國　立　交　通　大　學

電子工程學系　電子研究所碩士班

碩　士　論　文

在多天線系統下使用 Kalman 濾波器之

Tomlinson-Harashima 前置編碼設計

Design of Tomlinson-Harashima Precoding in MIMO Systems Using Kalman Filtering

研 究 生: 丁琬瑜

指導教授: 簡鳳村 博士

中 華 民 國 一〇〇 年 七 月

在多天線系統下使用 Kalman 濾波器之 Tomlinson-Harashima 前置編碼設計

Design of Tomlinson-Harashima Precoding in MIMO Systems Using Kalman Filtering

研 究 生: 丁琬瑜　　Student: Wan-Yu Ting

指導教授: 簡鳳村 博士　　Advisor: Dr. Feng-Tsun Chien

國 立 交 通 大 學

電子工程學系　　電子研究所碩士班

碩 士 論 文

A Thesis

Submitted to Department of Electronics Engineering & Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Electronics Engineering

July 2011

Hsinchu, Taiwan, Republic of China

中 華 民 國 一 ○ ○ 年 七 月

在多天線系統下使用 Kalman 濾波器之 Tomlinson-Harashima 前置編碼設計

研究生：丁琬瑜　　指導教授：簡鳳村 博士

國立交通大學

電子工程學系　電子研究所碩士班

摘要

在本篇論文中， 我們研究了如何在多天線(MIMO)系統下運用 Kalman 濾波器來對通道做追蹤估計，並且結合 Tomlinson-Harashima 前置編碼器來做通道等化的設計。在隨時間和距離變化而持續變動的多天線系統底下，多天線的 Tomlinson-Harashima 前置編碼器是一項可以用來消除不同訊號流之間的干擾的技術，這項技術是髒紙編碼(Dirty paper coding)的延伸應用，和其他等化技術不同的地方是，它可以在保持通道容量不變的前提下完成等化。而 Kalman 濾波器是一個運用了隨機過程觀念所延伸出來的估計方法，和其他估計技術不同的是，它是將以前到現在對要估計的變數所做的所有觀察集合起來當作估計的參考，而非只用要估計的當時所觀察到的資訊來做估計，因此可以獲得較為準確的結果。在假設當中，Tomlinson-Harashima 前置編碼器必須在傳輸端和接收端都完整的知道通道資訊的前提下才會是完美的，而這個前提卻是不實際的假設。因此，為了更貼近現實的情況，我們研究了在只有部分通道資訊的前提下，運用 Kalman 濾波器來估計通道並且在把通道估計誤差也考慮進去的狀況下來做 Tomlinson-Harashima 前置編碼器的

最佳化系統設計。在模擬結果的部分，我們比較了在 TDD 的系統下本篇論文所提出的方法和用線性最小均方差(LMMSE)估計法來結合 Tomlinson-Harashima 前置編碼器的結果，發現使用 Kalman 濾波器的 BER 會表現得比使用線性最小均方差(LMMSE)估計法還要好，並且當都卜勒速度改變時使用 THP-Kalman 會比 THP-LMMSE 還來得不易受影響。最後，我們也比較了在不同的模擬假設情形下 THP-Kalman 和 THP-LMMSE 的計算複雜度差異。

Design of Tomlinson-Harashima Precoding in MIMO Systems Using Kalman Filtering

Student: Wan-Yu Ting Advisor: Dr. Feng-Tsun Chien

Department of Electronics Engineering

Institute of Electronics

National Chiao Tung University

Abstract

In this thesis, we study the problem of combining Kalman filter for channel tracking and Tomlinson-Harashima precoding for channel equalization in MIMO systems. The multiple-input multiple-output (MIMO) Tomlinson-Harashima precoder (THP) is a well known equalization structure for mitigating inter-stream interference in fading MIMO systems, which is the application of “dirty paper coding“ and can reserve the channel capacity. Kalman filter is an estimator based on the conception of random process, compare to other estimators, Kalman filter collects all previous channel information for estimating, yet the other estimators estimate the variables by considering the present observation. THP is optimal by assuming perfect channel state information (CSI) at both transmitter and receiver. However, this assumption is not achievable in real world. In this work, under the assumption of partial channel state information (P-CSI), we use Kalman estimation for channel tracking and combine the estimation into THP optimization which considers the channel estimation error. In simulation results, we compare the proposed approach with earlier works and can show that the performance (BER) of THP system with Kalman estimation (THP-Kalman) for channel tracking is superior to THP with linear-minimum-mean-square-error (LMMSE) estimator (THP-LMMSE) in time-division-duplex (TDD) system. Besides, we also compare the computation

complexity of THP-Kalman and THP-LMMSE. By changing the Doppler rate (the parameter of mobility), THP-Kalman performs more flexible, while THP-LMMSE is sensitive to the varying rate of channel.

誌謝

這篇論文能夠順利完成，首先我要感謝我的指導教授簡鳳村老師，從我大四下學期開始就給我許多的指導，引領我進入通訊系統的領域。因為老師的細心教導和在專業領域的博學精深，讓我學習到不少研究上的方法和精神。而除了專業之外，也謝謝老師在報告以及製作簡報上面教給我許多專業的技巧，讓我的研究所生涯獲得的不只是課業上的進展，還有別的地方學習不到的表達能力訓練。

此外，感謝通訊電子與訊號處理實驗室所有的成員，包含各位師長、學長姐、同學和學弟妹們。感謝劉藹璇學姐、李重佑學長、邱頌恩學長和張傑堯學長給予我在研究上的指導與建議，以及怡茹、頌文、郁婷、俊言、兆軒、曉盈、威宇、卓翰、智凱、強丹、書緯、偵源、凱翔和復凱等同學願意分享研究和生活上的心得和建議，陪我度過這兩年的研究所生活。

最後，還要謝謝一直在背後強力支持我的父母以及願意無條件給予我幫助的哥哥，讓我在這六年中無後顧之憂來學習和完成我想做的事，你們永遠是我精神上最大的支柱。

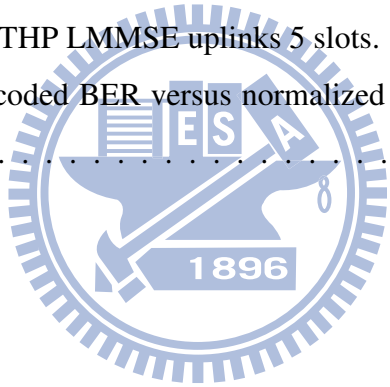
在此，將此篇論文獻給所有愛我和我愛的人。

丁琬瑜

民國一〇〇年七月 於新竹

Contents

List of Figures



Chapter 1

Introduction

1.1 Motivation

"MIMO broadcast channel" is a communication scenario with multiple cooperating transmitters (central base station), which can transmit a joint preprocessing of the signals, and multiple decentralized receivers (non-cooperating mobile stations), which process the received signals independently. MIMO techniques have been an important research topic due to their potential for high capacity, increased diversity and interference restrain. However, the parallel transmission of independent data streams introduces severe inter-stream interference (ISI). Thus, how to deal with the ISI in the MIMO systems has been one of the most important research topics in modern communication systems.

Over the last years, many transmitter and receiver structures for mitigating interference have been proposed, achieving various levels of performance with varying complexities. Equalization strategies for multi-antenna and multi-user transmission are studied recently, including linear equalizations and nonlinear equalizations. Common examples for linear equalizers include linear (pre)equalizer [2] and singular-value-decomposition (SVD) based equalizer [3][22], and the most famous nonlinear equalizer is perhaps the decision-feedback equalizer (DFE) [4] [5]. The DFE is an equalization technique at the receiver side that is easy to implement, but suffers from the drawback of possible error propagations. Recently, Tomlinson-Harashima Precoder (THP) has emerged as a feasible approach to maintain the channel capacity for eliminating ISI. The concept of THP was

first introduced by Tomlinson [6] and Harashima/Miyakawa [7] for single-input single-output (SISO) inter-symbol-interference (ISI) systems, which can be seen as the dual to DFE, i.e., moving the detection structure (feedback part) from the receiver side to the transmitter side. While the DFE feeds back already quantized symbols, the already precoded symbols are fed back in a THP system and modulo operations are applied to constrain the precoded symbol power at the transmitter. By moving the feedback part to the transmitter side, THP can overcome the main shortcoming of DFE,i.e., error propagations. Furthermore, adopting THP at the transmitter can achieve better bit-error-rate (BER) than the linear equalizers and DFE, as shown in [8].

On the other hand, THP can be considered as the simplest practical approximation of the "dirty paper coding" (DPC) [9]. In the DPC-based structure, on the condition of perfect CSI at the transmitter, the optimal transmitter adapts the transmit signal to the interference rather than cancel it. It has shown that the channel capacity is not decreased by the additional interference known at the transmitter. However, in the real world, perfect CSI is not available. Numerous researches have introduced different kinds of methods to obtain the channel information, and many estimation approaches have used to calculate the channel coefficients. More details will be described in Section 1.2.

In this work, to approach the real world situation, we are interested in the Tomlinson-Harashima precoding strategies with partial channel-state-information for multi-antenna and multi-user transmission in wireless communication systems with time-varying Rayleigh channels. In most studies, the accuracy of time-varying CSI is limited by the available number of past training sequences. To improve the quality of CSI, we adopt the idea of Kalman filtering to track the time-varying channel coefficients.

1.2 Related Work

Over the past years, the application of THP had been combined into MIMO system and numerous research studies have been working on this issue. The temporal inter-symbol interference mitigation of THP is applied in the $broadcast$ channel (i.e. point to multi-point transmission) by Ginis [10] and Yu [11]. Further, THP is applied into multiple-

input multiple-output (MIMO) systems by Fischer [12]. Also, he compares other popular equalization techniques with THP in MIMO system as in [8]. The optimization of THP in these research work are based on zero-forcing (ZF-THP) in frequency flat channels. Later, minimum mean-squared error based THP (MMSE-THP) is proposed in [13] in frequency selective channels, and optimum precoding order is combined into MMSE-THP in [14]. The maximum achievable information rate for ZF-THP and MMSE-THP is studied in [15]. Further, Lagunas [16] proposes a generalized structure of spatial THP which enables different transmission powers for each antenna, and the channel capacity bounds are also investigated. This achievable structure ensures that for variable power per transmitter, the precoding structure and its properties can be preserved. Also, an efficient algorithm for reducing the computational complexity of filter and precoding order based on symmetrically permuted Cholesky factorization is proposed in [17] and [18].

Most of the previous works assume that perfect CSI is given at the transmitter. However, since CSI uncertainties always exist in real world systems, this assumption is not realistic. Recently, systems with *mobility* in wireless communications is a major issue. In spite of THP's good performance, it is very sensitive to erroneous CSI, as the results shown in [19]. As CSI at the transmitter is never perfect, the system suffers from severe performance degradation. From the uplink in time-division-duplexing (TDD) systems or receivers' limited feedback, partial CSI (P-CSI) is available at the transmitter. TDD refers to a transmission scheme that allows an asymmetric flow for uplink and downlink transmission. In a TDD system, a general carrier is shared between the uplink and downlink, the resource is being switched in time. The time variations of channel and channel estimation error lead to significant outdated CSI at the transmitter in both cases. Lampe [20] considers THP without exact CSI, Dietrich [19] proposes a robust optimization for ZF-THP with erroneous CSI and use MMSE-based prediction for channel parameters, and Liavas [21] optimizes THP based on partial-CSI with cooperative receivers (which we do not discuss in our work). MMSE-THP with P-CSI has combined Kalman estimation for channel tracking and particle filtering techniques are introduced in [22]. Dietrich [23] introduces a robust optimization for MMSE-THP with P-CSI which consider channel estimation error, and a novel receiver based on CSI at the receivers is designed.

1.3 Contributions of the Research

Most of the previous works with the THP optimization methods are based on the assumption of perfect CSI. In this thesis, we relax the assumption of perfect CSI and attempt to track the time-varying channel coefficients using the Kalman filtering, which we refer to as the Kalman-THP method. By using the Kalman estimator, we can recursively update and predict the channel coefficients using all the previous training sequences from the uplink transmission in the TDD mode. Comparing to the approach proposed in [23], the proposed Kalman-THP method can achieve comparable bit error rate performance with close computational complexity when the channel is highly time-varying.

Chapter 2

Background Review

Usually, a MIMO transmission scheme is described by the basic relation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{x}$ denote the transmit vector which comprises the transmit symbols of n_T parallel data streams, and these streams are belong to different and independent users. The vectors $\mathbf{y}$ and $\mathbf{n}$ of dimension n_R designate the vector of received symbols and the vector of disturbances, respectively. The MIMO (flat-fading) channel is characterized by its $n_R \times n_T$ channel matrix $\mathbf{H}$.

The interference components of the transmitted vector $\mathbf{x}$ are present at the receiver sides. To be specific, the receive symbol of $i\text{-}th$ antenna/user is represent as

$$y_i = h_{ii}x_i + \sum_{j=1, j\neq i}^{n_T} h_{ij}x_j + n_i \tag{2.1}$$

where the second term of the equation is the $i\text{-}th$ antenna/user's interference from other users. Mathematically, after removing the interference, the ideal relation between the transmission scheme should be $\mathbf{y}' = \mathbf{x} + \mathbf{n}'$, where $\mathbf{y}'$ is the equalized signal vector. The main goal of "equalization strategies" is to eliminate the interference terms, and numerous types of equalizers had been published in the past.

In this chapter, we will first introduce the history of equalization techniques, then we'll focus on the details of Tomlinson-Harashima precoder. At the end of this chapter, autoregressive model will be discussed as the method of modeling the real channel.

2.1 Introduction of Previous Equalization Strategies

2.1.1 Linear Equalizations

As the name implies, "linear equalizer" is a techniques that remove the interference linearly, some of the popular examples are linear equalizer, linear pre-equalizer and singular value decomposition equalizer. Linear equalizer implement the equalization at the receiver side, on the contrary, linear pre-equalizer is proceed before the transmission, and singular value decomposition equalizer executes the equalization process jointly at transmitter and receiver sides. More details will be discussed as below.

Linear Equalizer

The fundamental linear equalization structure is shown in Fig. 2.1. In the figure, $\mathbf{a}[n] = [a_1[n] \dots a_{n_T}[n]]^T$ is the modulated data vector, $\mathbf{x}[n]$ is the transmit symbol vector, and $\mathbf{H}$ is the channel realization matrix, $\mathbf{n}[n]$ is the additive white Gaussian noise vector. In this case, $\mathbf{a}[n] = \mathbf{x}[n]$. The linear equalization strategy is to add an additional feedforward matrix $\mathbf{P} = \mathbf{H}_l^{-1}$ at the receiver, where $\mathbf{H}_l^{-1}$ is the left pseudo inverse of $\mathbf{H}$. Thus, the equalized signal vector is

$$\mathbf{y}'[n] = \mathbf{H}_l^{-1}\mathbf{H}\mathbf{x}[n] + \mathbf{H}_l^{-1}\mathbf{n}[n] = \mathbf{a}[n] + \mathbf{H}_l^{-1}\mathbf{n}[n] \tag{2.2}$$

As in (2.2), the interference is eliminated. However, linear equalizer suffer from noise enhancement, i.e., $\mathbf{H}_l^{-1}\mathbf{n}[n]$, and hence, lead to poor power efficiency.

