國 立 交 通 大 學

電信工程研究所

碩 士 論 文

大型天線陣列系統的秩值與複合通道估計
及閉迴路傳收機之設計

Rank Determination, Composite Channel Estimation and
Closed-loop Transceiver Design for Massive MIMO Systems

研 究 生：陳科夆

指導教授：蘇育德 教授

2013 年 7 月

大型天線陣列系統的秩值與複合通道估計及閉迴路傳收機之設計

Rank Determination, Composite Channel Estimation and Closed-loop Transceiver Design for Massive MIMO Systems

研 究 生：陳科峯　　Student：Ko-Feng Chen

指導教授：蘇育德　　Advisor：Dr. Yu T. Su

國 立 交 通 大 學

電信工程研究所

碩 士 論 文

A Thesis

Submitted to the Institute of Communications Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Communications Engineering

July 2013

Hsinchu, Taiwan

大型天線陣列系統的秩值與複合通道估計及閉迴路傳收機之設計

學生：陳科夆　　指導教授：蘇育德

國立交通大學

電信工程研究所碩士班

摘　　要

我們探討一個裝有大型天線陣列基地台(BS)來服務多個單天線用戶(UE)的多輸入多輸出(MIMO)分時多工系統。此系統藉由上行(UL)的領航信號(pilot)來估計包含大尺度衰減係數(LSFC)及小尺度衰減係數(SSFC)的通道狀態資訊(CSI)。雖然在MU-MIMO或是分散式MIMO系統之運作，LSFC資訊是不可或缺的，然而有關MIMO通道估計之研究往往被假設為已知或是被忽略。我們利用大型天線陣列之通道硬化(channel-hardening)並能同時收到大量空間樣本的特性，在不需SSFC資訊的前提下能有效的利用相對少量的領航符元精準地估計LSFC。

至於SSFC的估計，我們利用降秩(rank-reduced, RR)通道模型來完成。由於這種方法之降秩效應需選擇適當的基底並有準確的秩值估計，後者又需先知道通道的空間相關矩陣。針對這三項議題我們首先分析了最佳先設(predetermined)基底的選擇，證明兩種常用的基底之近優性(near-optimality)。接著我們探討秩值對SSFC估計的性能影響、設計一套秩值決定的演算法，最後並發展了估計空間相關矩陣的演算法。這些成果乃是以我們對SSFC估計法的均方誤差(mean squared error, MSE)性能的詳細分析為基礎。我們結合了大、小尺度衰減係數的估計並證明在接收信號之入射角度擴散(AS)不大時，還可利用適當的RR模型來一併估計平均的接收角度(AoA)。比起使用不含AoA資訊的模型之通道估計法，這種方法不但可降低MSE而且所估得的角度資訊可用來形成下傳鏈路的波束。

最後，我們推導了上行領航信號的設計方式，並且在含括前述的複合通道估計器後，提出了分別適用於分時多工和分頻多工模式的閉迴路傳收機(closed-loop transceiver)設計流程與細部演算法。由電腦實驗結果可以看出我們的複合通道估計器、秩值決定及空間相關矩陣估計等演算法在大型天線陣列系統中的優異表現。

Rank Determination, Composite Channel Estimation and Closed-loop Transceiver Design for Massive MIMO Systems

Student : Ko-Feng Chen Advisor : Yu T. Su

Institute of Communications Engineering

National Chiao Tung University

Abstract

We consider a multiuser (MU) multiple-input multiple-output (MIMO) time-division duplexing (TDD) system in which the base station (BS) is equipped with a large number of antennas for communicating with single-antenna mobile users. In such a system the BS has to estimate the channel state information (CSI) that includes large-scale fading coefficients (LSFCs) and small-scale fading coefficients (SSFCs) by uplink pilots. Although information about the former FCs are indispensable in a MU-MIMO or distributed MIMO system, they are usually ignored or assumed perfectly known when treating the MIMO CSI estimation problem. We take advantage of the large spatial samples of a massive MIMO BS to derive accurate LSFC estimates in the absence of SSFC information and with a training overhead no larger than that required by conventional LSFC estimators. With estimated LSFCs, SSFCs are then obtained using a rank-reduced (RR) channel model.

We analyze the mean squared error (MSE) performance of the proposed composite channel estimator and prove that the separable angle-of-arrival (AoA) information provided by the RR model is beneficial for enhancing the estimator's performance, especially when the mean angle spread of the uplink signal is not too large. To fully exploit

the RR channel model, we have to select a proper basis and determine associated rank (modeling order). We solve these two issues by developing a rank-determination algorithm based on two popular bases and verify the near-optimality through computer simulations. We discuss uplink pilot design and suggest closed-loop transceiver design flows for both TDD and FDD modes using the estimated AoA for downlink beamforming and the LSFC information for power allocation.

Some computer experiment results are provided to validate the efficiencies of the proposed estimators and the rank determination method.

誌　　謝

對於論文得以順利完成，首先感謝指導教授 蘇育德博士。老師的諄諄教誨使我對於通訊領域的研究有更深入的了解，也教導我們許多書上學不到的知識與人生道理，讓我受益良多。並感謝口試委員蘇賜麟教授、祁忠勇教授、趙啟超教授及林大衛教授給予的許多寶貴意見，以補足這份論文的缺失與不足之處。

也由衷感謝實驗室的劉彥成學長，在我研究上有問題時，能給予我建議以及討論，使我的研究能夠順利完成，從中學習到的經驗是很珍貴的。另外也感謝實驗的學長姐、同學、及學弟妹們，在這兩年內的互相支持與鼓勵。

最後，要感謝我的家人及朋友，他們總是在背後默默的關心與支持，使我有動力可以努力往前進，在此僅獻上此論文，以代表我最深的敬意。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A cellular mobile network in which each base station (BS) is equipped with an M-antenna array, is referred to as a large-scale multiple input, multiple output (MIMO) system or a massive MIMO system for short if $M \gg 1$ and $M \gg K$, where K is the number of active user antennas within its serving area. A massive MIMO system has the potentiality of achieving transmission rate much higher than those offered by current cellular systems with enhanced reliability and drastically improved power efficiency. It takes advantage of the so-called channel-hardening effect [1] which implies that the channel vectors seen by different users tend to be mutually orthogonal and frequency-independent [2]. As a result, linear receiver is almost optimal in the uplink and simple multiuser precoder are sufficient to guarantee satisfactory downlink performance. Although most investigation consider the co-located BS antenna array scenario [1], the use of a more general setting of massive distributed antennas has been suggested recently [4].

The Kronecker model [9], which assumes separable transmit and receive spatial statistics, is often used in the study of massive MIMO systems [16]. The spatial channel model (SCM) [7], which is adopted as the 3GPP standard, degenerates to the Kronecker model [8] when the number of subpaths approaches infinity. This model also implies that the distributions of angle of arrival (AoA) and angle of departure (AoD) are independent. In general such an assumption is valid if the antenna number is small and large cellular system is in question. But if one side of a MIMO link consists of multi-

ple single-antenna terminals, only the spatial correlation of the array side needs to be taken into account and thus the reduced Kronecker model and other spatial correlated channel models become equivalent. Throughout this paper our investigation focuses on this practical scenario, i.e., we consider a massive MIMO system where K is equal to the number of active mobile users.

We assume that the mobile users transmit orthogonal uplink pilots for the serving BS to estimate CSI that includes both small-scale fading coefficients (SSFCs) and large-scale fading coefficients (LSFCs). Besides data detection, CSI is needed for a variety of link adaptation applications such as precoder, modulation and coding scheme selection. The LSFCs, which summarize the pathloss and shadowing effect, are proportional to the average received signal strength (RSS) and are useful in power control, location estimation, hand-over protocol and other applications. While most existing works focus on the estimation of the channel matrix which ignores the LSFC [18] [29], it is desirable to know SSFCs and LSFCs separately. LSFCs are long-term statistics whose estimation is often more time-consuming than SSFCs estimation. Conventional MIMO CSI estimators usually assume perfect LSFC information and deal solely with SSFCs [4–6]. For co-located MIMO systems, it is reasonable to assume that the corresponding LSFCs remain constant across all spatial subchannels and the SSFC estimation can sometime be obtained without the LSFC information. Such an assumption is no longer valid in a multi-user MIMO system where the user-BS distances spread over a large range and SSFCs cannot be derived without the knowledge of LSFCs.

The estimation of LSFC has been largely neglected, assuming somehow perfectly known prior to SSFC estimation. When one needs to obtain a joint LSFC and SSFC estimate, the minimum mean square error (MMSE) or least squares (LS) criterion is not directly applicable. The expectation-maximization (EM) approach is a feasible alternate [32, Ch. 7] but it requires high computational complexity and cannot guarantee convergence. We propose an efficient algorithm for estimating LSFCs with no aid of

SSFCs by taking advantage of the channel hardening effect and large spatial samples available to a massive MIMO BS. Our LSFC estimator is of low computational complexity, requires relatively small training overhead and yields performance far superior to that of an EM-based estimator. Our analysis show that it is unbiased and asymptotically optimal.

Estimation of SSFCs, on the other hand, is more difficult as the associated spatial correlation is not as high as that among LSFCs. Nevertheless, given an accurate LSFC estimator, we manage to derive a reliable SSFC estimator which exploits the spatial correlation induced channel rank reduction and calls for estimation of much less channel parameters than that required by conventional method [29] when the angle spread (AS) of the uplink signals is small. The proposed SSFC estimator provides excellent performance and offer additional information about the average angle of arrival (AoA) which is very useful in designing a downlink precoder.

In this thesis we present a method to obtain estimates for both LS- and SSFCs. We first propose a uplink-pilot-based LSFC estimator for a massive MU-MIMO TDD system that requires a small training overhead. Based on the facts that i) the user channels tend to be mutually orthogonal and ii) the resolution of massive MIMO antenna array is high, thus the AoA spread (AS) at the BS from each uplink channel is small, we use the estimated LSFCs obtained in the first step to derive an efficient estimator incorporating the SSFC and mean AoA estimation through a rank-reduced (RR) channel model similar to that proposed in [12]. When considering SSFC estimator with RR channel modeling, [12] suggest the use of polynomial basis. Nevertheless, by connecting the basis selection issue here with that of image signal processing, we show that the Karhunen-Loève transformation (KLT) basis is optimal in terms of its outstanding energy compaction property; furthermore, type-2 discrete cosine transform (DCT-II) basis is the best approximation of KLT basis whereas has low computational complexity. Simulation results are conducted to show the superiority of our estimators in massive MU-MIMO system. Finally,

an low complexity iterative modeling order decision algorithm is investigated.

The rest of this thesis is organized as follows. In Chapter 2, we describe a massive MU-MIMO channel model that takes into account spatial correlations and large-scale fading. In Chapter 3, a novel uplink-pilot-based LSFC estimator is proposed and in Chapter 4, we devise an SSFC estimator by using the estimated LSFCs. After that, we analysis the effect of modeling order on SSFC estimation, and based on the analysis results, we propose a rank determination algorithm. Finally, we introduce the closed-loop transceiver design including the uplink pilot design issue for both TDD and FDD mode in Chapter 5. Our main contributions are summarized in Chapter 6.

The following notations are used throughout the thesis: upper case bold symbols denote matrices and lower case bold symbols denote vectors. $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ represent the transpose, conjugate transpose, and conjugate of the enclosed items, respectively. vec$(\cdot)$ is the operator that forms one tall vector by stacking columns of the enclosed matrix, whereas Diag$(\cdot)$ translate a vector into a diagonal matrix with the vector entries being the diagonal terms. While $\mathbb{E}\{\cdot\}$, $\|\cdot\|$, $\|\cdot\|_2$, and $\|\cdot\|_F$ denote the expectation, vector ℓ_2-norm, matrix spectral norm and Frobenius norm of the enclosed items, respectively, $\otimes$ and $\odot$ respectively denote the Kronecker and Hadamard product operator. Denote by $\mathbf{I}_L$, $\mathbf{1}_L$, and $\mathbf{0}_L$ respectively the $L \times L$ identity matrix and L-dimensional all-one and all-zero column vectors, whereas $\mathbf{1}_{L\times S}$, and $\mathbf{0}_{L\times S}$ are the matrix counterparts of the latter two. $\mathbf{e}_i$ and $\mathbf{E}_{ij}$ are all-zero vector and matrix except for their ith and (i,j)th element being 1, respectively.

Chapter 2

Preliminaries

2.1 Conventional MIMO System Model

Multiple input and multiple output antenna systems, or commonly referred as MIMO system, is a system with spatial separated antennas. Under suitable channel fading conditions, the MIMO channel provides an additional spatial dimension for communication and yields a degree-of-freedom gain [38]. These additional degree of freedom can be utilized by spatially multiplexing several data streams onto the MIMO channel, and increase the capacity.

2.1.1 Single-user MIMO System

For conventional single-user (SU) MIMO, the channel between transmitter and receiver at time k can be modeled as

$$\mathbf{H}(k,\tau) = \sum_{m=1}^{L} \mathbf{H}_m(k)\delta(t-\tau_m), \tag{2.1}$$

where L is the total number of paths between one antenna pair. τ_m is the delay of mth path, and δ is the Dirac delta function. This representation reduce to a $N_R \times N_T$ single-tap fading channel matrix as we consider a narrowband flat-fading MIMO channel with N_T transmit (Tx) antenna and N_R received (Rx) antennas. The $N_R \times 1$ received

signal vector is then given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.2}$$

where $\mathbf{x}$ is the data vector transmitted, and $\mathbf{H}$ is the Rayleigh-fading channel. The elements of $\mathbf{H}$ are independent and identically distributed (i.i.d.), zero-mean circularly symmetric complex Gaussian random variables. $\mathbf{n}$ is the additive white Gaussian noise (AWGN) vector, whose entries are of zero-mean and with σ_z^2 variance.

2.1.2 Multiuser MIMO System

Now we cast into a single-cell uplink (UL) multiuser MIMO (MU-MIMO) system having an M-antenna BS and K single-antenna mobile stations (MSs) as illustrated in Fig. 2.1. When referring to one of the K MSs, we focus on the "one-ring" model shown in Fig. 2.2 while the other channel models introduced in [12] are also considered. In such a MIMO setup, MS is surrounded by local scatterers and waveforms impending from the MS are richly scattered. On the other hand, BS is often unobstructed by local scatterers and has a mean angle of arrival (AoA) and small angle spread (AS) with respect to the transmitter. The clustered channel setup is typical in urban environments, and has been validated through field measurements [12].

The most apparent difference between SU-MIMO and MU-MIMO is that in addition to small-scale fading, the transmitted signals (from MSs) also suffer from different large-scale fading caused by pathloss and shadowing effect. The $M \times 1$ Rx signal vector can be written as

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\beta_k}\mathbf{h}_k x_k + \mathbf{n} = \mathbf{H}\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{x} + \mathbf{n} \tag{2.3}$$

where $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K] \in \mathbb{C}^{M\times K}$ and $\mathbf{D}_{\boldsymbol{\beta}} = \mathrm{Diag}(\boldsymbol{\beta})$ contain respectively the SSFCs and LSFCs that characterize the K uplink channels, and $\mathbf{n}$ is the noise vector whose entries are distributed according to $\mathcal{CN}(0,1)$. The vector $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_K]^T$ whose elements $\beta_k = \mathfrak{s}_k d_k^{-\alpha}$ describes the shadowing effect, parameterized by independent identically

distributed (i.i.d.) $\mathfrak{s}_k$'s with $10\log_{10}(\mathfrak{s}_k) \sim \mathcal{N}(0, \sigma_{\mathfrak{s}}^2)$, and the pathloss which depends on the distance between the BS and MS k, d_k, with $\alpha > 2$. We called $\mathbf{HD}_{\boldsymbol{\beta}}^{\frac{1}{2}}$ in (2.3) the *composite* fading channel matrix.

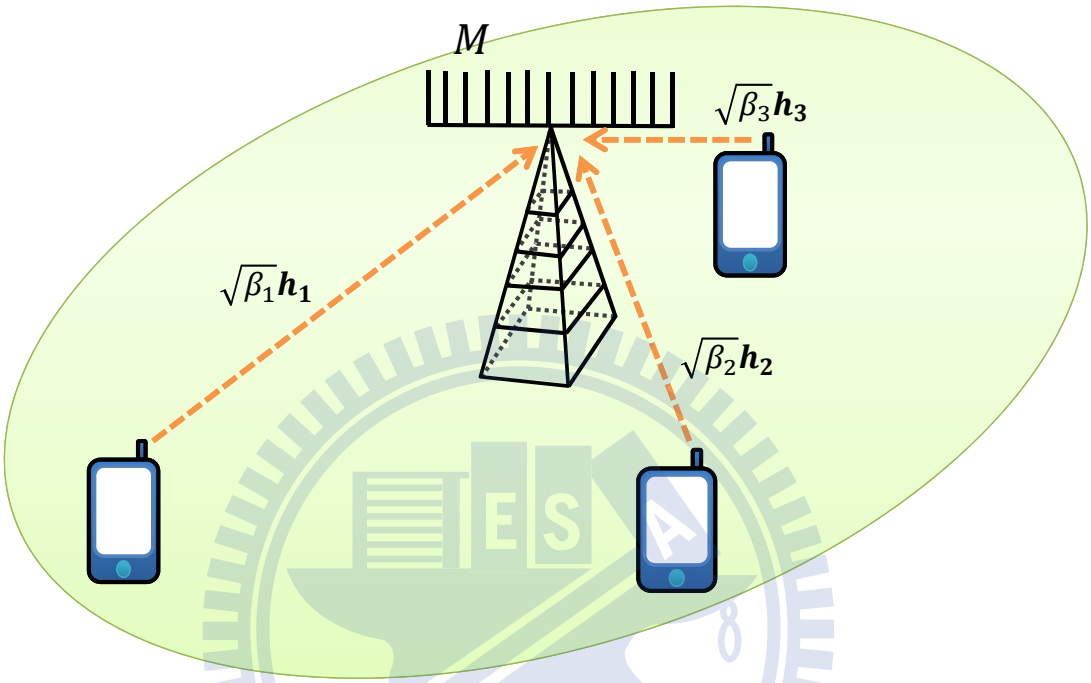

Figure 2.1: A co-located MU-MIMO system with an M-antenna BS and K single-antenna MSs.

2.2 Effect of Massive MIMO System

Massive MIMO, very-large MIMO, large-scale MIMO all refer to a system where the BS has an enormous number of antennas larger than the number of Rx antennas, that is, $N_T \gg N_R$ in SU-MIMO system and $M \gg K$ in MU-MIMO system. [1] and [16] say that M is about the magnitude of several hundreds but within a thousand.

Several observations on the effect of massive MIMO are given in the following sub-sections.

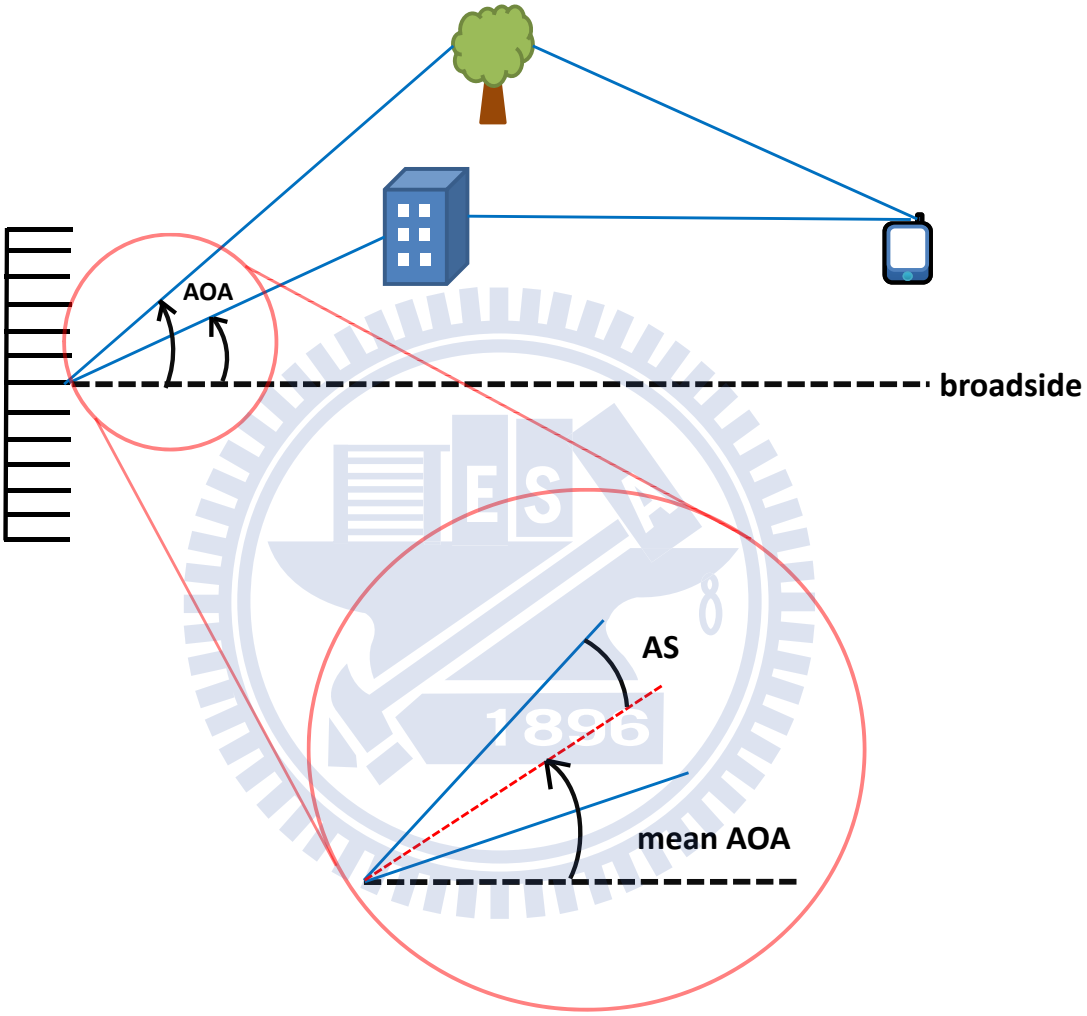

Figure 2.2: "One-ring" model with M transmit antennas and a single-antenna MS. The angle spread (AS) and mean angle of arrival (AoA) is depicted.

2.2.1 Channel Hardening Effect

Massive MIMO takes advantage of the channel hardening effect which says if the UL channel matrix representing the SSFCs between M transmit antennas and K single-antenna MSs, $M \gg K$, is $\mathbf{H} = [h_{ij}]$, whose entries are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with unit variance, $h_{ij} \sim \mathcal{CN}(0,1)$, then

$$\frac{1}{M}\mathbf{H}\mathbf{H}^H \xrightarrow{a.s.} \mathbf{I} \quad \text{and} \quad \sum_{\ell=1}^{K} \lambda_\ell^2 = ||\mathbf{H}||_F^2 \xrightarrow{a.s.} MK \quad \text{as } M \longrightarrow \infty$$

where λ_ℓ is the ℓth singular value of $\mathbf{H}$. This effect tell us that the channel gain is independent of frequency, thus, we can assign full bandwidth for all users, and no frequency domain scheduling is needed.

Moreover, the corresponding achievable sum rate is given by

$$\begin{aligned} R &= E\left\{\log_2 \det\left(\mathbf{I} + \frac{P}{M}\mathbf{H}\mathbf{H}^H\right)\right\} && (2.4)\\ &= \sum_{l=1}^{K} \log_2\left(1 + \frac{P}{M}\lambda_\ell^2\right) && (2.5)\\ &\xrightarrow{a.s.} K\log_2\left(1 + \frac{PM}{K}\right) && (2.6) \end{aligned}$$

which means ideally, thermal noise, interference all vanish asymptotically and to maintain a fixed sum rate, BS power can be scaled as $\frac{1}{M}$. It also implies that instead of reducing the cell size one can increase the system capacity by simply putting more antennas on the existing base stations [3]. Marzetta further showed that [3] the channel vectors seen by different users become mutually orthogonal and the simplest precoders–detectors, i.e., eigen beamforming (BF) and matched filter (MF), are asymptotically optimal. Furthermore, a simple regularized zero-forcing (RZF) precoding scheme can achieve the same performance as BF with one order of magnitude fewer antennas in both uncorrelated and correlated fading channels.

2.2.2 TDD and FDD system

The pilot sequences are needed for channel estimation and other synchronization purposes. Nevertheless, the time needed to transmit the DL pilot sequences to MSs equals to M, which may exceed the channel coherent interval, and result in an inaccurate estimation. Furthermore, as the CSI must be obtained through channel estimation, an FDD system would require the downlink receivers to estimate multitude of channels and feedback to the transmitter, costing immense overhead in bandwidth and power. A TDD system can largely bypasses such a need assuming reciprocity does hold and proper calibrations are readily available at the transmit side [19].

It is worth noting that the FDD option cannot become a serious contender unless we can greatly reduce the CSI requirement and pilot transmission time. In this thesis, we propose some approach to reduce the pilot dimension and feedback overhead simultaneously in FDD mode.

2.2.3 Large Antenna Aperture and Small Antenna Spacing

Owing to the large number of antennas, the beam resolution of massive antenna array is high [1], thus, high energy efficiency and power efficiency is achieved. More precisely, we have array power gain of M, hence the power per BS antenna can be scaled as $\frac{1}{M^2}$ when the total power is fixed [1]. Also, due to the high beam resolution, AoA spread at BS from each MS is small enough to helps us adopt the rank-reduced (RR) channel representation [12] of SSFCs; measurement results in [2] also suggest that the ASs from different users are small.

Due to the limited space that is available for BS installation, the implementation of a large amount of antennas on a BS forces them to be packed tightly (small antenna spacings). As a result, serious mutual coupling and spatial correlation may exist. Spatial correlation among BS antennas also enables the use of an RR model [12] that effectively decreases the number of SSFC parameters needed to be estimated.

2.3 Spatial-Correlated Small-Scale Fading Channel Models

2.3.1 Conventional Spatial-Correlated Channel Model

Consider a single-cell massive SU-MIMO system with an N_T-antenna BS and an N_R-antenna UE, where $N_T \gg N_R$. The signal received by the UE can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.7}$$

where $\mathbf{H} = [h_{ij}]$ is the $N_R \times N_T$ SSFC channel matrix with complex Gaussian entries, h_{ij}'s, $\mathbf{x}$ is the transmitted signal, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_R})$ represents the white noise.

Let $\mathbf{\Phi}$, $\mathbf{\Phi}_{\mathrm{T}}$, and $\mathbf{\Phi}_{\mathrm{R}}$ be the spatial correlation matrices of vec($\mathbf{H}$)

$$\mathbf{\Phi} \overset{def}{=} \mathbb{E}\left\{\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\right\}, \tag{2.8}$$

and those of the Tx and Rx antennas, respectively.

In general, a spatial-correlated Rayleigh fading MIMO channel can be modeled as

$$\mathrm{vec}(\mathbf{H}) = \mathbf{\Phi}^{\frac{1}{2}}\mathrm{vec}(\mathbf{H}_w), \tag{2.9}$$

where $\mathbf{H}_w$ is $N_R \times N_T$ with i.i.d., zero-mean, unit-variance complex Gaussian entries.

For a massive MU-MIMO channel where the receiving end consists of many single-antenna receivers at various locations, the above model must be modified in an user-by-user sense. The small-scale fading channel seen by user k can in general be written as

$$\mathbf{h}_k = \mathbf{\Phi}_k^{\frac{1}{2}}\tilde{\mathbf{h}}_k, \tag{2.10}$$

where $\mathbf{\Phi}_k$ is the transmit spatial correlation matrix with respect to the kth user and $\tilde{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$.

2.3.2 Kronecker Model

Kronecker model [9] is commonly used which assumes that the spatial correlations among Tx antennas and those among Rx antennas are separable so that

$$\mathbf{\Phi} = \mathbf{\Phi}_{\mathrm{T}} \otimes \mathbf{\Phi}_{\mathrm{R}} = \mathbf{\Phi}^{\frac{1}{2}} (\mathbf{\Phi}^{\frac{1}{2}})^{H}, \tag{2.11}$$

where the square root matrix $\mathbf{\Phi}^{\frac{1}{2}}$ has a similar decomposition

$$\mathbf{\Phi}^{\frac{1}{2}} = \mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \otimes \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}} \tag{2.12}$$

and therefore yields

$$\mathbf{H} = \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}} \mathbf{H}_{w} \left(\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \right)^{H}. \tag{2.13}$$

This model is reasonable accurate only when the main scattering is locally rich at the transmitter and receiver sides, respectively [**?**]. There are many field measurements and experiments that report inconsistencies with this model.

For a massive MU-MIMO channel where the receiving end consists of K single-antenna MSs, the Kronecker model is equivalent to conventional spatial-correlated channel model in 2.10.

2.3.3 Virtual Channel Representation

To solve the deficiencies of the Kronecker model, Sayeed [**?**] suggested a so-called virtual channel representation that takes the Tx-Rx cross-correlation into account. This model expands the spatial correlations by unitary matrices, $\mathbf{\Phi}_{\mathrm{T}}$, $\mathbf{\Phi}_{\mathrm{R}}$, relate the Tx and Rx spatial modes by a coupling matrix. Using the Fourier basis, one obtains

$$\mathbf{H} = \mathbf{F}_{\mathrm{R}} (\mathbf{\Omega} \odot \mathbf{H}_{w}) \mathbf{F}_{\mathrm{T}}^{H}, \tag{2.14}$$

where $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{F}_{\mathrm{T}}$ are DFT matrices of dimension N_R and N_T and $\mathbf{\Omega}$ is the coupling matrix. However, this model is only applicable for single-polarized ULAs and the Fourier basis is far from the optimal choice.

2.3.4 Weichselberger Model

An obvious optimal choice of basis that incurs no approximation error can be readily obtained by performing eigen-decomposition on the spatial correlation matrices. Weichselberger *et al.* therefore suggested the model [**?**]

$$\mathbf{H} = \mathbf{U}_{\mathrm{R}} \left(\tilde{\mathbf{\Omega}} \odot \mathbf{H}_w \right) \mathbf{U}_{\mathrm{T}}^H, \tag{2.15}$$

where $\mathbf{U}_{\mathrm{T}}$ and $\mathbf{U}_{\mathrm{R}}$ are the eigenbases of $\mathbf{\Phi}_{\mathrm{T}}$, and $\mathbf{\Phi}_{\mathrm{R}}$, respectively, i.e.,

$$\mathbf{\Phi}_{\mathrm{R}} = \mathbf{U}_{\mathrm{R}} \mathbf{\Lambda}_{\mathrm{R}} \mathbf{U}_{\mathrm{R}}^H, \quad \mathbf{\Phi}_{\mathrm{T}} = \mathbf{U}_{\mathrm{T}} \mathbf{\Lambda}_{\mathrm{T}} \mathbf{U}_{\mathrm{T}}^H. \tag{2.16}$$

with $\mathbf{\Lambda}_{\mathrm{R}}$ and $\mathbf{\Lambda}_{\mathrm{T}}$ being the diagonal matrices consist of the (nonnegative) eigenvalues of $\mathbf{\Phi}_{\mathrm{R}}$ and $\mathbf{\Phi}_{\mathrm{T}}$. $\tilde{\mathbf{\Omega}}$ is the element-wise square root of the coupling matrix $\mathbf{\Omega} = [\omega_{ij}]$ in which each entry specifies the average energy coupled from a transmit eigenmode to a receiver eigenmode [**?**].

