1 \le p \le K$. Similarly, $\mathbf{s}_m^{(p)}$ takes $\mathbf{d}_k^*$ if $\left|\left[\mathbf{X}_{B_k}\right]_{pm}\right|$ is one, for $K+1 \le p \le 2K$.

2.1.2 BGOSTBC Encoder (with CP Insertions)

Note that in Fig. 2, $(K,M_t) = (4,3)$ and CP insertions following the encoding have been taken into account. The $m^{\underline{th}}$ encoder output path is followed by a CP insertion whose output is connected to the $m^{\underline{th}}$ transmit antenna, for all values of m. CP is inserted prior to the transmission of each symbol block for mitigating the channel distortion. From here on each time epoch lasts for N+L symbol periods. Since the constituent matrices simply serve as designating the space-time ordering for signal blocks, CP insertions directly apply to $\mathbf{s}_m^{(p)}$, $1 \le p \le 2K$, $1 \le m \le M_t$. Let M=N+L. Then the signal blocks collected at the output of the CP insertions over $2K$ time epochs and across M_t transmit antennas can be represented in matrix from as

$$\begin{aligned} \widehat{\mathbf{G}}_B &\triangleq \left(\mathbf{I}_{2K} \otimes \mathbf{I}_{\mathrm{CP}}\right)\mathbf{G}_B \\ &= \sum_{k=1}^{K}\left[\mathbf{X}_{A_k} \otimes \left(\mathbf{I}_{\mathrm{CP}}\mathbf{c}_k\right) + \mathbf{X}_{B_k} \otimes \left(\mathbf{I}_{\mathrm{CP}}\mathbf{d}_k^*\right)\right] \\ &= \sum_{k=1}^{K}\left[\mathbf{X}_{A_k} \otimes \widehat{\mathbf{c}_k} + \mathbf{X}_{B_k} \otimes \widehat{\mathbf{d}_k^*}\right] \in \mathbb{C}^{2\mathrm{MK}\times\mathrm{M_t}}, \end{aligned} \tag{2.3}$$

where

$$\widehat{\mathbf{c}_k} \triangleq \mathbf{I}_{\mathrm{CP}}\mathbf{c}_k \in \mathbb{C}^{\mathrm{M}};$$

$$\widehat{\mathbf{d}_k^*} \triangleq \mathbf{I}_{\mathrm{CP}}\mathbf{d}_k^* \in \mathbb{C}^{\mathrm{M}};$$

$$\mathbf{I}_{\mathrm{CP}} \triangleq \left[\begin{array}{cc} \mathbf{0}_{L\times(N-L)} & \mathbf{I}_L \\ \hline \multicolumn{2}{c}{\mathbf{I}_N} \end{array}\right] \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}, \text{ the CP insertion matrix.}$$

As a building block, next we review the transmission of a single signal block with CP through a frequency-selective fading channel.

2.2 Block Transmission and IBI-free Model upon CP Removal

Let the sequence $\widehat{s}_m(pM)$, $\widehat{s}_m(pM+1),\cdots,$ $\widehat{s}_m(pM+M-1)$ denotes the symbols transmitted from $m^{\underline{th}}$ antenna during $p^{\underline{th}}$ time epoch. Let $\widehat{\mathbf{s}}_m^{(p)} \triangleq \left[\widehat{s}_m(pM) \;\; \widehat{s}_m(pM+1) \;\; \cdots \;\; \widehat{s}_m(pM+M-1)\right]^T \in \mathbb{C}^{\mathrm{M}}$. By noticing $\widehat{\mathbf{s}}_m^{(p)} = \mathbf{I}_{\mathrm{CP}}\mathbf{s}_m^{(p)}$, we know that $\widehat{\mathbf{s}}_m^{(p)}$ takes $\widehat{\mathbf{c}}_k$ if $\left|\left[\mathbf{X}_{A_k}\right]_{pm}\right|$ is one, for $1 \le p \le K$. Similarly, $\widehat{\mathbf{s}}_m^{(p)}$ takes $\widehat{\mathbf{d}_k^*}$ if $\left|\left[\mathbf{X}_{B_k}\right]_{pm}\right|$ is one, for $K+1 \le \;\; p \le 2K$.

2.2.1 MIMO Frequency-selective Fading Channel and Noise Model

We consider a frequency-selective channel where the channel impulse response of L+1 taps between $m^{\underline{th}}$ transmit antenna and $j^{\underline{th}}$ receive antenna is defined as $\mathbf{h}_{jm} \triangleq \left[h_{jm}(0) \;\; h_{jm}(1) \;\; \cdots \;\; h_{jm}(L)\right]^T \in \mathbb{C}^{\mathrm{L+1}}$. Throughout the thesis, we assume independence between channel taps for all $\mathbf{h}_{jm}$, $1 \le j \le M_r$, $1 \le m \le M_t$. Assume that $\mathbf{h}_{jm}$ is circular symmetric Gaussian distributed, i.e., $\mathbf{h}_{jm} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I}_{L+1})$. Hence the real and imaginary parts of each of the entries of $\mathbf{h}_{jm}$ are i.i.d. zero-mean Gaussian with variance $0.5\sigma_h^2$ each, i.e., $\mathcal{N}(0, 0.5\sigma_h^2)$. Assume that the channel remains fixed during $2K$ time epochs. Let the sequence $\eta_j(pM)$, $\eta_j(pM+1),\cdots,$ $\eta_j(pM+M-1)$ denotes the additive noise samples, circular symmetric Gaussian distributed, at the $j^{\underline{th}}$ receiver during the $p^{\underline{th}}$ time epoch. Hence, the noise vector defined as $\boldsymbol{\eta}_j^{(p)} \triangleq \left[\eta_j(pM) \;\; \eta_j(pM+1) \;\; \cdots \;\; \eta_j(pM+M-1)\right]^T \in \mathbb{C}^{\mathrm{M}}$ assumes $\mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_M)$.

2.2.2 Block Transmission via MIMO Frequency-Selective Fading Channels

Let the sequence $v_j(pM),\ v_j(pM+1),\cdots,\ v_j(pM+M-1)$ denote the received symbols at the $j^{\underline{th}}$ receiver during the $p^{\underline{th}}$ time epoch. Define $\mathbf{v}_j^{(p)} = \left[v_j(pM)\ v_j(pM+1)\ \cdots\ v_j(pM+M-1)\right]^T \in \mathbb{C}^{\mathrm{M}}$. For $0 \le n \le M-1$, $1 \le j \le M_r$,

$$v_j(pM+n) \triangleq \sum_{m=1}^{M_t}\sum_{l=0}^{L} h_{jm}(l)\widehat{s}_m(pM+n-l) + \eta_j(pM+n).$$

We can write $\mathbf{v}_j^{(p)}$ as

$$\mathbf{v}_j^{(p)} = \sum_{m=1}^{M_t}\left[\mathbf{H}_{jm}^{tr}\widehat{\mathbf{s}}_m^{(p)} + \mathbf{H}_{jm}^{IBI}\widehat{\mathbf{s}}_m^{(p-1)}\right] + \boldsymbol{\eta}_j^{(p)},$$

where

$$\mathbf{H}_{jm}^{tr} \triangleq \begin{pmatrix} h_{jm}(0) & 0 & \cdots & \cdots & 0 & 0 \\ h_{jm}(1) & h_{jm}(0) & 0 & & \vdots & 0 \\ \vdots & h_{jm}(1) & \ddots & \ddots & & \vdots \\ h_{jm}(L) & \vdots & & & 0 & \vdots \\ \vdots & \ddots & & & h_{jm}(0) & 0 \\ 0 & 0 & h_{jm}(L) & \cdots & h_{jm}(1) & h_{jm}(0) \end{pmatrix};$$

$$\mathbf{H}_{jm}^{IBI} \triangleq \begin{pmatrix} 0 & \cdots & h_{jm}(L) & \cdots & h_{jm}(2) & h_{jm}(1) \\ 0 & \cdots & 0 & h_{jm}(L) & \vdots & h_{jm}(2) \\ \vdots & & 0 & 0 & \ddots & \vdots \\ 0 & & \vdots & 0 & 0 & h_{jm}(L) \\ \vdots & \vdots & & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

After CP removal, we have the IBI-free received signal model, $\mathbf{y}_j^{(p)} \triangleq \left[\mathbf{0}_{N\times L}\quad \mathbf{I}_N\right]\mathbf{v}_j^{(p)} = \left[y_j(pN)\quad y_j(pN+1)\quad \cdots\quad y_j(pN+N-1)\right]^T \in \mathbb{C}^{\mathrm{N}}$, where $y_j(i)$, $pN \le i \le pN+N-1$, are the received symbols at the $j^{\underline{th}}$ receiver during the $p^{\underline{th}}$ time epoch after CP removals.

$$\begin{aligned}\mathbf{y}_j^{(p)} &= \left[v_j(pM+L)\quad v_j(pM+L+1)\quad \cdots\quad v_j(pM+M-1)\right] \\ &= \sum_{m=1}^{M_t}\left(\left[\mathbf{0}_{N\times L}\quad \mathbf{I}_N\right]\mathbf{H}_{jm}^{tr}\widehat{\mathbf{s}}_m^{(p)} + \left[\mathbf{0}_{N\times L}\quad \mathbf{I}_N\right]\mathbf{H}_{jm}^{IBI}\widehat{\mathbf{s}}_m^{(p-1)}\right) + \left[\mathbf{0}_{N\times L}\quad \mathbf{I}_N\right]\boldsymbol{\eta}_j^{(p)}\end{aligned}$$

$$
= \sum_{m=1}^{M_t} \begin{bmatrix} h_{jm}(L) & \cdots & h_{jm}(0) & 0 & \cdots & & 0 & 0 \\ 0 & \ddots & \vdots & & 0 & & & \vdots \\ 0 & & h_{jm}(L) & & \ddots & \ddots & & \\ \vdots & & 0 & \ddots & & \ddots & & 0 \\ & & \vdots & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & h_{jm}(L) & \cdots & h_{jm}(0) \end{bmatrix} \left[\frac{\mathbf{0}_{L\times(N-L)} \quad \mathbf{I}_L}{\mathbf{I}_N} \right] \mathbf{s}_m^{(p)} + \mathbf{w}_j^{(p)}
$$

$$
= \sum_{m=1}^{M_t} \mathbf{H}_{jm} \mathbf{s}_m^{(p)} + \mathbf{w}_j^{(p)},
$$

where $\mathbf{w}_j^{(p)} \triangleq \begin{bmatrix} \mathbf{0}_{N\times L} & \mathbf{I}_N \end{bmatrix} \boldsymbol{\eta}_j^{(p)}$ and

$$
\mathbf{H}_{jm} \triangleq \begin{bmatrix} h_{jm}(0) & 0 & \cdots & 0 & 0 & h_{jm}(L) & \cdots & h_{jm}(1) \\ h_{jm}(1) & h_{jm}(0) & \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & & \ddots & h_{jm}(0) & 0 & & h_{jm}(L) \\ h_{jm}(L) & h_{jm}(L-1) & & & \vdots & h_{jm}(0) & 0 & \vdots \\ \vdots & \vdots & \ddots & h_{jm}(L) & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & h_{jm}(L) & h_{jm}(L-1) & \cdots & h_{jm}(0) \end{bmatrix}.
$$

Let $\overline{\mathbf{h}}_{jm} = \begin{bmatrix} \mathbf{h}_{jm}^T & \mathbf{0}_{1\times(N-L-1)} \end{bmatrix}^T \in \mathbb{C}^{\mathrm{N}}$. We note that $\mathbf{H}_{jm}$ is an N by N circulant matrix with its first column being $\overline{\mathbf{h}}_{jm}$. It follows that, after CP removal, the signal received at the $j^{\underline{th}}$ receive antenna on the $p^{\underline{th}}$ time epoch is IBI-free and given by, for $1 \le p \le 2K$,

$$
\mathbf{y}_j^{(p)} = \sum_{m=1}^{M_t} \mathbf{H}_{jm} \mathbf{s}_m^{(p)} + \mathbf{w}_j^{(p)}. \tag{2.4}
$$

Let us define the following terms which will be used through out the following chapters:

$$
\begin{cases} \mathbf{h}_j \triangleq \begin{bmatrix} \mathbf{h}_{j1}^T & \mathbf{h}_{j2}^T & \cdots & \mathbf{h}_{jM_t}^T \end{bmatrix}^T \in \mathbb{C}^{\mathrm{M_t(L+1)}} \\ \overline{\mathbf{h}}_j \triangleq \begin{bmatrix} \overline{\mathbf{h}}_{j1}^T & \overline{\mathbf{h}}_{j2}^T & \cdots & \overline{\mathbf{h}}_{jM_t}^T \end{bmatrix}^T \in \mathbb{C}^{\mathrm{M_tN}} \\ \mathbf{h}_{all} \triangleq \begin{bmatrix} \mathbf{h}_1^T & \mathbf{h}_2^T & \cdots & \mathbf{h}_{M_r}^T \end{bmatrix}^T \in \mathbb{C}^{\mathrm{M_rM_t(L+1)}} \\ \overline{\mathbf{h}}_{all} \triangleq \begin{bmatrix} \overline{\mathbf{h}}_1^T & \overline{\mathbf{h}}_2^T & \cdots & \overline{\mathbf{h}}_{M_r}^T \end{bmatrix}^T \in \mathbb{C}^{\mathrm{M_rM_tN}} \end{cases}. \tag{2.5}
$$

Chapter 3

Frequency-Domain Equalization

In this chapter, we propose a generalized FDE scheme based on block-level extension of generalized complex orthogonal designs (GOSTBC). The extended block-level OSTBC scheme with more than two transmit antenna is capable of achieving full transmit-receive diversity using FDE over frequency-selective channels. This shows an alternative to the block-level extension resorting to QO-STBC[2], which can indeed achieve perfect FDE in more-than-two transmit antenna scenario but at the cost of additional hardware complexity accounting for adders. Aside from this complexity drawback, its spatial diversity is halved.

3.1 Overall Frequency-Domain Received Signal Model

In the subsequent discussion, we will only derive the received signal model at the j^{th} receive antenna as all receivers have the same signal model except for different channel impulse responses. First recall that $\mathbf{s}_m^{(p)}$ in (2.4) takes $\mathbf{c}_k$ if $\left|\left[\mathbf{X}_{A_k}\right]_{pm}\right|$ is one, for $1 \le p \le K$ and takes $\mathbf{d}_k^*$ if $\left|\left[\mathbf{X}_{B_k}\right]_{pm}\right|$ is one, for $K+1 \le p \le 2K$. With the above observations, we have the following received signal model after collecting the IBI-free received signal blocks over $2K$ time epochs (Without loss of generality, we have assumed the transmission started from time epoch index 1 and collect the received blocks all the way up to index $2K$)

$$\begin{aligned}
\mathbf{y}_j^{'} &\triangleq \left[\left(\mathbf{y}_j^{(1)}\right)^T \quad \left(\mathbf{y}_j^{(2)}\right)^T \quad \cdots \quad \left(\mathbf{y}_j^{(K)}\right)^T \middle| \left(\mathbf{y}_j^{(K+1)}\right)^T \quad \cdots \quad \left(\mathbf{y}_j^{(2K)}\right)^T\right]^T \\
&= \sum_{m=1}^{M_t}\left(\mathbf{I}_{2K} \otimes \mathbf{H}_{jm}\right)\left[\left(\mathbf{s}_m^{(1)}\right)^T \quad \left(\mathbf{s}_m^{(2)}\right)^T \quad \cdots \quad \left(\mathbf{s}_m^{(K)}\right)^T \middle| \left(\mathbf{s}_m^{(K+1)}\right)^T \quad \cdots \quad \left(\mathbf{s}_m^{(2K)}\right)^T\right]^T + \mathbf{w}_j^{'} \\
&= \sum_{m=1}^{M_t}\left(\mathbf{I}_{2K} \otimes \mathbf{H}_{jm}\right)\mathbf{G}_B(:,m) + \mathbf{w}_j^{'}, \qquad (3.1)
\end{aligned}$$

where $\mathbf{w}_j^{'} = \left[\left(\mathbf{w}_j^{(1)}\right)^T \quad \left(\mathbf{w}_j^{(2)}\right)^T \quad \cdots \quad \left(\mathbf{w}_j^{(2K)}\right)^T\right]^T$ denotes the stacked white noises over $2K$

time epochs.

3.1.1 Space-Time Combining Using Signal DFT's

Define $\boldsymbol{\mathcal{Y}}_j^{(p)} \triangleq \boldsymbol{\mathcal{F}}\mathbf{y}_j^{(p)},\ 1 \le p \le 2K$, $\boldsymbol{\mathcal{Y}}_j^{'} \triangleq (\mathbf{I}_{2K} \otimes \boldsymbol{\mathcal{F}})\mathbf{y}_j^{'}$ and $\boldsymbol{\mathcal{W}}_j^{'} \triangleq (\mathbf{I}_{2K} \otimes \boldsymbol{\mathcal{F}})\mathbf{w}_j^{'}$. So performing block-wise DFT on $\mathbf{y}_j^{'}$ in (3.1) leads to

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}_j^{'} &= \left[\left(\boldsymbol{\mathcal{Y}}_j^{(1)}\right)^T \quad \left(\boldsymbol{\mathcal{Y}}_j^{(2)}\right)^T \quad \cdots \quad \left(\boldsymbol{\mathcal{Y}}_j^{(K)}\right)^T \middle| \left(\boldsymbol{\mathcal{Y}}_j^{(K+1)}\right)^T \quad \cdots \quad \left(\boldsymbol{\mathcal{Y}}_j^{(2K)}\right)^T\right]^T \\
&= \sum_{m=1}^{M_t} (\mathbf{I}_{2K} \otimes \boldsymbol{\mathcal{F}})(\mathbf{I}_{2K} \otimes \mathbf{H}_{jm})\mathbf{G}_B(:,m) + (\mathbf{I}_{2K} \otimes \boldsymbol{\mathcal{F}})\mathbf{w}_j^{'} \\
&= \sum_{m=1}^{M_t} \left(\mathbf{I}_{2K} \otimes \left(\boldsymbol{\mathcal{F}}\mathbf{H}_{jm}\right)\right)\mathbf{G}_B(:,m) + \boldsymbol{\mathcal{W}}_j^{'}.
\end{aligned} \tag{3.2}$$

See the tailored version for 2-fold diversity in [9],[26]. Since $\mathbf{H}_{jm}$ is circulant, we can decompose it by the DFT matrix as follows: $\mathbf{H}_{jm} = \boldsymbol{\mathcal{F}}^H \Lambda_{jm} \boldsymbol{\mathcal{F}},\ \Lambda_{jm} = Diag\left(\sqrt{N}\boldsymbol{\mathcal{F}}\overline{\mathbf{h}}_{jm}\right)$, for all values of j and m. Let $\underline{\mathbb{X}}_k \triangleq \boldsymbol{\mathcal{F}}\mathbf{c}_k$. Since $\mathbf{d}_k$ is obtained from $\mathbf{c}_k$ by conjugated time reversal, we have[24, pp. 123-124], $\boldsymbol{\mathcal{F}}\mathbf{d}_k^* = \underline{\mathbb{X}}_k^*,\ 1 \le k \le K$. Hence we can rewrite (3.2) as

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}_j^{'} &= \sum_{m=1}^{M_t} \left(\mathbf{I}_{2K} \otimes \left(\Lambda_{jm}\boldsymbol{\mathcal{F}}\right)\right)\mathbf{G}_B(:,m) + \boldsymbol{\mathcal{W}}_j^{'} \\
&= \sum_{k=1}^{K}\sum_{m=1}^{M_t} \left(\mathbf{I}_{2K} \otimes \left(\Lambda_{jm}\boldsymbol{\mathcal{F}}\right)\right)\left(\mathbf{X}_{A_k}(:,m) \otimes \mathbf{c}_k + \mathbf{X}_{B_k}(:,m) \otimes \mathbf{d}_k^*\right) + \boldsymbol{\mathcal{W}}_j^{'} \\
&= \sum_{k=1}^{K}\sum_{m=1}^{M_t} \left\{\mathbf{X}_{A_k}(:,m) \otimes \left(\Lambda_{jm}\boldsymbol{\mathcal{F}}\mathbf{c}_k\right) + \mathbf{X}_{B_k}(:,m) \otimes \left(\Lambda_{jm}\boldsymbol{\mathcal{F}}\mathbf{d}_k^*\right)\right\} + \boldsymbol{\mathcal{W}}_j^{'} \\
&= \sum_{k=1}^{K}\sum_{m=1}^{M_t} \left\{\mathbf{X}_{A_k}(:,m) \otimes \left(\Lambda_{jm}\underline{\mathbb{X}}_k\right) + \mathbf{X}_{B_k}(:,m) \otimes \left(\Lambda_{jm}\underline{\mathbb{X}}_k^*\right)\right\} + \boldsymbol{\mathcal{W}}_j^{'}.
\end{aligned}$$