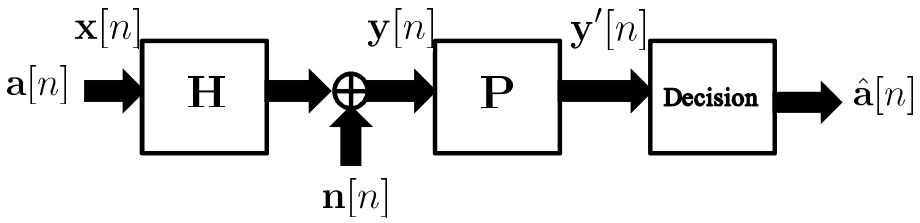


Figure 2.1: Linear Equalizer

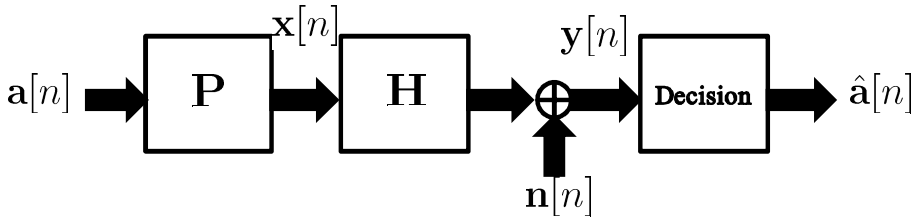


Figure 2.2: Linear Pre-equalizer

Linear Pre-equalizer [2]

The method of linear pre-equalization is the dual to linear equalizer, as in Fig. 2.2. In the condition of having the CSI at the transmitter, the data symbols are equalized prior to the transmission, i.e., $\mathbf{P} = \mathbf{H}_r^{-1}$, where $\mathbf{H}_r^{-1}$ is the right pseudo inverse of $\mathbf{H}$, hence, the transmit signal vector is $\mathbf{x}[n] = \mathbf{P}\mathbf{a}[n]$. The received signal vector in this scheme can be written as

$$\mathbf{y}[n] = \mathbf{H}\mathbf{H}_r^{-1}\mathbf{a}[n] + \mathbf{n}[n] = \mathbf{a}[n] + \mathbf{n}[n] \tag{2.3}$$

The linear pre-equalization scheme has overcome the disadvantage of linear equalization, nevertheless, it suffer from boosted transmit power, and also result in poor power efficiency.

Singular Value Decomposition Equalizer [3][22]

Since both linear equalization and linear pre-equalization have suffer from power efficiency, the SVD equalizer is presented in order to defeat the disadvantages of these previous works, the block diagram is shown in Fig. 2.3. In the condition of having the CSI at the transmitter, the channel matrix can be decomposed as [24] (singular value decom-

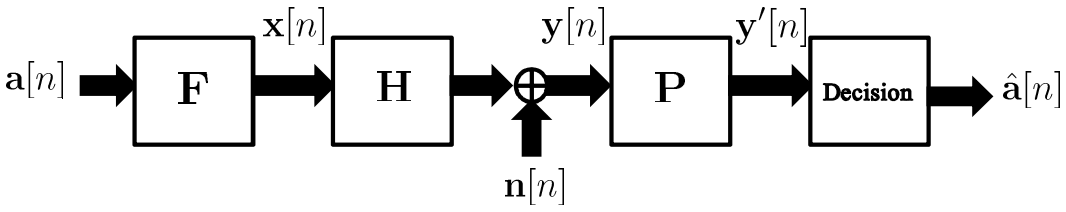


Figure 2.3: SVD Equalizer

position)

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{2.4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices which contain the eigenvectors of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{H}^H\mathbf{H}$, and the diagonal terms of $\mathbf{\Sigma} = diag(\sigma_1 \ldots \sigma_{n_R})$ are the positive square roots of the corresponding eigenvalues.

By applying $\mathbf{F} = \mathbf{V}$ at the transmitter and $\mathbf{P} = \mathbf{U}^H$ at the receiver, n_R independent and parallel sub-channels are present, and the overall signal vector $\mathbf{y}'[n]$ is

$$\mathbf{y}'[n] = \mathbf{P}\mathbf{y}[n] = \mathbf{\Sigma}\mathbf{a}[n] + \mathbf{U}^H\mathbf{n}[n] \tag{2.5}$$

Since both $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, compare to linear (pre)equalization, neither an increase of the channel noise, nor of transmit power occurs. However, as the diagonal terms of $\mathbf{\Sigma}$ represent the condition of the channel, i.e., the smallest eigenvalue stand for the illest channel, the worst channel will dominate the whole BER and lead to imperfect performance, as shown in [8].

2.1.2 Nonlinear Equalizations

In this section, we will introduce one main nonlinear equalization strategies: decision-feedback equalizer, which can be seen as the predecessor of Tomlinson-Harashima precoder. We will briefly describe the main goal of design and compare it to linear equalizers.

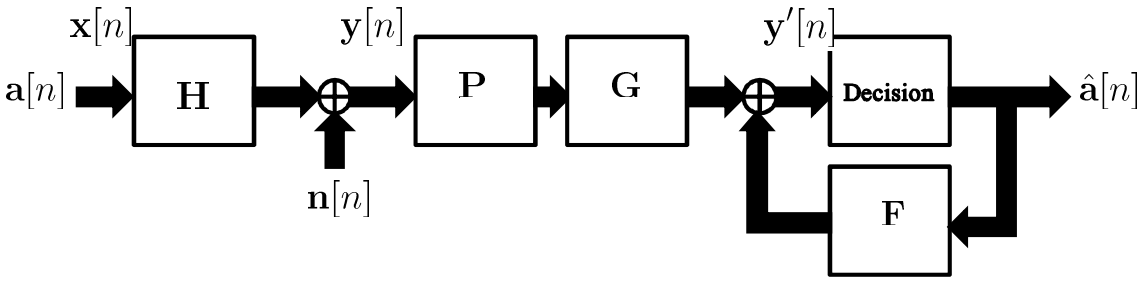


Figure 2.4: Decision Feedback Equalizer

Decision Feedback Equalizer [4] [5]

In order to improve the performance and to overcome the disadvantages at the receiver side of linear equalizations, DFE is occur, which is also called V-BLAST (Vertical-Bell Laboratories Layered Space-Time) system [25]. The block diagram of decision feedback equalizer is shown in Fig. 2.4, which is achievable if $joint\ processing$ is available at the receiver sides. The main idea of DFE in MIMO is that each receiver/user can communicate with others, and each user's signal is decided one after another. As the second element of $\mathbf{y}'[n]$ is shift into the decision device, the first input symbol(an already decided symbol) is then pass through the feedback matrix $\mathbf{F}$ and feed back in order to subtract the interference between the first and second user. The order of making decision can provide an additional degree of freedom for minimizing the mean-square error, i.e., the symbol which transmit over the better condition channel should be decided first and so on. More details can be found in [26].

In Fig. 2.4, $\mathbf{P}$ is the feedforward matrix and $\mathbf{F}$ is the feedback matrix, $\mathbf{G}$ is a gain-control matrix, which is diagonal $\mathbf{G} = diag(g_1 \dots g_{n_R})$. Note that the feedback matrix $\mathbf{F}$ must be strictly lower triangular to satisfy spatial causality. The signal $\mathbf{y}'[n]$ which is processed at the receiver in Fig. 2.4 can be represent as

$$\mathbf{y}'[n] = \mathbf{GPHa}[n] + \mathbf{GPn}[n] + \mathbf{F}\hat{\mathbf{a}}[n] \tag{2.6}$$

In DFE assumption, $perfect$ decision $\hat{\mathbf{a}}[n] = \mathbf{a}[n]$ is considered. Thus, the above equation can be rewritten as

$$\mathbf{y}'[n] = (\mathbf{GPH} + \mathbf{F})\mathbf{a}[n] + \mathbf{GPn}[n] \tag{2.7}$$

Assume that the CSI at the receiver is $perfect$, i.e., $\hat{\mathbf{H}} \triangleq \mathbf{H}$. Base on the equation (2.7), the feedback matrix is defined as $\mathbf{F} \triangleq \mathbf{I} - \mathbf{GPH}$ in order to remove the interference. For the choice of $\mathbf{G}$ and $\mathbf{P}$, note that to preserve the noise power, i.e., $E\{(\mathbf{GPn}[n])(\mathbf{GPn}[n])^H\} = E\{\mathbf{n}[n]\mathbf{n}^H[n]\}$, diagonal terms of $\mathbf{G}$ should be unit scalars, and $\mathbf{P}$ is constrain to a unitary matrix. A simple selection for $\mathbf{G}$ and $\mathbf{P}$ is applying the $QR\text{-}decomposition$ of the channel matrix

$$\mathbf{H} = \mathbf{Q}^H\mathbf{R}$$

where $\mathbf{Q}$ is the unitary matrix and $\mathbf{R} = [r_{ij}]$ is a lower triangular matrix. Thus, the scaling matrix and the feedforward matrix read $\mathbf{G} = diag(r_{11}^{-1} \dots r_{n_R n_R}^{-1})$ and $\mathbf{P} = \mathbf{Q}$, and finally, the feedback matrix is $\mathbf{F} = \mathbf{I} - \mathbf{GPH} = \mathbf{I} - \mathbf{GR}$.

The DFE strategy has outperform all linear equalization scheme [8]. However, even though the performance had further improved by reordering method, the $error\ propagation$ is still a main disadvantage of DFE. Also, for equalization, immediate decisions are required.

2.2 Tomlinson-Harashima Precoder

Initially, Tomlinson-Harashima Precoding [6] [7] was proposed to equalized the inter-stream-interference of severely distorting single-input single-output (SISO) channels, but in recent studies, it can be extended to MIMO channels. THP can be seen as moving the feedback part from receiver to transmitter under the condition of having CSI at the transmitter. In this case, compare to DFE, communication between different users is not needed(decentralized receivers). On the contrary, in downlink scheme, without having error propagation problem, user can still acquire other users' information and precoded adaptively to avoid the interference by feedback matrix and feedforward matrix. Also, no immediate decisions as in DFE are required.

The THP structure is shown in Fig. 2.5, "$\mathbf{Mod}$" denote the modulo operator, which is for constraining the precoded symbol power. $\mathbf{F} = [f_{ij}]$ is the feedback matrix and has to be a strictly lower triangular matrix in order to maintain causality, $\mathbf{P} = [p_{ij}]$ is the feedforward matrix and $\mathbf{G} = diag(g_1 \dots g_{n_R})$ is a diagonal scaling matrix (gain controller).

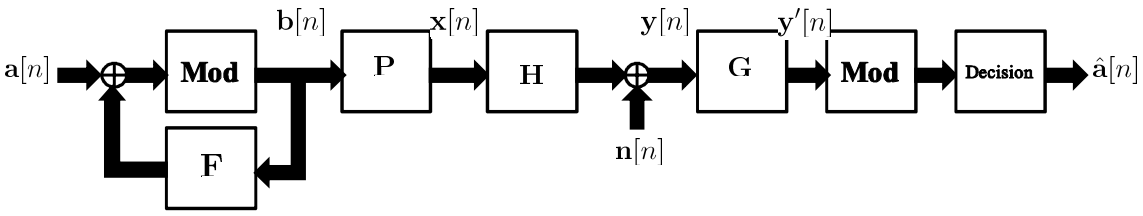


Figure 2.5: Tomlinson-Harashima Precoding scheme

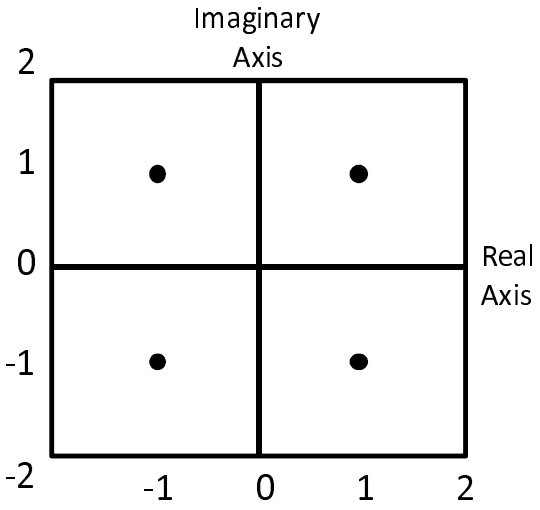


Figure 2.6: QPSK diagram (4-ary constellation) with real number and imaginary number axis.

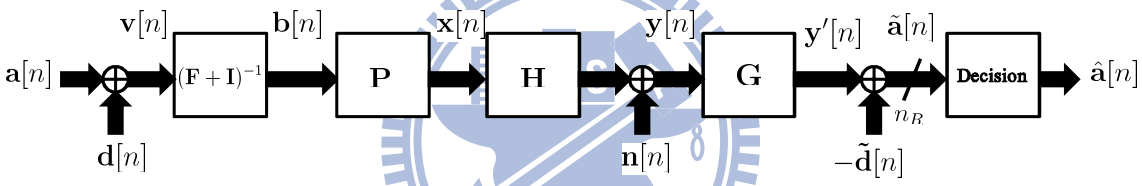


Figure 2.7: Linear representation of Tomlinson-Harashima Precoding scheme

The operation of these matrices will be described as follow. Similar in DFE, symbols are shifted into the precoding structure (modulo operator) one after another, therefore the immediate precoded symbol can have the information of all previously precoded symbols from feedback and thus adaptively precoded itself by feedback matrix $\mathbf{F}$ to avoid the interference from other users, i.e., i-th user have the information of 1-th...$(i-1)$-th users. The remain interference can be eliminated through the feedforward matrix $\mathbf{P}$. More details will be describe as follow.

Consider that the data symbol $a_i[n]$ is modulated as M-ary constellation, i.e., for QPSK, M=4. In the last paragraph, we interpret modulo operator as constraining the precoded symbol, this can be clearly explain by Fig. 2.6, which is a QPSK diagram (M=4). In QPSK, the data symbols $a_1[n] \ldots a_{n_R}[n]$ are modulated into these four points as in Fig. 2.6, and the data power is restricted in the $(2\sqrt{M} \times 2\sqrt{M})$-square. However,

without the modulo operator, the precoded symbol power may boost up due to the adding up feedback sequences. In order to limit the power, modulo operator is needed. Express the modulo operator mathematically, integer multiples of $2\sqrt{M}$ are added to the real and imaginary part of $a_i[n]$, the output of the modulo operator are given as

$$b_i[n] = \mathbf{Mod}(a_i[n] - \sum_{k=1}^{i-1} f_{ik} b_k[n]) = a_i[n] + d_i[n] - \sum_{k=1}^{i-1} f_{ik} b_k[n], \qquad i = 1, \ldots, n_R \tag{2.8}$$

where $d_i[n] \in \{2\sqrt{M} \cdot (d_I + jd_Q) \mid d_I, d_Q \in \mathbb{Z}\}$ is the precoding symbols. Modulo operator can be seen as to pull the precoded symbols back into the modulation square by adding or subtracting the real and imaginary part of integer multiples of $2\sqrt{M}$.