Let $\mathbf{u}_{\mathrm{R},i}$ and $\mathbf{u}_{\mathrm{T},j}$ be the ith and jth column vector of $\mathbf{U}_{\mathrm{R}}$ and $\mathbf{U}_{\mathrm{T}}$, respectively. Then

$$\omega_{ij} = \mathbb{E}\left\{ \left| \mathbf{u}_{\mathrm{R},i}^H \mathbf{H} \mathbf{u}_{\mathrm{T},j} \right|^2 \right\}, \; i = 1, \cdots, N_R, \; j = 1, \cdots, N_T \tag{2.17}$$

The Weichselberger model is perhaps more convenient to generate the SSFC matrix $\mathbf{H}$ 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 $\mathbf{H}$.

2.3.5 Rank-reduced Channel Representation

The rank-reduced (RR) model introduced in [12] is reviewed in this section. Singular value decomposition (SVD) of $\mathbf{H}$ gives

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H, \tag{2.18}$$

Let $\mathbf{Q}_1$ and $\mathbf{Q}_2$ be two *predefined* unitary matrices. Then

$$\mathbf{Q}_1 = \mathbf{U}\mathbf{P}_1, \quad \mathbf{Q}_2 = \mathbf{V}\mathbf{P}_2, \tag{2.19}$$

where $\mathbf{\Lambda}$ is a $N_R \times N_T$ rectangular diagonal matrix with nonnegative entries and both $\mathbf{P}_1$ and $\mathbf{P}_2$ are also unitary. This unitary transform leads to

$$\mathbf{H} = \mathbf{Q}_1\mathbf{P}_1^{-1}\mathbf{\Lambda}(\mathbf{P}_2^{-1})^H\mathbf{Q}_2^H = \mathbf{Q}_1\mathbf{C}\mathbf{Q}_2^H, \tag{2.20}$$

with a random matrix $\mathbf{C}$ characterizing the cross-coupling effect. It can be verified that all the aforementioned models are special cases of (2.20) which is valid for all slow-varying narrowband MIMO channels without any attached pre-assumptions needed.

It is equivalent to the Kronecker model if $\mathbf{C}$ satisfies

$$\mathrm{vec}(\mathbf{C}) = (\mathbf{\Xi}_{\mathrm{T}} \otimes \mathbf{\Xi}_{\mathrm{R}})\mathrm{vec}(\mathbf{H}_\omega), \tag{2.21}$$

where $\mathbf{\Xi}_{\mathrm{T}}$ and $\mathbf{\Xi}_{\mathrm{R}}$ are obtained via Gram-Schmidt orthonormalization with $\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} = \mathbf{Q}_2\mathbf{\Xi}_{\mathrm{T}}$ and $\mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}} = \mathbf{Q}_1\mathbf{\Xi}_{\mathrm{R}}$.

If $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are chosen to be composed of columns of DFT matrices, this generalized model is compatible with the virtual channel representation [**?**]. Finally, the RR model is related to the Weichselberger model via

$$\mathbf{U}_{\mathrm{T}} = \mathbf{Q}_2\mathbf{P}_{\mathrm{T}}^H, \quad \mathbf{U}_{\mathrm{R}} = \mathbf{Q}_1\mathbf{P}_{\mathrm{R}}^H \tag{2.22}$$

with $\mathbf{P}_{\mathrm{T}}$ and $\mathbf{P}_{\mathrm{R}}$ being the eigenbasis matrices of $\mathbb{E}\{\mathbf{C}\mathbf{C}^H\}$ and $\mathbb{E}\{\mathbf{C}^T\mathbf{C}^*\}$ whose eigenvalues are the same as those of $\mathbb{E}\{\mathbf{H}\mathbf{H}^H\}$ and $\mathbb{E}\{\mathbf{H}^T\mathbf{H}^*\}$.

The use of pre-determined basis matrices avoids the need for the channel estimator to perform the extra work of determining the eigenbasis (matrices) and provides a convenient way for RR representation. Furthermore, when the AoD spread is small, we can show that (2.20) can be expressed as

$$\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^H\mathbf{W}, \tag{2.23}$$

where $\mathbf{W} = \text{Diag}[w_1, w_2, \ldots, w_M]$ and $w_i = \exp(-j2\pi\frac{(i-1)\xi}{\lambda}sin\phi)$, ξ being the distance between neighboring elements at the BS linear array, and ϕ is the mean AoD.

As UL MU-MIMO introduced in Section 2.1.2 being considered, we can summarize the SSFC matrix (2.20) and (2.23) into the lemma below:

Lemma 2.3.1 (RR representations)**.** *The channel vector seen by kth user can be represented by*

$$\mathbf{h} = \mathbf{Q}_m^{(I)}\mathbf{c}^{(I)} \tag{2.24}$$

or alternately by

$$\mathbf{h} = \mathbf{W}(\phi)\mathbf{Q}_m^{(II)}\mathbf{c}^{(II)} \tag{2.25}$$

where $\mathbf{Q}_m^{(I)}, \mathbf{Q}_m^{(II)} \in \mathbb{R}^{M\times m}$ *are predetermined (unitary) basis matrices and* $\mathbf{c}^{(I)}, \mathbf{c}^{(II)} \in \mathbb{C}^{m\times 1}$ *are the channel vectors with respect to bases* $\mathbf{Q}_m^{(I)}$ *and* $\mathbf{Q}_m^{(II)}$ *for the user k-BS link and* $\mathbf{W}(\phi)$ *is diagonal with unit magnitude entries. The two equalities hold only if* $m = M$ *and become approximations if* $m < M$*. Furthermore, if the AS is small,* $[\mathbf{W}(\phi)]_{ii} = \exp\left(-j2\pi\frac{(i-1)\xi}{\lambda}\sin\phi\right)$ *with* ξ *and* λ *being the antenna spacing and signal wavelength and* ϕ *is the mean AoA.*

That is, *Lemma 2.3.1* suggests that the mean AoA (which is close to the incident angle of the strongest path) of each user is extractable if its AS is small. In addition, due to the large aperture massive MIMO antenna array has offered, good AoA resolution and thus accurate mean value extraction are guaranteed [1].

2.4 Antenna Selection

Antenna selection (ANS) is useful in lessening the RF and baseband implementation complexity. For conventional MIMO systems, the number of antennas is usually relative small and ANS is not a particular concern although proper ANS does save hardware cost

and power consumption. For massive, however, it becomes an important design consideration, as the associated number of RF chains reduction can be significant. Moreover, it not always a good policy to use more antennas if the number of antennas used is already large enough; employing more antennas may lead to degraded performance.

An outstanding low-complexity ANS scheme is especially important for massive MIMO systems as it can significantly reduce the hardware requirement of a BS without compromising performance. The solution, however, is challenging to say the least. This can be easily seen by simply counting the number of possible combinations $\binom{100}{50} \approx 10^{29}$ in selecting 50 out of 100 antennas that maximizes the system sum rate.

We will consider first the basic setting of an SU-MIMO system with large-scale BS antenna array and then extend the investigation to an MU-MIMO scenario. Furthermore, we focus the ANS study on the DL case, where a massive MIMO BS acts as the transmitter. The UL scenario can be similarly treated with some minor modifications.

2.4.1 ANS for Massive SU-MIMO systems

Consider a DL massive SU-MIMO system with N_T Tx antennas and N_R Rx antennas whose received vector is given by (2.38) and $N_T \gg N_R$.

In such a system, $\mathbf{H}$ is known by the receiver but unknown to the transmitter. The ergodic capacity is given by [21]

$$\begin{aligned} R(\mathbf{H}) &= \log_2 \det\left(\mathbf{I}_{N_T} + \frac{P}{N_T}\mathbf{H}^H\mathbf{H}\right) \\ &\approx \sum_{j=1}^{N_R} \log_2\left(1 + \frac{P}{N_T}\|\mathbf{h}_j\|^2\right). \end{aligned} \tag{2.26}$$

To facilitate subsequent discussion, we need the following definition [23].

Definition 2.4.1. *Let f be a function defined as $f : U \to \mathbb{R}^+$. Then f is called monotone if $f(S \cup \{a\}) - f(S) \geq 0$, $\forall a \in U$, $S \subseteq U$, $a \notin S$, and is called a sub-modular function if $f(S \cup \{a\}) - f(S) \geq f(T \cup \{a\}) - f(T)$, $\forall a \in U$, $a \notin T$ and $S \subseteq T \subseteq U$.*

A nesting property of $R(\mathbf{H})$ has been derived in [21] when one tries to select one more Rx antenna by a incremental capacity-based selection (CS) algorithm:

$$R(\mathbf{H}_{n+1}) = R(\mathbf{H}_n) + \log_2\left(1 + \frac{P}{M}\alpha_{J,n}\right) \quad (2.27)$$

$$\approx R(\mathbf{H}_n) + \log_2\left(1 + \frac{P}{N_T}\|\mathbf{h}_{J,n}\|^2\right) \quad (2.28)$$

where $\mathbf{H}_n$ is the channel matrix after selecting n Rx antennas and $\alpha_{j,n} = \mathbf{h}_j^H \left(\mathbf{I} + \frac{P}{N_T}\mathbf{H}_n^H\mathbf{H}_n\right)^{-1}\mathbf{h}_j$ with $\mathbf{h}_j$ being the channel seen by the jth user antenna. (2.28) implies that in an SU-MIMO system, the channel capacity is sub-modular over Rx antenna set $\{1, 2, \ldots, N_R\}$. Hence, the main design criterion of SU-MIMO Rx ANS is to reduce the hardware complexity which is dominated by the number of radio-frequency (RF) chains. Given the set of selected antennas, one selects, in each step, the Jth antenna that maximize $\|\mathbf{h}_{J,n}\|$ [21]

$$J = \arg\max_j \|\mathbf{h}_{j,n}\|, \quad (2.29)$$

which gives an important observation that the CS criterion is asymptotically equivalent to the norm-based selection (NS) criterion. Thus, low-complexity NS criterion will be good enough to retain the performance CS can achieve.

On the other hand, [22] and [23] reported that $\mathbf{H}_n$ is not sub-modular over the Tx antenna set $\{1, 2, \ldots, N_T\}$. This means, for a massive MIMO downlink the use of all Tx antennas may not offer better rate. In general, we have two Tx ANS selection criteria, i.e., capacity maximization and feedback overhead reduction for FDD mode.

2.4.2 ANS for Massive MU-MIMO systems

Owing to the results of [24], we know that when $M \gg K$, user selection is unnecessary. Hence, we focus more on the Tx ANS in DL massive MU-MIMO system provided that TDD mode is used.

For an MU-MIMO system with M BS antennas and K single-antenna MSs, $M \gg K$, the composite forward link (downlink) channel matrix consists of SSFCs and LSFCs [1]

is given by

$$\mathbf{G} = \mathbf{D}_{\beta}^{\frac{1}{2}}\mathbf{H}. \tag{2.30}$$

On the other hand, due to the *channel reciprocity* [19] in TDD mode, i.e., the forward link and reverse link channel are symmetric, the reverse link (uplink) composite channel matrix is simply the $M \times K$ matrix $\mathbf{G}^T$. Therefore, the DL system model is given by [42]

$$\mathbf{y} = \mathbf{G}\mathbf{s} + \mathbf{n} = \mathbf{G}\mathbf{W}\mathfrak{P}\mathbf{x} + \mathbf{n} \tag{2.31}$$

where $\mathbf{s}$ is the transmitted signal, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ the uncoded data, $\mathbf{W}$ the $M \times K$ precoding matrix, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ the received noise and $K \times K$ matrix $\mathfrak{P}$ diagonal with its kth entry being the square root of power allocated to user k. Thus, the achievable rate

$$R(\mathbf{H}) = \log_2 \det\left(\mathbf{I}_K + \mathbf{W}^H\mathbf{G}^H\mathbf{G}\mathbf{W}\mathfrak{P}^2\right), \tag{2.32}$$

subject to a total power constraint P

$$\mathrm{tr}\left(\mathbf{W}\mathfrak{P}^2\mathbf{W}^H\right) = \|\mathbf{W}\mathfrak{P}\|_F^2 \leq P, \tag{2.33}$$

can be obtained. It has been proved that, as opposed to SU-MIMO, MU-MIMO using linear precoding techniques, e.g., zero-forcing (ZF) and minimum mean square error (MMSE) precoding, has sub-modular property $R(\mathbf{H})$ over the Tx antenna set [42]. This is because in MU-MIMO, MSs cannot cooperate and no post-detection signal processing is allowed.

Assume decremental transmit antenna selection (TAS) algorithm [42] is used, the capacity loss when getting rid of one more Tx antenna, indexed α, can be computed by using (14) of [42]:

Lemma 2.4.2. *Let $\mathcal{S}$ and $\mathcal{S}'$ be two Tx antenna sets in MU-MIMO system and $\overline{\mathcal{S}} = \mathcal{S}' \setminus \mathcal{S} = r$, where $\mathcal{S} \subset \mathcal{S}' \subseteq \{1, \cdots, M\}$ and $|\overline{\mathcal{S}}| = 1$. Then, the difference of the sum*

rate (throughput loss) between these two sets is given by

$$R_D(r) \equiv R_D(\mathcal{S}') - R_D(\mathcal{S}) = K \log_2 \left(1 + \frac{\mathrm{SNR} \frac{\|\mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1}\|_2^2}{1 - \mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1} \mathbf{u}_r}}{[\mathrm{tr}(\mathbf{T}_{\mathcal{S}'})]^2 + \mathrm{tr}(\mathbf{T}_{\mathcal{S}'}) \left(\mathrm{SNR} + \frac{\|\mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1}\|_2^2}{1 - \mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1} \mathbf{u}_r} \right)} \right) \tag{2.34}$$

using a ZF precoder, where $\mathbf{T}_{\mathcal{S}'} = \left(\mathbf{G}_{\mathcal{S}'} \mathbf{G}_{\mathcal{S}'}^H \right)^{-1}$, $\mathbf{G}_{\mathcal{S}}$ *is the composite fading channel given the transmit antenna set* $\mathcal{S}$, $\mathbf{u}_\ell$ *is the channel vector seen by* ℓ*th Tx antenna,* $\mathrm{SNR} = \frac{P}{\sigma^2}$, P *is the total power constraint, and* $\sigma^2 = 1$ *is the noise power.*

Because $R_D(r) \geq 0$, we know that TAS in MU-MIMO system indeed has submodularity [23], hence the purpose of TAS is to reduce the hardware complexity and reduce the amount of feedback in FDD mode. Suppose that we want to select $M_F \gg K$ out of M Tx antennas, the capacity-based selection (CS) algorithm can be summarized in Algorithm 1.

Algorithm 1 Decremental CS algorithm

1: Let $\mathcal{S} = 1, 2, \cdots, M_\ell$;
2: **while** $|\mathcal{S}| > M_F$ **do**
3: $\quad \alpha = \arg\max_{r \in \mathcal{S}} R_D(\mathcal{S} - \{r\})$;
4: $\quad \mathcal{S} = \mathcal{S} - \{\alpha\}$;
5: **end while**
6: The resulting set $\mathcal{S}$ is the desired transmit antenna set.

From Lemma 4 of [42], we obtain

$$\begin{aligned} \alpha &= \arg\max_{r \in \mathcal{S}} R_D(\mathcal{S} \setminus \{r\}) \\ &= \arg\min_{r \in \mathcal{S}} R_D(r) \\ &= \arg\min_{r \in \mathcal{S}} \frac{\|\mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1}\|_2^2}{1 - \mathbf{u}_r^H (\mathbf{G}_{\mathcal{S}} \mathbf{G}_{\mathcal{S}}^H)^{-1} \mathbf{u}_r} \end{aligned} \tag{2.35}$$

Applying Lemma 3.3.1 to (2.35) and substituting $\mathbf{G}_{\mathcal{S}} = \mathbf{D}_{\boldsymbol{\beta}}^{1/2}\mathbf{H}_{\mathcal{S}}$ yields

$$\begin{aligned}
\alpha &= \arg\min_{r\in\mathcal{S}} \frac{\|\mathbf{u}_r^H(\mathbf{G}_{\mathcal{S}}\mathbf{G}_{\mathcal{S}}^H)^{-1}\|_2^2}{1-\mathbf{u}_r^H(\mathbf{G}_{\mathcal{S}}\mathbf{G}_{\mathcal{S}}^H)^{-1}\mathbf{u}_r} \\
&\overset{\text{a.s.}}{=} \arg\min_{r\in\mathcal{S}} \frac{\frac{1}{|\mathcal{S}|}\|\mathbf{u}_r^H\mathbf{D}_{\boldsymbol{\beta}}^{-1}\|_2^2}{1-\frac{1}{|\mathcal{S}|}\mathbf{u}_r^H\mathbf{D}_{\boldsymbol{\beta}}^{-1}\mathbf{u}_r} \\
&= \arg\min_{r\in\mathcal{S}} \frac{\|\mathbf{D}_{\boldsymbol{\beta}}^{-1}\mathbf{u}_r\|_2^2}{|\mathcal{S}|-\|\mathbf{D}_{\boldsymbol{\beta}}^{-\frac{1}{2}}\mathbf{u}_r\|_2^2}
\end{aligned} \tag{2.36}$$

We call (2.36) generalized norm-based selection (GNS), since the form of its selection metric is similar to norm-based selection (NS).

When the serving MSs are not far away from each other, or equivalently, $\mathbf{D}_{\boldsymbol{\beta}} \approx \beta\mathbf{I}_K$, which is in general the case because each RRH only serves the nearby users,

$$\begin{aligned}
\alpha &= \arg\min_{r\in\mathcal{S}} \frac{\|\mathbf{D}_{\boldsymbol{\beta}}^{-1}\mathbf{u}_r\|_2^2}{|\mathcal{S}|-\|\mathbf{D}_{\boldsymbol{\beta}}^{-\frac{1}{2}}\mathbf{u}_r\|_2^2} \\
&\approx \arg\min_{r\in\mathcal{S}} \frac{\|\mathbf{u}_r\|_2^2}{|\mathcal{S}|\beta^2-\beta\|\mathbf{u}_r\|_2^2} \\
&= \arg\min_{r\in\mathcal{S}} \|\mathbf{u}_r\|_2^2 \\
&= \arg\min_{r\in\mathcal{S}} \|\mathbf{u}_r\|_2
\end{aligned} \tag{2.37}$$

which is just the norm-based selection (NS).

Besides, [1] has shown that in massive MIMO system, ZF precoding is asymptotically equivalent to MF precoding, hence we can get the same result for the case of MF precoding.

2.5 System Model

Throughout this thesis, we consider a single-cell MU-MIMO system having an M-antenna BS and K single-antenna MSs, where $M \gg K$. For a muti-cell system, pilot contamination [13] may become a serious design concern in the worst case when the same pilot sequences (i.e., the same pilot symbols are placed at the same time-frequency locations) happen to be used simultaneously in several neighboring cells and are perfectly synchronized in both carrier and time. In practice, there are frequency, phase and

timing offsets between any pair of pilot signals and the number of orthogonal pilots is often sufficient to serve mobile users in neighboring cells. Moreover, neighboring cells may use the same pilot sequence but the pilot symbols are located in non-overlapping time-frequency units [15], hence a pilot sequence is more likely be interfered by uncorrelated asynchronous data sequences whose average effect is not as serious as the worst case and can be mitigated by proper inter-cell coordination, frequency planning and some interference suppression techniques [14]. We will, however, focus on the single-cell narrowband scenario throughout this thesis.

We assume a narrowband communication environment in which a transmitted signal suffers from both large- and small-scale fading. The K uplink packets place their pilot of length T at the same time-frequency locations so that, without loss of generality, the corresponding received samples, arranged in matrix form, $\mathbf{Y} = [y_{ij}]$ at the BS can be expressed as

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{\beta_k}\mathbf{h}_k\mathbf{p}_k^H + \mathbf{N} = \mathbf{H}\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P} + \mathbf{N} \tag{2.38}$$

where $\mathbf{N}$ is the noise matrix whose entries are distributed according to $\mathcal{CN}(0,1)$. The $K \times T$ matrix is defined by $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_K]^H$, where $T \geq K$ and $\mathbf{p}_k$ is the pilot sequence sent by MS k and $\mathbf{p}_j^H\mathbf{p}_k = 0$, $\forall\ j \neq k$. The optimality of using orthogonal pilots has been shown in [29].

We invoke the assumption that channels linking different users are independent as they are relatively far (with respect to the wavelength) apart and the kth uplink channel vector is

$$\mathbf{h}_k = \boldsymbol{\Phi}_k^{\frac{1}{2}}\tilde{\mathbf{h}}_k, \tag{2.39}$$

where $\boldsymbol{\Phi}_k$ is the transmit spatial correlation matrix with respect to the kth user and $\tilde{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. We assume that $\tilde{\mathbf{h}}_k$'s are i.i.d. and the SSFC $\mathbf{H}$ remains constant during a pilot sequence period, i.e., the channel's coherence time is greater than T, while the LSFC $\boldsymbol{\beta}$ varies much slower.

Chapter 3

Large-Scale Fading Coefficient Estimation

Unlike previous works on MIMO channel matrix estimation which either ignore LSFCs [18, 29] or assume perfect known LSFCs [4–6], we try to estimate $\mathbf{H}$ and $\mathbf{D}_{\boldsymbol{\beta}}$ jointly. We first introduce an efficient LSFC estimator without SSFCs information in this Chapter. We treat separately channels with and without spatial correlation at the BS side and show that both cases lead to same estimators when the BS is equipped with a large-scale linear antenna array.

3.1 Uncorrelated BS Antennas

It is known that if the BS antenna spacings are large enough, say greater than 5λ, where λ is the signal wavelength, spatial mode correlation can be neglected and thus $\boldsymbol{\Phi}_k = \mathbf{I}_M, \forall k$ [9]. A statistic based on the received sample matrix $\mathbf{Y}$ and is asymptotically independent of the SSFCs is derivable from the following property [31, Ch. 3]

Lemma 3.1.1. *Let $\mathbf{p}, \mathbf{q} \in \mathbb{C}^{M\times 1}$ be two independent M-dimensional random vectors whose elements are independent identically distributed (i.i.d.) according to $\mathcal{CN}(0,1)$. Then by the law of large number,*

$$\frac{1}{M}\mathbf{p}^H\mathbf{p} \xrightarrow{a.s.} 1 \quad \textit{and} \quad \frac{1}{M}\mathbf{p}^H\mathbf{q} \xrightarrow{a.s.} 0 \quad \textit{as} \quad M \to \infty.$$

For a massive MIMO system with $M \gg T \geq K$, we have, as $M \to \infty$, $\frac{1}{M}\mathbf{H}^H\mathbf{H} \xrightarrow{a.s.} \mathbf{I}_K$, $\frac{1}{M}\mathbf{N}^H\mathbf{N} \xrightarrow{a.s.} \mathbf{I}_T$, $\frac{1}{M}\mathbf{H}^H\mathbf{N} \xrightarrow{a.s.} \mathbf{0}_{K\times T}$, and thus

$$\begin{aligned}\frac{1}{M}\mathbf{Y}^H\mathbf{Y} - \mathbf{I}_T &= \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}\mathbf{P} + \frac{1}{M}\mathbf{N}^H\mathbf{N} - \mathbf{I}_T \\ &\quad + \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\left(\frac{1}{M}\mathbf{H}^H\mathbf{H} - \mathbf{I}_K\right)\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P} \\ &\quad + \frac{2}{M}\Re\left\{\mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{H}^H\mathbf{N}\right\} \\ &\xrightarrow{a.s.} \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}\mathbf{P} \end{aligned} \tag{3.1}$$

(3.1) indicates that the additive noise effect is reduced and the estimation of LSFCs can be decoupled from that of the SSFCs. Using the identity, $\mathrm{vec}(\mathbf{A}\cdot\mathrm{Diag}(\mathbf{c})\cdot\mathbf{F}) = \left((\mathbf{1}_S\otimes\mathbf{A})\odot(\mathbf{F}^T\otimes\mathbf{1}_T)\right)\mathbf{c}$ with $\mathbf{A}\in\mathbb{C}^{T\times K}$, $\mathbf{F}\in\mathbb{C}^{K\times S}$, and $\mathbf{c}\in\mathbb{C}^{K\times 1}$, we simplify [12](3.1) as

$$\mathrm{vec}\left(\frac{1}{M}\mathbf{Y}^H\mathbf{Y} - \mathbf{I}_T\right) \xrightarrow{a.s.} \left(\left(\mathbf{1}_T\otimes\mathbf{P}^H\right)\odot\left(\mathbf{P}^T\otimes\mathbf{1}_T\right)\right)\boldsymbol{\beta}$$

This equation suggests that we solve the following unconstrained convex problem

$$\min_{\boldsymbol{\beta}}\left\|\mathrm{vec}\left(\frac{1}{M}\mathbf{Y}^H\mathbf{Y} - \mathbf{I}_T\right) - \left(\left(\mathbf{1}_T\otimes\mathbf{P}^H\right)\odot\left(\mathbf{P}^T\otimes\mathbf{1}_T\right)\right)\boldsymbol{\beta}\right\|^2, \tag{3.2}$$

to obtain the LSFC estimate

$$\hat{\boldsymbol{\beta}} = \mathrm{Diag}\left(\|\mathbf{p}_1\|^{-4},\cdots,\|\mathbf{p}_K\|^{-4}\right)\cdot\left((\mathbf{1}_T^T\otimes\mathbf{P})\odot(\mathbf{P}^*\otimes\mathbf{1}_T^T)\right)\mathrm{vec}\left(\frac{1}{M}\mathbf{Y}^H\mathbf{Y} - \mathbf{I}_T\right). \tag{3.3}$$

This LSFC estimator is of low complexity as no matrix inversion is needed when orthogonal pilots are used and does not require any knowledge of SSFCs. Furthermore, the configuration of massive MIMO makes the estimator robust against noise, which is verified numerically later in Section 3.4.

3.2 Correlated BS Antennas

In practice, the spatial correlations are non-zero and $\mathbf{Y}$ is of the form

$$\mathbf{Y} = \tilde{\boldsymbol{\Phi}}\begin{bmatrix}\tilde{\mathbf{h}}_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \tilde{\mathbf{h}}_K\end{bmatrix}\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P} + \mathbf{N} \overset{def}{=} \tilde{\boldsymbol{\Phi}}\tilde{\mathbf{H}}\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P} + \mathbf{N}$$

where $\tilde{\mathbf{\Phi}} = [\mathbf{\Phi}_1^{\frac{1}{2}}, \cdots, \mathbf{\Phi}_K^{\frac{1}{2}}]$. Following [16, 17], we assume that

Assumption 1. *The spatial correlation at BS antennas seen by a user satisfies*

$$\limsup_{M\to\infty} \|\mathbf{\Phi}_k^{\frac{1}{2}}\|_2 < \infty, \quad \forall k;$$

or equivalently,

$$\limsup_{M\to\infty} \|\mathbf{\Phi}_k\|_2 < \infty, \quad \forall k.$$

Therefore, (3.1) becomes

$$\begin{aligned}\frac{1}{M}\mathbf{Y}^H\mathbf{Y} - \mathbf{I}_T &\xrightarrow{a.s.} \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}\mathbf{P} + \frac{2}{M}\Re\left\{\mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\tilde{\mathbf{H}}^H\tilde{\mathbf{\Phi}}^H\mathbf{N}\right\} \\ &\quad + \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\left(\frac{1}{M}\tilde{\mathbf{H}}^H\tilde{\mathbf{\Phi}}^H\tilde{\mathbf{\Phi}}\tilde{\mathbf{H}} - \mathbf{I}_K\right)\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P} \\ &\overset{def}{=} \mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}\mathbf{P} + \mathbf{N}'\end{aligned}$$

where $\mathbf{N}'$ is zero-mean with seemingly non-diminishing variance due to the spatial correlation. Nonetheless, we proved in Appendix A that

Theorem 3.2.1. *If* $\limsup_{M\to\infty} \sup_{1\le k\le K} \|\mathbf{\Phi}_k^{\frac{1}{2}}\|_2 < \infty$, *then*

$$\frac{1}{M}\tilde{\mathbf{H}}^H\tilde{\mathbf{\Phi}}^H\tilde{\mathbf{\Phi}}\tilde{\mathbf{H}} \xrightarrow{a.s.} \mathbf{I}_K, \tag{3.4}$$

$$\frac{1}{M}\tilde{\mathbf{H}}^H\tilde{\mathbf{\Phi}}^H\mathbf{N} \xrightarrow{a.s.} \mathbf{0}_{K\times T} \tag{3.5}$$

as $M \to \infty$.

This theorem implies that although the non-zero spatial correlation does cause the increase of variance of $\mathbf{N}'$, the channel hardening effect still exist and $\mathbf{N}'$ is asymptotically diminishing provided that *Assumption 1* holds. In this case, LS criterion also mandates the same estimator as (3.3). Several remarks are worth mentioning.

Remark 1. *If J consecutive coherence blocks in which the LSFCs remain constant are available, our estimator can be easily extended to*

$$\hat{\boldsymbol{\beta}} = \text{Diag}\left(\|\mathbf{p}_1\|^{-4}, \cdots, \|\mathbf{p}_K\|^{-4}\right)\left[\left(\mathbf{1}_T^T \otimes \mathbf{P}\right) \odot \left(\mathbf{P}^* \otimes \mathbf{1}_T^T\right)\right] \cdot \text{vec}\left(\frac{1}{MJ}\sum_{i=1}^{J}\mathbf{Y}_i^H\mathbf{Y}_i - \frac{1}{J}\mathbf{I}_T\right) \quad (3.6)$$

where $\mathbf{Y}_i$ is the ith received block. Moreover, the noise reduction effect becomes more evident as more received samples become available.

Remark 2. *The proposed LSFC estimators (3.3) and (3.6) render element-wise expressions as*

$$\hat{\beta}_k = \frac{\mathbf{p}_k^H\mathbf{Y}^H\mathbf{Y}\mathbf{p}_k - M\|\mathbf{p}_k\|^2}{M\|\mathbf{p}_k\|^4}, \quad \forall k, \quad (3.7)$$

$$\hat{\beta}_k = \frac{\sum_{i=1}^{J}\mathbf{p}_k^H\mathbf{Y}_i^H\mathbf{Y}_i\mathbf{p}_k - MJ\|\mathbf{p}_k\|^2}{MJ\|\mathbf{p}_k\|^4}, \quad \forall k. \quad (3.8)$$

Although these new expressions imply the same computational complexity, they are shown to be useful in Section 5.2 when designing uplink pilots.

Remark 3. *After the LSFC estimates are got, we can use the conventional least squares (LS) estimator presented in [29] to estimate the SSFCs:*

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{P}^H\text{Diag}\left(\frac{1}{\hat{\gamma}_1}, \cdots, \frac{1}{\hat{\gamma}_K}\right), \quad (3.9)$$

where $\hat{\gamma}_\ell = \sqrt{\hat{\beta}_\ell}\|\mathbf{p}_\ell\|^2$, or more desirable, adopt the SSFC estimator introduced later in Chapter 4.

