國 立 交 通 大 學

電信工程學系

博 士 論 文

應用於相關性多輸入多輸出系統
之通道建模、估測及前置編碼

Channel Representation, Estimation and Precoding for Correlated MIMO Systems

研 究 生：陳彥志

指導教授：蘇育德 博士

中 華 民 國 九十八 年 七 月

應用於相關性多輸入多輸出系統之通道建模、估測及前置編碼

Channel Representation, Estimation and Precoding for Correlated MIMO Systems

研究生：陳彥志　　Student: Yen-Chih Chen

指導教授：蘇育德　博士　　Advisor: Dr. Y. T. Su

國立交通大學

電信工程學系

博士論文

A Dissertation

Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in
Communication Engineering
Hsinchu, Taiwan

2009 年 7 月

應用於相關性多輸入多輸出系統之通道建模、估測及前置編碼

研究生：陳彥志　　指導教授：蘇育德 博士

國立交通大學 電信工程學系博士班

摘要

有別於傳統單一天線系統，多輸入多輸出(MIMO)技術由於能大幅提升通道傳輸容量，因此已被納入目前許多重要的無線通訊標準之中。藉由在傳輸端與接收端設置多根天線，我們可以透過分解通道矩陣來創造許多平行通道並用以同時傳輸多個資料流。為了充份發揮多天線系統的優勢，特別是要進行高速資料傳輸時，精準的通道狀態資訊通常是不可獲缺的。然而，隨著天線個數的增長，估測及處理龐大通道矩陣的工作顯得益發困難。在本論文中，我們提出一種簡潔有效的通道矩陣表示法，藉由此表示法我們可以減少在描述通道矩陣時所需的參數數量，同時也減輕後續信號的運算複雜度。在中至高度相關的多天線傳輸環境下，使用本文所提出的通道表示法將可大幅減少通道狀態資訊的參數個數，同時維持良好的資訊品質。

基於所提之通道描述，我們發展了一種遞迴最小平方法來估測幾種典型的多輸出入通道。所得到的通道估測值呈現一個緊緻的形式，該形式將有助於簡化許多需要利用通道矩陣估測值進行的後置信號處理程序。此外，藉由調整通道估測器中的一個模型階數，我們可以在演算法的估測準確性和計算複雜度之間取得平衡。值得一提的是，由於估測器中維度縮減特性所帶來額外的雜訊消除效果，我們將可得到比傳統最小平方估測法更優越的均方誤差表現。我們對所提的通道估測器之相關性能也作了理論分析並就不同通道環境進行數值模擬用以評估該通道估測器的效能並證明理論的正確性。

為了能充份發揮所提通道模式及估測器所帶來的好處，我們更進一步利用該模式發展新型的回饋前置編碼系統。由於在設計前置編碼器時引入前述的通道模式，我們可大幅減少回饋前置編碼系統所需的回授頻寬，並降低建構前置編碼器及後置等化器的

計算複雜度。相較於傳統上使用完整即時通道資訊的前置編碼系統，我們的系統只在非常高訊號雜音比時造成極輕微的性能損失，然而其在縮減回授頻寬和簡化計算複雜度上所帶來的好處卻是相對可觀。為了進一步評估系統因為模型通道模式的誤差所帶來可能的效能損失，我們在數學上推導了數個效能上界，用以評估訊號接收的均方誤差及回授訊號的資訊品質。同樣地，我們也提供了相關的數值結果用以驗証系統效能並証實所推導的效能上界的確可以準確預測系統的效能趨勢。

Channel Representation, Estimation and Precoding for Correlated MIMO Systems

Student: Yen-Chih Chen Advisor: Y. T. Su

Department of Communications Engineering

National Chiao Tung University

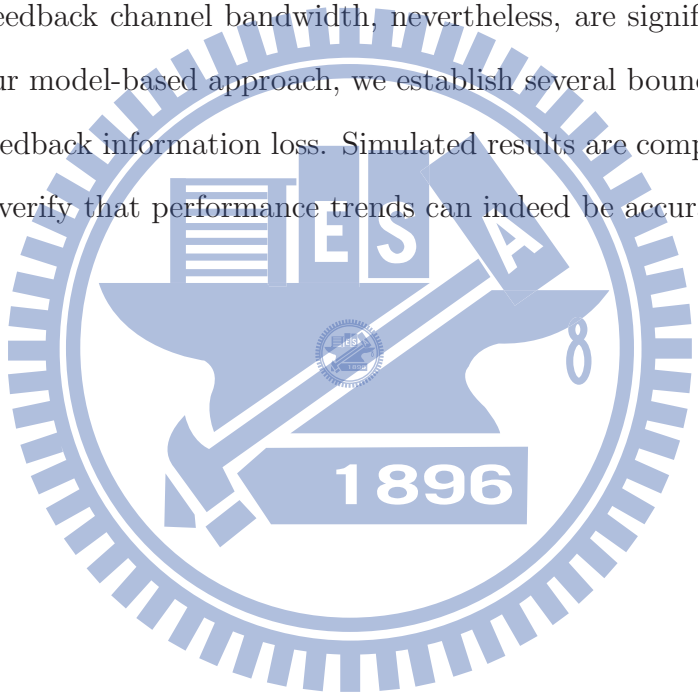
Abstract

Multiple-input multiple-output (MIMO) technology has been included in many industrial standards to achieve significant throughput enhancement compared with conventional single antenna systems. By using multi-element antennas at both transmit and receive sides, multiple data streams can be transmitted simultaneously through parallel spatial modes. To realize the advantages of MIMO systems, accurate channel state information (CSI) is indispensable, especially for high rate transmissions. With the increase of antenna number, the task of estimating or processing a MIMO channel matrix becomes more and more difficult. In this thesis, we propose an efficient channel representation such that the number of required parameters is reduced and the computation complexity can be lessened as well. For medially to highly correlated MIMO environments, the proposed representation can lead to significant parametric dimension reduction while maintaining good CSI quality.

Based on the proposed channel representation, we develop iterative least squared (LS) schemes to estimate several typical MIMO channels. The reduced-rank CSI representation is very useful for many post-channel-estimation operations that require processing the instantaneous channel matrices. Depending on the specified modelling order, the proposed channel estimators offer tradeoff between identification accuracy and computational complexity. Moreover, the dimension-reduction induced noise rejection effect enables the proposed model-based estimators to achieve superior mean squared error (MSE) performance over certain SNR region when compared with that of the conventional LS approach.

Theoretical analysis and numerical simulations of MSE performance are provided to assess the estimators' performance and validate the analytical predictions.

Taking advantage of the proposed compact CSI representation, we proceed to develop a model-based feedback precoded system. By incorporating our new channel representation into the precoder design, the resulting precoded system provides significant reductions on the feedback bandwidth and the computational complexity needed for constructing the precoder and equalizer matrices. Numerical results show that compared with the conventional approaches that need full knowledge of instantaneous CSI, our proposal suffers only negligible performance degradation at very high SNR region. The reductions on computing complexity and feedback channel bandwidth, nevertheless, are significant. To assess the performance of our model-based approach, we establish several bounds regarding the reception error and feedback information loss. Simulated results are compared with these analytical bounds to verify that performance trends can indeed be accurate predicted.



誌謝

本論文得以完成，首先要感謝恩師蘇育德教授。在漫長的博士班生涯中，老師對於學術領域的專注、熱情與不斷學習的態度，激勵著我在研究的道路上前進。除了學術之外，老師對於學生們不管是在生活上的幫助、經濟上的支援或精神上的開導，都使我深感慶幸能在老師的照顧下完成學位。此外，感謝在博士班口試期間蒞臨指導的口試委員，教授們專業的意見和想法使得本論文得以更加完整。感謝實驗室同儕鄭延修、李昌明、林淵斌、劉人仰在博士班就讀期間給予我學術上的交流和生活上的關照，我從你們身上學習到許多，同時也得到許多美好的生活回憶。

感謝最親愛的老婆翠如，妳無怨無悔的支持和愛，是我身心上最大的動力來源，我們一起經歷的磨練終於結為甜美的果實，獻上我最真摯的愛和感激。感謝父母親、家人及岳父岳母一直以來的鼓勵和照顧，暖暖的親情護佑著我完成攻讀博士學位的夢想。最後，謹以此論文獻給所有關心過我的人，在未來的人生道路上，我會更加努力的。

Contents

List of Figures

Chapter 1

Introduction

Increasing demand for higher wireless system capacity has catalyzed several groundbreaking transmission techniques, among which is the multiple-input/multiple-output (MIMO) technology that has attracted the great part of recent attention. It has been shown that in comparison with conventional single antenna systems, significant capacity gains are achievable when multi-element antennas (MEA) are used at both the transmit and receive sides [4],[5]. Spatial multiplexing techniques, for example, the BLAST (Bell-labs Layered Space-Time) system, was developed to attain very high spectral efficiencies in rich scattering environments.

Ideal rich-scattering environments decorrelate channels between different pairs of transmit and receive antennas so that maximum number of orthogonal subchannels is available. In practice, however, spatial correlations do exist and should be considered when designing a MIMO receiver for evaluating the corresponding system performance [6]. Spatial correlation depends on physical parameters such as antenna spacing, antenna arrangement, and scatters' distributions. Antenna correlations reduce the number of equivalent orthogonal subchannels, decrease spectral efficiency, making it more difficult to detect the transmitted data [4].

A coherent MIMO receiver requires an accurate channel estimate to perform critical operations and provide satisfactory performance. Not only is reliable channel estimation mandatory in guaranteeing signal reception quality but it is also needed in designing an adequate precoder at the transmit side to achieve maximum throughput or minimum bit error rate in feedback MIMO systems. Various pilot-assisted MIMO channel estimators

have been proposed [7, 8]. Unfortunately, few estimators are specifically designed for correlated MIMO channels and those few exploited only channel's time and frequency correlation characteristics by approximating the time- and/or frequency-domain responses by an analytic model [8, 9]. These analytic model based approaches can do without the channel information like covariance functions and signal-to-noise ratio which are required by most estimators and are to be obtained by on-line measurements. However, they fail to take into account and the advantage of the spatial structure of such channels which has significant impact on the system performance and should also be explored. The spatial correlation structure instead was often used to analyze the system capacity [10], to design beamformer [11] or pilot sequences [7, 12].

We present novel pilot-assisted channel estimation schemes on the basis of the proposed new general MIMO channel representation which does not require information of second-order channel statistics. Spatial and time covariance (or correlation) functions are described by nonparametric regression and the influence of the mean angle of departure (AoD) is related to other channel parameters via a regression model. This representation admits a reduced-rank channel model and compact channel state information (CSI) representation, making possible reduced feedback channel bandwidth requirement. It results in separable descriptions of channel correlations and mean AoD for correlated MIMO systems and enables us to develop efficient algorithms to identify the realistic channel responses. Although a model-based scheme inevitably induces a modelling error [9]-[13], as will be shown in Chapter 3, our algorithms are capable of describing realistic correlated MIMO channels with negligible modelling errors. The estimated CSI can be efficiently exploited for use in many channel estimation related operations such as MIMO data detection and optimal MIMO transceiver designs.

Optimal MIMO transceiver designs based on CSI at the transmitter (CSIT) have been thoroughly studied under several performance measures such as minimum mean squared error (MMSE) or maximum mutual information (MMI) [14]-[11]. When CSI is available at both ends of a link, conventional precoding-eigen-beamforming schemes can adapt to the channel condition to optimize the reception performance in the correlated

environment. However, in practice, downlink CSI (from base station to mobile unit) is often not available at transmit site and has to be estimated unless the channel transfer function can be assumed to be identical in both directions. Oftentimes, the downlink receiver has to send the information back to the transmitter through a feedback channel. It is thus critical that one control the amount of feedback information as the feedback channel usually has a very limited bandwidth.

To lessen the feedback load, several transmitter precoding/eigen-beamforming schemes based on partial channel information such as channel mean feedback and channel covariance feedback were proposed to reduce the feedback cost [15]-[16]. Mean feedback relies on the proposition that CSI resides in the mean of the distribution with white covariance. Therefore, only for very slowly faded channels can mean feedback adequately capture the channel behavior. On the other hand, covariance feedback models the channels as random vectors with zero mean and non-white covariance, which are only hold for rapid fading environments. Both feedback scenarios relies on imperfect long term statistical models and thus cannot well represent the instant or short term channel variations. Moreover, prior knowledge of channel statistics are often needed to compute the approximated feedback information. Generally, systems using statistical feedback come with a non-negligible performance loss compared with those using instantaneous channel realization.

Based on the proposed model-based channel representation, the instantaneous CSI is represented in a more compact form and estimated accordingly. With this efficient CSI estimation, we present a framework of transceiver design to render the advantage of the proposed model-based structure. For correlated MIMO channels, the proposed precoding scheme provides an alternative to reduce the requirement of instantaneous CSI feedback, while retaining or even improving the reception performance. Several performance bounds regarding reception error and feedback information loss are established to assess the system performance.

The rest of this thesis is arranged as follows. After a brief review of the typical space-time antenna setup and a general received MIMO signal model, we derive two new models [17] for spatial-correlated block-faded narrowband MIMO channels and their rela-

tions with some established analytic models in Section 2.3. We then propose single-block based iterative least squares (LS) channel estimators in Section 3.1 while the extension that takes the time-correlation and frequency-selective cases into account are given in Section 3.2 and 3.3, respectively. In Section 3.4, we analyze the mean squared error (MSE) of the proposed channel estimation algorithms. Numerical examples using industrial standard approved channel models are given in Section 3.5 to validate the proposed channel models and to demonstrate the effectiveness of our algorithms. In Chapter 4, we develop the basic framework of transceiver designs based on the reduced rank CSIT. Section 4.1 quickly reviews the channel representation proposed in Section 2.3 as the foundation for the proposed MIMO eigen-beamforming system. Section 4.2.1 and Section 4.2.2 give a brief review of some conventional MIMO precoder/beamforming systems with feedback CSI. Section 4.2.3 make use of the proposed channel representation to establish a nonparametric CSIT. With the nonparametric CSIT, the proposed eigen-beamforming design is developed and discussed in Section 4.2.4. Performance analysis of the proposed beamforming method is given in Section 4.3. In Section 4.5, we provide several numerical and simulation examples by using some well-established industrial channel models. Conclusion and remarks are given in Section 4.6. Chapter 5 summarizes the studies in this thesis and suggests some interesting research subjects under the framework of the proposed nonparametric scheme.

Chapter 2

MIMO Channel Representation

2.1 MIMO System

In this thesis, we focus on the clustered channel model. In such a MIMO setup, MS is surrounded by local scatterers and waveforms impending the receive antennas are richly scattered. On the other hand, BS is often unobstructed by local scatterers and has a mean angle of departure (AOD) with respect to the receiver cluster. The clustered channel setup is typical in urban environments, and has been validated through field measurements. A typical "one-ring" model is shown in Fig. 2.1, Δ denotes the azimuthal angle spread at the BS and ϕ denotes the mean AOD between BS and MS.

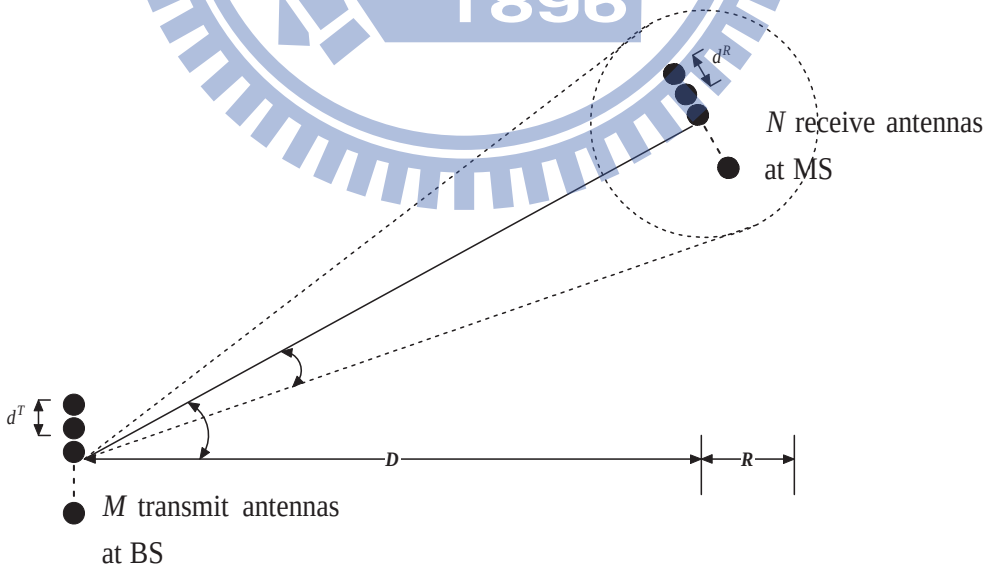


Figure 2.1: "One-ring" model with M transmit antennas at BS and N receive antennas at MS. D: distance from BS to MS. R: radius of the scatterer ring. ϕ: angle of departure. Δ: angle spread at BS. d^T: antenna spacing at BS. d^R: antenna spacing at MS.

2.2 Modelling Spatial-Correlated MIMO Channels

2.2.1 System Setup

Consider a cellular MIMO system in which the base station (BS) and a mobile station (MS) are equipped with linear arrays of M and N antennas, respectively. Independent data streams $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t), \cdots, x_M(t)]^T$ are transmitted from the BS at time t, where $x_m(t)$ denotes the source signal of the mth transmit antenna and the superscript T denotes vector (matrix) transposition. At the MS, the received baseband signals are given by $\mathbf{y}(t) = [y_1(t), y_2(t), y_3(t), \cdots, y_N(t)]^T$, where $y_n(t)$ is the signal received by the nth receive antenna at time t. With a sampling interval of $\triangle t$ seconds, the corresponding ith transmit and receive sample vectors are $\mathbf{x}_i = \mathbf{x}(i\triangle t)$, and $\mathbf{y}_i = \mathbf{y}(i\triangle t)$, respectively.

2.2.2 Wireless MIMO Channels

A general MIMO channel between BS and MS antennas is modelled as

$$\mathbf{H}(t) = \sum_{l=1}^{G} \mathbf{H}_l \delta(t - \tau_l), \tag{2.1}$$

where G is the maximum number of paths associated with any sub-channel between a transmit and receive antenna pair, τ_l is the delay of the lth path, and δ denotes the Dirac delta function. The complex channel gain matrix associated with the lth path is given by $\mathbf{H}_l = [h_{ij}^l]$, for $1 \leq i \leq N, 1 \leq j \leq M$, where h_{ij}^l is the complex sub-channel gain between the jth transmit and ith receive antennas. For a narrowband fading channel, (2.1) is reduced to a single-tape fading matrix and the received vector waveform is $\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{n}(t)$, where $\mathbf{H}(t)$ is an $N \times M$ complex channel matrix and $\mathbf{n}(t)$ a zero mean additive white Gaussian noise (AWGN) vector with covariance matrix $E\{\mathbf{n}\mathbf{n}^H\} = N_0\mathbf{I}_N$. We first consider the block fading case in which the channel gain matrix remains unchanged within a block of B symbol intervals and eliminate the time parameter t in related expressions. Section IV will discuss the case which takes the time-correlation among blocks into consideration.

2.2.3 Spatial-correlated block fading channels

Many analytic models for spatial-correlated MIMO channels have been proposed in the literatures. The Kronecker model [6] assumes separable statistics at transmitter and receiver so that the spatial correlation matrix $\mathbf{\Phi}$ of vec($\mathbf{H}$), vec($\cdot$) being the stacking operator, is given by the Kronecker product ($\otimes$) [18] of those of the transmit ($\mathbf{\Phi}_{\mathrm{T}}$) and receive ($\mathbf{\Phi}_{\mathrm{R}}$) antennas, $\mathbf{\Phi} = \mathbf{\Phi}_{\mathrm{R}} \otimes \mathbf{\Phi}_{\mathrm{T}} = \mathbf{\Phi}^{\frac{1}{2}}(\mathbf{\Phi}^{\frac{1}{2}})^H$, where the "square root" matrix $\mathbf{\Phi}^{\frac{1}{2}}$ has a similar decomposition $\mathbf{\Phi}^{\frac{1}{2}} = \mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \otimes \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}}$. The separable statistics assumption yields

$$\mathbf{H} = \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}} \mathbf{H}_w \mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}\ T}, \tag{2.2}$$

where $\mathbf{H}_w$ is an $N \times M$ channel matrix whose entries are i.i.d. complex zero-mean, unit-variance Gaussian random variables.

Although the Kronecker model is mathematical tractably, many measurement and theoretical results reveal that this separable model in general leads to misfits for capacity and error probability due to the smaller number of degrees of freedom (DF) [19],[20]. The Kronecker model has been generalized by Sayeed [21] and, more recently, by Weichselberger *et al.* [22] who considered joint correlation of both link ends and suggested the following analytic model

$$\mathbf{H} = \mathbf{U}_{\mathrm{R}} \left(\tilde{\mathbf{\Omega}} \odot \mathbf{R} \right) \mathbf{U}_{\mathrm{T}}^T, \tag{2.3}$$

where $\mathbf{U}_{\mathrm{T}}$ and $\mathbf{U}_{\mathrm{R}}$ are the eigenbases of the one-sided correlation matrices at the transmit and receive sites, respectively. Operator $\odot$ denotes the Hadamard product operation [18]. $\mathbf{R}$ denotes a random matrix whose elements are i.i.d. zero-mean, unit-variance complex Gaussian random variables. $\tilde{\mathbf{\Omega}}$ is the element-wise square root of the coupling matrix in which each entry specifies the mean amount of energy coupled with an eigenvector of the transmitter to that of the receiver. The Weichselberger model provides a more general framework of canonical modelling [22],[23],[24], where (2.3) can be represented by the following canonical form,

$$\mathbf{H} = \mathbf{U}_{\mathrm{R}} \mathbf{H}_{\mathrm{ind}} \mathbf{U}_{\mathrm{T}}^T. \tag{2.4}$$

$\mathbf{H}_{\mathrm{ind}}$ has independent, but not necssarily identically distributed entries. The Kronecker model can thus be understood as a special case of the Weichselberger and the canonical

model by the following equations

$$\begin{aligned}
\mathbf{H} &= \boldsymbol{\Phi}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_w\boldsymbol{\Phi}_{\mathrm{T}}^{\frac{1}{2}\,T} \\
&\overset{\text{canonical (2.4)}}{=} \mathbf{U}_{\mathrm{R}}\underbrace{\mathbf{D}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_w\mathbf{D}_{\mathrm{T}}^{\frac{1}{2}}}_{:\,\mathbf{H}_{\mathrm{ind}}}\mathbf{U}_{\mathrm{T}}^{T} \qquad (2.5) \\
&\overset{\text{Weichselberger (2.3)}}{=} \mathbf{U}_{\mathrm{R}}\underbrace{((\mathrm{diag}(\mathbf{D}_{\mathrm{R}}^{\frac{1}{2}})\mathrm{diag}(\mathbf{D}_{\mathrm{T}}^{\frac{1}{2}})^{T})}_{:\,\tilde{\boldsymbol{\Omega}}}\odot\mathbf{H}_w)\mathbf{U}_{\mathrm{T}}^{T}, \qquad (2.6)
\end{aligned}$$

where $\boldsymbol{\Phi}_{\mathrm{T}} = \mathbf{U}_{\mathrm{T}}\mathbf{D}_{\mathrm{T}}\mathbf{U}_{\mathrm{T}}^{H}$ and $\boldsymbol{\Phi}_{\mathrm{R}} = \mathbf{U}_{\mathrm{R}}\mathbf{D}_{\mathrm{R}}\mathbf{U}_{\mathrm{R}}^{H}$ denote the eigen decomposition of correlation matrices at transmitter and receiver, respectively. (2.5) follows from the isotropicity of an i.i.d. random matrix under an unitary transformation. Note, the DF of $\tilde{\boldsymbol{\Omega}}$ in (2.6) is $N + M$, while $\tilde{\boldsymbol{\Omega}}$ of the general Weichselberger model in (2.3) has DF equal to NM. The small number of DF explains the deficiency of the Kronecker model as described above and is mainly due to the lack of modelling the cross-correlation between transmitter and receiver sides. In the following, we will develop a channel representation which takes the Kronecker, Weichselberger and canonical model as special cases, and is useful for reduced-rank processing.

2.3 Channel Representation

An $N \times M$ matrix $\mathbf{H}$ always admit the singular value decomposition (SVD), $\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{T}$, where $\mathbf{U}$ is an $N \times N$ unitary matrix, $\mathbf{V}$ is an $M \times M$ unitary matrix, and the diagonal matrix $\boldsymbol{\Lambda}$ is $N \times M$ with non-negative entries. When $\mathbf{H}$ is random, its SVD component matrices are random and depend on the sample (matrix) value of $\mathbf{H}$. As $\mathbf{U}$ and $\mathbf{V}$ can be transformed into two predefined unitary matrices $\mathbf{Q}_{\mathrm{R}}$ and $\mathbf{Q}_{\mathrm{T}}$ by $\mathbf{U}\mathbf{P}_1 = \mathbf{Q}_{\mathrm{R}}$ and $\mathbf{V}\mathbf{P}_2 = \mathbf{Q}_{\mathrm{T}}$, with both transforms $\mathbf{P}_1$ and $\mathbf{P}_2$ being unitary, we have

$$\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{P}_1^{-1}\boldsymbol{\Lambda}(\mathbf{P}_2^{-1})^{T}\mathbf{Q}_{\mathrm{T}}^{T} = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^{T} \qquad (2.7)$$

and the only random component is $\mathbf{C}$. For the Weichselberger model, the predefined matrices are eigenbases of the one-sided correlation matrices while Sayeed's virtual channel representation uses the DFT bases.

Let $\boldsymbol{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \overset{def}{=} [\phi_{\mathrm{T}}(i,j)]$, where $\phi_{\mathrm{T}}(i,j)$ represents the root spatial correlation between ith and jth transmit antennas. As the M column vectors of $\boldsymbol{\Phi}_{\mathrm{T}}^{\frac{1}{2}}$ lie in a $K_T(\leq M)$ dimensional

subspace, we have

$$\mathbf{\Phi}_{\mathrm{T}}^{1/2} = \mathbf{Q}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}, \tag{2.8}$$

where $\mathbf{Q}_{\mathrm{T}}$ is an unitary matrix and the coefficient matrix $\mathbf{\Lambda}_{\mathrm{T}}$ can be obtained by the Gram-Schimdt orthonormalization procedure. The above equation implies $\phi_{\mathrm{T}}(i,j) = \sum_{k=1}^{K_T} \lambda_k^j \mathbf{q}_k(i)$, where $\mathbf{q}_k(i)$ is the ith element of the kth basis vector, λ_k^j is the projection of the jth column on $\mathbf{q}_k$.