The proof in Tomlinson's paper [6] had shown that the THP structure can be transformed into a linear scheme, as in Fig. 2.7. Instead of feeding the data symbols into the modulo operator and feedback, the effective data symbols $\mathbf{v}[n] = \mathbf{a}[n] + \mathbf{d}[n]$ are passed into $(\mathbf{F} + \mathbf{I})^{-1}$, where $\mathbf{d}[n] = [d_1[n] \ldots d_{n_R}[n]]^T$ and $d_i[n]$ is defined in (2.8). The signal at the receiver is given as

$$\mathbf{y}'[n] = \mathbf{GHP}(\mathbf{F} + \mathbf{I})^{-1}\mathbf{v}[n] + \mathbf{Gn}[n] \tag{2.9}$$

From the above equation, we aim to force $\mathbf{GHP}(\mathbf{F} + \mathbf{I})^{-1} = \mathbf{I}$, thus the feedback matrix is chosen as $\mathbf{F} = \mathbf{GHP} - \mathbf{I}$. Since $\mathbf{G}$ is a diagonal matrix, we can conclude that $\mathbf{HP}(\mathbf{F} + \mathbf{I})^{-1}$ is also a diagonal matrix (for interference elimination). As the operation of $\mathbf{F}$ is to remove the previously precoded users' interference, we can comprehend clearly that $\mathbf{P}$ is designed to eliminate the remain interference. Later on, the signal vector $\mathbf{y}'[n] = \mathbf{v}[n] + \mathbf{Gn}[n]$ can be restored to $\tilde{\mathbf{a}}[n]$ by another modulo operator at the receiver.

2.3 Autoregressive Model

A Rayleigh characterization of the mobile radio channel follows from the Gaussian wide-sense-stationary uncorrelated scattering model, where the fading process is modeled as a complex Gaussian process. The variability of the wireless channel over time in this model is reflected in its autocorrelation function (ACF), which is depend on the propagation geometry, the mobile velocity and antenna characteristics. A common and simple

assumption is that the propagation path consist of two-dimensional scattering with vertical monopole antennas at the receivers [27]. The briefly description of this model is in next section.

2.3.1 Correlated Fading Model

For Rayleigh fading, the channel coefficient $h^{(i,j)}(t)$ is a zero-mean, wide-sense-stationary complex Gaussian process, which is uncorrelated with $h^{(i',j')}(t)$. According to Jakes' model in [27], the channel coefficient satisfied the time-autocorrelation properties

$$E\left\{h^{(i,j)}(t_1)\left[h^{(i,j)}(t_2)\right]^*\right\} \sim \mathcal{J}_0\left(2\pi f_D^{(i,j)} T|t_1 - t_2|\right) \tag{2.10}$$

where $\mathcal{J}_0(\cdot)$ is the zero order Bessel function of the first kind, T is the slot period, and $f_D^{(i,j)}$ is the maximum Doppler rate from $j\text{-}th$ transmit antenna to $i\text{-}th$ receiver.

To simplify, we assume equal Doppler rate between all transmit-receive antenna pairs, i.e., $f_D^{(i,j)} = f_D$ for all $i \in \{1, \ldots, n_R\}$ and $j \in \{1, \ldots, n_T\}$. Thus, the corresponding Jakes' power spectrum density [27] (PSD) with maximum Doppler frequency f_D has the well-known U-shape bandlimit form

$$S(f) = \begin{cases} \frac{1}{\pi f_D T} \frac{1}{\sqrt{1-(\frac{f}{f_D T})^2}}, & |f| < f_D T \\ 0 \qquad\qquad , & otherwise \end{cases} \tag{2.11}$$

2.3.2 Autoregressive Model

In statistics and signal processing, an autoregressive (AR) model is a type of random process which is often used to model and predict various types of natural phenomena. The autoregressive model is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous outputs.

According to [28], the fading channel can be accurately modeled by a large order of autoregressive models, as in Fig. 2.8. However, large order leads to higher complexity; and further, as shown in [1], low order AR models can match the Bessel autocorrelation well for small lags $k = |t_1 - t_2|$, and can capture most of the channel dynamics, leading an efficient tracking algorithm. Thus, we use a low order AR model for channel tracking.

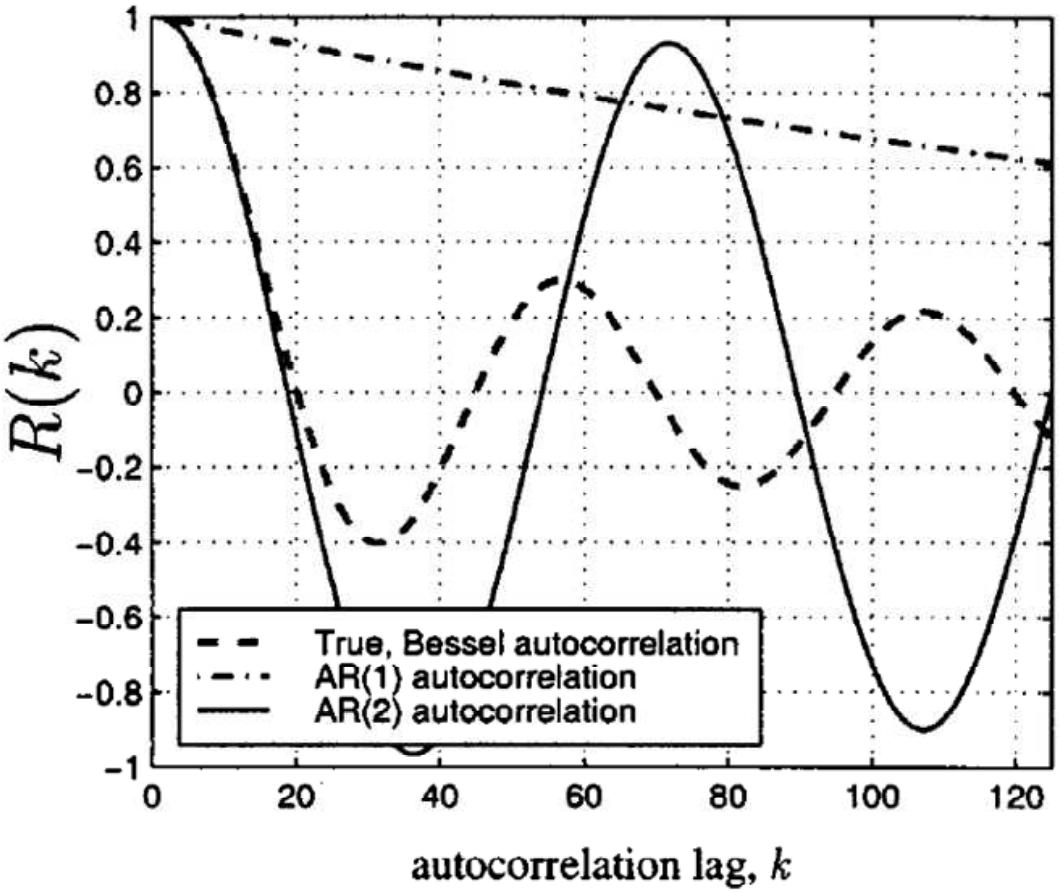


Figure 2.8: [1]Autocorrelation function $\mathbf{R}(k)$ true (Bessel) and for the AR(Q) model for $Q = 1, 2$ and Doppler rate $f_D T = 0.02$. The second-order AR model autocorrelation matches the true expression for lag < 20, although only the first three terms are exactly equal.

To approximate the time varying channel parameters $h^{(i,j)}(t)$, the following multi-channel AR process [29] is used

$$\mathbf{h}(t) = \sum_{q=1}^{Q} \mathbf{A}(l_q)\mathbf{h}(t-l_q) + \mathbf{B}_0\mathbf{w}(t) \tag{2.12}$$

where $\mathbf{h}(t) = [\mathbf{h}_1^T(t), \ldots, \mathbf{h}_{n_R}^T(t)]^T = [h^{(1,1)}(t), \ldots, h^{(n_T,n_R)}(t)]^T \in \mathbb{C}^{n_T n_R \times 1}$, $\mathbf{w}(t) \in \mathbb{C}^{n_T n_R \times 1}$ is a zero mean i.i.d circular complex Gaussian vector with correlation matrix $\mathbf{R}_{ww} = E\{\mathbf{w}(t_i)\mathbf{w}^*(t_j)\} = \mathbf{I}_{n_T n_R}\delta_{ij}$ for Rayleigh variate generation, δ_{ij} is the delta function. The matrices $\mathbf{A}(l_q) \in \mathbb{C}^{n_R n_T \times n_T n_R}$, $q = 1, \ldots, Q$ where Q is the order of AR model, and $\mathbf{B}_0 \in \mathbb{C}^{n_T n_R \times n_T n_R}$ is diagonal due to assumption in Section 2.3.1, i.e., $\mathbf{A}(l_q) = diag[a^{(i,j)}(l_q)]_{i,j=1}^{i=n_R,j=n_T}$ and $\mathbf{B}_0 = diag[b^{(i,j)}]_{i,j=1}^{i=n_R,j=n_T}$, (i,j) is the channel path index between different transmit-receive antenna pairs. l_q denote the number of outdated slots and q is the index of uplink slot.

The matrix form of (2.12) can be written as

$$\mathbf{h}_T = \mathbf{A}\mathbf{h}_{T,pre} + \mathbf{B}\mathbf{w}(t) \tag{2.13}$$

where $\mathbf{h}_T \in \mathbb{C}^{n_R n_T Q \times 1}$

$$\mathbf{h}_T = [\mathbf{h}(t)^T \quad \mathbf{h}(t-l_1)^T \quad \cdots \quad \mathbf{h}(t-l_{Q-1})^T]^T \tag{2.14}$$

$$\mathbf{h}_{T,pre} = [\mathbf{h}(t-l_1)^T \quad \mathbf{h}(t-l_2)^T \quad \cdots \quad \mathbf{h}(t-l_Q)^T]^T \tag{2.15}$$

and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}(l_1) & \ldots & \mathbf{A}(l_Q) \\ \mathbf{I}_{n_R n_T (Q-1)} & & \mathbf{0}_{n_T n_R (Q-1) \times n_T n_R} \end{bmatrix} \tag{2.16}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_0 \\ \mathbf{0}_{n_T n_R (Q-1) \times n_T n_R} \end{bmatrix} \tag{2.17}$$

After choosing the order Q for the AR model, we can fix $\mathbf{A}$ and $\mathbf{B}$ in (2.13), i.e., $a^{(i,j)}(l_q)$ and $b^{(i,j)}$. To simplify, assume $a^{(i,j)}(l_q) = a(l_q)$ for all $i \in \{1, \ldots, n_R\}$ and $j \in \{1, \ldots, n_T\}$, i.e., all channel paths varying at the same rate. Assume that the autocorrelation function (ACF) matrix is $\mathbf{R}$. Ignoring the ill condition, assume that the inverse $\mathbf{R}^{-1}$ exists and the Yule-Walker equations are generated to have the unique solution of $\mathbf{A}$ [28]

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{v} \tag{2.18}$$

where

$$\mathbf{R} = \begin{bmatrix} r[0] & r[-2] & \cdots & r[-2Q+2] \\ r[2] & r[0] & \cdots & r[-2Q+4] \\ \vdots & \vdots & \ddots & \vdots \\ r[2Q-2] & r[2Q-4] & \cdots & r[0] \end{bmatrix}$$

$$\mathbf{a} = \begin{bmatrix} a(l_1) & a(l_2) & \cdots & a(l_Q) \end{bmatrix}^T$$

$$\mathbf{v} = \begin{bmatrix} r[l_1] & r[l_2] & \cdots r[l_Q] \end{bmatrix}^T \tag{2.19}$$

where $r[\tau] = \mathcal{J}_0(2\pi f_D T|\tau|)$ is the time-autocorrelation function for given delay τ. Given the desired ACF sequences $r[\tau]$, the AR filter coefficients can be determined by solving the set of Q Yule-Walker equations in (2.18).

The channel varying $rate$ is fixed via **A**. From equation (2.12), the power of (i,j)-th channel path can be written as

$$E\{|h^{(i,j)}|^2\} = \frac{b^{(i,j)^2}}{(1 - a(l_1) - \cdots - a(l_Q))^2} \tag{2.20}$$

where $a(l_1), \ldots, a(l_Q)$ is determined by Doppler rate as in (2.18). Thus, $b^{(i,j)}$ is controlled by the channel path power. For example, a carrier frequency of 2GHz with a slot period of 0.675ms, and a normalized Doppler frequency of $f_D T = 0.08$ corresponds to a velocity of 64km/hr. Taking order $Q = 1$ and $l_q = 2q + 1$, channel power $\mathbf{C}_{\mathbf{h}_i} = \mathbf{I}_{n_T}$. Thus, this case sets $a(l_1) = 0.5074$ and $b^{(i,j)} = \sqrt{0.4926}$ for all $i \in \{1, \ldots, n_R\}$ and $j \in \{1, \ldots, n_T\}$ from (2.18) and (2.20).

Chapter 3

Channel Tracking and THP Optimization with Partial CSI

3.1 System Model

Consider a transceiver with n_T cooperative transmit antennas and n_R noncooperative receivers/users, each user is equipped with one antenna. Downlink and uplink takes place in TDD mode, as shown in Fig. 3.1. Here, uplink and downlink are assumed to be reciprocal, which means $\mathbf{H}_{DL} = \mathbf{H}_{UL}^T \triangleq \mathbf{H}$. Assume that data transmission is in downlink, i.e., $\mathbf{H}_{DL} \in \mathbb{C}^{n_R \times n_T}$, and n_T antennas transmit simultaneously in one fixed time slot. As shown in Fig. 3.1, $t = t_0$ denote the time slot index at t_0-th time slot, whereas n denote the symbol vector index transmitted in each time slot, and N is the total number of symbol vectors in each time slot, i.e., $n \in \{1, \ldots, N\}$. Thus, the absolute symbol index is given by $N(t-1)+n$.

3.1.1 Channel model

Consider a time-varying Rayleigh fading channel. Thus, downlink channel matrix $\mathbf{H}$ at time t is

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T(t) \\ \vdots \\ \mathbf{h}_{n_R}^T(t) \end{bmatrix} = \begin{bmatrix} h^{(1,1)}(t) & \ldots & h^{(1,n_T)}(t) \\ \vdots & \ddots & \vdots \\ h^{(n_R,1)}(t) & \ldots & h^{(n_R,n_T)}(t) \end{bmatrix}$$

$$\triangleq \mathbf{H}(t) \in \mathbb{C}^{n_R \times n_T} \tag{3.1}$$

where $h^{(i,j)}(t)$ is the channel coefficient from j-th transmit antenna to i-th receiver in time slot t. The channel coefficients are modeled as a stationary zero-mean complex Gaussian random vector with $\mathbf{h}_i(t) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{C}_{h_i})$.