3.3 Performance Analysis

As LSFC estimator (3.7) is unbiased because

$$\begin{aligned}\mathbb{E}\left\{\hat{\beta}_k\right\} &= \frac{\mathbf{p}_k^H(M\mathbf{P}^H\mathbf{D}_{\boldsymbol{\beta}}\mathbf{P} + M\mathbf{I}_K)\mathbf{p}_k - M\|\mathbf{p}_k\|^2}{M\|\mathbf{p}_k\|^4} \\ &= \frac{M\beta_k\|\mathbf{p}_k\|^4 + M\|\mathbf{p}_k\|^2 - M\|\mathbf{p}_k\|^2}{M\|\mathbf{p}_k\|^4} \\ &= \beta_k, \ \forall k, \end{aligned} \quad (3.10)$$

the mean squared error (MSE) of $\hat{\beta}_k$ is thus

$$\mathbb{E}\left\{\left|\hat{\beta}_k - \beta_k\right|^2\right\} \;=\; \mathrm{Var}\left\{\hat{\beta}_k\right\}. \tag{3.11}$$

Lemma 3.3.1 (*Lemma 4* of [16]). *Let* $\mathbf{A} \in \mathbb{C}^{M\times M}$ *and* $\mathbf{p}$ *and* $\mathbf{q}$ *be two vectors with i.i.d. elements drawn from* $\mathcal{CN}(0,1)$. *If* $\limsup_{M\to\infty} \|\mathbf{A}\|_2 < \infty$, *then*

$$\mathbf{p}^H\mathbf{A}\mathbf{p} \xrightarrow{a.s.} \mathrm{tr}(\mathbf{A}) \quad \textit{and} \quad \frac{1}{M}\mathbf{p}^H\mathbf{A}\mathbf{q} \xrightarrow{a.s.} 0 \quad \textit{as} \quad M \to \infty.$$

Remark 4. *Using [39, Lemma B.26], we can prove that the convergence rates in the aforementioned asymptotic formulae follow* $\mathcal{O}(\|\mathbf{A}\|_F/M)$. *More precisely,*

$$\mathbb{E}\left\{\left|\frac{\mathbf{p}^H\mathbf{A}\mathbf{p} - \mathrm{tr}(\mathbf{A})}{M}\right|\right\} \;=\; \mathcal{O}(\|\mathbf{A}\|_F/M) \tag{3.12}$$

$$\mathbb{E}\left\{\left|\frac{\mathbf{p}^H\mathbf{A}\mathbf{q}}{M}\right|\right\} \;=\; \mathcal{O}(\|\mathbf{A}\|_F/M). \tag{3.13}$$

By reformulating (3.7) as

$$\hat{\beta}_k = \beta_k + \underbrace{\frac{\mathbf{p}_k^H(\mathbf{N}^H\mathbf{N} - M\mathbf{I}_K)\mathbf{p}_k}{M\|\mathbf{p}_k\|^4}}_{r_1} + \underbrace{\frac{\beta_k(\mathbf{h}_k^H\mathbf{h}_k - M)}{M}}_{r_2} + \underbrace{\frac{\sqrt{\beta_k}\left(2\Re\left\{\mathbf{h}_k^H\mathbf{N}\mathbf{p}_k\right\}\right)}{M\|\mathbf{p}_k\|^2}}_{r_3},$$

and invoking *Assumption 1*, *Lemmas 3.1.1* and *3.3.1*, and the fact that $\mathbf{h}_k = \mathbf{\Phi}_k^{\frac{1}{2}}\tilde{\mathbf{h}}_k$, we conclude that $r_1, r_2, r_3 \xrightarrow{a.s.} 0$ as $M \to \infty$, and thus

$$\mathrm{Var}\left\{\hat{\beta}_k\right\} \;=\; \mathbb{E}\left\{|r_1 + r_2 + r_3|^2\right\} \xrightarrow{a.s.} 0. \tag{3.14}$$

As $\mathbf{p}_k, \mathbf{N}$ and $\mathbf{h}_k$ are uncorrelated, we have

$$\mathbb{E}\left\{|r_1 + r_2 + r_3|^2\right\} \approx \mathbb{E}\left\{|r_1|^2\right\} + \mathbb{E}\left\{|r_2|^2\right\} + \mathbb{E}\left\{|r_3|^2\right\} \tag{3.15}$$

Lemma 3.3.2. *When the pilot length is $T = K$, the MSE convergence rates for $\mathbb{E}\{|r_1|^2\}$, $\mathbb{E}\{|r_2|^2\}$ and $\mathbb{E}\{|r_3|^2\}$ follow $\mathcal{O}\left(\frac{\beta_k^2}{T\cdot \mathrm{SNR}_k^2}\frac{1}{M}\right)$, $\mathcal{O}\left(\beta_k^2\frac{\|\mathbf{\Phi}_k\|_F^2}{M^2}\right)$ and $\mathcal{O}\left(\frac{4\beta_k^2}{T\cdot \mathrm{SNR}_k}\frac{1}{M}\right)$, respectively, where $\mathrm{SNR}_k \overset{def}{=} \frac{\beta_k\|\mathbf{p}_k\|^2}{T}$. Thus the convergence rate of Var$\{\hat{\beta}_k\}$ is dominated by $\mathbb{E}\{|r_2|^2\}$.*

Proof.

$$\begin{aligned}
\mathbb{E}\left\{|r_1|^2\right\} &= \frac{1}{\|\mathbf{p}_k\|^8}\mathbb{E}\left\{\mathrm{tr}\left(\mathbf{p}_k\mathbf{p}_k^H\left(\frac{\mathbf{N}^H\mathbf{N}-M\mathbf{I}_K}{M}\right)\right)^2\right\} \\
&\leq \frac{1}{\|\mathbf{p}_k\|^8}\mathbb{E}\left\{\mathrm{tr}\left(\mathbf{p}_k\mathbf{p}_k^H\right)^2\mathrm{tr}\left(\frac{\mathbf{N}^H\mathbf{N}-M\mathbf{I}_K}{M}\right)^2\right\} \\
&= \frac{1}{\|\mathbf{p}_k\|^4}\mathbb{E}\left\{\left(\mathrm{tr}\left(\frac{\mathbf{N}^H\mathbf{N}-M\mathbf{I}_K}{M}\right)\right)^2\right\} \\
&= \frac{1}{\|\mathbf{p}_k\|^4}\mathbb{E}\left\{\left(\sum_{i=1}^{K}\frac{\mathbf{n}_i^H\mathbf{n}_i-M}{M}\right)^2\right\} \\
&= \frac{1}{\|\mathbf{p}_k\|^4}\sum_{i=1}^{K}\mathbb{E}\left\{\left(\frac{\mathbf{n}_i^H\mathbf{n}_i-M}{M}\right)^2\right\} \\
&= O\left(\frac{K}{\|\mathbf{p}_k\|^4}\frac{1}{M}\right) = O\left(\frac{\beta_k^2}{T\cdot \mathrm{SNR}_k^2}\frac{1}{M}\right) \qquad (3.16)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left\{|r_2|^2\right\} &= \beta_k^2\mathbb{E}\left\{\left|\frac{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}_k\tilde{\mathbf{h}}_k-M}{M}\right|^2\right\} \\
&= O\left(\beta_k^2\frac{\|\mathbf{\Phi}_k\|_F^2}{M^2}\right) \\
&\gg O\left(\beta_k^2\frac{M}{M^2}\right) = O\left(\beta_k^2\frac{1}{M}\right) \qquad (3.17)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left\{|r_3|^2\right\} &= \frac{4\beta_k}{\|\mathbf{p}_k\|^4}\mathbb{E}\left\{\left|\sum_{i=1}^{K}\frac{p_{ki}\Re\{\mathbf{h}_k^H\mathbf{n}_i\}}{M}\right|^2\right\} \\
&= \frac{4\beta_k}{\|\mathbf{p}_k\|^4}\sum_{i=1}^{K}p_{ki}^2\mathbb{E}\left\{\left|\frac{\Re\{\mathbf{h}_k^H\mathbf{n}_i\}}{M}\right|^2\right\} \\
&= \frac{4\beta_k}{\|\mathbf{p}_k\|^4}\sum_{i=1}^{K}p_{ki}^2\mathbb{E}\left\{\left|\frac{\Re\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}_k^{\frac{1}{2}}\mathbf{n}_i\}}{M}\right|^2\right\} \\
&= \frac{4\beta_k}{\|\mathbf{p}_k\|^2}O\left(\frac{\|\mathbf{\Phi}_k^{\frac{1}{2}}\|_F^2}{M^2}\right) = O\left(\frac{4\beta_k}{\|\mathbf{p}_k\|^2}\frac{1}{M}\right) \\
&= O\left(\frac{4\beta_k^2}{T\cdot\mathrm{SNR}_k}\frac{1}{M}\right), \qquad (3.18)
\end{aligned}
$$

□

Lemma 3.3.3. *The LSFC estimators (3.3) and (3.6) approach the minimum mean square error (MMSE) estimator with asymptotically diminishing MSE as $M \to \infty$.*

Remark 5 (Fisher Information Matrix)**.** *Denote $\boldsymbol{\theta}^T = \left[\sqrt{\beta}_k \;\; \mathbf{h}_k^T\right]$, the Fisher information matrix for estimating $\boldsymbol{\theta}$ is given by*

$$
\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix}
2\|\mathbf{p}_k\|^2\|\mathbf{h}_k\|^2 & \sqrt{\beta}_k\|\mathbf{p}_k\|^2 h_{k1} & \cdots & \sqrt{\beta}_k\|\mathbf{p}_k\|^2 h_{kM} \\
\sqrt{\beta}_k\|\mathbf{p}_k\|^2 h_{k1}^\star & \beta_k\|\mathbf{p}_k\|^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{\beta}_k\|\mathbf{p}_k\|^2 h_{kM}^\star & 0 & \cdots & \beta_k\|\mathbf{p}_k\|^2
\end{pmatrix}. \qquad (3.19)
$$

Moreover, the Cramér-Rao lower bound (CRLB) of estimating β_k is given by

$$
\mathrm{Var}\left\{\sqrt{\hat{\beta}}_k\right\} \geq \left[\mathbf{I}(\boldsymbol{\theta})\right]_{11}, \qquad (3.20)
$$

thus, (3.14) means that our LSFC estimator asymptotically achieves the CRLB.

Remark 6. *From Lemma 3.3.2, the only term related to spatial correlation in (3.15) is $\mathbb{E}\left\{|r_2|^2\right\}$, and thus, for cases with finite M, the MSE-minimizing spatial correlation matrix $\mathbf{\Phi}_k^\star$ is the solution of*

$$
\begin{aligned}
\min_{\mathbf{A}} \quad & \mathbb{E}\left\{\left|\tilde{\mathbf{h}}_k^H\mathbf{A}\tilde{\mathbf{h}}_k\right|^2\right\} - \mathrm{tr}(\mathbf{A}) \\
\text{s.t.} \quad & [\mathbf{A}]_{ii} = 1, \;\forall\; i. \qquad (3.21)
\end{aligned}
$$

Following the method of Lagrange multiplier, we obtain $\mathbf{\Phi}_k^\star = \mathbf{I}_M$. *The convexity of (3.21) implies that* $\mathrm{Var}\{\hat{\beta}_k\}$ *is an increasing function of* $\|\mathbf{\Phi}_k - \mathbf{I}_M\|_F$, *i.e., the variance of the LSFC estimator decreases as the channel becomes less correlated; meanwhile, Lemma 3.3.2 says the error convergence rate also improves.*

Remark 7 (Finite M scenarios)**.** *Low normalized MSE, in the order of* 10^{-5} *to* 10^{-4}, *is obtainable with not-so-large BS antenna numbers (e.g.,* 50*). The above MSE performance analysis is validated via simulation in Section 3.4.*

3.4 Numerical Results and Discussion

Simulation results reported here using the channel generated by [40] whose spatial correlation at the BS is related to AoA distribution and antenna spacings. In addition, the environment surrounding a user is of rich scattering with AoD's uniformly distributed in $[-\pi, \pi)$ making spatial correlation at MSs negligible. This setting accurate describes the environment where the BS with large-scale antenna array are mounted on an elevated tower or building. Throughout this section, we assume that there are 8 uniformly distributed users in a circular cell of radius R with the mean AoAs equispaced within $[-60°, 60°]$. The other simulation parameters are listed in Table 3.1 unless explicitly stated otherwise. We define average received signal-to-noise power ratio as $\mathrm{SNR} \overset{def}{=} \beta_k \|\mathbf{p}_k\|^2/T$ with k the index of farthest MS from the BS, and normalized mean squared error (NMSE) as the MSE between the real and estimated vectors normalized by the former's dimension and entry variance.

First, in Fig. 3.1 we compare the performance of the proposed LSFC estimator (3.3) with that of a conventional LS estimator [32, Ch. 8]

$$\hat{\boldsymbol{\beta}} = \left(\left[(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \mathrm{vec}(\mathbf{Y})\right]\right)^2 \tag{3.22}$$

where $\mathbf{A} = (\mathbf{1}_T \otimes \mathbf{H}) \odot \left(\mathbf{P}^T \otimes \mathbf{1}_M\right)$. As opposed to (3.3), the conventional estimator needs to know SSFCs beforehand, hence a full knowledge of SSFCs is assumed for the

Table 3.1: Simulation parameters

Parameters	*Values*
Operating frequency	2.6 GHz [2]
Cell radius R	100 meters
Pathloss exponent α	3
Shadow fading variance σ_s	10 dB
Number of BS antennas M	100
BS antenna spacing ξ	0.5λ
Number of MSs K	8

latter. Figure 3.1 shows that our proposed estimator outperforms the conventional significantly even if the latter has full knowledge of SSFCs. It is because the former takes advantage of the noise reduction effect that massive MIMO systems have offered. Moreover, as antenna spacing increases, the channel decorrelates and thus the estimation error due to spatial correlation decreases, which verifies *Theorem 3.2.1.* Figure 3.2 illustrates the effect of massive antennas to MSE. Owing to the fact that we have assumed perfect SSFC knowledge for the conventional LSFC estimator, MSE decreases with increasing sample amount as M increases. Unlike the conventional, the amount of known information does not grow with M for the proposed LSFC estimator. However, it still can be utilized to improve estimation and thus enables the proposed to outperform the conventional. In addition, the proposed is robust to SNR degradation due to noise or user-BS distance increment.

When comparing our proposed estimators with EM-based estimators, Fig. 3.3-3.7 shows the superiority of the proposed estimators. The details of EM approach is as follows: (i)set the initial value of $\widehat{\boldsymbol{\beta}}$;

(ii)evaluate the LMMSE estimator of SSFC,

$$\widehat{\mathrm{vec}(\mathbf{H})} = \mathrm{Diag}\left((\boldsymbol{\Phi}_1 + \|\mathbf{p}_1\|^2\widehat{\beta}_1\mathbf{I}_M)^{-1}\cdots(\boldsymbol{\Phi}_K + \|\mathbf{p}_K\|^2\widehat{\beta}_K\mathbf{I}_M)^{-1}\right)\left(\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}}\mathbf{P}^{\star}\otimes\mathbf{I}_M\right)\mathrm{vec}(\mathbf{Y});$$

(ii)evaluate the LMMSE estimator of LSFC,

$$\widehat{\sqrt{\boldsymbol{\beta}}} = \mathbb{E}\left\{\sqrt{\boldsymbol{\beta}}\right\} + \left(\mathbf{C}^{-1}_{\sqrt{\boldsymbol{\beta}}\sqrt{\boldsymbol{\beta}}} + \mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\left(\mathrm{vec}(\mathbf{Y}) - \mathbf{A}\mathbb{E}\left\{\sqrt{\boldsymbol{\beta}}\right\}\right)$$

where $\mathbf{A} = \left(\mathbf{1}_T \otimes \widehat{\mathbf{H}}\right) \odot \left(\mathbf{P}^T \otimes \mathbf{1}_M\right)$;

(iv)recursively compute until convergence occur.

Moreover, the modified EM (MEM) approach is obtained by replacing

$$\mathbf{A}^H\mathbf{A} = (\mathbf{H}^H\mathbf{H}) \odot (\mathbf{P}^\star\mathbf{P}^T) \xrightarrow{a.s.} \text{Diag}\left(M\|\mathbf{p}_1\|^2 \cdots M\|\mathbf{p}_K\|^2\right)$$

in the EM approach.

The results in Figs. 3.8 and 3.9 demonstrate the large-system performance of the proposed LSFC and full-order SSFC estimator. As can be seen, similar to the results in Fig. 3.2, the accurate LSFC estimates due to large received samples make its compensation prior to the SSFC estimation reliable. Furthermore, such large sample size clearly improves the performance of the SSFC estimators directly. In addition, the proposed LSFC estimator has noise-robustness in the sense that the MSE performance is nearly the same when SNR = 0 dB to 15 dB, which gives similar result as Fig. 3.2.

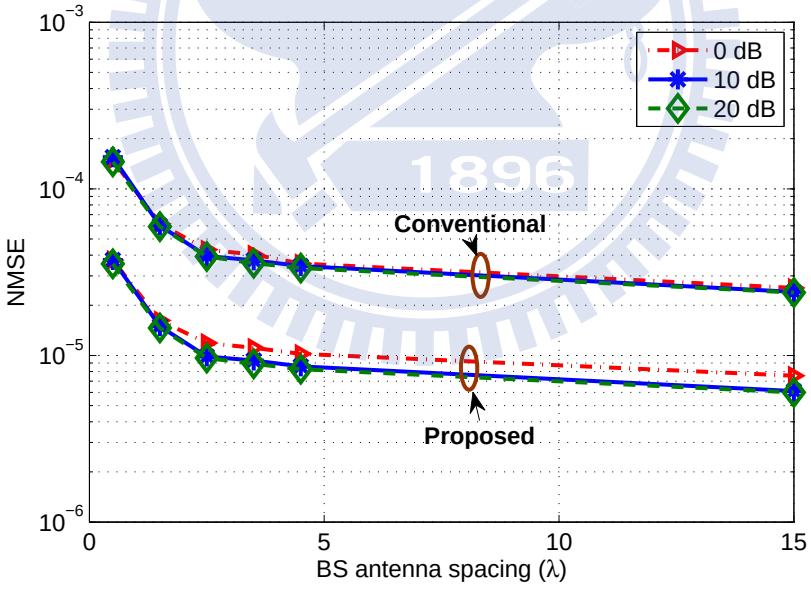

Figure 3.1: MSE performance of the conventional and proposed LSFC estimator with perfect SSFC knowledge assumed for the former, AS= 15°.

After discussing some results about the average MSE performance of MSs, we recast to compare the MSE performance of MSs at different locations, or equivalently, different

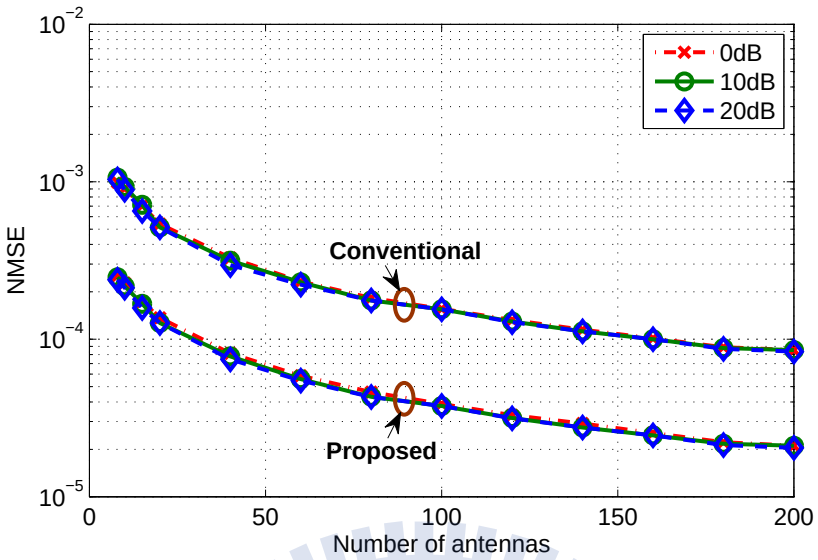

Figure 3.2: MSE performance of the conventional and proposed LSFC estimator with perfect SSFC knowledge assumed for the former, AS= 15°.

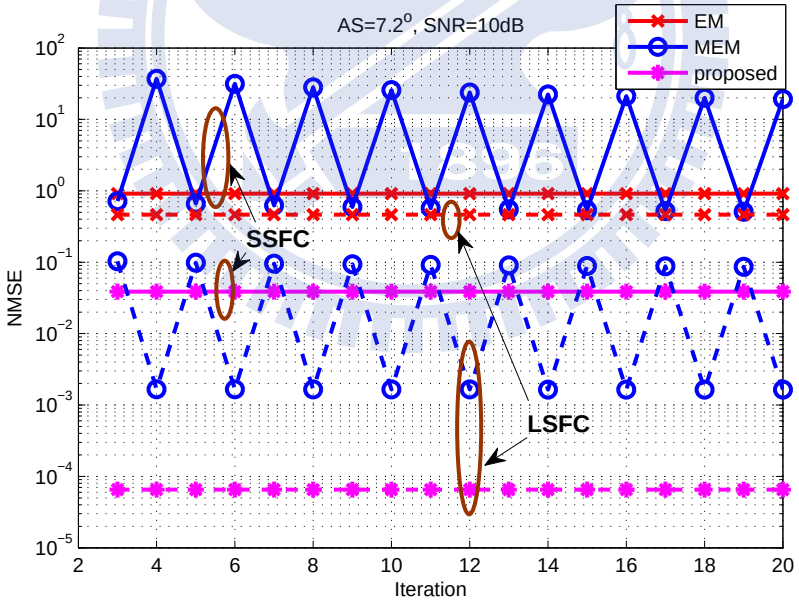

Figure 3.3: MSE performance comparison between the proposed estimators (LSFC, SSFC) and the EM-based estimators (LSFC, SSFC) versus iteration number of EM-based estimators, where AS= 7.2°, SNR=10dB, and full modeling order is used. Initial $\widehat{\boldsymbol{\beta}}$ is chosen as $\mathbb{E}\{\boldsymbol{\beta}\}$.

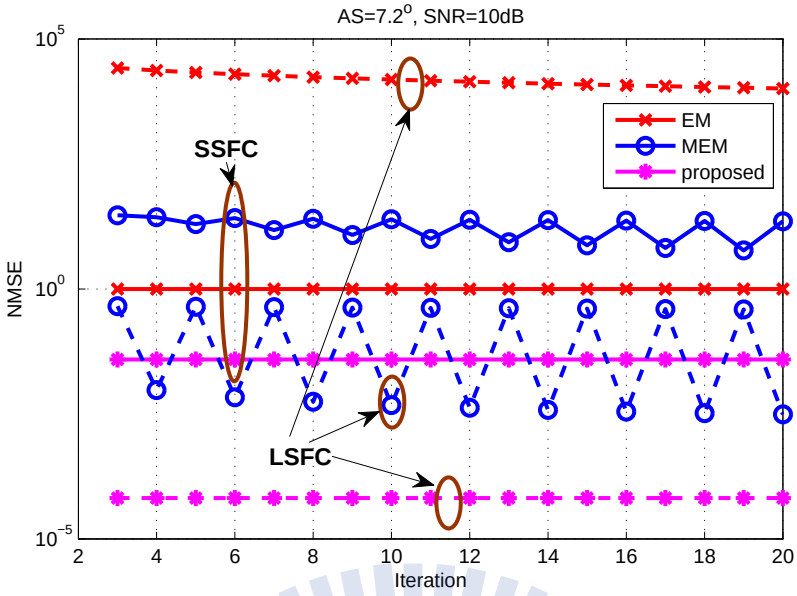

Figure 3.4: MSE performance comparison between the proposed estimators (LSFC, SSFC) and the EM-based estimators (LSFC, SSFC) versus iteration number of EM-based estimators, where AS= 7.2°, SNR=10dB, and full modeling order is used. Initial $\widehat{\boldsymbol{\beta}}$ is chosen as $\frac{1}{2}\mathbf{1}_K$.

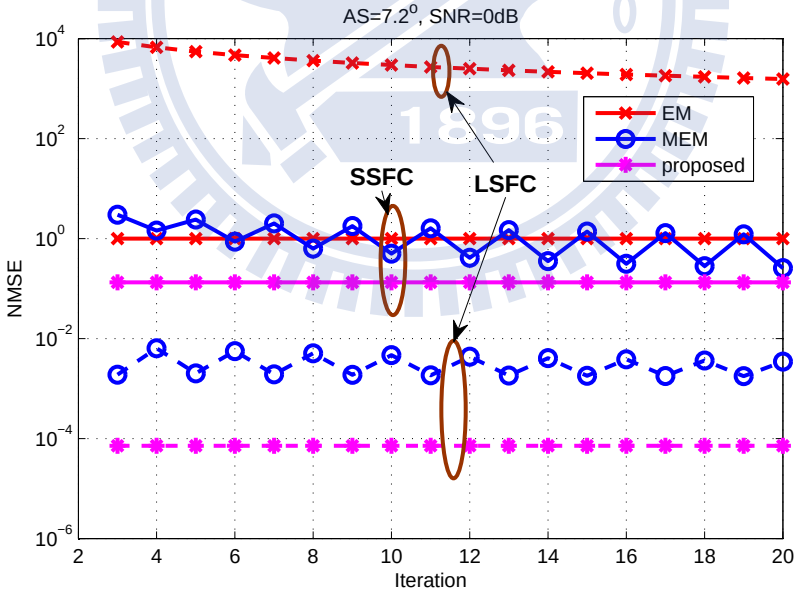

Figure 3.5: MSE performance comparison between the proposed estimators (LSFC, SSFC) and the EM-based estimators (LSFC, SSFC) versus iteration number of EM-based estimators, where AS= 7.2°, SNR=0dB, and full modeling order is used. Initial $\widehat{\boldsymbol{\beta}}$ is chosen as $\frac{1}{2}\mathbf{1}_K$.

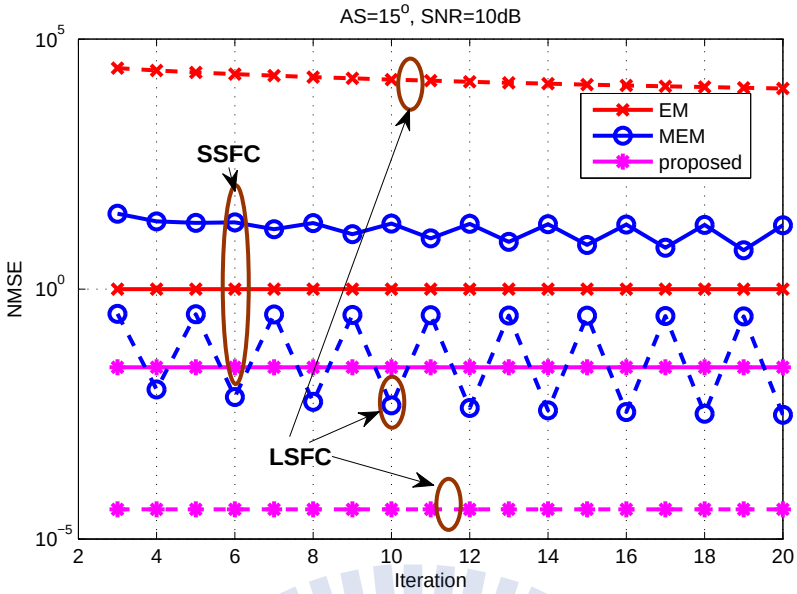

Figure 3.6: MSE performance comparison between the proposed estimators (LSFC, SSFC) and the EM-based estimators (LSFC, SSFC) versus iteration number of EM-based estimators, where AS= 15°, SNR=10dB, and full modeling order is used. Initial $\widehat{\boldsymbol{\beta}}$ is chosen as $\frac{1}{2}\mathbf{1}_K$.

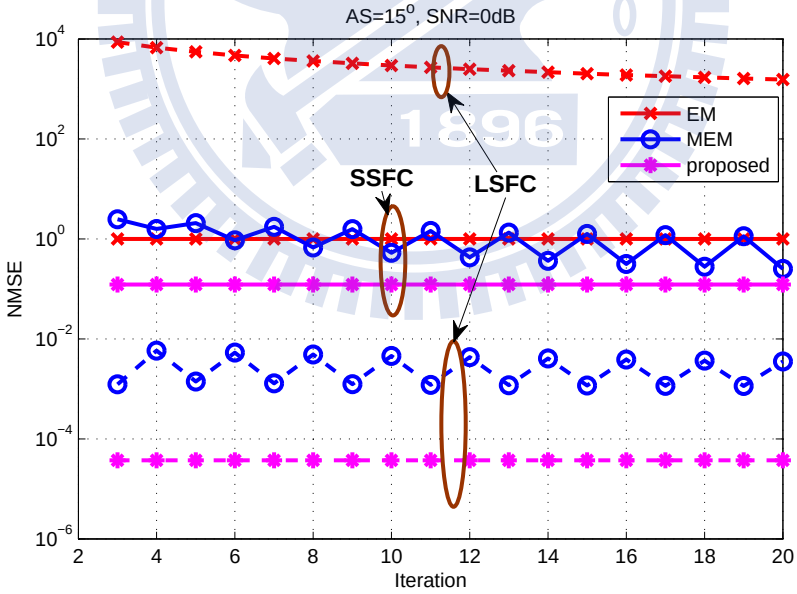

Figure 3.7: MSE performance comparison between the proposed estimators (LSFC, SSFC) and the EM-based estimators (LSFC, SSFC) versus iteration number of EM-based estimators, where AS= 15°, SNR=0dB, and full modeling order is used. Initial $\widehat{\boldsymbol{\beta}}$ is chosen as $\frac{1}{2}\mathbf{1}_K$.

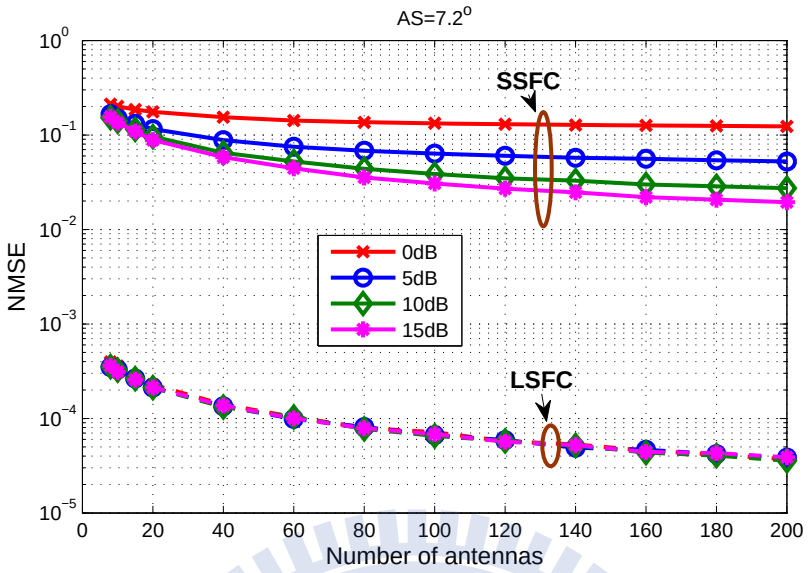

Figure 3.8: MSE performance of the proposed LSFC and SSFC estimator versus number of BS antennas and received SNR, where AS= 7.2°, and full modeling order is used.