Using a similar decomposition for $\mathbf{\Phi}_{\mathrm{R}}^{1/2}$ leads to

$$\mathbf{\Phi}^{1/2} = (\mathbf{Q}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}) \otimes (\mathbf{Q}_{\mathrm{R}}\mathbf{\Lambda}_{\mathrm{R}}) = (\mathbf{Q}_{\mathrm{T}} \otimes \mathbf{Q}_{\mathrm{R}})\,(\mathbf{\Lambda}_{\mathrm{T}} \otimes \mathbf{\Lambda}_{\mathrm{R}})\,,$$

where we have invoked the identity [18],

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2)\cdots(\mathbf{A}_k \otimes \mathbf{B}_k) = (\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k) \otimes (\mathbf{B}_1\mathbf{B}_2\cdots\mathbf{B}_k). \tag{2.9}$$

From the canonical representation, $\mathrm{vec}(\mathbf{H}) = \mathbf{\Phi}^{\frac{1}{2}}\mathrm{vec}(\mathbf{H}_w)$, we obtain

$$\mathrm{vec}(\mathbf{H}) = (\mathbf{Q}_{\mathrm{T}} \otimes \mathbf{Q}_{\mathrm{R}})\,(\mathbf{\Lambda}_{\mathrm{T}} \otimes \mathbf{\Lambda}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{H}_w) \stackrel{def}{=} (\mathbf{Q}_{\mathrm{T}} \otimes \mathbf{Q}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{C}). \tag{2.10}$$

The identity

$$\mathrm{vec}\,(\mathbf{ABD}) = (\mathbf{D}^T \otimes \mathbf{A})\,\mathrm{vec}\,(\mathbf{B}) \tag{2.11}$$

implies $\mathrm{vec}(\mathbf{H}) = \mathrm{vec}\left(\mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T\right)$, and so $\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T$, which is the same as (2.7).

We summarize the above derivation on the relation between the proposed analytic model with the Kronecker, Sayeed, and Weichselberger models in

Proposition 2.1. *An $N \times M$ MIMO channel matrix $\mathbf{H}$, can always be expressed as*

$$\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T \tag{2.12}$$

where $\mathbf{C}$ is a complex random coefficient matrix, $\mathbf{Q}_{\mathrm{R}}$ and $\mathbf{Q}_{\mathrm{T}}$ are predefined unitary matrices. The above model is equivalent to the Kronecker model if the matrix $\mathbf{C}$ satisfies the separable correlation condition

$$\mathrm{vec}(\mathbf{C}) = (\mathbf{\Lambda}_{\mathrm{T}} \otimes \mathbf{\Lambda}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{H}_w) \tag{2.13}$$

where $\mathbf{\Lambda}_{\mathrm{T}}$ *and* $\mathbf{\Lambda}_{\mathrm{R}}$ *are coefficient matrices that depend on the spatial correlations among the transmit and the receive antenna arrays, respectively. (2.12) is related to the Weichselberger model via*

$$\mathbf{U}_{\mathrm{T}} = \mathbf{Q}_{\mathrm{T}}\mathbf{P}_{\mathrm{T}}^{H}, \quad \mathbf{U}_{\mathrm{R}} = \mathbf{Q}_{\mathrm{R}}\mathbf{P}_{\mathrm{R}}^{H} \tag{2.14}$$

$$\mathbf{P}_{\mathrm{R}}^{H}\Gamma_{\mathrm{R}}\mathbf{P}_{\mathrm{R}} = E\left\{\mathbf{C}\mathbf{C}^{H}\right\}, \quad \mathbf{P}_{\mathrm{T}}^{H}\Gamma_{\mathrm{T}}\mathbf{P}_{\mathrm{T}} = E\left\{\mathbf{C}^{T}\mathbf{C}^{*}\right\} \tag{2.15}$$

where $\mathbf{P}_{\mathrm{T}}$, $\mathbf{P}_{\mathrm{R}}$ *are unitary matrices and* $\Gamma_{\mathrm{R}}, \Gamma_{\mathrm{T}}$ *have the same eigenvalues of the matrices* $E\left\{\mathbf{H}\mathbf{H}^{H}\right\}$ *and* $E\left\{\mathbf{H}^{T}\mathbf{H}^{*}\right\}$, *respectively. When the predefined matrices are the same as* $\mathbf{U}_{\mathrm{R}}$ *and* $\mathbf{U}_{\mathrm{T}}$, $\mathbf{C}$ *has the special form* $\tilde{\mathbf{\Omega}} \odot \mathbf{R}$. *Moreover, (2.12) is equivalent to the virtual representation of Sayeed if columns of* $\mathbf{Q}_{\mathrm{R}}$ *and* $\mathbf{Q}_{\mathrm{T}}$ *are DFT basis vectors and entries of* $\mathbf{C}$ *are independent complex Gaussian random variables.*

[1] suggested and [25] verified through field measurements that the mean direction of arrival (DoA) can be embedded in the channel model by pre-multiplying the channel matrix $\mathbf{H}$ by a diagonal matrix which is a function of the DoA. We can derive a similar model by invoking the fact that if $\mathbf{W}$ is a diagonal matrix with unit modulus entries and $\mathbf{V}$ is unitary then both $\mathbf{V}\mathbf{W}$ and $\mathbf{W}^{-1}\mathbf{V}$ are also unitary, to obtain the alternative representation (2.16).

Corollary 2.1. *An equivalent channel matrix for stationary frequency-flat MIMO channel is given by*

$$\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\overline{\mathbf{Q}}_{\mathrm{T}}^{T}\mathbf{W} \tag{2.16}$$

where $\overline{\mathbf{Q}}_{\mathrm{T}}^{T}\mathbf{W} = \mathbf{Q}_{\mathrm{T}}^{T}$ *and* $\mathbf{W}$ *is a diagonal matrix with unit modulus entries.*

Several remarks and observations on the channel models (2.12) and (2.16) are given below.

R1. The Kronecker model requires that $\mathbf{C}$ has the special structure (2.13) while the Weichselberger, Sayeed and canonical models demand that the entries of $\mathbf{C}$ be independent (but not identical) random variables. In contrast, the proposed model does not impose any constraint on the coefficient matrix $\mathbf{C}$ and is valid for arbitrary block-faded $\mathbf{H}$.

R2. The Weichselberger model is perhaps more convenient to generate the matrix channel $\mathbf{H}$ and for evaluating the channel capacity of correlated MIMO channels as the coefficient matrix has independent entries. It is also useful to analyze MIMO system performance. However, it is not suitable for channel estimation applications because the number of parameters, including the unknown eigenbases, is even larger than that of $\mathbf{H}$.

R3. For practical correlated MIMO channels, which are of particular concern, the entries of $\mathbf{H}$ are not i.i.d. but correlated random variables and $\mathbf{H}$ admits reduced-rank representation. That is, although $\mathbf{H}$ is likely to be of full rank, one can approximate it by reduced-rank unitary matrices (so is the coefficient matrix), ignoring the weaker eigenmodes. The rank-reduction is most obvious for typical urban macro-cellular environments in which an MS is surrounded by local scatterers, and waveforms impending the receive antennas are richly scattered, while the BS is not obstructed by the local scatterers [6][26]. Appendix A shows that, if the angle spread (AS) Δ is not too large, the diagonal matrix $\mathbf{W}$

$$\mathbf{W} = \text{diag}\left[w_1, w_2, \cdots, w_M\right], \tag{2.17}$$

has entries of the form $w_i = \exp\left[-j2\pi\frac{(i-1)d}{\lambda}\sin\phi\right]$, d being the inter-element distance, that bear the mean AoD information. As will become clear later, the separability of channel correlation and angle information characterizations has some useful implications.

R4. Given predefined bases $\mathbf{Q}_{\text{R}}$, $\mathbf{Q}_{\text{T}}$, or $\overline{\mathbf{Q}}_{\text{T}}$, the statistic properties of the corresponding coefficient matrix is completely determined by those of $\mathbf{H}$. Identification of the unknown channel $\mathbf{H}$ is equivalent to the estimation of $\mathbf{C}$ or the pair $(\mathbf{C}, \mathbf{W})$, which usually has a lower rank and much smaller number of entries than those of $\mathbf{H}$ for the link environment of interest. Thus, using model (2.12) or (2.16) reduces the number of parameters to be estimated and enhances the performance. Moreover, as the bases in both (2.12) and (2.16) are pre-defined, these two models can be easily extended to time-varying block fading and frequency-selective fading environments.

R5. There are several classes of basis functions to choose from. The Taylor and Weierstrass arguments and the results of [27] suggest the use of polynomial bases. If we use polynomials of degree P as basis functions in expanding a spatial correlation function of length P, the corresponding basis matrix $\mathbf{P}_P$ has entries

$$[\mathbf{P}]_{i,j} = (i-1)^{j-1}, \ i,j = 1,2,\ldots,P, \tag{2.18}$$

Although the column vectors in (2.18) form a basis, they are not orthogonal. Furthermore, these vectors have different norms, which might result in numerical instability. By applying the QR decomposition to the corresponding $\mathbf{P}_P$ [28], we obtain an orthonormalized polynomial basis matrix $\mathbf{P}_o$. The basis matrices $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ of (3.2) or $\mathbf{Q}_{L,K_L}$ of (3.23) are obtained by selecting the first K_T, K_R or K_L columns of the corresponding $\mathbf{P}_o$.

R6. For a fixed base one needs to determine the modelling orders, K_T and K_R. Either the Akaike information criterion (AIC) and the minimum description length (MDL) approach can be used to determine the optimal modelling orders that trade-off the system complexity and performance [29]. Time domain modelling order K_L discussed in Section IV can also be similarly determined. Depending on the application scenario, these order parameter values can be obtained by an one-shot open loop estimate or should be periodically updated.

R7. The model (2.16) is especially useful for channel estimation application because, as will be shown in the next section, it allows very efficient (in terms of convergence rate) channel estimation algorithms that iteratively estimate $\mathbf{C}$ and $\mathbf{W}$ separately, and, at low SNR's, the reduced-rank model gives performance superior to that of the full-rank model. Furthermore, for a small-to-medium AS, which occurs quite often in cellular downlinks, the extracted AoD information can be feedback for downlink beamforming.

R8. Our simulation experiments indicate that, when the AS $\triangle$ becomes large, the rank of $\mathbf{C}$ increases and there is no dominant spatial angle. The steering matrix $\mathbf{W}$

becomes an identity matrix which gives no AoD information and (2.16) degenerates to (2.12).

R9. The proposed channel representation for single-block frequency-flat MIMO channels, i.e., (2.16), can also be extended to the cases of time-variant frequency-flat and time-variant frequency-selective fading channels by properly modelling the time-domain correlation. These extensions are given in Section 3.2 and Section 3.3 respectively.

Chapter 3

Model-based MIMO Channel Estimation

This chapter presents novel schemes for estimating correlated multiple-input multiple-output (MIMO) fading channels. Our schemes are based on an analytic correlated block fading model and its time-variant extension which encompass the popular Kronecker model and the more general Weichselberger model as special cases. Both static and time-variant models offer compact representations of spatial- and/or time-correlated channels. When the transmit antenna array is such that the associated MIMO channel has a small angle spread (AS), which occurs quite often in a cellular downlink, our models admit reduced-rank channel representations. They also enable us to develop effective estimators and provide compact channel state information (CSI) descriptions which are needed in feedback systems and for many post channel estimation applications. The latter has the important implication of reduced feedback channel bandwidth requirement and lower post-processing complexity.

We propose iterative algorithms for estimating static and time-variant MIMO channels. The proposed models make it natural to decompose each iteration into two successive stages that are responsible for estimating the correlation coefficients and the signal direction, respectively. Both spatial- and time-correlated fadings are considered. The mean-squared error (MSE) performance of our estimators are analyzed as well. Using popular industry-approved standard channel models, we verify through simulations that our algorithms yield offer good MSE performance which, in many practical cases, is better than that achievable by a conventional least-square estimator.

3.1 Single-Block Based Channel Estimation

In this section we consider estimation schemes which are based on a single block of observation without taking into account the (time-)correlation among blocks. We propose two iterative schemes in which an iteration consists of two phases. The first phase is responsible for the estimation of the coefficient matrix, $\mathbf{C}$, while the directional matrix, $\mathbf{W}$ in (2.16), is estimated in the second phase. Both tentative estimates are updated as one proceeds with each new iteration until the stopping criterion is met. The two schemes differs in the second phase only.

Consider the $M \times B$ matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_B]$ formed by B length-M input symbol vectors, where $B \geq M$. Assuming $\mathbf{H}$ remains static during a B-block period, we express the received sample block, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_B]$ as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \tag{3.1}$$

where $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_B]$ is the corresponding noise matrix whose entries are i.i.d. zero mean complex Gaussian random variables. In estimating $\mathbf{H}$, $\mathbf{X}$ is assumed to be composed of either the pilot vectors or some decision feedback results. Substituting two known unitary matrices $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ with ranks $K_T(\leq M)$ and $K_R(\leq N)$ for $\overline{\mathbf{Q}}_T$ and $\mathbf{Q}_R$ in (2.16), we want to find the optimal solution $\{\mathbf{W}_{opt}, \mathbf{C}_{opt}\}$ to the problem

$$\arg\min_{\mathbf{W},\mathbf{C}} \|\mathbf{Y} - \mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}_{M,K_T}^T\mathbf{W}\mathbf{X}\|^2 \tag{3.2}$$

We express the corresponding optimal (least-squares) channel estimate in terms of $\mathbf{W}_{opt}$ and $\mathbf{C}_{opt}$

$$\mathbf{H}_{opt} = \mathbf{Q}_{N,K_R}\mathbf{C}_{opt}\mathbf{Q}_{M,K_T}^T\mathbf{W}_{opt} \tag{3.3}$$

so that (3.1) can be rewritten as

$$\mathbf{Y} = \mathbf{H}_{opt}\mathbf{X} + \Delta\mathbf{H}\mathbf{X} + \mathbf{N} \stackrel{def}{=} \mathbf{H}_{opt}\mathbf{X} + \widetilde{\mathbf{N}}, \tag{3.4}$$

where $\widetilde{\mathbf{N}}$ represents the sum of the modelling error $\Delta\mathbf{H}\mathbf{X}$ due to the reduced rank representation and the AWGN vector, $\mathbf{N}$.

To derive an iterative algorithm for obtaining the joint directional and channel solution $\{\mathbf{W}_{opt}, \mathbf{H}_{opt}\}$, we assume that, at the $(i-1)$th iteration,

$$\mathbf{Y} = \widehat{\mathbf{H}}_{i-1}\mathbf{X} + \Delta\widehat{\mathbf{H}}_{i-1}\mathbf{X} + \widetilde{\mathbf{N}} \tag{3.5}$$

where $\Delta\widehat{\mathbf{H}}_{i-1} \stackrel{def}{=} \mathbf{H}_{opt} - \widehat{\mathbf{H}}_{i-1}$ is the residual error at the end of the $(i-1)$th iteration, and consider the estimation of the channel (coefficients) and AoD in two separate phases.

3.1.1 Phase I - Coefficient Estimation

Assume that the directional matrix in this phase is optimum, i.e., $\mathbf{W} = \mathbf{W}_{opt}$. From (3.1) and (3.3), we have

$$\mathrm{vec}(\mathbf{Y}) = \left\{\left((\mathbf{W}_{opt}\mathbf{X})^T\mathbf{Q}_{M,K_T}\right) \otimes \mathbf{Q}_{N,K_R}\right\}\mathrm{vec}(\mathbf{C}) + \mathrm{vec}(\mathbf{N}). \tag{3.6}$$

Substituting the definition $\mathbf{Z} \stackrel{def}{=} \left((\mathbf{W}_{opt}\mathbf{X})^T\mathbf{Q}_{M,K_T}\right) \otimes \mathbf{Q}_{N,K_R}$ into (3.6), we have the LS solution

$$\mathrm{vec}(\widehat{\mathbf{C}}) = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathrm{vec}(\mathbf{Y}) \stackrel{def}{=} F(\mathbf{W}_{opt}). \tag{3.7}$$

While the optimal directional matrix estimate is not available, we replace it by the tentative estimation from the previous iteration, $\mathbf{W}_{i-1}$. $\mathrm{vec}(\widehat{\mathbf{C}})$ is then obtained by computing $F(\mathbf{W}_{i-1})$ instead, and the corresponding tentative estimate is denoted by $\widehat{\mathbf{C}}_i$. Initially, we can arbitrarily set $\mathbf{W}_0$ to be an identity matrix.

3.1.2 Phase II - Direction Estimation

Similar to Phase I, we begin with the assumption that the coefficient matrix in this estimation phase is optimum. The directional information is to be obtained by estimating a diagonal matrix $\mathbf{W}$ with unit modulus entries; see (2.17). Setting

$$\mathbf{G} \stackrel{def}{=} \mathbf{Q}_{N,K_R}\mathbf{C}_{opt}\mathbf{Q}_{M,K_T}^T \tag{3.8}$$

and invoking (3.3), we have $\widehat{\mathbf{H}}_{i-1} = \mathbf{G}\widehat{\mathbf{W}}_{i-1}$. As $\mathbf{C}_{opt}$ is unavailable, $\mathbf{C}_{opt}$ is replaced by the previous estimate $\widehat{\mathbf{C}}_{i-1}$ in computing $\mathbf{G}$ during the ith iteration. In the following, we propose two algorithms to estimate the phase of the unit modulus diagonal entries of $\mathbf{W}$.

3.1.2.1 Algorithm A - Maximum Matching Output

To estimate $\widehat{\mathbf{W}}_i$ in diagonal form, we start with the following lemma whose proof is given in Appendix B.

Lemma 3.1. *For two matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *of size* $N \times M$ *and* $M \times E$ *respectively, and an arbitrary vector* $\mathbf{c}$ *of size* $M \times 1$*, the following identity holds.*

$$vec\left(\mathbf{A} \cdot diag(\mathbf{c}) \cdot \mathbf{B}\right) = \left[(\mathbf{1}_E \otimes \mathbf{A}) \odot \left(\mathbf{B}^T \otimes \mathbf{1}_N\right)\right] \mathbf{c}, \tag{3.9}$$

where "diag" denotes the diagonal operation used to translate a vector into a diagonal matrix, with its diagonal terms being the elements of the original vector.

Combined with matrix $\mathbf{G}$ defined in (3.8), (3.4) is rewritten as

$$\mathbf{Y} = \mathbf{G}\mathbf{W}_{opt}\mathbf{X} + \widetilde{\mathbf{N}}. \tag{3.10}$$

Let $\mathbf{w}_{opt}$ be the column vector that consists of the diagonal elements of $\mathbf{W}_{opt}$, i.e., $\mathbf{w}_{opt}(i) = \mathbf{W}_{opt}(i,i)$, for any $1 \le i \le M$. Then, by *Lemma* 3.1, we have

$$\text{vec}\left(\mathbf{Y}\right) = \left[(\mathbf{1}_B \otimes \mathbf{G}) \odot \left(\mathbf{X}^T \otimes \mathbf{1}_N\right)\right] \mathbf{w}_{opt} + \text{vec}\left(\widetilde{\mathbf{N}}\right) \tag{3.11}$$

and the LS estimate of $\mathbf{w}_{opt}$ is given by $\widehat{\mathbf{w}}_{LS} = \mathbf{T}^{\dagger} \cdot \text{vec}(\mathbf{Y})$, where $\left[(\mathbf{1}_B \otimes \mathbf{G}) \odot \left(\mathbf{X}^T \otimes \mathbf{1}_N\right)\right] \overset{def}{=} \mathbf{T}$.

In order to extract the steering vector $\widehat{\mathbf{w}}$, we introduce $\mathbf{v}(\theta) \overset{def}{=} \left[1, v(\theta), \ldots, v^{M-1}(\theta)\right]^T$, where $v(\theta) = \exp\left[-j2\pi\frac{d}{\lambda}\sin(\theta)\right]$. The AoD information $\widehat{\phi}$ is retrieved by maximizing the matching output

$$\widehat{\phi} = \arg \max_{-\pi \le \theta \le \pi} \text{Re}\left\{\mathbb{P}\left(\widehat{\mathbf{w}}_{LS}\right)^H \mathbf{v}(\theta)\right\}, \tag{3.12}$$

where $\mathbb{P}(\cdot)$ is defined by the following phase extraction operator,

$$\mathbb{P}\left([a_0 e^{jb_0}, a_1 e^{jb_1}, \cdots, a_K e^{jb_K}]\right) \overset{def}{=} [1, e^{j(b_1 - b_0)}, \cdots, e^{j(b_K - b_0)}],$$
$$\text{for } \{a_i\}_{i=0}^K \in \mathcal{R}^{K+1} \text{ and } \{b_i\}_{i=0}^K \in [0,\ 2\pi). \tag{3.13}$$

Once $\widehat{\phi}$ is available, it is straightforward to obtain $\widehat{\mathbf{W}} = \text{diag}(\mathbf{v}(\widehat{\phi}))$. Solving (3.12) over $[0, 2\pi)$ can be accomplished by using the conventional line searching algorithm.

Computing $\widehat{\mathbf{w}}_{LS}$ in (3.12) involves a pseudo-inverse operation of matrix $\mathbf{T}$, and is thus computational expansive. Thanks to the special structure of $\mathbf{T}$, a training matrix with orthogonal rows can be used to bypass the calculation of pseudo-inversion. Note that

$$\mathbf{T}^H\mathbf{T} \;=\; (B\mathbf{G}^H\mathbf{G}) \odot (N\mathbf{X}^*\mathbf{X}^T), \tag{3.14}$$

the right-hand side of (3.14) will become a diagonal matrix with nonnegative real elements if $\mathbf{X}^*\mathbf{X}^T = B\mathbf{I}$. Such an orthogonality is guaranteed provided that the optimal training matrix for LS estimator is used [7]. Under the assumption of orthogonal training matrix, we have

$$\mathbb{P}\left(\widehat{\mathbf{w}}_{LS}\right) = \mathbb{P}\left(\mathbf{T}^{\dagger}\cdot \mathrm{vec}(\mathbf{Y})\right) = \mathbb{P}\left(\mathbf{T}^{H}\cdot \mathrm{vec}(\mathbf{Y})\right) \stackrel{def}{=} \mathbb{P}\left(\widetilde{\widehat{\mathbf{w}}}_{LS}\right), \tag{3.15}$$

where $\widetilde{\widehat{\mathbf{w}}}_{LS}$ doesn't require the cumbersome matrix inversion. The AoD information can thus be obtained simply by substituting $\widetilde{\widehat{\mathbf{w}}}_{LS}$ for $\widehat{\mathbf{w}}_{LS}$ in (3.12).

3.1.2.2 Algorithm B - Root Finding Method

An alternative way to find the optimal phase is to convert (3.12) into a root finding problem. Note that the elements of $\mathbf{w}_{opt}$ are of geometric progression, i.e., they form a row vector of a Vandermonde matrix. Hence if we define the correlation polynomial

$$P(z) \stackrel{def}{=} \mathbb{P}(\widehat{\mathbf{w}}_{LS})^H\mathbf{z} - M, \tag{3.16}$$

where $\mathbf{z} = [1, z, \ldots, z^{M-1}]$ and let $\mathcal{Z}$ be the set of its zeros in the complex plane, then solving (3.12) is equivalent to

$$\widehat{z} = \arg\min_{z\in\mathcal{Z}} |(|z| - 1)| \quad \text{and} \quad \widehat{\phi} = \sin^{-1}\left(\frac{-\mathrm{Arg}\{\widehat{z}\}\lambda}{2\pi d}\right) \tag{3.17}$$

and the directional matrix is reconstructed by $\widehat{\mathbf{W}} = \mathrm{diag}(\widehat{\mathbf{z}})$, where $\widehat{\mathbf{z}} = [1, \widehat{z}, \cdots, \widehat{z}^{M-1}]$. Unlike *Algorithm A* whose solution accuracy relies on the resolution the numerical search algorithm used, this algorithm gives the exact analytic solution once (3.16) is solved. Similar to *Algorithm A*, $\widetilde{\widehat{\mathbf{w}}}_{LS}$ can be substituted for $\widehat{\mathbf{w}}_{LS}$ in (3.16) to simplify the computation.

Since the object function in (3.2) is jointly convex with respect to $\mathbf{C}$ and $\mathbf{W}$ and the proposed algorithms have the form of a nonlinear Gauss-Seidel algorithm, the convergences of our algorithms are guaranteed [30]. All the simulation examples reported

in Section 3.2 converge and achieve the theoretical performance lower bound derived in Section 3.4.

The computation complexity of the proposed algorithm is dominated by the LS operations in *Phase I*. The flop count of the LS operation in *Phase I* is $O(BK_T^2)$, $K_T \leq M$ while the conventional LS estimator needs $O(BM^2)$ flops [31]. The complexity of *Phase II* is mainly contributed by the product of $\mathbf{T}$ and vec($\mathbf{Y}$), and is in general much less than that of *Phase I*, thanks to Eq. (3.15) and the special structure of $\mathbf{T}$. Therefore, the total complexity of the proposed algorithm is less than that of the conventional LS. Moreover, except for static channels, the estimates for both $\mathbf{W}$ and $\mathbf{C}$ need to be updated periodically. Let each $B-$symbol interval be called an estimation interval (EI). Since the mean AoD usually change much slower than the channel coefficients (gains) variation, updating frequencies for $\mathbf{W}$ and $\mathbf{C}$ can and should be different, i.e., if the two estimates are updated every T_o^c and T_o^w EIs, respectively, then $T_o^w \gg T_o^c$ (see Fig. 3.8 of Section 3.5). This dual updating frequency option is unique to our approach and implies that *Phase II* may be disabled most of the time while *Phase I* needs single iteration per update EI, hence our algorithm can be computational more efficient than the conventional LS approach for many non-static channels.

The major advantage of our channel model and estimator lies not only in the computational efficiency of the channel estimator but also in the compactness of CSI representation which is needed in a feedback system and that of post processing operations. As mentioned in R3 and R4 in the previous chapter, a small K_T is often sufficient to accurately describe a MIMO channel with high transmit spatial correlation. For any post channel estimation operation associated with $\mathbf{H}$, e.g., taking pseudo-inverse or eigen-decomposition of $\mathbf{H}$, the computing load is reduced as it involves the $K_R \times K_T$ coefficient matrix and the estimated AoD instead of the original $N \times M$ channel matrix.

3.1.3 Order Determination for Block-Fading Channels

The remark **R4** in Chapter 2 tells us that we can choose to use low-rank bases to closely approximate $\mathbf{H}$ provided that the modelling orders K_T and K_R are properly selected. We use Akaike information criterion (AIC) to determine the optimal modelling orders as it

tends to give robust and reliable results, especially for small sample size. Taking into account the proposed channel estimator, the AIC-based order determination scheme is given by [29],

$$[K_T, K_R] = \arg \min_{1 \le K_T \le M, 1 \le K_R \le N} N \log \left(\frac{\mathrm{RSS}_{K_T,K_R}}{N} \right) + 2(K_T + K_R). \tag{3.18}$$

where RSS_{K_T,K_R} is the squared error of (3.2) associated with the modelling orders K_T and K_R. Instead of using the instant sample error, we can use the time-average squared error in calculating the AIC solution to obtain a more reliable estimate.