3.1.2 Downlink Training Channel

According to [23], in order to precisely design receivers, we need receivers' channel knowledge. Thus, downlink training transmission is assumed, which are transmitted orthogonally to the data with in the same time slot, as shown in Fig. 3.2. The receivers' channel knowledge is determined by linear precoder $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{n_R}] \in \mathbb{C}^{n_T \times n_R}$ and known training sequence $\mathbf{b}[n]$, and we assume each receiver has perfect channel knowledge of $\mathbf{h}_i^T \mathbf{q}_i$.

$\mathbf{Q}$ provides an additional degree of freedom in system design, the choice of $\mathbf{q}_i$ and the receivers' design based on $\mathbf{h}_i^T \mathbf{q}_i$ will be discussed in Section 3.1.4.

3.1.3 Uplink Training Channel

The transmitter channel state information is offered by uplink training channel, as in Fig. 3.3. Assumed that the worst time delay is three time slots. Each training sequence $\mathbf{s}[n] \in \mathbb{C}^{n_R \times 1}$ is used to uplink at a time, and the total training sequences $\mathbf{S}_{up} = [\mathbf{s}[0] \quad \mathbf{s}[1] \ldots \mathbf{s}[N-1]] \in \mathbb{C}^{n_R \times N}$ will be transmitted in every time slot, which is known

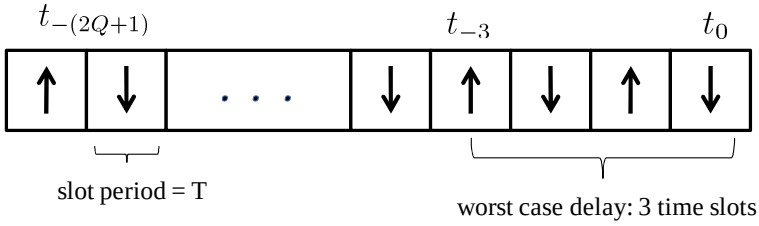


Figure 3.1: TDD structure: Uplink(↑) and downlink(↓) for data transmission in fixed time slot, each time slot period is T and the time slot index is t. Three time slots delay is assumed.

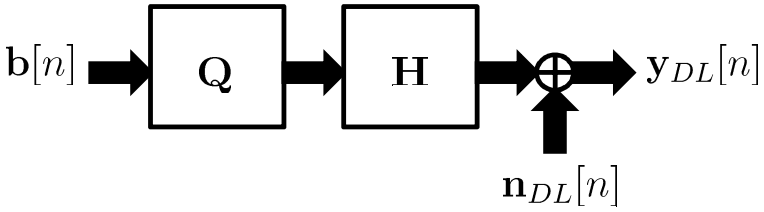


Figure 3.2: Downlink Training channel: $\mathbf{Q} \in \mathbb{C}^{n_T \times n_R}$ is the linear precoder which offers the receiver channel knowledge, and $\mathbf{H} \triangleq \mathbf{H}[t]$.

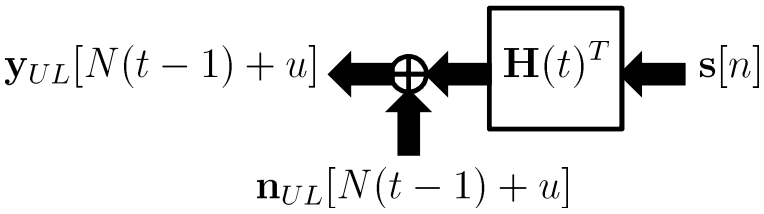


Figure 3.3: Uplink Training channel: The uplink training channel transmit at t-th time slot, and the absolute symbol is $N(t-1)+n$.

by both transmitter and receiver. Thus, the single receive signal in one time slot can be written as

$$\mathbf{y}_{UL}(N(t-1)+n) = \mathbf{H}(t)^T\mathbf{s}[n] + \mathbf{n}_{UP}(N(t-1)+n) \in \mathbb{C}^{n_T \times 1} \tag{3.2}$$

where $n \in (1, \ldots, N)$ and the additive white complex Gaussian noise is $\mathbf{n}_{UP}[N(t-1)+n] \sim \mathcal{N}_c(0, \sigma_n^2\mathbf{I}_{n_T})$.

Collecting N receive signal that transmitted in one uplink time slot, we obtain

$$\mathbf{Y}_{UL}(t) = \mathbf{H}(t)^T\mathbf{S}_{up} + \mathbf{N}(t) \in \mathbb{C}^{n_T \times N} \tag{3.3}$$

Reshape the matrix form into column form

$$\bar{\mathbf{y}}_{UL}(t) = vec(\mathbf{Y}_{UL}(t)) = (\mathbf{S}_{up}^T \otimes \mathbf{I}_{n_T})\mathbf{h}(t) + \bar{\mathbf{n}}(t) \in \mathbb{C}^{n_T N \times 1} \tag{3.4}$$

where $\mathbf{h}(t) \triangleq vec(\mathbf{H}(t)^T) = [\mathbf{h}_1^T(t) \ldots \mathbf{h}_{n_R}^T(t)]^T \in \mathbb{C}^{n_R n_T \times 1}$.

As we had discussed in Section 2.3.2, higher AR order Q offers higher precision of modeling time varying channel. Thus, we collect Q previous uplink slots into a column,

which are outdated by l_q, $q \in \{1, \ldots, Q\}$, slots. The total observation at the transmitter can be written as

$$\mathbf{y}_{T,pre} = \mathbf{S}\mathbf{h}_{T,pre} + \mathbf{n}_{T,pre} \in \mathbb{C}^{n_T NQ \times 1} \tag{3.5}$$

where $\mathbf{h}_{T,pre} = [\mathbf{h}(t-l_1)^T \ldots \mathbf{h}(t-l_Q)^T]^T \in \mathbb{C}^{n_T n_R Q \times 1}$, $\mathbf{y}_{T,pre} = [\bar{\mathbf{y}}(t-l_1)^T \ldots \bar{\mathbf{y}}(t-l_Q)^T]^T \in \mathbb{C}^{n_T n_R Q \times 1}$, $\mathbf{n}_{T,pre} = [\bar{\mathbf{n}}(t-l_1)^T \ldots \bar{\mathbf{n}}(t-l_Q)^T]^T \in \mathbb{C}^{n_T n_R Q \times 1}$ and $\mathbf{S} = \mathbf{I}_Q \otimes \mathbf{S}_{up} \otimes \mathbf{I}_{n_T} \in \mathbb{C}^{n_T NQ \times n_T n_R Q}$.

3.1.4 Downlink Data Channel

The downlink data channel model is shown in Fig. 3.4. The models of transmitter and receiver side are as follow:

Transmitter Model

The transmit data symbol $\mathbf{a}[n] = [a_1[n] \ldots a_{n_R}[n]]^T$ is modulated as M-ary constellation, and is precoded symbol by symbol. Modulo operator "$\mathbf{Mod}$" is required in order to constrain the precoded symbol power. In Section 2.2, we had evidently explain how the modulo operator works. First, we represent the linear THP model, as shown in Fig. 3.5. The feedback matrix $\mathbf{F} = [\mathbf{f}_1 \ldots \mathbf{f}_{n_R}] \in \mathbb{C}^{n_R \times n_R}$ is used to feedback the information of previous precoded symbols, and has to be a $strictly$ lower triangular matrix in order to ensure spatial causality. The output of the modulo operator $\mathbf{b}[n] \in \mathbb{C}^{n_R \times 1}$ can be written as

$$\begin{aligned} b_i[n] &= \mathbf{Mod}(a_i[n] - \sum_{l=1}^{i-1} f_{il} d_l[n]) \\ &= a_i[n] + d_i[n] - \sum_{l=1}^{i-1} f_{il} d_l[n] \end{aligned} \tag{3.6}$$

where $i = 1, \ldots, n_R$, f_{il} is the (i,l)th element in $\mathbf{F}$, and $d_i \in \{2\sqrt{M} \cdot (d_I + j d_Q) \mid d_I, d_Q \in \mathbb{Z}\}$ is the precoding symbols. $\mathbf{b}[n]$ is then pass into the feedforward matrix $\mathbf{P} = [\mathbf{p}_1 \ldots \mathbf{p}_{n_R}] \in \mathbb{C}^{n_T \times n_R}$, the output of $\mathbf{P}$ reads as $\mathbf{x}[n] = \mathbf{P}(\mathbf{F}+\mathbf{I})^{-1}\mathbf{v}[n] \in \mathbb{C}^{n_T \times 1}$, and the received signal at the receivers is

$$\mathbf{y}[n] = \mathbf{H}\mathbf{P}(\mathbf{F}+\mathbf{I})^{-1}\mathbf{v}[n] + \mathbf{n}[n] \in \mathbb{C}^{n_T \times 1} \tag{3.7}$$

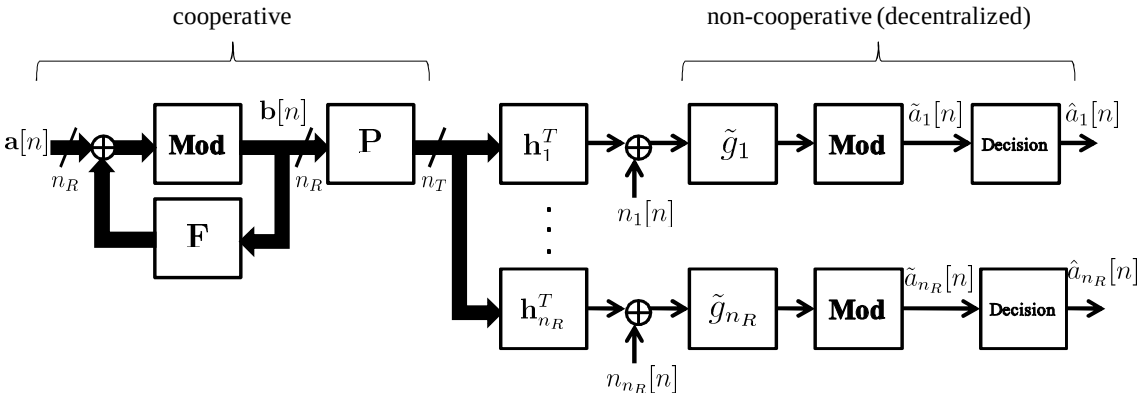


Figure 3.4: THP model with n_T transmit antennas and n_R users, each user is equipped with one antenna.

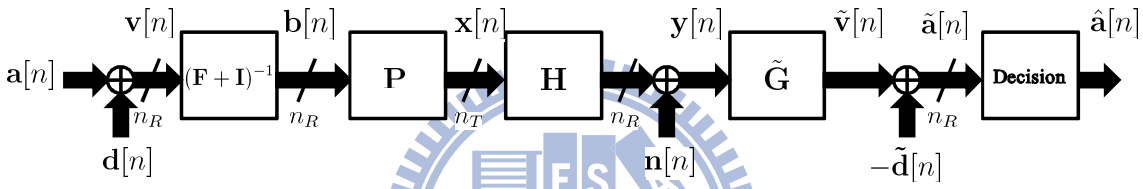


Figure 3.5: THP model with linear representation

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{n_R}]^T \triangleq \mathbf{H}[t] \in \mathbb{C}^{n_R \times n_T}$ and $\mathbf{n}[n] \in \mathbb{C}^{n_R \times 1}$ is the additive white complex Gaussian noise that satisfy $\mathbf{n}[n] \sim \mathcal{N}_c(\mathbf{0}, \sigma_n^2 \mathbf{I}_{n_R})$. $\mathbf{P}$ is designed to eliminate the interference, i.e., making the effective channel $\mathbf{HP}(\mathbf{F}+\mathbf{I})^{-1}$ a *diagonal* matrix.

Receiver Model

The (noncooperative) receivers are models as $\tilde{\mathbf{G}} = diag[\tilde{g}_i]_{i=1}^{n_R} \in \mathbb{C}^{n_R \times n_R}$. Traditionally, the receivers' design is a real value scaling

$$\tilde{\mathbf{G}} = \beta^{-1}\mathbf{G} = \beta^{-1}\mathbf{I}_{\mathbf{n_R}} \tag{3.8}$$

β is an amplitude scaling and provides the necessary degree of freedom to allow for the transmit power constraint. However, according to [23], an additional channel knowledge for the receivers can offer a degree of freedom for designing THP, as we had discussed in Section 3.1.2. By considering the correction of the channel phase, a simple and more

precise design of receivers is (As in [23])

$$\tilde{g}_i = \beta^{-1} f(\mathbf{h}_i^T \mathbf{q}_i) = \beta^{-1} \frac{(\mathbf{h}_i^T \mathbf{q}_i)^*}{|\mathbf{h}_i^T \mathbf{q}_i|} \tag{3.9}$$

where $\tilde{\mathbf{G}} = diag[\tilde{g}_i]_{i=1}^{n_R}$, and choosing $\mathbf{q}_i$ as the complex conjugate principal eigenvector of the conditional correlation matrix $E_h[\mathbf{h}_i \mathbf{h}_i^H | \mathbf{y}_T]$ which based on the idea of combining (phase correction).

3.2 Problem Setup

The THP optimization is based on the MMSE criterion and the knowledge of the current channel. The total MSE of all users between $\mathbf{v}[n]$ and $\tilde{\mathbf{v}}[n]$ in Fig. 3.5 is

$$MSE(\mathbf{P}, \mathbf{F}, \beta; \mathbf{H}) = E_{\mathbf{w},\mathbf{n}}\left[\|\mathbf{v}[n] - \tilde{\mathbf{v}}[n]\|_2^2\right] \tag{3.10}$$

where $\mathbf{v}[n] = (\mathbf{F} + \mathbf{I})\mathbf{b}[n]$, $\tilde{\mathbf{v}}[n] = \beta^{-1}\mathbf{G}(\mathbf{HPb}[n] + \mathbf{n}[n])$, and $\mathbf{H} \triangleq \mathbf{H}[t]$.

The THP optimization problem is to design the optimum feedforward matrix $\mathbf{P}$, feedback matrix $\mathbf{F}$, and receiver parameter β which can minimize the mean square error, as follow

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{F},\beta} MSE(\mathbf{P}, \mathbf{F}, \beta; \mathbf{H}) \\ &\quad \text{s.t.} \quad 1) \quad tr(\mathbf{PC_bP}^H) \le P_T \\ &\qquad\qquad 2) \quad \mathbf{F} : \text{strictly lower triangular matrix} \end{aligned} \tag{3.11}$$

where P_T is the average transmit power constraint, and $\mathbf{C_b}$ is the covariance matrix of vector $\mathbf{b}[n]$.

3.3 Kalman Estimation

Since channel matrix $\mathbf{H}$ is unknown, THP optimization in (3.11) can not be analyzed. Thus, we first estimate channel using Kalman estimation. The estimation is based on AR model in (2.13)

$$\mathbf{h}_T = \mathbf{A}\mathbf{h}_{T,pre} + \mathbf{B}\mathbf{w}(t)$$

and the observation given by uplink training channel in (3.5)

$$\mathbf{y}_{T,pre} = \mathbf{S}\mathbf{h}_{T,pre} + \mathbf{n}_{T,pre}$$

Where the covariance matrices of $\mathbf{w}(t)$ and $\mathbf{n}_{T,pre}$ are $\mathbf{R_{ww}} = \delta_{ij}\mathbf{I}_{n_T n_R}$ and $\mathbf{R_{nn}} = \sigma_n^2\mathbf{I}_{n_T n_R Q}$. Kalman estimation can be seen as building a channel model first, and then revise it by the channel information (uplink observation in this case) based on MMSE.