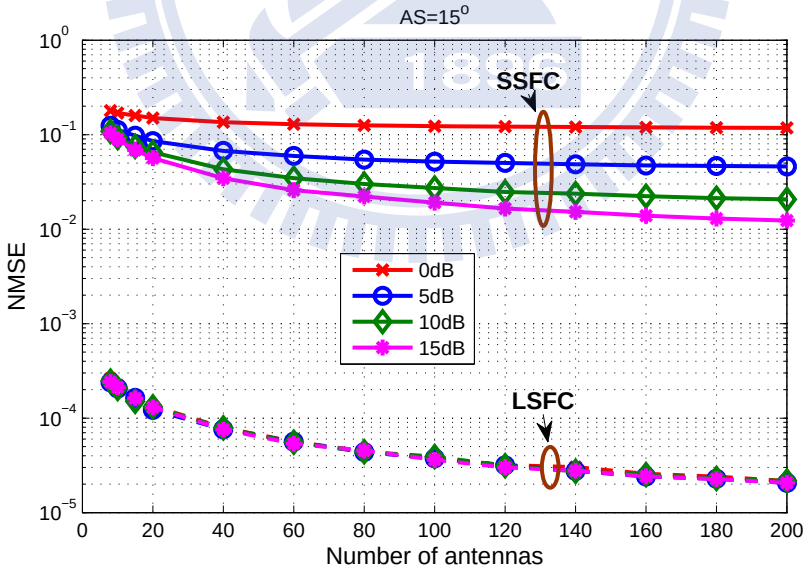

Figure 3.9: MSE performance of the proposed LSFC and SSFC estimator versus number of BS antennas and received SNR, where AS= 15°, and full modeling order is used.

LSFC. Consider Fig. 3.10, the red lines with $\times$ represents the nearest MS while the blue lines with $\circ$ represents the farthest MS from the BS. Assume the pilot power is such that the received SNR from farthest MS at BS equals to 10.22dB for left figure and 20.22dB for right figure. It can be seen clearly that the pilot power has little effect on the MSE performance of LSFC estimator due to the robustness of our estimator to noise. Nevertheless, when considering the full-order SSFC estimator, the nearest MS indeed has better MSE performance.

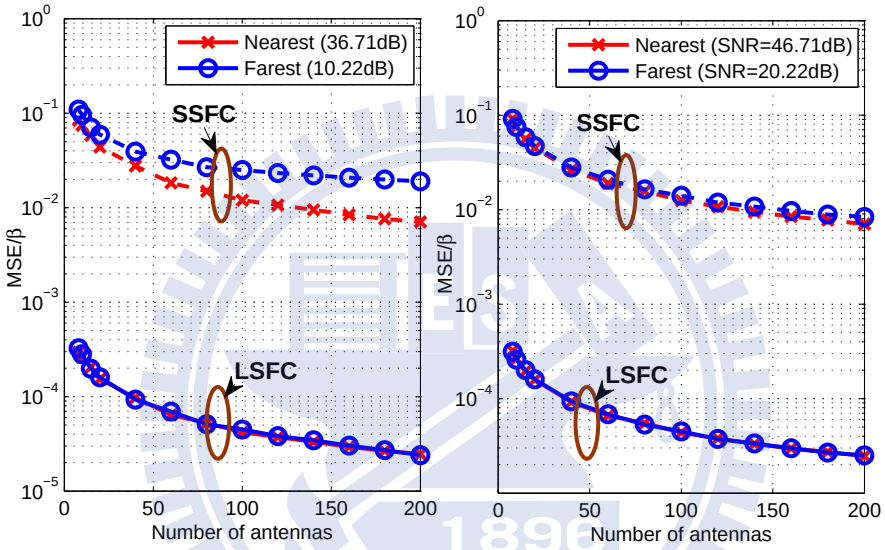

Figure 3.10: MSE performance of the proposed LSFC and SSFC estimator versus number of BS antennas with different user location, hence different received SNR at BS (indicated in the legend), where AS= 15°, and full modeling order is used.

Chapter 4

Estimation of Small-Scale Fading Coefficients

Since the SSFC estimation scheme is valid for any user-BS link, for the sake of brevity, we omit the user index k in the ensuing discussion.

4.1 Reduced-Rank Channel Modeling

In [12], two analytic correlated MIMO channel models were proposed. These models generalize and encompass as special cases, among others, the Kronecker [9, 40], virtual representation [11] and Weichselberger [10] models. They often admit flexible reduced-rank representations. Moreover, if the angle spread (AS) of the transmit signal is small, which, as reported in a recent measurement campaign [2], is the case when a large uniform linear array (ULA) is used at the BS, one of the models can provide angle of arrival (AoA) information. In other words, since the ASs from uplink users in a massive MIMO system are relatively small (say, less than 15°), the following RR model is easily derivable from [12, *Proposition 1*]

Lemma 4.1.1 (RR representations)**.** *The channel vector seen by kth user can be represented by*

$$\mathbf{h} = \mathbf{Q}_m^{(I)} \mathbf{c}^{(I)} \tag{4.1}$$

or alternately by

$$\mathbf{h} = \mathbf{W}(\phi)\mathbf{Q}_m^{(II)}\mathbf{c}^{(II)} \tag{4.2}$$

where $\mathbf{Q}_m^{(I)}, \mathbf{Q}_m^{(II)} \in \mathbb{R}^{M\times m}$ *are predetermined (unitary) basis matrices and* $\mathbf{c}^{(I)}, \mathbf{c}^{(II)} \in \mathbb{C}^{m\times 1}$ *are the channel vectors with respect to bases* $\mathbf{Q}_m^{(I)}$ *and* $\mathbf{Q}_m^{(II)}$ *for the user k-BS link and* $\mathbf{W}(\phi)$ *is diagonal with unit magnitude entries. The two equalities hold only if* $m = M$ *and become approximations if* $m < M$*. Furthermore, if the AS is small,* $[\mathbf{W}(\phi)]_{ii} = \exp\left(-j2\pi\frac{(i-1)\xi}{\lambda}\sin\phi\right)$ *with* ξ *and* λ *being the antenna spacing and signal wavelength and* ϕ *is the mean AoA.*

This Lemma suggests that the mean AoA (which is approximately equal to the incident angle of the strongest path) of each user link is extractable if the associated AS is small. Having a large aperture, a massive MIMO antenna array may provide fine AoA resolution and a channel estimator based on the model (4.2) can therefore offer accurate mean AoA information [1].

Remark 8. *The estimated mean AoAs can be used by the BS to perform downlink beamforming. The use of predetermined basis matrices, as the virtual representation [11], avoids the need of the spatial correlation information required by [10].*

Remark 9. *For large-scale ULAs, the spatial correlation can be high, small modeling order* m *may be sufficient to capture the spatial variance of the SSFCs.*

4.2 SSFC Estimation

We begin with the channel model (4.1) and denote by $\tilde{\boldsymbol{\epsilon}}$ the modeling error. Let $\gamma = \sqrt{\beta}\|\mathbf{p}\|^2$ and assume for the moment that LSFCs are known. Then

$$\begin{aligned}\mathbf{Yp} &= \sqrt{\beta}\|\mathbf{p}\|^2\mathbf{h} + \mathbf{Np} \\ &= \gamma\left(\mathbf{Q}_m^{(\mathrm{I})}\mathbf{c}^{(\mathrm{I})} + \tilde{\boldsymbol{\epsilon}}\right) + \mathbf{Np}\end{aligned} \tag{4.3}$$

which brings about the following LS problem

$$\min_{\mathbf{c}} \quad \left\|\mathbf{Yp} - \gamma \mathbf{Q}_m^{(\mathrm{I})} \mathbf{c}^{(\mathrm{I})}\right\|^2 \tag{4.4}$$

The optimal solution can be shown as

$$\hat{\mathbf{c}}^{(\mathrm{I})} = \frac{1}{\gamma} \left(\mathbf{Q}_m^{(\mathrm{I})}\right)^H \mathbf{Yp}. \tag{4.5}$$

Replacing γ by $\hat{\gamma} = \hat{\beta}^{\frac{1}{2}} \|\mathbf{p}\|^2$ for the case when LSFCs have to be estimated, we have

$$\hat{\mathbf{h}}^{(\mathrm{I})} = \mathbf{Q}_m^{(\mathrm{I})} \hat{\mathbf{c}}^{(\mathrm{I})} = \frac{1}{\gamma} \mathbf{Q}_m^{(\mathrm{I})} \left(\mathbf{Q}_m^{(\mathrm{I})}\right)^H \mathbf{Yp}. \tag{4.6}$$

On the other hand, if (4.2) is the channel model and $\boldsymbol{\epsilon}$ is the corresponding modeling error, then

$$\begin{aligned} \mathbf{Yp} &= \sqrt{\beta} \|\mathbf{p}\|^2 \mathbf{h} + \mathbf{Np} \\ &= \gamma \left(\mathbf{W}(\phi) \mathbf{Q}_m^{(\mathrm{II})} \mathbf{c}^{(\mathrm{II})} + \boldsymbol{\epsilon}\right) + \mathbf{Np} \end{aligned} \tag{4.7}$$

which suggests the LS formulation

$$\begin{aligned} \min_{\phi, \mathbf{c}} \quad & \left\|\mathbf{Yp} - \gamma \mathbf{W}(\phi) \mathbf{Q}_m^{(\mathrm{II})} \mathbf{c}^{(\mathrm{II})}\right\|^2 \\ \text{s.t.} \quad & \mathbf{W}(\phi) = \mathrm{Diag}\left(\omega_1(\phi), \cdots, \omega_M(\phi)\right), \\ & \omega_i(\phi) = \exp\left(-j2\pi \frac{(i-1)\xi}{\lambda} \sin\phi\right). \end{aligned} \tag{4.8}$$

With $\mathbf{F}_m(\phi) \overset{def}{=} \mathbf{W}(\phi) \mathbf{Q}_m^{(\mathrm{II})}$ and $\mathbf{A}^\dagger \overset{def}{=} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$, the optimal solution to (4.8) is given as

$$\begin{aligned} \hat{\phi} &= \arg\max_{\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]} \mathbf{p}^H \mathbf{Y}^H \mathbf{F}_m(\phi) \mathbf{F}_m^\dagger(\phi) \mathbf{Yp} \\ &= \arg\max_{\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]} \left\| \left(\mathbf{Q}_m^{(\mathrm{II})} \mathbf{W}(\phi)\right)^H \mathbf{Yp} \right\|^2. \end{aligned} \tag{4.9}$$

$$\hat{\mathbf{c}}^{(\mathrm{II})} = \frac{1}{\gamma} \mathbf{F}_m^\dagger(\hat{\phi}) \mathbf{Yp} = \frac{1}{\gamma} \left(\mathbf{Q}_m^{(\mathrm{II})}\right)^H \mathbf{W}^H(\hat{\phi}) \mathbf{Yp}, \tag{4.10}$$

When the true LSFCs are not available we use their estimates, $\hat{\gamma} = \hat{\beta}^{\frac{1}{2}} \|\mathbf{p}\|^2$, and obtain the SSFCs estimate

$$\hat{\mathbf{h}}^{(\mathrm{II})} = \mathbf{W}(\hat{\phi}) \mathbf{Q}_m^{(\mathrm{II})} \hat{\mathbf{c}}^{(\mathrm{II})}. \tag{4.11}$$

Both (4.9) and (4.10) require no matrix inversion while $\hat{\phi}$ can be easily found by a simple line search. However, in *Section 4.5* we show

Theorem 4.2.1. *The (4.2)-based channel estimator* $\hat{\mathbf{h}}^{(II)}$ *defined by (4.11), which combines the mean AoA estimate* $\hat{\phi}$ *and the RR representation vector estimate* $\hat{\mathbf{c}}$*, yields better MSE performance than that provided by* $\hat{\mathbf{h}}^{(I)}$*, which is based on (4.1). In other words,*

$$\mathrm{MSE}_m(\hat{\mathbf{h}}^{(I)}) > \mathrm{MSE}_m(\hat{\mathbf{h}}^{(II)}) \tag{4.12}$$

where $\mathrm{MSE}_m(\cdot)$ *stands for the MSE of the enclosed estimate with modeling order* m*.*

Because of the superiority of $\hat{\mathbf{h}}^{(\mathrm{II})}$ we henceforth discuss the properties of this SSFC estimator only; the superscript (II) is omitted.

As estimation performance depends on the basis matrix $\mathbf{Q}_m$ and the modeling order m used, we address related issues prior to the performance analysis.

4.3 Basis Selection for RR Channel Modeling

Define the rotated channel vector by $\mathbf{r} = \mathbf{W}^H(\phi)\mathbf{h}$ with the covariance matrix given by

$$\mathbf{C} \stackrel{def}{=} \mathbb{E}\left\{\mathbf{W}^H(\phi)\mathbf{h}\mathbf{h}^H\mathbf{W}(\phi)\right\} = \mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$$

and can be shown to be close to a real matrix. Denote by $\mathcal{U}(M)$ the unitary group of degree M, which consists of all $M \times M$ unitary matrix, and $\mathbf{Q} \stackrel{def}{=} \mathbf{Q}_M$ the complete basis matrix. With $\mathbf{Q} \in \mathcal{U}(M)$, we perform a unitary transform on the rotated channel vector $\mathbf{r}$ and obtain the so-called transformed channel vector

$$\mathbf{u} = \mathbf{Q}^H\mathbf{r} = \mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{h} \tag{4.13}$$

which is equal to $\mathbf{c}^{(\mathrm{II})}$ in (4.2) when $m = M$. For a reason to become clear later we call its covariance matrix

$$\mathbf{B} \stackrel{def}{=} \mathbb{E}\left\{\mathbf{Q}^H\mathbf{r}\mathbf{r}^H\mathbf{Q}\right\} = \mathbf{Q}^H\mathbf{C}\mathbf{Q}. \tag{4.14}$$

the *bias matrix*, which is the separable two-dimensional (2D) unitary transform of the covariance matrix $\mathbf{C}$.

As an unitary transform is energy preserving

$$\sum_{\ell=1}^{M} \sigma_\ell^2 \;\; = \;\; \mathrm{tr}(\mathbf{B}) = \mathrm{tr}(\mathbf{Q}^H \mathbf{C} \mathbf{Q}) = \sum_{\ell=1}^{M} [\mathbf{C}]_{\ell\ell} = M \tag{4.15}$$

where $\sigma_\ell^2 \stackrel{def}{=} [\mathbf{B}]_{\ell\ell}$ and we have use the identity

$$[\mathbf{C}]_{\ell\ell} = \mathrm{tr}(\mathbf{e}_i^H \mathbf{W}^H(\phi) \mathbf{\Phi} \mathbf{W}(\phi) \mathbf{e}_i) = \mathrm{tr}(\mathbf{\Phi} \mathbf{E}_{\ell\ell}) = [\mathbf{\Phi}]_{\ell\ell} = 1.$$

the transformation $\mathbf{Q}$ in effect redistributes the energy (or variance) of and decorrelates $\mathbf{r}$.

To accurately describe a spatially-correlated channel via (4.2) with minimal modeling order m, we have to find a unitary transform that distributes most of the power of $\mathbf{r}$ in a subspace of minimal dimension. To this end, we make use of the notion of *energy compactness* of an unitary transform in image processing [34, 35, 41] which is measured by the (subband) coding gain [34, Ch.1]. In other words, by maximizing the coding gain, $\tau(\mathbf{Q})$, of the unitary transform $\mathbf{Q}$ defined as

$$\tau(\mathbf{Q}) = \frac{\frac{1}{M}\sum_{\ell=1}^{M} \sigma_\ell^2}{\left(\prod_{i=1}^{M} \sigma_\ell^2\right)^{\frac{1}{M}}} = \frac{1}{\left(\prod_{\ell=1}^{M} \sigma_\ell^2\right)^{\frac{1}{M}}}. \tag{4.16}$$

or, equivalently, by solving the minimization problem

$$\min_{\mathbf{Q} \in \mathcal{U}(M)} \left(\prod_{i=1}^{M} \sigma_\ell^2 \right)^{\frac{1}{M}}. \tag{4.17}$$

We obtain the unitary transform with highest energy compactness. The optimal solution, which depends on $\mathbf{C}$, is called Karhunen-Loève transformation (KLT) and satisfies

$$\mathbf{B}_{\mathrm{KL}} = \mathbf{Q}_{\mathrm{KL}}^H \mathbf{C} \mathbf{Q}_{\mathrm{KL}} = \mathrm{Diag}(\lambda_1, \cdots, \lambda_M) \tag{4.18}$$

where $\lambda_1 \geq \cdots \geq \lambda_M$ are the eigenvalues of $\mathbf{C}$ and $\mathbf{Q}_{\mathrm{KL}} \stackrel{def}{=} [\mathbf{v}_1, \cdots, \mathbf{v}_M]$ with $\mathbf{v}_\ell$ being the eigenvector corresponding to λ_ℓ [34, 35].

Corollary 4.3.1. *The use of KLT basis guarantees*

$$\tau(\mathbf{Q}_{KL})\big|_{M_1} > \tau(\mathbf{Q}_{KL})\big|_{M_2} \tag{4.19}$$

if BS antenna array sizes $M_1 > M_2$.

$\mathbf{Q}_{\mathrm{KL}}$ [34] is optimal in energy-compaction as it minimizes the truncation error

$$\varepsilon_m \overset{def}{=} \frac{1}{M}\mathbb{E}\left\{\left\|\mathbf{u}-\mathbf{u}^{(m)}\right\|^2\right\} \tag{4.20}$$

where $\mathbf{u}^{(m)}$ is the truncated version of $\mathbf{u}$ by nulling the last $M-m$ elements of the latter.

Remark 10. *The truncation error* ε_m *is equivalent to the normalized mean squared error, i.e.,*

$$\begin{aligned}\varepsilon_m &= \frac{1}{M}\mathbb{E}\left\{\left\|\mathbf{W}(\phi)\mathbf{Q}\left(\mathbf{u}-\mathbf{u}^{(m)}\right)\right\|^2\right\} \\ &= \frac{1}{M}\mathbb{E}\left\{\left\|\mathbf{h}-\hat{\mathbf{h}}\right\|^2\right\}\end{aligned} \tag{4.21}$$

Corollary 4.3.2 (Coding gain upper bound [35])**.** *For a massive MIMO system with array size* $M \to \infty$*, the corresponding coding gain, which is also the coding gain upper bound due to the strictly monotonic increasing nature of Corollary 4.3.1, can be written as*

$$\lim_{M\to\infty}\tau(\mathbf{Q}) = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{\mathbf{r}}\left(e^{j\omega}\right)d\omega}{\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(S_{\mathbf{r}}\left(e^{j\omega}\right)\right)d\omega\right\}} \geq 1, \tag{4.22}$$

where $S_{\mathbf{r}}\left(e^{j\omega}\right)$*, assumed nonzero, is the power spectral density (PSD) of the rotated channel vector* $\mathbf{r}$*. The equality is achieved when* $\mathbf{r}$ *is white and the asymptotic coding gain, whose reciprocal is a measure of the spectral flatness, increases as the spectrum becomes less and less flat [35], i.e., as the degree of spatial correlation increases.*

The above two corollaries say that the efficiency of the RR channel model can increases with M and/or spatial correlation. [18, Lemma 2] also obtains a similar but more explicit result on the spatial correlation case.

However, KLT is computationally expensive and nonflexible in that $\mathbf{Q}_{\mathrm{KL}}$ is channel dependent and one needs to collect sufficient observations before SSFC estimation to compute the spatial correlation matrix $\mathbf{C}$, which is then eigen-decomposed to derive the basis matrix. Even worse, there is no fast algorithm for the KLT. Our channel model (4.2) uses a predetermined signal-independent basis $\mathbf{Q}_m$ which requires far less complexity. Two candidate bases are of special interest to us for their proximity to the KL basis.

4.3.1 Polynomial Basis [12]

As the BS antenna spatial correlation is often reasonably smooth, polynomial basis of dimension $m < M$ may be sufficient to track the channel variation. To construct an orthonormal discrete polynomial basis we perform standard QR decomposition $\mathbf{P} = \mathbf{QR}$, where $[\mathbf{P}]_{ij} = (i-1)^{j-1}$, $\forall i, j = 1, \cdots, M$, Since the polynomial degree of each column of $\mathbf{Q}$ are arranged in an ascending order, the RR basis $\mathbf{Q}_m$ is obtained by keeping the first m columns.

4.3.2 Type-2 Discrete Cosine Transform (DCT) Basis [41]

DCT, especially Type-2 DCT (DCT-2 or simply DCT), is a widely used for image coding for its excellent energy compaction capability [36, 41]. For a smooth finite-length sequence, its DCT is often energy-concentrated in lower-indexed coefficients. Hence the DCT basis matrix

$$[\mathbf{Q}_m]_{ij} = q_j \cos\left(\frac{\pi(2i-1)(j-1)}{2M}\right), \tag{4.23}$$

for $1 \le i \le M$ and $1 \le j \le m$, where $1 \le m \le M$ and

$$q_j = \begin{cases} \sqrt{1/M}, & j = 1; \\ \sqrt{2/M}, & j = 2, \cdots, M. \end{cases} \tag{4.24}$$

is an excellent candidate RR basis for our channel estimation purpose. Some comments on the predetermined basis selection are provided [34, 35].

Remark 11. *The complexity of computing KLT, $\mathcal{O}(M^2)$, is much greater than that of DCT's $\mathcal{O}(M \log_2 M)$.*

Remark 12. *The fact that the energy compaction efficiency of DCT is near-optimal make it the closest KLT approximation in the high correlation regime among the following unitary transforms: Walsh-Hadamard, Slant, Haar, and discrete Legendre transform. The last one is equivalent to a polynomial-based transform and is slightly inferior to DCT in energy compaction capability.*

The above claims have been verified in the context of image compression [34, 35]. In terms of RR MIMO channel representation, we show in Section 4.8 that, for the same modeling order m, the DCT basis does outperform the polynomial basis in MS channel estimation error regardless of the correlation level.

4.4 Optimal SSFC Modeling Order Determination

With the aforementioned basis matrices $\mathbf{Q}$, we analyze the performance of the proposed SSFC estimator and its RR capability, assuming perfectly known LSFCs and mean AoAs. We first substitute (4.10) into (4.11) to obtain

$$\begin{aligned} \hat{\mathbf{h}} &= \frac{1}{\gamma}\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{Y}\mathbf{p} \\ &= \frac{1}{\gamma}\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)(\gamma\mathbf{h}+\mathbf{N}\mathbf{p}) \\ &= \mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{h}+\boldsymbol{\nu} \end{aligned} \tag{4.25}$$

where $\boldsymbol{\nu} = \frac{1}{\gamma}\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{N}\mathbf{p}$, and the decomposition

$$\begin{aligned} \mathrm{MSE}_m(\hat{\mathbf{h}}) &= \mathbb{E}\left\{\left\|\hat{\mathbf{h}}-\mathbf{h}\right\|^2\right\} \\ &= \underbrace{\mathbb{E}\left\{\left\|\hat{\mathbf{h}}-\mathbb{E}\{\hat{\mathbf{h}}\}\right\|^2\right\}}_{\stackrel{def}{=}\mathrm{Var}\{\hat{\mathbf{h}}\}} + \underbrace{\mathbb{E}\left\{\left\|\mathbb{E}\{\hat{\mathbf{h}}\}-\mathbf{h}\right\|^2\right\}}_{\stackrel{def}{=}b(\hat{\mathbf{h}})} \end{aligned} \tag{4.26}$$

$\mathrm{Var}\{\hat{\mathbf{h}}\}$ and $b(\hat{\mathbf{h}})$ represent respectively the variance and bias of estimator $\hat{\mathbf{h}}$. For these two error terms we prove in Appendix B

Theorem 4.4.1 (Proposed SSFC Estimator MSE)**.** *For SSFC estimator* $\hat{\mathbf{h}}$ *based on the representation (4.2),*

$$\mathrm{Var}\left\{\hat{\mathbf{h}}\right\} = \frac{m}{\beta\|\mathbf{p}\|^2}, \tag{4.27}$$

$$b(\hat{\mathbf{h}}) = \mathrm{tr}\left(\mathbf{D}_m\mathbf{B}\right) \tag{4.28}$$

where $\mathbf{D}_m = \mathrm{Diag}\left(\left[\mathbf{0}_{1\times m}\ \mathbf{1}_{1\times(M-m)}\right]^T\right)$ *and* $\mathbf{B}$ *the bias matrix defined by (4.14) with nonnegative diagonal entries.*

Remark 13. *The estimation bias* $b(\hat{\mathbf{h}})$ *is equal to the sum of the last* $M-m$ *diagonal terms of the positive semidefinite matrix* $\mathbf{B}$*, i.e.,*

$$b(\hat{\mathbf{h}}) = \sum_{\ell=m+1}^{M} [\mathbf{B}]_{\ell\ell}.$$

and is minimized by the semiunitary $\mathbf{Q}_{KL}^{(m)}$ *associated with the KLT of* $\mathbf{C}$*. That is, if we denote by* $\mathcal{U}(M,m)$ *the set of all* $M\times m$ *semiunitary matrices, and by* $\mathbf{Q}^{(m)}$ *the semiunitary matrix obtained by deleting the rightmost* $M-m$ *columns of the* $M\times M$ *unitary matrix* $\mathbf{Q}$*, then we have*

$$\arg\min_{\mathbf{Q}\in\mathcal{U}(M,m)} b(\hat{\mathbf{h}}) = \arg\min_{\mathbf{Q}\in\mathcal{U}(M,m)} \mathbb{E}\left\{\left\|\mathbf{u}-\mathbf{u}^{(m)}\right\|^2\right\} = \mathbf{Q}_{KL}^{(m)}.$$

Corollary 4.4.2 (Effect of M on RR estimation [34, Ch. 1])**.** *For two BS antenna array sizes* $M_1 > M_2$ *and a fixed* m*, the biases associated with the channel estimator (4.11) satisfy*

$$\frac{1}{M_1}\mathrm{tr}(\mathbf{D}_m\mathbf{B})\big|_{M_1} < \frac{1}{M_2}\mathrm{tr}(\mathbf{D}_m\mathbf{B})\big|_{M_2}. \tag{4.29}$$

if $\mathbf{Q}_{KL}^{(m)}$ *is used. The normalized MSE (NMSE) also has the monotonicity property:*

$$\mathrm{NMSE}_m(\hat{\mathbf{h}})\big|_{M_1} \overset{def}{=} \frac{1}{M_1}\mathrm{MSE}_m(\hat{\mathbf{h}})\big|_{M_1} < \mathrm{NMSE}_m(\hat{\mathbf{h}})\big|_{M_2} \tag{4.30}$$

If full rank model is used, we have

Lemma 4.4.3. *The combined SSFC estimator,* $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_k]_{k=1}^K$*, is equivalent to the conventional unbiased LS estimator given in [29]*

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{P}^H \text{Diag}\left(\frac{1}{\gamma_1}, \cdots, \frac{1}{\gamma_K}\right) \tag{4.31}$$

when $\gamma_l = \sqrt{\beta_l}\|\mathbf{p}\|^2$ *and the full modeling order,* $m = M$*, is used.*

Proof. The proposed SSFC estimator (4.11) is biased when $m < M$, since with (4.25) we have

$$\mathbb{E}\left\{\hat{\mathbf{h}}_k\right\} = \mathbf{W}(\phi_k)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi_k)\mathbf{h}_k. \tag{4.32}$$

However, when $m = M$, the fact that $\mathbf{Q}_m\mathbf{Q}_m^H = \mathbf{W}(\phi_k)\mathbf{W}^H(\phi_k) = \mathbf{I}_M$ makes the estimator unbiased and

$$\begin{aligned}\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1, \cdots, \hat{\mathbf{h}}_K\right] &= \mathbf{Y}\left[\mathbf{p}_1, \cdots, \mathbf{p}_K\right] \text{Diag}\left(\frac{1}{\gamma_1}, \cdots, \frac{1}{\gamma_K}\right) \\ &= \mathbf{Y}\mathbf{P}^H \text{Diag}\left(\frac{1}{\gamma_1}, \cdots, \frac{1}{\gamma_K}\right),\end{aligned} \tag{4.33}$$

a same result given in [29]. □

In the following, we analyze the estimator's MSE performance for both uncorrelated and correlated channels.

Definition 4.4.4. *The optimal modeling order* $m^\star$ *is the one that minimizes* $\text{MSE}_m(\hat{\mathbf{h}})$.

4.4.1 Uncorrelated Channels

Let SNR be the (normalized) signal-to-noise power ratio (SNR) with user k's signal at the BS.

$$\text{MSE}_m(\hat{\mathbf{h}}) = \frac{m}{\beta\|\mathbf{p}\|^2} + \text{tr}(\mathbf{D}_m) = M - m\left(1 - \frac{1}{T \cdot \text{SNR}}\right) \tag{4.34}$$

where the first equality is due to the zero spatial correlation assumption ($\mathbf{B} = \mathbf{I}_M$) and $\text{SNR} = \beta\|\mathbf{p}\|^2/T$. Obviously, for this case, MSE improves with increasing modeling order m if $\beta\|\mathbf{p}\|^2 > 1$. We conclude that

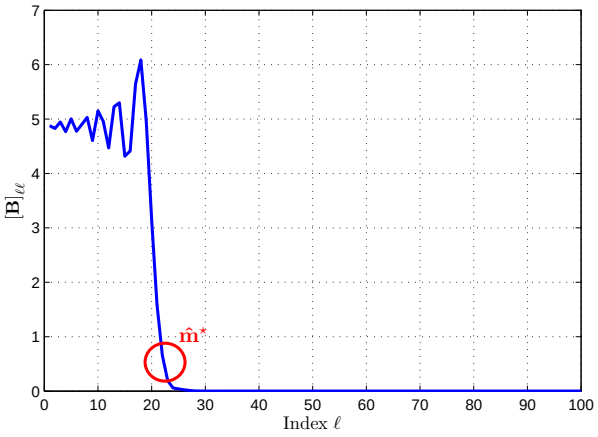

Figure 4.1: An illustration of the σ_ℓ^2 distribution with respect to ℓ for $M = 100$.

Corollary 4.4.5. *The optimal modeling order for uncorrelated channels is M if the symbol SNR is greater than 1.*

Note that (4.34) also predicts that, for low symbol SNR ($\beta\|\mathbf{p}\|^2 < 1$), higher modeling order does help even for uncorrelated spatial channels as we are just trying to fit more parameters to noise-dominant observations.

4.4.2 Correlated Channels

As indicated by (4.26)–(4.28) and the discussion on basis selection, to minimize MSE we have to consider the energy compactness of $\mathbf{B}$. Recall that $[\mathbf{B}]_{\ell\ell} = 1$ when $\mathbf{\Phi} = \mathbf{I}_M$ and, for a given basis, the diagonal terms becomes more and more 'nonflat' with increasing spatial correlation.