Since the channel statistics varies much slower than the instantaneous channel strength, the update period of the modelling order is much longer than that of the instant channel estimate; reducing the overhead required by order estimation. Similar order determination scheme can be used to estimate the modelling order for time-correlated fading channels, provided that RSS_{K_T,K_R} of (3.2) is calculated using the time-correlated model (3.24) given in the next section and the frequency-selective case (3.38) discussed in Section 3.3. Moreover, the optimal time domain modelling order can also be determined by incorporating K_L discussed in the next section into the degrees of freedom in AIC's formula.

3.2 Channel Estimation with Time Correlation Consideration

We now extend our investigation to the case that considers the time correlation among blocks. Similar to our spatial modelling approach, we use a set of orthonormal basis functions to describe a snap shot of a fading channel's time domain behavior. We assume an equally spaced pilot-block arrangement. The issue of the optimal pilot arrangement that minimizes the MSE or bit error rate (BER) was addressed in [12] and [32].

Assume the two leading pilot symbol vectors of two consecutive pilot block is T symbol intervals away. The receive signal block at time nT can be written as

$$\mathbf{Y}_n = \mathbf{H}_n \mathbf{X}_n + \mathbf{N}_n \tag{3.19}$$

where $\mathbf{Y}_n = \mathbf{Y}(nT)$ and $\mathbf{X}_n = \mathbf{X}(nT)$ are the $N \times B$ receive matrix at time nT and the corresponding $M \times B$ transmit block, respectively. $\mathbf{H}_n$ is the $N \times M$ matrix whose (i, j)th

entry represents the link gain between the ith transmit and the jth receive antennas at time nT.

We consider the time-variant behavior of a MIMO channel within a fixed observation window of L blocks (EIs). The received sample blocks from nT to $(n+L-1)T$ can be cascaded into the matrix

$$\mathbf{Y}_{n,L} \overset{def}{=} [\mathbf{Y}_n, \mathbf{Y}_{n+1}, \ldots, \mathbf{Y}_{n+L-1}]. \tag{3.20}$$

Using (2.11), we obtain

$$\mathrm{vec}(\mathbf{Y}_{n,L}) = \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \cdot \mathrm{vec}\left(\mathbf{H}_{n,L}\right) + \mathrm{vec}\left(\mathbf{N}_{n,L}\right) \tag{3.21}$$

where $\mathrm{vec}(\mathbf{H}_{n,L}) \overset{def}{=} \left[\mathrm{vec}(\mathbf{H}_n)^T, \ldots \mathrm{vec}(\mathbf{H}_{n+L-1})^T\right]^T$, $\mathrm{vec}(\mathbf{N}_{n,L}) \overset{def}{=} \left[\mathrm{vec}(\mathbf{N}_n)^T, \ldots \mathrm{vec}(\mathbf{N}_{n+L-1})^T\right]^T$, and

$$\mathbf{X}_{n,L}^T \overset{def}{=} \begin{bmatrix} \mathbf{X}_n^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{n+1}^T & \vdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{n+L-1}^T \end{bmatrix}.$$

Substituting (2.12) for each $\mathbf{H}_n$ and assuming the eigenbases $\mathbf{Q}_\mathrm{T}$ and $\mathbf{Q}_\mathrm{R}$ remain invariant during an estimation period, we obtain

$$\mathrm{vec}(\mathbf{H}_{n,L}) = \left(\mathbf{I}_L \otimes \mathbf{Q}_\mathrm{T} \otimes \mathbf{Q}_\mathrm{R}\right) \mathbf{\Gamma}_{n,L}. \tag{3.22}$$

Each component of the vector $\mathbf{\Gamma}_{n,L} = [\gamma_n, \gamma_{n+1}, \cdots, \gamma_{n+L-1}]^T$ is itself an $(NM)\times 1$ column vector $\gamma_n = \left(\gamma_{1n}, \gamma_{2n}, \cdots, \gamma_{(NM)n}\right)^T$ that represents the complex fading coefficients for all NM MIMO subchannels at time nT and, γ_{pn}, $1 \le p \le NM$, are independent.

The stacked vector, $\gamma(p) = \left[\gamma_{pn}, \gamma_{p(n+1)}, \cdots, \gamma_{p(n+L-1)}\right]^T$, represents a finite-duration sample of the complex random process associated with the pth subchannel [25]. Such a process can also be expanded by a set of smooth functions [8, 33], and thus its estimation can be obtained by using a method similar to that developed in the previous section. Similar to the approach used in Section 2.2.3, we can first apply the orthogonal transform $\gamma(p) = \mathbf{Q}_L \mathbf{b}_{pn}$, where $\mathbf{Q}_L$ is an $L \times L$ orthogonal matrix, and $\mathbf{b}_{pn}$ is the transform domain coefficient vector. Then, the time domain channel correlation can be approximated by using the reduced basis matrix $\mathbf{Q}_{L,K_L}$

$$\gamma(p) \approx \mathbf{Q}_{L,K_L} \mathbf{c}_{pn}, \quad \text{and} \quad \mathbf{\Gamma}_{n,L} \approx \left(\mathbf{Q}_{L,K_L} \otimes \mathbf{I}_{LMN}\right) \cdot \mathbf{c}_{\mathrm{coef}}, \tag{3.23}$$

where K_L denotes the time domain modelling order, and $\mathbf{c}_{pn}$ is a $K_L \times 1$ coefficient vector.

By using (2.16), (3.22) and the approximation (3.23), we decouple the signal part of (3.21) into the product of two modelling domains - space and time domains

$$\begin{aligned} \mathrm{vec}(\bar{\mathbf{Y}}_{n,L}) &\approx \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right)\left[\mathbf{Q}_{L,K_L} \otimes (\mathbf{W}^T\overline{\mathbf{Q}}_{\mathrm{T}}) \otimes \mathbf{Q}_{\mathrm{R}}\right]\mathbf{c}_{\mathrm{coef}} \\ &\approx \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right)\left[\mathbf{Q}_{L,K_L} \otimes (\mathbf{W}^T\mathbf{Q}_{\mathrm{T},K_T}) \otimes \mathbf{Q}_{\mathrm{R},K_R}\right]\tilde{\mathbf{c}}_{\mathrm{coef}} \\ &\overset{def}{=} \left(((\mathbf{W}_L\mathbf{X}_{n,L})^T\widetilde{\mathbf{Q}}_{\mathrm{T},K_T}) \otimes \mathbf{Q}_{\mathrm{R},K_R}\right)\tilde{\mathbf{c}}_{\mathrm{coef}} \end{aligned} \tag{3.24}$$

where $\mathbf{W}_L \overset{def}{=} (\mathbf{I}_L \otimes \mathbf{W})$, $\widetilde{\mathbf{Q}}_{\mathrm{T},K_T} \overset{def}{=} \mathbf{Q}_{L,K_L} \otimes \mathbf{Q}_{\mathrm{T},K_T}$ and $\mathbf{Q}_{\mathrm{T},K_T}$ and $\mathbf{Q}_{\mathrm{R},K_R}$ are composed of K_T and K_R column vectors of $\overline{\mathbf{Q}}_{\mathrm{T}}$ and $\mathbf{Q}_{\mathrm{R}}$, respectively. $\mathbf{W}$ is the steering matrix defined in (2.17). Since the mean AoD usually varies slowly with respect to a sub-channel's coherent time, we assume that $\mathbf{W}$ remains the same during a period of L data blocks. Similar to the narrowband model (2.16), we do not impose the implicit Kronecker structure and Gaussian assumption on $\tilde{\mathbf{c}}_{\mathrm{coef}}$.

As (3.24) can be obtained by replacing $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, $\mathrm{vec}(\mathbf{C})$, $\mathbf{Q}_{M,K_T}$, and $\mathbf{Q}_{N,K_R}$ in (3.6) by $\mathbf{X}_{n,L}$, $\mathbf{Y}_{n,L}$, $\mathbf{W}_L$, $\tilde{\mathbf{c}}_{\mathrm{coef}}$, $\widetilde{\mathbf{Q}}_{\mathrm{T},K_T}$, and $\mathbf{Q}_{\mathrm{R},K_R}$, we conclude that both spatial and time correlations can be described by similar models. Hence, the two-phase iterative estimation scheme developed in Section 3.1 can be extended to estimate the coefficient vector $\tilde{\mathbf{c}}_{\mathrm{coef}}$, and the directional matrix $\mathbf{W}_L$ in (3.24). In the following, we describe two-phase channel estimation schemes with time correlation consideration.

3.2.1 Phase I - Coefficient Estimation

Following an argument similar to that used in Section 3.1, we assume that the directional matrix $\mathbf{W}_L$ is optimal in the coefficient estimation phase and define

$$\widetilde{\mathbf{Z}} \overset{def}{=} \left((\mathbf{W}_{L,opt}\mathbf{X}_{\mathbf{n,L}})^T\widetilde{\mathbf{Q}}_{\mathrm{T},K_T}\right) \otimes \mathbf{Q}_{\mathrm{R},K_R}. \tag{3.25}$$

The LS estimate of $\tilde{\mathbf{c}}_{\mathrm{coef}}$ is

$$\widehat{\tilde{\mathbf{c}}}_{\mathrm{coef}} = (\widetilde{\mathbf{Z}}^H\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^H\mathrm{vec}\left(\mathbf{Y}_{n,L}\right) \overset{def}{=} \widetilde{F}(\mathbf{W}_{L,opt}), \tag{3.26}$$

which is a function of the optimal directional matrix $\mathbf{W}_{L,opt}$. At the ith iteration, since the optimal directional matrix is not available, the tentative estimation, $\mathbf{W}_{L,i-1}$, is substituted for $\mathbf{W}_{L,opt}$.

3.2.2 Phase II - Direction Estimation

Similar to the single-block based case, we propose two AoD estimation algorithms. Again, we assume the optimal coefficient vector is available, i.e., $\tilde{\mathbf{c}}_{\text{coef}} = \tilde{\mathbf{c}}_{\text{coef},opt}$, when estimating the directional information.

Define a new matrix $\widetilde{\mathbf{G}} \overset{def}{=} \mathbf{Q}_{\text{R},K_R}\widetilde{\mathbf{C}}_{\text{coef},opt}\widetilde{\mathbf{Q}}^T_{\text{T},K_T}$, where $\widetilde{\mathbf{C}}_{\text{coef},opt}$ is a $K_R \times K_L K_T$ matrix derived from $\tilde{\mathbf{c}}_{\text{coef},opt}$ by $\widetilde{\mathbf{C}}_{\text{coef},opt}(i,j) = \tilde{\mathbf{c}}_{\text{coef},opt}\,(K_R(j-1)+i)$, $1 \le i \le K_R, 1 \le j \le K_L K_T$. We rewrite the received matrix in vector form

$$\begin{aligned} \text{vec}(\mathbf{Y}_{n,L}) &= \text{vec}\left(\widetilde{\mathbf{G}}\mathbf{W}_L\mathbf{X}_{n,L}\right) + \widetilde{\mathbf{N}}_{n,L} \\ &= \left(\mathbf{X}^T_{n,L} \otimes \widetilde{\mathbf{G}}\right)\text{vec}(\mathbf{I}_L \otimes \mathbf{W}) + \widetilde{\mathbf{N}}_{n,L}, \end{aligned} \tag{3.27}$$

where $\widetilde{\mathbf{N}}_{n,L}$ represents the sum of the modelling error associated with $\widetilde{\mathbf{G}}$ and the AWGN term $\mathbf{N}_{n,L}$.

3.2.2.1 Algorithm A - Maximum Matching Output

If $\mathbf{W}$ is constrained to be a diagonal matrix, i.e., $\mathbf{W} = \text{diag}(\mathbf{w})$, then $\mathbf{I}_L \otimes \mathbf{W} = \text{diag}(\mathbf{1}_L \otimes \mathbf{w})$ and therefore

$$\text{vec}(\mathbf{Y}_{n,L}) = \text{vec}\left(\widetilde{\mathbf{G}} \cdot \text{diag}(\mathbf{1}_L \otimes \mathbf{w}) \cdot \mathbf{X}_{n,L}\right) + \widetilde{\mathbf{N}}_{n,L}. \tag{3.28}$$

From *Lemma* 3.1, we have

$$\begin{aligned} &\text{vec}\left(\widetilde{\mathbf{G}} \cdot \text{diag}(\mathbf{1}_L \otimes \mathbf{w})) \cdot \mathbf{X}_{n,L}\right) \\ &\qquad = \left(\left(\mathbf{1}_{BL} \otimes \widetilde{\mathbf{G}}\right) \odot \left(\mathbf{X}^T_{n,L} \otimes \mathbf{1}_N\right)\right)(\mathbf{1}_L \otimes \mathbf{I}_M)\mathbf{w} \overset{def}{=} \widetilde{\mathbf{T}}\mathbf{w}. \end{aligned} \tag{3.29}$$

Similar to *Algorithm A* presented in the previous subsection, the LS estimate of $\mathbf{w}_{opt}$ is $\widehat{\mathbf{w}}_{LS} = \widetilde{\mathbf{T}}^{\dagger} \cdot \text{vec}(\mathbf{Y}_{n,L})$. To improve the estimate and reconstruct a steering vector $\widehat{\mathbf{w}}$, we analogously define a steering vector $\mathbf{v}(\theta) \overset{def}{=} \left[1, v(\theta), \cdots, v^{M-1}(\theta)\right]^T$, where $v(\theta) = \exp(-j2\pi\frac{d}{\lambda}\sin(\theta))$. The AoD information $\widehat{\phi}$ can be retrieved by

$$\widehat{\phi} = \arg \max_{-\pi \le \theta \le \pi} \text{Re}\left\{\mathbb{P}(\widehat{\mathbf{w}}_{LS})^H \mathbf{v}(\theta)\right\}, \tag{3.30}$$

where $\mathbb{P}$ denotes the phase extraction operator defined by (3.13). Having obtained $\widehat{\phi}$, we then proceed to compute $\widehat{\mathbf{W}}_L = \mathbf{I}_L \otimes \mathbf{V}(\widehat{\phi})$, where $\mathbf{V}(\widehat{\phi}) = \text{diag}(\mathbf{v}(\widehat{\phi}))$.

Also, the pseudo-inverse operation $\widetilde{\mathbf{T}}^{\dagger}$ is not necessary if orthogonal training matrix is used for $\mathbf{X}_n$, i.e., $\mathbf{X}_n\mathbf{X}_n^H = B\mathbf{I}$ for each n. Following the discussion given in Section 3.1.2, for orthogonal training matrices, we have

$$\mathbb{P}(\ \widetilde{\mathbf{T}}^{\dagger} \cdot \mathrm{vec}\,(\mathbf{Y}_{n,L})\ ) = \mathbb{P}(\ \overbrace{\widetilde{\mathbf{T}}^{H} \cdot \mathrm{vec}\,(\mathbf{Y}_{n,L})}^{\stackrel{def}{=}\widetilde{\widehat{\mathbf{w}}}_{LS}}\ ), \tag{3.31}$$

and $\widehat{\mathbf{w}}_{LS}$ can be replaced by $\widetilde{\widehat{\mathbf{w}}}_{LS}$ in (3.30).

3.2.2.2 Algorithm B - Root-Finding Method

The root-finding approach for the block fading case can be used as well. It is easy to see that (3.30) is equivalent to searching for the root of the correlation polynomial $P(z)$ which is the closest to the unit circle, i.e.,

$$\widehat{z} = \arg\min_{z} \ ||z| - 1|, \ \text{ subject to } P(z) \stackrel{def}{=} \mathbb{P}(\widehat{\mathbf{w}}_{LS})^H\mathbf{z} - M = 0 \tag{3.32}$$

and then retrieving the AoD information from $\widehat{z} = \exp\left[-j2\pi\frac{d}{\lambda}\sin(\widehat{\phi})\right]$. The directional matrix is to be reconstructed by $\widehat{\mathbf{W}}_L = \mathbf{I}_L \otimes \mathrm{diag}(\widehat{\mathbf{z}})$, where $\widehat{\mathbf{z}} = [1, \widehat{z}, \ldots, \widehat{z}^{M-1}]$. Also, for orthogonal training matrices, $\widetilde{\widehat{\mathbf{w}}}_{LS}$ can be substituted for $\widehat{\mathbf{w}}_{LS}$ to skip the pseudo-inverse computation.

The total complexity per block of the proposed algorithm, like the single-block based case in Section 3.1, is smaller than that of the conventional LS estimator. Given a fixed iteration number, the flop count of the proposed algorithm is decided by *Phase I* and is of the order $O(BK_T^2L)$, while the conventional LS estimator needs $O(BM^2L)$ flops. Thus, we can save the computational complexity up to the ratio $\frac{K_T^2}{M^2}$. Moreover, if the operating scenario allows the use of the dual updating frequencies option and $T_o^w \ll T_o^c$, the total complexity can be reduced further. For slowing time-variant channels, the required time domain modelling order, K_L, is small, the number of channel representation parameters is reduced from LMN to $K_LK_TK_R+1$. Such a reduction yields compact CSI representation and benefits many post channel estimation operations involving $\mathbf{H}$, as was discussed at the end of last section.

3.3 Channel Estimation for Frequency-Selective Time-Varying Fading Channels

For estimating a correlated frequency-selective time-varying fading channel, approaches used in the previous sections are extended to accommodate frequency-selective characteristics. Assume that the power delay profiles for the sub-channels between transmit and receive pairs are independent but of the same form. With a sampling interval of T, the receive signals at time nT can be written as

$$\mathbf{Y}_n = \sum_{d=0}^{D-1} \mathbf{H}_n^{(d)} \mathbf{X}_n^{(d)} + \mathbf{N}_n \tag{3.33}$$

where $\mathbf{Y}_n = \mathbf{Y}(nT)$ is the $N \times B$ receive vector at time nT, $\mathbf{X}_n^{(d)} = \mathbf{X}(nT - dT)$ is the $M \times B$ transmit block, at time $(n-l)T$, and B denotes the length of training block. $\mathbf{H}_n^{(d)}$ is the $N \times M$ matrix whose (i,j)th entry represents the fading coefficient of the dth delay path, at time nT for the channel between the ith transmit and jth receive antennas.

In the estimation of the time-variant fading MIMO channel, an observation window is included to take into account the channel variation. For an observation window of size L, the stacked receive sample vector from time nT to $(n+L-1)T$ can be expressed as

$$\mathbf{Y}_{n,L} \stackrel{def}{=} [\mathbf{Y}_n, \mathbf{Y}_{n+1}, \cdots, \mathbf{Y}_{n+L-1}]. \tag{3.34}$$

Applying (2.11) to the stacked version of (3.33), we have

$$\mathrm{vec}(\mathbf{Y}_{n,L}) = \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \cdot \mathrm{vec}\left(\mathbf{H}_{n,L}\right) + \mathrm{vec}\left(\mathbf{N}_{n,L}\right) \tag{3.35}$$

where

$$\begin{aligned}
\mathrm{vec}(\mathbf{H}_{n,L}) &\stackrel{def}{=} \left[\mathrm{vec}(\mathbf{H}_n)^T, \ldots \mathrm{vec}(\mathbf{H}_{n+L-1})^T\right]^T, \\
\mathrm{vec}(\mathbf{H}_n) &\stackrel{def}{=} \left[\mathrm{vec}(\mathbf{H}_n^{(0)})^T, \cdots, \mathrm{vec}(\mathbf{H}_n^{(D-1)})^T\right]^T, \\
\mathrm{vec}(\mathbf{N}_{n,L}) &\stackrel{def}{=} \left[\mathrm{vec}(\mathbf{N}_n)^T, \ldots \mathrm{vec}(\mathbf{N}_{n+L-1})^T\right]^T,
\end{aligned}$$

and

$$\mathbf{X}_{n,L}^T \stackrel{def}{=} \begin{bmatrix}
\mathbf{X}_n^{(0)^T}, \cdots, \mathbf{X}_n^{(D-1)^T} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{X}_{n+1}^{(0)^T}, \cdots, \mathbf{X}_{n+1}^{(D-1)^T} & \vdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{n+L-1}^{(0)^T}, \cdots, \mathbf{X}_{n+L-1}^{(D-1)^T}
\end{bmatrix}.$$

Substituting (2.12) for each $\mathbf{H}_n$ and assuming the eigenbases $\mathbf{Q}_{\mathrm{T}}$ and $\mathbf{Q}_{\mathrm{R}}$ remain invariant during an estimation period, we obtain

$$\mathrm{vec}(\mathbf{H}_{n,L}) = (\mathbf{I}_{LD} \otimes \mathbf{Q}_{\mathrm{T}} \otimes \mathbf{Q}_{\mathrm{R}})\, \mathbf{\Gamma}_{n,L}. \tag{3.36}$$

Each component of the vector $\mathbf{\Gamma}_{n,L} = \left[\gamma_n^{(0)^T}, \cdots, \gamma_n^{(D-1)^T}, \cdots, \gamma_{n+L-1}^{(0)^T}, \cdots, \gamma_{n+L-1}^{(D-1)^T}\right]^T$ is itself an $(NM) \times 1$ column vector $\gamma_n^{(d)} = \left(\gamma_{1n}^{(d)}, \gamma_{2n}^{(d)}, \cdots, \gamma_{(NM)n}^{(d)}\right)^T$ that represents the complex fading coefficients of the dth path for the pth MIMO subchannel at time nT, $p \in \{1, 2, \cdots, NM\}$, and $d \in \{0, 1, \cdots, D-1\}$.

The stacked vector, $\gamma(p) = \left[\gamma_{pn}^{(d)}, \gamma_{p(n+1)}^{(d)}, \cdots, \gamma_{p(n+L-1)}^{(d)}\right]^T$, represents a finite-duration sample of the complex random process associated with the dth delay path which has a fixed Doppler spectrum [25]. Such a process can also be expanded by a set of smooth functions [33],[8], and thus its estimation can be obtained by using a method similar to that developed in the previous section. Similarly, as described in Section 2.2.3, we can first apply the orthogonal transform $\gamma(p) = \mathbf{Q}_L \mathbf{b}_{pn}^{(d)}$, where $\mathbf{Q}_L$ is a full-rank orthogonal matrix, and $\mathbf{b}_{pn}^{(d)}$ is the transform domain coefficient. Then, the time domain channel correlation is approximated in the following equation by using the reduced bases matrix $\mathbf{Q}_{L,K_L}$,

$$\gamma(p) \approx \mathbf{Q}_{L,K_L} \mathbf{c}_{pn}^{(l)}, \quad \text{and} \quad \mathbf{\Gamma}_{n,L} \approx (\mathbf{Q}_{L,K_L} \otimes \mathbf{I}_{DMN}) \cdot \mathbf{c}_{\mathrm{coef}}, \tag{3.37}$$

where K_L denotes the time domain modelling order, and $\mathbf{c}_{pn}^{(d)}$ is a $K_L \times 1$ coefficient vector.

By using (3.35), (3.36) and the approximation (3.37), we decouple the signal part of (3.35) into the product of two modelling domains - space and time domains

$$\begin{aligned}
\mathrm{vec}(\mathbf{Y}_{n,L}) &\approx \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) (\mathbf{I}_{LD} \otimes \mathbf{U}_{\mathrm{LS}} \otimes \mathbf{U}_{\mathrm{R}}) \left(\mathbf{Q}_{L,K_L} \otimes \mathbf{I}_{DMN}\right) \mathbf{c}_{\mathrm{coef}} \\
&= \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \left[(\mathbf{Q}_{L,K_L} \otimes \mathbf{I}_D) \otimes \left((\mathbf{W}^T \mathbf{Q}_{\mathrm{T},K_T}) \otimes \mathbf{Q}_{\mathrm{R},K_R}\right)\right] \tilde{\mathbf{c}}_{\mathrm{coef}} \\
&\overset{def}{=} \left(((\mathbf{W}_{DL} \mathbf{X}_{n,L})^T \widetilde{\mathbf{Q}}_{\mathrm{T},K_T}) \otimes \mathbf{Q}_{\mathrm{R},K_R}\right) \tilde{\mathbf{c}}_{\mathrm{coef}}
\end{aligned} \tag{3.38}$$

where $\mathbf{W}_{DL} \overset{def}{=} (\mathbf{I}_{LD} \otimes \mathbf{W})$, $\widetilde{\mathbf{Q}}_{\mathrm{T},K_T} \overset{def}{=} \mathbf{Q}_{L,K_L} \otimes \mathbf{I}_D \otimes \mathbf{Q}_{\mathrm{T},K_T}$, $\tilde{\mathbf{c}}_{\mathrm{coef}} = (\mathbf{I}_{K_L D} \otimes \mathbf{P}_{\mathrm{T}} \otimes \mathbf{P}_{\mathrm{R}})\, \mathbf{c}_{\mathrm{coef}}$. $\mathbf{W}$ is the steering matrix defined in (2.17). Since the mean AoD usually varies slowly with respect to a sub-channel's coherent time, we assume that $\mathbf{W}$ remains the same during a period of L data blocks. Similar to the narrowband model (21), we do not impose the implicit Kronecker structure and Gaussian assumption on $\tilde{\mathbf{c}}_{\mathrm{coef}}$.

As (3.38) can be obtained by replacing $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, vec($\mathbf{C}$), $\mathbf{Q}_{M,K_T}$, and $\mathbf{Q}_{N,K_R}$ in (3.6) by $\mathbf{X}_{n,L}$, $\mathbf{Y}_{n,L}$, $\mathbf{W}_{DL}$, $\tilde{\mathbf{c}}_{\text{coef}}$, $\widetilde{\mathbf{Q}}_{\text{T},K_T}$, and $\mathbf{Q}_{\text{R},K_R}$, we conclude that both block fading and time variant frequency selective fading channels can be described by similar models. Hence, the two-phase iterative estimation scheme developed in Section 3.1 can be extended to estimate the coefficient vector $\tilde{\mathbf{c}}_{\text{coef}}$, and the directional matrix $\mathbf{W}_{DL}$ in (3.38). In the following, we describe a two-phase channel estimation approach for frequency-selective time-variant MIMO fading channels.

3.3.1 Phase I - Coefficient Estimation

Following an argument similar to that used in Section 3.1, we assume that the directional matrix $\mathbf{W}_{DL}$ is optimal in the coefficient estimation phase and define

$$\widetilde{\mathbf{Z}} \stackrel{def}{=} \left((\mathbf{W}_{LD,opt}\mathbf{X}_{n,L})^T \widetilde{\mathbf{Q}}_{\text{T},K_T} \right) \otimes \mathbf{Q}_{\text{R},K_R}. \tag{3.39}$$

The LS estimate of $\tilde{\mathbf{c}}_{\text{coef}}$ is

$$\widehat{\tilde{\mathbf{c}}}_{\text{coef}} = (\widetilde{\mathbf{Z}}^H\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^H \text{vec}\left(\mathbf{Y}_{n,B}\right) \stackrel{def}{=} \widetilde{F}(\mathbf{W}_{LD,opt}), \tag{3.40}$$

which is a function of the optimal directional matrix $\mathbf{W}_{LD,opt}$. At the ith iteration, since the optimal directional matrix is not available, we substitute the tentative estimation at $(i-1)$th iteration, $\mathbf{W}_{LD,i-1}$, for $\mathbf{W}_{LD,opt}$.