Since the randomness (3.5) has introduced into the model (2.13), the equation (2.13) have to be written as

$$\hat{\mathbf{h}}_T = \mathbf{A}\hat{\mathbf{h}}_{T,pre} + \mathbf{K}_{pre}(\mathbf{y}_{T,pre} - \hat{\mathbf{y}}_{T,pre}) \tag{3.12}$$

where

$$\begin{aligned}\mathbf{y}_{T,pre} - \hat{\mathbf{y}}_{T,pre} &= [\mathbf{S}\mathbf{h}_{T,pre} + \mathbf{n}_{T,pre}] - [\mathbf{S}\hat{\mathbf{h}}_{T,pre}] \\ &= \mathbf{S}(\mathbf{h}_{T,pre} - \hat{\mathbf{h}}_{T,pre}) + \mathbf{n}_{T,pre} \\ &= \mathbf{S}\tilde{\mathbf{h}}_{T,pre} + \mathbf{n}_{T,pre} = \tilde{\mathbf{y}}_{T,pre}\end{aligned}$$

$\tilde{\mathbf{h}}$ can be seen as the estimation error of $\mathbf{h}$. Now, we aim to find the correction item $\mathbf{K}_{pre}$.

The error of the channel coefficients at time t (in matrix form) is

$$\begin{aligned}\tilde{\mathbf{h}}_T &= \mathbf{h}_T - \hat{\mathbf{h}}_T \\ &= \mathbf{A}\tilde{\mathbf{h}}_{T,pre} + \mathbf{B}\mathbf{w}(t) - \mathbf{K}_{pre}(\mathbf{y}_{T,pre} - \hat{\mathbf{y}}_{T,pre}) \\ &= (\mathbf{A} - \mathbf{K}_{pre}\mathbf{S})\tilde{\mathbf{h}}_{T,pre} + \mathbf{B}\mathbf{w}(t) - \mathbf{K}_{pre}\mathbf{n}_{T,pre}\end{aligned} \tag{3.13}$$

Applying equation (3.13) to the matrix form

$$\tilde{\mathbf{h}}_T = (\mathbf{A} - \mathbf{K}_{pre}\mathbf{S})\tilde{\mathbf{h}}_{T,pre} + \mathbf{B}\mathbf{w}(t) - \begin{bmatrix}\mathbf{B} & -\mathbf{K}_{pre}\end{bmatrix}\begin{bmatrix}\mathbf{w}(t) \\ \mathbf{n}_{T,pre}\end{bmatrix} \tag{3.14}$$

The mean-square-error of $\mathbf{h}_T$ can be written as

$$\begin{aligned}&MSE = E[\tilde{\mathbf{h}}_T\tilde{\mathbf{h}}_T^H] \\ &= (\mathbf{A} - \mathbf{K}_{pre}\mathbf{S})E[\tilde{\mathbf{h}}_{T,pre}\tilde{\mathbf{h}}_{T,pre}^H](\mathbf{A} - \mathbf{K}_{pre}\mathbf{S})^H \begin{bmatrix}\mathbf{B} & -\mathbf{K}_{pre}\end{bmatrix}\begin{bmatrix}\mathbf{R}_{ww} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{nn}\end{bmatrix}\begin{bmatrix}\mathbf{B}^H \\ -\mathbf{K}_{pre}^H\end{bmatrix}\end{aligned}$$

$$= \begin{bmatrix} \mathbf{I}_{n_T NQ} & \mathbf{K}_{pre} \end{bmatrix} \begin{bmatrix} \mathbf{A}\mathbf{P}_{pre}\mathbf{A}^H + \mathbf{B}\mathbf{I}_{n_T n_R}\mathbf{B} & -(\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H) \\ -(\mathbf{S}\mathbf{P}_{pre}\mathbf{A}^H) & \mathbf{R}_{nn} + \mathbf{S}\mathbf{P}_{pre}\mathbf{S}^H \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_T NQ} \\ \mathbf{K}_{pre}^H \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I}_{n_T NQ} & \mathbf{K}_{pre} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_T NQ} & -(\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1} \\ \mathbf{0} & \mathbf{I}_{n_T NQ} \end{bmatrix} \times$$

$$\begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{e,pre} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_T NQ} & \mathbf{0} \\ -(\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1} & \mathbf{I}_{n_T NQ} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_T NQ} \\ \mathbf{K}_{pre}^H \end{bmatrix} \tag{3.15}$$

where $\mathbf{P}_{pre} = E[\tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t]$, $\mathbf{R}_{e,pre} = E[\tilde{\mathbf{y}}_{T,pre}\tilde{\mathbf{y}}_{T,pre}^H] = \mathbf{R}_{nn} + \mathbf{S}\mathbf{P}_{pre}\mathbf{S}^H$ and $\Delta = \mathbf{A}\mathbf{P}_{pre}\mathbf{A}^H + \mathbf{B}\mathbf{I}_{n_T n_R}\mathbf{B} - (\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1}(\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)^H$.

By the technique of "completing the square method", the the ideal result in equation (3.15) can be written as

$$MSE = \Delta + (\mathbf{K}_{pre} - \mathbf{K}_{opt,pre})\mathbf{R}_{e,pre}(\mathbf{K}_{pre} - \mathbf{K}_{opt,pre})^H \tag{3.16}$$

where $\mathbf{K}_{opt,pre}$ denote the optimum solution of $\mathbf{K}_{pre}$ for minimizing MSE. Comparing equation (3.15) and (3.16), we finally get

$$\mathbf{K}_{pre} = \mathbf{K}_{opt,pre} = (\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1}$$

Thus, the relation between $\mathbf{h}_T$ and $\hat{\mathbf{h}}_{T,pre}$ has been solved. The recursive equations are as follows

$$\hat{\mathbf{h}}_T = \mathbf{A}\hat{\mathbf{h}}_{T,pre} + \mathbf{K}_{\mathbf{pre}}\tilde{\mathbf{y}}_{T,pre} \tag{3.17}$$

where

$$\tilde{\mathbf{y}}_{T,pre} = \mathbf{y}_{T,pre} - \mathbf{S}\hat{\mathbf{h}}_{T,pre}$$

$$\mathbf{K}_{pre} = (\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1}$$

$$\mathbf{R}_{e,pre} = \mathbf{R}_{nn} + \mathbf{S}\mathbf{P}_{pre}\mathbf{S}^H$$

$$\mathbf{P}_t = E\{\tilde{\mathbf{h}}_\mathbf{t}\tilde{\mathbf{h}}_t^H\} = \mathbf{A}\mathbf{P}_{pre}\mathbf{A}^H + \mathbf{B}\mathbf{B}^H - \mathbf{K}_{pre}\mathbf{R}_{e,pre}\mathbf{K}_{pre}^H$$

where $\mathbf{R}_{nn}$ is the covariance matrix of $\mathbf{n}_{T,pre}$, $\mathbf{P}_t$ is the minimum value of the error correlation matrix at time t where $\tilde{\mathbf{h}_\mathbf{t}} = \mathbf{h}_t - \hat{\mathbf{h}}_t$, and $\hat{\mathbf{h}}_{T,pre}$ is obtained by the Kalman filtering in previous time slots.

Thus, the estimation of channel coefficients reads

$$\hat{\mathbf{h}}_t = [[\hat{\mathbf{h}}_T]_1 \cdots [\hat{\mathbf{h}}_T]_{n_T n_R}]^T$$

where $[\hat{\mathbf{h}}_T]_k$ denote the k-th element in vector $\hat{\mathbf{h}}_T$.

3.4 THP Optimization

The traditional method for solving THP optimization is to substitute $\hat{\mathbf{H}}$ in (3.17) into the MSE function (3.10) as in Fig. 3.6, but we do not consider this method in our work. Since $\mathbf{H}$ is a unknown matrix for partial CSI, the MSE function can be seen as a random variable. To make an accurate design of THP, instead of estimating $\mathbf{H}$ then substitute $\hat{\mathbf{H}}$ into the MSE function, we can further conclude the estimation error into the optimization, as in Fig. 3.7.

First, we assume the conditional mean (CM) estimator [30]

$$\begin{aligned}
&MSE_{CM}(\mathbf{P}, \mathbf{F}, \beta; \mathbf{y}_{T,pre}) \\
&= E_{\mathbf{h}}[MSE(\mathbf{P}, \mathbf{F}, \beta; \mathbf{H})|\mathbf{y}_{T,pre}] \\
&= tr\big((\mathbf{I}_{n_R} - \mathbf{F})\mathbf{C}_{\mathbf{b}}(\mathbf{I}_{n_R} - \mathbf{F})^H + \beta^{-2}E_{\mathbf{h}}[\mathbf{G}\mathbf{G}^H|\mathbf{y}_{T,pre}]\sigma_n^2 \\
&\qquad +\beta^{-2}\mathbf{P}\mathbf{C}_{\mathbf{b}}\mathbf{P}^H E_{\mathbf{h}}[\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}|\mathbf{y}_{T,pre}]\big) \\
&\qquad -2\beta^{-1}Re\{tr\big(E_{\mathbf{h}}[\mathbf{G}\mathbf{H}|\mathbf{y}_{T,pre}]\mathbf{P}\mathbf{C}_{\mathbf{b}}(\mathbf{I}_{n_R} - \mathbf{F})^H\big)\}
\end{aligned} \tag{3.18}$$

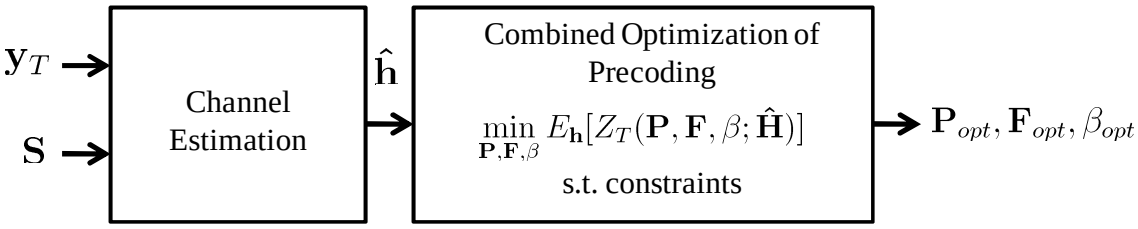


Figure 3.6: Traditional Optimization: Separate optimization of channel estimation and THP

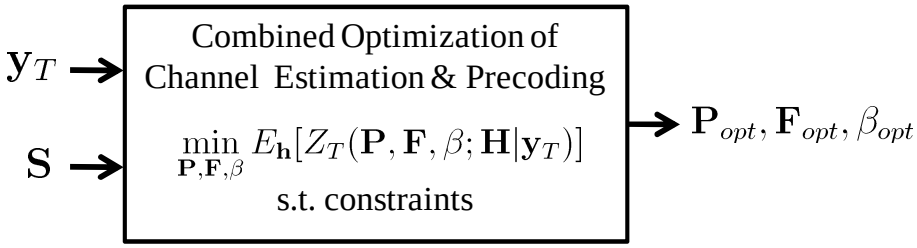


Figure 3.7: Non-traditional Optimization: Combine optimization of channel estimation and THP

Thus, the new optimization problem for partial-CSI at the transmitter reads

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{F},\beta} MSE_{CM}(\mathbf{P},\mathbf{F},\beta;\mathbf{y}_{T,pre}) \\ &\quad s.t. \quad 1) \quad tr(\mathbf{P}\mathbf{C}_{\mathbf{b}}\mathbf{P}^H) \leq P_T \\ &\qquad\quad 2) \quad \mathbf{F}: \text{strictly lower triangular matrix} \end{aligned} \tag{3.19}$$

where P_T is the average transmit power constraint and $\mathbf{F}$ is a strictly lower triangular matrix.

The optimization problem is solved by first minimize $\mathbf{F}$ with fixed but unknown $\mathbf{P}$, then find the solution of $\mathbf{P}$ and β by Lagrangian approach, as shown in [23]

$$\mathbf{p}_i = \beta_P \left(\mathbf{L}_{\mathbf{y}_{T,pre}} + \hat{\mathbf{A}}_i^H \hat{\mathbf{A}}_i + \frac{\hat{G}_N \sigma_n^2}{P_T} \mathbf{I}_{n_T} \right)^{-1} \hat{\mathbf{A}}_i^H \mathbf{e}_i \tag{3.20}$$

$$\mathbf{f}_i = -\beta^{-1} \begin{bmatrix} \mathbf{0}_{i \times n_T} \\ \hat{\mathbf{B}}_i \end{bmatrix} \mathbf{p}_i \tag{3.21}$$

$$\begin{aligned} \mathbf{L}_{\mathbf{y}_{T,pre}} &= E_{\mathbf{h}} \Big[(\mathbf{GH} - E_{\mathbf{h}}[\mathbf{GH}|\mathbf{y}_{T,pre}])^H \mathbf{GH} - E_{\mathbf{h}}[\mathbf{GH}|\mathbf{y}_{T,pre}]) | \mathbf{y}_{T,pre} \Big] \\ &= E_{\mathbf{h}}[\mathbf{H}^H \mathbf{G}^H \mathbf{GH}|\mathbf{y}_{T,pre}] - E_{\mathbf{h}}[\mathbf{H}^H \mathbf{G}^H|\mathbf{y}_{T,pre}] E_{\mathbf{h}}[\mathbf{GH}|\mathbf{y}_{T,pre}] \end{aligned} \tag{3.22}$$

where $\mathbf{F} = [\mathbf{f}_1 \dots \mathbf{f}_{n_R}] \in \mathbb{C}^{n_R \times n_R}$, and $\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_{n_R}] \in \mathbb{C}^{n_T \times n_R}$. $\mathbf{L}_{\mathbf{y}_{T,pre}}$ is the conditional covariance matrix of $(\mathbf{GH})^H$, β is chosen to satisfy the transmit power constraint, $\hat{\mathbf{A}}_i$ denotes the first i rows and $\hat{\mathbf{B}}_i$ the last $n_R - i$ rows of $E_{\mathbf{h}}[\mathbf{GH}|\mathbf{y}_{T,pre}]$, $\hat{G}_N = tr(E_{\mathbf{h}}[\mathbf{GG}^H|\mathbf{y}_{T,pre}])$, and $\mathbf{e}_i$ is vector with one on the ith element and zeros elsewhere.