Based on (2.2), we have

$$\boldsymbol{\mathcal{Y}}_j^{'} = \sum_{k=1}^{K}\sum_{m=1}^{M_t} \left\{\left(\frac{\boldsymbol{\mathcal{G}}_{A_k}(:,m) \otimes \left(\Lambda_{jm}\underline{\mathbb{X}}_k\right)}{\mathbf{0}_{NK\times 1}}\right) + \left(\frac{\mathbf{0}_{NK\times 1}}{\boldsymbol{\mathcal{G}}_{A_k}(:,m) \otimes \left(\Lambda_{jm}\underline{\mathbb{X}}_k^*\right)}\right)\right\} + \boldsymbol{\mathcal{W}}_j^{'}. \tag{3.3}$$

Define $\boldsymbol{\mathcal{Y}}_j \triangleq \left[\left(\boldsymbol{\mathcal{Y}}_j^{'}(0:KN-1)\right)^T \middle| \left(\boldsymbol{\mathcal{Y}}_j^{'*}(KN:2KN-1)\right)^T\right]^T \in \mathbb{C}^{2KN}$, then we have

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_j &= \sum_{k=1}^{K}\sum_{m=1}^{M_t}\left\{\left(\frac{\boldsymbol{\mathcal{G}}_{A_k}(:,m)\otimes\left(\Lambda_{jm}\underline{\mathbb{X}}_k\right)}{\mathbf{0}_{NK\times 1}}\right)+\left(\frac{\mathbf{0}_{NK\times 1}}{\boldsymbol{\mathcal{G}}_{A_k}(:,m)\otimes\left(\Lambda_{jm}^*\underline{\mathbb{X}}_k\right)}\right)\right\}+\boldsymbol{\mathcal{W}}_j \\ &= \sum_{k=1}^{K}\left\{\sum_{m=1}^{M_t}\left(\left(\frac{\boldsymbol{\mathcal{G}}_{A_k}(:,m)}{\mathbf{0}_{K\times 1}}\right)\otimes\Lambda_{jm}+\left(\frac{\mathbf{0}_{K\times 1}}{\boldsymbol{\mathcal{G}}_{A_k}(:,m)}\right)\otimes\Lambda_{jm}^*\right)\underline{\mathbb{X}}_k\right\}+\boldsymbol{\mathcal{W}}_j. \end{aligned} \tag{3.4}$$

Note that the definition of $\boldsymbol{\mathcal{Y}}_j$ follows from the reversed transmission mode by virtue of GOSTBC. Let $\mathbf{O}_{jk} \triangleq \sum_{m=1}^{M_t}\left[\left(\frac{\boldsymbol{\mathcal{G}}_{A_k}(:,m)}{\mathbf{0}_{K\times 1}}\right)\otimes\Lambda_{jm}+\left(\frac{\mathbf{0}_{K\times 1}}{\boldsymbol{\mathcal{G}}_{A_k}(:,m)}\right)\otimes\Lambda_{jm}^*\right]$, $\underline{\mathbb{X}} \triangleq \left[\underline{\mathbb{X}}_1^T \quad \underline{\mathbb{X}}_2^T \quad \cdots \quad \underline{\mathbb{X}}_K^T\right]^T$ and $\Lambda_j \triangleq \left[\mathbf{O}_{j1} \quad \mathbf{O}_{j2} \quad \cdots \quad \mathbf{O}_{jK}\right] \in \mathbb{C}^{2\mathrm{KN}\times\mathrm{KN}}$. We can then rewrite (3.4) more compactly as, for $1 \le j \le M_r$,

$$\boldsymbol{\mathcal{Y}}_j = \Lambda_j\underline{\mathbb{X}} + \boldsymbol{\mathcal{W}}_j. \tag{3.5}$$

Equation (3.5) represents the frequency-domain input-output relation between the j^{th} receive antenna and the transmit antennas, where Λ_j, $\underline{\mathbb{X}}$ and $\boldsymbol{\mathcal{W}}_j$ are embedded with knowledge of channel state information (CSI), of transmitted signal blocks and of channel noise at the j^{th} receive antenna, respectively. Let $\boldsymbol{\mathcal{Y}} = \left[\boldsymbol{\mathcal{Y}}_1^T \quad \boldsymbol{\mathcal{Y}}_2^T \quad \cdots \quad \boldsymbol{\mathcal{Y}}_{M_r}^T\right]^T \in \mathbb{C}^{2\mathrm{KNM_r}}$. By stacking the received signals in (3.5) across M_r receive antennas, we arrive at the following MIMO channel model in frequency domain

$$\boldsymbol{\mathcal{Y}} = \Lambda\underline{\mathbb{X}} + \boldsymbol{\mathcal{W}}, \tag{3.6}$$

where $\boldsymbol{\mathcal{W}} = \left[\boldsymbol{\mathcal{W}}_1^T \quad \boldsymbol{\mathcal{W}}_2^T \quad \cdots \quad \boldsymbol{\mathcal{W}}_{M_r}^T\right]^T$ and $\Lambda \triangleq \left[\Lambda_1^T \quad \Lambda_2^T \quad \cdots \quad \Lambda_{M_r}^T\right]^T \in \mathbb{C}^{2\mathrm{KNM_r}\times\mathrm{KN}}$ is the equivalent frequency-domain MIMO channel matrix for block transmission.

3.1.2 Frequency-Domain Equalization with Perfect CSI

Let $\widehat{\underline{\mathbb{X}}} = \Lambda^+\boldsymbol{\mathcal{Y}}$ and $\widehat{\boldsymbol{\mathcal{W}}} = \Lambda^+\boldsymbol{\mathcal{W}}$ and assume that $rank(\Lambda) = KN$. Hence $\Lambda^H\Lambda$ is invertible, and the Moore-Penrose pseudo-inverse of Λ can be written as $\Lambda^+ = \left(\Lambda^H\Lambda\right)^{-1}\Lambda^H \in \mathbb{C}^{\mathrm{KN}\times 2\mathrm{KNM_r}}$. So by pre-multiplying both sides of (3.6) by the linear equalizer Λ^+, we have

$$\underline{\widehat{\mathbb{X}}} = \underline{\mathbb{X}} + \widehat{\boldsymbol{\mathcal{W}}}, \tag{3.7}$$

where $\underline{\widehat{\mathbb{X}}}$ contains the decoded DFT's of the transmitted signal blocks. Therefore the transmitted blocks can be recovered through FDE. It is seen from (3.7) that by nonsingularity of $\Lambda^H\Lambda$, FDE can be achieved up to a zero-mean equalization error term $\widehat{\boldsymbol{\mathcal{W}}}$ when perfect CSI is available. The proposed BGOSTBC achieves block-level orthogonality in frequency domain, that is, $\Lambda^H\Lambda$ is a diagonal matrix and diagonally loaded with diversity weightings. More precisely, we have the following results.

Theorem 3.1 $\Lambda^H\Lambda = 2\left(\mathbf{I}_K \otimes \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)$.

(*Proof:* See Appendix A1.)

□

Remark 3.1 With perfect CSI, the FDE can be carried out *efficiently*. By efficiency we mean the involved computation for matrix inverse $\left(\Lambda^H\Lambda\right)^{-1}$ reduces to merely computing the reciprocals of each of the diagonal entries of $\Lambda^H\Lambda$. This can be seen from *Theorem 3.1* that $\Lambda^H\Lambda$ is diagonal.

Remark 3.2 The double summation of the elementary square block of $\Lambda^H\Lambda$ implicitly indicates that the *full M_rM_t-fold spatial diversity* is achieved through the proposed FDE scheme (See also chapter 5 for an explicit validation of full diversity via PEP analysis). Notice that had the encoder been realized with QO-STBC, the system would enjoy halved spatial diversity by carefully decoupling the received signals upon DFT transform[2]. The QO-STBC is more redundant than OSTBC in nature by inspection and relies on additional arithmetic operations for exploiting quasi-orthogonality at the receiving end, with approximately full transmission rate though.

Remark 3.3 Equation (3.3) is hereby regarded as a building block as we will exploit more of it towards the discussion of training-based channel estimation in chapter 5. Such discussion is as practical as essential to applying the OSTBC to the frequency-selective fading environment. Our proposed scheme can be adopted to generalize the channel estimation techniques in [3] to the frequency-selective fading scenario in a quite *straightforward* manner, in contrast to [16], with similar conclusion on optimal training design obtained in [3].

Remark 3.4 By adopting the Alamouti scheme for the construction of $\mathbf{G}_B$ with proper changes in dimension and scalar scaling factor makes the 2-fold diversity FDE (Alamouti-like) scheme a special case.

Remark 3.5 Notice that the FDE behind (3.7) is based on the premise that $rank(\Lambda) = KN$, and hence $rank\left(\Lambda^H\Lambda\right) = KN$. This in turn requires that $rank\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right) = N$ be satisfied as we can see from ***Remark 3.1***. Notice that each of the diagonal entries of $\Lambda_{jm}^H\Lambda_{jm}$ is nonnegative. To proceed further, let us define $\tilde{\bar{\mathbf{h}}}_{jm} \triangleq \boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_{jm} = \left[\tilde{\bar{h}}_{jm}(0) \quad \tilde{\bar{h}}_{jm}(0) \quad \cdots \quad \tilde{\bar{h}}_{jm}(N-1)\right]^T \in \mathbb{C}^{\mathrm{N}}$, where $\tilde{\bar{h}}_{jm}(\bar{n})$ is the $\bar{n}^{\underline{th}}$ entry of the DFT, $0 \le \bar{n} \le N-1$, $1 \le j \le M_r$ and $1 \le m \le M_t$.

Theorem 3.2 If there exist at least one pair of values $\left(j^*, m^*\right)$, $1 \le j^* \le M_r$, $1 \le m^* \le M_t$, such that $\left|\tilde{\bar{h}}_{j^*m^*}(\bar{n})\right|^2 > 0$, $0 \le \bar{n} \le N-1$, then $rank(\Lambda) = KN$ holds.

□

Let us denote the sufficient condition in ***Theorem 3.2*** as (c1.R). That is, if there is a subchannel whose transfer function is free from zeros at the frequencies $e^{j\frac{2\pi n}{N}}$, $0 \le n \le N-1$, then $\Lambda^H\Lambda$ is nonsingular. The condition is very weak and is generically satisfied. It is interesting to consider the following special case. Suppose $\left|\tilde{\bar{h}}_{j^*m^*}(n)\right|^2 > 0$, for $0 \le n \le N-1$ while $\tilde{\bar{\mathbf{h}}}_{jm} = \mathbf{0}_{N\times1}$, for $1 \le j \ne j^* \le M_r$, $1 \le m \ne m^* \le M_t$. Consequently equation (3.4) becomes

$$\boldsymbol{\mathcal{Y}}_j = \sum_{k=1}^{K}\left\{\mathbf{O}_{jk}\underline{\mathbb{X}}_k\right\} + \boldsymbol{\mathcal{W}}_j \text{ with } \mathbf{O}_{jk} \triangleq \left[\left(\frac{\mathcal{G}_{A_k}(:,m^*)}{\mathbf{0}_{K\times1}}\right)\otimes\Lambda_{j^*m^*} + \left(\frac{\mathbf{0}_{K\times1}}{\mathcal{G}_{A_k}(:,m^*)}\right)\otimes\Lambda_{j^*m^*}^H\right].$$

By stacking across all M_r receive antennas, we have Λ with the block-level orthogonality that $\Lambda^H\Lambda = 2\left(\mathbf{I}_K \otimes \Lambda_{j^*m^*}^H\Lambda_{j^*m^*}\right)$, which clearly lacks of spatial diversity, compared to (3.6) whose subchannels are all active. Intuitively, the decoded signal $\hat{\underline{\mathbb{X}}} = \Lambda^+\boldsymbol{\mathcal{Y}}$ becomes much more erroneous without the spatial diversity than with any. The above scenario arises when our MIMO channel model degenerates to a single-input-single-output (SISO) frequency-domain equivalent. So far, even though we consider a scenario which is more likely than (c1.R) to occur, such situation is far from common, not to mention that only one pair of values $\left(j^*, m^*\right)$, $1 \le j^* \le M_r$ and $1 \le m^* \le M_t$ such that $\left|\tilde{\bar{h}}_{j^*m^*}(\bar{n})\right|^2 > 0$, $0 \le \bar{n} \le N-1$. This is why we call (c1.R) a weak condition as it can be easily satisfied in practice.

Notice that the reason why we are able to recover the transmitted symbols even under the above mentioned harsh situations lies in that the space-time redundancy transmitted a signal block through every transmit antennas over different time epochs. Even though the MIMO channels degenerates to an SISO channel, all the transmitted blocks can still manage to the single receiver, provided that the reception lasts for $2K$ time epochs.

Chapter 4

Pairwise Error Probability Analysis

In this chapter, we will derive upper bounds for the average pairwise error probability (PEP) for the FDE-based scheme. The essential assumption under which the derivation is justified would be stated with numerical support. Our perspective reveals that at high SNR, the proposed system dose have the potential of delivering maximum possible spatial as well as multipath diversity.

4.1 PEP Analysis for Suboptimal Detection Problem

First, we derive an upper bound for the average PEP, assuming that the decoding consists of a linear equalization followed by an symbol-wise quantization into the signal constellation $\boldsymbol{\mathcal{A}}$. See also [11] and [17]. To see that, let us formulate the detection problem as follows. Based on (3.6), we have $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Omega}\bar{\mathbf{c}} + \boldsymbol{\mathcal{W}}$ with the matrix $\boldsymbol{\Omega} \triangleq \Lambda\left(\mathbf{I}_K \otimes \boldsymbol{\mathcal{F}}\right) \in \mathbb{C}^{2\mathrm{KNM_r} \times \mathrm{KN}}$ being responsible for yielding the frequency-domain output of the time-domain input signal blocks $\bar{\mathbf{c}} \triangleq \begin{bmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T & \cdots & \mathbf{c}_K^T \end{bmatrix}^T \in \mathbb{C}^{\mathrm{KN}}$. Hence we can address the ML-detection problem as follows:

$$\hat{\bar{\mathbf{c}}}_{ML} = \underset{\mathbf{c} \in \boldsymbol{\mathcal{A}}^{\mathrm{KN}}}{\operatorname{argmin}} \ \left\|\boldsymbol{\mathcal{Y}} - \boldsymbol{\Omega}\mathbf{c}\right\|_2, \tag{4.1}$$

where $\hat{\bar{\mathbf{c}}}_{ML}$ is the optimal decoded symbol block in ML sense. However, this ML exhaustive search yields infeasible computational cost of order $\left|\boldsymbol{\mathcal{A}}\right|^{KN}$ for practical values of K and N, where $\left|\boldsymbol{\mathcal{A}}\right|$ denotes size of the employed constellation. So we seek a suboptimal linear equalization approach. Note that $\boldsymbol{\mathcal{W}}$ is a white Gaussian vector since DFT serves as a unitary operation only. Let us define $\boldsymbol{\mathcal{Z}} \triangleq \Lambda^H \boldsymbol{\mathcal{Y}} \in \mathbb{C}^{\mathrm{KN}}$, $\widetilde{\boldsymbol{\mathcal{W}}} \triangleq \Lambda^H \boldsymbol{\mathcal{W}} \in \mathbb{C}^{\mathrm{KN}}$ and $\boldsymbol{\Theta} \triangleq \Lambda^H \boldsymbol{\Omega} \in \mathbb{C}^{\mathrm{KN} \times \mathrm{KN}}$. Then we can rewrite (3.6) as

$$\boldsymbol{\mathcal{Z}} = \boldsymbol{\Theta}\bar{\mathbf{c}} + \widetilde{\boldsymbol{\mathcal{W}}}. \tag{4.2}$$

We assume $rank(\Lambda) = KN$. Since $\mathbf{I}_K \otimes \boldsymbol{\mathcal{F}}$ is of full rank, $rank(\boldsymbol{\Theta}) = KN$ by definition of $\boldsymbol{\Theta}$. Hence the inverse $\boldsymbol{\Theta}^{-1}$ exists and can be used as a linear equalizer. Let $\widetilde{\boldsymbol{\mathcal{Z}}} = \boldsymbol{\Theta}^{-1}\boldsymbol{\mathcal{Z}}$ denote the output of the linear equalizer $\boldsymbol{\Theta}^{-1}$, then from (4.2) we readily arrive at

$$\widetilde{\boldsymbol{\mathcal{Z}}} = \bar{\mathbf{c}} + \boldsymbol{\Theta}^{-1}\widetilde{\boldsymbol{\mathcal{W}}}. \tag{4.3}$$

Notice that the diversity gain weighting has been normalized at the output of $\boldsymbol{\Theta}^{-1}$ and hence the detection problem can be formulated with respect to the employed constellation. Based on signal model of (4.3), we obtain an suboptimal detection problem:

$$\hat{\bar{\mathbf{c}}} = \underset{\mathbf{c} \in \boldsymbol{\mathcal{A}}^{\mathrm{KN}}}{\operatorname{argmin}} \left\| \widetilde{\boldsymbol{\mathcal{Z}}} - \mathbf{c} \right\|_2, \tag{4.4}$$

where $\hat{\bar{\mathbf{c}}}$ is an estimate of $\bar{\mathbf{c}}$ after equalization. Note that the so obtained estimate minimizing the metric in (4.4) is suboptimal since the underlying noise is no longer white and its covariance depends on the matrix $\boldsymbol{\Theta}^{-1}\Lambda^H$. We see that minimizing the metric in (4.4) amounts to a symbol-wise hard-decision into $\boldsymbol{\mathcal{A}}$ since $\left\| \widetilde{\boldsymbol{\mathcal{Z}}} - \mathbf{c} \right\|_2$ is the sum of KN nonnegative terms and each of which can be minimized with respect to an entry of $\mathbf{c}$, i.e., minimization on symbol-level. Therefore, using $\boldsymbol{\Theta}^{-1}$ as an equalizer and (4.4) as the detector amounts to having a linear equalization scheme followed by a hard decision on each entry of $\widetilde{\boldsymbol{\mathcal{Z}}}$ into the constellation $\boldsymbol{\mathcal{A}}$. It is noteworthy that the computational cost of (4.4) is linear in composite block length KN, which is computationally cheaper than that of (4.1).