3.3.2 Phase II - Direction Estimation

The two AoD estimation algorithms established in Section 3.2 can be directly extended here for wide-band MIMO channels. Again, we assume the optimal coefficient vector is available, i.e., $\tilde{\mathbf{c}}_{\text{coef}} = \tilde{\mathbf{c}}_{\text{coef},opt}$, when estimating the directional information.

Define a new matrix $\widetilde{\mathbf{G}} \stackrel{def}{=} \mathbf{Q}_{\text{R},K_R}\widetilde{\mathbf{C}}_{\text{coef},opt}\widetilde{\mathbf{Q}}^T_{\text{T},K_T}$, where $\widetilde{\mathbf{C}}_{\text{coef},opt}$ is a $K_R \times DK_LK_T$ matrix derived from $\tilde{\mathbf{c}}_{\text{coef},opt}$ by $\widetilde{\mathbf{C}}_{\text{coef},opt}(i,j) = \tilde{\mathbf{c}}_{\text{coef},opt}\left(K_R(j-1)+i\right)$, $1 \le i \le K_R$ and $1 \le j \le DK_LK_T$. We rewrite the received matrix in vector form

$$\text{vec}(\mathbf{Y}_{n,L}) = \text{vec}\left(\widetilde{\mathbf{G}}\mathbf{W}_{DL}\mathbf{X}_{n,L}\right) + \mathbf{N}'_{n,L} = \left(\mathbf{X}^T_{n,L} \otimes \widetilde{\mathbf{G}}\right)\text{vec}(\mathbf{I}_{DL} \otimes \mathbf{W}) + \mathbf{N}'_{n,L}, \tag{3.41}$$

where $\mathbf{N}'_{n,L}$ represents the sum of the modelling error associated with $\widetilde{\mathbf{G}}$ and the AWGN term $\mathbf{N}_{n,L}$.

If $\mathbf{W}$ is constrained to be a diagonal matrix, i.e., $\mathbf{W} = \text{diag}(\mathbf{w})$, then $\mathbf{I}_{DL} \otimes \mathbf{W} = \text{diag}(\mathbf{1}_{DL} \otimes \mathbf{w})$ and therefore

$$\text{vec}(\mathbf{Y}_{n,L}) = \text{vec}\left(\widetilde{\mathbf{G}} \cdot \text{diag}(\mathbf{1}_{DL} \otimes \mathbf{w}) \cdot \mathbf{X}_{n,L}\right) + \mathbf{N}'_{n,L}. \tag{3.42}$$

From Lemma 3.1, we have

$$\begin{aligned} \text{vec}\left(\widetilde{\mathbf{G}} \cdot \text{diag}(\mathbf{1}_{DL} \otimes \mathbf{w})) \cdot \mathbf{X}_{n,L}\right) &= \left(\left(\mathbf{1}_{BL} \otimes \widetilde{\mathbf{G}}\right) \odot \left(\mathbf{X}_{n,L}^T \otimes \mathbf{1}_N\right)\right)(\mathbf{1}_{LD} \otimes \mathbf{I}_M)\mathbf{w} \\ &\stackrel{def}{=} \widetilde{\mathbf{T}}_D \mathbf{w}. \end{aligned} \tag{3.43}$$

Here, we can extend the proposed two direction estimation algorithms developed in Section 3.2.2 to extract the AoD information for frequency-selective channel, simply by replacing the $\widetilde{\mathbf{T}}$ in (3.29) with $\widetilde{\mathbf{T}}_D$ in the above equation.

3.4 Performance Analysis

In analyzing the MSE performance

$$\epsilon \stackrel{def}{=} E\left\{\|\mathbf{H} - \widehat{\mathbf{H}}\|_F^2\right\} = E\left\{\|\text{vec}(\mathbf{H}) - \text{vec}(\widehat{\mathbf{H}})\|_2^2\right\}. \tag{3.44}$$

of the proposed $\widehat{\mathbf{H}}$, we first make the optimistic assumptions that the optimal orthogonal pilot matrix [7] for conventional LS channel estimator is used and the directional matrix estimate $\widehat{\mathbf{W}}$ is perfect.

Notations

For notational simplicity and when there is no danger of ambiguity, $\mathbf{H}$ and $\mathbf{W}$ in this section denote the channel and directional matrices of (3.1)/(3.2) or (3.21)/ (3.24) for single-block based or time-correlated based estimators, and $\mathbf{X}_p$ represents $\mathbf{X}$ in (3.6) or $\mathbf{X}_{n,L}$ in (3.24). Furthermore, $\mathbf{Q}_{\text{T}}$ and $\mathbf{Q}_{\text{R}}$ denote either the modelling bases $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ in (3.6), or $\widetilde{\mathbf{Q}}_{\text{T},K_T}$ and $\mathbf{Q}_{\text{R},K_R}$ in (3.24).

Then (3.44) can be expressed as

$$\begin{aligned} \epsilon(\mathbf{X}_p; \mathbf{W}) &= E\left\{\|\text{vec}(\mathbf{H}) - \text{vec}(\mathbf{Q}_{\text{R}}\widehat{\mathbf{C}}\mathbf{Q}_{\text{T}}^T\mathbf{W})\|_2^2\right\} \\ &= E\left\{\|\text{vec}(\mathbf{H}) - \mathbf{\Psi}\mathbf{\Omega}_z\text{vec}(\mathbf{H}\mathbf{X}_p + \mathbf{N})\|_2^2\right\} \end{aligned} \tag{3.45}$$

where $\mathbf{\Psi} \stackrel{def}{=} (\mathbf{W}^T\mathbf{Q}_{\text{T}}) \otimes \mathbf{Q}_{\text{R}}$ and $\mathbf{\Omega}_z \stackrel{def}{=} (\widehat{\mathbf{Z}}^H\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}^H$, $\widehat{\mathbf{Z}}$ being the LS estimate of $\mathbf{Z}$ defined in 3.1.1, i.e.,

$$\widehat{\mathbf{Z}} \stackrel{def}{=} ((\mathbf{W}\mathbf{X}_p)^T\mathbf{Q}_{\text{T}}) \otimes \mathbf{Q}_{\text{R}} \tag{3.46}$$

As $\mathbf{HX}_p$ and $\mathbf{N}$ are statistically independent, the MSE can be separated into two terms which are contributed by modelling error (reduced-rank basis matrices) and AWGN, respectively.

$$\begin{aligned}\epsilon(\mathbf{X}_p;\mathbf{W}) &= E\left\{\|\mathrm{vec}(\mathbf{H})-\mathbf{\Psi\Omega}_z\mathrm{vec}(\mathbf{HX}_p)\|_2^2\right\}+E\left\{\|\mathbf{\Psi\Omega}_z\mathrm{vec}(\mathbf{N})\|_2^2\right\}\\ &\stackrel{def}{=} \epsilon_h(\mathbf{X}_p,\mathbf{W})+\epsilon_n(\mathbf{X}_p,\mathbf{W}).\end{aligned} \tag{3.47}$$

Define the following projections

$$\begin{aligned}\mathbf{P}_{\mathbf{W}} &\stackrel{def}{=} \left[\mathbf{W}^T\mathbf{Q}_{\mathrm{T}}(\mathbf{Q}_{\mathrm{T}}^T\mathbf{W}^*\mathbf{X}_p^*\mathbf{X}_p^T\mathbf{W}^T\mathbf{Q}_{\mathrm{T}})^{-1}\mathbf{Q}_{\mathrm{T}}^T\mathbf{W}^*\mathbf{X}_p^*\mathbf{X}_p^T\right]\otimes\mathbf{Q}_{\mathrm{R}}\mathbf{Q}_{\mathrm{R}}^T\\ \tilde{\mathbf{P}}_{\mathbf{W}} &\stackrel{def}{=} \left[\mathbf{W}^T\mathbf{Q}_{\mathrm{T}}(\mathbf{Q}_{\mathrm{T}}^T\mathbf{Q}_{\mathrm{T}})^{-1}\mathbf{Q}_{\mathrm{T}}^T\mathbf{W}^*\right]\otimes\mathbf{Q}_{\mathrm{R}}\mathbf{Q}_{\mathrm{R}}^T.\end{aligned}$$

The first term on the RHS of (3.47) becomes

$$\begin{aligned}\epsilon_h(\mathbf{X}_p;\mathbf{W}) &= E\|(\mathbf{I}-\mathbf{P}_{\mathbf{W}})\,\mathrm{vec}(\mathbf{H})\|_2^2\\ &= \mathrm{tr}\left((\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}}^H)(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{R}_h\right)\\ &= \sum_{k=1}^{\chi}\lambda_k\|(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|_2^2\end{aligned} \tag{3.48}$$

where $\mathbf{R}_h = E\left\{\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\right\}$ is the channel correlation matrix and $\mathbf{f}_k$ is $\mathbf{R}_h$'s eigenvector associated with the eigenvalue λ_k, $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_\chi$; χ being the degree of freedom of $\mathbf{H}$. For the single-block based case, $\chi = NM$ and it is equal to NML when the estimator considers the time correlation effect. (3.48) is valid since the orthogonal training matrix $\mathbf{X}_p$ is used. Let $1<K\leq\chi$ be the rank of the dominant signal subspace of the channel covariance matrix. Then $\mathbf{R}_h=\sum_{k=1}^{\chi}\lambda_k\mathbf{f}_k\mathbf{f}_k^H\simeq\sum_{k=1}^{K}\lambda_k\mathbf{f}_k\mathbf{f}_k^H$, with $\lambda_k\ll 1$ for $K<k\leq\chi$. Since $\|(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|_2^2\leq 1$, we have $\sum_{k=K+1}^{\chi}\lambda_k\|(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|_2^2\leq\sum_{k=K+1}^{K_s}\lambda_k\ll 1$. Let the compound modelling order K_s be equal to K_TK_R and $K_TK_RK_L$ for the two cases under investigation. If K_s is chosen to be larger than K, the rank of $\mathbf{R}_h$, i.e., $K<K_s\leq\chi$, and the basis matrices $\mathbf{Q}_{\mathrm{T}}$ and $\mathbf{Q}_{\mathrm{R}}$ span the dominant signal subspace of $\mathbf{R}_h$, then the matrix $\tilde{\mathbf{P}}_{\mathbf{W}}$ is a projection operator whose range lies mostly in the space spanned by $\{\mathbf{f}_k\}, 1\leq k\leq K$ and we conclude that $\|(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|_2^2\stackrel{def}{=}\|\tilde{\mathbf{P}}_{\mathbf{W}}^{\perp}\mathbf{f}_k\|_2^2\ll 1,\ \text{for } 1\leq k\leq K$. Therefore, the modelling error ϵ_h is negligible in this case. On the other hand, if the modelling order is not enough to span the signal subspace, there is under-modelling error contributed by those non-negligible terms $\lambda_k\|(\mathbf{I}-\tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|_2^2$ which will dominate the mean squared error when the AWGN is small (high SNR region).

As for the MSE due to thermal noise–the second term on the RHS of (3.47), we can show that

$$\epsilon_n(\mathbf{X}_p, \mathbf{W}) = E\left\{\|\mathbf{\Psi}\mathbf{P}_z \mathrm{vec}(\mathbf{N})\|_2^2\right\} = \mathrm{tr}\left(\frac{N_0}{B}\tilde{\mathbf{P}}_{\mathbf{W}}\right) = \frac{N_0}{B}K_s, \tag{3.49}$$

where we have invoked the facts that (i) the training signal $\mathbf{X}_p$ and the noise $\mathbf{N}$ are independent, (ii) unitary pilot matrix is used, $\mathbf{X}_p^*\mathbf{X}_p^T = B\mathbf{I}$ and (iii) elements of $\mathbf{N}$ is i.i.d. complex white Gaussian noise with variance $\sigma_n^2 = N_0$. (3.49) implies that thermal noise induced MSE can be reduced by using a small modelling order. In Section 3.5 (Figs. 3.3 -3.5), we find that this noise-reduction effect is significant in low SNR environments where thermal noise dominates the MSE performance while the modelling error of (3.48) dominates in high SNR region.

If $\widehat{\mathbf{W}}$ is not perfect and $\mathbf{W} = \widehat{\mathbf{W}} + \Delta\mathbf{W}$, then

$$\widehat{\mathbf{Z}} \stackrel{def}{=} \mathbf{Z} + \Delta\mathbf{Z} = ((\mathbf{W}\mathbf{X}_p)^T\mathbf{Q}_{\mathrm{T}}) \otimes \mathbf{Q}_{\mathrm{R}} + ((\Delta\mathbf{W}\mathbf{X}_p)^T\mathbf{Q}_{\mathrm{T}}) \otimes \mathbf{Q}_{\mathrm{R}}. \tag{3.50}$$

The coefficient vector estimation $\mathrm{vec}(\widehat{\mathbf{C}})$ can be approximated up to the first order of $\Delta\mathbf{Z}$ as [34]

$$\mathrm{vec}(\widehat{\mathbf{C}}) \simeq \mathrm{vec}(\mathbf{C}) - \mathbf{Z}^\dagger\Delta\mathbf{Z}\mathrm{vec}(\mathbf{C}) + \mathbf{Z}^\dagger\mathrm{vec}(\mathbf{N}) + (\mathbf{Z}^H\mathbf{Z})^{-1}\Delta\mathbf{Z}^H P_{\mathbf{Z}}^{\perp}\mathrm{vec}(\mathbf{N}) - \mathbf{Z}^\dagger\Delta\mathbf{Z}\mathbf{Z}^\dagger\mathrm{vec}(\mathbf{N}), \tag{3.51}$$

where $P_{\mathbf{Z}}^{\perp} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}$. The above equation indicates that, besides the terms that have to do with the noise $\mathbf{N}$, the coefficient vector estimation error is determined by the projection error $\Delta\mathbf{Z}$. Hence, when the projection error $\Delta\mathbf{W}$ is small (and thus $\Delta\mathbf{Z}$ is small), $\mathrm{vec}(\widehat{\mathbf{C}})$ is a good approximation of $\mathrm{vec}(\mathbf{C})$ at high SNR region.

3.5 Numerical Results and Discussion

Simulation results reported here use the reference MIMO channel model [2], the IEEE 802.11 TGn channel model [3], and the SCM model [35]. The former two are stochastic models whose spatial correlation matrices are generated by the power azimuth spectrum (PAS) at the BS and MS, respectively. The SCM model generates the channel coefficients according to a set of selected parameters (e.g., AS, AoD, AoA, etc.). It is a popular parametric stochastic model whose spatial cross correlations are functions of the joint

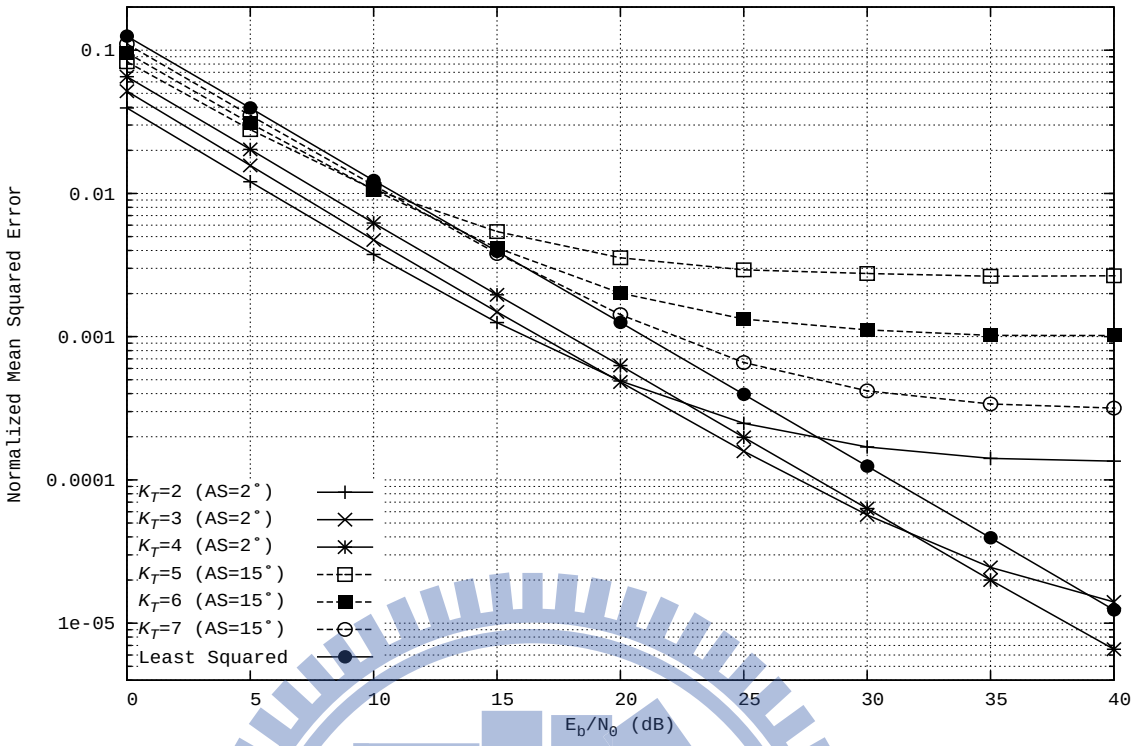


Figure 3.1: MSE performance of *Algorithm B* as a function of SNR with different modelling orders; solid curves: AS=2°, dotted curves: AS=15°.

distribution of the AoD at the transmit side and the AoA at the receive side. We assume that the environment surrounding MS is rich scattering with negligible spatial correlations. Hence, a full rank basis matrix is used to characterize the spatial correlation at the receive side. Other assumptions and conditions used in our simulation are: (i) the antenna spacings at transmit and receive arrays are both 0.5λ, (ii) an orthogonal pilot matrix is used, (iii) 10 iterations are used for all simulations (although in most cases convergence occurs in less than 3 iterations), and (iv) SNR (E_b/N_0) is defined as the average signal to noise power ratio at the input of each receive antenna, (v) orthonormal polynomial basis matrices are used. Both algorithms compute $\widehat{\mathbf{H}}$ by substituting the final result of *Phase I*–estimated coefficient matrix $\widehat{\mathbf{C}}$–and that of *Phase II*–$\widehat{\mathbf{W}}$–into (3.3).

Solid curves in Fig. 3.1 are the MSE performance of *Algorithm B* in Section 3.1 for an 8×8 MIMO system with $\Delta = 2^\circ$ and are based on the channel model of [2]. The channel is a block fading channel with an approximated rank of two. Since the BS spatial correlations are high, the corresponding correlation function lies in a low-dimension subspace so that a small K_T is sufficient to describe the channel. Dotted curves in Fig. 3.1 show the system

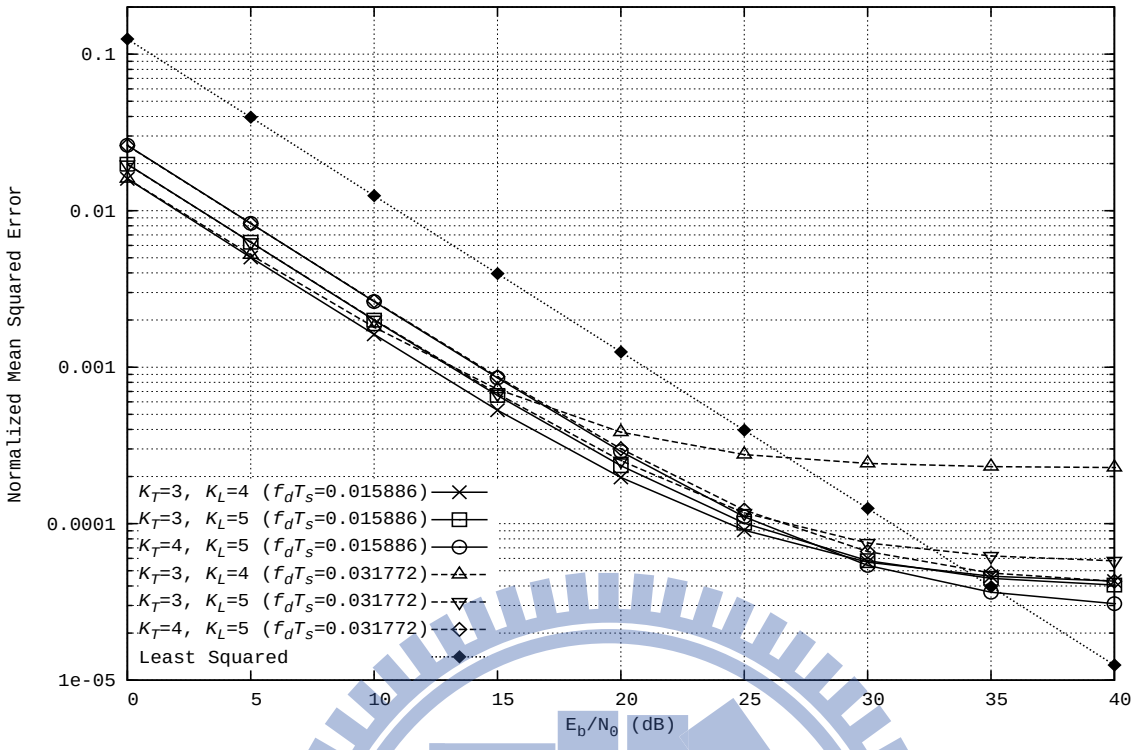


Figure 3.2: The effect of the modelling order on *Algorithm B*'s MSE performance in a channel generated by the model described in [1] with AS=2°.

performance when $\Delta = 15°$. It is obvious that as Δ increases, the spatial correlations among the transmit antennas elements decrease and a higher modelling order is necessary to describe rapid-changing spatial waveforms at the transmitter side. However, as can be seen from Figs. 2-5, an optimal K_T exists for any given SNR and Δ; increasing the modelling order does not necessary improves the channel estimator's performance. As expected, we find that modelling errors dominate the MSE performance when SNR is high. Such a behavior is consistent with what the performance analysis given in Section 3.4 has predicted and is similar to those observed in other model-based approaches [9]-[13].

The MSE performance of *Algorithm B* of Section 3.2 for a time-correlated fading channel [2] are depicted in Fig. 3.2 and Fig. 3.3 using an observation window of 12 EIs and $f_dT_s = 0.031772$ or 0.015886. Similar to the single-blocked based case (Fig. 3.1), the processing dimension (K_T) can be drastically reduced provided that either the spatial or time domain correlation is high enough. Performance degradation occurs when the modelling order is not large enough to capture the channel characteristics. In Fig. 3.4, we compare the theoretical MSE derived in Section 3.4 with the simulated performance

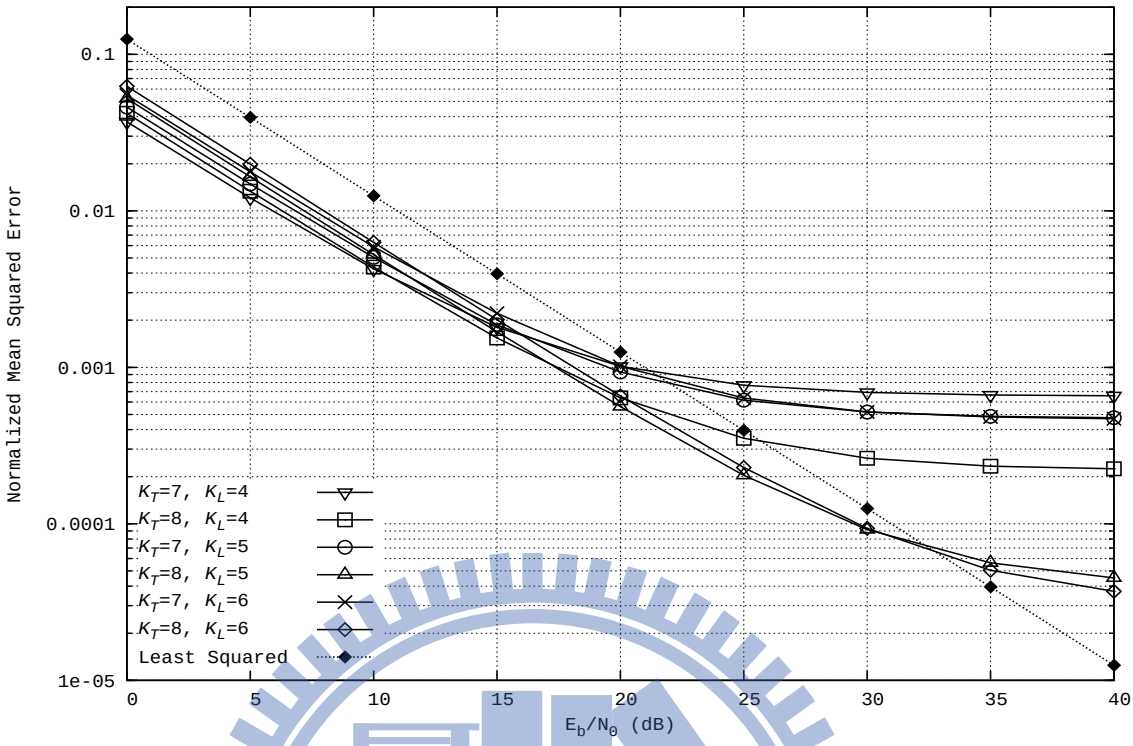


Figure 3.3: The effect of modelling order on *Algorithm B*'s MSE performance in a channel generated by the model described in [1] with AS=15° and f_dT_s=0.031772.