To calculate the CM estimate of the effective channel $\mathbf{GH}$ [23], we first define the complex Gaussian random variable $z_i = \mathbf{h}_i^T \mathbf{q}_i$ and consider the ith row of $\mathbf{GH}$. Using the properties of the conditional expectation [31], we obtain

$$E_{\mathbf{h}}[g_i \mathbf{h}_i | \mathbf{y}_{T,pre}] = E_{\mathbf{h}}\left[\frac{(\mathbf{h}_i^T \mathbf{q}_i)^*}{|\mathbf{h}_i^T \mathbf{q}_i|}\mathbf{h}_i | \mathbf{y}_{T,pre}\right] = E_{\mathbf{h},z_i}\left[\frac{z_i^*}{|z_i|}\mathbf{h}_i | \mathbf{y}_{T,pre}\right]$$

$$= E_{\mathbf{h}}\left[\frac{z_i^*}{|z_i|}E_{\mathbf{h}}[\mathbf{h}_i | \mathbf{y}_{T,pre}, z_i] | \mathbf{y}_{T,pre}\right] \tag{3.23}$$

As $p_{\mathbf{h}_i|\mathbf{y}_{T,pre},z_i}(\mathbf{h}_i|\mathbf{y}_{T,pre}, z_i)$ is complex Gaussian, the CM estimator $E_{\mathbf{h}}[\mathbf{h}_i|\mathbf{y}_{T,pre}, z_i]$ is then [32]

$$E_{\mathbf{h}}[\mathbf{h}_i|\mathbf{y}_{T,pre}, z_i] = E_{\mathbf{h}}[\mathbf{h}_i|\mathbf{y}_{T,pre}] + \mathbf{c}_{\mathbf{h}_i z_i|\mathbf{y}_{T,pre}} c^{-1}_{z_i|\mathbf{y}_{T,pre}}(z_i - E_{z_i}[z_i|\mathbf{y}_{T,pre}]) \tag{3.24}$$

with covariances

$$\mathbf{c}_{\mathbf{h}_i z_i|\mathbf{y}_{T,pre}} = E_{\mathbf{h},z_i}[(\mathbf{h}_i - E_{\mathbf{h}}[\mathbf{h}_i|\mathbf{y}_{T,pre}])(z_i - E_{z_i}[z_i|\mathbf{y}_{T,pre}])^*|\mathbf{y}_{T,pre}]$$

$$= \mathbf{C}_{\mathbf{h}_i|\mathbf{y}_{T,pre}}\mathbf{q}_i^*$$

$$c_{z_i|\mathbf{y}_{T,pre}} = E_{z_i}[|z_i - E_{z_i}[z_i|\mathbf{y}_{T,pre}]|^2|\mathbf{y}_{T,pre}] = \mathbf{q}_i^H \mathbf{C}^*_{\mathbf{h}_i|\mathbf{y}_{T,pre}}\mathbf{q}_i$$

and

$$\mu_{z_i|\mathbf{y}_{T,pre}} = E_{z_i}[z_i|\mathbf{y}_{T,pre}] = \hat{\mathbf{h}}_i^T \mathbf{q}_i$$

where $\hat{\mathbf{h}}_i$ is the Kalman estimator of $\mathbf{h}_i$.

Applying (3.24) to (3.23) yields

$$E_{\mathbf{h}}[g_i \mathbf{h}_i|\mathbf{y}_{T,pre}] = E_{\mathbf{h}}[g_i|\mathbf{y}_{T,pre}]E_{\mathbf{h}}[\mathbf{h}_i|\mathbf{y}_{T,pre}]$$

$$+\mathbf{c}_{\mathbf{h}_i z_i|\mathbf{y}_{T,pre}} c^{-1}_{z_i|\mathbf{y}_{T,pre}}(E_{z_i}[|z_i||\mathbf{y}_{T,pre}] - E_{\mathbf{h}}[g_i|\mathbf{y}_{T,pre}]E_{z_i}[z_i|\mathbf{y}_{T,pre}])$$

with [33], the remaining terms

$$\hat{g}_i = E_{\mathbf{h}}[g_i|\mathbf{y}_{T,pre}] = \frac{\sqrt{\pi}}{2}\frac{|\mu_{z_i|\mathbf{y}_{T,pre}}|}{c^{1/2}_{z_i|\mathbf{y}_{T,pre}}}\frac{\mu^*_{z_i|\mathbf{y}_{T,pre}}}{|\mu_{z_i|\mathbf{y}_{T,pre}}|}\,{}_1F_1\left(\frac{1}{2}, 2, -\frac{|\mu_{z_i|\mathbf{y}_{T,pre}}|^2}{c_{z_i|\mathbf{y}_{T,pre}}}\right)$$

$$E_{z_i}[|z_i||\mathbf{y}_{T,pre}] = \frac{\sqrt{\pi}}{2}c^{1/2}_{z_i|\mathbf{y}_{T,pre}}\,{}_1F_1\left(-\frac{1}{2}, 1, -\frac{|\mu_{z_i|\mathbf{y}_{T,pre}}|^2}{c_{z_i|\mathbf{y}_{T,pre}}}\right)$$

where ${}_1F_1(\alpha, \beta, \gamma)$ is the confluent hypergeometric function.

Summarizing the derivation, we obtain

$$E_{\mathbf{h}}[\mathbf{GH}|\mathbf{y}_{T,pre}] = \hat{\mathbf{G}}\hat{\mathbf{H}} + \mathbf{U}_{H|\mathbf{y}_{T,pre}} \tag{3.25}$$

where $\hat{\mathbf{G}} = E_{\mathbf{h}}[\mathbf{G}|\mathbf{y}_{T,pre}] = diag[\hat{g}_i]_{i=1}^{n_R}$ and $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_{n_R}]^T$ is the Kalman estimation of $\mathbf{H}$ in (3.17).

The $i\text{-}th$ row of $\mathbf{U}_{\mathbf{H}|\mathbf{y}_{T,pre}}$ is

$$\mathbf{e}_i^T\mathbf{U}_{\mathbf{H}|\mathbf{y}_{T,pre}} = \mathbf{q}_i^H\mathbf{C}^*_{\mathbf{h}_i|\mathbf{y}_{T,pre}}c^{-1}_{z_i|\mathbf{y}_T}(E_{z_i}[|z_i||\mathbf{y}_{T,pre}] - \mu_{z_i|\mathbf{y}_{T,pre}}\hat{g}_i) \tag{3.26}$$

where $\mathbf{C}_{\mathbf{h}_i|\mathbf{y}_{T,pre}}$ is the covariance matrix of $\mathbf{h}_i$ given $\mathbf{y}_{T,pre}$, and

$$z_i = \mathbf{h}_i^T\mathbf{q}_i$$

$$\begin{aligned} c_{z_i|\mathbf{y}_{T,pre}} &= E_{z_i}[|z_i - E_{z_i}[z_i|\mathbf{y}_{T,pre}]|^2|\mathbf{y}_{T,pre}] \\ &= \mathbf{q}_i^H\mathbf{C}^*_{\mathbf{h}_i|\mathbf{y}_{T,pre}}\mathbf{q}_i \end{aligned}$$

$$\mu_{z_i|\mathbf{y}_{T,pre}} = E_{z_i}[z_i|\mathbf{y}_{T,pre}] = \hat{\mathbf{h}}_i^T\mathbf{q}_i$$

The solutions of THP optimization is completed.

3.5 Computation Complexity Comparison

In this section, we will analyze the computation complexity of Kalman-THP and LMMSE-THP, and then compare the differences between them. By using the same THP optimization solutions, the computation complexity variation is based on the method of estimation. Thus, we will calculate the complexities of Kalman filter and LMMSE estimation only. One way to quantify this is with the notation of a $flop$ [24]. A flop is a floating point operation. For example, a dot product of length n involves $2n$ flops because there are n multiplications and n adds in either of these vector operations.

The common operations for the following calculations are matrix multiplication, matrix addition, Kronecker product and matrix pseudo inverse. For $complex$-number matrix multiplication, $\mathbf{XY}$ where $\mathbf{X} \in \mathbb{C}^{m\times p}$, $\mathbf{Y} \in \mathbb{C}^{p\times n}$ involves $mn(8p-2)$ flops, which include mnp complex multiplication ($6 \times mnp$ flops) and $mn(p-1)$ complex addition

$(2 \times mn(p-1)$ flops). To simplify the calculation, we regard the complexity as $8mnp$ as in [24]. For $complex$-number matrix addition, $\mathbf{Z} + \mathbf{V}$ where $\mathbf{Z} \in \mathbb{C}^{m \times n}$, $\mathbf{V} \in \mathbb{C}^{m \times n}$ involves $2mn$ flops. For $complex$-number Kronecker product, $\mathbf{X} \otimes \mathbf{Y}$ involves $6mnp^2$ flops. Here, for inversion, we only consider $real$-number case (we only need real-number matrix inversion in the following), $\mathbf{E}^{-1}$ where $\mathbf{E} \in \mathbb{R}^{m \times m}$ involves $(2/3)m^3$ flops (For more details, see [24]).

3.5.1 Computation Complexity of Kalman Filter

As we had described in Section 3.3, the solutions of Kalman filter are

$$\hat{\mathbf{h}}_T = \mathbf{A}\hat{\mathbf{h}}_{T,pre} + \mathbf{K}_{\mathbf{pre}}\tilde{\mathbf{y}}_{T,pre} \tag{3.27}$$

$$\tilde{\mathbf{y}}_{T,pre} = \mathbf{y}_{T,pre} - \mathbf{S}\hat{\mathbf{h}}_{T,pre} \tag{3.28}$$

$$\mathbf{K}_{pre} = (\mathbf{A}\mathbf{P}_{pre}\mathbf{S}^H)\mathbf{R}_{e,pre}^{-1} \tag{3.29}$$

$$\mathbf{R}_{e,pre} = \mathbf{R}_{nn} + \mathbf{S}\mathbf{P}_{pre}\mathbf{S}^H \tag{3.30}$$

$$\mathbf{P}_t = E\{\tilde{\mathbf{h}}_\mathbf{t}\tilde{\mathbf{h}}_t^H\} = \mathbf{A}\mathbf{P}_{pre}\mathbf{A}^H + \mathbf{B}\mathbf{B}^H - \mathbf{K}_{pre}\mathbf{R}_{e,pre}\mathbf{K}_{pre}^H \tag{3.31}$$

We first calculate the computation complexity of equation (3.31). We start with reminding the dimension of the matrices, $\mathbf{A} \in \mathbb{R}^{n_T n_R Q \times n_T n_R Q}$, $\mathbf{B} \in \mathbb{R}^{n_T n_R Q \times n_T n_R}$, $\mathbf{P}_{pre} \in \mathbb{C}^{n_T n_R Q \times n_T n_R Q}$, $\mathbf{K}_{pre} \in \mathbb{C}^{n_T n_R Q \times n_T NQ}$ and $\mathbf{R}_{e,pre} \in \mathbb{C}^{n_T NQ \times n_T NQ}$. Two $real$-number matrices $\mathbf{A}$ and $\mathbf{B}$ are given from autoregressive model in Section 2.3.2. The elements in $\mathbf{A}$ are calculated from equation (2.18) where $\mathbf{R} \in \mathbb{R}^{Q \times Q}$ and $\mathbf{v} \in \mathbb{R}^{Q \times 1}$, and thus involves $(2/3)Q^3 + 2Q^2$ flops. For matrix $\mathbf{B}$, the elements $b^{(i,j)}$ where $i \in \{1, \ldots, n_R\}$ and $j \in \{1, \ldots, n_T\}$ are calculated from equation (2.20), which involves $n_T n_R(Q+5)$ flops. $\mathbf{A}\mathbf{P}_{pre}\mathbf{A}^H$ involves $16(n_T n_R Q)^3$ flops, $\mathbf{B}\mathbf{B}^H$ involves $8(n_T n_R)^3 Q^2$ flops, $\mathbf{K}_{pre}\mathbf{R}_{e,pre}\mathbf{K}_{pre}^H$ involves $8n_T^3 n_R Q^3 N^2 + 8n_T^3 n_R^2 Q^3 N$ flops and the number of flops for addition and subtraction in (3.31) is $4(n_T n_R Q)^2$.

Next, for equation (3.30), the covariance matrix of $\mathbf{n}_{T,pre}$ is $\mathbf{R}_{nn} \in \mathbb{C}^{n_T NQ \times n_T NQ}$, the total training sequences in one time slot is $\mathbf{S} \in \mathbb{C}^{n_T NQ \times n_T n_R Q}$ and the minimum error covariance matrix at the previous time is $\mathbf{P}_{pre} \in \mathbb{C}^{n_T n_R Q \times n_T n_R Q}$. The computation complexity of $\mathbf{S} = \mathbf{I}_Q \otimes \mathbf{S}_{up}^T \otimes \mathbf{I}_{n_T}$ can be omitted since this parameter will also appeared

in LMMSE calculation. The addition in (3.30) involves $2(n_T NQ)^2$ flops and $\mathbf{SP}_{pre}\mathbf{S}^H$ involves $8n_T^3 n_R^2 Q^3 N + 8n_T^3 n_R Q^3 N^2$ flops.

Then, for equation (3.29), $\mathbf{AP}_{pre}\mathbf{S}^H$ involves $8(n_T n_R Q)^3 + 8n_T^3 n_R^2 Q^3 N$ flops, the pseudo inverse in $\mathbf{R}_{e,pre}^{-1}$ can also omitted since there's a same size ($n_T NQ \times n_T NQ$) pseudo inverse in LMMSE calculation. The multiplication between $(\mathbf{AP}_{pre}\mathbf{S}^H)$ and $\mathbf{R}_{e,pre}^{-1}$ involves $8n_T^3 n_R Q^3 N^2$ flops. Similarly, $8n_T^2 n_R QN + 2n_T NQ$ flops is needed in equation (3.28) and $8(n_t n_R Q)^2 + 8n_T^2 n_R Q^2 N + 2n_T n_R Q$ flops for equation (3.27).