The PEP analysis considers the probability that a symbol block $\bar{\mathbf{c}} \in \boldsymbol{\mathcal{A}}^{\mathrm{KN}}$ is transmitted while another $\tilde{\mathbf{c}}$ is detected in the minimum-distance perspective. Given the channel realization $\mathbf{h}_{all}$, and hence the matrix Λ, the conditional PEP is defined as

$$\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right] = \Pr\left[\left\|\tilde{\mathbf{c}}\text{-}\widetilde{\boldsymbol{\mathcal{Z}}}\right\|_2 < \left\|\bar{\mathbf{c}}\text{-}\widetilde{\boldsymbol{\mathcal{Z}}}\right\|_2 \middle| \Lambda\right]. \tag{4.5}$$

Note that proving that the average PEP has maximum diversity is equivalent to proving that the error rate performance exhibits maximum diversity, by virtue of the union bound[9]. For this purpose, let us define $\mathcal{D} \triangleq \|\bar{\mathbf{c}} - \tilde{\mathbf{c}}\|_2$ and $\hat{\mathbf{e}} \triangleq (\bar{\mathbf{c}} - \tilde{\mathbf{c}}) / \|\bar{\mathbf{c}} - \tilde{\mathbf{c}}\|_2$ to be the Euclidean distance between $\bar{\mathbf{c}}$ and $\tilde{\mathbf{c}}$, as well as the corresponding normalized difference, respectively. Let $\xi \triangleq \mathrm{Re}\left\{\left(\boldsymbol{\Theta}^{-1}\widetilde{\boldsymbol{\mathcal{W}}}\right)^H \widehat{\mathbf{e}}\right\}$ [18, pp. 508], then (4.5) becomes

$$\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right] = \Pr\left[\xi > \frac{\mathcal{D}}{2} \middle| \Lambda\right]. \tag{4.6}$$

As the variable ξ, when conditioned on channel realizations, is a zero-mean Gaussian random variable, the error probability in (4.6) is completely determined by the variance σ_ξ^2. To compute σ_ξ^2, first let us write $\boldsymbol{\mathcal{W}} = \left(\mathbf{I}_{2KM_r} \otimes \boldsymbol{\mathcal{F}}\right)\overline{\boldsymbol{w}}$ for notational purpose, where $\overline{\boldsymbol{w}}$ is related to the white noise $\mathbf{w}_j^{'}$ in (3.1) by $\overline{\boldsymbol{w}} = \left[\mathbf{w}_j^{'}(0:KN-1)^T \quad \widehat{\mathbf{w}}^T\right]^T$, $\widehat{\mathbf{w}} = \mathbf{w}_j^{'}(KN:2KN-1)^*_{\mathrm{mod}-N}$. Then we can factorize $\left(\boldsymbol{\Theta}^{-1}\widetilde{\boldsymbol{\mathcal{W}}}\right)^H \widehat{\mathbf{e}}$ as $\overline{\boldsymbol{w}}^H \hat{\mathbf{e}}$, where $\hat{\mathbf{e}} \triangleq \left(\boldsymbol{\Theta}^{-1}\Lambda^H\left(\mathbf{I}_{2KM_r} \otimes \boldsymbol{\mathcal{F}}\right)\right)^H \widehat{\mathbf{e}}$. The variance is computed as follows:

$$\begin{aligned}
\sigma_\xi^2 &= E\left\{\xi^2 \middle| \Lambda\right\} = E\left\{\mathrm{Re}\left\{\left(\boldsymbol{\Theta}^{-1}\widetilde{\boldsymbol{\mathcal{W}}}\right)^H \widehat{\mathbf{e}}\right\}^2 \middle| \Lambda\right\} \\
&= E\left\{\mathrm{Re}\left\{\overline{\boldsymbol{w}}^H \hat{\mathbf{e}}\right\}^2 \middle| \Lambda\right\} \\
&= E\left\{\left[\mathrm{Re}\{\overline{\boldsymbol{w}}\}^T \mathrm{Re}\{\hat{\mathbf{e}}\}\right]^2 \middle| \Lambda\right\} + \\
&\quad E\left\{\left[\mathrm{Im}\{\overline{\boldsymbol{w}}\}^T \mathrm{Im}\{\hat{\mathbf{e}}\}\right]^2 \middle| \Lambda\right\} + \\
&\quad 2E\left\{\left(\mathrm{Re}\{\overline{\boldsymbol{w}}\}^T \mathrm{Re}\{\hat{\mathbf{e}}\}\right)\left(\mathrm{Im}\{\overline{\boldsymbol{w}}\}^T \mathrm{Im}\{\hat{\mathbf{e}}\}\right) \middle| \Lambda\right\}.
\end{aligned}$$

Notice that since $\overline{\boldsymbol{w}}$ is white, the entries of $\overline{\boldsymbol{w}}$ are i.i.d. circular symmetric Gaussian, with real part and imaginary part of each of the entries being $\mathcal{N}(0, 0.5\sigma_w^2)$ distributed. Also notice that $\mathrm{Re}\{\overline{\boldsymbol{w}}\}^T \mathrm{Re}\{\hat{\mathbf{e}}\}$ and $\mathrm{Im}\{\overline{\boldsymbol{w}}\}^T \mathrm{Im}\{\hat{\mathbf{e}}\}$ are statistically independent of each other by virtue of the circular nature of $\overline{\boldsymbol{w}}$. For notational purpose, let $[\mathbf{y}]_i$ denotes the $i^{\underline{th}}$ entry of a vector $\mathbf{y}$. So we can rewrite the above equation as

$$\begin{aligned}
\sigma_\xi^2 &= E\left\{\left[\mathrm{Re}\{\overline{\boldsymbol{w}}\}^T \mathrm{Re}\{\hat{\mathbf{e}}\}\right]^2 \middle| \Lambda\right\} + E\left\{\left[\mathrm{Im}\{\overline{\boldsymbol{w}}\}^T \mathrm{Im}\{\hat{\mathbf{e}}\}\right]^2 \middle| \Lambda\right\} \\
&= E\left\{\left[\sum_{i=1}^{2KNM_r}\left[\mathrm{Re}\{\overline{\boldsymbol{w}}\}\right]_i\left[\mathrm{Re}\{\hat{\mathbf{e}}\}\right]_i\right]^2 + \left[\sum_{i=1}^{2KNM_r}\left[\mathrm{Im}\{\overline{\boldsymbol{w}}\}\right]_i\left[\mathrm{Im}\{\hat{\mathbf{e}}\}\right]_i\right]^2 \middle| \Lambda\right\} \\
&= \sum_{i=1}^{2KNM_r}\left[\mathrm{Re}\{\hat{\mathbf{e}}\}\right]_i^2 E\left\{\left[\mathrm{Re}\{\overline{\boldsymbol{w}}\}\right]_i^2 \middle| \Lambda\right\} + \sum_{i=1}^{2KNM_r}\left[\mathrm{Im}\{\hat{\mathbf{e}}\}\right]_i^2 E\left\{\left[\mathrm{Im}\{\overline{\boldsymbol{w}}\}\right]_i^2 \middle| \Lambda\right\}
\end{aligned}$$

$$= \left(\frac{1}{2}\sigma_w^2\right)\left\{\sum_{i=1}^{2KNM_r}\left[\mathrm{Re}\{\hat{\mathbf{e}}\}\right]_i^2 + \sum_{i=1}^{2KNM_r}\left[\mathrm{Im}\{\hat{\mathbf{e}}\}\right]_i^2\right\}$$

$$= \left(\frac{1}{2}\sigma_w^2\right)\|\hat{\mathbf{e}}\|_2^2$$

$$= \frac{1}{2}\sigma_w^2\left\|\left[\mathbf{\Theta}^{-1}\Lambda^H\left(\mathbf{I}_{2KM_r}\otimes\boldsymbol{\mathcal{F}}\right)\right]^H\hat{\mathbf{e}}\right\|_2^2$$

$$= \frac{1}{2}\sigma_w^2\left\|\left[\left(\mathbf{I}_K\otimes\boldsymbol{\mathcal{F}}^H\right)\Lambda^+\left(\mathbf{I}_{2KM_r}\otimes\boldsymbol{\mathcal{F}}\right)\right]^H\hat{\mathbf{e}}\right\|_2^2,$$

where we have used independence between each of the entries of $\mathrm{Re}\{\overline{\boldsymbol{w}}\}$ due to $\overline{\boldsymbol{w}}$ being white and zero-mean. Notice that $\mathbf{\Theta}^{-1}\Lambda^H = \left(\mathbf{I}_K\otimes\boldsymbol{\mathcal{F}}^H\right)\Lambda^+$. Similar reasoning applies to $\mathrm{Im}\{\overline{\boldsymbol{w}}\}$. We know that the PEP in (4.6) can be expressed in terms of the Q-function, $Q(\alpha)\triangleq\int_\alpha^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$, as

$$\Pr\left[\bar{\mathbf{c}}\to\tilde{\mathbf{c}}\,|\,\mathbf{h}_{all}\right] = Q\left(\frac{\mathcal{D}}{\sqrt{4\sigma_\xi^2}}\right)$$

$$= Q\left(\frac{\mathcal{D}}{\sqrt{2\sigma_w^2\left\|\left[\left(\mathbf{I}_K\otimes\boldsymbol{\mathcal{F}}^H\right)\Lambda^+\left(\mathbf{I}_{2KM_r}\otimes\boldsymbol{\mathcal{F}}\right)\right]^H\hat{\mathbf{e}}\right\|_2^2}}\right)$$

$$\le Q\left(\frac{\mathcal{D}}{\sqrt{2\sigma_w^2\left\|\left(\mathbf{I}_K\otimes\boldsymbol{\mathcal{F}}^H\right)\Lambda^+\left(\mathbf{I}_{2KM_r}\otimes\boldsymbol{\mathcal{F}}\right)\right\|_F^2}}\right),$$

where we have invoked the inequality that $\|\mathbf{A}\hat{\mathbf{e}}\|_2 \le \|\mathbf{A}\|_F\|\hat{\mathbf{e}}\|_2 = \|\mathbf{A}\|_F$ for any matrix $\mathbf{A}$ and a unit vector $\hat{\mathbf{e}}$. Since multiplication by unitary matrix dose not change Frobenius norm, we arrive at

$$\Pr\left[\bar{\mathbf{c}}\to\tilde{\mathbf{c}}\,|\,\mathbf{h}_{all}\right] \le Q\left(\frac{\mathcal{D}}{\sqrt{2\sigma_w^2\left\|\Lambda^+\right\|_F^2}}\right).$$

Then by using $Q(\alpha)\le\frac{1}{2}\exp(-\frac{\alpha^2}{2})$, we have the PEP of interest upper bounded as

$$\Pr\left[\bar{\mathbf{c}}\to\tilde{\mathbf{c}}\,|\,\mathbf{h}_{all}\right] \le \frac{1}{2}\exp\left(-\frac{\mathcal{D}^2}{4\sigma_w^2\left\|\Lambda^+\right\|_F^2}\right). \tag{4.7}$$

Let us define the following terms:

$$
\begin{cases}
\overline{\mathbf{A}}_{jk}^{(m)} \triangleq \left(\begin{array}{c} \boldsymbol{\mathcal{G}}_{A_k}(:,m) \\ \hline \mathbf{0}_{K\times 1} \end{array}\right)^T \otimes \mathbf{I}_N; \\
\overline{\overline{\mathbf{A}}}_{jk}^{(m)} \triangleq \left(\begin{array}{c} \mathbf{0}_{K\times 1} \\ \hline \boldsymbol{\mathcal{G}}_{A_k}(:,m) \end{array}\right)^T \otimes \mathbf{I}_N; \\
\overline{\mathbf{A}}_{jk} \triangleq \begin{bmatrix} \overline{\mathbf{A}}_{jk}^{(1)} & \overline{\mathbf{A}}_{jk}^{(2)} & \cdots & \overline{\mathbf{A}}_{jk}^{(M_t)} \end{bmatrix}; \\
\overline{\overline{\mathbf{A}}}_{jk} \triangleq \begin{bmatrix} \overline{\overline{\mathbf{A}}}_{jk}^{(1)} & \overline{\overline{\mathbf{A}}}_{jk}^{(2)} & \cdots & \overline{\overline{\mathbf{A}}}_{jk}^{(M_t)} \end{bmatrix}; \\
\overline{\boldsymbol{\Upsilon}}_j \triangleq \begin{bmatrix} \mathbf{I}_{2K} \otimes \Lambda_{j1}^H \left(\Lambda^H \Lambda\right)^{-1} & \mathbf{I}_{2K} \otimes \Lambda_{j2}^H \left(\Lambda^H \Lambda\right)^{-1} & \cdots & \mathbf{I}_{2K} \otimes \Lambda_{jM_t}^H \left(\Lambda^H \Lambda\right)^{-1} \end{bmatrix}^T; \\
\overline{\overline{\boldsymbol{\Upsilon}}}_j \triangleq \begin{bmatrix} \mathbf{I}_{2K} \otimes \Lambda_{j1} \left(\Lambda^H \Lambda\right)^{-1} & \mathbf{I}_{2K} \otimes \Lambda_{j2} \left(\Lambda^H \Lambda\right)^{-1} & \cdots & \mathbf{I}_{2K} \otimes \Lambda_{jM_t} \left(\Lambda^H \Lambda\right)^{-1} \end{bmatrix}^T.
\end{cases}
\tag{4.8}
$$

From (4.8), $\Lambda^+ \triangleq \left(\Lambda^H \Lambda\right)^{-1} \Lambda^H$ can be rewritten as

$$
\Lambda^+ = \begin{bmatrix} \overline{\mathbf{A}} & \overline{\overline{\mathbf{A}}} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{\Upsilon}} \\ \overline{\overline{\boldsymbol{\Upsilon}}} \end{bmatrix}, \tag{4.9}
$$

where

$$
\begin{cases}
\overline{\mathbf{A}} \triangleq \begin{bmatrix} \overline{\mathbf{A}}_{11} & \cdots & \overline{\mathbf{A}}_{M_r 1} \\ \vdots & \ddots & \vdots \\ \overline{\mathbf{A}}_{1K} & \cdots & \overline{\mathbf{A}}_{M_r M_t} \end{bmatrix}; \; \overline{\overline{\mathbf{A}}} \triangleq \begin{bmatrix} \overline{\overline{\mathbf{A}}}_{11} & \cdots & \overline{\overline{\mathbf{A}}}_{M_r 1} \\ \vdots & \ddots & \vdots \\ \overline{\overline{\mathbf{A}}}_{1K} & \cdots & \overline{\overline{\mathbf{A}}}_{M_r M_t} \end{bmatrix}; \\
\overline{\boldsymbol{\Upsilon}} \triangleq \begin{bmatrix} \overline{\boldsymbol{\Upsilon}}_1 & & \\ & \ddots & \\ & & \overline{\boldsymbol{\Upsilon}}_{M_r} \end{bmatrix}; \quad \overline{\overline{\boldsymbol{\Upsilon}}} \triangleq \begin{bmatrix} \overline{\overline{\boldsymbol{\Upsilon}}}_1 & & \\ & \ddots & \\ & & \overline{\overline{\boldsymbol{\Upsilon}}}_{M_r} \end{bmatrix}.
\end{cases}
$$

Hence, $\left\|\Lambda^+\right\|_F^2 \le \left\|\begin{bmatrix} \overline{\mathbf{A}} & \overline{\overline{\mathbf{A}}} \end{bmatrix}\right\|_F^2 \left(\left\|\overline{\boldsymbol{\Upsilon}}\right\|_F^2 + \left\|\overline{\overline{\boldsymbol{\Upsilon}}}\right\|_F^2\right)$. Notice that $\begin{bmatrix} \overline{\mathbf{A}} & \overline{\overline{\mathbf{A}}} \end{bmatrix}$ is solely determined by the encoder structure, and by (4.8) it can be easily verified that $\left\|\overline{\boldsymbol{\Upsilon}}\right\|_F^2 = \left\|\overline{\overline{\boldsymbol{\Upsilon}}}\right\|_F^2$ and

$$
\left\|\overline{\boldsymbol{\Upsilon}}\right\|_F^2 = 2K \cdot \sum_{j=1}^{M_r} \sum_{m=1}^{M_t} \left\| \left(\sum_{j=1}^{M_r} \sum_{m=1}^{M_t} \Lambda_{jm}^H \Lambda_{jm} \right)^{-1} \Lambda_{jm}^H \right\|_F^2. \tag{4.10}
$$

With the definition of $\tilde{\bar{\mathbf{h}}}_{jm}$ in ***Remark 3.5***, (4.10) can be rewritten as

$$\begin{aligned}\left\|\overline{\boldsymbol{\Upsilon}}\right\|_F^2 &= 2K\cdot\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left(\sum_{n=0}^{N-1}\frac{\left|\tilde{\bar{h}}_{jm}(n)\right|^2}{N\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\bar{h}}_{jm}(n)\right|^2\right)^2}\right)\\ &= \frac{2K}{N}\cdot\sum_{n=0}^{N-1}\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\bar{h}}_{jm}(n)\right|^2\right)^{-1}.\end{aligned}\tag{4.11}$$

Now, let us define an N by N matrix $\mathbf{D}_{ps}$ as

$$\mathbf{D}_{ps}=\begin{bmatrix}\sqrt{\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\bar{h}}_{jm}(0)\right|^2} & & & \\ & \sqrt{\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\bar{h}}_{jm}(1)\right|^2} & & \\ & & \ddots & \\ & & & \sqrt{\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\bar{h}}_{jm}(N-1)\right|^2}\end{bmatrix}.\tag{4.12}$$

We assume $rank(\Lambda)=KN$, or equivalently $rank\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)=N$. Hence, we see that $\mathbf{D}_{ps}$ is invertible. Hence, by defining $C_E\triangleq\left\|\left[\overline{\mathbf{A}}\quad\overline{\overline{\mathbf{A}}}\right]\right\|_F^2$ and from (4.9) up to (4.12) we arrive at

$$\left\|\Lambda^+\right\|_F^2\le C_E\cdot\frac{4K}{N}\cdot\left\|\mathbf{D}_{ps}^{-1}\right\|_F^2.\tag{4.13}$$

Before we proceed further, let us make one more assumption deduced from the rank premise: condition number of $\mathbf{D}_{ps}$, denoted as $\mathcal{K}\left(\mathbf{D}_{ps}\right)$, is upper bounded by a finite real number $\mathcal{K}_u\in\mathbb{R}$ for all possible channel realizations, i.e.,

$$\mathcal{K}\left(\mathbf{D}_{ps}\right)\le\mathcal{K}_u \text{ for all } \mathbf{D}_{ps}.\tag{4.14}$$

Notice that by orthogonality,

$$\mathcal{K}(\Lambda)=\sqrt{\frac{\lambda_{\max}\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)}{\lambda_{\min}\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)}}=\sqrt{\frac{\max\left\{Diag\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)\right\}}{\min\left\{Diag\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\Lambda_{jm}^H\Lambda_{jm}\right)\right\}}},$$

where $\lambda_{\max}(\Lambda)$ and $\lambda_{\min}(\Lambda)$ denote the maximal and minimal eigenvalues of matrix Λ respectively. Hence $\mathcal{K}(\mathbf{D}_{ps})=\mathcal{K}(\Lambda)$. By regarding the simulation results from Fig. 3-1. to Fig. 3-5, we see that

i) For all possible system configurations (N,M_r,M_t,L,K), the probability of singular occurrence $rank(\Lambda)<KN$ is very low in practice. As one can perform as many simulations as possible to find the above highly plausible, our rank premise is reasonable. Hence $\mathcal{K}(\mathbf{D}_{ps})<\infty$.

ii) Given known channel order L and desired transmission rate, selecting high values for M_t and M_r can effectively suppress $\mathcal{K}(\mathbf{D}_{ps})$.

iii) It agrees with the intuition that as N increases, the probability of $\mathcal{K}(\mathbf{D}_{ps})\rightarrow\infty$ drops. This is because each of the diagonal entries, which are correlated, of $\mathbf{D}_{ps}$ is sum of nonnegative terms.

From the above three observations we *deduce* that $\mathcal{K}(\mathbf{D}_{ps})$ can be universally upper bounded by a finite real number in practice.

On the other hand, suppose $\mathbf{B}\in\mathbb{R}^{n\times n}$ is a nonsingular diagonal matrix, then for any full rank $\mathbf{A}\in\mathbb{R}^{n\times m}$, $m<n$, such that $\|\mathbf{A}\|_F^2\le\frac{m}{n(1+n\mathcal{K}(\mathbf{B}))}$, $\|\mathbf{B}^{-1}\|_F^2\le\|(\mathbf{B}\mathbf{A})^+\|_F^2$.

Proof:

$$\|\mathbf{B}^{-1}\|_F^2=\frac{\sum_{i=0}^{n-1}[\mathbf{B}]_{ii}^{-2}\cdot\sum_{i=0}^{n-1}[\mathbf{B}]_{ii}^2}{\|\mathbf{B}\|_F^2}\le\frac{n\left(1+n\mathcal{K}(\mathbf{B})\right)}{\|\mathbf{B}\|_F^2}\le\frac{m}{\|\mathbf{B}\|_F^2\|\mathbf{A}\|_F^2}\le\|(\mathbf{B}\mathbf{A})^+\|_F^2. \quad (4.15)$$

□

Therefore, by assuming (4.14) and along with (4.15), there exist infinitely many full rank $\{\mathbf{\Psi}_i\}_{i=1}^{\infty}\in\mathbb{R}^{N\times(N-L)}$ satisfying $\|\mathbf{\Psi}_i\|_F^2\le\frac{m}{n(1+n\mathcal{K}_u)}$ such that $\|\mathbf{D}_{ps}^{-1}\|_F^2\le\|(\mathbf{D}_{ps}\mathbf{\Psi}_i)^+\|_F^2$ for all channel realizations. Now let us find such a matrix $\mathbf{\Psi}$ among these possible candidates so that any (N-L) of the rows of $\mathbf{\Psi}$ are linearly independent. Note that this additional constraint can be very easily satisfied and hence does not in any way contradict (4.14). Then, (4.13) can be written as

$$\left\|\Lambda^{+}\right\|_F^2 \le C_E \cdot \frac{4K}{N} \cdot \left\|\left(\mathbf{D}_{ps}\mathbf{\Psi}\right)^{+}\right\|_F^2. \tag{4.16}$$

Let $\beta \triangleq \{n_0, n_1, \cdots, n_{L-1}\}$ be the set of indexes corresponding to the smallest L composite frequency responses such that

$$\begin{cases} \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_p\right)\right|^2 \le \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_q\right)\right|^2 \text{ for } n_p \in \beta \text{ and } n_q \not\in \beta \\ \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_0\right)\right|^2 \le \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_1\right)\right|^2 \le \cdots \le \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_{L-1}\right)\right|^2 \end{cases}, \tag{4.17}$$

where $0 \le n_i \le N-1$ and $0 \le i \le L-1$. Define $\mathbf{D}_\beta$ as an N by N diagonal matrix whose entries are given by

$$\begin{cases} \left[\mathbf{D}_\beta\right]_{nn} = \dfrac{1}{\sqrt{\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}(n)\right|^2}}, \text{ if } n \not\in \beta \\ \left[\mathbf{D}_\beta\right]_{nn} = 0 \qquad\qquad\qquad\qquad , \text{ if } n \in \beta \end{cases}. \tag{4.18}$$

Then we construct an (N-L) by N matrix $\mathbf{\Psi}_\beta$ by the following two-fold step:

i) Remove all the rows, whose row indexes belong to set β, of $\mathbf{\Psi}$ to obtain an (N-L) by (N-L) matrix $\overline{\mathbf{\Psi}}$.
ii) Insert L zero columns to $\overline{\mathbf{\Psi}}^{-1}$ so that after the insertion, the inserted columns have indexes belong to set β. Let $\mathbf{\Psi}_\beta$ denote the so obtained matrix. Note that the constraint on linear independence of any (N-L) of rows of $\mathbf{\Psi}$ ensures existence of $\overline{\mathbf{\Psi}}^{-1}$.