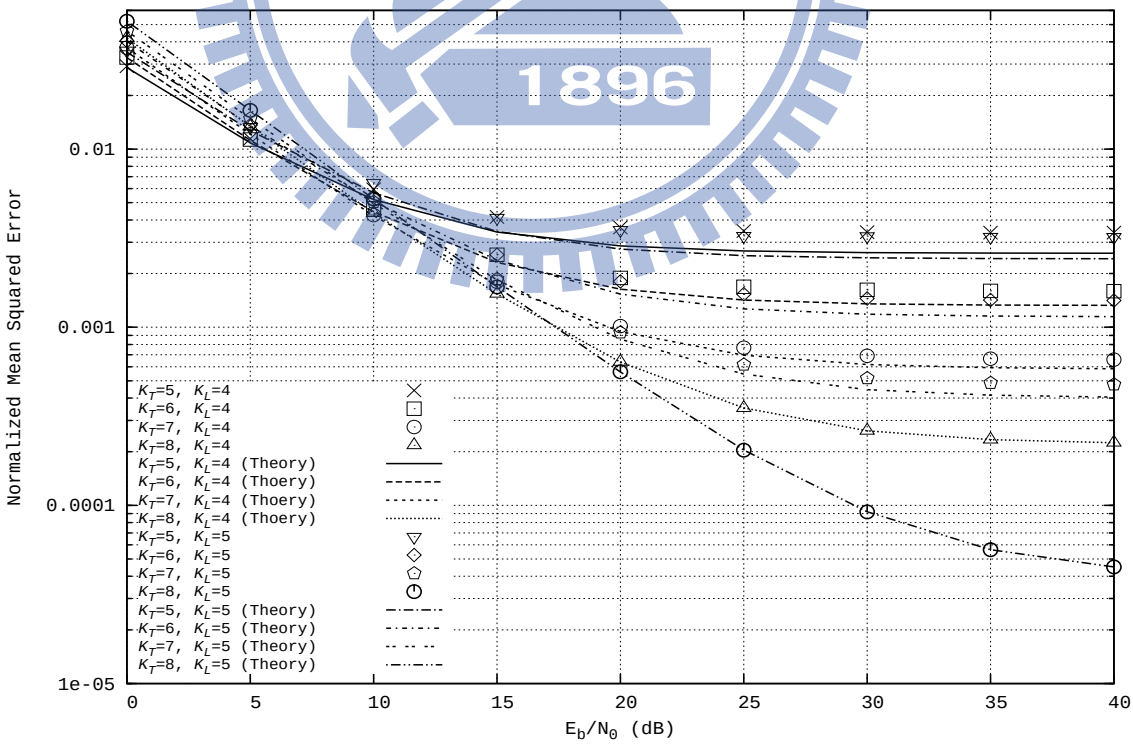


Figure 3.4: Comparison of theoretical and simulated MSE performance of *Algorithm B* in a channel generated by the model described in [1]; AS=15° and f_dT_s=0.031772.

and find that the latter is very close to the theoretical bound which assumes a perfect $\widehat{\mathbf{W}}$. When used for estimating other reference channels, the proposed estimators exhibit similar performance behaviors. Fig. 3.5 depicts the MSE performance in an IEEE 802.11 TGn channel [3] with $L = 12$, $\Delta = 15°$, and $f_d T_s = 0.0022$, while Fig. 3.6 shows the MSE performance in a 3GPP-SCM channel [35] with $L = 12$, $\Delta = 15°$ and $f_d T_s = 0.02844$. When K_T is large enough, the time-domain modelling order needed to characterize a slow fading channel like the IEEE 802.11 TGn channel is smaller than that for a fast fading SCM channel. Note that in all cases, the performance becomes independent of the AS when the full modelling order is used (i.e., $K_T = 8$) and is equivalent to that of the conventional LS approach.

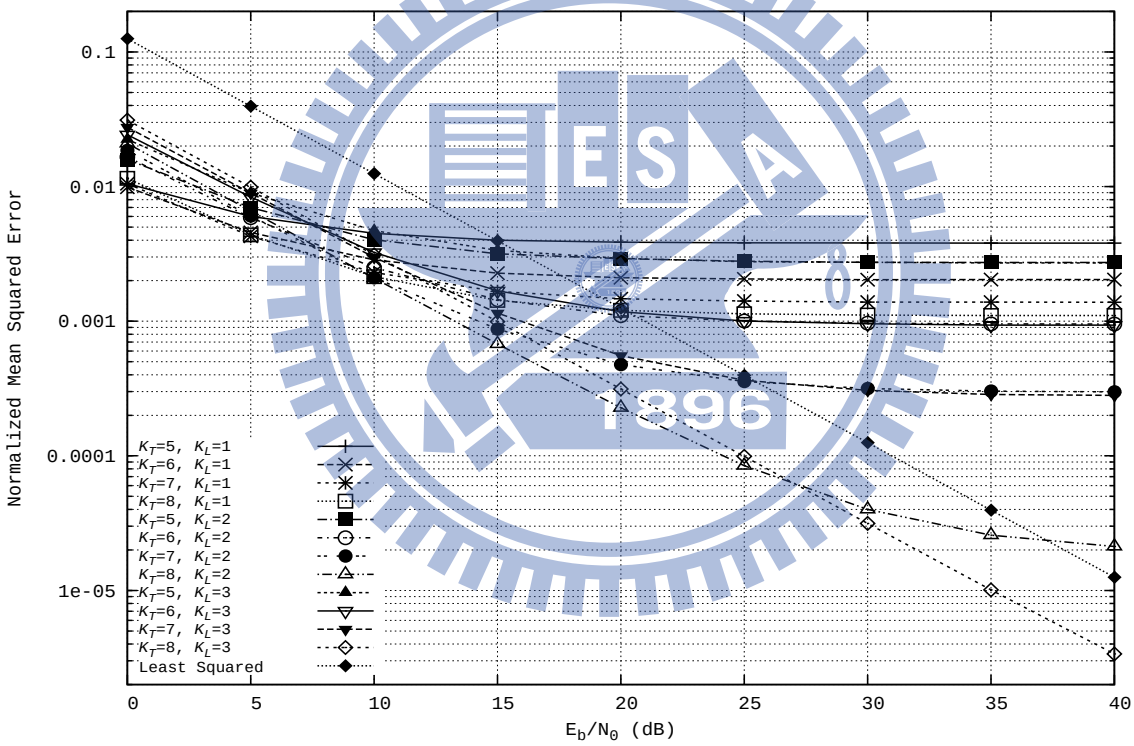


Figure 3.5: The effect of the modelling order on the MSE performance of *Algorithm B* in a channel generated by IEEE 802.11 TGn channel model A; AS=15°, and $f_d T_s$ =0.0022.

The next two numerical results assume that the algorithms developed in Section 3.1 are used and, except for Fig. 3.8, the same channel model as that used for Fig. 3.1. Fig. 3.7 compares the MSE performance of *Algorithms A* and *B* developed in Section 3.1 when $\Delta = 15°$. If the maximum matching output is obtained by selecting the best one from the outputs using 100 candidate phases uniformly distributed within $[-\pi, \pi)$,

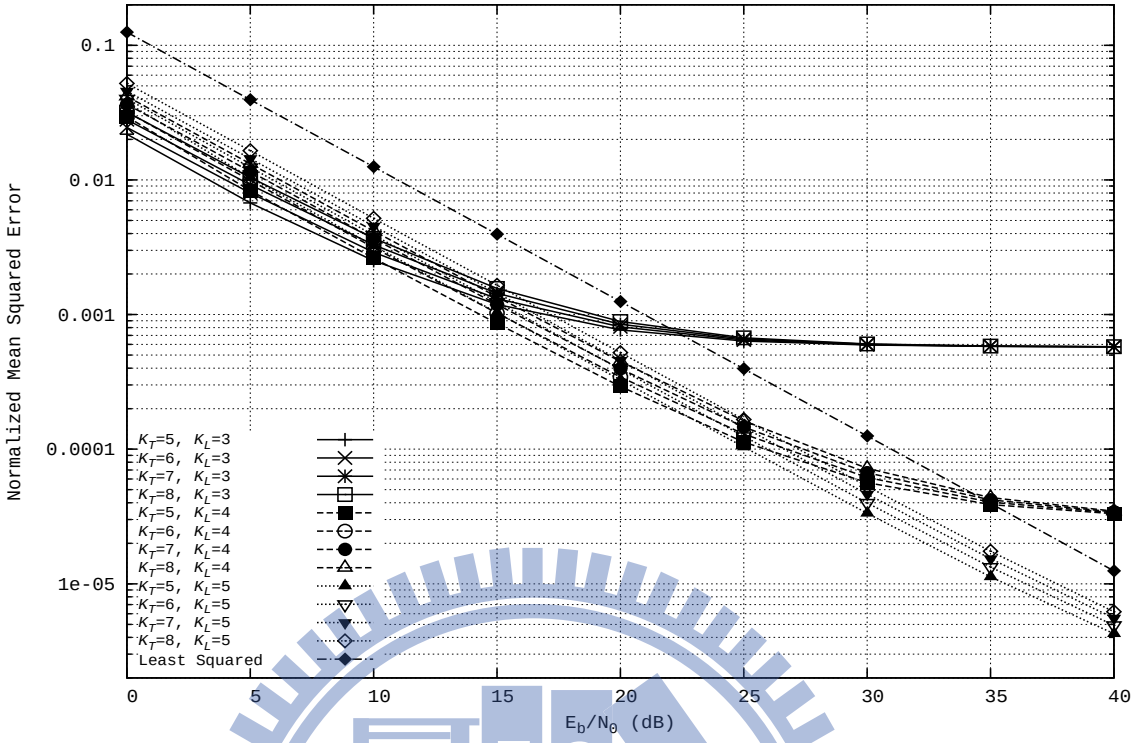


Figure 3.6: The effect of the modelling order (K_T) on the MSE performance of *Algorithm B* in a 3GPP-SCM channel; AS=15° and f_dT_s=0.02844.

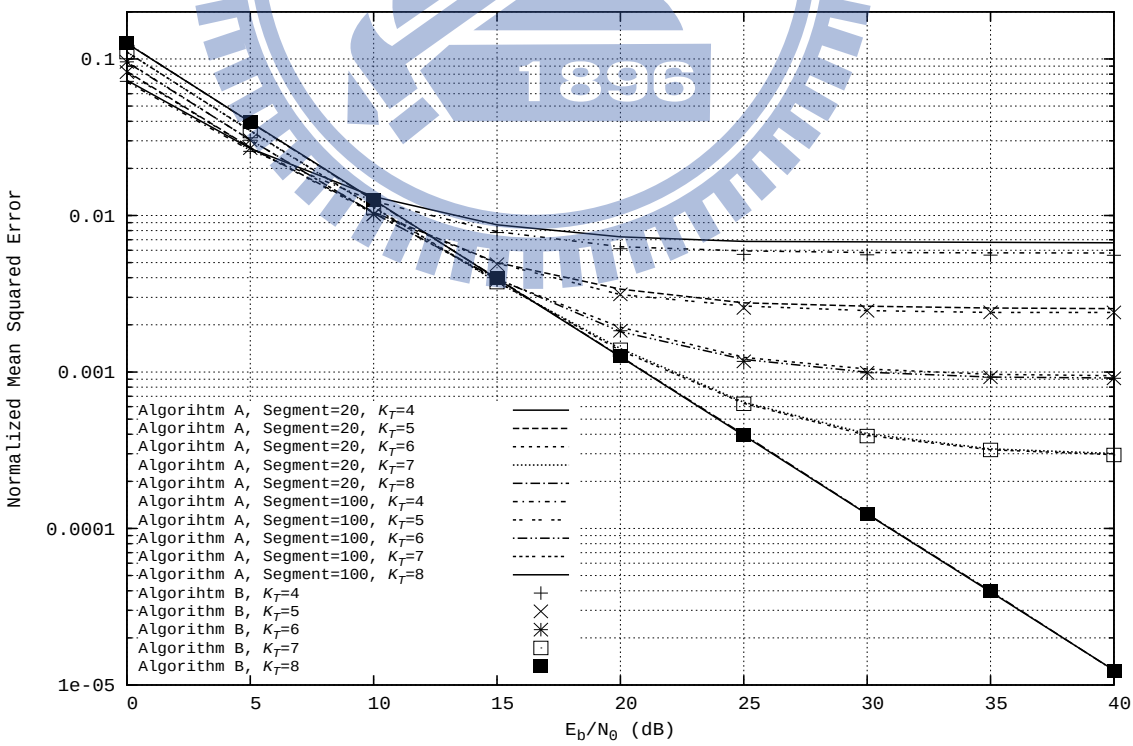


Figure 3.7: MSE performance comparison of *Algorithm A* (−−) and *Algorithm B* (−); AS=15°.

Algorithm A and *Algorithm B* give almost identical performance. However, if only 20 candidate phases are used, *Algorithm A* results in a little performance degradation with respect to that obtained by *Algorithm B* when SNR is high. Fig. 3.8 examines the MSE performance when $\widehat{\mathbf{W}}$ is updated with different EI lengths for various channel settings. Smaller performance loss results if the channel is more static or less dynamic (smaller $f_d T_s$). When $K_T \geq 3$ for channel 1 and $KT \geq 2$ for channel 2, the performance loss is negligible for all the update frequencies. Recall that more computation saving is obtained by a larger T_o^w. It is clear that our reduced-order modelling approach outperform the conventional LS estimator for most $\frac{E_b}{N_0}$ when a proper K_T is used.

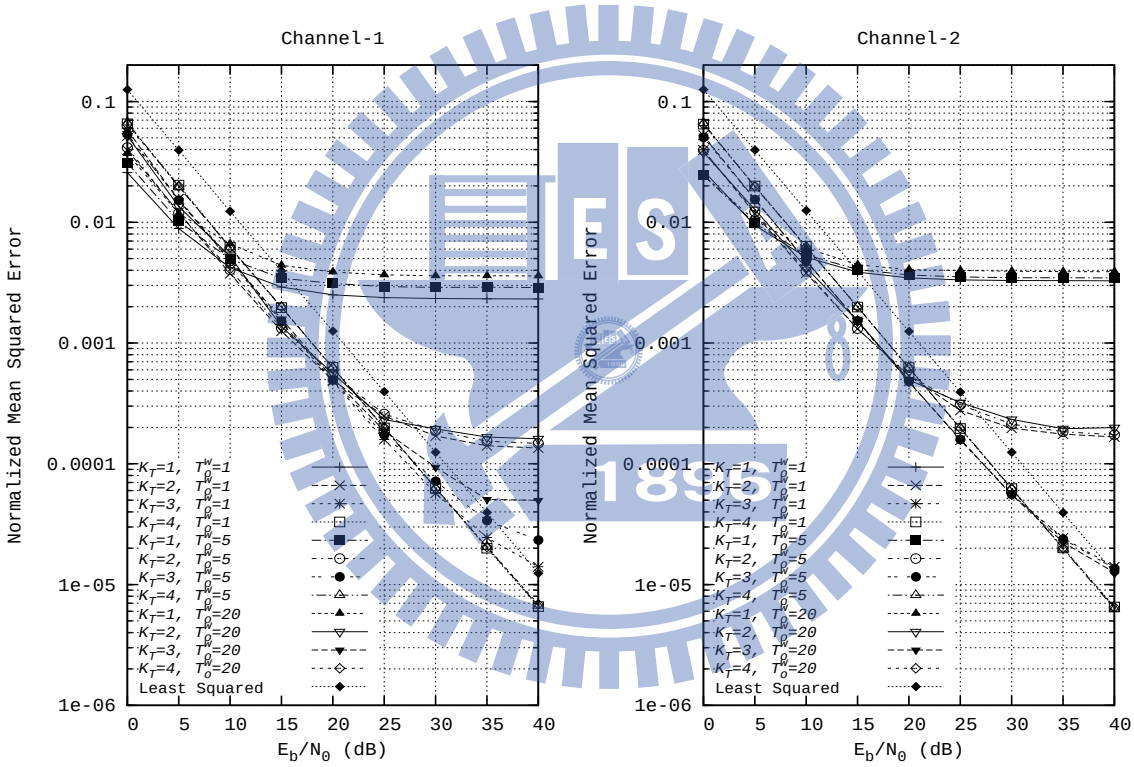


Figure 3.8: The effect of the update period on the MSE performance of *Algorithm B*. Channel-1 is based on [2] with $f_d T_s$=0.015886 while Channel-2 is based on [3] with $f_d T_s$ = 0.0022. AS=2°, $T_o^c = 1$; both T_o^c and T_o^w are measured in EIs.

Fig. 3.9 and 3.10 illustrate the MSE performance when the target MIMO channels are time-variant and frequency-selective. We use *Algorithm B* developed in Section 3.3 to estimate the instant channel waveform. Fig. 3.9 is simulated under the MIMO channels generated by [2] where the power delay profile of six independent paths is given by [0, -1, -9, -10, -15, -20] dB with relative delays of [0, 310, 610, 1090, 1730, 2510] nanoseconds.

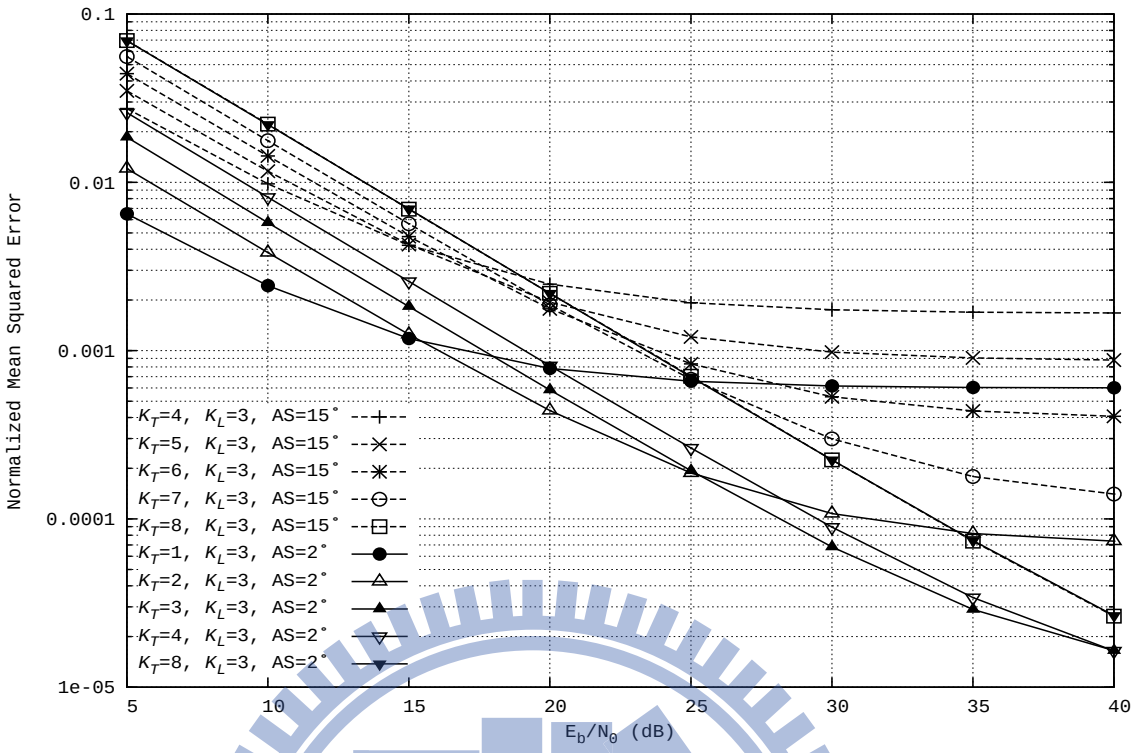


Figure 3.9: The effect of the angle spread on the MSE performance of *Algorithm B* in a channel generated by the model described in [2]; $L = 12$ and $f_d T_s$=0.015886.

Fig. 3.10 depicts the identification error under the fadings generated by [35]. In both Fig. 3.9 and 3.10, $K_L = 3$ is used to model the time-domain correlation. Compared with the frequency-nonselective cases reported above, we can also reach similar conclusions about the relationship between the performance trend and the underlying modelling order.

3.6 Summary

This chapter presents novel schemes to estimate spatial correlated MIMO fading channels based on new compact analytic models which can span the spatial and/or time correlation functions over the dominant signal subspace and provides additional directional information. Iterative algorithms are proposed for estimating spatial-correlated MIMO channels. We then extend our work to model both spatial- and time-correlated link gains associated with a MIMO channel and derive efficient estimators when the time-correlation is taken into account. We simulate the estimators' performance in various popular industry-approved and standardized channels to validate the accuracy of our model and the usefulness of our channel estimators. Numerical results show that in many instants the proposed

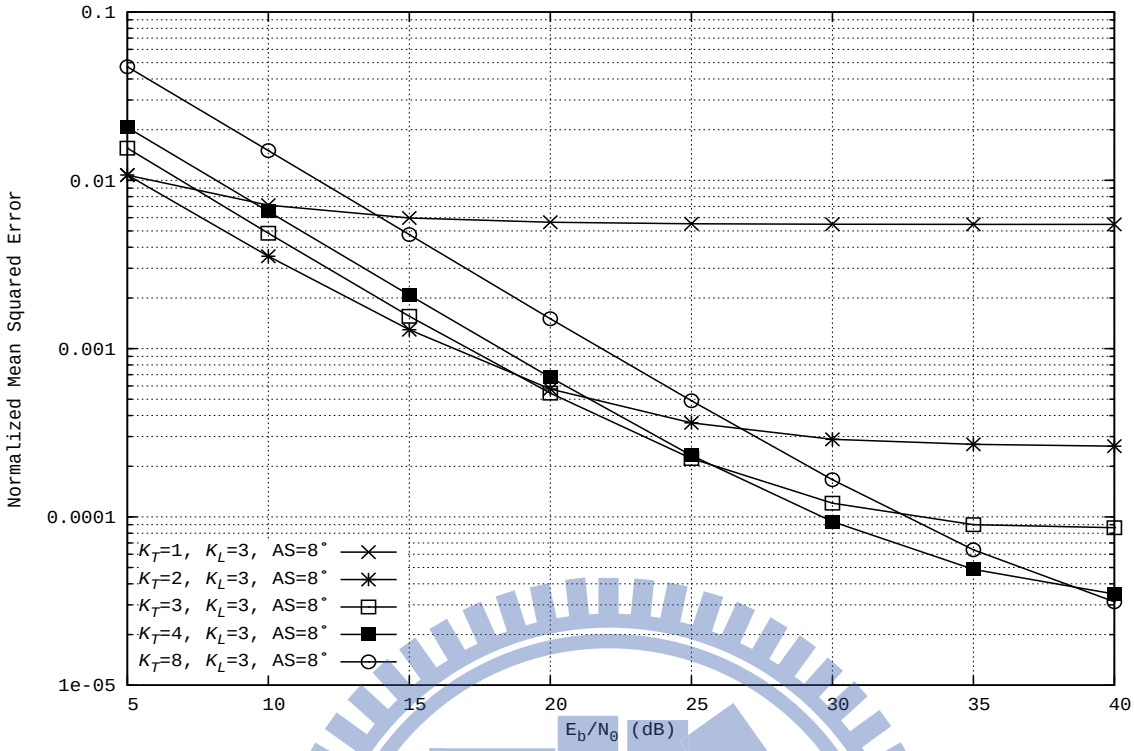


Figure 3.10: MSE performance of *Algorithm B* in an SCM channel; AS=8°, $L = 12$ and $f_d T_s$=0.02844.

algorithms give superior MSE performance. Our estimators offer tradeoffs between performance and complexity. They are easily extendable for use in wideband MIMO systems and are most effective when the channel's AS is small, i.e., when the dimension of the dominant subspace is much smaller than full channel correlation rank. Not only do they offer fast and accurate estimates, give MSE performance improvement due to the noise reduction effect but, more importantly, also provide compact and useful CSI that lead to significant feedback channel bandwidth reduction and other potential post processing complexity cutbacks.

Chapter 4

Model-Based Eigen-Beamforming

In this chapter, we present a novel transceiver design based on a nonparametric MIMO channel estimator established in the previous chapter. Optimal MIMO system performance is achieved when CSI is available at both sites of the communication link. This is usually accomplished by deriving the CSI at the receiving site and feeding it back to the transmitting site. To maintain the promised system performance, large amount of CSI must be regularly updated at the transmit side through a feedback channel. Providing channel tracking information to the transmitter either consumes feedback bandwidth or increases the feedback delay. By using a reduced-order nonparametric MIMO channel model that characterizes the channel spatial correlations, we are able to reduce the feedback requirement while compromising no system performance. We obtain bounds of the reception mean squared error and feedback information loss that can be used to assess the system performance. Numerical and simulation results based on several environment settings are given to validate the proposed method.

4.1 Modelling of Correlated MIMO Channels

4.1.1 Notations

Boldface upper-case and boldface lower-case letters denote matrices and column vectors, respectively. Italics denote scalars. $\mathbb{R}^{i\times j}$ and $\mathbb{C}^{i\times j}$ denotes the set of $i \times j$ real and complex matrices, respectively. The super-scripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote transpose, complex conjugate, and Hermitian operations, respectively. $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix and $\det(\cdot)$ denotes the determine of a square matrix. $\|\mathbf{X}\|_F$ is the Frobenius norm

of a matrix $\mathbf{X}$, and $\|\mathbf{Y}\|_2$ is the 2-norm of a matrix $\mathbf{Y}$. $\mathcal{R}(\mathbf{A})$ denotes the range (column space) of the matrix $\mathbf{A}$. diag($\mathbf{x}$) is a diagonal matrix with its diagonal entries given by the elements of vector $\mathbf{x}$, while diag($\mathbf{Y}$) denotes the column vector whose entries are the diagonal elements of matrix $\mathbf{Y}$. vec($\mathbf{X}$) denotes a column vector obtained by stacking the columns of matrix $\mathbf{X}$ into a single vector. Operator $\otimes$ denotes the Kronecker product. $\mathbf{I}_K$ denotes a $K \times K$ identity matrix. $[\mathbf{X}]_{i,j}$ denotes the (i,j)th element of $\mathbf{X}$ while $[\mathbf{X}]_{L\times M}$ signifies that $\mathbf{X}$ is an $L \times M$ matrix. Operator $(x)^+$ is defined as $\max(x, 0)$.

4.1.2 System Setup

Following the same system configuration used in Chapter 2 and Chapter 3, the base station (BS) and mobile station (MS) are equipped with linear arrays of M and N antennas, respectively. Independent data streams $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t), \cdots, x_M(t)]^T$ are transmitted at BS at time t, where $x_m(t)$ denotes the source signal at the mth transmit antenna. At the MS, the received baseband waveform is given by $\mathbf{y}(t) = [y_1(t), y_2(t), y_3(t), \cdots, y_N(t)]^T$, where $y_n(t)$ is the received signal at the nth receive antenna at time t. For notational simplicity, we define the two M-dimensional vectors $\mathbf{x}_i = \mathbf{x}(i\triangle t)$ and $\mathbf{y}_i = \mathbf{y}(i\triangle t)$, where $\triangle t$ is the sampling interval.

4.1.3 Nonparametric Channel Modelling

For the convenience of reference, we summarize the channel representation proposed in Chapter 2 as following,

$$\mathbf{H} \simeq \mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}_{M,K_T}^T\mathbf{W}, \tag{4.1}$$

where $\mathbf{W}$ is the matrix bearing the directional information, $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ are $M \times K_T$ and $N \times K_R$ matrices whose column vectors are the orthonormal bases used to describe the discrete root power correlations, K_T and K_R being the associated modelling orders. Reminding that when the angle spread Δ at BS is zero and antennas at BS are fully correlated, the waveform transmitted from the BS MEA can be regarded as a plane wave with a fixed AOD ϕ and $\mathbf{W}$ is therefore equivalent to a diagonal steering matrix,

$$\mathbf{W} = \text{diag}\left([w_1, w_2, \cdots, w_M]\right) \tag{4.2}$$

where $w_i = \exp\left[-j2\pi\frac{(i-1)d}{\lambda}\sin(\phi)\right]$, d is the inter-element distance. On the other hand, if the angle spread is large enough and the MEA at the BS tends to be fully uncorrelated, the resulting modelling order K_T used in $\mathbf{Q}_{M,K_T}$ equals M.