Fig. 4.1 exemplifies a plausible distribution of $[\mathbf{B}]_{\ell\ell} = \sigma_\ell^2$ as a function of the index ℓ. As can be seen, most of the bias power lies in the region $\ell \leq \hat{m}^\star$, with $\hat{m}^\star$ located at the bottom of the water-fall curve. The dominant subspace of the column space of the bias matrix is that spanned by the first $\hat{m}^\star$ columns of $\mathbf{Q}$. We define the optimal modeling order as

Definition 4.4.6. *The optimal modeling order $\hat{m}^\star$ for the threshold η satisfies*

$$\begin{aligned} \min_{m} \quad & m \\ \text{s.t.} \quad & [\mathbf{B}]_{\ell\ell} < \eta, \ \forall \ \ell \geq m. \end{aligned} \tag{4.35}$$

As optimal modeling order $m^\star$ needs to be acquired via exhaustively searching over all m's, $\hat{m}^\star$ eases such burden and is shown to be near-optimal in the rest of this subsection with a proper choice of η.

Corollary 4.4.7. *Modeling order $\hat{m}^\star$ decreases with increasing spatial correlation or decreasing* $\text{rank}(\mathbf{\Phi})$.

Proof. With λ_ℓ's being the ordered eigenvalues of $\mathbf{\Phi}$ and

$$\mathbf{B} \ = \ \mathbf{Q}_{\text{KL}}^H \mathbf{W}(\phi)^H \mathbf{\Phi} \mathbf{W}(\phi) \mathbf{Q}_{\text{KL}} = \text{Diag}(\lambda_1, \cdots, \lambda_M),$$

the fact that $\text{rank}(\mathbf{B}) = \text{rank}(\mathbf{\Phi})$ implies that increasing spatial correlation reduces the rank of $\mathbf{B}$ and thus $\hat{m}^\star$. □

As mentioned in *Remark 12* that DCT is near-optimal, it can be shown that DCT basis exhibits the same behavior trend. On the other hand, since spatial waveforms smoothen with increasing correlation and can be reproduced by lower order polynomials, modeling order $\hat{m}^\star$ decreases accordingly.

Remark 14. *Let $\eta = 1$. For a channel of moderate-to-high correlation, the use of polynomial or DCT basis matrix guarantees the diagonal entries $[\mathbf{B}]_{\ell\ell}$ that are greater than or equal to η be concentrated at the lower indices ℓ. More precisely, for $m > \hat{m}^\star$,*

$$\frac{\sum_{\ell=m+1}^{M} [\mathbf{B}]_{\ell\ell}}{M} \ll 1, \tag{4.36}$$

and thus bias $b(\hat{\mathbf{h}}) \ll 1$. For more details, see Appendix C.

Consequently, the MSE performance is dominated by $\mathrm{Var}\{\hat{\mathbf{h}}\}$ and

Remark 15. *The increase of modeling order m beyond $\hat{m}^\star$ cannot improve or, even worse, can do harm to the MSE performance. This is because for $m > \hat{m}^\star$, the MSE of the RR channel estimate*

$$\mathrm{MSE}_m(\hat{\mathbf{h}}) \approx \mathrm{Var}\{\hat{\mathbf{h}}\} = \frac{m}{\beta\|\mathbf{p}\|^2} \tag{4.37}$$

approximately grows linearly with m. Therefore, if $\hat{m}^\star < m_1 < m_2$, then

$$\mathrm{MSE}_{m_1}(\hat{\mathbf{h}}) < \mathrm{MSE}_{m_2}(\hat{\mathbf{h}}). \tag{4.38}$$

Remark 16. *On the other hand, for moderate to high SNRs, catastrophic MSE performance degradation is induced with m decreasing towards 0. This is due to the fact that the MSE is bias-dominated for $m < \hat{m}^\star$ and thus*

$$\mathrm{MSE}_m(\hat{\mathbf{h}}) \approx \underbrace{\sum_{\ell=m+1}^{\hat{m}^\star} [\mathbf{B}]_{\ell\ell}}_{>\hat{m}^\star - m} + \underbrace{\sum_{\ell=\hat{m}^\star+1}^{M} [\mathbf{B}]_{\ell\ell}}_{\approx 0} \approx \sum_{\ell=m+1}^{\hat{m}^\star} [\mathbf{B}]_{\ell\ell} > \hat{m}^\star - m. \tag{4.39}$$

Thus, if $m_1 < m_2 < \hat{m}^\star$, then

$$\mathrm{MSE}_{m_1}(\hat{\mathbf{h}}) - \mathrm{MSE}_{m_2}(\hat{\mathbf{h}}) = \sum_{\ell=m_1+1}^{m_2} [\mathbf{B}]_{\ell\ell} > m_2 - m_1 \gg 0 \tag{4.40}$$

where the last inequality is owing to the energy compaction property of the basis used.

Verified via simulation (see Section 4.8), for practical correlated channels, similar convergence outcome is attained with any $\eta \in [0.5, 1]$.

4.5 RR Model I Based Performance Analysis

Theorem 4.2.1 says that separate estimation of mean AoA matrix $\mathbf{W}(\phi)$ and RR representation vector $\mathbf{c}$ using RR Model II (4.2) outperforms $\hat{\mathbf{h}}^{(\mathrm{I})}$ based on Model I (4.1) which estimates only $\mathbf{c}$. We now prove this claim.

We called the estimator based on the Model I the SSFC-2 estimator. When the LSFCs are known and $\tilde{\boldsymbol{\epsilon}}$ is the modeling error, we have

$$\begin{aligned}\mathbf{Yp} &= \sqrt{\beta}\|\mathbf{p}\|^2\mathbf{h} + \mathbf{Np} = \gamma\mathbf{h} + \mathbf{Np} \\ &= \gamma\left(\mathbf{Q}_m\mathbf{c} + \tilde{\boldsymbol{\epsilon}}\right) + \mathbf{Np} \end{aligned} \tag{4.41}$$

which brings about the following LS problem

$$\min_{\mathbf{c}} \quad \|\mathbf{Yp} - \gamma\mathbf{Q}_m\mathbf{c}\|^2 \tag{4.42}$$

The optimal solution is shown by [32, Ch. 8] as

$$\hat{\mathbf{c}}^{(\mathrm{I})} = \frac{1}{\gamma}\mathbf{Q}_m^H\mathbf{Yp}. \tag{4.43}$$

Finally, replace γ as $\hat{\gamma} = \sqrt{\hat{\beta}}\|\mathbf{p}\|^2$ for the cases where LSFCs have to be estimated beforehand. The channel estimates $\forall k$ are retained via

$$\hat{\mathbf{h}}^{(\mathrm{I})} = \mathbf{Q}_m\hat{\mathbf{c}}^{(\mathrm{I})} = \frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{Yp}, \tag{4.44}$$

where $\hat{\mathbf{h}}^{(\mathrm{I})}$ is the SSFC-2 estimator.

Suppose perfect LSFCs and mean AoAs estimates, we obtain

Theorem 4.5.1 (Proposed SSFC-2 Estimator MSE)**.** *For SSFC estimator based on Form-I RR model* $\hat{\mathbf{h}}^{(\mathrm{I})}$*,*

$$\mathrm{Var}\left\{\hat{\mathbf{h}}^{(\mathrm{I})}\right\} = \frac{m}{\beta\|\mathbf{p}\|^2}, \tag{4.45}$$

$$b(\hat{\mathbf{h}}^{(\mathrm{I})}) = \mathrm{tr}\left(\mathbf{D}_m\tilde{\mathbf{B}}\right) \tag{4.46}$$

where $\mathbf{D}_m = \mathrm{Diag}\left(\left[\mathbf{0}_{1\times m}\ \mathbf{1}_{1\times(M-m)}\right]^T\right)$ *and* $\tilde{\mathbf{B}} = \mathbf{Q}^H\boldsymbol{\Phi}\mathbf{Q} \in \mathbb{C}^{M\times M}$ *the bias matrix with non-negative diagonal entries.*

Proof. See Appendix D. □

Although the variance term in MSE is exactly the same as that of SSFC estimator in (4.11), the bias term in MSE is different from that in (4.11). Obviously, the only difference lies in the form of bias matrix. For SSFC estimator, $\mathbf{B} = \mathbf{Q}^H\mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q} \in \mathbb{C}^{M\times M}$ while for SSFC-2, $\tilde{\mathbf{B}} = \mathbf{Q}^H\mathbf{\Phi}\mathbf{Q} \in \mathbb{C}^{M\times M}$. Hence, we firstly explain *Theorem 4.2.1* by the viewpoint of bias matrix. We introduce the definition of "reverse modeling order" beforehand.

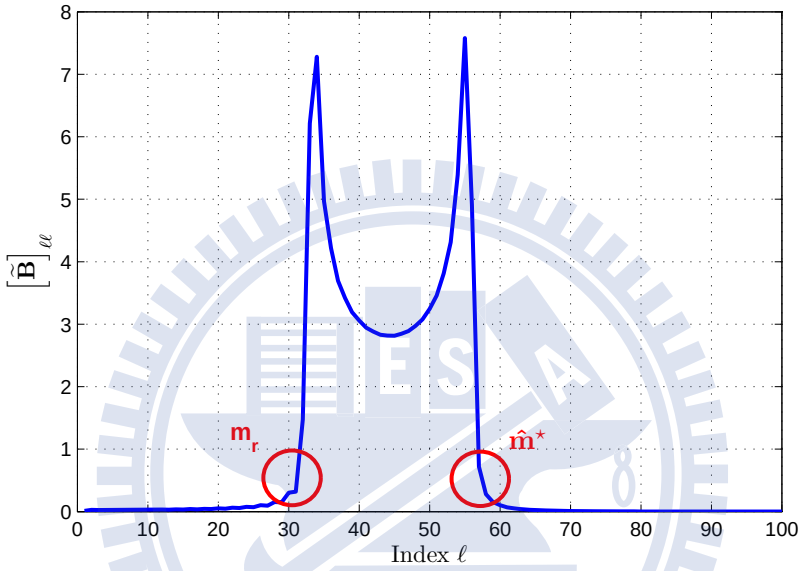

Figure 4.2: An illustration of the $[\tilde{\mathbf{B}}]_{\ell\ell}$ distribution with respect to ℓ when $M = 100$.

Definition 4.5.2. *Denote by η a prescribed threshold, say $\eta = 1$. Define the "reverse modeling order", m_r, as the index of first entry larger than η on the diagonal of $\tilde{\mathbf{B}}$. In more detail, m_r is the index m satisfies*

$$\begin{aligned} \max_{m} \quad & m \\ \text{s.t.} \quad & [\tilde{\mathbf{B}}]_{\ell\ell} < \eta, \ \forall \ell < m. \end{aligned} \tag{4.47}$$

Similar to *Remark 14*, we have to discuss the distribution of diagonal entries in $\tilde{\mathbf{B}}$ to obtain

Remark 17. *Let $\eta = 1$. For a channel of moderate-to-high correlation, the use of polynomial or DCT basis matrix guarantees the diagonal entries $[\tilde{\mathbf{B}}]_{\ell\ell}$ that are greater than or equal to η be concentrated at some medium indices ℓ. In other words,*

$$1 - \frac{\sum_{\ell=m_r-1}^{\hat{m}^\star}[\tilde{\mathbf{B}}]_{\ell\ell}}{M} \ll 1 \tag{4.48}$$

with $\hat{m}^\star - m_r \ll M$.

We illustrate this property in Fig. 4.2. For more details, see Appendix E.

Verified via simulation (see Section 4.8), for practical correlated channels, similar convergence outcome is attained with any $\eta \in [0.5, 1]$.

In accordance with *Remark 17*, because of the frequency translation property , we know that the (estimated) optimal modeling order, $\hat{m}^\star$, of SSFC-2 estimator must be larger than that of SSFC estimator, which verifies *Theorem 4.2.1*.

Finally, we would like to proof *Theorem 4.2.1* mathematically.

Lemma 4.5.3. *Denote $\mathbf{B}$ and $\tilde{\mathbf{B}}$ the bias matrices of SSFC estimator and SSFC-2 estimator, respectively. Then the bias term in MSE satisfies*

$$\mathrm{tr}(\mathbf{D}_m\tilde{\mathbf{B}}) \geq \mathrm{tr}(\mathbf{D}_m\mathbf{B}) \tag{4.49}$$

where $\mathbf{D}_m = \begin{bmatrix} \mathbf{O}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{I}_{M-m} \end{bmatrix} \in \mathbb{R}^{M\times M}$

Proof. See Appendix F. □

Since the variance term in MSE are the same for both SSFC and SSFC-2 estimator, *Lemma 4.5.3* suggests that separately estimate $\mathbf{W}(\phi)$ and $\mathbf{c}$ using SSFC estimator results in smaller MSE than estimate $\mathbf{c}$ only using SSFC-2 estimator, hence proves *Theorem 4.2.1*.

4.6 SNR Effect on Modeling Order

As MSE is also a function of received SNR, we shall investigate its influence on the optimal modeling order.

Lemma 4.6.1. *For* $\mathrm{SNR}_1 < \mathrm{SNR}_2$*, the optimal modeling order*

$$m_1 < m_2 \tag{4.50}$$

where

$$m_j \;=\; \arg\min_m \mathrm{MSE}_m(\hat{\mathbf{h}})\big|_{\mathrm{SNR}_j}. \tag{4.51}$$

Proof. Since $\mathrm{MSE}_{m_1}(\hat{\mathbf{h}})\big|_{\mathrm{SNR}_1} < \mathrm{MSE}_{m_2}(\hat{\mathbf{h}})\big|_{\mathrm{SNR}_1}$,

$$\frac{m_1}{T\cdot\mathrm{SNR}_1} + \sum_{\ell=m_1+1}^{M}[\mathbf{B}]_{\ell\ell} < \frac{m_2}{T\cdot\mathrm{SNR}_1} + \sum_{\ell=m_2+1}^{M}[\mathbf{B}]_{\ell\ell}$$

which can be reduced to

$$\frac{m_2-m_1}{T\cdot\mathrm{SNR}_1} \;>\; \sum_{\ell=m_1+1}^{M}[\mathbf{B}]_{\ell\ell} - \sum_{\ell=m_2+1}^{M}[\mathbf{B}]_{\ell\ell} = \begin{cases} \sum_{\ell=m_1+1}^{m_2}[\mathbf{B}]_{\ell\ell}, & \text{if } m_1 < m_2; \\ -\sum_{\ell=m_2+1}^{m_1}[\mathbf{B}]_{\ell\ell}, & \text{if } m_1 > m_2. \end{cases} \tag{4.52}$$

For SNR_2, $\mathrm{MSE}_{m_2}(\hat{\mathbf{h}})\big|_{\mathrm{SNR}_2} < \mathrm{MSE}_{m_1}(\hat{\mathbf{h}})\big|_{\mathrm{SNR}_2}$ gives

$$\frac{m_2-m_1}{T\cdot\mathrm{SNR}_2} \;<\; \begin{cases} \sum_{\ell=m_1+1}^{m_2}[\mathbf{B}]_{\ell\ell}, & \text{if } m_1 < m_2; \\ -\sum_{\ell=m_2+1}^{m_1}[\mathbf{B}]_{\ell\ell}, & \text{if } m_1 > m_2. \end{cases} \tag{4.53}$$

Thus, we have

$$\begin{cases} \mathrm{SNR}_1 < \dfrac{m_2-m_1}{T\sum_{\ell=m_1+1}^{m_2}[\mathbf{B}]_{\ell\ell}} < \mathrm{SNR}_2, & \text{if } m_1 < m_2; \\ \mathrm{SNR}_2 < \dfrac{m_1-m_2}{T\sum_{\ell=m_2+1}^{m_1}[\mathbf{B}]_{\ell\ell}} < \mathrm{SNR}_1, & \text{if } m_1 > m_2. \end{cases} \tag{4.54}$$

While the case $m_1 < m_2$ results in a contradiction to our assumption that $\mathrm{SNR}_1 < \mathrm{SNR}_2$, the other is true. We can conclude that $m_1 < m_2$. □

We may see later via simulation that the optimal modeling order $m^\star(0\text{ dB})$ when $\mathrm{SNR} = 0$ dB, is indeed smaller than $\hat{m}^\star$, which is optimal when $\mathrm{SNR} \approx 10$ dB.

Remark 18. *Since* $\mathrm{Var}\{\hat{\mathbf{h}}\} = m/(T\cdot\mathrm{SNR})$*, we have*

$$\mathrm{Var}\{\hat{\mathbf{h}}\}\big|_{\mathrm{SNR}_1} > \mathrm{Var}\{\hat{\mathbf{h}}\}\big|_{\mathrm{SNR}_2}, \tag{4.55}$$

for $\mathrm{SNR}_1 < \mathrm{SNR}_2$. *In addition, as bias* $b(\hat{\mathbf{h}})$ *is independent of SNR, the extent that MSE is dominated by* $\mathrm{Var}\{\hat{\mathbf{h}}\}$ *aggravates with decreasing SNR in the low SNR regime for any* m. *Therefore, as* m *decreases,* $\mathrm{Var}\{\hat{\mathbf{h}}\}\big|_{\mathrm{SNR}_1}$ *loses its dominance slower than* $\mathrm{Var}\{\hat{\mathbf{h}}\}\big|_{\mathrm{SNR}_2}$.

4.7 Rank Determination

In Section 4.4, the RR estimation performance with respect to modeling order $\hat{m}^\star$ (given in *Definition 4.4.6*) is analyzed assuming perfectly known bias matrix $\mathbf{B}$ or at least $[\mathbf{B}]_{\ell\ell}$'s. This is not the case in practice. We investigate the case when only the spatial correlation is known and proposed a (channel) rank (modeling order) determination method on which the estimates for mean AoA and RR channel SLFCs are based.

Lemma 4.7.1. *The optimal modeling order for an uplink user channel is a decreasing function of the mean AoA to the BS (with respect to the array broadside). That is to say, if* $0 \leq |\phi_2| < |\phi_1| \leq \frac{\pi}{2}$, *then* $m^\star(\phi_1) < m^\star(\phi_2)$, *where* $m^\star(\phi)$ *denotes the optimal modeling order with mean AoA being* ϕ.

As ϕ and thus $\mathbf{B} = \mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q}$ are still unknown, except when $\phi = 0$, we propose the iterative modeling order determination (IMOD) algorithm which initializes $\hat{\phi}$ as 0 and accordingly the maximal possible $\hat{m}^\star = \hat{m}^\star(0)$ and finds the modeling order $\hat{m}^\star$ recursively. The following steps are executed sequentially:

1. (Initialization) Initialize $\hat{\phi} = 0$.

2. (Updating bias matrix) Calculate

$$\hat{\mathbf{B}} = \mathbf{Q}^H\mathbf{W}^H(\hat{\phi})\mathbf{\Phi}\mathbf{W}(\hat{\phi})\mathbf{Q} \tag{4.56}$$

   and solve (4.35) with $\mathbf{B} = \hat{\mathbf{B}}$ to obtain the estimate $\hat{m}^\star$.

3. (Updating mean AoA) With $m = \hat{m}^\star$, find $\hat{\phi}$ via (4.9).

4. (Recursion) Go to Step 2); Terminate and output $\hat{m}^\star$ if the stopping criterion is met.

while once the modeling order is determined, we can proceed to the estimation of RR SSFC representations with the mean AoA $\hat{\phi}$ obtained here. Two remarks are given to conclude this section.

Remark 19. *The modeling order $\hat{m}^\star$ obtained via IMOD algorithm requires much less complexity yet achieves near-optimal performance as compared with $m^\star$ via exhaustive search [12] or some SVD-based methods [26] [27]. This is because it does not need to know the perfect SSFC matrix (which is impossible) nor the eigen-structure of spatial correlation matrix, and the energy compaction nature of the basis used guarantees fast convergence. In addition, IMOD algorithm is basis-independent while SVD-based methods can only apply to KLT basis case. It is shown in Section 4.8 that the algorithm can achieve convergence within two iterations.*

Remark 20. *If the system requirement mandates all user channel vectors be estimated with a single modeling order $\hat{m}^\star$, we choose to minimize the error of the user who is the most sensitive to the modeling order mismatch. As suggested by Remark 16, the performance suffers more from insufficient than redundant order.*

$$\hat{m}^\star = \max_k m_k^\star. \tag{4.57}$$

Remark 21. *As a matter of fact, the correlation matrix used in IMOD algorithm is not perfect, we should doing correlation estimation before apply IMOD algorithm. It is worth noting that sufficient samples from different OFDM subcarriers help to improve the estimation accuracy. The spatial correlation matrix can be written as*

$$\mathbf{\Phi} = \frac{1}{\gamma^2}\left(\mathbb{E}\{\mathbf{Y}\mathbf{p}\mathbf{p}^H\mathbf{Y}^H\} - \|\mathbf{p}\|^2\mathbf{I}_M\right) \tag{4.58}$$

where the term $\mathbb{E}\{\mathbf{Y}\mathbf{p}\mathbf{p}^H\mathbf{Y}^H\} \overset{def}{=} \mathbf{\Psi}$ can be accurately estimated by either ML estimation

$$\widehat{\mathbf{\Psi}} = \frac{1}{n-1}\sum_{\ell=1}^{n}\mathbf{Y}_\ell\mathbf{p}\mathbf{p}^H\mathbf{Y}_\ell^H \tag{4.59}$$

or shrinkage estimation method introduced in [20], where ℓ denotes the subcarrier index and n is the number of subcarriers.

4.8 Numerical Results and Discussion

In this section, we present the performance of our SSFC estimator and investigate the effect of modeling order as well as basis matrices. Assume the same simulation scenario as Section 3.4 is used here, wherein parameters are listed in Table 3.1 unless explicitly stated otherwise. Recall that the discussion in this chapter does not vary from user to user, so we focus only on an MS here, and define the average Rx SNR as $\text{SNR} \stackrel{def}{=} \beta\|\mathbf{p}\|^2/T$. Besides, the mean AoA is assumed to be uniformly distributed in $[-60°, 60°]$.

We study the SSFC estimation performance with respect to modeling order and basis matrix with estimated or perfectly-known $\boldsymbol{\beta}$. Since the spatial correlation increases with reducing AS, the spatial waveform of a user is anticipated to be smoother. As a result, in the case where AS is comparatively small, the estimation performance due to over-modeling a channel not only cannot improve, but also may degrade because the number of parameters to be estimated grows with the same amount of available data. As can be observed in Fig. 4.3, the estimation accuracy with polynomial basis degrades as modeling order increases from 20 to 100 for SNR smaller than 9 dB. Besides, the optimal modeling order increases with SNR, e.g., optimal order at SNR = 5 and 10 dB are respectively 20 and 30. This is because in the low SNR regime, MSE performance is noise-limited as suggested in Section 4.6, while modeling error shows its importance for high SNRs. Similar trend is also observed with DCT basis in Fig. 4.5. Figure 4.6 shows that for 15° angle spread and SNR smaller than 8 dB, the MSE performance deteriorates as modeling order goes up from 30 to 100.

All the three cases mentioned above verify Remark 15 in previous discussion, however, it shall be noted that as polynomial basis is used and the angle spread is 15°, the best

MSE performance happened when full modeling order is used as depicted in Fig. 4.4. This is owing to the fact that when AS goes large from 7.2° to 15°, the spatial correlation dropped and hence steepen the spatial waveform, for which higher modeling order (in this case, 100) is needed. In other words, m_c defined in Definition 4.4.6 gets large toward 100 when spatial correlation reduced, which also verifies Remark 4.4.7. Furthermore, Section 4.4.2 has shown that DCT-II outperforms polynomial basis in sense of energy compaction capability, thus, a smaller optimal modeling order must be found in Fig. 4.6 compared to that in Fig. 4.4, which is equal to 100.

For the sake of finding the optimal modeling order more precisely, we use the exhaustive search method similar in [12], which needs the exact SSFC information for all users, to calculate the NMSE for all possible modeling orders, and depict them in Fig. 4.7 and Fig. 4.8 for a fixed SNR, 10dB. The simulation results are obviously consistent with that in Fig. 4.3 to 4.6. Moreover, we observe that when angle spread is 7.2° (large spatial correlation), the optimal modeling order when using DCT-II basis, which is about 15, is only slightly inferior to that when using polynomial basis, which is around 25, as claimed in [35]; however, the performance gap between using respectively the two bases becomes large when angle spread is 15° (small spatial correlation), that is, 30 for the former and 100 for the latter.

We then turn into investigating the impact of different SNR on modeling order, and then validate *Lemma 4.6.1*. When polynomial basis is used, Fig. 4.3 and Fig. 4.7 shows that when angle spread equals to 7.2°, the optimal modeling order is about 20 as SNR = 9 dB while larger than 20 as SNR > 9 dB. As DCT-II basis is used, Fig. 4.5 and 4.7 shows that the optimal modeling order at SNR = 5 and 16 dB are respectively 15 and 25 when angle spread is 7.2°; besides, Fig. 4.6 and 4.8 shows that the optimal modeling order at SNR = 10 and 16 dB are respectively 30 and 40 when angle spread is 15°. This is because in low SNR regime, MSE performance is noise-dominant, while modeling error shows its importance for high SNRs as we have discussed.

Note that the proposed SSFC (and SSFC-2) estimators with order 100 gives the identical performance that a conventional LS estimator in (4.31) can offer as presented in *Lemma 4.4.3*, and depicted in Fig. 4.7-4.8. Although, in the case where AS= 15° and polynomial basis is used, (4.10) and (4.9) fail to give better performance than that of the conventional LS estimator as depicted in Fig. 4.8, the proposed estimator offers direct performance-complexity and/or performance-feedback rate trade-off.

Succeedingly, we inspect the effect of modeling order on spatial waveform of ULA as DCT-II basis being chosen, mean AoA being $\frac{\pi}{21}$ and AS being 7.2°. Fig. 4.9-4.12 show the spatial waveform (real part) of $\hat{\mathbf{h}}^{(\mathrm{I})}$, $\hat{\mathbf{h}}^{(\mathrm{II})}$, and exact $\mathbf{h}$ when modeling order is 5, 15, 80 and 100, respectively. It is clear that modeling order of 15, being correspond to the optimal modeling order, is sufficient to capture all channel behavior, while modeling order of 80 and 100 over-model the channel behavior, and modeling order of 5 is not enough to represent the spatial waveform. Furthermore, the SSFC-2 estimator is always insufficient to model the channel aside from using full modeling order.

The comparison of $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$ are illustrated in Fig. 4.7-4.8. As being proved in Section 4.5, no matter what modeling order we used and how angle spread is, the MSE performance of SSFC-2 estimator is indeed much worse than that of SSFC estimator (except for the case when full modeling order is used, they both equivalent to conventional LS estimator).

Fig. 4.13-4.18 show the diagonal entries of bias matrix, when KLT, DCT-II and polynomial basis is used respectively. The solid line represents diagonal distribution of $\mathbf{B}$ (SSFC estimator being used) while the dotted line represents diagonal distribution of $\tilde{\mathbf{B}}$ (SSFC-2 estimator being used). Firstly, we focus on the solid lines in the figures, when using the KLT basis, diagonal terms larger than $\eta = 1$ are indeed concentrated at (low) indices smaller than 15 and 29 (which are exactly the optimal modeling orders found by IMOD algorithm) when AS is 7.2° and 15°, respectively; thus, energy compaction nature of optimal KLT basis is verified. Nonetheless, when DCT-II basis is used, diagonal terms

larger than $\eta = 1$ are also concentrated at indices smaller than 15 and 29 when AS is 7.2° and 15°, respectively, which guarantees that the energy compaction capability of DCT-II basis is really near-optimal. On the other hand, the optimal modeling order of polynomial basis, is slight larger than that of DCT-II basis ($23 > 15$) when AS is 7.2°, yet much larger than DCT-II basis ($100 \gg 29$) when AS is 15°; this observation is consistent with what we have in Fig. 4.7-4.8. Secondly, we turn to inspect the dotted lines in the figures, whose energy compaction parts are frequency-shifted to some medium orders as we shown in *Remark 17*. Recall we have proves in Section 4.7 that the energy compaction property of each solid line becomes more apparent while the mean AoA increases, we depict this phenomenon in Fig. 4.19-4.21. However, dotted lines in these figures also illustrate the fact that the energy compaction part is frequency-shifted to higher orders when the mean AoA increases, we have also shown this in *Remark 17*. After that, by comparing the solid lines and dotted lines, we can make the same conclusion as Fig. 4.7-4.8, that is to say, SSFC estimator outperforms SSFC-2 estimator a lot.

Fig. 4.22 investigate the convergence speed of our IMOD algorithm given DCT-II basis is used, SNR= 10 dB and $\eta = 1$. It is clear that IMOD algorithm converge within two iterations regardless of how large the AS is. Furthermore, we can seen in Table 4.1 that the convergence is quite accurate on account of the fact that when $\eta = 0.5 \sim 1$, IMOD algorithm achieves almost the same optimal modeling order compared with the exhaustive search (or brute force) methods, which is optimal, used in [12]. Besides, the iteration number needed for convergence seems to have little to do with the choice of η. Therefore, we assume $\eta = 1$ henceforth.

Remember that

$$\mathrm{NMSE}_m(\hat{\mathbf{h}})\big|_M = \frac{m}{MT \cdot \mathrm{SNR}} + \frac{1}{M}\mathrm{tr}\left(\mathbf{D}_m\mathbf{B}\right). \tag{4.60}$$

To compare the NMSE given different BS antenna number, M, we want to emphasize here that the bias term $\frac{1}{M}\mathrm{tr}\left(\mathbf{D}_m\mathbf{B}\right)$ is almost equal for any M if we choose the optimal modeling order, $m^\star$, by IMOD algorithm. Thus, a simplified metric to measure the

NMSE of different M is $\frac{m^\star}{M}$. Table 4.2-4.3 presents the effect of large system to optimal modeling order found by IMOD algorithm given DCT-II and KLT basis is used, and verifies Corollary 4.4.2 since the NMSE metric, $\frac{m^\star}{M}$, becomes smaller when the number of BS antennas M gets larger. Moreover, we see again that the DCT-II basis gives almost the same optimal modeling order as KLT basis whatever M we used. On the other hand, [18, Lemma 2.] has derived the upper bound of NMSE metric, $\frac{m^\star}{M}$, as $M \to \infty$, being 0.0187 and 0.0386 when AS= 7.2° and AS= 15°, respectively.

We than consider the impact of imperfect (estimated) spatial correlation matrix to our IMOD algorithm. The energy distribution of rotated channel vector, which is just the diagonal distribution of bias matrix $\mathbf{B}$, is depicted in Fig. 4.23-4.30 when the number of subcarriers is 100, 50, 30 or 2 and AS is 7.2° or 15°, respectively. It is clear that when the number of subcarriers larger than the *true* optimal modeling order, than KLT and DCT-II basis yields almost the same estimated dominant rank as the true one. However, when the number of subcarriers reduced to 2, the estimated dominant rank equals to the number of subcarriers while remains the same (29 for 15° AS and 15 for 7.2° AS) when KLT and DCT-II basis are used respectively. The reason why this is so is due to the fact that when DCT-II basis is used, we can find out the *largest* frequency components even if the number of samples (subcarriers) are small, this phenomenon can also be observed when using polynomial basis; on the other hand, KLT basis is irregular with the number of nonzero terms on the diagonal of $\mathbf{B}$ less or equals to the number of samples (subcarriers), hence, when using KLT basis, we need sufficient samples to accurately estimate the dominant rank, which is called the "sample-deficient problem". It is worth mentioning that common SVD-based methods [26] [27] is equivalent to the KLT basis case, thus, they also suffer from the sample-deficient problem when number of subcarriers are smaller than the *true* optimal modeling order. Last but not least, the shrinkage correlation estimation seems to be unnecessary when doing rank determination, since from the above simulation results, it performs not better than ML

estimation while a heavy-tail problem is discovered, which may influence the accuracy of rank determination. We summarize in Table 4.4 and Table 4.5 the effect of imperfect spatial correlation matrix estimated by ML estimator (4.59) to optimal modeling order found by IMOD algorithm when using KLT, DCT-II or polynomial basis.