From (4.17) up to the definition of $\mathbf{\Psi}_\beta$, it can be verified that

$$\left(\mathbf{\Psi}_{\beta}\mathbf{D}_{\beta}\right)\left(\mathbf{D}_{ps}\mathbf{\Psi}\right)=\mathbf{I}_{N-L}\,. \tag{4.19}$$

By the basic property of the pseudo-inverse we know that $\left\|\left(\mathbf{D}_{ps}\mathbf{\Psi}\right)^{+}\right\|_F \le \left\|\mathbf{\Psi}_{\beta}\mathbf{D}_{\beta}\right\|_F$ due to (4.19)[27, pp. 257]. Hence, (4.16) becomes

$$\left\|\Lambda^{+}\right\|_F^2 \le C_E\cdot\frac{4K}{N}\cdot\left\|\mathbf{\Psi}_{\beta}\mathbf{D}_{\beta}\right\|_F^2 \le C_E\cdot\frac{4K}{N}\cdot\left\|\mathbf{\Psi}_{\beta}\right\|_F^2\left\|\mathbf{D}_{\beta}\right\|_F^2. \tag{4.20}$$

Let $C_1 \triangleq \max_{\beta}\left\|\mathbf{\Psi}_{\beta}\right\|_F^2$, where the maximization is taken over all subsets of $\{0,1,\cdots,N-1\}$. Now, let us define $n_L=\arg\min_{n\notin\beta}\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}(n)\right|^2$. Then from (4.18) and (4.20), we have

$$\left\|\Lambda^{+}\right\|_F^2 \le C_E C_1\cdot\frac{4K}{N}\cdot(N-L)\left(\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_L\right)\right|^2\right)^{-1}. \tag{4.21}$$

Define $\underline{\mathbf{h}} \triangleq \left[\underline{\mathbf{h}}_{11}^T \;\; \cdots \;\; \underline{\mathbf{h}}_{M_rM_t}^T\right]^T$, where $\underline{\mathbf{h}}_{jm} \triangleq \left[\tilde{\tilde{h}}_{jm}\left(n_0\right) \;\; \tilde{\tilde{h}}_{jm}\left(n_1\right) \;\; \cdots \;\; \tilde{\tilde{h}}_{jm}\left(n_L\right)\right]^T$. From (4.17) and the definition of n_L, we see that

$$\left\|\underline{\mathbf{h}}\right\|_2^2 \le (L+1)\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_L\right)\right|^2. \tag{4.22}$$

Note that $\underline{\mathbf{h}}=\mathbf{V}\mathbf{h}_{all}$, where $\mathbf{V}$ is a $M_rM_t(L+1)$ by $M_rM_t(L+1)$ Vandermonde matrix with nonzero smallest singular value, $\sqrt{\lambda_{\min}\left(\mathbf{V}^H\mathbf{V}\right)}$. Since $\lambda_{\min}\left(\mathbf{V}^H\mathbf{V}\right)\cdot\left\|\mathbf{h}_{all}\right\|_2^2 \le \left\|\underline{\mathbf{h}}\right\|_2^2$, along with (4.22) we have

$$\left\|\mathbf{h}_{all}\right\|_2^2 \le \frac{(L+1)}{\lambda_{\min}\left(\mathbf{V}^H\mathbf{V}\right)}\sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\left|\tilde{\tilde{h}}_{jm}\left(n_L\right)\right|^2. \tag{4.23}$$

Hence, from (4.23), we can rewrite (4.21) as

$$\left\|\Lambda^{+}\right\|_F^{-2} \ge \frac{N\cdot\lambda_{\min}\left(\mathbf{V}^H\mathbf{V}\right)}{4K(N-L)(L+1)C_EC_1}\left\|\mathbf{h}_{all}\right\|_2^2. \tag{4.24}$$

Let $\kappa = \frac{N \cdot \lambda_{\min}\left(\mathbf{V}^H \mathbf{V}\right) \cdot \mathcal{D}^2}{16K(N-L)(L+1)C_E C_1}$ and SNR= $1/\sigma_w^2$. Then we see that the upper bound in (4.7) can be expressed as

$$\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right] \le \frac{1}{2} \exp\left(-\kappa \times \text{SNR} \left\|\mathbf{h}_{all}\right\|_2^2\right). \tag{4.25}$$

Let $\overline{m} = M_r M_t (L+1)$ and write $\mathbf{h}_{all} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{\overline{m}} \end{bmatrix}^T$. We assume that all the MIMO multipaths, $\mathbf{h}_{jm}$, are statistically independent of each other. The distribution of the channel vector $\mathbf{h}_{all}$ can then be expressed as a multi-dimensional Gaussian pdf[18, pp. 502],

$$\begin{aligned} f_{\mathbf{h}_{all}}\left(\mathbf{h}_{all}\right) &= \frac{1}{(\pi)^{\mathrm{M_r M_t (L+1)}} \left(\sigma_h^2\right)^{\mathrm{M_r M_t (L+1)}}} \exp\left(-\frac{1}{\sigma_h^2} \left\|\mathbf{h}_{all}\right\|_2^2\right) \\ &= \prod_{i=1}^{\overline{m}} \frac{1}{\pi \sigma_h^2} \exp\left(-\frac{\left|\alpha_i\right|^2}{\sigma_h^2}\right). \end{aligned} \tag{4.26}$$

Now we can average the upper bound with respect to the channel pdf:

$$\begin{aligned} E_{\bar{\mathbf{h}}}\left\{\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right]\right\} &\le \frac{1}{2} \int_{-\infty}^{\infty} \exp\left(-\kappa \times \text{SNR} \left\|\mathbf{h}_{all}\right\|_2^2\right) f_{\mathbf{h}_{all}}\left(\mathbf{h}_{all}\right) d\mathbf{h}_{all} \\ &= \frac{1}{2} \prod_{i=1}^{\overline{m}} \int_{-\infty}^{\infty} \frac{1}{\pi \sigma_h^2} \exp\left(-\left(\frac{1}{\sigma_h^2} + \kappa \times \text{SNR}\right) \left|\alpha_i\right|^2\right) d\alpha_i. \end{aligned}$$

Note that the integrand is to be integrated with respect to a *complex* variable, α_i, which is essentially a circular symmetric Gaussian random variable. Let $q_i = \text{Re}\left\{\alpha_i\right\}$, and $\bar{q}_i = \text{Im}\left\{\alpha_i\right\}$. Therefore by symmetry between the real part and imaginary part of α_i, we can further factorize the integration as

$$\begin{aligned} E_{\bar{\mathbf{h}}}\left\{\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right]\right\} &\le \frac{\prod_{i=1}^{\overline{m}} \left[\int_{-\infty}^{\infty} e^{-\left(\sigma_h^{-2} + \kappa \times \text{SNR}\right) p_i^2} dp_i\right] \prod_{i=1}^{\overline{m}} \left[\int_{-\infty}^{\infty} e^{-\left(\sigma_h^{-2} + \kappa \times \text{SNR}\right) q_i^2} dq_i\right]}{2\left(\pi \sigma_h^2\right)^{\overline{m}}} \\ &= \frac{\left\{\prod_{i=1}^{\overline{m}} \left[\sqrt{\frac{\pi}{\sigma_h^{-2} + \kappa \times \text{SNR}}}\right]\right\}^2}{2\left(\pi \sigma_h^2\right)^{\overline{m}}} \end{aligned}$$

$$= \frac{1}{2}\left(\sigma_h^2\right)^{-M_r M_t (L+1)} \prod_{i=1}^{M_r M_t (L+1)} \left(\sigma_h^{-2} + \kappa \times \mathrm{SNR}\right)^{-1}.$$

Hence, at high SNR region, we can expect the PEP be bounded by

$$E_{\bar{\mathbf{h}}}\left\{\Pr\left[\bar{\mathbf{c}} \to \tilde{\mathbf{c}} \middle| \mathbf{h}_{all}\right]\right\} \leq \frac{1}{2}\left(\sigma_h^2\right)^{-M_r M_t (L+1)} (\kappa \times \mathrm{SNR})^{-M_r M_t (L+1)}. \tag{4.27}$$

It is seen from (4.27) that at high SNR, full spatial (transmit-receive) diversity gain of order $M_r M_t$ and full multipath diversity of order $L+1$ are achieved. Note that the above results counterbalance the deduction in [26] that the "CP-only" scheme with extended GOSTBC cannot achieve full multipath diversity.

It is noteworthy that the derivation from (4.7) to (4.25) really relies on the structure of GOSTBC. From the perspective in [17], the particular relation $\left\|\Lambda^{+}\right\|_F^{-2} \geq C\left\|\mathbf{h}_{all}\right\|_2^2$ for some real constant C in (4.24) requires certain precoder design, whose purpose is to compensate the detrimental effect due to channel zeros/nulls. However, as we see from the above derivation, the precoding mechanism *could* be not necessary. Note that $\mathbf{\Psi}$ can be regarded as a *virtual* precoder introduced for the purpose of deriving PEP.

Remark 4.1 By noting the slope change at relatively higher SNR region within Fig. 5, we see that the possibility of delivering full multipath diversity under the proposed scheme is evidently noticeable even with small numbers of antennas, i.e., as the number of taps (L+1) increases by 1.5 times, so does the slope of error probability above 4 dB. This in turn justifies the deduction behind (4.16) upon which our PEP analysis is built. For simulations in Fig. 5, we set N=250 and QPSK for 1000 consecutive transmission during which the channel remains fixed. The channel taps of each of them assume power delay profiles whose sum are normalized to 2 for a fair comparison. Other details for simulations are stated in Chapter 6.

Chapter 5

Training-Based Channel Estimation

In this chapter, the training-based channel estimation by M. Biguesh[3] will be generalized to frequency-selective fading scenario by exploiting the structure in (3.3). Optimality criteria for the frequency-domain channel estimation in LS sense and under a given power constraint will be derived. Also, conditions under which the training blocks are designed so as to apply a linear MMSE approach for time-domain channel estimation are derived. First, let us denote the training matrix as $\Pi_t = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_K \end{bmatrix}$, where $\mathbf{t}_k \in \mathbb{C}^{\mathrm{N}}$ is the $k^{\underline{th}}$ training block. For simplicity, we reuse the notations adopted in chapter 3 for the subsequent discussion on channel estimation. Hence the encoder output becomes

$$\mathbf{G}_B \triangleq \sum_{k=1}^{K} \left[\mathbf{X}_{A_k} \otimes \mathbf{c}_k + \mathbf{X}_{B_k} \otimes \mathbf{d}_k^* \right] \in \mathbb{C}^{2\mathrm{NK} \times \mathrm{M_t}},$$

where

$$\begin{cases} \mathbf{c}_k = \mathbf{t}_k = \begin{bmatrix} c_k(0) & c_k(1) & \cdots & c_k(N-1) \end{bmatrix}^T. \\ \mathbf{d}_k = \begin{bmatrix} d_k(0) & d_k(1) & \cdots & d_k(N-1) \end{bmatrix}^T. \\ d_k(n) = c_k((\text{-}n)_N),\ 0 \le n \le N-1. \end{cases}$$

Since the channel estimation is processed on a per receiver basis, we consider MISO in the subsequent sections of this chapter, i.e., the index j equals one and one only, and will hence be discarded.

5.1 Least-Squares Channel Estimate

With the training matrix Π_t being transmitted, the received signal block in (3.1) becomes, for $1 \le p \le 2K$,

$$\mathbf{y}^{(p)} = \sum_{m=1}^{M_t} \mathbf{H}_m \mathbf{s}_m^{(p)} + \mathbf{w}^{(p)} \in \mathbb{C}^{\mathrm{N}},$$

where $\mathbf{s}_m^{(p)}$ takes $\mathbf{t}_k$ if $\left|\left[\mathbf{X}_{A_k}\right]_{pm}\right|$ is one, for $1 \le p \le K$. Similarly, $\mathbf{s}_m^{(p)}$ takes $\mathbf{d}_k^*$ if $\left|\left[\mathbf{X}_{B_k}\right]_{pm}\right|$ is one, for $K+1 \le p \le 2K$. Following the same derivations as those from (3.1) to (3.3), we have the following frequency-domain received signal model

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}' &\triangleq (\mathbf{I}_{2K} \otimes \boldsymbol{\mathcal{F}})\mathbf{y}' \\
&= \left[\left(\boldsymbol{\mathcal{Y}}^{(1)}\right)^T \quad \left(\boldsymbol{\mathcal{Y}}^{(2)}\right)^T \quad \cdots \quad \left(\boldsymbol{\mathcal{Y}}^{(K)}\right)^T \middle| \left(\boldsymbol{\mathcal{Y}}^{(K+1)}\right)^T \quad \cdots \quad \left(\boldsymbol{\mathcal{Y}}^{(2K)}\right)^T\right]^T \\
&= \sum_{k=1}^{K}\sum_{m=1}^{M_t}\left\{\begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_k}(:,m) \\ \mathbf{0}_{K\times 1}\end{pmatrix} \otimes \left(\Lambda_m \underline{\mathbb{X}}_k\right) + \begin{pmatrix} \mathbf{0}_{K\times 1} \\ \boldsymbol{\mathcal{G}}_{A_k}(:,m)\end{pmatrix} \otimes \left(\Lambda_m \underline{\mathbb{X}}_k^*\right)\right\} + \boldsymbol{\mathcal{W}}'.
\end{aligned} \tag{5.1}$$

Notice that $\Lambda_m \underline{\mathbb{X}}_k = Diag\left(\sqrt{N}\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)\left(\boldsymbol{\mathcal{F}}\mathbf{c}_k\right) = \sqrt{N}Diag\left(\boldsymbol{\mathcal{F}}\mathbf{c}_k\right)\left(\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)$. So if we define $\Lambda_{\underline{\mathbb{X}}_k}$ as $Diag\left(\boldsymbol{\mathcal{F}}\mathbf{c}_k\right)$, then we have $\Lambda_m \underline{\mathbb{X}}_k = \sqrt{N}\Lambda_{\underline{\mathbb{X}}_k}\left(\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)$ [4]. Since $\boldsymbol{\mathcal{F}}\mathbf{d}_k^* = \left(\boldsymbol{\mathcal{F}}\mathbf{c}_k\right)^*$, we have $Diag\left(\boldsymbol{\mathcal{F}}\mathbf{d}_k^*\right) = \Lambda_{\underline{\mathbb{X}}_k}^H$. So $\boldsymbol{\mathcal{Y}}'$ in (5.1) can be rewritten as

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}' &= \sum_{k=1}^{K}\left\{\sum_{m=1}^{M_t}\begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_k}(:,m) \\ \mathbf{0}_{K\times 1}\end{pmatrix} \otimes \left(\sqrt{N}\Lambda_{\underline{\mathbb{X}}_k}\left(\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)\right) + \begin{pmatrix} \mathbf{0}_{K\times 1} \\ \boldsymbol{\mathcal{G}}_{A_k}(:,m)\end{pmatrix} \otimes \left(\sqrt{N}\Lambda_{\underline{\mathbb{X}}_k}^H\left(\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)\right)\right\} + \boldsymbol{\mathcal{W}}' \\
&= \sqrt{N}\left\{\sum_{m=1}^{M_t}\left\{\sum_{k=1}^{K}\left(\begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_k}(:,m) \\ \mathbf{0}_{K\times 1}\end{pmatrix} \otimes \Lambda_{\underline{\mathbb{X}}_k} + \begin{pmatrix} \mathbf{0}_{K\times 1} \\ \boldsymbol{\mathcal{G}}_{A_k}(:,m)\end{pmatrix} \otimes \Lambda_{\underline{\mathbb{X}}_k}^H\right)\right\}\left(\boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m\right)\right\} + \boldsymbol{\mathcal{W}}'.
\end{aligned} \tag{5.2}$$

Let $\mathbf{O}_m \triangleq \sum_{k=1}^{K}\left\{\begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_k}(:,m) \\ \mathbf{0}_{K\times 1}\end{pmatrix} \otimes \Lambda_{\underline{\mathbb{X}}_k} + \begin{pmatrix} \mathbf{0}_{K\times 1} \\ \boldsymbol{\mathcal{G}}_{A_k}(:,m)\end{pmatrix} \otimes \Lambda_{\underline{\mathbb{X}}_k}^H\right\}$ and $\Lambda_T \triangleq \left[\mathbf{O}_1 \quad \mathbf{O}_2 \quad \cdots \quad \mathbf{O}_{M_t}\right] \in \mathbb{C}^{2\mathrm{KN}\times \mathrm{M_t N}}$. Since $\bar{\mathbf{h}} = \left[\bar{\mathbf{h}}_1^T \quad \bar{\mathbf{h}}_2^T \quad \cdots \quad \bar{\mathbf{h}}_{M_t}^T\right]^T \in \mathbb{C}^{\mathrm{M_t N}}$ by (2.5), we define $\tilde{\bar{\mathbf{h}}} \triangleq \left(\mathbf{I}_{M_t} \otimes \boldsymbol{\mathcal{F}}\right)\bar{\mathbf{h}} \in \mathbb{C}^{\mathrm{M_t N}}$. Then (5.2) becomes

$$\boldsymbol{\mathcal{Y}}' = \sqrt{N}\Lambda_T \tilde{\bar{\mathbf{h}}} + \boldsymbol{\mathcal{W}}', \tag{5.3}$$

Given training signal blocks, and hence Λ_T, it is seen from (5.3) that if Π_t is designed such that $rank\left(\Lambda_T\right) = M_t N$, then by nonsingularity of $\Lambda_T^H \Lambda_T$ we can obtain an estimate for the channel DFT by using the pseudo-inverse $\Lambda_T^+ \triangleq \left(\Lambda_T^H \Lambda_T\right)^{-1}\Lambda_T^H$. Note that for generalized complex orthogonal design, $2K \ge M_t$. Define $\hat{\tilde{\bar{\mathbf{h}}}} \triangleq \frac{1}{\sqrt{N}}\Lambda_T^+ \boldsymbol{\mathcal{Y}}'$ to be an estimate of $\tilde{\tilde{\mathbf{h}}}$, which contains the subchannel DFT's and let $\widehat{\mathcal{W}}' = \frac{1}{\sqrt{N}}\Lambda_T^{+}\mathcal{W}'$. Then we have, from (5.3),

$$\hat{\tilde{\tilde{\mathbf{h}}}} = \tilde{\tilde{\mathbf{h}}} + \widehat{\mathcal{W}}'. \tag{5.4}$$

Hence, $\tilde{\tilde{\mathbf{h}}}$ can be estimated up to an zero-mean estimation error term $\widehat{\mathcal{W}}'$. Note that equation (5.4) can be regarded as a consequence from FDE as well. The only difference lies in that during the training mode, we take vectorized channel DFT $\tilde{\tilde{\mathbf{h}}}$ as input block into the system constructed by training signal DFT's along with GOSTBC constituent matrices. More precisely, we have the following results.

Theorem 5.1 $\Lambda_T^H\Lambda_T = 2\sum_{k=1}^{K}\left(\mathbf{I}_{M_t}\otimes\Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right)$.

(*Proof:* See Appendix A2.)

Remark 5.1 With known training blocks, this FDE-based channel estimation can be carried out *efficiently*. By efficiency we mean the involved computation for matrix inverse $\left(\Lambda_T^H\Lambda_T\right)^{-1}$ reduces to merely computing the reciprocals of each of the diagonal entries of $\Lambda_T^H\Lambda_T$. This can be seen from ***Theorem 5.1*** that $\Lambda_T^H\Lambda_T = 2\sum_{k=1}^{K}\left(\mathbf{I}_{M_t}\otimes\Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right)$ is diagonal.

Remark 5.2 Notice that the FDE behind (5.4) is based on the premise that $rank\left(\Lambda_T\right) = M_tN$, and hence $rank\left(\Lambda_T^H\Lambda_T\right) = M_tN$. This in turn require that $rank\left(\sum_{k=1}^{K}\Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right) = N$ must be satisfied as we can see from ***Remark 5.1***. Notice that each of the diagonal entries of $\Lambda_T^H\Lambda_T$ is nonnegative. To proceed further, let us first define $\underline{\mathbb{X}}_k(\overline{n})$ as the $\overline{n}^{th}$ entry of $\underline{\mathbb{X}}_k$, where $1\le\overline{n}\le N$ and $1\le k\le K$. As long as there is at least one value k^*, $1\le k^*\le K$ such that $\left|\underline{\mathbb{X}}_{k^*}(\overline{n})\right|^2 > 0$ among all the $\overline{n}^{th}$ entries, the premise holds. Let us denote this sufficient condition as (c2.R). However, unlike the conclusion we obtained in ***Remark 3.5***, (c2.R) could fail hadn't the training blocks been properly chosen, e.g., the arbitrary training we adopted for simulations in chapter 6. We will elaborate the meaning of *arbitrariness* at the end of this subsection. For the simulations in chapter 6, we simply bypass it with programming.