There are several classes of basis functions to choose from. The Taylor and Weierstrass theorem arguments and the results of [27] suggest the use of polynomial regression estimators. If we use polynomials of degree K as basis functions for estimating spatial correlation functions of length L, the corresponding basis matrix $\mathbf{P}_{L,K}$ has entries

$$[\mathbf{P}]_{l,k} = (l-1)^{k-1},\ l = 1, 2, \ldots, L, \text{and } k = 1, 2, \ldots, K, \tag{4.3}$$

where the modelling order is $K \leq L$. Although the column vectors in (4.3) can be used as bases, they are not orthogonal. Furthermore, these vectors have different norms, which might result in numerical instability. By applying the QR decomposition to $\mathbf{P}_{L,K}$ [28], we obtain the orthonormalized basis matrices for $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$.

Another class of candidate basis matrices is the discrete cosine transform (DCT) matrices. The reasons for using DCT are twofold. Firstly, DCT is very good at energy compaction for most correlated sources, especially for Markov sources with high correlation coefficient. Furthermore, the channel correlation matrix $\mathbf{R}_{\mathbf{H}}$ defined in (4.11) below tends to be a toeplitz matrix, which can be approximately diagonalized by DCT. Secondly, DCT has several well established computing structures that are both efficient and robust. A typical $L \times (K+1)$ DCT matrix is defined as

$$[\mathbf{Q}]_{l,k} = q(k)\cos\frac{\pi(2l-1)(k-1)}{2L},\ l = 1, 2, \ldots, L, \text{and } k = 1, 2, \ldots, K, \tag{4.4}$$

where $q(k) = \sqrt{\frac{1}{L}}$ for $k = 1$ and $q(k) = \sqrt{\frac{2}{L}}$ for $2 \leq k \leq K$. If the modelling order K equals to L, both the orthonormalized polynomial basis matrix and the DCT matrix become full-ranked.

4.1.4 Nonparametric Space-Time Channel Estimation

For a receiver to extract both the coefficient matrix $\mathbf{C}$ and the directional information $\mathbf{W}$ in (4.1), in Chapter 3, we develop iterative schemes which consist of two processing phases to estimate several typical channels, such as block fading, time-variant frequency non-selective, and time-variant frequency-selective channels. The proposed channel estimators

incorporate the following two steps : (i) at the ith iteration, estimate the coefficient matrix $\widehat{\mathbf{W}}_i$ based on the estimate $\widehat{\mathbf{C}}_{i-1}$ obtained from the previous iteration, and (ii) estimate the directional matrix $\widehat{\mathbf{C}}_i$ based on the tentative estimate $\widehat{\mathbf{W}}_i$. Both estimators improve as one proceeds with more iterations. The proposed nonparametric estimators can be summarized as follows.

- Given a data block of B length-M input symbol vectors, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_B]$, and the channel output block, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_B]$, the proposed channel estimator outputs the optimal solution $\{\hat{\phi}, \widehat{\mathbf{C}}\}$ for the least squared (LS) problem,

$$\left\{\hat{\phi}, \widehat{\mathbf{C}}\right\} = \arg\min_{\phi, \mathbf{C}} \; ||\mathbf{Y} - \mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}^T_{M,K_T}\mathbf{W}(\phi)\mathbf{X}||^2_F, \tag{4.5}$$

  where $\widehat{\mathbf{C}}$ is a $K_R \times K_T$ complex matrix.

- Let $\hat{w}_i = \exp\left[-j2\pi\frac{(i-1)d}{\lambda}\sin(\hat{\phi})\right]$ and define $\widehat{\mathbf{W}} \stackrel{def}{=} \mathbf{W}(\hat{\phi}) = \text{diag}\left([\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_M]\right)$, a diagonal steering matrix associated with the mean AOD estimate $\hat{\phi}$. The LS channel estimate is obtained by $\widehat{\mathbf{H}} = \mathbf{Q}_{N,K_R}\widehat{\mathbf{C}}\mathbf{Q}^T_{M,K_T}\widehat{\mathbf{W}}$.

It is clear that once the low dimensional channel representation $\left\{\hat{\phi}, \widehat{\mathbf{C}}\right\}$ becomes available, we can use them to synthesize the required CSI for feedback. In other words, these two matrices, $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{W}}$, serve as an alternative CSI that provide potential saving of feedback information.

4.2 Model-Based Optimal Transceiver Design

Based on the nonparametric channel model developed in the previous section, we will propose a basic structure for MIMO transceiver, and present the corresponding optimal minimum mean squared error (MMSE) designs. In a sense, the proposed transceiver design is a generalization of the optimal multi-dimensional eigen-beamforming systems based on the general precoding framework developed in [14], [11]. In this thesis, the terms "precoder" and "eigen-beamforming" are used interchangeably. For convenience of reference, some notations used follow closely with those used therein [11]. To begin with, we review conventional optimum designs under various performance criteria and then proceed to present the proposed framework.

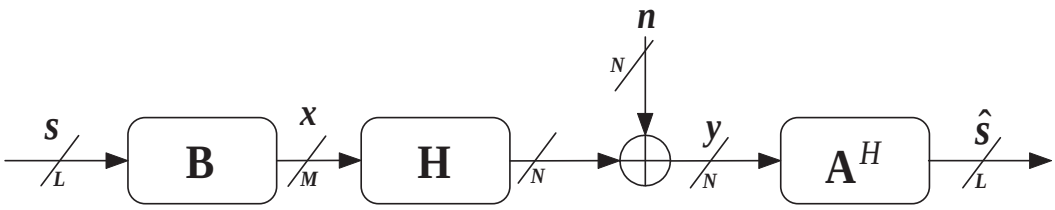


Figure 4.1: Basic structure of a general MIMO transceiver.

4.2.1 Basic Transceiver Structure

Fig. 4.1 illustrates the basic structure of a conventional MIMO transceiver in which a $L \times 1$ source vector $\mathbf{s}$ is precoded by

$$\mathbf{x} = \mathbf{B}\mathbf{s}, \tag{4.6}$$

where $\mathbf{B} \in \mathbb{C}^{M \times L}$ is a linear precoder, and $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the encoded output.

The received baseband signal can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.7}$$

where $\mathbf{y} \in \mathbb{C}^{N \times 1}$ is the channel output, $\mathbf{H} \in \mathbb{C}^{N \times M}$ represents the fading channel matrix, $\mathbf{n} \in \mathbb{C}^{N \times 1}$ is the additive noise, which is assumed to be a zero-mean circularly symmetric complex Gaussian vector with covariance matrix $\mathbf{R_n}$. At the receiving site, the source vector is estimated via

$$\hat{\mathbf{s}} = \mathbf{A}^H\mathbf{y} = \mathbf{A}^H\mathbf{H}\mathbf{B}\mathbf{s} + \mathbf{A}^H\mathbf{n}, \tag{4.8}$$

where $\mathbf{A}^H \in \mathbb{C}^{L \times N}$ represents the linear equalization operator.

4.2.2 Optimal Design under MMSE Criterion

Considering the mean squared error matrix associated with source signal $\mathbf{s}$,

$$\text{MSE} \stackrel{def}{=} \mathcal{E}\left[(\hat{\mathbf{s}} - \mathbf{s})(\hat{\mathbf{s}} - \mathbf{s})^H\right]. \tag{4.9}$$

Given fixed precoder $\mathbf{B}$, the optimum receive matrix based on MMSE criterion $\min \mathcal{E}\left\{\|\hat{\mathbf{s}} - \mathbf{s}\|_2^2\right\} = \min \text{Tr}\left\{\text{MSE}\left(\mathbf{A}, \mathbf{B}\right)\right\}$ is given by the Wiener solution [36][11],

$$\mathbf{A}_{\text{opt}} \;=\; \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_\mathbf{H}\mathbf{B}\right)^{-1}, \tag{4.10}$$

where

$$\mathbf{R_H} \stackrel{def}{=} \mathbf{H}^H \mathbf{R_n}^{-1} \mathbf{H}. \tag{4.11}$$

Since the optimal equalizer in (4.10) is deterministic when $\mathbf{B}$ is given, we express the MSE in (4.9) as a function of matrix $\mathbf{B}$

$$\begin{aligned} \mathbf{MSE}(\mathbf{B}) &\stackrel{def}{=} \mathbf{MSE}\left(\mathbf{A}_{\text{opt}}, \mathbf{B}\right) \\ &= (\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \mathbf{R_n}^{-1} \mathbf{H} \mathbf{B})^{-1}, \end{aligned} \tag{4.12}$$

with the diagonal element $[\mathbf{MSE}(\mathbf{B})]_{i,i}$ representing the achievable MSE of s_i. The individual signal to interference-plus-noise ratio (SINR) is shown to be closely related with individual MSE [11] by,

$$\text{SINR}_i = \frac{1}{[\mathbf{MSE}(\mathbf{B})]_{i,i}} - 1, \tag{4.13}$$

which says that minimizing the MSE is equivalent to maximize the SINR and thus minimize the bit error rate (BER). Since $\mathbf{MSE}(\mathbf{B})$ depends on the transmit matrix (precoder) only, the following constrained optimization problem arises.

$$\begin{cases} \min_{\mathbf{B}} \text{Tr}(\mathbf{MSE}(\mathbf{B})) \\ s.t.\ \text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, \end{cases} \tag{4.14}$$

where P_T is the maximum total transmit power.

The corresponding solution $\mathbf{B}$ can be derived by using the Karush-Kuhn-Tucker (KKT) conditions. From the eigenvalue decomposition (EVD) of $\mathbf{R_H}$

$$\mathbf{R_H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{4.15}$$

we collect the eigenvectors of $\mathbf{U}$ corresponding to the first L largest eigenvalues in descending order to form $\mathbf{U}_L \in \mathbb{C}^{M \times L}$. Note that L also represents the maximal number of parallel streams provided by the eigen-beamforming system. The optimum solution of (4.14) is given as

$$\mathbf{B}_{\text{opt}} = \mathbf{U}_L \mathbf{\Phi}. \tag{4.16}$$

$\mathbf{\Phi} = \text{diag}(\{\phi_i\})$ is a $L \times L$ diagonal matrix with its (i,i)th element being decided by the water-filling principle [14]-[11],

$$\phi_i^2 = \left(\mu \lambda_i^{-1/2} - \lambda_i^{-1}\right)^+, \tag{4.17}$$

where ϕ_i is the ith diagonal element of $\mathbf{\Phi}$ and λ_i is the ith eigenvalue of $\mathbf{\Lambda}$ in descending order. Parameter μ denotes the water level, which is chosen to satisfy $\sum_i \phi_i^2 = P_T$.

It was proved [37], [11] that, under the same fixed transmit power constraint, other design criteria such as zero forcing (ZF), minimum weighted sum of MSEs, minimum product of MSEs, minimum det ($\mathbf{MSE}(\mathbf{B})$), or maximum mutual information, give transmit precoder solutions differ only in how power is loaded on the eigen-channels associated with $\mathbf{R_H}$. This implies that if the CSI, $\mathbf{R_H}$ or $\mathbf{H}$ and $\mathbf{R}_n^{-1}$, can be perfectly estimated and fed back to the transmitter, the ideal system performance bound will be achieved. It becomes impractical when the transmit array size is large and/or the CSI must be updated for every block. This is where our proposed method comes in play.

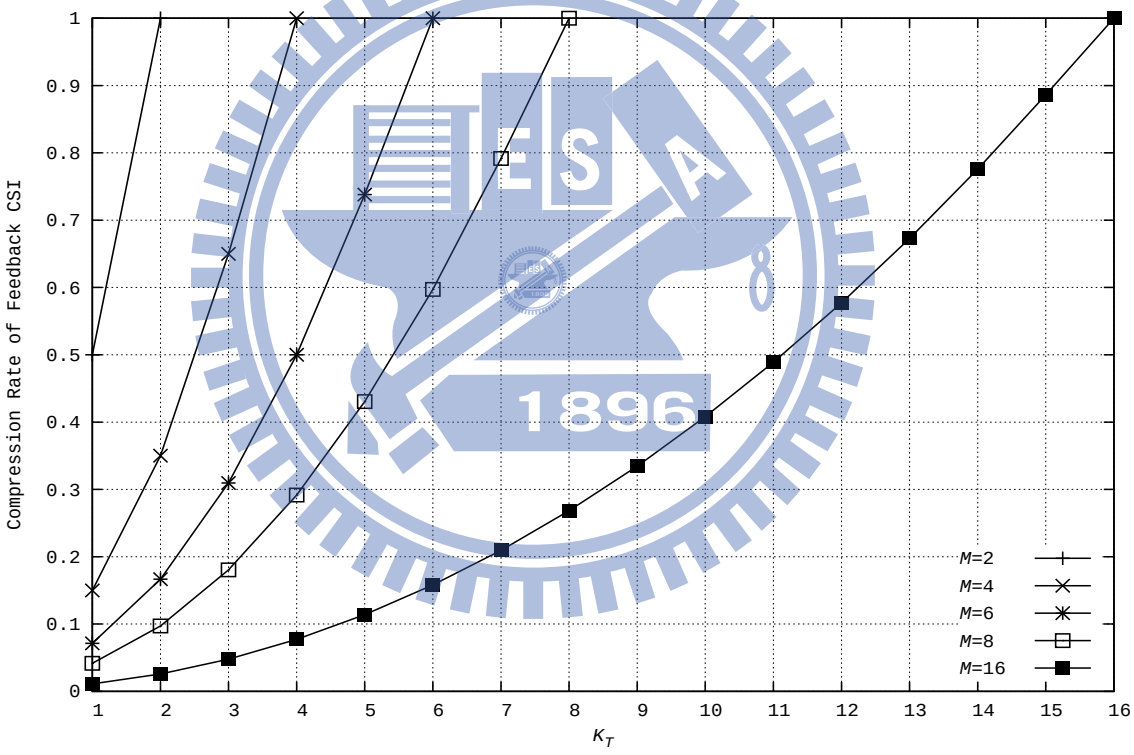


Figure 4.2: CSI compression rate of the proposed transceiver; M = number of transmit antennas, K_T = modelling order of the transmit spatial correlation.

4.2.3 CSI Compression by nonparametric channel representation

By substituting the proposed channel representation (4.1) into (4.11), we have

$$\mathbf{R_H} \simeq \mathbf{W}^H\mathbf{Q}_{M,K_T}\mathbf{C}^H\mathbf{Q}_{N,K_R}^T\mathbf{R_n}^{-1}\mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}_{M,K_T}^T\mathbf{W} \tag{4.18}$$

$$\stackrel{def}{=} \mathbf{W}^H\mathbf{Q}_{M,K_T}\mathbf{R_C}\mathbf{Q}_{M,K_T}^T\mathbf{W}. \tag{4.19}$$

The modelling bases $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ are deterministic, and are available at both transmit and receive sites. Therefore, the coefficient matrix $\mathbf{R_C} \stackrel{def}{=} \mathbf{C}^H\mathbf{Q}_{N,K_R}^T\mathbf{R_n}^{-1}\mathbf{Q}_{N,K_R}\mathbf{C}$ and the steering matrix $\mathbf{W}$ contain the same information as that in the full matrix $\mathbf{R_H}$. Since $\mathbf{R_C}$ is a Hermitian matrix, it can be represented by $\frac{K_T(K_T+1)}{2} \times 2$ real floating numbers. Moreover, only a single mean AOD ϕ is needed to fully characterize the steering matrix $\mathbf{W}$, which has a diagonal structure as described in (4.2). The total number of the feedback information in our approach is $K_T(K_T+1)+1$ while feedback of $\mathbf{R_H}$ needs $M(M+1)$ real floating numbers. The information compression rate, which is defined by the ratio between the floating numbers of $\mathbf{R_C}$ and that required by $\mathbf{R_H}$, is given by $\min\{\frac{K_T^2+K_T+1}{M^2+M}, 1\}$. Fig. 4.2 shows the compression rate performance of the proposed approach. Note that if K_T equals to M, $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ are simply orthogonal transforms, no rank reduction is obtained and so the compression ratio is just one. In such a case, the CSI is fully characterized by $\mathbf{R_C}$ and the estimate directional information $\hat{\phi}$ can be discarded.

When the additive noise at the receiver is white complex Gaussian then $\mathbf{R_n}$ is a scaled identity matrix, and the pair $\{\mathbf{H}, \mathbf{W}\}$ can fully represent the feedback information. (4.1) and (4.5) indicate that $\mathbf{C}$ and ϕ are all we need to represent the channel information. The compression rate in this case is $\min\{\frac{2K_TK_R+1}{2MN}, 1\}$, which says the pair $\{\mathbf{C}, \phi\}$ also offers a compressed CSI representation and is particularly useful when K_T is small.

For both CSI representations, the compression rate improves as the spatial correlation between antenna elements increases or as the modelling order K_T decreases.

4.2.4 Model-based Transceiver Design

From (4.18), the optimum MMSE precoder $\mathbf{B}_{\text{opt}}$ in (4.16) can be approximated according to the eigenvectors and eigenvalues of $\mathbf{W}^H\mathbf{Q}_{M,K_T}\mathbf{R_C}\mathbf{Q}_{M,K_T}^T\mathbf{W}$. After performing the

eigenvalue decomposition, $\mathbf{R_C}$ can be rewritten as

$$\mathbf{R_C} = \mathbf{U_C}\mathbf{\Lambda_C}\mathbf{U_C}^H, \tag{4.20}$$

where $\mathbf{U_C}$ is the unitary matrix composed of eigenvectors and $\mathbf{\Lambda_C}$ is a diagonal matrix with its elements being the eigenvalues of $\mathbf{R_C}$ in descending order. From (4.19) and (4.20), we rewrite $\mathbf{R_H}$ as

$$\mathbf{R_H} \simeq \underbrace{\mathbf{W}^H\mathbf{Q}_{M,K_T}\mathbf{U_C}}_{\stackrel{def}{=}\widetilde{\mathbf{U}}_\mathbf{C}}\mathbf{\Lambda_C}\underbrace{\mathbf{U_C}^H\mathbf{Q}_{M,K_T}^T\mathbf{W}}_{=\widetilde{\mathbf{U}}_\mathbf{C}^H} \tag{4.21}$$

Since $\mathbf{W}$ and $\mathbf{U_C}$ are unitary matrices, and $\mathbf{Q}_{M,K_T} \in \mathbb{R}^{M\times K_T}$ has orthonormal columns, $\widetilde{\mathbf{U}}_\mathbf{C} \in \mathbb{C}^{M\times K_T}$ also has orthonormal columns, and there exists $\widetilde{\mathbf{U}}_\mathbf{C}^\perp \in \mathbb{C}^{M\times(M-K_T)}$ such that $\widetilde{\mathbf{U}} = \left[\widetilde{\mathbf{U}}_\mathbf{C}\ \widetilde{\mathbf{U}}_\mathbf{C}^\perp\right]$ is unitary [18]. We can rewrite (4.21) as

$$\begin{aligned}\mathbf{R_H} &\simeq \widetilde{\mathbf{U}}\begin{bmatrix}\mathbf{\Lambda_C} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\widetilde{\mathbf{U}}^H\\ &\stackrel{def}{=} \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^H.\end{aligned} \tag{4.22}$$

Equation (4.22) represents the approximated eigenvalue decomposition of $\mathbf{R_H}$. The approximated precoder $\widetilde{\mathbf{B}}_{\text{opt}}$ in (4.16) can be computed as

$$\widetilde{\mathbf{B}}_{\text{opt}} = \widetilde{\mathbf{U}}_L\widetilde{\mathbf{\Phi}}, \tag{4.23}$$

where $\widetilde{\mathbf{U}}_L$ is composed of the columns in $\widetilde{\mathbf{U}}$ that are associated with the L largest eigenvalues of $\widetilde{\mathbf{\Lambda}}$. $\widetilde{\mathbf{\Phi}} = \text{diag}(\{\tilde{\phi}_i\})$ is a $L\times L$ diagonal matrix, with its (i,i)th element $\tilde{\lambda}_i$ being computed by the water-filling strategy,

$$\tilde{\phi}_i^2 = \left(\tilde{\mu}\tilde{\lambda}_i^{-1/2} - \tilde{\lambda}_i^{-1}\right)^+. \tag{4.24}$$

$\tilde{\phi}_i$ is the (i,i)th element of $\widetilde{\mathbf{\Phi}}$, and $\tilde{\lambda}_i$ is the ith largest eigenvalues of $\widetilde{\mathbf{\Lambda}}$. $\tilde{\mu}$ denotes the water level under which the total transmit power is preserved, i.e., $\sum_i \tilde{\phi}_i^2 = P_T$. From (4.22), we know that there is at most $\min\{K_T,\ \text{rank}(\mathbf{R_H})\}$ nonzero eigenvalues associated with $\widetilde{\mathbf{\Lambda}}$. Therefore, the maximum number of equivalent active subchannels (defined as $\tilde{\phi}_i^2 > 0$) of model-based transceiver systems is less than or equal to the maximum number of equivalent active subchannels (defined as $\phi_i^2 > 0$) of the transceiver systems with perfect

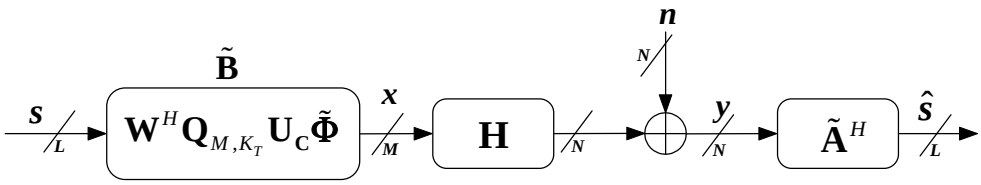


Figure 4.3: Proposed structure of the model-based MIMO transceiver.

CSI $\mathbf{R_H}$. Hence, the size of source symbols $\mathbf{s}$ should satisfy $L \leq \min\{K_T,\ \text{rank}(\mathbf{R_H})\}$ for not losing any information when the SNR is high and P_T is large.

From (4.1) and (4.10), the optimum linear equalizer $\widetilde{\mathbf{A}}_{\text{opt}}$ based on the proposed channel estimation is given as

$$\widetilde{\mathbf{A}}_{\text{opt}} \simeq \mathbf{R}_{\mathbf{n}}^{-1}\mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}_{M,K_T}^{T}\mathbf{W}\widetilde{\mathbf{B}}_{\text{opt}}\left(\mathbf{I}+\widetilde{\mathbf{B}}_{\text{opt}}^{H}\mathbf{W}^{H}\mathbf{Q}_{M,K_T}\mathbf{R}_{\mathbf{C}}\mathbf{Q}_{M,K_T}^{T}\mathbf{W}\widetilde{\mathbf{B}}_{\text{opt}}\right)^{-1}. \quad (4.25)$$

Fig. 4.3 illustrates the basic structure of the proposed model-based transceiver. Following the basic framework in this section, many optimum precoder designs in [11] can be translated to model-based designs with $\mathbf{U}$ and λ_i in (4.15) being replaced by $\widetilde{\mathbf{U}}$ and $\tilde{\lambda}_i$ in (4.22), respectively. The MSE performance of the equalization, however, depends on not only the realistic channel state but also the difference between the true channel information $\mathbf{R_H}$ and the approximated channel subspace decided by $\mathbf{W}$, $\mathbf{Q}_{M,K_T}$ and $\mathbf{R}_C$.

4.3 Performance of the Model-Based Designs

In this section, we derive performance bounds of the reception MSE in the presence of CSI modelling error. These bounds can give us a measure of how much the MSE drifts away from the optimum value when using the model-based schemes. We also suggest a distance function to assess the effects of both channel estimation error and regression modelling error.

4.3.1 MSE performance in the presence of modelling error

From (4.12) and (4.23), the MSE performance based on $\widetilde{\mathbf{B}}_{\text{opt}}$ is written as

$$\begin{aligned}
\text{Tr}\left\{\text{MSE}(\widetilde{\mathbf{B}}_{\text{opt}})\right\} &= \text{Tr}\left\{\left(\mathbf{I}+\widetilde{\mathbf{B}}_{\text{opt}}^{H}\mathbf{H}^{H}\mathbf{R}_{n}^{-1}\mathbf{H}\widetilde{\mathbf{B}}_{\text{opt}}\right)^{-1}\right\} \\
&= \text{Tr}\left\{\left(\mathbf{I}+\widetilde{\mathbf{\Phi}}\widetilde{\mathbf{U}}_{L}^{H}\mathbf{R}_{\mathbf{H}}\widetilde{\mathbf{U}}_{L}\widetilde{\mathbf{\Phi}}\right)^{-1}\right\} \\
&= \text{Tr}\left\{\left(\mathbf{I}+\widetilde{\mathbf{\Phi}}\widetilde{\mathbf{U}}_{L}^{H}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^{H}\widetilde{\mathbf{U}}_{L}\widetilde{\mathbf{\Phi}}\right)^{-1}\right\} \\
&\leq \text{Tr}\left\{\left(\mathbf{I}+\mathbf{\Phi}\mathbf{U}_{L}^{H}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^{H}\mathbf{U}_{L}\mathbf{\Phi}\right)^{-1}\right\} \qquad (4.26) \\
&= \sum_{i=1}^{L}\frac{1}{1+\gamma_i}. \qquad (4.27)
\end{aligned}$$

where $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_L$ are eigenvalues of $\mathbf{\Phi}\mathbf{U}_{L}^{H}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^{H}\mathbf{U}_{L}\mathbf{\Phi}$. The inequality of (4.26) results from the fact that the optimal precoder for transform channel $\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^{H}$ is given by $\widetilde{\mathbf{B}}_{\text{opt}}$ in (4.23). On the other hand, $\text{Tr}\left\{\text{MSE}(\widetilde{\mathbf{B}}_{\text{opt}})\right\}$ is obviously lower bounded by the optimal system performance $\text{Tr}\left\{\text{MSE}(\mathbf{B}_{\text{opt}})\right\}$ given by (4.12).

Let $\mathbf{E} \in \mathbb{C}^{M\times M}$ denotes the modelling error between the perfect CSI and the transform subspace such that $\mathbf{E} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^{H} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{H}$. By applying matrix perturbation theory [18], [38], we show that the MMSE performance is bounded by following theorem.