Totally, the computation complexity of Kalman filter is written as

$$\begin{aligned} Number-of-flops_{Kalman} = 24(n_T n_R Q)^3 + 8(n_T n_R)^3 Q^2 + 24n_T^3 n_R Q^3 N^2 + 8(n_T n_R Q)^2 \\ + 24n_T^3 n_R^2 Q^3 N + 2(n_T NQ)^2 + 8n_T^2 n_R Q^2 N + \frac{2}{3}Q^3 \\ + 2n_T NQ + 2n_T n_R Q + 2Q^2 + n_T n_R (Q+5) \end{aligned} \tag{3.32}$$

3.5.2 Computation Complexity of LMMSE

In this section, before we analyze the computation complexity, we first introduce the LMMSE estimation shown in [23]

$$\hat{\mathbf{h}} = \mathbf{W}\mathbf{y}_{T,pre} \tag{3.33}$$

where

$$\mathbf{W} = \mathbf{C}_{\mathbf{h}\mathbf{h}_T}\mathbf{S}^H(\mathbf{S}\mathbf{C}_{\mathbf{h}_T}\mathbf{S}^H + \sigma_n^2\mathbf{I}_{n_T NQ})^{-1} \tag{3.34}$$

$$C_{\mathbf{h}|\mathbf{y}_T} = \mathbf{C}_{\mathbf{h}} - \mathbf{W}\mathbf{S}\mathbf{C}_{\mathbf{h}\mathbf{h}_T}^H \tag{3.35}$$

where $C_{\mathbf{h}|\mathbf{y}_T}$ is the covariance matrix of $\mathbf{h}$ given $\mathbf{y}_T$, $\mathbf{C}_{\mathbf{h}\mathbf{h}_T} = E_{\mathbf{h}}[\mathbf{h}\mathbf{h}_T^H] = [r[l_1] \dots r[l_Q]] \otimes \mathbf{R}_{\mathbf{h}}$, $\mathbf{h} = vec(\mathbf{H}^T)$ and $\mathbf{C}_{\mathbf{h}} = E_{\mathbf{h}}[\mathbf{h}(t)\mathbf{h}(t)^H]$, which is block diagonal assuming channels from different receivers are uncorrelated, i.e., $E_{\mathbf{h}}[\mathbf{h}_i(t)\mathbf{h}_{i'}(t)^H] = \mathbf{C}_{\mathbf{h}_i}\delta_{ii'}$. The i-th column of $\mathbf{H}(t)^T$ is $\mathbf{h}_i(t) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{C}_{\mathbf{h}_i})$ as in Section 3.1.1. $\mathbf{C}_{\mathbf{h}_T} = E_{\mathbf{h}}[\mathbf{h}_T\mathbf{h}_T^H] = \mathbf{C}_T \otimes \mathbf{C}_{\mathbf{h}}$ where $\mathbf{C}_T$ is Toeplitz with first column $[r[0] \quad r[2] \dots r[2Q-2]]^T$. Since $\mathbf{C}_{\mathbf{h}} \in \mathbb{C}^{n_T n_R \times n_T n_R}$ and $\mathbf{C}_T \in \mathbb{C}^{Q \times Q}$, the number of flops for calculating $\mathbf{C}_{\mathbf{h}\mathbf{h}_T}$ is $2(n_T n_R)^2 Q$ and $\mathbf{C}_{\mathbf{h}_T}$ is $2(n_T n_R Q)^2$.

The dimension of matrices in (3.34) are $\mathbf{C}_{\mathbf{hh}_T} \in \mathbb{C}^{n_T n_R \times n_T n_R Q}$, $\mathbf{C}_{\mathbf{h}_T} \in \mathbb{C}^{n_T n_R \times n_T n_R Q}$ and $\mathbf{S} \in \mathbb{C}^{n_T NQ \times n_T n_R Q}$. The generation of $\mathbf{C}_{\mathbf{hh}_T}$ involves $2(n_T n_R)^2 Q$ flops and $\mathbf{C}_{\mathbf{h}_T}$ involves $2(n_T n_R Q)^2$ flops. The multiplication of $\mathbf{C}_{\mathbf{hh}_T}\mathbf{S}^H$ involves $8n_T^3 n_R^2 Q^2 N$ flops, $\mathbf{S}\mathbf{C}_{\mathbf{h}_T}\mathbf{S}^H$ involves $8n_T^3 n_R Q^3 N^2 + 8n_T^3 n_R^2 Q^3 N$ flops, and the multiplication between $\mathbf{C}_{\mathbf{hh}_T}\mathbf{S}^H$ and $(\mathbf{S}\mathbf{C}_{\mathbf{h}_T}\mathbf{S}^H + \sigma_n^2 \mathbf{I}_{n_T NQ})^{-1}$ involves $8n_T^3 n_R Q^3 N^2$ flops. $2(n_T QN)^2$ flops are included for the addition. As in Kalman filter, pseudo inverse in $(\mathbf{S}\mathbf{C}_{\mathbf{h}_T}\mathbf{S}^H + \sigma_n^2 \mathbf{I}_{n_T NQ})^{-1}$ is omitted. As the complexity of $\mathbf{W}$ is produced, the number of flops in equation (3.33) is $8n_T^2 n_R QN$.

Similarly, the total number of flops in equation (3.35) is $2(n_T n_R)^2 + 8n_T^3 n_R^2 Q^2 N + 8n_T^3 n_R^3 Q$. Finally, the overall computation complexity of LMMSE reads

$$\begin{aligned} Number-of-flops_{LMMSE} = {} & 2(n_T n_R)^2 Q + 2(n_T n_R Q)^2 + 16 n_T^3 n_R^2 Q^2 N + 2(n_T QN)^2 \\ & + 8n_T^3 n_R Q^3 N^2 + 8n_T^2 n_R QN + 8n_T^3 n_R Q^2 N^2 \\ & + 2(n_T n_R)^2 + 8(n_T n_R)^3 Q + 8n_T^3 n_R^2 Q^3 N \end{aligned} \quad (3.36)$$

Since n_T and n_R denote the number of transmit antennas and receivers, and Q denote the order of AR model (which is assumed to low order in our work), these three parameters' value are close. Even though the total number of training sequences in one time slot N is much larger than n_T, n_R and Q, we don't see the dominant items appear in either equation (3.32) or (3.36). The comparison is difficult by assuming variables, thus we will further compare them case by case in simulation results.

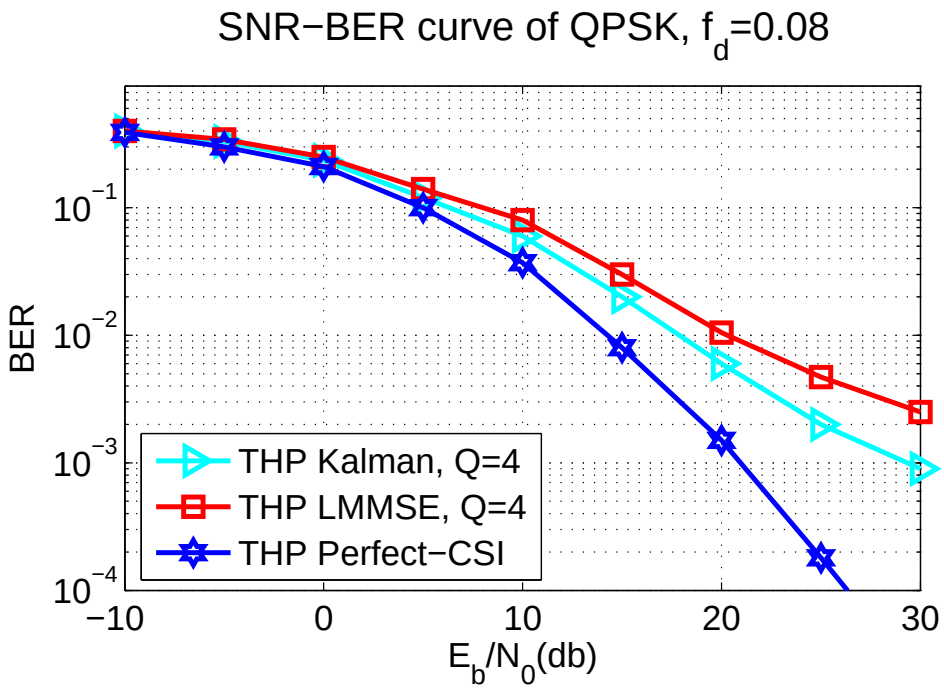
Chapter 4

Simulations

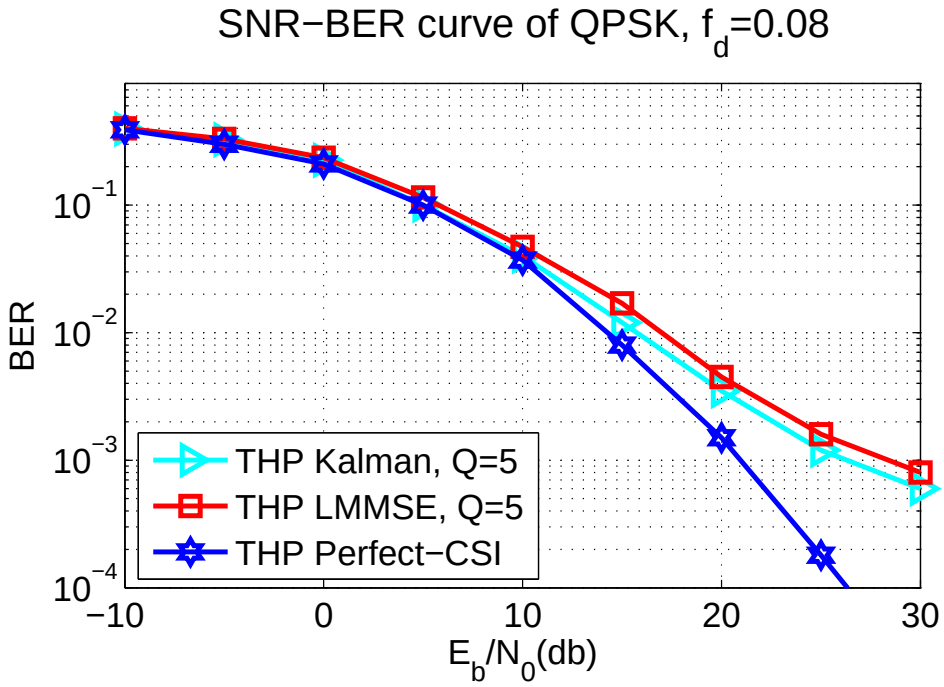
4.1 Simulation Setup

In this chapter, we consider an example with $n_T = 4$ transmit antennas and $n_R = 3$ receivers with one antenna each, and data streams are transmitted through a Rayleigh flat fading channel. The temporal autocorrelation of the complex Gaussian channel coefficients is identical for all coefficients, and is corresponds to Jakes' power spectrum with maximum normalized Doppler frequency f_d, i.e. Doppler frequency is normalized to the time slot period T, $f_d = f_D/(1/T)$ where f_D is the maximum Doppler frequency. System parameters for UMTS UTRA TDD systems is taken from [34], i.e. a carrier frequency of 2 GHz and a slot period of 0.675 ms. As an example, a maximum normalized Doppler frequency $f_d = 0.08$ corresponds to a velocity of 64 km/hr for these parameters.

Assuming an alternating uplink/downlink slots as shown in Fig. 3.1, and a worst-case delay of three time slots to the first uplink slot available with the training sequence. Thus, the observation is $\mathbf{y}_{T,pre} = [\bar{\mathbf{y}}(t-3)^T, \bar{\mathbf{y}}(t-5)^T \ldots, \bar{\mathbf{y}}(t-(2Q+1))^T]^T \in \mathbb{C}^{n_T NQ \times 1}$ for $Q = 5$ uplink slots and sequences of length $N = 32$ are used for the training uplink channel. To simplify, we assume $\mathbf{C_a} = \mathbf{I}_{n_R}$, $\mathbf{C}_{\mathbf{h}_i} = \mathbf{I}_{n_T}$ and $\mathbf{C_n} = \sigma_n^2 \mathbf{I}_{n_R}$ where SNR=$10\log_{10}(P_T/\sigma_n^2)$ and $P_T = 1$. For the results shown in the following figures, 300 QPSK data symbols were transmitted over 100 time slots per channel realization and averaged over 100 independent channel realizations, i.e., 300 symbols are totally average over 100 slots$\times$ 100 channel realizations. "THP LMMSE" is the THP optimization with



(a)



(b)

Figure 4.1: Performances of uncoded BER versus SNR for f_d=0.08. (a) Both THP Kalman and THP LMMSE had uplink 4 time slots. (b) Both THP Kalman and THP LMMSE had uplink 5 time slots.

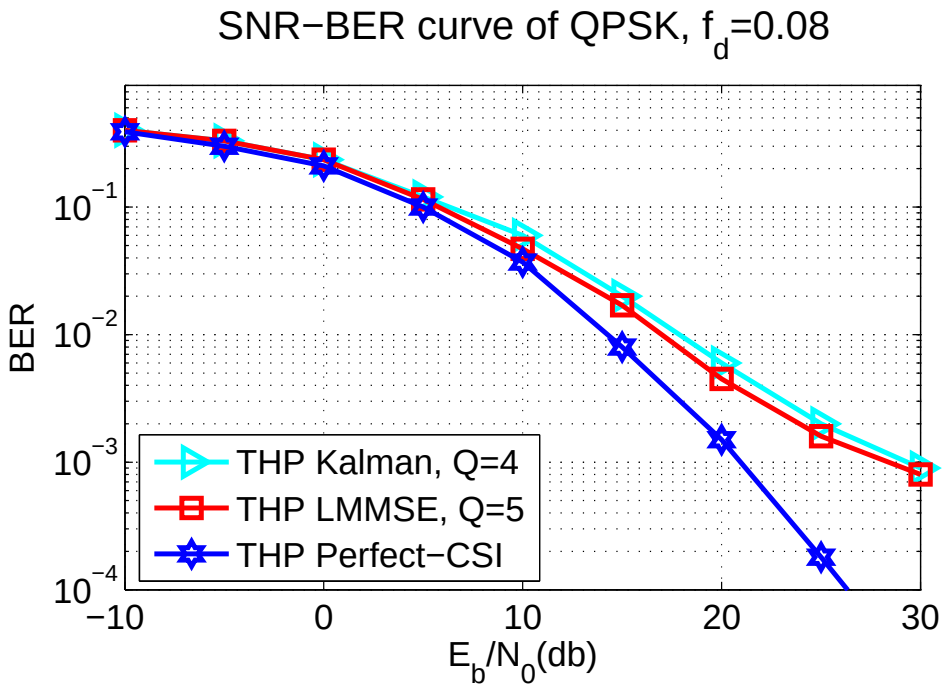


Figure 4.2: Performances of uncoded BER versus SNR for f_d=0.08. In this case, THP Kalman uplinks 4 slots and THP LMMSE uplinks 5 slots.

LMMSE for channel estimation as in [23] and "THP Kalman" is the THP optimization with Kalman estimation for channel tracking.