5.1.1 Optimality Criterion for LS Estimate under a Power Constraint

We use (5.3) to derive optimality criterion for optimal frequency-domain least-squares (LS) channel estimation under a power constraint. To this end, let us define $\mathbf{P} \triangleq \Lambda_T$, $\tilde{\bar{\mathbf{h}}}_m \triangleq \boldsymbol{\mathcal{F}}\bar{\mathbf{h}}_m \in \mathbb{C}^{N}$ and $\boldsymbol{\mathcal{H}} \triangleq \begin{bmatrix} \tilde{\bar{\mathbf{h}}}_1 & \tilde{\bar{\mathbf{h}}}_2 & \cdots & \tilde{\bar{\mathbf{h}}}_{M_t} \end{bmatrix} \in \mathbb{C}^{M_t N}$. Hence $\tilde{\bar{\mathbf{h}}} = \mathrm{vec}(\boldsymbol{\mathcal{H}})$. Then we can rewrite (5.3) as

$$\boldsymbol{\mathcal{Y}}' = \sqrt{N}\mathbf{P}\mathrm{vec}(\boldsymbol{\mathcal{H}}) + \boldsymbol{\mathcal{W}}'. \tag{5.5}$$

From (5.5), the LS estimate of $\mathrm{vec}(\boldsymbol{\mathcal{H}})$ is $\mathrm{vec}\left(\left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right) \triangleq \underset{\mathrm{vec}(\boldsymbol{\mathcal{H}})}{\arg\min} \left\|\sqrt{N}\mathbf{P}\mathrm{vec}(\boldsymbol{\mathcal{H}}) - \boldsymbol{\mathcal{Y}}'\right\|_2^2$. The solution is known as, from (5.4),

$$\mathrm{vec}\left(\left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right) = \left(\tfrac{1}{\sqrt{N}}\mathbf{P}^{+}\right)\boldsymbol{\mathcal{Y}}' = \mathrm{vec}(\boldsymbol{\mathcal{H}}) + \widehat{\boldsymbol{\mathcal{W}}}', \tag{5.6}$$

where $\mathrm{vec}\left(\left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right)$ is equal to $\hat{\tilde{\bar{\mathbf{h}}}}$ in (5.4) by definition. Now suppose we are given a power constraint,

$$\sum_{k=1}^{K}\left\|\mathbf{t}_k\right\|_2^2 = \frac{\mathbb{P}_o}{2M_t}. \tag{5.7}$$

This means that the total transmitted signal block power per transmit antenna shall be fixed, by virtue of GOSTBC structure. Note that $\left\|\mathbf{t}_k\right\|_2 = \left\|\mathbf{c}_k\right\|_2 = \left\|\underline{\mathbb{X}}_k\right\|_2$. We can transform the power constraint into frequency domain, by ***Theorem 5.1***,

$$\begin{aligned} \mathbb{P}_o &= 2M_t\sum_{k=1}^{K}\left\|\underline{\mathbb{X}}_k\right\|_2^2 \\ &= 2\mathrm{Tr}\left\{\sum_{k=1}^{K}\left(\mathbf{I}_{M_t}\otimes\Lambda_{\underline{\mathbb{X}}_k}^{H}\Lambda_{\underline{\mathbb{X}}_k}\right)\right\} \\ &= \mathrm{Tr}\left\{\mathbf{P}^{H}\mathbf{P}\right\} \\ &= \left\|\mathbf{P}\right\|_F^2, \end{aligned} \tag{5.8}$$

where $\|\mathbf{P}\|_F$ denotes the Frobenius norm of the matrix $\mathbf{P}$. From (5.6) and (5.8), we wish to find the optimality condition for training design to achieve lowest LS estimation error. This amounts to the following optimization problem:

$$\min_{\mathbf{P}} E\left\{\left\|\boldsymbol{\mathcal{H}} - \left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right\|_F^2\right\} \text{ subject to } \|\mathbf{P}\|_F^2 = \mathbb{P}_o, \tag{5.9}$$

where $\mathbb{P}_o \in \mathbb{R}_+$ is a positive constant. With (5.6), we can rewrite the mean-square error in (5.9) as

$$\begin{aligned} E\left\{\left\|\boldsymbol{\mathcal{H}} - \left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right\|_F^2\right\} &= E\left\{\left\|\mathrm{vec}(\boldsymbol{\mathcal{H}}) - \mathrm{vec}\left(\left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right)\right\|_2^2\right\} \\ &= E\left\{\left\|\widehat{\boldsymbol{\mathcal{W}}}'\right\|_2^2\right\} \\ &= E\left\{\left\|\tfrac{1}{\sqrt{N}}\mathbf{P}^+\left(\mathbf{I}_{2K}\otimes\boldsymbol{\mathcal{F}}\right)\mathbf{w}'\right\|_2^2\right\}. \end{aligned} \tag{5.10}$$

For simplicity, let us define $\boldsymbol{\mathcal{F}}_K \triangleq \mathbf{I}_{2K}\otimes\boldsymbol{\mathcal{F}}$. By noting that $E\left\{\mathbf{w}'\mathbf{w}'^H\right\} = \sigma_w^2\mathbf{I}_{2KN}$, $\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}^H = \mathbf{I}_N$ and $\mathbf{P}^+ = \left(\mathbf{P}^H\mathbf{P}\right)^{-1}\mathbf{P}^H$, we can rewrite (5.10) as

$$\begin{aligned} E\left\{\left\|\boldsymbol{\mathcal{H}} - \left(\widehat{\boldsymbol{\mathcal{H}}}\right)_{LS}\right\|_F^2\right\} &= \frac{1}{N}\mathrm{Tr}\left\{\left(\mathbf{P}^+\right)\boldsymbol{\mathcal{F}}_K E\left\{\mathbf{w}'\mathbf{w}'^H\right\}\boldsymbol{\mathcal{F}}_K^H\left(\mathbf{P}^+\right)^H\right\} \\ &= \frac{\sigma_w^2}{N}\mathrm{Tr}\left\{\left(\mathbf{P}^+\right)\boldsymbol{\mathcal{F}}_K\mathbf{I}_{2KN}\boldsymbol{\mathcal{F}}_K^H\left(\mathbf{P}^+\right)^H\right\} \\ &= \frac{\sigma_w^2}{N}\mathrm{Tr}\left\{\left(\mathbf{P}^H\mathbf{P}\right)^{-1}\right\}. \end{aligned} \tag{5.11}$$

Hence the optimization problem in (5.9) becomes

$$\min_{\mathbf{P}} \frac{\sigma_w^2}{N}\mathrm{Tr}\left\{\left(\mathbf{P}^H\mathbf{P}\right)^{-1}\right\} \text{ subject to } \mathrm{Tr}\left\{\mathbf{P}^H\mathbf{P}\right\} = \mathbb{P}_o. \tag{5.12}$$

Since $\mathbf{P}^H\mathbf{P}$ is Hermitian and positive semi-definite (p.s.d.), we can decompose it into $\mathbf{P}^H\mathbf{P} = \mathbf{Q}\Delta\mathbf{Q}^H$, where $\Delta = Diag\left(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_{M_tN}\right)$, $\mathbf{Q}^H\mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}_{M_tN}$ and λ_i, $1 \le i \le M_tN$ are the eigenvalues of the matrix $\mathbf{P}^H\mathbf{P}$. Therefore, (5.12) becomes minimizing $\sum_{i=1}^{M_tN}\frac{1}{\lambda_i}$ under the constraint that $\sum_{i=1}^{M_tN}\lambda_i = \mathbb{P}_o$. By the arithmetic-geometric mean inequality[25, pp. 75], we know that $\sum_{i=1}^{M_tN}\frac{1}{\lambda_i} \geq M_tN\left(\prod_{i=1}^{M_tN}\frac{1}{\lambda_i}\right)^{1/(M_tN)}$, where equality holds when all λ_i are identical. Hence,

$$\lambda_i = \frac{\mathbb{P}_o}{M_tN},\ 1 \leq i \leq M_tN, \tag{5.13}$$

is the solution to (5.12). From (5.13), we have

$$\mathbf{P}^H\mathbf{P} = \frac{\mathbb{P}_o}{M_tN}\mathbf{I}_{M_tN}. \tag{5.14}$$

Notice that for any matrix **P**, regardless of its structure, which satisfies (5.14) is an optimal solution to (5.12), and hence optimal to (5.9). However, the structure of matrix **P** under consideration here is determined by the BGOSTBC encoder as well according to ***Theorem 5.1***. We summarize what we have so far.

(i) For a given training matrix Π_t, which completely determines $\mathbf{P} = \Lambda_T$, the LS estimate of the channel vector $\tilde{\bar{\mathbf{h}}}$ is given by $\hat{\tilde{\bar{\mathbf{h}}}} = \left(\frac{1}{\sqrt{N}}\mathbf{P}^+\right)\boldsymbol{\mathcal{Y}}'$, where $\boldsymbol{\mathcal{Y}}'$ is defined in (5.3).

(ii) If the training matrix Π_t is chosen such that $\frac{\mathbb{P}_o}{M_tN}\mathbf{I}_{M_tN} = 2\sum_{k=1}^{K}\left(\mathbf{I}_{M_t} \otimes \Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right)$, then the training blocks satisfies the power constraint in (5.7) and minimizes the average channel estimation error $E\left\{\left\|\tilde{\bar{\mathbf{h}}} - \hat{\tilde{\bar{\mathbf{h}}}}\right\|_2^2\right\}$.

(iii) In particular, the lowest LS channel estimation cost is given by

$$(J)_{LS} = \frac{\sigma_w^2}{N}\mathrm{Tr}\left\{\left(\frac{\mathbb{P}_o}{M_tN}\mathbf{I}_{M_tN}\right)^{-1}\right\} = \frac{N\sigma_w^2M_t^2}{\mathbb{P}_o}. \tag{5.15}$$

From (5.15) we can see that the LS objective per transmit antenna is proportional to M_t. Therefore, as number of transmit antennas increases, the performance of LS channel estimate deteriorates.

5.1.2 Selection of training signal blocks for optimal LS channel estimate

Notice that the power criterion, $\frac{\mathbb{P}_o}{M_t N}\mathbf{I}_{M_t N} = 2\sum_{k=1}^{K}\left(\mathbf{I}_{M_t}\otimes\Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right)$, can be rephrased as follows:

$$\frac{\mathbb{P}_o}{M_t N} = 2\sum_{k=1}^{K} a_k(n),\ 0\le n\le N-1,\ \text{where } \mathbf{a}_k = Diag\left(\Lambda_{\underline{\mathbb{X}}_k}^H\Lambda_{\underline{\mathbb{X}}_k}\right). \tag{5.16}$$

If we allow the employed symbol constellation to be arbitrary, the choices for training blocks which satisfy (5.16) are simply any matrix Π_{t^*} satisfying $\|\mathbf{b}_n\|_2^2 = \frac{\mathbb{P}_o}{2M_t N}$, where $\Pi_{t^*} = \boldsymbol{\mathcal{F}}^H\left[\mathbf{b}_0^T\quad \mathbf{b}_1^T\quad\cdots\quad\mathbf{b}_{N-1}^T\right]^T$ and $\mathbf{b}_n = \left[b_n(1)\quad b_n(2)\quad\cdots\quad b_n(K)\right]\in\mathbb{C}^{1\times K}$. This is true because (5.16) can be rephrased as: $\frac{\mathbb{P}_o}{2M_t N} = \|\mathbf{b}_n\|_2^2$, $0\le n\le N-1$, where the relation between $\underline{\mathbb{X}}_k$ and $\mathbf{b}_n$ is given by $\underline{\mathbb{X}}_k = \left[b_0(k)\quad b_1(k)\quad\cdots\quad b_{N-1}(k)\right]^T$. In particular, when $N=K$, (5.16) can be easily satisfied by letting $\mathbf{t}_k = \sqrt{\frac{\mathbb{P}_o}{2M_t N}}\boldsymbol{\mathcal{F}}^H(:,k)$, $1\le k\le K=N$, which is essentially a general PSK constellation.

In chapter 6, we will use $\mathbf{t}_k = \sqrt{\frac{\mathbb{P}_o}{2M_t N}}\boldsymbol{\mathcal{F}}^H(:,k)$, $1\le k\le K=N$, for comparing the performances of optimal/orthogonal training and arbitrary training. By optimality we mean that (5.8) along with (5.16) must be satisfied a priori, and by arbitrariness the simulation subjects to (5.8) only. Though both of them must guarantee that Λ_T is full rank. Note that the word "orthogonal" we adopted here is to underline (5.14).

5.2 Linear Minimum-Mean-Square Error Estimate

In this subsection, we are interested in finding the *time-domain* linear MMSE channel estimate. Suppose $\mathbf{h}_{all}$ is zero-mean for simplicity. Before we proceed further, let us define a ZP removal matrix $\mathbb{M}\triangleq\left[\mathbf{I}_{L+1}\quad \mathbf{0}_{(L+1)\times(N-L-1)}\right]$. Note that $\mathbb{M}\mathbb{M}^T = \mathbf{I}_{L+1}$. Hence $\mathbb{M}\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{H}} = \mathbb{M}\left[\bar{\mathbf{h}}_1\quad\bar{\mathbf{h}}_2\quad\cdots\quad\bar{\mathbf{h}}_{M_t}\right] = \left[\mathbf{h}_1\quad\mathbf{h}_2\quad\cdots\quad\mathbf{h}_{M_t}\right]\in\mathbb{C}^{(L+1)\times M_t}$ and $\text{vec}\left(\mathbb{M}\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{H}}\right) = \mathbf{h}$ by definition. Let $\hat{\mathbf{h}}$ be an estimate of $\mathbf{h}$. Now we define the mean-square error to be minimized as,

$$\varepsilon\triangleq E\left\{\left\|\mathbf{h}-\hat{\mathbf{h}}\right\|_2^2\right\} = E\left\{\left\|\text{vec}\left(\mathbb{M}\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{H}}\right)-\hat{\mathbf{h}}\right\|_2^2\right\}. \tag{5.17}$$

From (5.17) and by definition, the linear MMSE estimate of $\mathbf{h}$ is $\left(\hat{\mathbf{h}}\right)_{MMSE} \triangleq \arg\min_{\hat{\mathbf{h}}} \varepsilon$. Let Φ be a linear channel estimator such that the input-output relation between the received signal blocks and the estimate is given by $\hat{\mathbf{h}} = \Phi\boldsymbol{\mathcal{Y}}'$. Then we turn to find $(\Phi)_{opt} = \arg\min_{\Phi} \varepsilon$ and

$$\left(\hat{\mathbf{h}}\right)_{MMSE} = (\Phi)_{opt}\,\boldsymbol{\mathcal{Y}}'. \tag{5.18}$$

Now, for notational purpose, let us define $\mathbf{P} \triangleq \sqrt{N}\Lambda_T$ and $\boldsymbol{\mathcal{F}}_M \triangleq \mathbf{I}_{M_t} \otimes \boldsymbol{\mathcal{F}}$. Hence, $\tilde{\bar{\mathbf{h}}} = \boldsymbol{\mathcal{F}}_M \bar{\mathbf{h}}$ by definition. We know $\bar{\mathbf{h}} = \left(\mathbf{I}_{M_t} \otimes \mathbb{M}^T\right)\mathbf{h}$. Rewrite (5.17) as, from (5.3),

$$\begin{aligned}
\varepsilon &= E\left\{\left\|\mathbf{h} - \Phi\boldsymbol{\mathcal{Y}}'\right\|_2^2\right\} \\
&= E\left\{\mathrm{Tr}\left\{\left(\mathbf{h} - \Phi\boldsymbol{\mathcal{Y}}'\right)\left(\mathbf{h} - \Phi\boldsymbol{\mathcal{Y}}'\right)^H\right\}\right\} \\
&= \mathrm{Tr}\Big\{E\left\{\mathbf{h}\mathbf{h}^H\right\} - E\left\{\mathbf{h}\bar{\mathbf{h}}^H\right\}\boldsymbol{\mathcal{F}}_M^H\mathbf{P}^H\Phi^H + \\
&\quad \Phi\boldsymbol{\mathcal{F}}_K E\left\{\mathbf{w}'\mathbf{w}'^H\right\}\boldsymbol{\mathcal{F}}_K^H\Phi^H - \Phi\mathbf{P}\boldsymbol{\mathcal{F}}_M E\left\{\bar{\mathbf{h}}\mathbf{h}^H\right\} + \\
&\quad \Phi\mathbf{P}\boldsymbol{\mathcal{F}}_M E\left\{\bar{\mathbf{h}}\bar{\mathbf{h}}^H\right\}\boldsymbol{\mathcal{F}}_M^H\mathbf{P}^H\Phi^H\Big\}.
\end{aligned}$$

By defining $\mathbf{R}_{\boldsymbol{\mathcal{H}}} \triangleq E\left\{\mathbf{h}\mathbf{h}^H\right\}$ and using the identity $E\left\{\mathbf{w}'\mathbf{w}'^H\right\} = \sigma_w^2\mathbf{I}_{2KN}$ along with $\bar{\mathbf{h}} = \left(\mathbf{I}_{M_t} \otimes \mathbb{M}^T\right)\mathbf{h}$, we have the following equation:

$$\begin{aligned}
\varepsilon = \Big\{&\mathrm{Tr}\left\{\mathbf{R}_{\boldsymbol{\mathcal{H}}}\right\} - \mathrm{Tr}\left\{\mathbf{R}_{\boldsymbol{\mathcal{H}}}\left(\mathbf{I}_{M_t} \otimes \mathbb{M}\boldsymbol{\mathcal{F}}^H\right)\mathbf{P}^H\Phi^H\right\} - \mathrm{Tr}\left\{\Phi\mathbf{P}\left(\mathbf{I}_{M_t} \otimes \boldsymbol{\mathcal{F}}\mathbb{M}^T\right)\mathbf{R}_{\boldsymbol{\mathcal{H}}}\right\} + \\
&\mathrm{Tr}\left\{\Phi\left(\sigma_w^2\mathbf{I}_{2KN} + \mathbf{P}\left(\mathbf{I}_{M_t} \otimes \boldsymbol{\mathcal{F}}\mathbb{M}^T\right)\mathbf{R}_{\boldsymbol{\mathcal{H}}}\left(\mathbf{I}_{M_t} \otimes \mathbb{M}\boldsymbol{\mathcal{F}}^H\right)\mathbf{P}^H\right)\Phi^H\right\}\Big\}.
\end{aligned} \tag{5.19}$$

Note that $\mathbf{R}_{\boldsymbol{\mathcal{H}}}$ is nonsingular since $\mathbf{h}$ is not zero-padded. We have assumed that both $\mathbf{h}$ and $\mathbf{w}'$ are zero-mean and they are statistically independent of each other, that is, $E\left\{\mathbf{h}\mathbf{w}'^H\right\} = \mathbf{0}_{M_t(L+1)\times 2KN}$ and $E\left\{\mathbf{w}'\mathbf{h}^H\right\} = \mathbf{0}_{2KN\times M_t(L+1)}$. In the subsequent discussion regarding linear MMSE approach, we assume perfect knowledge of $\mathbf{R}_{\boldsymbol{\mathcal{H}}}$ and σ_w^2. Now we let $\frac{\partial\varepsilon}{\partial\Phi} \equiv \mathbf{0}$ [5],[6] to find $(\Phi)_{opt}$:

$$\begin{aligned}
\frac{\partial\varepsilon}{\partial\Phi} = \Big\{&\mathbf{0} - 2\mathbf{R}_{\boldsymbol{\mathcal{H}}}\left(\mathbf{I}_{M_t} \otimes \mathbb{M}\boldsymbol{\mathcal{F}}^H\right)\mathbf{P}^H + \\
&2\Phi\left(\sigma_w^2\mathbf{I}_{2KN} + \mathbf{P}\left(\mathbf{I}_{M_t} \otimes \boldsymbol{\mathcal{F}}\mathbb{M}^T\right)\mathbf{R}_{\boldsymbol{\mathcal{H}}}\left(\mathbf{I}_{M_t} \otimes \mathbb{M}\boldsymbol{\mathcal{F}}^H\right)\mathbf{P}^H\right)\Big\}\Big|_{\Phi=(\Phi)_{opt}} \equiv 0
\end{aligned} \tag{5.20}$$