Finally, Fig. 4.31 shows the effectiveness of the large aperture of the ULA to resolve user AoAs even when the ASs of different MSs overlap.

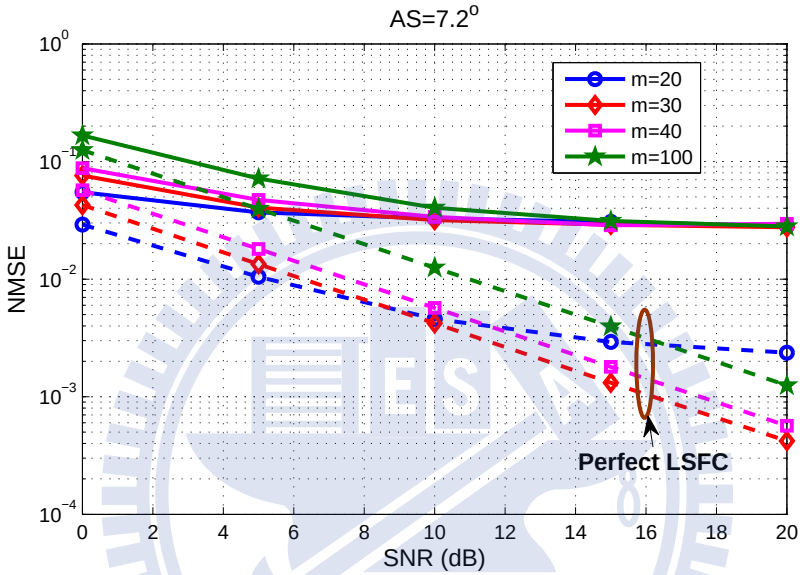

Figure 4.3: MSE performance of the proposed SSFC estimator versus received SNR and modeling order with estimated and perfect LSFC, where AS= 7.2°, and polynomial basis is used.

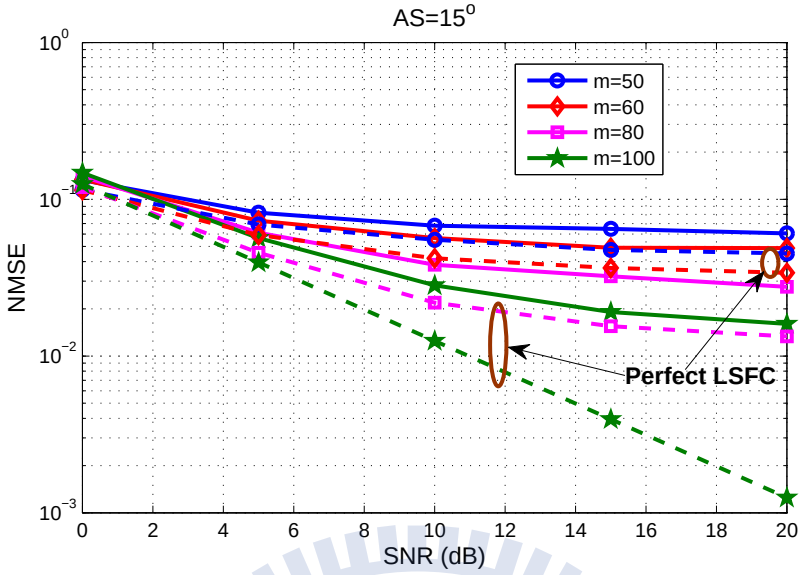

Figure 4.4: MSE performance of the proposed RR SSFC estimator versus received SNR and modeling order with estimated and perfect LSFC, where AS= 15°, and polynomial basis is used.

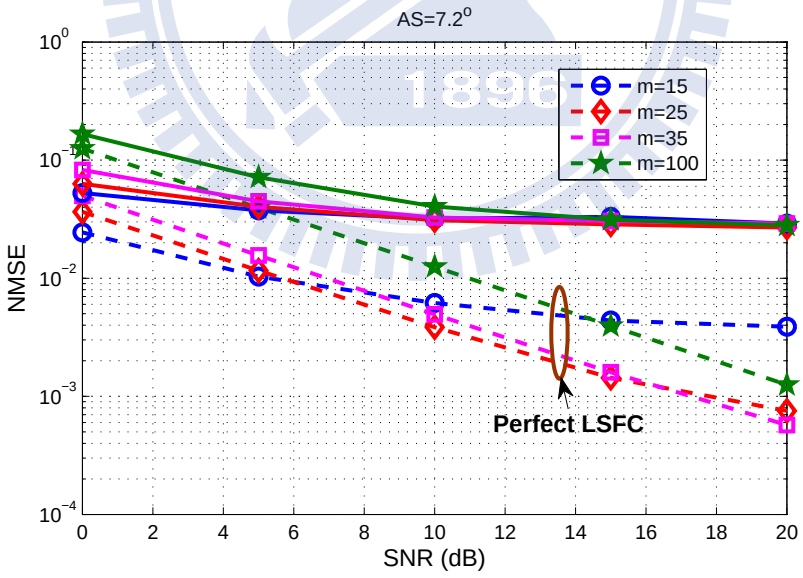

Figure 4.5: MSE performance of the proposed SSFC estimator versus received SNR and modeling order with estimated and perfect LSFC, where AS= 7.2°, and DCT-II basis is used.

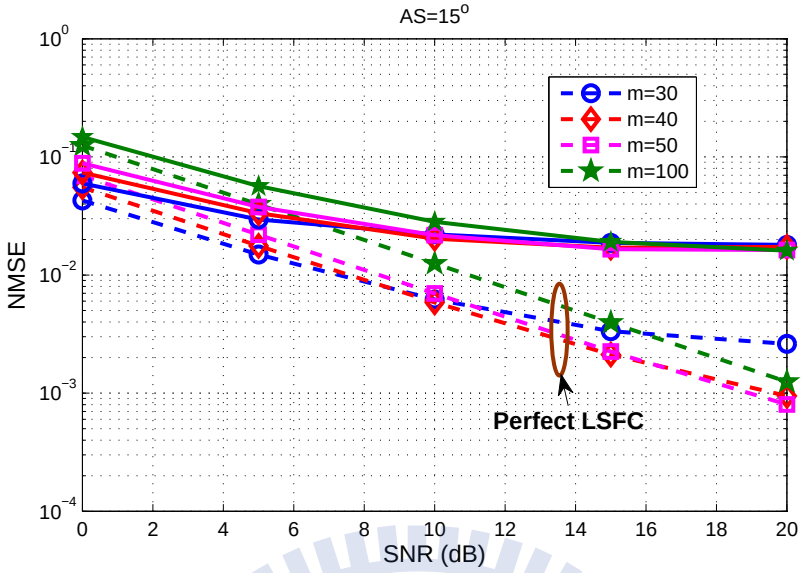

Figure 4.6: MSE performance of the proposed RR SSFC estimator versus received SNR and modeling order with estimated and perfect LSFC, where AS= 15°, and DCT-II basis is used.

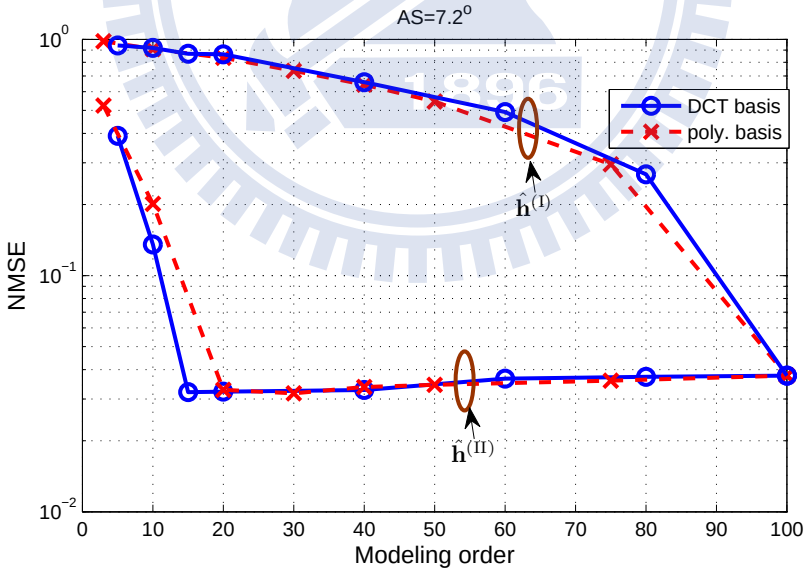

Figure 4.7: MSE performance of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, versus modeling order when using respectively DCT-II and polynomial basis, where AS= 7.2°, SNR= 10 dB, and imperfect LSFC is used.

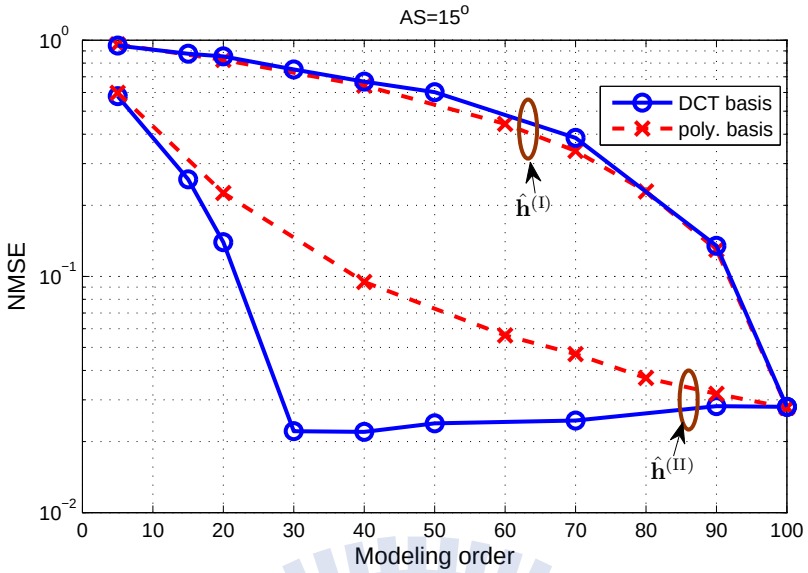

Figure 4.8: MSE performance of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, versus modeling order when using respectively DCT-II and polynomial basis, where AS= 15°, SNR= 10 dB, and imperfect LSFC is used.

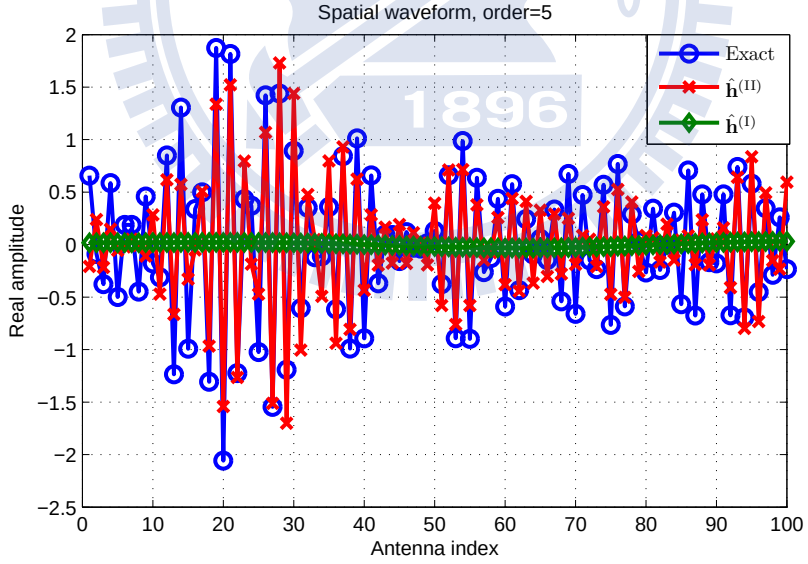

Figure 4.9: Spatial waveform (real part) of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, compared with true (exact) spatial waveform when DCT-II basis being chosen, where AS= 7.2°, modeling order=5, SNR= 10 dB, mean AoA= $\frac{\pi}{21}$, and imperfect LSFC is used.

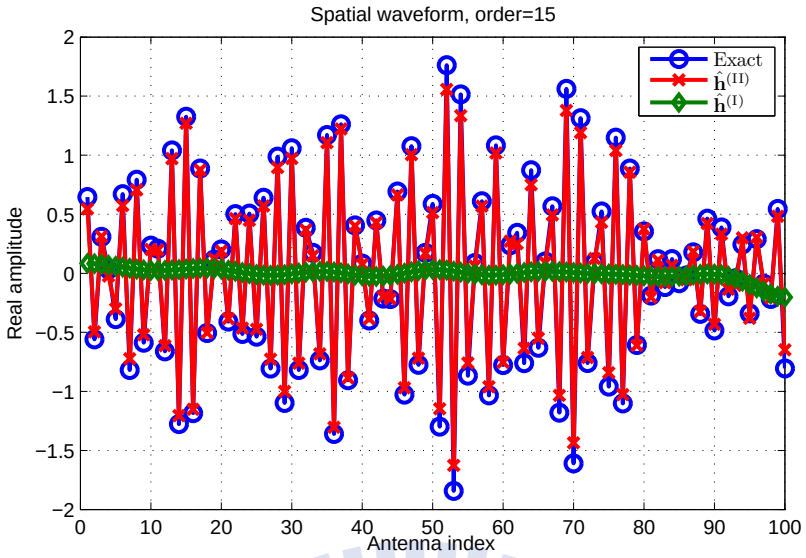

Figure 4.10: Spatial waveform (real part) of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(I)}$ and $\hat{\mathbf{h}}^{(II)}$, compared with true (exact) spatial waveform when DCT-II basis being chosen, where AS= 7.2°, modeling order=15, SNR= 10 dB, mean AoA= $\frac{\pi}{21}$, and imperfect LSFC is used.

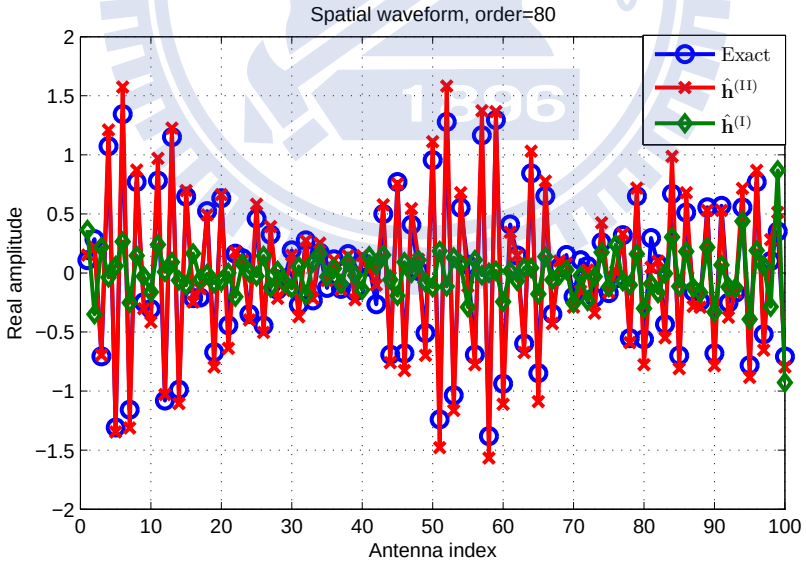

Figure 4.11: Spatial waveform (real part) of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(I)}$ and $\hat{\mathbf{h}}^{(II)}$, compared with true (exact) spatial waveform when DCT-II basis being chosen, where AS= 7.2°, modeling order=80, SNR= 10 dB, mean AoA= $\frac{\pi}{21}$, and imperfect LSFC is used.

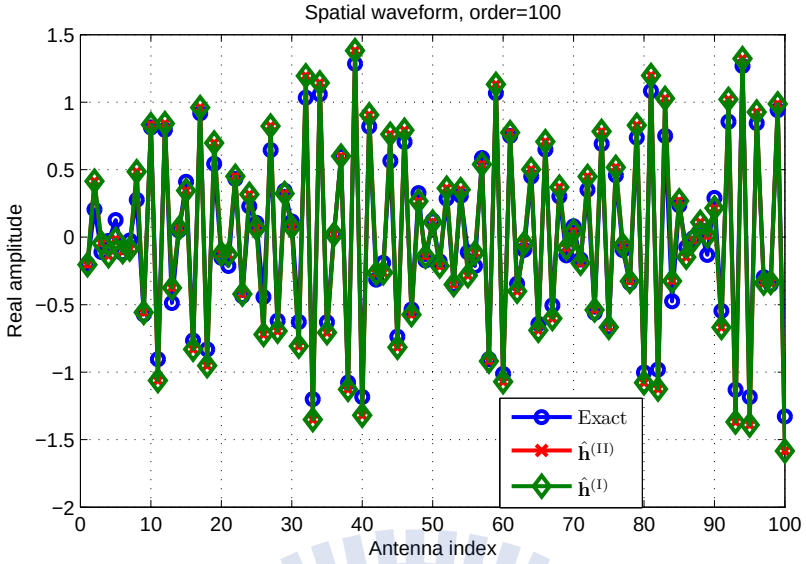

Figure 4.12: Spatial waveform (real part) of the proposed SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, compared with true (exact) spatial waveform when DCT-II basis being chosen, where AS= 7.2°, modeling order=100, SNR= 10 dB, mean AoA= $\frac{\pi}{21}$, and imperfect LSFC is used.

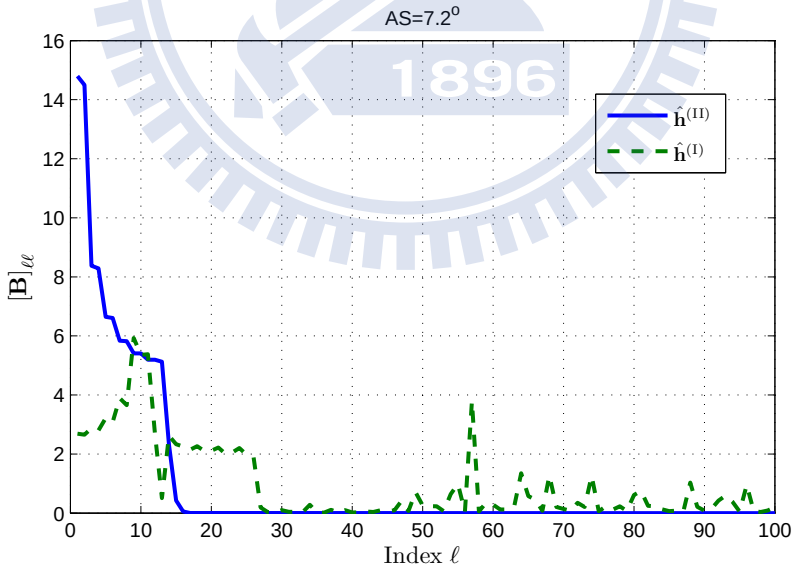

Figure 4.13: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and KLT basis is used.

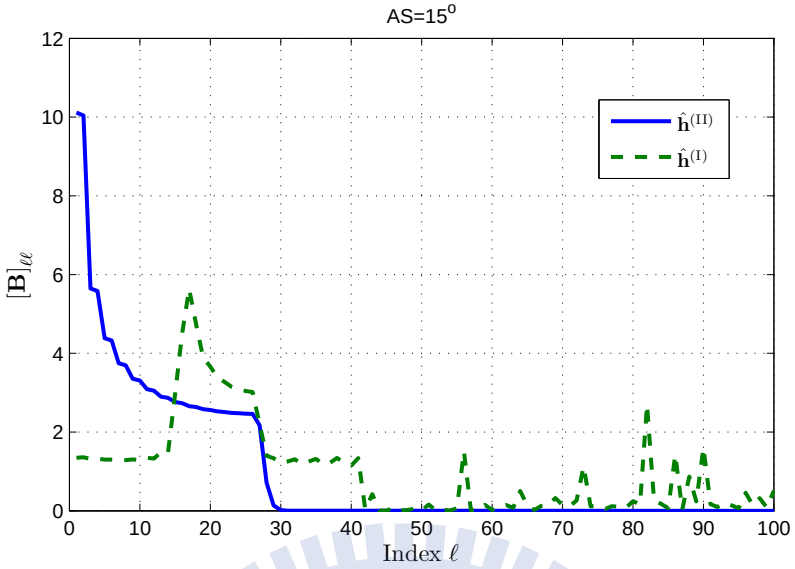

Figure 4.14: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and KLT basis is used.

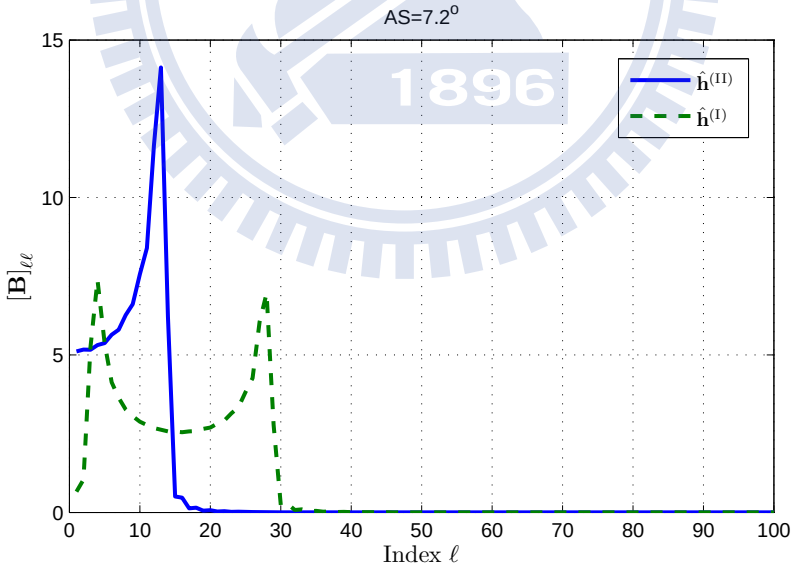

Figure 4.15: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and DCT-II basis is used.

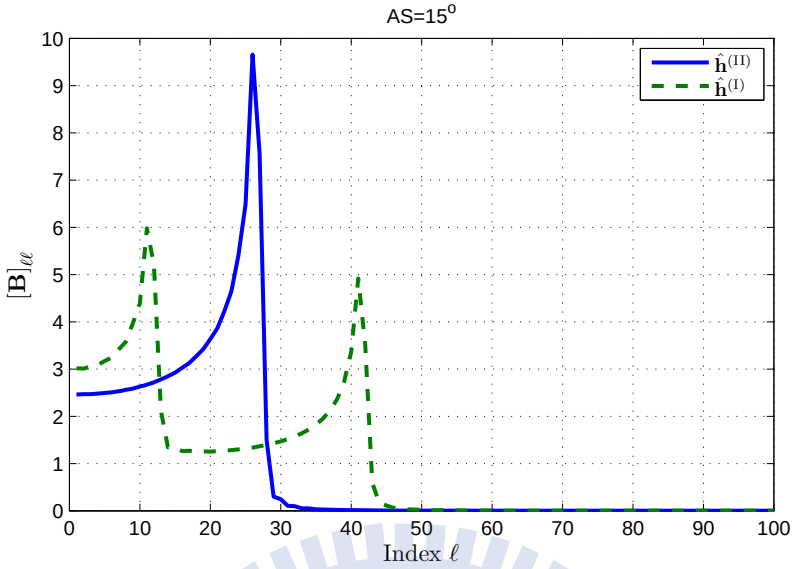

Figure 4.16: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and DCT-II basis is used.

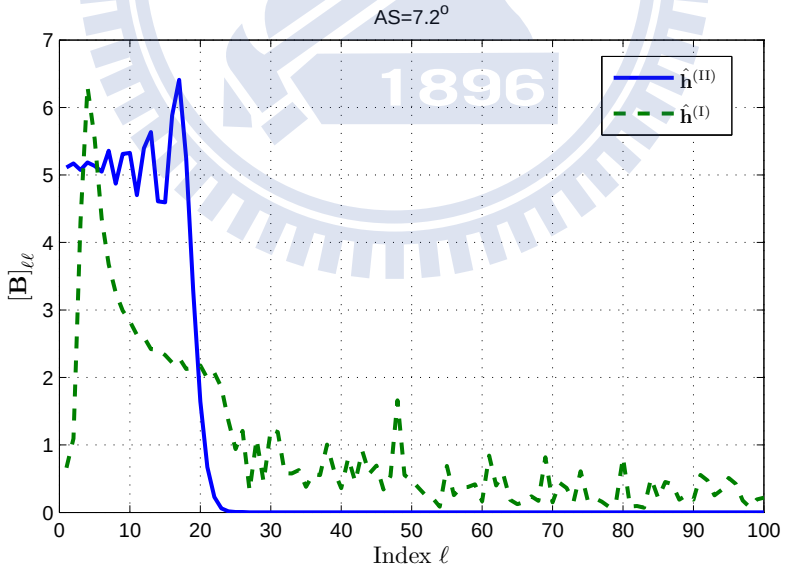

Figure 4.17: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and polynomial basis is used.

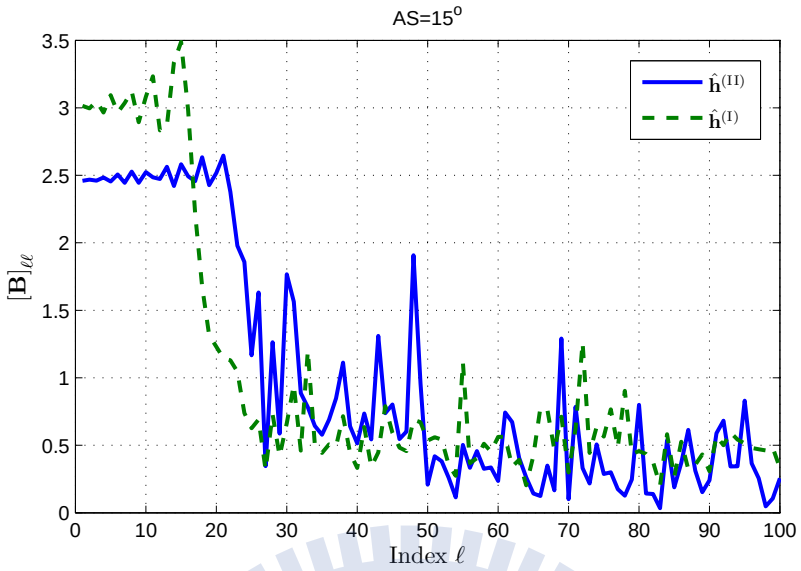

Figure 4.18: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and polynomial basis is used.

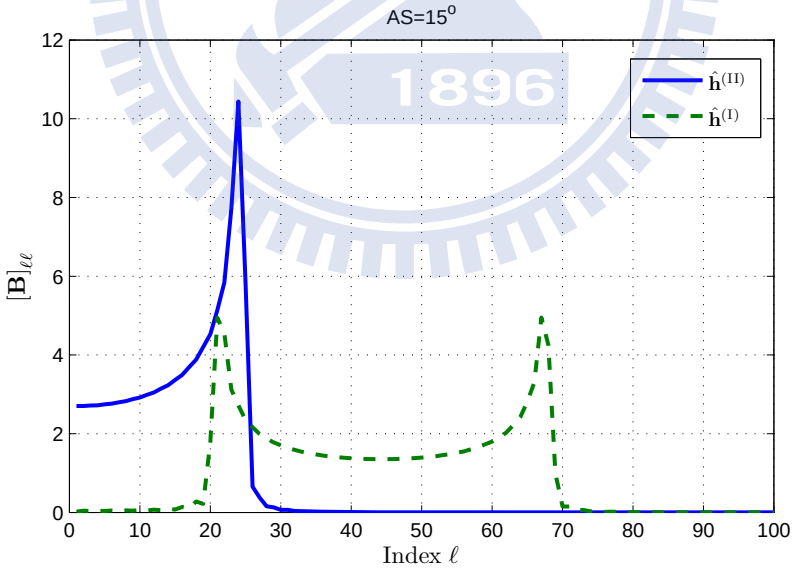

Figure 4.19: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{3\pi}{21}$, SNR= 10 dB and DCT-II basis is used.

Table 4.1: Convergence speed and accuracy of proposed IMOD algorithm for different choice of η assume DCT-II basis is used and SNR= $10dB$.

AS = 15°	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1.0$
$\hat{m}^\star$	32	29	29
Iteration Number	2	2	2
Exhaustive Search	30	30	30
AS = 7.2°	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1.0$
$\hat{m}^\star$	19	16	15
Iteration Number	1	2	1
Exhaustive Search	15	15	15

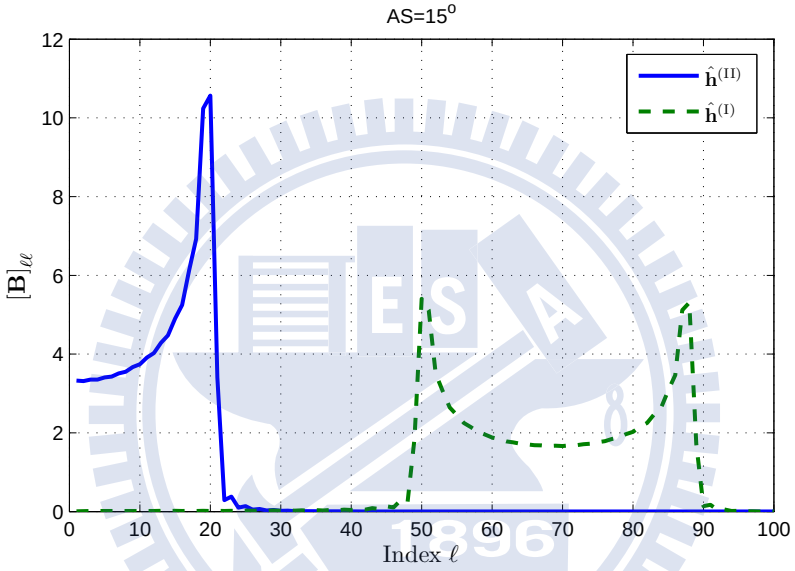

Figure 4.20: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{5\pi}{21}$, SNR= 10 dB and DCT-II basis is used.