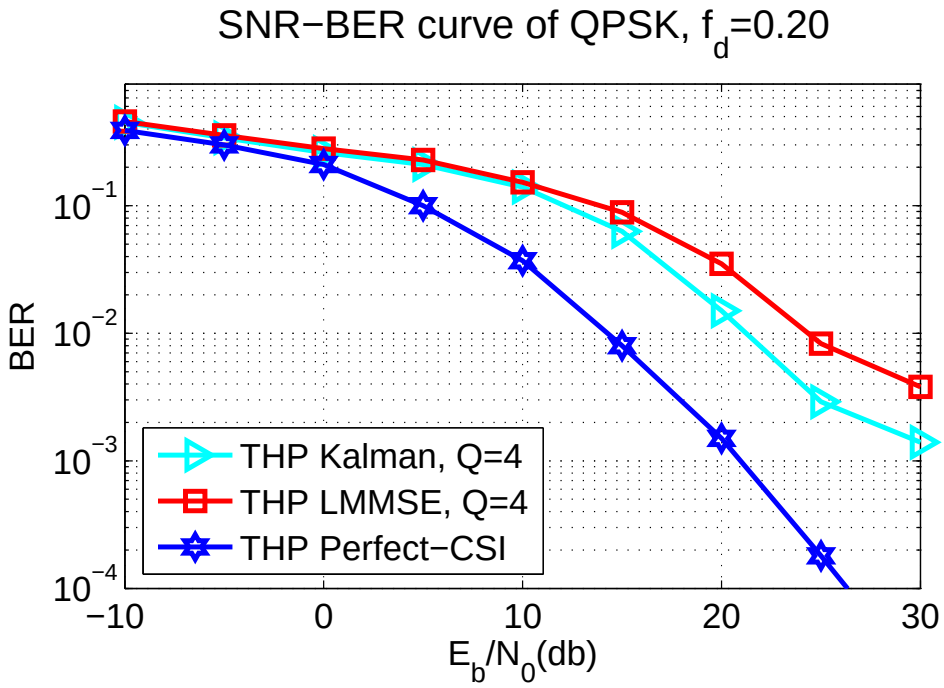
By setting $n_T = 4$, $n_R = 3$ and $N = 32$, the computation complexity of THP-Kalman and THP-LMMSE can be written as a function of Q

$$Number - of - flops_{Kalman} = (5,202,432 + \frac{2}{3})Q^3 + 72,322Q^2 + 292Q + 60 \quad (4.1)$$

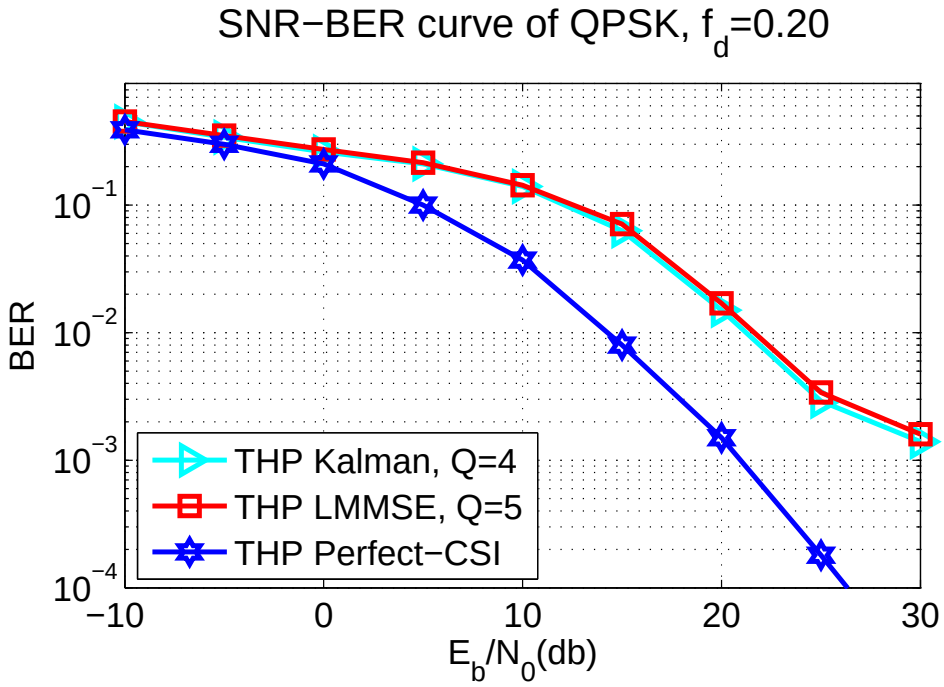
$$Number - of - flops_{LMMSE} = 1,720,320Q^3 + 1,900,832Q^2 + 26,400Q + 144 \quad (4.2)$$

4.2 Numerical Results

Fig. 4.1(a), Fig. 4.1(b) and Fig. 4.2 show results for comparing two kinds of estimators with different number pairs of uplink slots for normalized Doppler frequency f_d=0.08. In Fig. 4.1(a), the BER of THP-Kalman is lower due to the adding up channel information



(a)



(b)

Figure 4.3: Performances of uncoded BER versus SNR for f_d=0.20. (a) Both THP Kalman and THP LMMSE had uplink 4 time slots. (b) THP Kalman uplinks 4 slots and THP LMMSE uplinks 5 slots.

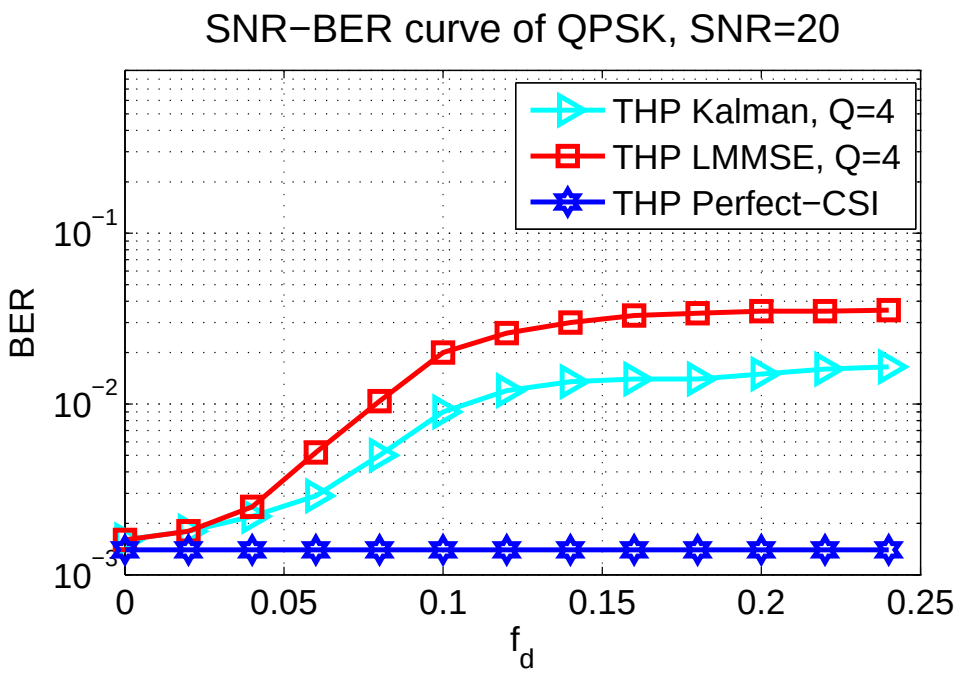


Figure 4.4: Performance of uncoded BER versus normalized Doppler frequency for SNR=20dB.

by considering Kalman estimation, on the other hand, the channel information for THP-LMMSE is only obtained by 4 presently uplink slots. The computation complexity of THP-Kalman and THP-LMMSE in this case is $334,114,070$ flops and $140,619,680$ flops, both of the computation complexity are in the same order, i.e.,10^8, the cost of THP-Kalman is almost two times larger than THP-LMMSE, thus it's a trade off between BER and complexity. In Fig. 4.1(b), the number of uplink slots for both THP-LMMSE and THP-Kalman is adding up to 5. In this case, it is not surprising that the complexity of THP-Kalman is still larger than THP-LMMSE, THP-Kalman involves $652,113,653$ flops and THP-LMMSE involves $262,693,088$ flops, which is still a trade off between BER and complexity. As we increase the number of uplink slots of THP-LMMSE to Q=5, as shown in Fig. 4.2, the BER shows better performance for all SNR. By comparing Fig. 4.1(a) and Fig. 4.1(b), it is clear to find out that THP-LMMSE is more sensitive to the number of uplink slots, and the BER of THP-Kalman achieves almost the same value for Q=4 and Q=5. The computation complexity of THP-Kalman and THP-LMMSE in this

case is $334,114,070$ flops and $262,693,088$ flops. Even though the complexity variation between THP-Kalman and THP-LMMSE are smaller, the BER of THP-Kalman still has small loss compared to THP-LMMSE. Thus, THP-Kalman is not a appropriate choice in $f_d = 0.08$ case.

Fig. 4.3(a) and Fig. 4.3(b) show results for comparing two kinds of estimators with different number pairs of uplink slots for normalized Doppler frequency f_d=0.20. In Fig. 4.3(a), the differential value between the BER of THP-LMMSE and THP-Kalman is larger than in Fig. 4.1(a), and this shows that THP-Kalman performs the flexibility in fast fading channel. The computation complexity of THP-Kalman and THP-LMMSE in this case is $334,114,070$ flops and $140,619,680$ flops, both of the computation complexity are in the same order. This is still a trade off between BER and computation complexity. Further, THP-Kalman can achieve even better performance for less number of uplink slots, as shown in Fig. 4.3(b). The computation complexity of THP-Kalman and THP-LMMSE in this case is $334,114,070$ flops and $262,693,088$ flops, this variation is acceptable. To sum up, the proposed method is appropriate in fast fading channel($f_d = 0.20$) case for a acceptable complexity.

Fig. 4.4 shows result for a fixed SNR=20dB versus normalized Doppler frequency f_d. Evidently, the performance of THP-Kalman perform better in all f_d but having more cost in computation complexity.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

We had studies Tomlinson-Harashima precoding optimization problem with partial channel-state-information under a Rayleigh flat fading channel. Kalman estimator is introduced for channel tracking and is combined into THP design. Kalman estimator is on the basis of random process, which has collect up all previous information and performs a robust estimation. Simulation results has shows better results compare to LMMSE-based THP in BER. Also, Kalman-based THP acts flexible as the Doppler frequency changes. From the view of calculating computation complexity, the Kalman-based THP achieve better BER and a close complexity in fast fading channel. With these features, the proposed Kalman-based THP is acceptable for application in the fast fading wireless broadcast channel.

5.2 Future Work

While in our work, we have proposed an effective Kalman-based THP algorithm, which decrease the BER evidently, we have not discuss about the accuracy for designing the autoregressive model, which can further improve the correctness of Kalman estimator. Also, the complexity of the THP optimization has not been analyzed.

Future works might consider the design of the AR model, i.e., in (2.13), instead of using the traditional AR parameters $\mathbf{A}$ and $\mathbf{B}$, which are chosen based on Jakes' model. In

order to approximate the real channel, the parameters can be further designed by considering the channel information at the transmitter, which can be obtained by uplink training sequences. Also, future works could take account of the reduction of THP optimization. The complexity of THP optimization can be reduced by slipping the interference mitigation into two parts, which is named “cascaded THP-BD system” in [35], i.e., mitigating multi-user interference using Block Diagonalization (To orthogonalize the channels) and then mitigate inter-stream interference (the effect of multi-antenna interference) using a simpler THP.

Bibliography

[1] C. Komninakis, C. Fragouli, A. Sayed, and R. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1065–1076, May 2002.

[2] B. R. Vojcicand W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, pp. 1346–1355, Octobor 1998.

[3] R. Knopp and G. Caire, "Power control schemes for TDD systems with multiple transmit and receive antennas," *in Proc. IEEE Global Telecommunications Conference (Globecom)*, pp. 2326–2330, Rio de Janeiro, November 1999.

[4] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Processing*, pp. 2921–2936, Octobor 2000.

[5] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multiinput multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, pp. 1334–1347, September 1994.

[6] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, March 1971.

[7] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. COM-20, no. 8, pp. 774–780, Augest 1972.

[8] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in Multiantenna and Multiuser Communications," *IEEE Trans. Wireless Commun*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[9] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[10] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," *in Proc. Asilomar Conf. Signals, Systems,Computers*, vol. 2, pp. 1627–1631, October 2000.

[11] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," *in Proc. IEEE Global Telecommunications (IEEE GLOBECOM) Conf,*, vol. 2, pp. 1344–1348, November 2001.

[12] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, "Space-time transmission using Tomlinson-Harashima precoding," *in Proc. Proceedings of 4th ITG Conference on Source and Channel Coding*, January 2002.

[13] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," *in Proc. ITG Conf. Source Channel Coding (ITG SCC04)*, pp. 387–394, January 2004.

[14] M. Joham and W. Utschick, "Ordered spatial Tomlinson-Harashima precoding," *Smart Antennas–State-of-the-Art, ser. EURASIP Book Series on Signal Processing and Communications. New York: EURASIP, Hindawi Publishing Corporation*, vol. 3, 2005.

[15] R. D. Wesel and J. Cioffi, "Achievable Rates for Tomlinson-Harashima Precoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, March 1998.

[16] M. Payaro, A. P. Neira, and M. A. Lagunas, "Achievable Rates for Generalized Spatial Tomlonson-Harashima Precoding in MIMO Systems," *IEEE Vehicular Technology Conference (VTC)*, vol. 4, pp. 2462–2466, 2004.

[17] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel," *in Proc. Int. Conf. Communications (ICC)*, vol. 3, pp. 2021–2025, May 2005.

[18] "Cholesky Factorization with Symmetric Permutation applied to Detecting and Precoding Spatially Multiplexed Data Streams," *IEEE Trans. Signal Processing*, 2006, accepted for publication.

[19] F. A. Dietrich and W. Utschick, "Robust Tomlinson-Harashima Precoding," *in Proc. 16th IEEE Symp. Personal, Indoor Mobile Radio Communications, Berlin, Germany*, pp. 136–140, 2005.

[20] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," *in Proc. Int. Zurich Seminar Broadband Communications*, pp. 7–1–7–6, February 2002.

[21] A. P. Liavas, "Tomlinson-Harashima precoding with partial channel knowledge," *IEEE Trans. Commun.*, vol. 53, pp. 5–9, January 2005.

[22] P. M. Castro, L. Castedo, and J. Migue, "Adaptive Precoding in MIMO Wireless Communication Systems Using Blind Channel Prediction over Frequency Selective Channels," *Universidad de A Coruna, Facultad de Informatica Departamento de Electronica y Sistemas Campus de Elvina, s/n. 15071, A Coruna, SPAIN*, 2005.

[23] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima Precoding for the Wireless Broadcast Channel," *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 631–644, February 2007.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore, MA: The Johns Hopkins Univ. Press, 3rd ed edition, 1996.

[25] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *presented at the ISSSE-98, Pise, Italy,*, September 1998.

[26] J. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841–1852, November 1999.

[27] W. C. Jakes Jr, *Microwave Mobile Communizations*, New York: Wiley, 1974.

[28] K. E. Baddour and N. C. Beaulieu, "Autoregressive models for fading channel simulation," *in Proc. IEEE Global Telecommun. Conf.*, vol. 2, pp. 1187–1192, 2001.

[29] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Processing*, vol. 53, pp. 211–229, 1996.

[30] F. A. Dietrich, F. Hoffmann, and W. Utschick, "Conditional mean estimator for the Gramian matrix of complex Gaussian random variables," *in Proc. Int. Conf. Acoustic, Speech, Signal Processing(ICASSP), Philadelphia, PA*, vol. 3, pp. 1137–1140, March 2005.

[31] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Englewood Cliffs, NJ: Prentice- Hall, 2nd ed edition, 1994.

[32] S. M. Kay, *Fundamentals of Statistical Signal ProcessingXEstimation Theory*, Englewood Cliffs, NJ: PTR Prentice-Hall, 1st ed edition, 1993.

[33] K. S. Miller, *Complex Stochastic Processes*, Reading, MA: Addison-Wesley, 1st ed edition, 1974.

[34] Version 4.3.0 3GPP 2001[Online]. Available: http://www.3gpp.org Technical Specifications, 3GPP TS 25.221, ," .

[35] D. Sharma and S. N. Kizhakkemadam, "Cascaded Tomlinson-Harashima Precoding and Block Diagonalization for Multi-User MIMO," *2011 National Conference on Communications (NCC)*, January 2011.