For brevity, let $\overline{\mathcal{F}}_M \triangleq \left(\mathbf{I}_{M_t} \otimes \mathcal{F}\mathbb{M}^T\right)$. Notice that $\overline{\mathcal{F}}_M^H \overline{\mathcal{F}}_M = \left(\mathbf{I}_{M_t} \otimes \mathbb{M}\mathcal{F}^H \mathcal{F}\mathbb{M}^T\right) = \mathbf{I}_{M_t(L+1)}$ by definition of $\mathbb{M}$. Define $\widetilde{\mathbf{P}} \triangleq \mathbf{P}\overline{\mathcal{F}}_M$. Hence, from (5.20), we have

$$(\Phi)_{opt} = \mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H \left[\sigma_w^2 \mathbf{I}_{2KN} + \widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\right]^{-1}. \tag{5.21}$$

Hence, with $(\Phi)_{opt}$ in (5.21), we can rewrite (5.18) as

$$\left(\hat{\mathbf{h}}\right)_{MMSE} = \mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H \left[\sigma_w^2 \mathbf{I}_{2KN} + \widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\right]^{-1} \mathcal{Y}', \tag{5.22}$$

with the corresponding lowest linear MMSE cost denoted as $(\varepsilon)_{MMSE} = \varepsilon|_{\Phi=(\Phi)_{opt}}$. By substituting (5.22) into (5.17), we arrive at

$$\begin{aligned}
(\varepsilon)_{MMSE} &= \varepsilon|_{\Phi=\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\left[\sigma_w^2\mathbf{I}_{2KN}+\widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\right]^{-1}} \\
&= \mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}\right\} - \mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H \left[\sigma_w^2 \mathbf{I}_{2KN} + \widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\right]^{-1} \widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\right\} \\
&= \mathrm{Tr}\left\{\left[\mathbf{R}_{\mathcal{H}}^{-1} + \frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right]^{-1}\right\}.
\end{aligned} \tag{5.23}$$

5.2.1 Optimality Criterion for Linear MMSE Estimate under a Power Constraint

First, we need to specify a reasonable power constraint at our disposal. The given power constraint in section 5.1.1, i.e., $\sum_{k=1}^{K}\left\|\mathbf{t}_k\right\|_2^2 = \frac{\mathbb{P}_o}{2M_t}$, can be transformed into $\left\|\mathbf{P}\right\|_F^2 = N\mathbb{P}_o$ with $\mathbf{P} \triangleq \sqrt{N}\Lambda_T$ for linear MMSE derivation. Notice that the matrix $\mathbf{P}^H\mathbf{P}$ subjects to ***Theorem 5.1*** and hence it is diagonal. It can be verified that for any diagonal matrix $\mathbf{A} \in \mathbb{C}^{M_tN\times M_tN}$, the matrix $\mathcal{F}_M\left(\mathbf{I}_{M_t} \otimes \mathbb{M}^T\mathbb{M}\right)\mathcal{F}_M^H$ serves as spreading the power of $\mathbf{A}$ only, hence the following equality holds:

$$\mathrm{Tr}\left\{\mathbf{A}\mathcal{F}_M\left(\mathbf{I}_{M_t} \otimes \mathbb{M}^T\mathbb{M}\right)\mathcal{F}_M^H\right\} = \frac{L+1}{N}\mathrm{Tr}\left\{\mathbf{A}\right\}. \tag{5.24}$$

With (5.8) and (5.24), we herein define a reasonable power constraint for our linear

MMSE estimate as

$$\begin{aligned}\left\|\widetilde{\mathbf{P}}\right\|_F^2 &= \operatorname{Tr}\left\{\left(\mathbf{P}^H\mathbf{P}\right)\boldsymbol{\mathcal{F}}_M\left(\mathbf{I}_{M_t}\otimes\mathbb{M}^T\mathbb{M}\right)\boldsymbol{\mathcal{F}}_M^H\right\}\\ &= \frac{L+1}{N}\operatorname{Tr}\left\{\mathbf{P}^H\mathbf{P}\right\}\\ &= (L+1)\mathbb{P}_o. \end{aligned} \tag{5.25}$$

Therefore, given $\left\|\widetilde{\mathbf{P}}\right\|_F^2=(L+1)\mathbb{P}_o$, we are to find optimality conditions over the training matrix minimizing $(\varepsilon)_{MMSE}$ in (5.23). The optimization problem can be formulated as

$$\min_{\widetilde{\mathbf{P}}}\ (\varepsilon)_{MMSE}\ \text{ subject to } \left\|\widetilde{\mathbf{P}}\right\|_F^2=(L+1)\mathbb{P}_o. \tag{5.26}$$

Define the Lagrangian $\mathcal{L}\left(\widetilde{\mathbf{P}},\ \mu\right)\triangleq(\varepsilon)_{MMSE}+\mu\left[\operatorname{Tr}\left\{\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right\}-(L+1)\mathbb{P}_o\right]$, where $\mu\in\mathbb{R}_+$ is a Lagrangian multiplier. Let $\frac{\partial\mathcal{L}(\widetilde{\mathbf{P}},\ \mu)}{\partial\left(\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)}\equiv\mathbf{0}$ [5],[6]:

$$\begin{aligned}\frac{\partial\mathcal{L}\left(\widetilde{\mathbf{P}},\ \mu\right)}{\partial\left(\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)} &= \frac{\partial(\varepsilon)_{MMSE}}{\partial\left(\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)}+\mu\left[\frac{\partial\operatorname{Tr}\left\{\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right\}}{\partial\left(\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)}-0\right]\\ &= \mu\mathbf{I}_{M_tN}+\frac{\partial\operatorname{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right\}}{\partial\left(\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)}\frac{\partial\operatorname{Tr}\left\{\left[\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right]^{-1}\right\}}{\partial\left(\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right)}\\ &= \mu\mathbf{I}_{M_tN}-\frac{1}{\sigma_w^2}\mathbf{I}_{M_tN}\left[\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right]^{-2}\equiv\mathbf{0}\\ &\Rightarrow \mu\mathbf{I}_{M_tN}=\frac{1}{\sigma_w^2}\left[\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right]^{-2}. \end{aligned} \tag{5.27}$$

Since $\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}$ is Hermitian, we can decompose it into $\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}=\Psi\Delta\Psi^H$, where $\Psi^H\Psi=\Psi\Psi^H=\mathbf{I}_{M_tN}$ and Δ is a diagonal matrix with the eigenvalues of $\mathbf{R}_{\mathcal{H}}^{-1}+\frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}$ on its diagonal. Substituting this decomposition into (5.27), we have $\Psi\Delta^{-2}\Psi^H=\sigma_w^2\mu\mathbf{I}_{M_tN}$. Hence $\Delta=\frac{1}{\sigma_w\sqrt{\mu}}\mathbf{I}_{M_tN}$, and the decomposition becomes

$$\mathbf{R}_{\mathcal{H}}^{-1} + \frac{1}{\sigma_w^2}\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \frac{1}{\sigma_w\sqrt{\mu}}\mathbf{I}_{M_tN} \tag{5.28}$$

Imposing $\mathrm{Tr}\left\{\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}}\right\} = (L+1)\mathbb{P}_o$ on (5.28), we have $\frac{1}{\sigma_w\sqrt{\mu}} = \frac{1}{M_tN}\left(\frac{(L+1)\mathbb{P}_o}{\sigma_w^2} + \mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}\right\}\right)$, which in turn can be substituted back into (5.28) to obtain

$$\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \frac{\mathbb{P}_o}{M_t}\left\{\left(\frac{L+1}{N} + \frac{\sigma_w^2}{N\mathbb{P}_o}\mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}\right\}\right)\mathbf{I}_{M_tN} - \frac{M_t\sigma_w^2}{\mathbb{P}_o}\mathbf{R}_{\mathcal{H}}^{-1}\right\}. \tag{5.29}$$

Note that the left-hand side of (5.29) is p.s.d., which in turn requires the right-hand side to be p.s.d. as well. So (5.29) is much more *plausible* at high SNR. That is, when $\frac{\sigma_w^2}{\mathbb{P}_o} \to 0$,

$$\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \frac{L+1}{N}\frac{\mathbb{P}_o}{M_t}\mathbf{I}_{M_tN} \tag{5.30}$$

is generically true. Since $(L+1)<N$, there is no way we can recover $\mathbf{P}^H\mathbf{P}$ from (5.30) with the identities that $\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \overline{\mathcal{F}}_M^H\left(\mathbf{P}^H\mathbf{P}\right)\overline{\mathcal{F}}_M$ and $\overline{\mathcal{F}}_M^H\overline{\mathcal{F}}_M = \mathbf{I}_{M_t(L+1)}$. However, if we *relax* the condition of (5.30) into $\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \frac{\mathbb{P}_o}{M_t}\mathbf{I}_{M_tN}$, it can be easily verified that the following choice satisfies both $\widetilde{\mathbf{P}}^H\widetilde{\mathbf{P}} = \overline{\mathcal{F}}_M^H\left(\mathbf{P}^H\mathbf{P}\right)\overline{\mathcal{F}}_M$ and $\|\mathbf{P}\|_F^2 = N\mathbb{P}_o$:

$$\mathbf{P}^H\mathbf{P} = \frac{\mathbb{P}_o}{M_t}\mathbf{I}_{M_tN}. \tag{5.31}$$

Hence, we can say that (5.31) is an optimal solution to (5.26) at high SNR. Surprisingly, at high SNR, the optimal training design for linear MMSE estimate converges to that of the LS estimate. Hence, subsequent discussions similar to those for the LS estimate follow accordingly. We give the following summary.

(i) For a given training matrix Π_t, which completely determines $\mathbf{P} = \sqrt{N}\Lambda_T$, the linear MMSE estimate of the channel vector $\mathbf{h}$ is given by $\left(\hat{\mathbf{h}}\right)_{MMSE} = (\Phi)_{opt}\,\boldsymbol{\mathcal{Y}}'$, where $\boldsymbol{\mathcal{Y}}'$ is defined in (5.3) and $(\Phi)_{opt} = \mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\left[\sigma_w^2\mathbf{I}_{2KN} + \widetilde{\mathbf{P}}\mathbf{R}_{\mathcal{H}}\widetilde{\mathbf{P}}^H\right]^{-1}$ with $\widetilde{\mathbf{P}} \triangleq \mathbf{P}\left(\mathbf{I}_{M_t} \otimes \mathcal{F}\mathbb{M}^T\right)$ and $\mathbb{M} \triangleq \left[\mathbf{I}_{L+1} \quad \mathbf{0}_{(L+1)\times(N-L-1)}\right]$.

(ii) If the training matrix Π_t is chosen such that $\frac{\mathbb{P}_o}{M_t N}\mathbf{I}_{M_t N} = 2\sum_{k=1}^{K}\left(\mathbf{I}_{M_t} \otimes \Lambda_{\mathbb{X}_k}^H \Lambda_{\mathbb{X}_k}\right)$, then the training blocks satisfies the power constraint in (5.7) and minimizes the average channel estimation error $E\left\{\left\|\mathbf{h} - \left(\hat{\mathbf{h}}\right)_{MMSE}\right\|_2^2\right\}$ at high SNR. In chapter 6, through simulation we will show that the choice in (5.31) for the linear MMSE estimate yields good performance to our expectation, though we can only approximate the solution to (5.30) by (5.31).

(iii) In particular, the lowest linear MMSE channel estimation cost is given by

$$\begin{aligned}
\min(\varepsilon)_{MMSE} &= \mathrm{Tr}\left\{\left[\mathbf{R}_{\mathcal{H}}^{-1} + \frac{1}{\sigma_w^2}\left\{\left(\frac{L+1}{N}\frac{\mathbb{P}_o}{M_t} + \frac{\sigma_w^2}{NM_t}\mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}\right\}\right)\mathbf{I}_{M_t N} - \sigma_w^2 \mathbf{R}_{\mathcal{H}}^{-1}\right\}\right]^{-1}\right\} \\
&= \mathrm{Tr}\left\{\left[\frac{1}{M_t}\left(\frac{L+1}{N}\frac{\mathbb{P}_o}{\sigma_w^2} + \frac{1}{N}\mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}\right\}\right)\mathbf{I}_{M_t N}\right]^{-1}\right\} \\
&= \frac{M_t^2 N^2}{(L+1)\frac{\mathbb{P}_o}{\sigma_w^2} + \mathrm{Tr}\left\{\mathbf{R}_{\mathcal{H}}^{-1}\right\}}.
\end{aligned} \tag{5.32}$$

From (5.32) we can see that the linear MMSE objective per transmit antenna is proportional to M_t. Therefore, as number of transmit antennas increases, the performance of linear MMSE channel estimate deteriorates.

Remark 5.3 The channel estimation techniques in [3] include scaled-LS (SLS) estimate and relaxed-MMSE (RMMSE) estimate. Just as what we have done to generalize the results for LS and MMSE scenarios in [3] to the frequency- selective fading environment, we can do the same for SLS and RMMSE channel estimates by following similar routine derivations. However we choose to omit these two generalizations since our primary goal is to validate that the channel estimation techniques in [3] can indeed be generalized and to obtain the optimal training design counterpart for frequency-selective fading under the proposed scheme.

Chapter 6

Simulation Results

In this chapter, the performances of LS and linear MMSE channel estimators are compared through numerical simulations for both arbitrary training and optimal/ orthogonal training. By optimality, we mean that (5.8) along with (5.16) must be satisfied a priori, and by arbitrariness the simulation subjects to (5.8) only. Though both of them must guarantee that Λ_T is full rank.

6.1 Assumptions for Numerical Simulation Runs

As we know from previous discussion that when $N = K$, the power condition in (5.16), can be easily satisfied by letting $\mathbf{t}_k = \sqrt{\frac{\mathbb{P}_o}{2M_t N}}\boldsymbol{\mathcal{F}}^H(:,k)$, $1 \le k \le K = N$. It can be verified that adopting the above scaled DFT blocks for optimal training always satisfies the sufficient condition (c2.R) in ***Remark 5.2.*** Throughout our simulations, the channel coefficients and the additive channel noise are assumed to be circular symmetric complex Gaussian distributed, as we assumed in section 2.2.1, which are randomly generated for each of the following simulations with 1000 independent runs. The channel noise is assumed to be white. We assume that the channel noise variance and $\mathbf{R}_{\boldsymbol{\mathcal{H}}_j}$ are known a priori or have been somehow accurately estimated for the linear MMSE estimate, while the LS estimate utilizes the received signal only. QPSK constellation is employed for all the following simulations.

6.2 Result Overview and Discussion

In Fig. 3, the performances of frequency-domain LS and time-domain linear MMSE channel estimates, both with arbitrary training, are compared in terms of NMSE. Let SNR= $\mathbb{P}_o / \sigma_w^2$. The performances of arbitrary training and its optimal/orthogonal training counterpart is compared in Fig. 4. The FDE on which the symbol decoding is based relies on the channel estimates acquired from training blocks prior to data transmission, during which the channel state remains static. Denote the trio (M_t, M_r, K) as the system configuration referred to. The normalized MSE (NMSE) of channel estimate is given by

$$\text{NMSE} \triangleq \frac{\sum_{j=1}^{M_r} \left\| \mathbf{e}_j - \hat{\mathbf{e}}_j \right\|_2^2}{\left\| \mathbf{e}_j \right\|_2^2},$$

where $\mathbf{e} = \tilde{\tilde{\mathbf{h}}}$ for calculating the frequency-domain LS channel estimation error while $\mathbf{e} = \mathbf{h}$ for the time-domain linear MMSE counterpart.

6.2.1 Arbitrary Training: NMSE Performance

As shown in Fig. 3, the time-domain linear MMSE estimate delivers substantially lower NMSE than that of frequency-domain LS estimate as the former exploits more CSI. The examples can be divided into two major groups depending on value of K. The group with K=4 consists of (3, 2, 4) and (4, 2, 4) configurations while another one consists of (5, 2, 8), (6, 2, 8) and (7, 2, 8). For both of the groups, simulation results validate the summaries that the higher the value of M_t is, the worst the NMSE of channel estimate becomes as we expected. We used N=10 and L=5 for simulations in Fig. 3.

6.2.2 Optimal/Orthogonal Training: NMSE Performance

As depicted in Fig. 4, the performances of channel estimates based on optimal/ orthogonal training outperform those based on arbitrary training in terms of NMSE generally. The minor inconsistency for linear MMSE at larger antenna numbers is due to the fact that the FDE-based trainings already enjoy high diversity plus partial channel information and that (5.31) is a rough approximation. In general, the results verify the summaries in chapter 5. Here we simply make N=K for simulations in Fig. 4 due to the fact that (5.16) can be easily satisfied in this situation, as we explained in the previous chapters. Note that we set L=2 for both (3, 2, 4) and (6, 2, 8) system configurations.

Chapter 7

Conclusion

In this thesis, we propose an instructive derivation for the generalized block-level orthogonal space-time block encoder, capable of achieving full spatial diversity via frequency-selective fading environment, provided that channel order is known. Instead of dealing with special case and then extending the results intuitively, we provide an alternative by starting with the general signal model with multiple transmit and multiple receive antennas, from which a general form of block-level orthogonality is established. In particular, transmit diversity with more than two transmit antennas can be achieved without compromise by means of frequency-domain equalization, in contrast to the QO-STBC-based approach. However, the cost is that the proposed scheme only has nearly 1/2 symbol rate when discarding CP overheads. Pairwise error probability analysis are derived, under certain assumption which is numerically supported by simulation results, for analytical verifications of our claim on full diversity, inclusive of transmit-receive diversity and the multipath one. Hence, we are able to counterbalance the deduction[26] that the "CP-only" scheme based on the GOSTBC extension cannot exploit full multipath diversity. It is seen from the simulation results that the proposed scheme does stand a big chance of delivering full multipath diversity. Moreover, the encoder structure enables us to generalize a training-based channel estimation technique, originally proposed for flat-fading scenario, to the frequency-selective fading scenario. Surprisingly we even obtain similar optimality criteria for optimal training block design which in our case, the signal block are fixed as OSTBC-based and the design derivation reduces to derive optimal power constraint over the training blocks. The optimality criteria for the training blocks are easy to satisfy when randomness of signal constellation is not a concern. Simulation results validate our discussion of the behaviors of the least-squares and linear MMSE channel estimates. In contrast to the involved derivations provided in [16], which considers general training blocks for MIMO frequency-selective fading, we provide an alternative by generalizing the work in [3] to achieve the same purpose in a straightforward manner by fixing training block structure as BGOSTBC.