Theorem 4.1. *Let $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_L$ be the eigenvalues of $\mathbf{\Phi}\mathbf{U}^{H}\mathbf{E}\mathbf{U}\mathbf{\Phi}$, $\{\lambda_i\phi_i^2\}$ the diagonal elements of $\mathbf{\Phi}\mathbf{\Lambda}_L\mathbf{\Phi}$. The MMSE in (4.27) is bounded by*

$$\text{Tr}\left\{\text{MSE}(\widetilde{\mathbf{B}}_{opt})\right\} \leq \sum_{i=1}^{L}\frac{1}{1+\lambda_i\phi_i^2+\epsilon_L}, \qquad (4.28)$$

if $\min_i(\lambda_i\phi_i^2)+\epsilon_L > -1$ is satisfied.

Proof. see Appendix C □

To obtain a tighter upper bound, we need to invoke the notion of majorization.

Definition 4.1. *A descending sequence $\{b_i\}_{i=1}^{L}$ majorizes another descending sequence $\{a_i\}_{i=1}^{L}$, denoted by $\{a_i\}_{i=1}^{L} \prec \{b_i\}_{i=1}^{L}$, if $\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i$, $1 \leq k \leq L-1$ and $\sum_{i=1}^{L} a_i = \sum_{i=1}^{L} b_i$.*

Applying the majorization theory developed in [39], we obtain a refined upper bound.

Theorem 4.2. *If* $\min_i\{\lambda_i\phi_i^2\}+\epsilon_L > -1$, *the minimum mean square error (4.27) is upper bounded by*

$$\mathrm{Tr}\left\{\mathrm{MSE}(\widetilde{\mathbf{B}}_{opt})\right\} \leq \sum_{i=1}^{L} \frac{1}{1+\lambda_{\pi,i}\phi_{\pi,i}^2+\epsilon_i}. \tag{4.29}$$

Proof. see Appendix D □

The above theorems reveal that the MSE performance upper bound of any reduced-rank precoders of the form (4.23) is the same as long as its CSI modelling error results in the same dominant eigenvalue values $\epsilon_i, i = 1, 2, \cdots, L$.

4.3.2 Impact of Imperfect CSI

Having established the relationship between the MSE and the CSI modelling error, we now use a geometric measure to compare the sensitivities of our approach and conventional EVD approach against various sources of error.

Define the distance between two equidimensional subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ as [40]

$$\mathrm{dist}(\mathcal{S}_1,\mathcal{S}_2) = \sin\Theta(\mathcal{S}_1,\mathcal{S}_2) \overset{def}{=} \|(I-P_{\mathcal{S}_1})P_{\mathcal{S}_2}\|_2 = \|(I-P_{\mathcal{S}_2})P_{\mathcal{S}_1}\|_2, \tag{4.30}$$

where $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$ denote the projectors on the subspace $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.

Rewrite (4.15) in the partition form,

$$\mathbf{R_H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \overset{def}{=} [\mathbf{U}_1\ \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_2\end{bmatrix}[\mathbf{U}_1\ \mathbf{U}_2]^H \tag{4.31}$$

where $\mathbf{\Lambda}_1$ is a $K_T \times K_T$ diagonal matrix, and let $\widetilde{\mathbf{R}}_\mathbf{H} = \mathbf{R_H} + \delta\mathbf{R_H}$, where $\delta\mathbf{R_H}$ denotes the perturbation due to quantization error, channel estimation error, and uplink transmission error. Denote the eigenvalue decomposition of $\widetilde{\mathbf{R}}_\mathbf{H}$ by

$$\widetilde{\mathbf{R}}_\mathbf{H} = [\widetilde{\mathbf{U}}_1\ \widetilde{\mathbf{U}}_2]\begin{bmatrix}\widetilde{\mathbf{\Lambda}}_1 & \mathbf{0}\\ \mathbf{0} & \widetilde{\mathbf{\Lambda}}_2\end{bmatrix}\left[\widetilde{\mathbf{U}}_1\ \widetilde{\mathbf{U}}_2\right]^H. \tag{4.32}$$

Assume that $\|\cdot\| = \|\cdot\|_2$. The following lemma from the results of Wedin [41] and Fierro [40] can be used to bound the distance between the subspaces of $\mathcal{R}(\mathbf{R_H})$ and $\mathcal{R}(\widetilde{\mathbf{R}}_\mathbf{H})$.

Lemma 4.1. *Assume there exists a* $\delta > 0$ *and* $\alpha \geq 0$ *such that* $\min(\lambda(\widetilde{\mathbf{\Lambda}}_1)) \geq \alpha+\delta$, *and* $\max(\lambda(\mathbf{\Lambda}_2)) \leq \alpha$. *If* $\|\delta\mathbf{R_H}\| < \lambda_{K_T} - \lambda_{K_T+1}$, *then*

$$\sin\Theta\left(\mathcal{R}(\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^H), \mathcal{R}(\widetilde{\mathbf{U}}_1\widetilde{\mathbf{\Lambda}}_1\widetilde{\mathbf{U}}_1^H)\right) \leq \frac{\|\delta\mathbf{R_H}\|}{\lambda_{K_T}-\lambda_{K_T+1}-\|\delta\mathbf{R_H}\|}. \tag{4.33}$$

The above theorem shows that the optimal EVD-based design is relatively insensitive to small perturbation, i.e., $\|\delta\mathbf{R_H}\| \ll \lambda_{K_T} - \lambda_{K_T+1}$.

On the other hand, the channel model we use implies that $\mathbf{R_H}$ can be decomposed as

$$\mathbf{R_H} = [\overbrace{\mathbf{W}^H\mathbf{Q}_{M,K_T}}^{\stackrel{def}{=}\mathbf{Q}_1}\ \overbrace{\mathbf{W}^H\mathbf{Q}^{\perp}_{M,K_T}}^{\stackrel{def}{=}\mathbf{Q}_2}] \begin{bmatrix} \mathbf{R_C} & \mathbf{F}^H \\ \mathbf{F} & \mathbf{G} \end{bmatrix} [\mathbf{W}^H\mathbf{Q}_{M,K_T}\ \mathbf{W}^H\mathbf{Q}^{\perp}_{M,K_T}]^H \tag{4.34}$$

where $\mathbf{F}$ and $\mathbf{G}$ are the residual coefficient matrices that cause the modelling error. The matrix $\mathbf{Q}^{\perp}_{M,K_T}$ contains the residual $M - K_T$ columns of $\mathbf{Q}_{M,M}$ excluding those columns in $\mathbf{Q}_{M,K_T}$. $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have orthonormal columns that span the whole column/row space of $\mathbf{R}_H$, and $\mathbf{Q}_2 = \mathbf{Q}_1^{\perp}$.

$\widetilde{\mathbf{R}}_{\mathbf{H}}$ can also be represented by the unitary decomposition

$$\widetilde{\mathbf{R}}_{\mathbf{H}} = [\mathbf{Q}_1\ \mathbf{Q}_2] \begin{bmatrix} \widetilde{\mathbf{R}}_{\mathbf{C}} & \widetilde{\mathbf{F}}^H \\ \widetilde{\mathbf{F}} & \widetilde{\mathbf{G}} \end{bmatrix} [\mathbf{Q}_1\ \mathbf{Q}_2]^H . \tag{4.35}$$

Based upon the above observations, we obtain an upper bound on the distance between the eigen-subspace $\mathcal{R}(\mathbf{U}_1\mathbf{\Lambda_1}\mathbf{U}_1^H)$ and the perturbed modelling subspace $\mathcal{R}(\mathbf{Q}_1\widetilde{\mathbf{R}}_{\mathbf{C}}\mathbf{Q}_1^H)$ as a lemma to Fierro [40].

Lemma 4.2. *Assume there exists a* $\delta > 0$ *and* $\alpha \geq 0$ *such that* $\min(\lambda(\widetilde{\mathbf{R}}_{\mathbf{C}})) \geq \alpha + \delta$ *and* $\max(\lambda(\mathbf{G})) \leq \alpha$, *we have*

$$\sin\Theta(\mathcal{R}(\mathbf{U}_1\mathbf{\Lambda_1}\mathbf{U}_1^H), \mathcal{R}(\mathbf{Q}_1\widetilde{\mathbf{R}}_{\mathbf{C}}\mathbf{Q}_1^H)) \leq \frac{2\|\widetilde{\mathbf{F}}\| + \|\delta\mathbf{R_H}\|}{\min(\lambda(\widetilde{\mathbf{R}}_{\mathbf{C}})) - \lambda_{K_T+1}} \tag{4.36}$$

$$\leq \frac{2\|\widetilde{\mathbf{F}}\| + \|\delta\mathbf{R_H}\|}{\lambda_{K_T} - \lambda_{K_T+1} - 2\|\widetilde{\mathbf{F}}\| - \|\delta\mathbf{R_H}\|}. \tag{4.37}$$

From *Lemmas* 4.1 and 4.2, we conclude that, compared with the EVD-based optimal design, our approach is slightly more sensitive to perturbation $\delta\mathbf{R_H}$ because of the existence the off diagonal block matrix $\widetilde{\mathbf{F}}$. Hence, we should select a basis for expanding the channel spatial correlation that minimizes $\|\widetilde{\mathbf{F}}\|$ under the modelling order constraint.

4.4 Limited Feedback using Model-Based Estimated CSI

Before feeding back the instantaneous CSI estimation to the transmitter side, proper quantization schemes can be applied to reduce the data rate needed on the reverse link

[42]. To do this, conventionally, we can quantize the channel directly, feed back the quantized CSI and let the transmitter calculate the precoder assuming that the feedback CSI is perfect [43],[44]. Or, the recevier can pick the linear precoder/eigen-beamforming matrix from a set of pre-calculated codebooks according to the channel estimation results and sent the index back to the transmitter over a feedback channel [45],[46],[47]. Generally, to design a codebook aims to find a finite set, or packing, of subspaces that represent the CSI or precoders [47].

Conventionally, we can use the theory of Grassmannian subspace packing [47] to establish such codebooks. Grassmannian subspace packing method attempts to design finite sets of matrices that maximize the minimum subspace distance. Besides the subspace distance presented in the previous section, various distance measures can also be used [48] to optimize different performance criteria for various linear receivers. Alternatively, vector quantization (VQ) is an technique that we can use to construct the codebooks [49]. Basically, VQ minimizes a selected distortion function by using an iterative numerical method, such as the conventional Lloyd algorithm. After several iterations, the algorithm tends to converge to a near optimal solution and generate a candidate codebook. Both Grassmannian and VQ methods can generate good enough codebooks that achieve satisfactory error rate performance under spatially uncorrelated Rayleigh channels.

Theoretically, the distortion cost functions which serve as the design criteria of Grassmannian or VQ grow exponentially with the number of transmit antennas M provided that the number of spatial streams L is fixed [47]. Under the framework of the proposed channel representation scheme developed in Chapter 2, the equivalent CSI, $\mathbf{C}$, in (4.1) is of size $K_R \times K_T$. The exponent of the distortion cost function is thus scaled by a factor of $\frac{K_T}{M}$. For example, considering the mean squared error selection criterion (MSE-SC) [47],

$$\mathbf{B} = \arg\min_{B_i \in \mathcal{B}} \operatorname{tr}\left(\mathrm{MSE}(\mathbf{B}_i)\right) \tag{4.38}$$

where $\mathrm{MSE}(\cdot)$ is defined in (4.12) and $\mathbf{B}_i$ is the ith precoding matrix of the set of codebook matrices $\mathcal{B} = \{\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_S\}$. Here, S denotes the total number of precoding matrices used in the codebook and the number of feedback bits equals to $\log_2(S)$. Under the condition of high SNR and white Gaussian noise, we can bound $\operatorname{tr}\left(\mathrm{MSE}(\mathbf{B}_i)\right)$ in (4.38), and

characterize the distortion of the Grassmannian subspace packing by using the following cost function modified from the results of [47]

$$E_{\mathbf{H}}\left[\lambda_{\min}\{\mathbf{H}\mathbf{B}_{\text{opt}}\} - \max_{\mathbf{B}_i\in\mathcal{B}} \lambda_{\min}\{\mathbf{H}\mathbf{B}_i\}\right] \tag{4.39}$$

$$\lessapprox E_{\mathbf{H}}\left[\lambda_L\{\mathbf{H}\}\phi_L^2\right] \cdot \left(1 + S \cdot \left(\frac{\delta_{\text{proj}}}{2\sqrt{L}}\right)^{2ML+o(M)} \left(\frac{\delta_{\text{proj}}^2}{4} - 1\right)\right), \tag{4.40}$$

where ϕ_L is given in (4.17), $\mathbf{B}_{\text{opt}}$ is defined in (4.16), and $\lambda_L(\mathbf{M})$ denotes the Lth largest eigenvalue of matrix $\mathbf{M}^H\mathbf{M}$. δ_{proj} is the minimum distance between any two subspaces

$$\delta_{\text{proj}} = \min_{1\le i<j\le S} d_{\text{proj}}(\mathbf{B}_i, \mathbf{B}_j) \tag{4.41}$$

where

$$d_{\text{proj}}(\mathbf{B}_i, \mathbf{B}_j) = \|\mathbf{B}_i\mathbf{B}_i^* - \mathbf{B}_j\mathbf{B}_j^*\|_2 = \sqrt{1 - \lambda_{\min}\{\mathbf{B}_i^*\mathbf{B}_j\}} \tag{4.42}$$

denotes the projection two-norm distance. Since we always have $\delta_{\text{proj}} < 1$, the distortion cost function is an increasing function of M. Assuming that the modelling order K_T can fully characterize the signal space of $\mathbf{R}_{\mathbf{H}}$, the exponent, $2ML+o(M)$, inside the distortion cost function (4.40) can be replaced by $2K_TL+o(K_T)$. Thus, the distortion cost function decreases for any $K_T < M$. Accordingly, either the proposed model-based system can achieve smaller distortion by using the codebook of the same size as that used in the original precoding system, or we can have similar distortion with a codebook of smaller size such that feedback data rate is reduced.

Although both Grassmannian and VQ methods can generate good enough codebooks under spatially i.i.d. Rayleigh channels, the reception performance degrades when the i.i.d. codebook design is used for spatially correlated channels [42]. Under the i.i.d. assumption, the isotropicity of the dominant right singular vector of $\mathbf{H}$ allows a Grassmannian subspace packing solution [47]. For spatial correlated MIMO channels, this isotropicity, however, is destroyed and large degradation of reception performance is possible when the i.i.d. codebook is used in the correlated environment [50]. This implies that for spatially correlated MIMO channels, an i.i.d. codebook of larger size is necessary to achieve similar error rate performance as that of the spatially uncorrelated channels. Otherwise, we have to quantize the dominant space nonuniformly [50]. Under the framework of the

proposed model-based system, since the subspace packing is more compact (i.e., smaller distortion can be achieved using the same number of feedback bits), the reception performance can be improved as well. Moreover, even if exact spatial correlation is used in designing a statistical precoding systems, there will be significant performance loss for the so-called "mismatched channels" [50],[20],[51]. Here, the term "mismatched channel" means that: 1) the eigenvalues of the transmit covariance matrix can not be partitioned into two components: a dominant component of L eigenvalues that is well-conditioned* and a sub-dominant component of $M - L$ eigenvalues that is ill-conditioned away from the dominant component, and 2) the eigenvalues of the receive covariance matrix are ill-conditioned. In practice, however, the "mismatched channel" is very common for spatial correlated environments. Under the framework of proposed model-based scheme, the orthogonal matrix $\mathbf{Q}_{M,K_T}$ plays the role of pre-conditioned matrix and the correlation matrix of the coefficient matrix $\mathbf{C}$ is generally better conditioned compared with that of the original channel matrix $\mathbf{H}$. In other words, the reception performance of statistical precoding systems can be improved if the Grassmannian subspace packing is performed on $\mathbf{C}$ rather than $\mathbf{H}$ in spatial correlated environments.

4.5 Simulation Results and Discussions

Simulation results reported in this section use the reference MIMO channel model of [52], [2]. We consider an 8×8 MIMO system. Spatial correlation matrices are generated by the power azimuth spectrum (PAS) at the BS and MS respectively according to the specific physical settings. To comply with the *one-ring* model, we assume that the environment surround MS is rich scattering and uncorrelated. Different physical settings at BS such as angle spread and nonzero AOD are used to examine the effect of different degrees of spatial correlation. The size of source vector, L, is 2, and the BPSK constellation is assumed. Each element of the channel matrix is normalized to $E\left[|\mathbf{H}_{i,j}|^2\right] = 1$, for $1 \leq i \leq N$, and $1 \leq j \leq M$. The SNR is defined as the total transmitted power over the noise variance at each received antenna. Other assumptions used in our simulation are: (i) the antenna

*If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$ denote the first L eigenvalues and $\frac{\lambda_1}{\lambda_L}$ is not significantly larger than 1, we can roughly say that these eigenvalues are well-conditioned. Oppositely, we say that they are ill-conditioned.

spacing at transmit and receive antennas are both half wavelength, and (ii) additive noise distribution at the receiver is complex white Gaussian. The DCT bases are used in the proposed transceiver since we found that the DCT and polynomial bases give almost the same performance. For comparison purpose, we also present performance of the optimal transceiver with a full rank estimated CSI feedback (F-CSI).

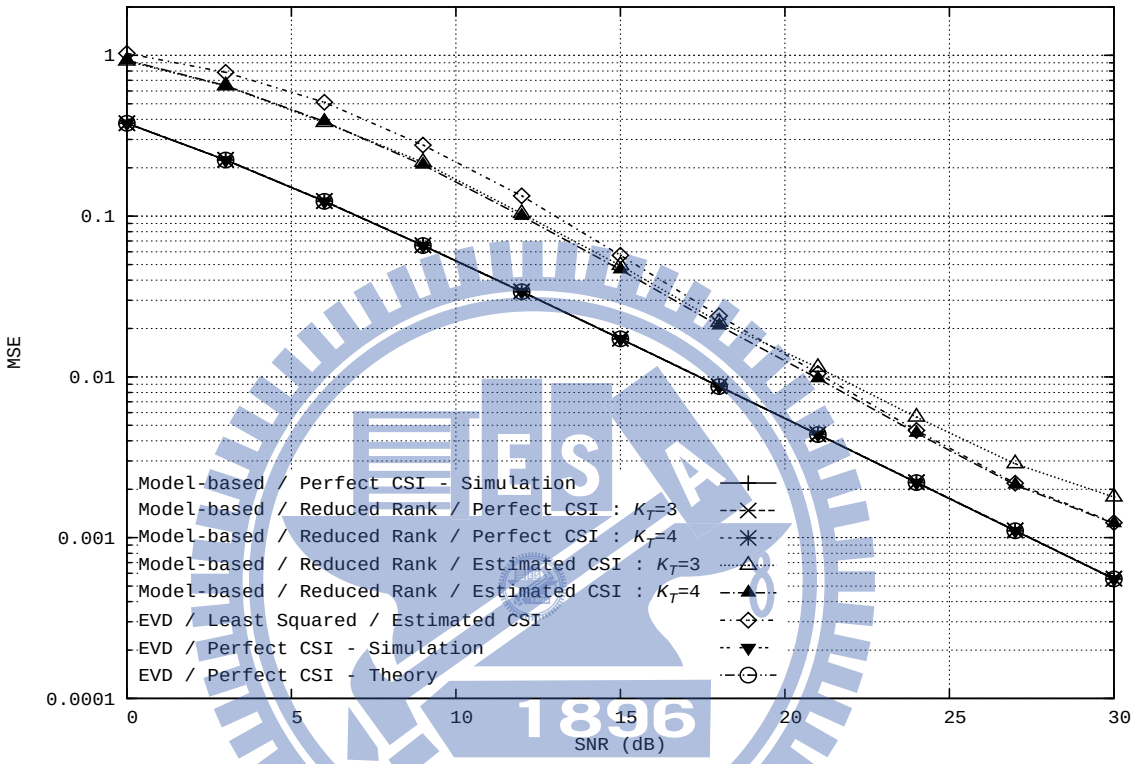


Figure 4.4: MSE performance of DCT-based transceiver; angel spread = 4°, AOD = 45°, $L = 2$.

Performance curves in Fig. 4.4 represent the MSE performance for the MIMO system with AOD = 45° and angle spread $\Delta = 4°$. The simulation results are obtained by averaging over 100 random channels. Since the channel correlation at BS is high, the corresponding correlation function lies in functional subspace of small dimension. Hence, even the number of antennas at BS increases, the amount of feedback information required by the proposed technique remains low and the performance degradation with respect to that of the optimal transceiver with F-CSI remains negligible. When the modelling order is only 3, which gives a compression rate of 0.18, the MSE performance of the proposed model-based system is still very close to that of the system with F-CSI. The MSE gap

between the performance of the optimal transceiver with perfect CSI and that of the transceivers with estimated CSI is obviously due to channel estimation error.

As the angle spread Δ increases, correlation between the transmit antennas diminishes and a higher modelling order is necessary to describe the rapid-changing spatial correlation at the transmit site. MSE curves in Fig. 4.5 are the performance of an 8×8 MIMO system with AOD$=45°$, angle spread $\Delta = 15°$, and different modelling orders. Simulation results indicate that bases of order less than 4 tend to incur larger modelling errors while those of order larger or equal to 4 provide performance which is almost the same as that of the optimal transceiver with F-CSI.

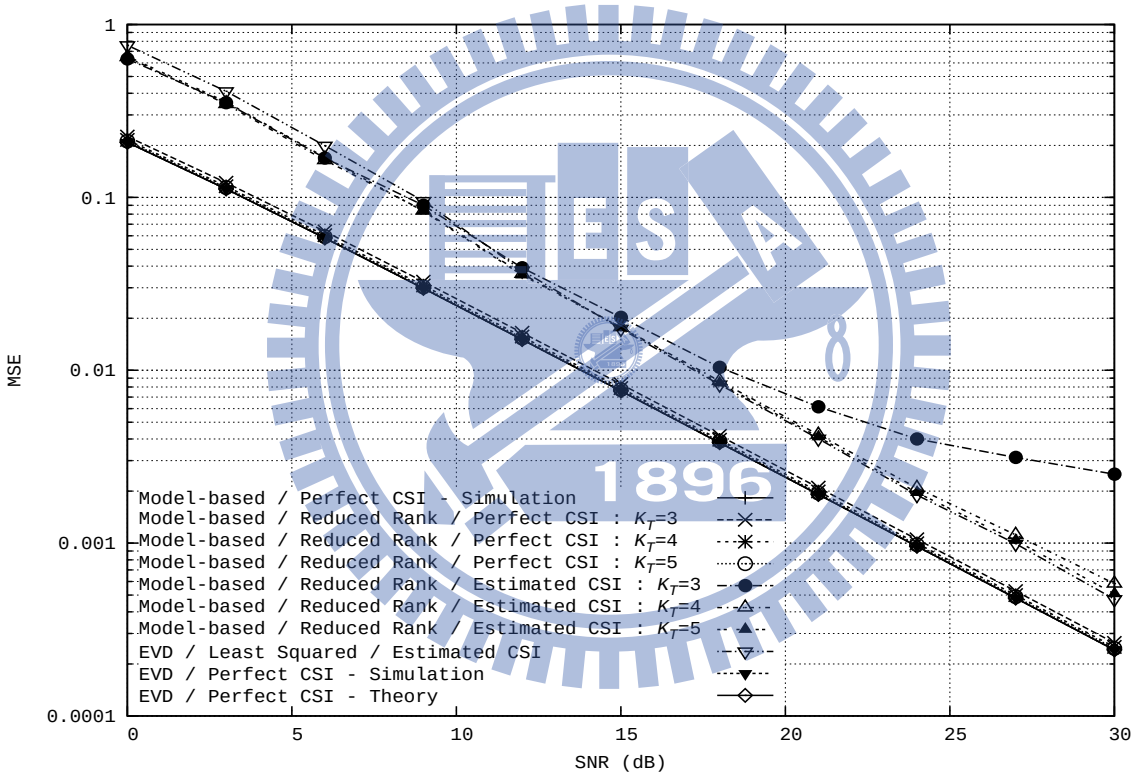


Figure 4.5: MSE performance of DCT-based transceiver; angel spread $= 15°$, AOD $= 45°$, $L = 2$.

In both figures, we notice that the proposed approach outperform the F-CSI approach at low SNRs. This observation indicates that although our approach introduces modelling error due to reduced-rank regression model it also reject the noise outside the modelling space. At higher SNRs, like other model-based methods, the modelling error dictates the performance whence this advantage gradually disappears.

The usefulness of the upper bounds derived in *Theorems* 4.1 and 4.2 are demonstrated in Fig. 4.6. It is shown that both bounds predict correct trend of the MSE performance of the system. These bounds becomes tighter as the modelling order increases. Although the bound of (4.29) is tighter than (4.28), (4.28) needs only the knowledge of the smallest eigenvalue, ϵ_L.

We use the two perturbation upper bounds given in *Lemma* 4.2 to review the effect of CSI error from a geometric perspective in Fig. 4.7(a), assuming AS = 2°. The distance between the subspace associated with perturbed CSI using a rank 1 ($K_T = 1$) approximation and that associated with the perfect CSI increases as the quality of channel estimation deteriorates at lower SNRs, and remains steady for the case of good channel estimation (high SNR).

In Fig. 4.7(b), we plot the distance between the above-mentioned two subspaces for the cases of AS equals to 4°. These curves show that in higher correlation case, a rank-1 model is sufficient to describe the spatial correlation and thus the corresponding distance is small. For the larger AS case, since the channel correlation decreases the subspace distance increases for the rank 1 system and a larger modelling order is needed.