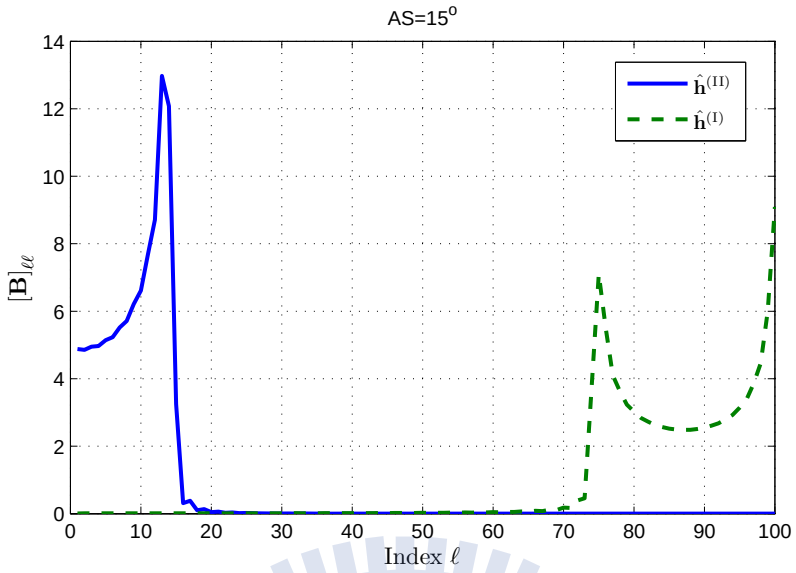

Figure 4.21: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to the SSFC estimators, $\hat{\mathbf{h}}^{(\mathrm{I})}$ and $\hat{\mathbf{h}}^{(\mathrm{II})}$, where AS= 15°, mean AoA= $\frac{7\pi}{21}$, SNR= 10 dB and DCT-II basis is used.

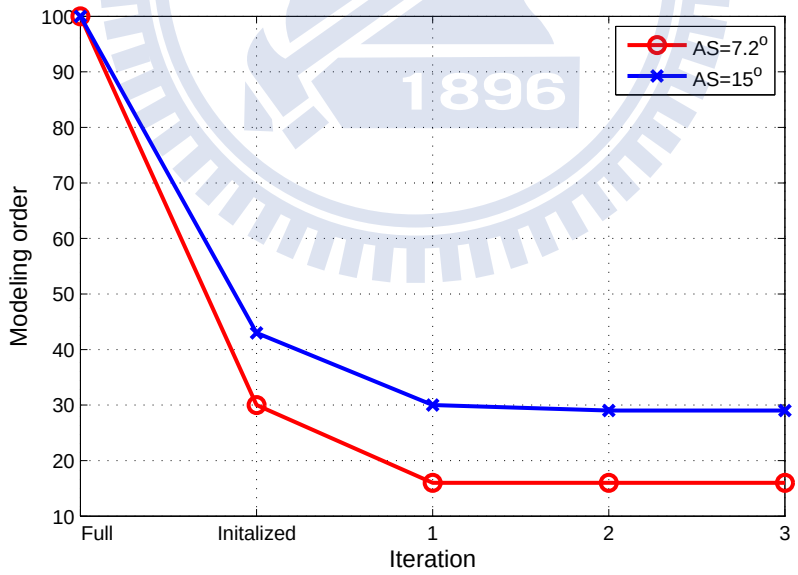

Figure 4.22: Convergence speed of proposed IMOD algorithm. Iteration number "*Full*" represents the full modeling order and "*Initialized*" means the initialization. Assume DCT-II basis is used, SNR= 10 dB and $\eta = 1$.

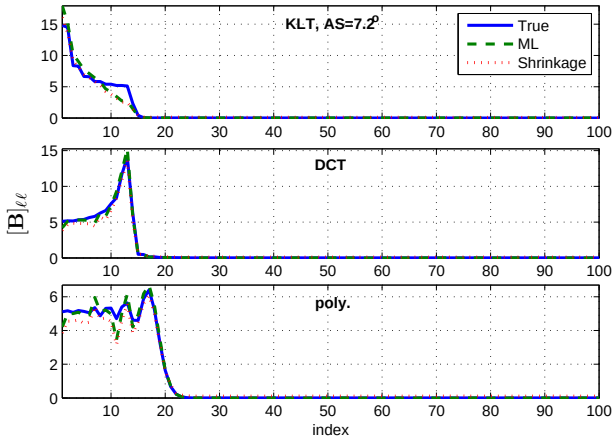

Figure 4.23: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 100. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

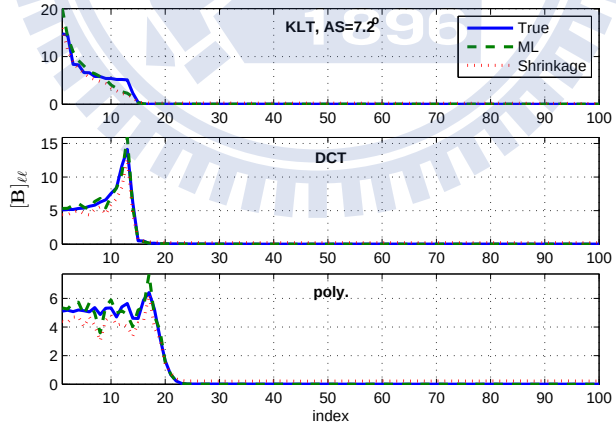

Figure 4.24: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 50. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

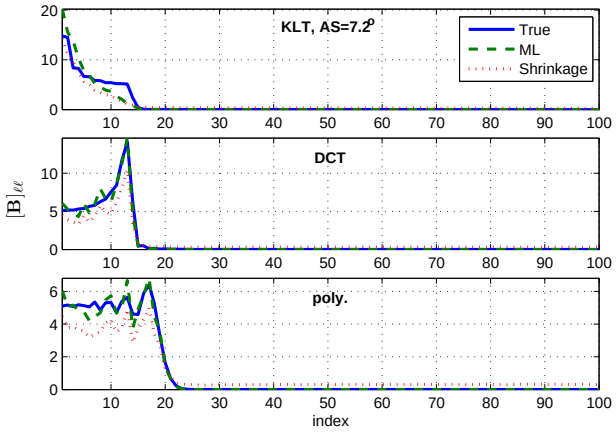

Figure 4.25: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 30. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

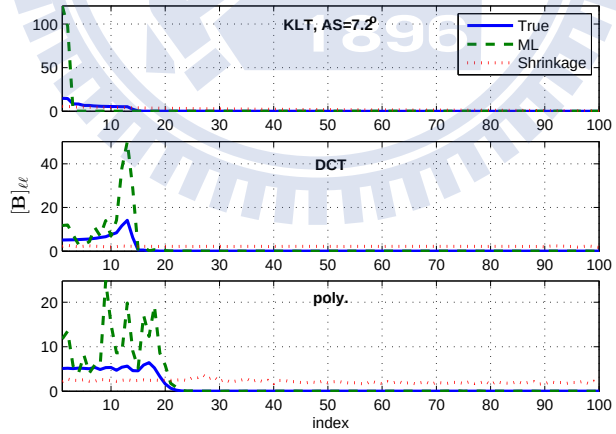

Figure 4.26: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 7.2°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 2. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

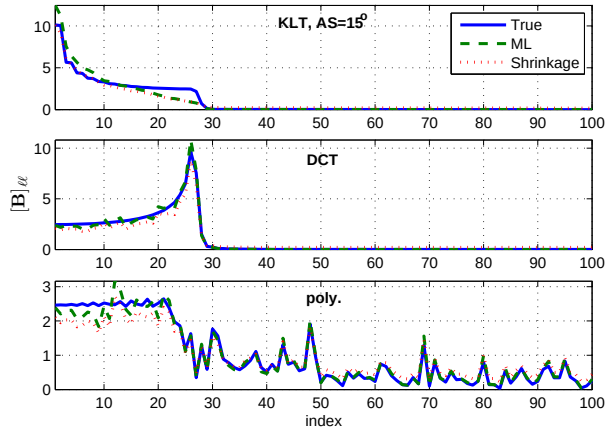

Figure 4.27: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 100. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

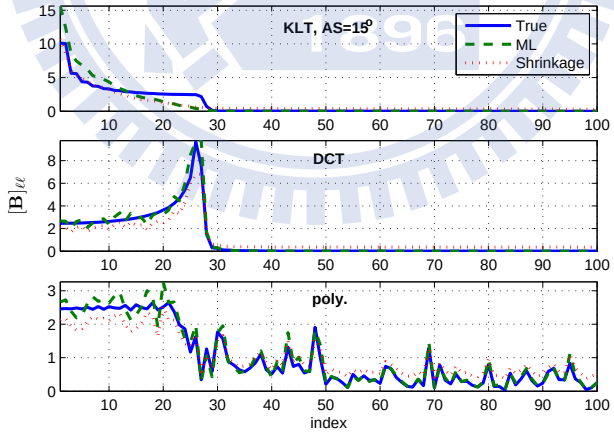

Figure 4.28: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 50. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

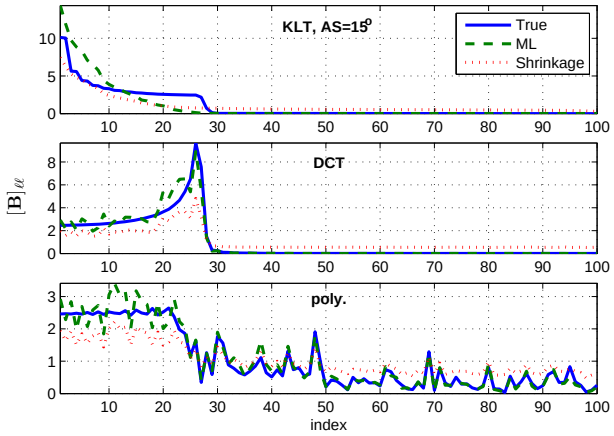

Figure 4.29: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 30. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

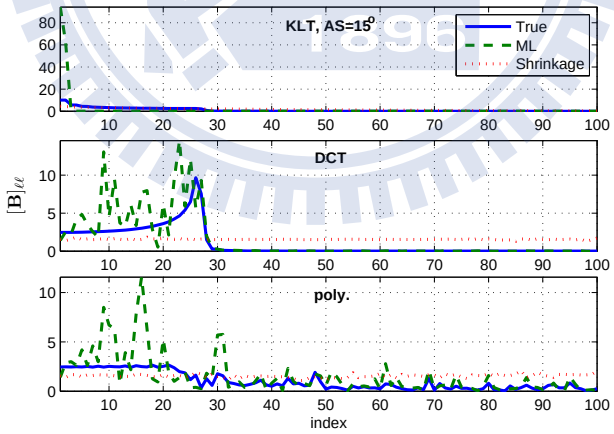

Figure 4.30: Diagonal distribution of the bias matrix, $\mathbf{B}$, with respect to KLT, DCT-II, and polynomial basis, where AS= 15°, mean AoA= $\frac{\pi}{21}$, SNR= 10 dB and number of subcarriers is 2. Assume imperfect spatial correlation matrix estimated by ML or shrinkage [20] method is used here.

Table 4.2: Effect of large system to optimal modeling order found by IMOD algorithm given DCT-II or KLT basis is used. Assume SNR= $10dB$, AS= $7.2°$ and $\eta = 1$.

M	30	50	100	200	300	400
$\hat{m}^\star$(DCT)	6	9	15	28	40	54
$\hat{m}^\star$(KLT)	6	9	15	28	40	53
$\frac{\hat{m}^\star(\text{KLT})}{M}$	0.2	0.18	0.15	0.14	0.133	0.1325

Table 4.3: Effect of large system to optimal modeling order found by IMOD algorithm given DCT-II or KLT basis is used. Assume SNR= $10dB$, AS= $15°$ and $\eta = 1$.

M	30	50	100	200	300	400
$\hat{m}^\star$(DCT)	10	16	29	55	81	107
$\hat{m}^\star$(KLT)	10	15	28	54	80	106
$\frac{\hat{m}^\star(\text{KLT})}{M}$	0.333	0.3	0.28	0.27	0.2667	0.265

Table 4.4: Effect of imperfect spatial correlation matrix estimated by ML estimator (4.59) to optimal modeling order found by IMOD algorithm when using KLT, DCT-II or polynomial basis. Assume that SNR= $10dB$, AS= $7.2°$, $\eta = 1$ and that there are total n subcarriers.

n	2	10	30	50	100	perfect
$\hat{m}^\star$(KLT)	2	10	14	15	15	15
$\hat{m}^\star$(DCT)	15	15	15	15	15	15
$\hat{m}^\star$(Poly.)	22	21	21	21	21	21

Table 4.5: Effect of imperfect spatial correlation matrix estimated by ML estimator (4.59) to optimal modeling order found by IMOD algorithm when using KLT, DCT-II or polynomial basis. Assume that SNR= $10dB$, AS= $15°$, $\eta = 1$ and that there are total n subcarriers.

n	2	10	30	50	100	perfect
$\hat{m}^\star$(KLT)	2	10	21	29	29	29
$\hat{m}^\star$(DCT)	29	29	29	29	29	29
$\hat{m}^\star$(Poly.)	100	100	100	100	100	100

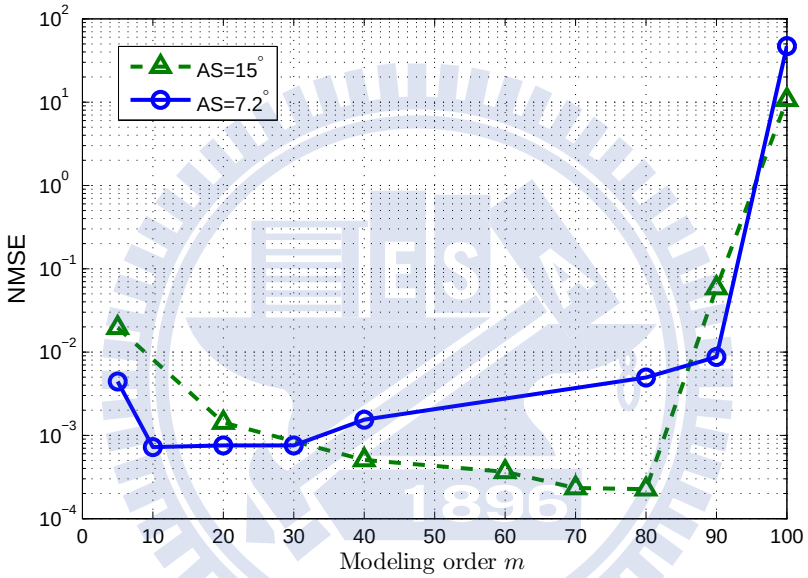

Figure 4.31: Performance of the estimated mean AoA (in radians) with different ASs versus modeling order, where SNR= 10 dB, and polynomial basis is used.

Chapter 5

Closed-Loop Transceiver Design

5.1 Distributed Massive MIMO

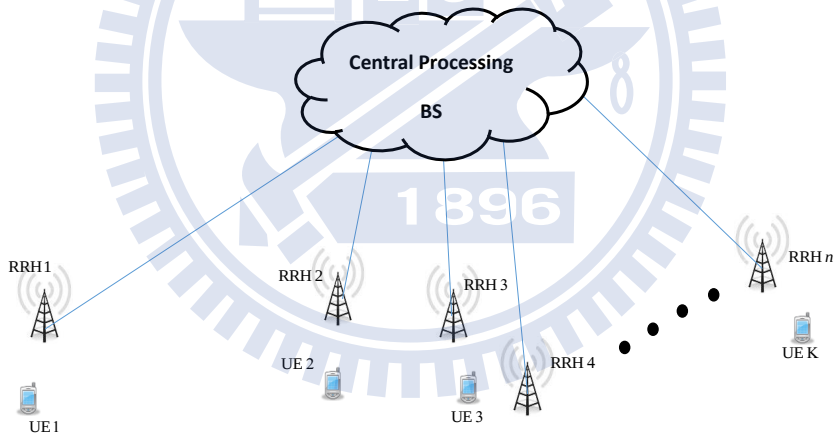

Figure 5.1: Illustration of Distributed Massive MIMO System.

We now present closed-loop transceiver design for large-scale MIMO systems in both TDD and FDD modes based on the proposed LSFC and SSFC estimators, where a same frequency band is shared by both UL and DL transmission for TDD mode. First, we generalize the co-located multiantenna scenario to one that distributes its M BS antennas into clusters (remote radio heads, RRHs) within the coverage area. The distributed

antennas are linked via optical fiber-based backhaul to the central BS (CBS) responsible for joint signal processing [4]; this scenario is illustrated in Fig. 5.1. Nevertheless, the proposed application is different from [4] in twofold: i) A total of N clusters are assumed in the serving cell of which the nth cluster is equipped with $M_n \gg K$ antennas; ii) The selection of clusters and antennas within is employed to reduce computational complexity. This is especially true for the systems operating in FDD mode for the reason that a great amount of CSI must be fed back for higher sum rate.

As a result, each cluster serves only its nearby rather than all users in the cell. Denote by K_n the number of UEs served by RRH n. To this end, knowledge of LSFCs may be useful for selecting serving RRHs [4]. As for the transmit antenna selection (TAS) for an RRH, we consider an MU-MIMO system employing maximum ratio transmission (MRT) or zero-forcing (ZF) precoding. Denote by $\mathcal{S}$ a set of selected transmit antennas. A decremental antenna selection scheme aims at removing antennas one by one based on either the generalized norm- (GNS) or capacity-based selection (CS) until reaching a prescribed number of active antennas. Among them, the CS turns off transmit antenna s based on

$$s = \arg\max_{i \in \mathcal{S}} R_D\left(\mathcal{S} \setminus \{i\}\right), \tag{5.1}$$

where $R_D(\mathcal{S})$ stands for the sum rate with selected antennas $\mathcal{S}$, and GNS removes

$$s = \arg\min_{i \in \mathcal{S}} \frac{\|\mathbf{D}_{\boldsymbol{\beta}}^{-1}\mathbf{u}_i\|^2}{|\mathcal{S}| - \|\mathbf{D}_{\boldsymbol{\beta}}^{-\frac{1}{2}}\mathbf{u}_i\|^2}, \tag{5.2}$$

with $\mathbf{u}_i$ being the SSFC channel seen by antenna i.

Lemma 5.1.1. *For an RRH n performing decremental TAS, if all its served UEs are closely located, then both the GNS and CS converges to*

$$s = \arg\min_{i \in \mathcal{S}_n} \|\mathbf{u}_i\| \tag{5.3}$$

almost surely as $M_n \to \infty$, where $\mathcal{S}_n$ is a selected set.

Proof. It follows directly from Section 2.4.2. □

While [1] has shown that the performance of MRT asymptotically achieves that of ZF and MMSE precoding, shown in the following lemma, for practical array size M, MRT is far from achieving ZF precoder.

Lemma 5.1.2. *For $\frac{M}{K} < \infty$, the SINR performance and thus the achievable sum rate of MRT is strictly inferior than ZF precoder.*

Proof. See Appendix G. □

5.2 Uplink Pilot Design

To minimize the required training period, we let $T = K$ and thus have square $\mathbf{P}$'s. Besides, as the derivation of the aforementioned estimators relies on the use of mutually orthogonal uplink pilot sequences for their optimality, we will focus on the design of such category of pilots [29]. With fixed transmission power, the MSE performance of the estimators using different orthogonal pilots are shown to the same, provided that the pilot power $\|\mathbf{p}_k\|^2$ is fixed, in Section 3.3 and *Theorem 4.4.1*. Therefore, the main objective of pilot sequence design falls into the minimization of computational complexity.

Definition 5.2.1 ([30, Ch. 2])**.** *A Hadamard matrix is a square matrix whose rows (or columns) are mutually orthogonal and of ± 1 entries. It is conjectured that a Hadamard matrix of order $4n$ exists $\forall n \in \mathbb{N}$.*

Remark 22 (Hadamard Matrix-Based Pilots)**.** *By choosing a Hadamard matrix as the pilot matrix $\mathbf{P}$, the computation effort can be reduced significantly due to the fact that the calculation of $\mathbf{Y}\mathbf{p}_k$ in (3.7), (4.10), and (4.9) involves only column additions and subtractions of $\mathbf{Y}$.*

Remark 23 (Diagonal Matrix-Based Pilots)**.** *An alternative choice of pilot matrix is it being diagonal, specifically,* $\mathbf{P} = \mathrm{Diag}(s_1, \cdots, s_K)$*, which is able to simplify (3.7) as*

$$\hat{\beta}_k = \frac{s_k \mathbf{y}_k^H \mathbf{y}_k s_k - M s_k^2}{M s_k^4} = \frac{\|\mathbf{y}_k\|^2 - M}{M s_k^2}, \tag{5.4}$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_K]$*. This estimator coincides with our intuition in that the instantaneous received signal strength minus the noise power,* $\|\mathbf{y}_k\|^2 - M$*, is approximately equal to the strength of the desired signal and thus fairly reflects the gain provided by large-scale fading if it is divided by* $M s_k^2$*, the total power emitted by user* k *times the number of copies received at the BS. On the other hand, for SSFC estimation:*

$$\begin{aligned}
\hat{\mathbf{c}}_k &= \frac{1}{\sqrt{\hat{\beta}_k} s_k} \mathbf{Q}_m^H \mathbf{W}^H(\hat{\phi}_k) \mathbf{y}_k \\
\hat{\phi}_k &= \underset{\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]}{\arg\max}\ s_k^2 \mathbf{y}_k^H \mathbf{W}(\phi) \mathbf{Q}_m \mathbf{Q}_m^H \mathbf{W}(\phi)^H \mathbf{y}_k \\
&= \underset{\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]}{\arg\max}\ \mathbf{y}_k^H \mathbf{W}(\phi) \mathbf{Q}_m \mathbf{Q}_m^H \mathbf{W}(\phi)^H \mathbf{y}_k
\end{aligned}$$

Remark 24. *Although diagonal matrix-based pilots give lower computational burden then the ones being Hadamard, the requirement that an MS needs to transmit all pilot power in a time slot to achieve the same performance shows a risk of disobeying the maximum user output power constraint. For instance, [25] defines the maximum output power for any transmission bandwidth within the channel bandwidth to be* 23 *dBm. In other words, the decision of a suitable uplink pilot pattern is a trade-off between the computational complexity and maximum user output power.*

5.3 TDD Mode

Throughout this section, we devise a transceiver procedure for the massive MIMO operating in TDD mode assuming UL-DL channel reciprocity holds [1]; see also Fig. 5.2 for illustration. The downlink channel is thus $\mathbf{D}_{\boldsymbol{\beta}}^{\frac{1}{2}} \mathbf{H}^T$ for the reason that UL and DL transmission are via the same frequency band.

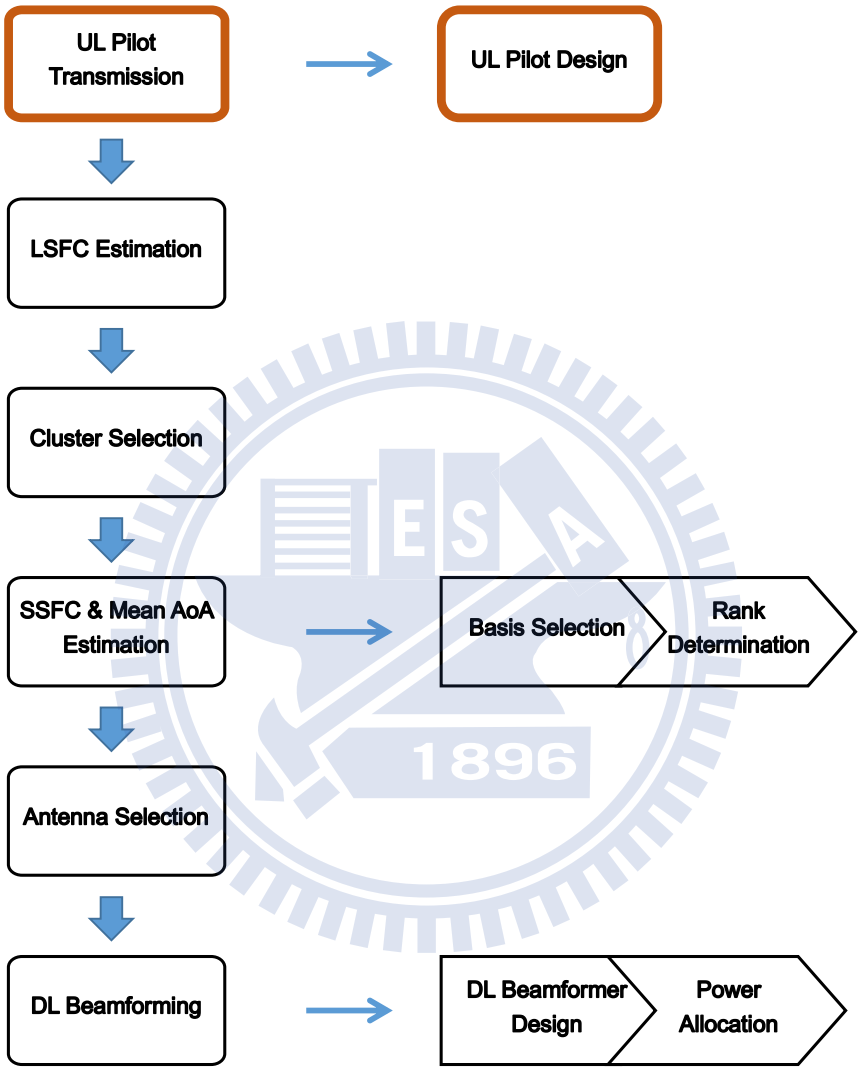

Figure 5.2: The flow chart of closed-loop transceiver design in TDD mode. Colored blocks are done by MSs while others done by BS.

T1 All K UEs broadcast their mutually orthogonal pilot sequences to all N RRHs simultaneously.

T2 All N RRHs estimate K LSFCs with (3.7) and forward these estimates to the CBS.

T3 The CBS performs cluster selection algorithm given in [4] with the gathered estimates.

T4 Each selected RRH chooses a basis (DCT or polynomial-based) for RR SSFC estimation.

T5 Each selected RRH n determines the modeling order for all K_n serving UEs via the IMOD algorithm and obtains their RR channel and mean AoA estimates using (4.10)–(4.11) with this order.

T6 RRHs determine whether or not to perform antenna selection with (5.3).

T7 Each RRH performs either MRT or ZF precoding with LSFC and SSFC estimates obtained in T2 and T5, respectively.

T8 Each RRH doing power allocation via compensating the LSFCs for all serving UEs.

5.4 FDD Mode

As suggested in [1] and [16], massive MIMO operating in the FDD mode is unrealistic due to the fact that: i) a DL pilot of duration equals to the RRH antenna number is called for; ii) a vector CSI of the same length needs to be fed back from each UE. In this subsection, we develop an RR DL SSFC estimator and its corresponding pilot design that makes significant amount of feedback reduction possible.

While shadow fading is frequency-invariant [43]– [46], unlike in the TDD mode, there is no reciprocity between UL and DL pathloss in the FDD mode. Fortunately, a translation can be done to obtain DL pathloss with the estimated UL ones.

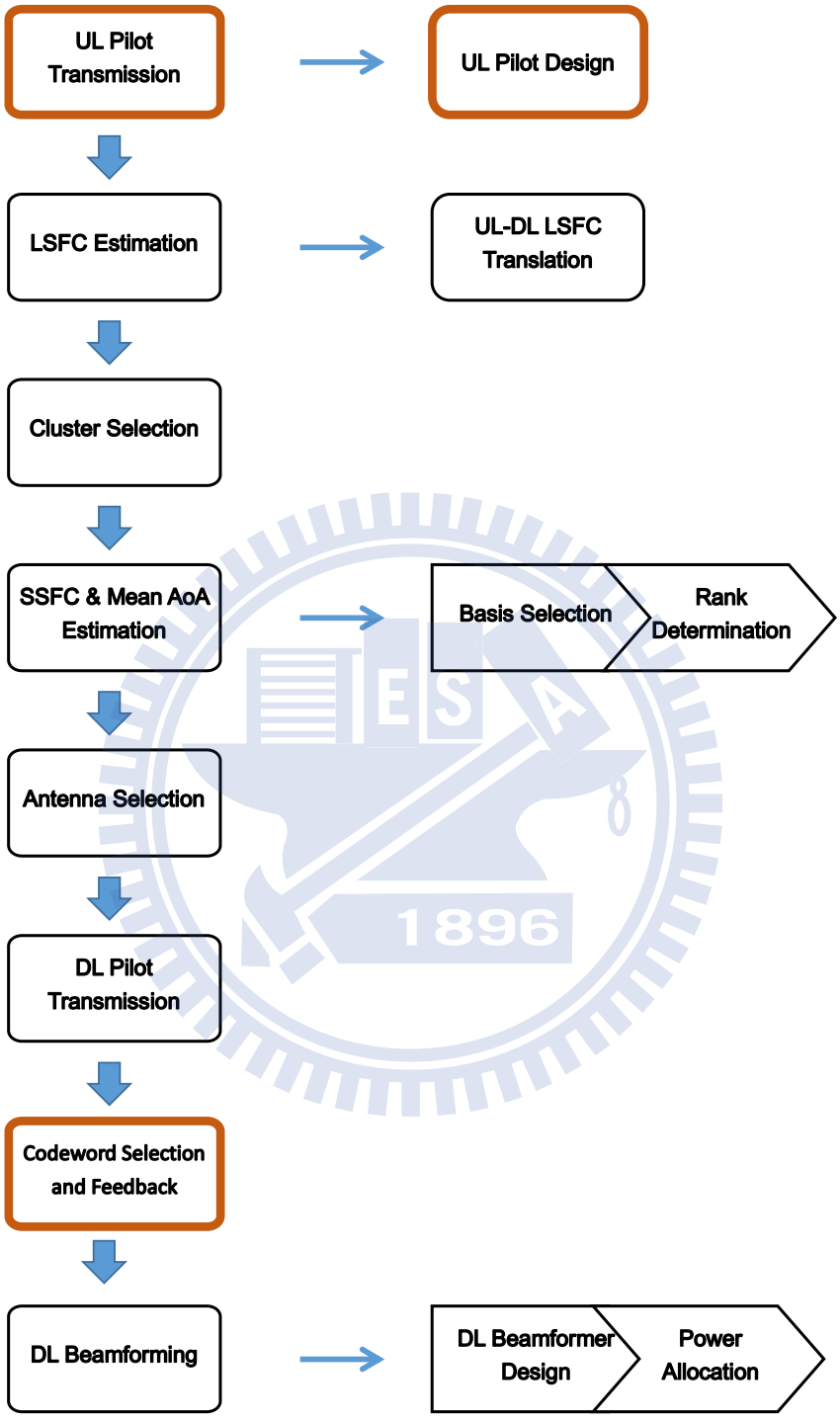

Figure 5.3: The flow chart of closed-loop transceiver design in FDD mode. Colored blocks are done by MSs while others done by BS.