Appendix

A1. Proof of Block-Level Orthogonality part I

Note that the constituent matrices have the following properties:

$$\begin{cases} \sum_{m=1}^{M_t}\sum_{\substack{n=1\\ n\neq m}}^{M_t} \mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{A_l}(:,n) = \sum_{m=1}^{M_t}\sum_{\substack{n=1\\ n\neq m}}^{M_t} \mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{B_l}(:,n) = 0, \text{ for } 1 \le k \neq l \le K. \\ \mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{A_l}(:,n) = \mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{B_l}(:,n) = 1, \ k = l \text{ and } m = n. \\ \mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{B_l}(:,n) = 0, \ \forall \ k,l,m,n. \end{cases}$$

$$\begin{aligned} \therefore\ & \Lambda_j(:,(k-1)N : kN-1)^H \Lambda_j(:,(l-1)N : lN-1) \\ &= \sum_{m=1}^{M_t}\sum_{n=1}^{M_t} \Big\{ \big(\mathbf{X}_{A_k}(:,m)^T \otimes \Lambda_{jm}^H + \mathbf{X}_{B_k}(:,m)^T \otimes \Lambda_{jm}\big) \\ &\qquad \big(\mathbf{X}_{A_l}(:,n) \otimes \Lambda_{jm} + \mathbf{X}_{B_l}(:,n) \otimes \Lambda_{jm}^H\big) \Big\} \\ &= \sum_{m=1}^{M_t}\sum_{n=1}^{M_t} \Big\{ \big(\mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{A_l}(:,n)\big) \otimes \big(\Lambda_{jm}^H \Lambda_{jm}\big) + \\ &\qquad \big(\mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{B_l}(:,n)\big) \otimes \big(\Lambda_{jm}^H \Lambda_{jm}^H\big) + \\ &\qquad \big(\mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{A_l}(:,n)\big) \otimes \big(\Lambda_{jm} \Lambda_{jm}\big) + \\ &\qquad \big(\mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{B_l}(:,n)\big) \otimes \big(\Lambda_{jm} \Lambda_{jm}^H\big) \Big\} \\ &= 2\sum_{m=1}^{M_t} \Lambda_{jm}^H \Lambda_{jm}, \text{ for } 1 \le k = l \le K. \end{aligned}$$

$\therefore$ Λ_j is a matrix with orthogonal *block - wise* column, e.g.,
$\Lambda_j(:,(l-1)N : lN-1)$ is $l^{\underline{th}}$ block-wise column of Λ_j,
which is denoted by $\Lambda_j^{(l)}$.

$$\Rightarrow \Lambda_j^H \Lambda_j = \begin{bmatrix} \big(\Lambda_j^{(1)}\big)^H \\ \big(\Lambda_j^{(2)}\big)^H \\ \vdots \\ \big(\Lambda_j^{(K)}\big)^H \end{bmatrix} \begin{bmatrix} \Lambda_j^{(1)} & \Lambda_j^{(2)} & \cdots & \Lambda_j^{(K)} \end{bmatrix} = 2\left(\mathbf{I}_K \otimes \sum_{m=1}^{M_t} \Lambda_{jm}^H \Lambda_{jm} \right).$$

A2. Proof of block-level orthogonality part II

By the same orthogonalities used in Appendix A1.

$$\begin{aligned}
&\Lambda_T(:,(m-1)N:mN-1)^H \Lambda_T(:,(n-1)N:nN-1) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\Big\{\Big(\mathbf{X}_{A_k}(:,m)^T \otimes \Lambda_{\mathbb{X}_k}^H + \mathbf{X}_{B_k}(:,m)^T \otimes \Lambda_{\mathbb{X}_k}\Big) \\
&\qquad\qquad \Big(\mathbf{X}_{A_l}(:,n) \otimes \Lambda_{\mathbb{X}_l} + \mathbf{X}_{B_l}(:,n) \otimes \Lambda_{\mathbb{X}_l}^H\Big)\Big\} \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\Big\{\Big(\mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{A_l}(:,n)\Big) \otimes \Big(\Lambda_{\mathbb{X}_k}^H \Lambda_{\mathbb{X}_l}\Big) + \\
&\qquad\qquad \Big(\mathbf{X}_{A_k}(:,m)^T \mathbf{X}_{B_l}(:,n)\Big) \otimes \Big(\Lambda_{\mathbb{X}_k}^H \Lambda_{\mathbb{X}_l}^H\Big) + \\
&\qquad\qquad \Big(\mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{A_l}(:,n)\Big) \otimes \Big(\Lambda_{\mathbb{X}_k} \Lambda_{\mathbb{X}_l}\Big) + \\
&\qquad\qquad \Big(\mathbf{X}_{B_k}(:,m)^T \mathbf{X}_{B_l}(:,n)\Big) \otimes \Big(\Lambda_{\mathbb{X}_k} \Lambda_{\mathbb{X}_l}^H\Big)\Big\} \\
&= 2\sum_{k=1}^{K} \Lambda_{\mathbb{X}_l}^H \Lambda_{\mathbb{X}_k}, \text{ for } 1 \le m = n \le M_t.
\end{aligned}$$

$\Rightarrow \Lambda_T$ is a matrix with orthogonal *block - wise* column, e.g., $\Lambda_T(:,(m-1)N:mN-1)$ is m^{th} block-wise column of Λ_T, which is denoted by $\Lambda_T^{(m)}$.

$$\Rightarrow \Lambda_T^H \Lambda_T = \begin{bmatrix} \left(\Lambda_T^{(1)}\right)^H \\ \left(\Lambda_T^{(2)}\right)^H \\ \vdots \\ \left(\Lambda_T^{(M_t)}\right)^H \end{bmatrix} \begin{bmatrix} \Lambda_T^{(1)} & \Lambda_T^{(2)} & \cdots & \Lambda_T^{(M_t)} \end{bmatrix} = 2\left(\mathbf{I}_{M_t} \otimes \sum_{k=1}^{K} \Lambda_{\mathbb{X}_l}^H \Lambda_{\mathbb{X}_k}\right).$$

A3. OSTBC construction

To be self-contained, we hereby review some basics about orthogonal design. Recall the OSTBC proposed by V. Tarokh [1].

$$\boldsymbol{\mathcal{G}}_{M_t}(\mathbf{x}) = \sum_{k=1}^{K} \mathbf{X}_k x_k \in \mathbb{R}^{\mathrm{T}\times\mathrm{M_t}},$$

where

$$\begin{cases} \mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \end{bmatrix}^T \in \mathbb{R}^{\mathrm{K}} \text{ is real-valued symbol block.} \\ \mathbf{X}_k \in \mathbb{R}^{\mathrm{T}\times\mathrm{M_t}} \text{ are the constituent matrices, } 1 \le k \le K. \\ \mathbf{X}_k^T \mathbf{X}_l = \begin{cases} \mathbf{I}_{M_t}, \ 1 \le k = l \le K \\ -\mathbf{X}_l^T \mathbf{X}_k, \ 1 \le k \ne l \le K \end{cases}. \\ \mathrm{T} : \text{the block time length of } \boldsymbol{\mathcal{G}}_{M_t}(\mathbf{x}). \end{cases}$$

It is easy to verify that $\boldsymbol{\mathcal{G}}_{M_t}^T(\mathbf{x})\boldsymbol{\mathcal{G}}_{M_t}(\mathbf{x}) = \sum_{k=1}^{K} |x_k|^2 \mathbf{I}_{M_t}$. As stated in [1], OSTBC exists in various dimensions smaller than or equal to 8, so does its complex counterpart, GOSTBC. Therefore the subsequent block-level discussion applies whenever the corresponding symbol-level GOSTBC design exists. The construction of GOSTBC is as follows:

$$\mathbf{G}_{M_t}(\mathbf{x}) = \sum_{k=1}^{K} \mathbf{X}_{A_k} x_k + \mathbf{X}_{B_k} x_k^* \ \in \mathbb{C}^{2\mathrm{T}\times\mathrm{M_t}},$$

where

$$\begin{cases} \mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \end{bmatrix}^T \in \mathbb{C}^{\mathrm{K}} \text{ is complex-valued symbol block.} \\ \mathbf{X}_{A_k}, \mathbf{X}_{B_k} \in \mathbb{R}^{2\mathrm{T}\times\mathrm{M_t}} \text{ are the constituent matrices, } 1 \le k \le K. \\ \mathbf{X}_{A_k}^T \mathbf{X}_{A_l} = \begin{cases} \mathbf{I}_{M_t}, \ 1 \le k = l \le K \\ -\mathbf{X}_{A_l}^T \mathbf{X}_{A_k}, \ 1 \le k \ne l \le K \end{cases}. \\ \mathbf{X}_{B_k}^T \mathbf{X}_{B_l} = \begin{cases} \mathbf{I}_{M_t}, \ 1 \le k = l \le K \\ -\mathbf{X}_{B_l}^T \mathbf{X}_{B_k}, \ 1 \le k \ne l \le K \end{cases}. \\ \mathbf{X}_{A_k}^T \mathbf{X}_{B_l} = \mathbf{0}_{M_t\times M_t}, \ 1 \le k, l \le K. \end{cases}$$

Again it is easy to verify the orthogonality, $\mathbf{G}_{M_t}^T(\mathbf{x})\mathbf{G}_{M_t}(\mathbf{x}) = 2\sum_{k=1}^{K}|x_k|^2\mathbf{I}_{M_t}$. Here, we simply adopt the complex orthogonal design introduced in [1], where $\mathbf{G}_{M_t}(\mathbf{x})$ is closely related to $\boldsymbol{\mathcal{G}}_{M_t}(\mathbf{x})$ by $\mathbf{G}_{M_t}(\mathbf{x}) = \left[\boldsymbol{\mathcal{G}}_{M_t}^T(\mathbf{x}) \quad \boldsymbol{\mathcal{G}}_{M_t}^T(\mathbf{x}^*)\right]^T$. Here $\boldsymbol{\mathcal{G}}_{M_t}(\mathbf{x})$ is an OSTBC design having identical constituent matrices as the aforementioned ones but taking complex objectives in lieu of real-valued ones. It can be shown that the construction of $\mathbf{X}_{A_i}$ is basically $\mathbf{X}_i$ concatenated with an all-zero matrix of the same dimension, i.e., $\mathbf{X}_{A_i} = \left[\mathbf{X}_i^T \quad \mathbf{0}_{M_t \times T}\right]^T$ and $\mathbf{X}_{B_i} = \left[\mathbf{0}_{M_t \times T} \quad \mathbf{X}_i^T\right]^T$. Here, for notational purpose we introduce a notation $\boldsymbol{\mathcal{G}}_{A_i}$ to replace $\mathbf{X}_i$ such that $\mathbf{X}_{A_i}$ and $\mathbf{X}_{B_i}$ can be expressed as follows: $\mathbf{X}_{A_i} = \left[\boldsymbol{\mathcal{G}}_{A_i}^T \quad \mathbf{0}_{M_t \times T}\right]^T$ and $\mathbf{X}_{B_i} = \left[\mathbf{0}_{M_t \times T} \quad \boldsymbol{\mathcal{G}}_{A_i}^T\right]^T$. By replacing the ordinary multiplication for the constituent matrices and transmitted symbols with kronecker product and using index K in lieu of T, we immediately arrive at the BGOSTBC structure. For illustrational purposes in the following chapters, we give two examples:

ex.1: $T = K = 4,\ M_t = 3$

$$\boldsymbol{\mathcal{G}}_3 = \begin{pmatrix} x_1 & x_2 & x_3 \\ -x_2 & x_1 & -x_4 \\ -x_3 & x_4 & x_1 \\ -x_4 & -x_3 & x_2 \end{pmatrix}, \quad \mathbf{G}_3 = \begin{pmatrix} \boldsymbol{\mathcal{G}}_3(\mathbf{x}) \\ \boldsymbol{\mathcal{G}}_3(\mathbf{x}^*) \end{pmatrix},$$

where

$$\begin{cases} \boldsymbol{\mathcal{G}}_{A_1} = \begin{pmatrix} \mathbf{I}_3 \\ \mathbf{0}_{1\times 3} \end{pmatrix}, \ \mathbf{X}_{A_1} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_1} \\ \mathbf{0}_{4\times 3} \end{pmatrix}; \ \boldsymbol{\mathcal{G}}_{A_2} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \mathbf{X}_{A_2} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_2} \\ \mathbf{0}_{4\times 3} \end{pmatrix}; \\ \boldsymbol{\mathcal{G}}_{A_3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}, \ \mathbf{X}_{A_3} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_3} \\ \mathbf{0}_{4\times 3} \end{pmatrix}; \ \boldsymbol{\mathcal{G}}_{A_4} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \ \mathbf{X}_{A_4} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_4} \\ \mathbf{0}_{4\times 3} \end{pmatrix}; \\ \mathbf{X}_{B_i} = \begin{pmatrix} \mathbf{0}_{4\times 4} & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{0}_{4\times 4} \end{pmatrix} \mathbf{X}_{A_i}, \ 1 \le i \le 4. \end{cases}$$

ex.2: $T = K = 8,\ M_t = 5$

$$\boldsymbol{\mathcal{G}}_5 = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ -x_2 & x_1 & x_4 & -x_3 & x_6 \\ -x_3 & -x_4 & x_1 & x_2 & x_7 \\ -x_4 & x_3 & -x_2 & x_1 & x_8 \\ -x_5 & -x_6 & -x_7 & -x_8 & x_1 \\ -x_6 & x_5 & -x_8 & x_7 & -x_2 \\ -x_7 & x_8 & x_5 & -x_6 & -x_3 \\ -x_8 & -x_7 & x_6 & x_5 & -x_4 \end{pmatrix},\ \mathbf{G}_5 = \begin{pmatrix} \boldsymbol{\mathcal{G}}_5(\mathbf{x}) \\ \boldsymbol{\mathcal{G}}_5(\mathbf{x}^*) \end{pmatrix},$$

where

$$\begin{cases} \boldsymbol{\mathcal{G}}_{A_1} = \begin{pmatrix} \mathbf{I}_5 \\ \mathbf{0}_{3\times 5} \end{pmatrix},\ \mathbf{X}_{A_1} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_1} \\ \mathbf{0}_{8\times 5} \end{pmatrix};\ \boldsymbol{\mathcal{G}}_{A_2} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},\ \mathbf{X}_{A_2} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_{A_2} \\ \mathbf{0}_{8\times 5} \end{pmatrix}; \\ \mathbf{X}_{A_3},\ \mathbf{X}_{A_4}, \cdots,\ \mathbf{X}_{A_8} \text{ can be visualized as well and } \mathbf{X}_{B_i} = \begin{pmatrix} \mathbf{0}_{8\times 8} & \mathbf{I}_8 \\ \mathbf{I}_8 & \mathbf{0}_{8\times 8} \end{pmatrix} \mathbf{X}_{A_i},\ 1 \le i \le 8. \end{cases}$$

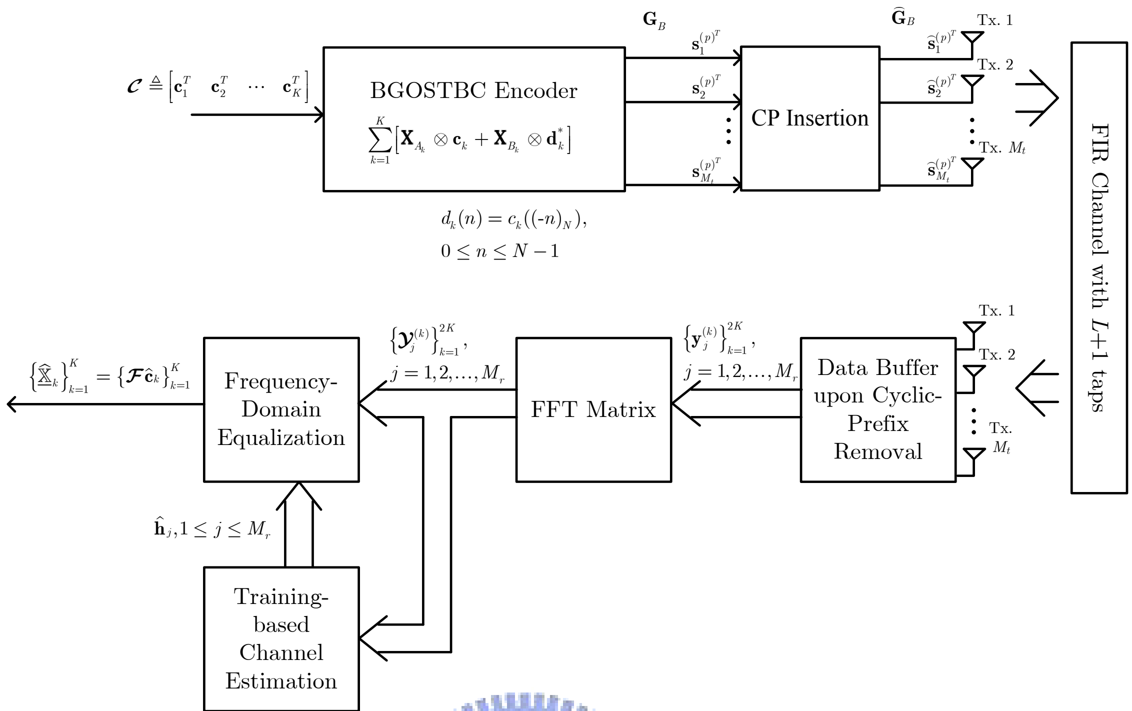

Fig. 1 Transmission scheme based on BGOSTBC encoder and the associated FDE based on training-based channel estimation.

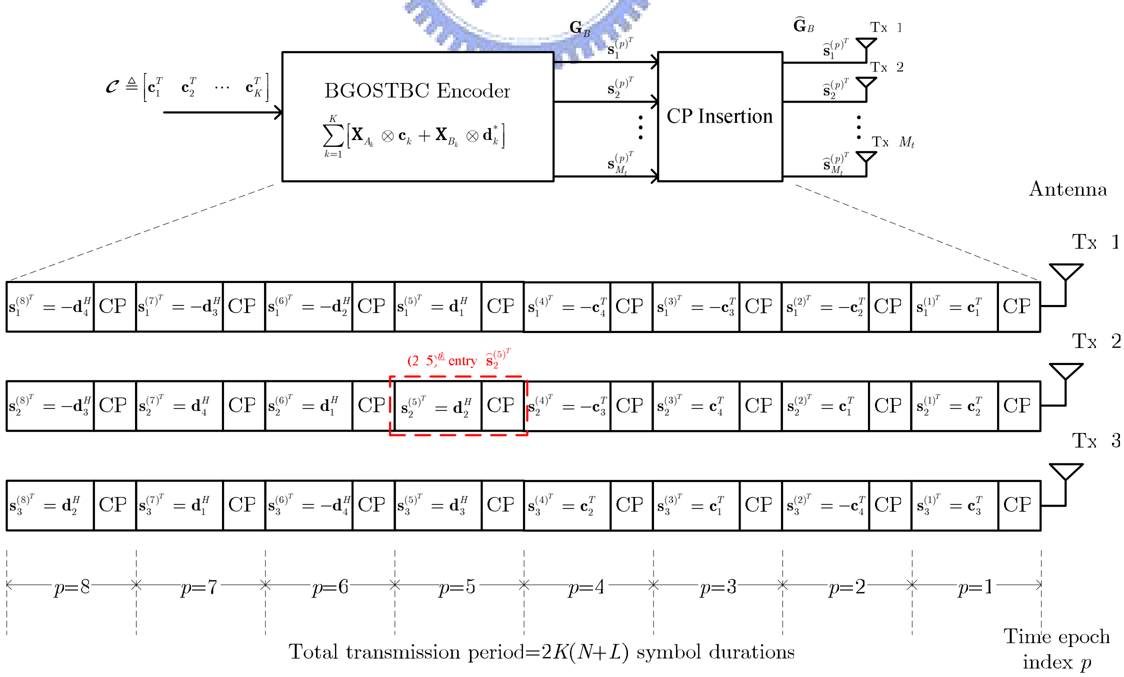

Fig. 2 A BGOSTBC encoder employing K=4, M_t=3 configuration.

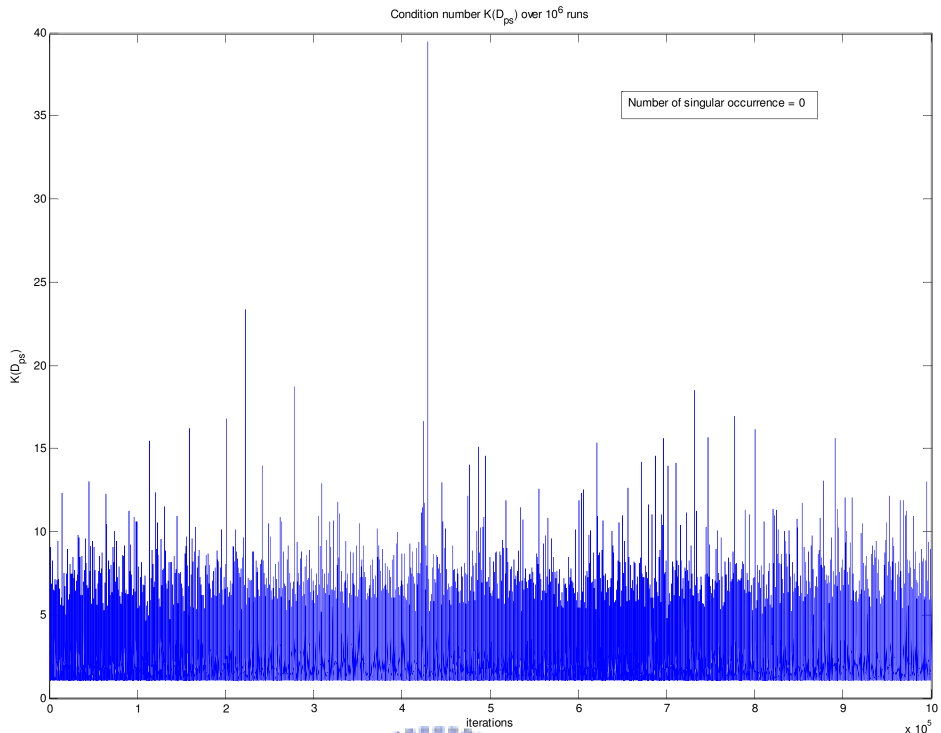

Fig. 3-1 Condition numbers generated for 10^6 independent simulation runs under (N, M_r, M_t, L, K)=(10,1,3,1,4).