4.6 Summary

This chapter presents a novel regression model-based transceiver design for spatial correlated MIMO fading channels. Orthogonal bases and an additional AOD information are used to model the spatial correlation functions associated with MEAs in BS and MS so that compact CSI representation can be obtained. Optimum precoding strategies are provided based on the proposed channel representation. Computer simulation results show that excellent performance is attainable if proper modelling basis and order are used. The modelling order provides trade-off between reception performance and feedback complexity. Significant feedback compression is achieved if the channel spatial correlation is high. To analyze the performance loss, we derive perturbation bounds for the reception MMSE caused by CSI modelling errors. We also provide bounds for the distance between the signal subspace associated with perfect CSI and that associated with the proposed

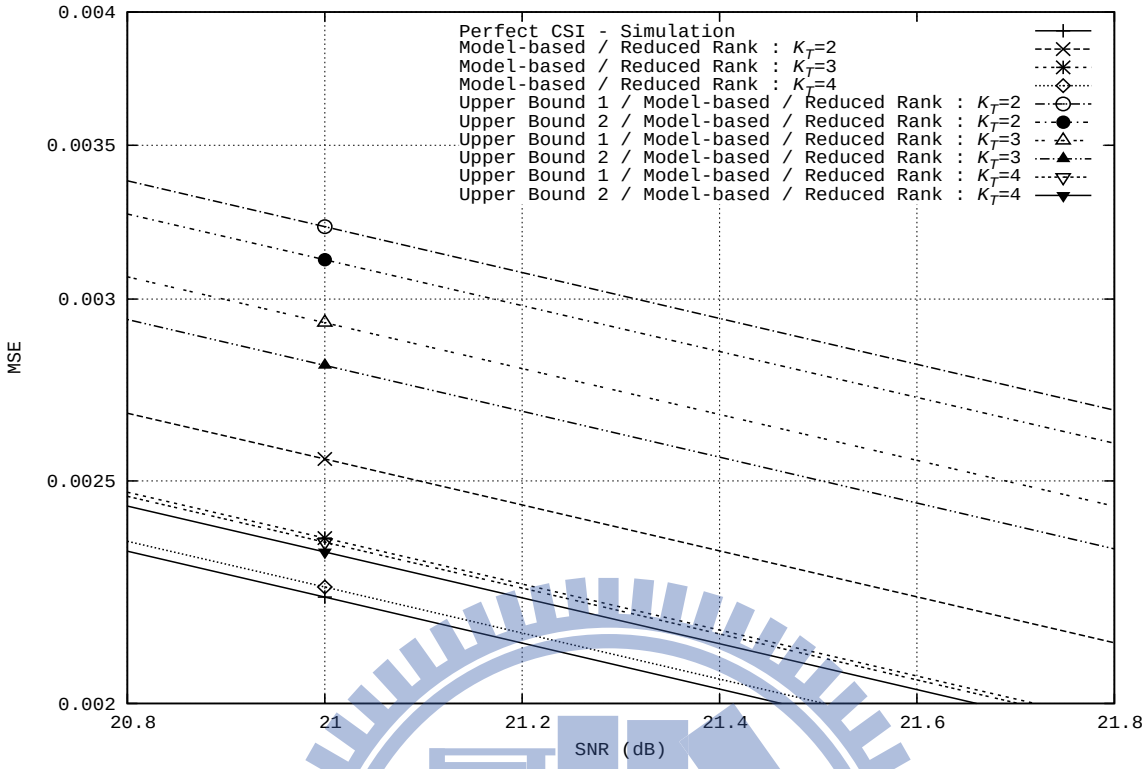


Figure 4.6: MSE upper bounds of DCT-based transceiver; angel spread = 10°, AOD = 45°, $L = 2$.

approach for which only imperfect CSI is available. Numerical results for these bounds are given to show that performance trends can be accurate predicted.

4.7 Acknowledgement

We would like to thank the European Union IST project ISI-2000-30148 I-METRA Deliverable D2 for providing a MIMO simulation package as the reference space-time channels.

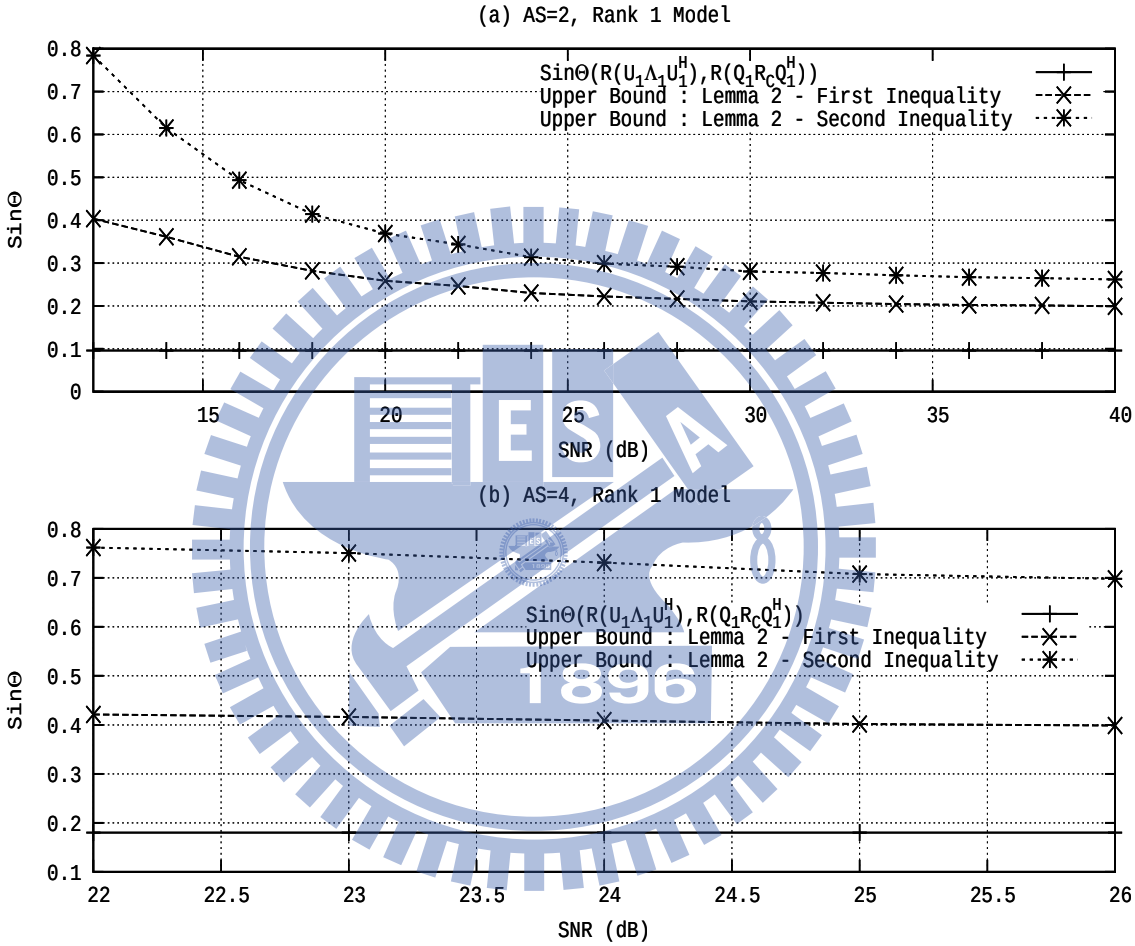


Figure 4.7: (a): angle spread = 2°, AOD = 45°, and (b): angle spread = 4°, AOD = 45°. □ : distance between signal subspaces of perfect CSI and model-based CSI using a rank-1 approximation; △ : perturbation bound of (4.36). ∘ : perturbation bound of (4.37).

Chapter 5

Conclusion and Future Work

This thesis presents a framework of nonparametric model-based MIMO systems which are basically based on the proposed compact analytic models for correlated MIMO fading channels. The proposed work models both spatial- and time-correlated link gains associated with a MIMO channel and derive efficient estimators when the spatial and/or time-correlation is taken into account. For correlated MIMO channels, by spanning the spatial and/or time correlation functions over the dominant signal subspace using a set of orthogonal modelling bases, we obtain an efficient channel representation that can alleviate the processing complexity for many post-channel-estimation processes and reduce the feedback bandwidth requirement for MIMO precoding systems as well. Tremendous computation saving and large reduction of feedback data rate are accessible especially for large MIMO systems and highly correlated environments. Based on the proposed models, we develop channel estimation schemes against several typical channel situations. Iterative batch algorithms are proposed to accomplish the task of channel estimation, and the sequential adaptive algorithms are also available for channel tracking [17]. Various popular industry-approved and standardized channels are simulated to validate the accuracy of our model and the usefulness of our channel estimators. Numerical results show that the proposed algorithms can provide tradeoffs between performance and complexity. Moreover, we also show that under different channel conditions, the modelling order that leads to dimension reduction may also achieve the best MSE performance due to the noise reduction effect discussed in Chapter 3. In this situation, we can provide compact and useful CSI that lead to significant feedback channel bandwidth reduction and other po-

tential post processing complexity cutbacks while retaining good reception performance at the same time.

Based on the proposed channel representation, optimum precoding/eigen-beamforming strategies are provided for feedback MIMO systems. Computer simulation results show that excellent performance is attainable provided proper modelling basis and order are used. Over a wide range of interested SNR, the reception performance of the model-based eigen-beamforming system using optimal modelling order is shown to be better than that of the conventional eigen-beamforming systems, which use full dimensional LS channel estimation results as feedback CSI. Significant feedback compression is achieved if the channel spatial correlation is high. We derive perturbation bounds for the reception MMSE to analyze the performance loss caused by CSI modelling errors. We also provide bounds for the distance between the signal subspace associated with perfect CSI and that associated with the proposed approach for which only imperfect CSI is available. Numerical results for these bounds are given to show that performance trends can be accurate predicted. For limited feedback MIMO systems, we prove that the proposed model will lead to fewer distortion or compact compression if a conventional quantization scheme, such as Grassmannian packing, is used.

The framework of the proposed model-based MIMO system leads to a new class of model-based MIMO processing techniques. Similar to the proposed precoding/eigen-beamforming systems, MIMO detection schemes that incorporate the MIMO channel matrices, such as *LLL* aglorithm which is based on the theory of lattice reduction [53], and the sphere decoding algorithm that performs the QR factorization of the channel matrix [54],[55], can be derived accordingly. Besides the potential saving of computational complexity, the proposed channel representation may lead to essential improvement of the lattice structure embedded in the underlying channel matrix. Such improvement is suspected to bring additional reception benefits. Hence, experimental algorithms and theoretical studies are going to be conducted to reveal the potential usages of the proposed model-based schemes.

The performance improvement made by the proposed model-based systems somehow

depends on the accuracy of the selected modelling order. Although, in general, we can use a modelling order larger than is needed to guarantee a negligible modelling error, representation efficiency and processing advantages will diminish for an over-modelling system. The order determination scheme provided in this thesis calculates the optimal order based on the long-term channel statistics. In practice, for a more non-stationary channel, the modelling order should be update in a short period. Hence, a order determination scheme, which is computational more efficient and works reliably under non-stationary environment is what we try to explore next.

Appendix A

AoD Information Extraction

For small Δ, the correlation between two transmit antennas i, j can be approximated by [6]

$$E\left\{h_{mi}h_{mj}^{*}\right\} \approx \exp\left\{-j\frac{2\pi}{\lambda}(i-j)d\sin\phi\right\} J_0\left(\Delta\frac{2\pi}{\lambda}(i-j)d\cos\phi\right). \tag{A.1}$$

In addition, correlation between two receive antennas p, q can be approximated by $E\left\{h_{pi}h_{qi}^{*}\right\} \approx J_0\left(\frac{2\pi}{\lambda}(p-q)d\right)$, for $\frac{d}{R} \ll 1$. By using the $\mathbf{W}$ defined in (2.17), the definition $\tilde{\Delta} \overset{def}{=} \frac{2\pi d}{\lambda}\Delta\cos\phi$ and (A.1) implies

$$\mathbf{\Phi}_{\mathrm{T}} = \begin{bmatrix} 1 & e^{j\frac{2\pi}{\lambda}d\sin\phi}J_0(\tilde{\Delta}) & \cdots & e^{j\frac{2\pi}{\lambda}|M-1|d\sin\phi}J_0(|M-1|\tilde{\Delta}) \\ e^{-j\frac{2\pi}{\lambda}d\sin\phi}J_0(\tilde{\Delta}) & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{\lambda}|M-1|d\sin\phi}J_0(|M-1|\tilde{\Delta}) & \cdots & \cdots & 1 \end{bmatrix},$$

which can be further decomposed by using the $\mathbf{W}$ defined in (2.17)

$$\begin{aligned} \mathbf{\Phi}_{\mathrm{T}} &= \mathbf{W}\cdot\begin{bmatrix} 1 & J_0(\tilde{\Delta}) & \cdots & J_0(|M-1|\tilde{\Delta}) \\ J_0(\tilde{\Delta}) & 1 & \cdots & J_0((|M-1|-1)\tilde{\Delta}) \\ \vdots & \vdots & \ddots & \vdots \\ J_0(|M-1|\tilde{\Delta}) & J_0((|M-1|-1)\tilde{\Delta}) & \cdots & 1 \end{bmatrix}\cdot\mathbf{W}^{H} \\ &\overset{def}{=} \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}\mathbf{W}^{H} = \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}H}\mathbf{W}^{H}, \text{ such that } \mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} = \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}}. \end{aligned} \tag{A.2}$$

The correlation matrix at the receive site can also be decomposed as

$$\mathbf{\Phi}_{\mathrm{R}} = \begin{bmatrix} 1 & J_0(\tilde{d}) & \cdots & J_0(|M-1|\tilde{d}) \\ J_0(\tilde{d}) & 1 & \cdots & J_0((|M-1|-1)\tilde{d}) \\ \vdots & \vdots & \ddots & \vdots \\ J_0(|M-1|\tilde{d}) & J_0((|M-1|-1)\tilde{d}) & \cdots & 1 \end{bmatrix} \overset{def}{=} \bar{\mathbf{\Phi}}_{\mathrm{R}}^{\frac{1}{2}}\bar{\mathbf{\Phi}}_{\mathrm{R}}^{\frac{1}{2}\,H} \tag{A.3}$$

where $\tilde{d} \overset{def}{=} \frac{2\pi d}{\lambda}$. The above two equations immediately lead to (A.4) and (A.5). Here, the separable model (2.2) is equivalent to

$$\mathbf{H} = \bar{\mathbf{\Phi}}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_{w}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}\,T}\mathbf{W}, \tag{A.4}$$

and the canonical model (2.4) is equivalent to

$$\mathbf{H} = \bar{\mathbf{\Phi}}_{\mathrm{R}}^{\frac{1}{2}} \mathbf{H}_{\mathrm{ind}} \bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}\ T} \mathbf{W}, \tag{A.5}$$

where $\bar{\mathbf{\Phi}}_{\mathrm{T}}$ and $\bar{\mathbf{\Phi}}_{\mathrm{R}}$ denote the power correlation matrices at the transmit and the receive sites, respectively. Using $\mathbf{\Phi}_{\mathrm{T}}^{1/2} = \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{1/2}$ and following a procedure similar to (2.12)–(2.15), we obtain (2.16) of the main text.

The separable $\mathbf{W}$ can also be obtained directly from the physical model [56],[57]. In [56], the directional term $\exp\left(\frac{-j2\pi(i-j)d\sin\phi}{\lambda}\right)$ is also shown to be separable in the expression of spatial correlation (i.e., Eq. (9) in [56]), and has the similar form of (A.1). Note that Forenza *et al.* [58] have recently showed that, for a clustered MIMO channel with uniform linear or circular array, the cross-correlation coefficients also have a regression form similar to (A.1). Hence, if we assume a similar environment, we will obtain an analytical model of the same form as (2.16).

In the above single-directional model, the AoD from the transmitting antennas at the transmitter can be captured by a mean AoD. In contrast, the principle of maximum entropy [59] assumes i.i.d. uniformly distributed AoA angles over $[0, 2\pi]$ and leaves no mean arriving direction being modelled at the mobile side. It models the separate power azimuthal spectra (PAS) of AoA and AoD, with a common direction being described by the mean AoD at the base station [57].

Appendix B

Proof of Lemma 3.1

According to *Lemma* 5.1.3 of [18], the ith entry of the vector $\left[(\mathbf{1}_E \otimes \mathbf{A}) \odot (\mathbf{B}^T \otimes \mathbf{1}_N)\right] \mathbf{c}$ is identical to the (i,i)th diagonal entry of the square matrix $\left[(\mathbf{1}_E \otimes \mathbf{A})\text{diag}(\mathbf{c})(\mathbf{B} \otimes \mathbf{1}_N^T)\right]$, for $i = 1, 2, \ldots, NE$. Define $\tilde{\mathbf{A}} = [\tilde{a}_{m,n}] \stackrel{def}{=} (\mathbf{1}_E \otimes \mathbf{A})$ and $\tilde{\mathbf{B}} = [\tilde{b}_{m,n}] \stackrel{def}{=} (\mathbf{B}^T \otimes \mathbf{1}_N)$. Then, for $i = N(p-1)+q$, $p \in \{1, \ldots, E\}$ and $q \in \{1, \ldots, N\}$, we have

$$\left[(\mathbf{1}_E \otimes \mathbf{A})\text{diag}(\mathbf{c})(\mathbf{B} \otimes \mathbf{1}_N^T)\right]_{i,i} = \sum_{j=1}^{M} \tilde{a}_{i,j} c_j \tilde{b}_{i,j} = \sum_{j=1}^{M} a_{q,j} c_j b_{j,p} = \left[\text{vec}\left(\mathbf{A}\text{diag}(\mathbf{c})\mathbf{B}\right)\right]_i, \quad \text{(B.1)}$$

where $a_{q,j}$ is the (q,j)th entry of $\mathbf{A}$ and $b_{j,p}$ is the (j,p)th entry of $\mathbf{B}$. $[\mathbf{D}]_{i,i}$ denotes the (i,i)th entry of the matrix $\mathbf{D}$, while $[\mathbf{e}]_i$ denotes the ith entry of the vector $\mathbf{e}$. Hence we conclude that

$$\left[\left((\mathbf{1}_E \otimes \mathbf{A}) \odot (\mathbf{B}^T \otimes \mathbf{1}_N)\right) \mathbf{c}\right]_i = \left[\text{vec}\left(\mathbf{A}\text{diag}(\mathbf{c})\mathbf{B}\right)\right]_i, \quad \forall\, i = 1, \ldots, NE, \quad \text{(B.2)}$$

which proves the *Lemma* 3.1.

Appendix C

Proof of Theorem 4.1

Proof. Since $\mathbf{\Phi\Lambda}_L\mathbf{\Phi}$ is a diagonal matrix, its eigenvalues are the permutation of the diagonal elements $\{\lambda_i\phi_i^2\}$. Let $\{\lambda_{\pi,i}\phi_{\pi,i}^2\}$ be the permutation of $\{\lambda_i\phi_i^2\}$ such that $\lambda_{\pi,1}\phi_{\pi,1}^2 \geq \lambda_{\pi,2}\phi_{\pi,2}^2 \geq \ldots \geq \lambda_{\pi,L}\phi_{\pi,L}^2$ are in descending order. Due to the fact that $\mathbf{\Phi U}^H\mathbf{EU\Phi}$ is Hermitian, we have the following perturbation bound from Weyl theorem [18],

$$\lambda_{\pi,i}\phi_{\pi,i}^2 + \epsilon_L \leq \gamma_i. \tag{C.1}$$

Since the function $f(z) \overset{def}{=} \frac{1}{1+z}$ is monotonically decreasing for $z > -1$, we have

$$\sum_i f(a_i) \geq \sum_i f(b_i), \text{ if } -1 \leq a_i \leq b_i, \forall\ i. \tag{C.2}$$

Combining (4.27), (C.1) and (C.2), we complete the proof by recognizing that

$$\sum_{i=1}^{L} \frac{1}{1+\lambda_{\pi,i}\phi_{\pi,i}^2 + c} = \sum_{i=1}^{L} \frac{1}{1+\lambda_i\phi_i^2 + c}, \forall\ c \notin \{-1-\lambda_i\phi_i^2\}. \tag{C.3}$$

In order to keep $f(\cdot)$ being monotonically decreasing, we must have $\min_i(\lambda_i\phi_i^2) + \epsilon_L > -1$, which implicitly satisfies the constraint in (C.3). □

Appendix D

Proof of Theorem 4.2

Proof. Notice that $\mathbf{\Phi\Lambda\Phi}$ is a diagonal matrix, its descending ordered eigenvalues are exactly the permuted diagonal elements $\{\lambda_{\pi,i}\phi_{\pi,i}^2\}_{i=1}^L$. From theorem (9.G.1) of [39], we have

$$\begin{aligned}
& (\gamma_1, \gamma_2, \ldots, \gamma_L) \\
= \; & \left(\mu_1(\mathbf{\Phi U}^H\mathbf{R_H U\Phi}), \mu_2(\mathbf{\Phi U}^H\mathbf{R_H U\Phi}), \ldots, \mu_L(\mathbf{\Phi U}^H\mathbf{R_H U\Phi})\right) \\
\prec \; & \left(\mu_1(\mathbf{\Phi\Lambda\Phi}) + \mu_1(\mathbf{\Phi U}^H\mathbf{EU\Phi}), \ldots, \mu_L(\mathbf{\Phi\Lambda\Phi}) + \mu_L(\mathbf{\Phi U}^H\mathbf{EU\Phi})\right) \\
= \; & \left(\lambda_{\pi,1}\phi_{\pi,1}^2 + \epsilon_1, \lambda_{\pi,2}\phi_{\pi,2}^2 + \epsilon_2, \ldots, \lambda_{\pi,L}\phi_{\pi,L}^2 + \epsilon_L\right), \qquad \text{(D.1)}
\end{aligned}$$

where $\mu_i(\mathbf{G})$ is the ith largest eigenvalues of the matrix $[\mathbf{G}]_{L\times L}$ and $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_L$.

Let $h(\{z_i\}) = \sum_i f(z_i) = \sum_i \frac{1}{1+z_i}$, where $z_i \geq z_{i+1}$. Since $f'(z) = \frac{-1}{(1+z)^2}$ is negative, continuous and monotonically increasing for $z > -1$, $f(z)$ is convex for $z > -1$. From (3.C.1) of [39], $h(\{z_i\})$ is thus Schur-convex, hence, $h(\{x_i\}) \leq h(\{y_i\})$, if $\{x_i\} \prec \{y_i\}$.

Combining the above results with (D.1), we complete the proof by restricting the domain to $\min\left(\{\gamma_i\}, \{\lambda_{\pi,i}\phi_{\pi,i}^2 + \epsilon_i\}\right) = \lambda_{\pi,L}\phi_{\pi,L}^2 + \epsilon_L = \min_i\{\lambda_i\phi_i^2\} + \epsilon_L > -1$. □

Bibliography

[1] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. E. Mogensen, "A stochastic multiple-input multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. Vehicular Technology Conf.*, Boston, MA, Sep. 2000, pp. 893–897.

[2] J. P. Kermoal, L. Schumacher, and P. Mogensen, "MIMO channel characterisation," IST Project IST-2000-30148 I-METRA Deliverable D2, Oct. 2002. [Online]. Available: http://www.ist-imetra.org/index_deliverables.html

[3] V. Erceg, L. Schumacher, and P. Kyritsi, "TGn channel models," IEEE 802.11 TGn, Tech. Rep. 03/940r4, Jan. 2004.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[5] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[6] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[7] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[8] Y. H. Kho and D. P. Taylor, "MIMO channel estimation and tracking based on polynomial prediction with application to equalization," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1585–1595, May 2008.

[9] X. Wang and K. J. R. Liu, "Model-based channel estimation framework for MIMO multicarrier communication systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1050–1063, May 2005.

[10] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO wireless channels: capacity, optimal signaling, and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2058–2072, Jun. 2005.

[11] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[12] L. Y. X. Ma 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.

[13] M.-X. Chang and Y. T. Su, "Model-based channel estimation for OFDM signals in rayleigh fading," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 540–544, Apr. 2002.

[14] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[15] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[16] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[17] Y.-C. Chen and Y.-T. Su, "Model-based MIMO channel estimation in spatially correlated environments," in *IEEE PIMRC*, vol. 1, Sep. 2004, pp. 498–502.

[18] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*. Cambridge, U. K.: Cambridge Univ. Press, 1991.

[19] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of 'Kronecker' MIMO radio channel model," *Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.

[20] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Why does the Kronecker model result in misleading capacity estimates?" Nov. 2007. [Online]. Available: http://www.arxiv.org/abs/0808.0036

[21] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[22] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[23] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Canonical statistical modeling and capacity analysis of correlated MIMO fading channels," 2008. [Online]. Available: http://www.ifp.uiuc.edu/~vasanth

[24] A. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[25] L. Schumacher, J. P. Kermoal, F. Frederiksen, and K. I. Pedersen, "MIMO channel characterisation," IST Project IST-1999-11729 METRA Deliverable D2, Feb. 2001. [Online]. Available: http://www.ist-imetra.org/metra/index_deliverables.html

[26] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.

[27] F. B. Gross, "New approximation to J_0 and J_1 Bessel functions," *IEEE Trans. Antennas Propag.*, vol. 43, no. 8, pp. 904–907, Aug. 1995.

[28] L. S. Scharf, *Statistical Signal Processing.* Reading, MA: Addison-Wesley, 1991.

[29] A. D. R. McQuarrie and C.-L. Tsai, *Regression and time series model selection.* World Scientific, 1998.

[30] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computations.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[31] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Johns Hopkins, 1996.

[32] W. Zhang, X.-G. Xia, and P. C. Ching, "Optimal training and pilot pattern design for OFDM systems in Rayleigh fading," *IEEE Trans. Broadcast.*, vol. 52, no. 4, pp. 505–514, Dec. 2006.

[33] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.

[34] A. Swindlehurst and J. Yang, "Using least squares to improve blind signal copy performance," *IEEE Signal Process. Lett.*, vol. 1, pp. 80–82, May 1994.

[35] "Spatial channel model for multiple input multiple output (MIMO) simulations (Rel. 6)," 3GPP, Tech. Rep. TR 25.996, 2003. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25996.htm

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation/Detection Theory.* Englewood Cliffs, NJ: Prentice Hall, 1993, vol. 1,2.

[37] D. P. Palomar, A. P-Iserte, J. M. Cioffi, and M. A. Lagunas, *Convex Optimization Theory Applied to Joint Transmitter-Receiver Design in MIMO Channels.* New York: Wiley, 2005.

[38] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory.* New York: Academic Press, 1990.

[39] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press, 1979.

[40] R. D. Fierro, "Perturbation analysis for two-sided (or complete) orthogonal decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 2, pp. 384–400, Apr. 1996.

[41] P.-A. Wedin, "Perturbation bounds in connection with the singular value decomposition," *BIT Numerical Math.*, vol. 12, no. 1, pp. 99–111, Mar. 1972.

[42] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," vol. 26, no. 8, pp. 1341–1365, October 2008.

[43] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Trans. Veh. Technol.*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.

[44] C. Jotten, P. Baier, M. Meurer, T. Weber, and M. Haardt, "Efficient representation and feedback signaling of channel state information in frequency division duplexing MIMO systems," in *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, vol. 2, Oct. 2002, pp. 444–448 vol.2.

[45] D. Love, J. Heath, R.W., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[46] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[47] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[48] A. Barg and D. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2450–2454, Sep 2002.

[49] J. Roh and B. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a vq-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, March 2006.

[50] V. Raghavan, V. Veeravalli, and A. Sayeed, "Quantized multimode precoding in spatially correlated multiantenna channels," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6017–6030, Dec. 2008.

[51] V. Raghavan and A. Sayeed, "Impact of spatial correlation on statistical precoding in MIMO channels with linear receivers," in *Proc. Annual Allerton Conference on Communications, Control and Computing*, 2006.

[52] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[53] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice reduction," in *Communications, 2004 IEEE International Conference on*, vol. 2, June 2004, pp. 798–802 Vol.2.

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

[55] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, March 2003.

[56] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr 2002.

[57] H. Xu, D. Chizhik, H. Huang, and R. Valenzuela, "A generalized space-time multiple-input multiple-output (MIMO) channel model," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 966–975, May 2004.

[58] A. Forenza, D. J. Love, and J. R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1933, Jul. 2007.

[59] M. Debbah and R. R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667–1690, May 2005.