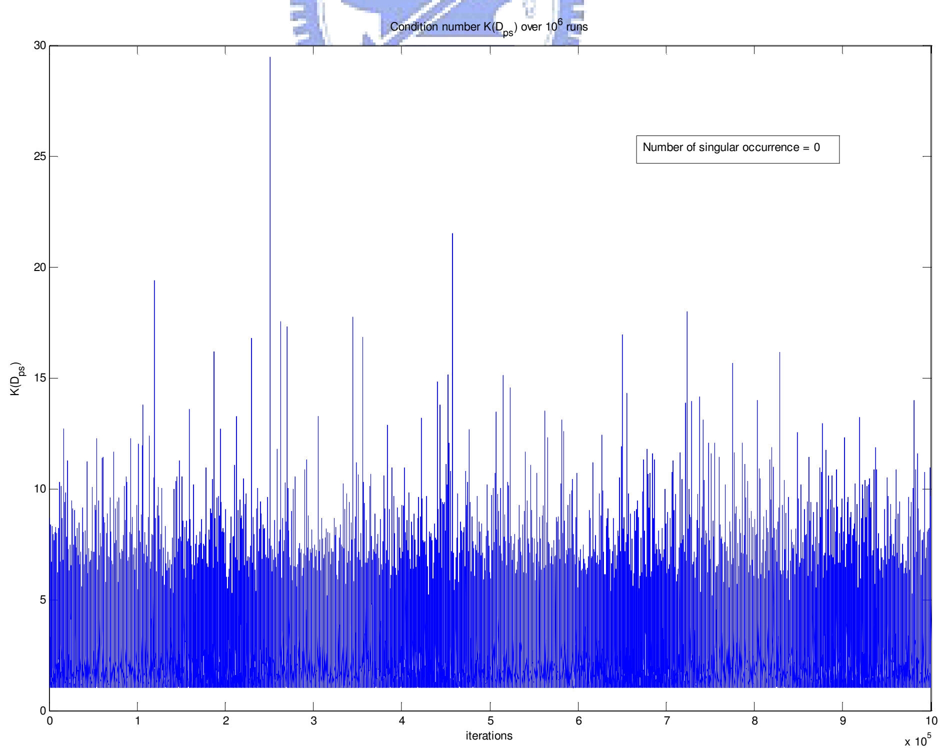

Fig. 3-2 Condition numbers generated for 10^6 independent simulation runs under (N, M_r, M_t, L, K)=(16,1,3,1,4).

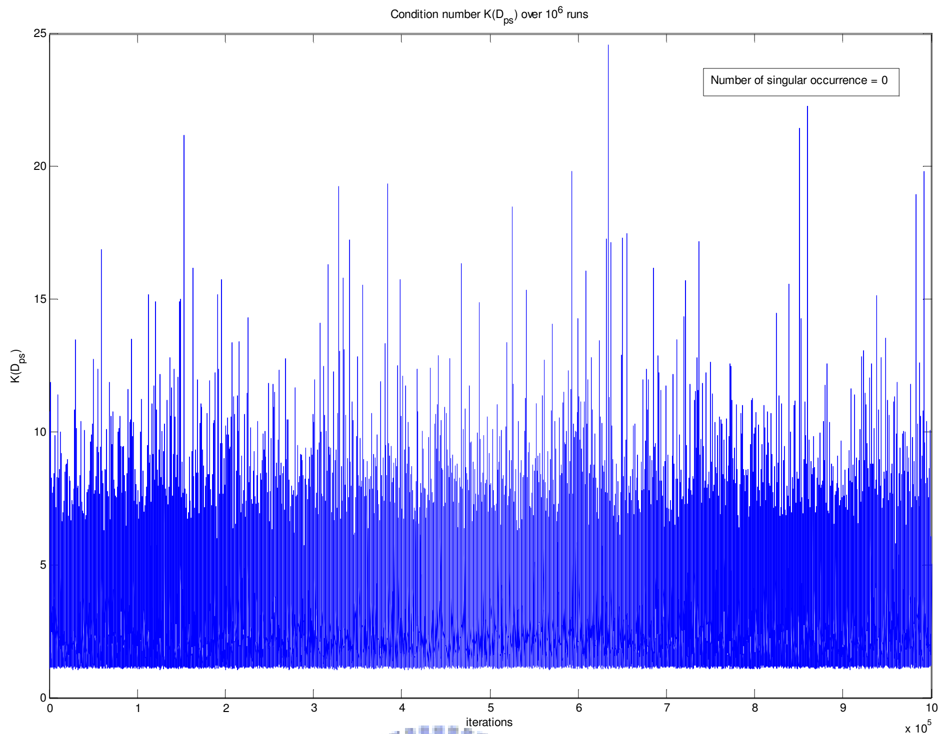

Fig. 3-3 Condition numbers generated for 10^6 independent simulation runs under (N,M_r,M_t,L,K)=(16,1,3,2,4).

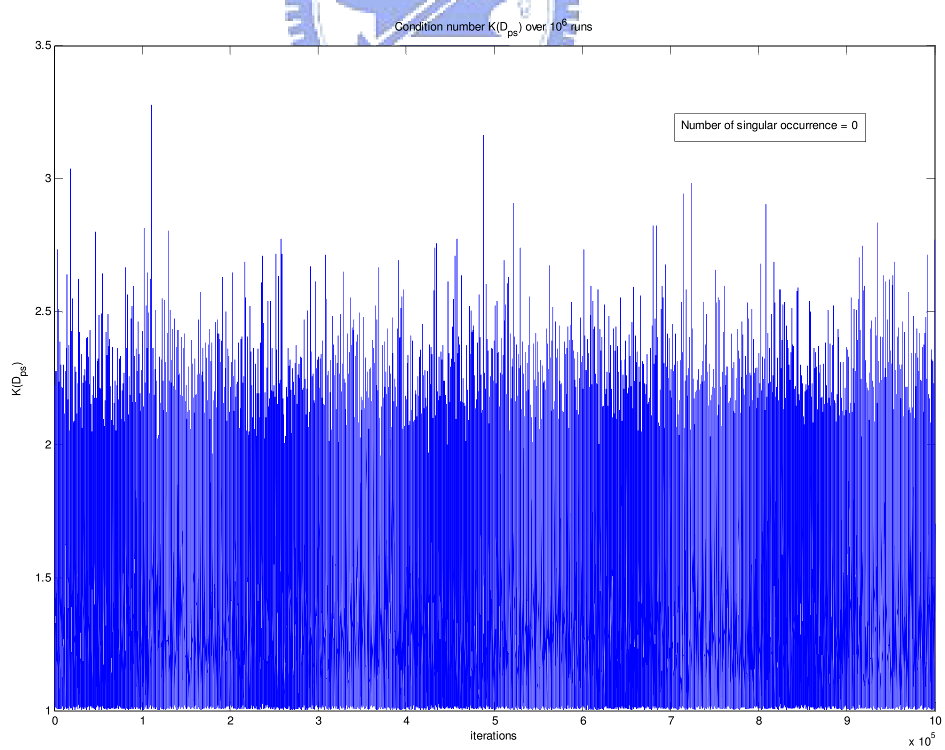

Fig. 3-4 Condition numbers generated for 10^6 independent simulation runs under (N,M_r,M_t,L,K)=(10,2,6,1,8).

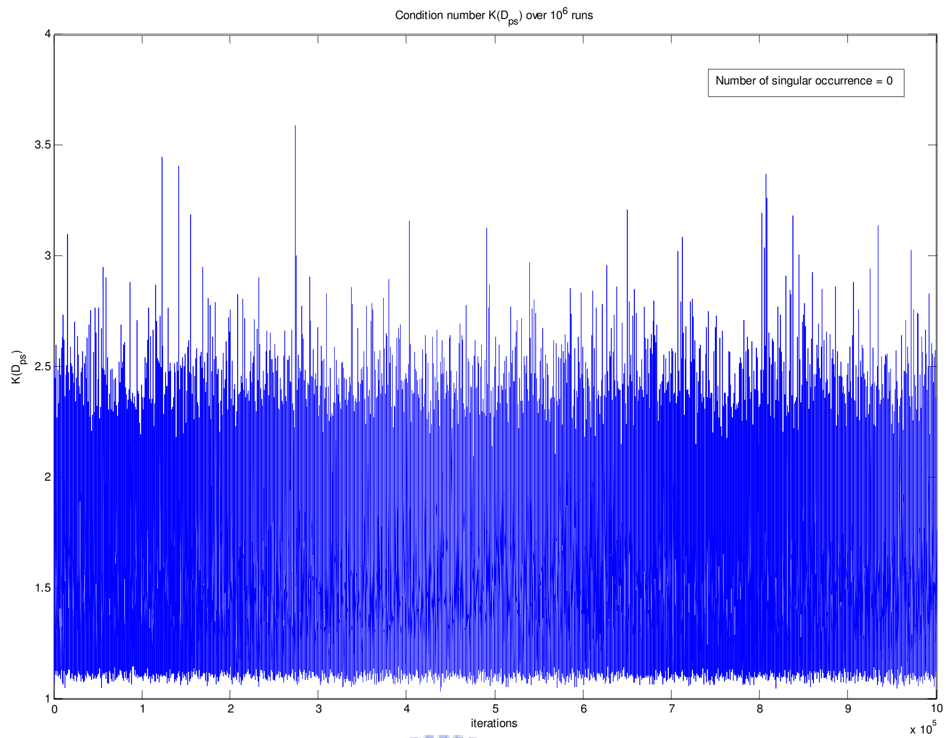

Fig. 3-5 Condition numbers generated for 10^6 independent simulation runs under (N,M_r,M_t,L,K)=(10,2,6,3,8).

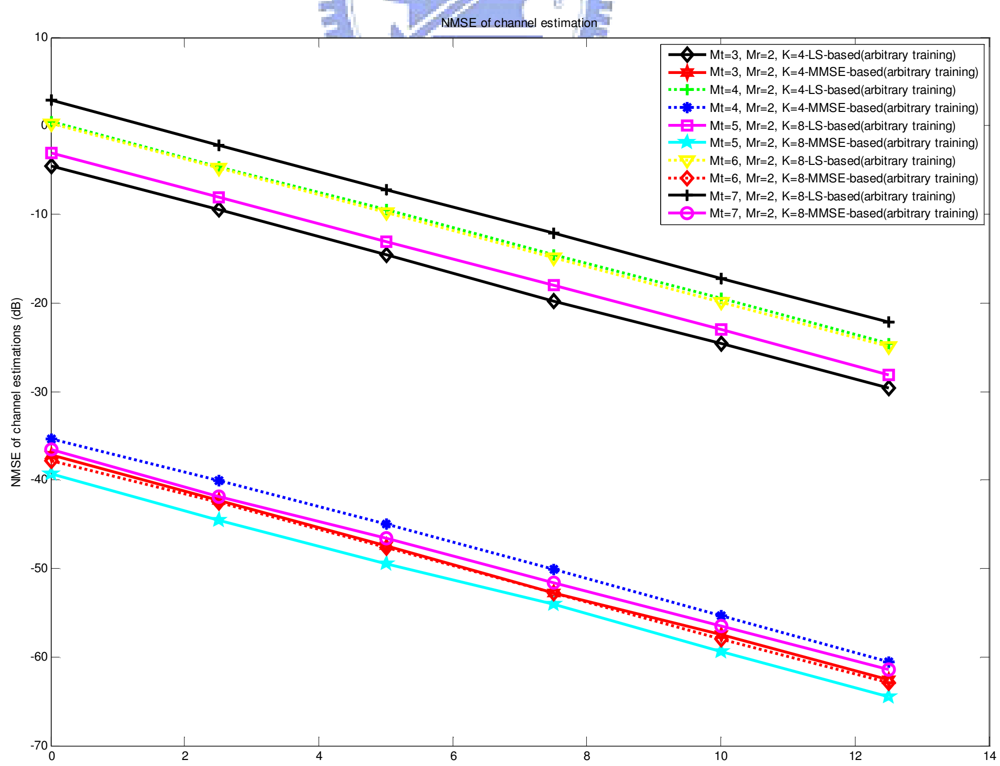

Fig. 3 Compare NMSE of LS and linear MMSE estimates, with arbitrary training.

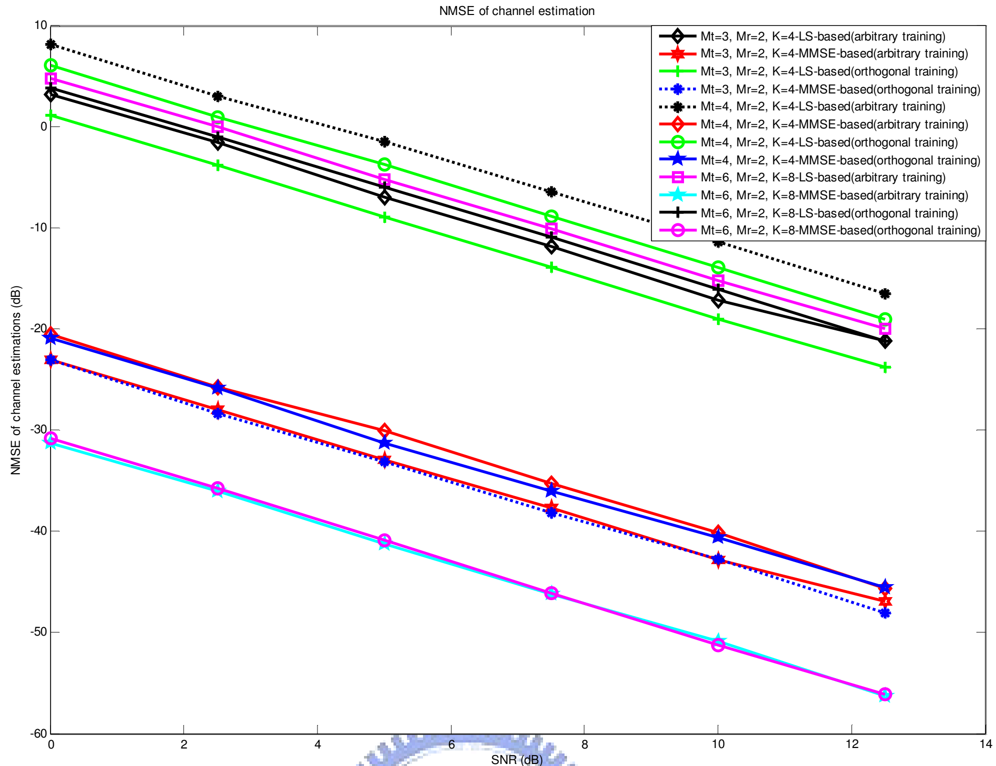

Fig. 4 Compare NMSE of arbitrary training and optimal/orthogonal training.

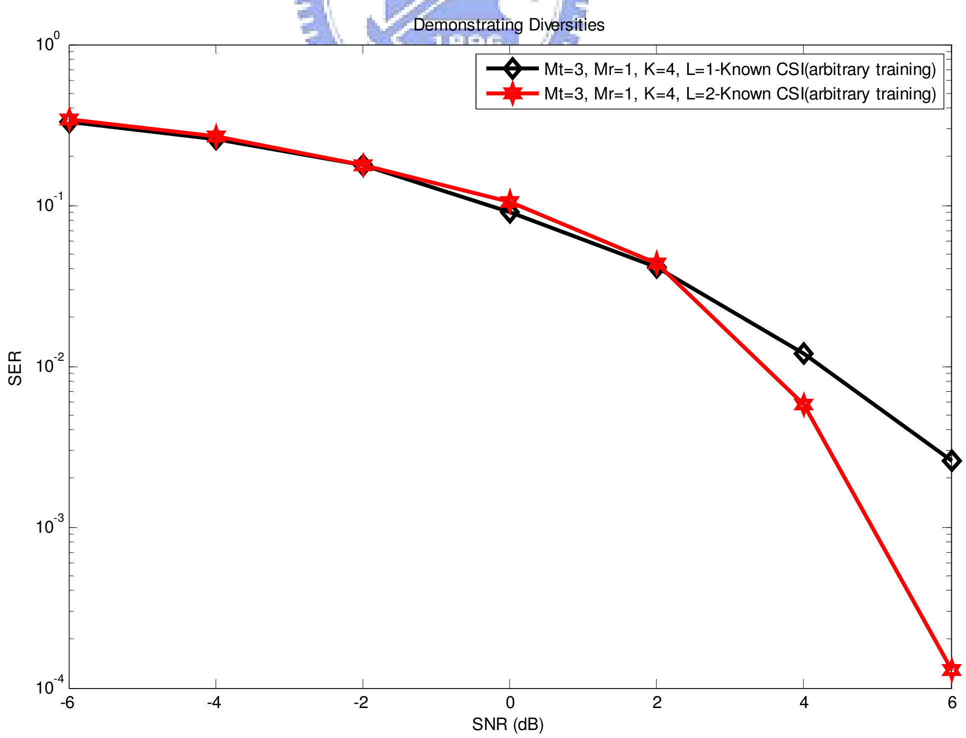

Fig. 5 Demonstration of full multipath diversity at high SNR region, averaged over 100 independent runs.

References

[1] V. Tarokh, H. Jafarkhani and A. R. Calderbank, “Space-Time Block Codes from Orthogonal Designs”, *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456-1467, Jul. 1999.

[2] H. Mheidat, M. Uysal and N. Al-Dhahir, “Time- and Frequency-Domain Equalization for Quasi-Orthogonal STBC over Frequency-Selective Channels”, in *Proc. IEEE 2004 Int. Conf. Commun.*, vol. 2, pp. 697-701, Jun. 2004.

[3] M. Biguesh and Alex B. Gershman, “Training-Based MIMO channel Estimation: A study of Estimator tradeoffs and Optimal training Signals”, *IEEE Trans. Signal Process.*, vol. 54, pp. 884-893, Mar. 2006.

[4] Qi Ling and Tongtong Li, “Blind Channel Estimation for MIMO Systems with Structured Transmit Delay Diversity”, in *Proc. IEEE 2005 Global Commun. Conf.*, vol. 5, pp. 2964-2968, Nov-Dec. 2005.

[5] A. Graham, *Kronecker Products and Matrix Calculus: with Applications*. West Sussex, U.K.: Ellis Horwood Limited, 1981.

[6] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[7] E. Lindskog and A. Paulraj, “A Transmit Diversity Scheme for Channel with Intersymbol Interference”, *ICC 00*, p. 307-311, Jun. 2000.

[8] S. M. Alamouti, “A Simple Transmit Diversity Scheme for Wireless Communications”, *IEEE JSAC*, p. 1451-1458, Oct., 1998.

[9] N. Al-Dhahir, “Single-carrier frequency-domain equalization for space-time block coded transmission over frequency-selective fading channels”, *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 304-306, Jul. 2001.

[10] H. Jafarkhani, “A Quasi-Orthogonal Space-Time Block Code”, *IEEE Trans. Commun.*, vol. 49, pp. 1-4, Jan. 2001.

[11] J. Y. Wu and T. S. Lee, “Non-Redundant Precoding Assisted Blind Channel Estimation for Single-Carrier Space-Time Block Coded Transmission with Frequency-Domain Equalization”, accepted by *IEEE*, 2005.

[12] F. Pancaldi and Giorgio M. Vitetta, “Frequency-Domain Equalization for Space-Time Block-Coded Systems”, *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2907-2916, Nov. 2005.

[13] K. Takeda, T. Itagaki and F. Adachi, “Application of space-time transmit diversity to single-carrier transmission with frequency-domain equalization and receive antenna diversity in a frequency-selective fading channel”, in *Proc. IEE Commun.*, vol. 151, no.6, pp. 627-632, Dec. 2004.

[14] Z. Liu and G. B. Giannakis, “Space-Time Block-Coded Multiple Access Through Frequency-selective Fading Channels”, *IEEE Trans. Commun.* , vol. 49, no. 6, pp. 1033-1044, Jun. 2001.

[15] Xiaoli Ma, G. B. Giannakis and Shuichi Ohno, “Optimal training for Block Transmission Over Doubly Selective Wireless Fading Channels”, *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351-1366, May 2003.

[16] Xiaoli Ma, Liuqing Yang and G. B. Giannakis, “Optimal training for MIMO Frequency-Selective Fading Channels”, *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453-466, Mar. 2005.

[17] C. Tepedelenlioğlu, “Maximum multipath diversity with linear equalization in precoded OFDM systems”, *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 232-234, Jan. 2004.

[18] David Tse and Pramod Viswanath, *Fundamentals of Wireless Commmunication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[19] W. K. Ma, B. N. Vo, Timothy. N. Davidson and P. C. Ching, “Blind ML Detection of Orthogonal Space-Time BlockCodes: Efficient High-Performance Implementations”, *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 738-751, Feb. 2006.

[20] S. Shahbazpanahi, Alex B. Gershman and H. Manton, “Closed-form blind MIMO channel estimation for orthogonal space-time block codes”, *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 738-751, Dec. 2005.

[21] C. Helmberg, F. Rendl, R. J. Vanderbei and H. Wolkowicz, “An Interior-Point Method for Semidefinite Programming”, *SIAM, J. Optimiz.,* vol. 6, no. 2, pp. 342 -361, 1996.

[22] A. L. Swindlehurst and G. Leus, “Blind and semi-blind equalization for generalized space-time block codes,” *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2589–2498, 2002.

[23] M. Gharavi–Alkhansari and A. B. Gershman, “Constellation space invariance of orthogonal space–time block codes,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 331–334, Jan. 2005.

[24] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. New Jersey,

U.S.: Ellis Prentice-Hall, Inc., 1999.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] S. Zhou and G. B. Giannakis, "Single-Carrier Space-Time Block-Coded Transmissions Over Frequency-Selective Fading Channels", *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 164-179, Jan. 2003.

[27] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.