Remark 25. *Suppose the frequency bands used for UL and DL transmission are respectively ω_{UL} and ω_{DL}. The frequency dependence of pathloss is approximately characterized by*

$$\frac{\mathrm{PL}(\omega_{UL})}{\mathrm{PL}(\omega_{DL})}[dB] \approx -20\log\left(\frac{\omega_{UL}}{\omega_{DL}}\right). \tag{5.5}$$

Therefore, for a same link using different bands,

$$\frac{\beta^{UL}}{\beta^{DL}} \approx \frac{\omega_{DL}^2}{\omega_{UL}^2}. \tag{5.6}$$

Besides, the fact that UE k's UL mean AoA ϕ_k and SSFC modeling order m_k equal to the DL counterparts enables RR DL SSFC estimation and even feedback reduction. Specifically, UE k receives the pilot $\boldsymbol{\mathcal{P}}_k \in \mathbb{C}^{M_n \times T}$ emitted by its serving RRH n:

$$\begin{aligned}\mathbf{x}_k^H &= \sqrt{\beta_k^{\mathrm{DL}}}\mathbf{g}_k^H\boldsymbol{\mathcal{P}}_k + \mathbf{n}_k^H \\ &= \sqrt{\beta_k^{\mathrm{DL}}}\left(\mathbf{b}_k^H\mathbf{Q}_{m_k}^H\mathbf{W}^H(\phi_k) + \boldsymbol{\epsilon}_k^H\right)\boldsymbol{\mathcal{P}}_k + \mathbf{n}_k^H,\end{aligned}$$

where $\mathbf{g}_k^H \approx \mathbf{b}_k^H\mathbf{Q}_{m_k}^H\mathbf{W}^H(\phi_k)$ is the RR approximation of DL channel $\mathbf{g}_k^H$, $\boldsymbol{\epsilon}_k$ the modeling error, and noise $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}_T, \mathbf{I}_T)$, and performs LS channel estimation:

$$\hat{\mathbf{b}}_k = \frac{1}{\sqrt{\beta_k^{\mathrm{DL}}}}\left(\boldsymbol{\mathcal{P}}_k^H\mathbf{W}(\phi_k)\mathbf{Q}_{m_k}\right)^\dagger\mathbf{x}_k. \tag{5.7}$$

However, as β_k^{DL} and ϕ_k are not known to UE k, the following DL pilot structure is proposed with $T = m_k$:

$$\boldsymbol{\mathcal{P}}_k = \frac{1}{\sqrt{\beta_k^{\mathrm{DL}}}}\mathbf{W}(\phi_k)\mathbf{Q}_{m_k} \tag{5.8}$$

which results in $\hat{\mathbf{b}}_k = \mathbf{x}_k$, simply the received signal that requires minimal extra computation. Note that in (5.8), the multiplication of $\mathbf{W}(\phi_k)$ can be regarded as RRH beamforms DL pilot (or signal) to the ϕ_k direction for better signal power concentration. This RR DL SSFC estimate potentially reduce the amount of CSI needed to be fed back since vector $\mathbf{g}_k$ of dimension M is represented by length-m $\hat{\mathbf{b}}_k$. While for practical systems as LTE, vector quantization (or codeword selection) is done to the feedback

vector [28], we assume in the following that the vector is fed back without quantization and, furthermore, via a noise-free channel.

F1 All K UEs broadcast their mutually orthogonal pilot sequences to all N RRHs simultaneously.

F2 All N RRHs estimate K UL LFSCs with (3.7) and translate them into

$$\hat{\beta}^{\mathrm{DL}} = \hat{\beta}^{\mathrm{UL}} \frac{\omega_{\mathrm{UL}}^2}{\omega_{\mathrm{DL}}^2}, \tag{5.9}$$

information intended for the CBS.

F3 The CBS performs cluster selection algorithm given in [4] with the gathered DL LSFC estimates.

F4 Each selected RRH chooses a basis (DCT or polynomial-based) for RR UL SSFC estimation.

F5 Each selected RRH determines the modeling order $\hat{m}_k^\star$ for each served UE k via the IMOD algorithm and obtains its RR UL SSFC and mean AoA estimates using (4.10)–(4.11) with this order.

F6 RRHs determine whether or not to perform antenna selection with (5.3).

F7 Each RRH beamforms the DL pilot intended for a served UE to AoD $\hat{\phi}_k$ obtained in F4 and compensate for the DL LSFC $\hat{\beta}_k^{\mathrm{DL}}$ in F2, i.e., (5.8).

F8 All served UEs feed their received signal $\mathbf{x}_k$ back to serving RRHs.

F9 RRHs recover DL SSFC estimates via

$$\hat{\mathbf{g}}_k^H = \mathbf{x}_k^H \mathbf{Q}_{\hat{m}_k^\star}^H \mathbf{W}^H(\hat{\phi}_k). \tag{5.10}$$

F10 Each RRH performs either MRT or ZF precoding with DL LSFC and SSFC estimates obtained in F2 and F9, respectively.

F11 Each RRH doing power allocation via compensating the LSFCs for all serving UEs.

5.5 Numerical Results and Discussion

In this section, we present the simulation results about sum rate performance of antenna selection algorithms (CS and GNS) and different precoding methods (ZF and MRT). Here we denote M_T the total number of RRH antennas and M_F the number of selected RRH antennas. From Fig. 5.4-5.5, we can see that when M_F is about more than two times larger than K with K the number of MSs, GNS algorithm has almost the same sum rate performance as CS as we have discussed in *Lemma 5.1.1*. In other words, we only have to use low complexity GNS algorithm to do TAS in massive MIMO system. Moreover, both figures tell us the fact that when M_F is finite, or $\frac{M_F}{K} < \infty$, the achievable sum rate of MRT is strictly inferior than ZF precoder, as we have shown in *Lemma 5.1.2*.

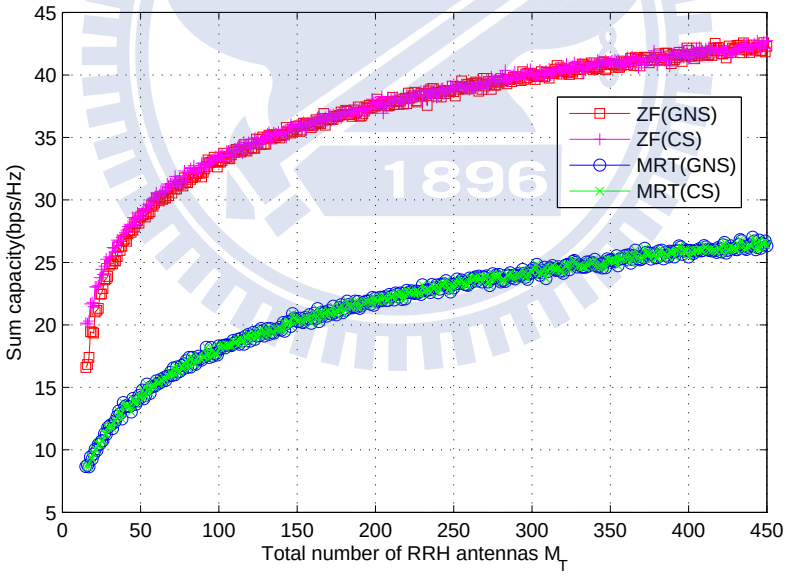

Figure 5.4: Sum rate performance of antenna selection algorithms (CS and GNS) versus M_T provided that ZF and MRT precoding are used, where $M_T = 3M_F$ represents the number of RRH antennas, M_F is the number of selected antennas, number of users is 4, and SNR= 11.76 dB.

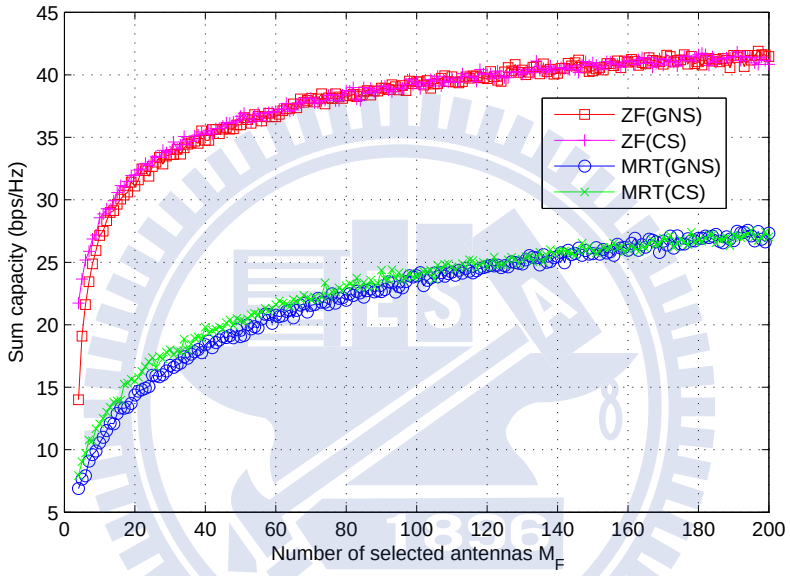

Figure 5.5: Sum rate performance of antenna selection algorithms (CS and GNS) versus M_F provided that ZF and MRT precoding are used, where $M_T = 200$ represents the number of RRH antennas, M_F is the number of selected antennas, number of users is 4, and SNR= 11.76 dB.

Chapter 6

Conclusion

Due to the effect of noise reduction in massive MIMO systems, a novel LSFC estimator for both i.i.d. and spatial correlated channel is proposed. This estimator can extended to the one considering multiple pilot blocks to improve performance. Advantages of the proposed LSFC estimator includes: low complexity and no prior knowledge of SSFCs and spatial correlation is needed. Furthermore, it can be shown that the proposed LSFC estimator has asymptotically zero-MSE and hence approaches to the MMSE estimator when the number of BS antennas is large. The fact that it even significantly outperforms the conventional LSFC estimator with known SSFCs is revealed by the simulation results.

By using the estimated LSFCs, a estimator incorporating individual estimation of mean AoA and SSFC combined with rank-reduced channel model is also presented. The estimated mean AoA is helpful for downlink beamforming, while the RR characteristics enables us to reduce the amount of feedback overhead, pilot dimension and hence pilot transmission time in FDD mode. We analysis the effect of modeling order on MSE performance of the proposed estimator by mathematical approaches and find out that the "bias matrix" is an important metric in modeling order selection. Moreover, we present some candidate for basis selection as KLT, DCT, and polynomial basis, after connecting the RR model to image signal processing, we show that the DCT-II basis is a practical, low computational complexity and near-optimal choice owing to the so-called energy

compaction property. Eventually, IMOD algorithm is devised to determine the optimal modeling order used in the SSFC estimator, it offers a low computational-complexity approach to quickly determine the optimal modeling order (or, rank indicator) for each user as shown in our simulation results.

Numerical results show the effectiveness of the proposed estimator to enhance accuracy by reducing the number of parameters needed to be estimated and that an optimal modeling order is related to the value of AS and SNR. In addition, massive MIMO system is shown to be helpful in terms of reducing the NMSE of the proposed SSFC estimator. After that, with the aid of the large aperture offered by a massive MIMO BS, a precise estimation of mean AoA is seen, even when the AS is not small.

Finally, we present a closed-loop transceiver design of distributed massive MIMO system for both TDD and FDD mode. The transceiver combines the estimators proposed in the previous sections with a practical MIMO system including antenna selection, cluster selection and transmit beamforming. With the aid of massive number of antennas, the antenna selection metric has been simplified a lot even if we consider the LSFC and spatial correlation into this metric. Finally, we show that in FDD mode, all MSs need to do is to feedback their received signal instead of doing channel estimation and a limited feedback effect can help us to construct a smaller codebook.

Appendix A

Proof of Theorem 3.2.1

Lemma 3.3.1 implies that if

$$\limsup_{M\to\infty} \sup_{1\le i,j\le K} \|\mathbf{\Phi}_i^{\frac{1}{2}}\mathbf{\Phi}_j^{\frac{1}{2}}\|_2 < \infty, \tag{A.1}$$

we have

$$\frac{1}{M}\tilde{\mathbf{H}}^H\mathbf{A}\tilde{\mathbf{H}} = \frac{1}{M}\begin{bmatrix} \tilde{\mathbf{h}}_1^H\mathbf{\Phi}_1^{\frac{1}{2}}\mathbf{\Phi}_1^{\frac{1}{2}}\tilde{\mathbf{h}}_1 & \cdots & \tilde{\mathbf{h}}_1^H\mathbf{\Phi}_1^{\frac{1}{2}}\mathbf{\Phi}_K^{\frac{1}{2}}\tilde{\mathbf{h}}_K \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{h}}_K^H\mathbf{\Phi}_K^{\frac{1}{2}}\mathbf{\Phi}_1^{\frac{1}{2}}\tilde{\mathbf{h}}_1 & \cdots & \tilde{\mathbf{h}}_K^H\mathbf{\Phi}_K^{\frac{1}{2}}\mathbf{\Phi}_K^{\frac{1}{2}}\tilde{\mathbf{h}}_K \end{bmatrix} \xrightarrow{a.s.} \mathbf{I}_K, \tag{A.2}$$

and if $\limsup_{M\to\infty} \sup_{1\le i\le K} \|\mathbf{\Phi}_i^{\frac{1}{2}}\|_2 < \infty$, or equivalently

$$\limsup_{M\to\infty} \sup_{1\le i\le K} \|\mathbf{\Phi}_i\|_2 < \infty, \tag{A.3}$$

$$\frac{1}{M}\tilde{\mathbf{H}}^H\tilde{\mathbf{\Phi}}^H\mathbf{N} = \frac{1}{M}\begin{bmatrix} \tilde{\mathbf{h}}_1^H\mathbf{\Phi}_1^{\frac{1}{2}}\mathbf{n}_1 & \cdots & \tilde{\mathbf{h}}_1^H\mathbf{\Phi}_1^{\frac{1}{2}}\mathbf{n}_K \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{h}}_K^H\mathbf{\Phi}_K^{\frac{1}{2}}\mathbf{n}_1 & \cdots & \tilde{\mathbf{h}}_K^H\mathbf{\Phi}_K^{\frac{1}{2}}\mathbf{n}_K \end{bmatrix} \xrightarrow{a.s.} \mathbf{0}_{K\times T}.$$

Note that *Assumption 1* is equivalent to condition (A.3) and can imply condition (A.1) because if $\forall i,\ \|\mathbf{\Phi}_i^{\frac{1}{2}}\|_2 < \infty$,

$$\|\mathbf{\Phi}_i^{\frac{1}{2}}\mathbf{\Phi}_j^{\frac{1}{2}}\|_2 \le \|\mathbf{\Phi}_i^{\frac{1}{2}}\|_2\|\mathbf{\Phi}_j^{\frac{1}{2}}\|_2 < \infty,\ \forall 1 \le i,j \le K.$$

Appendix B

Proof of Theorem 4.4.1

We first derive the variance term in MSE. Substitute (4.32) into (4.25) yields

$$\hat{\mathbf{h}} = \mathbb{E}\left\{\hat{\mathbf{h}}\right\} + \frac{1}{\gamma}\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{N}\mathbf{p}$$

thus,

$$\begin{aligned}
\mathrm{Var}\{\hat{\mathbf{h}}\} &= \mathbb{E}\left\{\|\hat{\mathbf{h}} - \mathbb{E}\left\{\hat{\mathbf{h}}\right\}\|^2\right\} \\
&= \mathbb{E}\left\{\|\frac{1}{\gamma}\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{N}\mathbf{p}\|^2\right\} \\
&= \frac{1}{\gamma^2}\mathbb{E}\left\{\mathbf{p}^H\mathbf{N}^H\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{N}\mathbf{p}\right\} \\
&= \frac{1}{\gamma^2}\mathbf{p}^H\mathbb{E}\left\{\mathbf{N}^H\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{N}\right\}\mathbf{p} \\
&\overset{(a)}{=} \frac{1}{\gamma^2}\mathbf{p}^H(\mathrm{tr}\left(\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\right)\mathbf{I}_K)\mathbf{p} \\
&\overset{(b)}{=} \frac{1}{\gamma^2}\mathbf{p}^H\mathrm{tr}\left(\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi)\mathbf{W}(\phi)\right)\mathbf{p} \\
&\overset{(c)}{=} \frac{1}{\gamma^2}\mathbf{p}^H\mathrm{tr}\left(\mathbf{Q}_m^H\mathbf{Q}_m\right)\mathbf{p} \\
&\overset{(d)}{=} \frac{1}{\gamma^2}\mathbf{p}^H(m\mathbf{I}_K)\mathbf{p} \\
&= \frac{m}{\gamma^2}\|\mathbf{p}\|^2 = \frac{m}{\beta\|\mathbf{p}\|_2^2}
\end{aligned} \tag{B.1}$$

where step (a) follows from the fact that

$$\begin{aligned}\mathbb{E}\left\{\mathbf{N}^H\mathbf{A}\mathbf{N}\right\} &= \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij}\mathbb{E}\left\{\mathbf{x}_i\mathbf{x}_j^H\right\} \\ &= \sum_{i=1}^{M} a_{ii}\mathbb{E}\left\{\mathbf{x}_i\mathbf{x}_i^H\right\} \\ &= \sum_{i=1}^{M} a_{ii}\mathbf{I}_K \\ &= \mathrm{tr}(\mathbf{A})\mathbf{I}_K \end{aligned} \tag{B.2}$$

with $\mathbf{N} = [\mathbf{x}_1, \cdots, \mathbf{x}_M]^H$ and any square matrix $\mathbf{A} = [a_{ij}]$; step (b) follows from the property that the matrices in a trace of a product can be switched, step (c) follows from the equality $\mathbf{W}(\phi)\mathbf{W}^H(\phi) = \mathbf{I}_M$, and step (d) utilizes the orthogonality of a basis matrix.

The bias term, (4.26), renders

$$\begin{aligned} b(\hat{\mathbf{h}}) &= \mathbb{E}\left\{\|\mathbb{E}\left\{\hat{\mathbf{h}}\right\} - \mathbf{h}\|^2\right\} \\ &\overset{(a)}{=} \mathbb{E}\left\{\|(\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)\mathbf{h}\|^2\right\} \\ &= \mathbb{E}\left\{\mathbf{h}^H(\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2\mathbf{h}\right\} \\ &\overset{(b)}{=} \mathbb{E}\left\{\mathrm{tr}(\mathbf{h}^H(\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2\mathbf{h})\right\} \\ &\overset{(c)}{=} \mathbb{E}\left\{\mathrm{tr}((\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2\mathbf{h}\mathbf{h}^H)\right\} \\ &= \mathrm{tr}((\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2\mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\}) \end{aligned} \tag{B.3}$$

In the above derivation, step (a) is obtained by substituting (4.32) into (4.26), step (b) follows from the fact that trace of a scalar is equal to the scalar itself and step (c) invokes the property that the vectors (matrices) in a trace of a product are commutative.

Since

$$\mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\} = \mathbf{\Phi}^{\frac{1}{2}}\mathbb{E}\left\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\right\}\mathbf{\Phi}^{\frac{1}{2}} = \mathbf{\Phi}, \tag{B.4}$$

(B.3) can be simplified as

$$\begin{aligned} b(\hat{\mathbf{h}}) &= \mathrm{tr}((\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2\mathbf{\Phi}) \\ &\overset{def}{=} \mathrm{tr}(\mathbf{A}\mathbf{\Phi}) \end{aligned} \tag{B.5}$$

Note that when full modeling order is used, $m = M$, and $\mathbf{A} = \mathbf{O}_M$ and the estimator becomes an unbiased estimator, which is consistent with *Lemma 4.4.3*. Nevertheless, for the general case that $m \leq M$, we use the decomposition

$$\begin{aligned} \mathbf{Q}_m\mathbf{Q}_m^H &= \mathbf{Q}\begin{bmatrix} \mathbf{I}_m \\ \mathbf{O}_{(M-m)\times m} \end{bmatrix}\begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{(M-m)\times m} \end{bmatrix}\mathbf{Q}^H \\ &= \mathbf{Q}\begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{O}_{(M-m)} \end{bmatrix}\mathbf{Q}^H \end{aligned} \tag{B.6}$$

where $\mathbf{Q} \in \mathbb{C}^{M\times M}$ the "complete" basis matrix (i.e. no rank reduction), to express $\mathbf{A}$ as

$$\begin{aligned} \mathbf{A} &= (\mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi) - \mathbf{I}_M)^2 \\ &= (\mathbf{I}_M - \mathbf{W}(\phi)\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{W}^H(\phi))^2 \\ &= (\mathbf{W}(\phi)\mathbf{Q}\underbrace{\left(\mathbf{I}_M - \begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{O}_{(M-m)} \end{bmatrix}\right)}_{\mathbf{D}_m} \\ &\qquad \mathbf{Q}^H\mathbf{W}^H(\phi))^2 \\ &= \mathbf{W}(\phi)\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{W}(\phi)\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\mathbf{W}^H(\phi) \\ &= \mathbf{W}(\phi)\mathbf{Q}\mathbf{D}_m^2\mathbf{Q}^H\mathbf{W}^H(\phi) \end{aligned} \tag{B.7}$$

The matrix $\mathbf{D}_m$ is an orthogonal projection and idempotent matrix since

$$\begin{aligned} \mathbf{D}_m &= \left(\mathbf{I}_M - \begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{O}_{(M-m)} \end{bmatrix}\right) \\ &= \begin{bmatrix} \mathbf{O}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{I}_{M-m} \end{bmatrix} \in \mathbb{R}^{M\times M} \end{aligned}$$

and

$$\begin{aligned} \mathbf{D}_m^2 &= \begin{bmatrix} \mathbf{O}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{I}_{M-m} \end{bmatrix}^2 \\ &= \begin{bmatrix} \mathbf{O}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{I}_{M-m} \end{bmatrix} \\ &= \mathbf{D}_m \end{aligned} \tag{B.8}$$

Suppose $\mathbf{e}_i$ is an all-zero vector except for the ith entry being 1, and $\mathcal{W}$ is a subspace of $\mathbb{R}^M$ that is spanned by the orthonormal set of vectors $\mathbf{e}_{m+1}, \cdots, \mathbf{e}_M$, that is, $\mathcal{W} = \text{span}\left\{ \begin{bmatrix} \mathbf{O}_{m\times(M-m)} \\ \mathbf{I}_{M-m} \end{bmatrix} \right\}$, then $\mathbf{D}_m$ is clearly the orthogonal projection matrix on $\mathcal{W}$. Hence, (B.7) and (B.5) can be further simplified as

$$\mathbf{A} = \mathbf{W}(\phi)\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\mathbf{W}^H(\phi)$$

and

$$\begin{aligned} b(\hat{\mathbf{h}}) &= \text{tr}(\mathbf{A}\mathbf{\Phi}) \\ &= \text{tr}(\mathbf{W}(\phi)\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}) \\ &= \text{tr}(\mathbf{D}_m\mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q}) \\ &\overset{def}{=} \text{tr}(\mathbf{D}_m\mathbf{B}) \end{aligned} \tag{B.9}$$

where the *bias matrix* $\mathbf{B}$ is a positive semi-definite matrix because of the facts that

$$\mathbf{B}^H = \mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q} = \mathbf{B}$$

and

$$\begin{aligned} \mathbf{x}^H\mathbf{B}\mathbf{x} &= \mathbf{x}^H\mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q}\mathbf{x} \\ &= \mathbf{x}^H\mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{W}(\phi)\mathbf{Q}\mathbf{x} \\ &= \|\mathbf{\Phi}^{\frac{1}{2}}\mathbf{W}(\phi)\mathbf{Q}\mathbf{x}\|_2^2 \geq 0, \quad \forall\ \mathbf{x} \end{aligned} \tag{B.10}$$

When substituting $\mathbf{e}_i,\ \forall\ 1 \leq i \leq M$ into the above $\mathbf{x}$, it is obvious that all diagonal entries in $\mathbf{B}$ are non-negative. Furthermore, the definition of Frobenius inner product [33, Ch. 10.4] implies that

$$b(\hat{\mathbf{h}}) = \text{tr}(\mathbf{D}_m\mathbf{B}) = \langle \mathbf{D}_m, \mathbf{B} \rangle_F \tag{B.11}$$

where $\langle \cdot \rangle_F$ denotes Frobenius inner product. The two equivalent expressions conclude that the bias term is the sum of the last $(M-m)$ diagonal terms (which are all non-negative) of $\mathbf{B}$.

Appendix C

On Remark 14

If the mean AoA is not 0°, the spatial correlation matrix is a complex matrix, that is, $\mathbf{\Phi} \in \mathbb{C}^{M\times M}$. However, [12] has shown that when the AoA spread is small (say, less than 15°),

$$\mathbf{\Phi} \approx \mathbf{W}(\phi)\bar{\mathbf{\Phi}}\mathbf{W}^H(\phi) \tag{C.1}$$

where $\bar{\mathbf{\Phi}} \in \mathbb{R}^{M\times M}$ is a real matrix with $\left[\bar{\mathbf{\Phi}}_{ij}\right] = J_0(|i-j|\frac{2\pi d}{\lambda}\Delta\cos\phi)$, Δ being the AS, λ being the wavelength and d being the antenna spacing. Hence, we can treat $\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$ as an operation that rotate the phase of $\mathbf{\Phi}$ to make it a real matrix.

From the "energy compaction property" [36] [41] of DCT-2, if we transform a length-M real sequence (or, real vector), $\mathbf{x}$, by DCT-2, then in frequency domain, the larger coefficients (or strictly speaking, coefficients larger than a benchmark, $\eta = 1$) must be more highly concentrated at low indices much smaller than M. More specifically,

$$1 - \frac{\sum_{\ell=1}^{\hat{m}^\star} x_l}{\sum_{\ell=1}^{M} x_l} \ll 1 \tag{C.2}$$

with $\hat{m}^\star \ll M$ being defined as the index of first entry smaller than η on $\mathbf{x}$, and all entries whose index larger than $\hat{m}^\star$ are also smaller than η. Note that DCT-2 is a separable transform [34, Ch. 4], or, to put it in another way, multidimensional DCT-2 can be decomposed into successive application of one-dimensional DCT-2 in the appropriate directions.

Since $\mathbf{B} = \mathbf{Q}^H\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi_k)\mathbf{Q}$, we can use a two-step interpretation to describe $\mathbf{B}$:

Step 1 Phase-rotating $\mathbf{\Phi}$ to be a real matrix, $\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$.

Step 2 Using two-dimensional DCT-2, applied subsequently to rows and columns of the real matrix $\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$, to transform it into frequency domain, $\mathbf{B}$, where

$$1 - \frac{\sum_{\ell=1}^{\hat{m}^\star}[\mathbf{B}]_{\ell\ell}}{\sum_{\ell=1}^{M}[\mathbf{B}]_{\ell\ell}} = 1 - \frac{\sum_{\ell=1}^{\hat{m}^\star}[\mathbf{B}]_{\ell\ell}}{M} \ll 1 \tag{C.3}$$

where $\hat{m}^\star << M$, owing to the energy compaction property [36, Ch. 8] [41] of DCT-2.

Besides, the concept of using polynomial basis is to approximate the smooth spatial waveform by a polynomial with lower order, m, than the full order, M. When the polynomial basis is used, we can replace Step 2 as

Step 2 Using two polynomial to describe respectively the rows and columns of the smooth real matrix $\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$. Because the spatial waveform (either along rows or columns of $\mathbf{W}^H(\phi)\mathbf{\Phi}\mathbf{W}(\phi)$) is smooth, the larger coefficients are also highly concentrated at low polynomial orders (i.e. low indices), while the coefficients corresponds to higher polynomial order are all small than η (since we do not need such high-order polynomial to describe the smooth spatial waveform). In other words,

$$1 - \frac{\sum_{\ell=1}^{\hat{m}^\star}[\mathbf{B}]_{\ell\ell}}{\sum_{\ell=1}^{M}[\mathbf{B}]_{\ell\ell}} = 1 - \frac{\sum_{\ell=1}^{\hat{m}^\star}[\mathbf{B}]_{\ell\ell}}{M} \ll 1 \tag{C.4}$$

where $\hat{m}^\star << M$.

Consequently, the diagonal entries in $\mathbf{B}$ must also be highly concentrated at low indices much smaller than M as in the DCT-2 basis case. We can say that, the polynomial basis we used, which is a discrete polynomial transform, also has the so-called "energy compaction property". And the compaction performance of it is only slightly worse than that of DCT-2 owing to Remark 11.

Appendix D

Proof of Theorem 4.5.1

Above all, we rewrite $\hat{\mathbf{h}}^{(\mathrm{I})}$ in (4.44) as

$$\begin{aligned}\hat{\mathbf{h}}^{(\mathrm{I})} &= \frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{Y}\mathbf{p} \\ &= \frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H(\gamma\mathbf{h}+\mathbf{N}\mathbf{p}) \\ &= \mathbf{Q}_m\mathbf{Q}_m^H\mathbf{h}+\frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{N}\mathbf{p} \end{aligned} \tag{D.1}$$

Take expectation to (4.44) yields

$$\begin{aligned}\mathbb{E}\left\{\hat{\mathbf{h}}^{(\mathrm{I})}\right\} &= \mathbf{Q}_m\mathbf{Q}_m^H\mathbf{h}+\mathbb{E}\left\{\frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{N}\mathbf{p}\right\} \\ &= \mathbf{Q}_m\mathbf{Q}_m^H\mathbf{h} \end{aligned} \tag{D.2}$$

Substituting (D.2) into (D.1), we rewrite $\hat{\mathbf{h}}^{(\mathrm{I})}$ as

$$\hat{\mathbf{h}}^{(\mathrm{I})} = \mathbb{E}\left\{\hat{\mathbf{h}}^{(\mathrm{I})}\right\}+\frac{1}{\gamma}\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{N}\mathbf{p}.$$

Thus,

$$\begin{aligned}
\mathrm{Var}\{\hat{\mathbf{h}}^{(\mathrm{I})}\} &= \mathbb{E}\left\{\|\hat{\mathbf{h}}^{(\mathrm{I})} - \mathbb{E}\left\{\hat{\mathbf{h}}^{(\mathrm{I})}\right\}\|^2\right\} \\
&= \frac{1}{\gamma^2}\mathbb{E}\left\{\mathbf{p}^H\mathbf{N}^H\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{N}\mathbf{p}\right\} \\
&= \frac{1}{\gamma^2}\mathbf{p}^H\mathbb{E}\left\{\mathbf{N}^H\mathbf{Q}_m\mathbf{Q}_m^H\mathbf{N}\right\}\mathbf{p} \\
&\overset{(a)}{=} \frac{1}{\gamma^2}\mathbf{p}^H(\mathrm{tr}\left(\mathbf{Q}_m\mathbf{Q}_m^H\right)\mathbf{I}_K)\mathbf{p} \\
&= \frac{1}{\gamma^2}\mathbf{p}^H\mathrm{tr}\left(\mathbf{Q}_m^H\mathbf{Q}_m\right)\mathbf{p} \\
&= \frac{1}{\gamma^2}\mathbf{p}^H(m\mathbf{I}_K)\mathbf{p} \\
&= \frac{m}{\gamma^2}\|\mathbf{p}\|^2 = \frac{m}{\beta\|\mathbf{p}\|_2^2}
\end{aligned} \tag{D.3}$$

where step (a) follows from the property (B.2).

After that, the bias term in the MSE can be derived:

$$\begin{aligned}
b(\hat{\mathbf{h}}^{(\mathrm{I})}) &= \mathbb{E}\left\{\|\mathbb{E}\left\{\hat{\mathbf{h}}\right\} - \mathbf{h}\|^2\right\} \\
&\overset{(a)}{=} \mathbb{E}\left\{\|(\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)\mathbf{h}\|^2\right\} \\
&= \mathbb{E}\left\{\mathbf{h}^H(\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2\mathbf{h}\right\} \\
&\overset{(b)}{=} \mathbb{E}\left\{\mathrm{tr}(\mathbf{h}^H(\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2\mathbf{h})\right\} \\
&\overset{(c)}{=} \mathbb{E}\left\{\mathrm{tr}((\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2\mathbf{h}\mathbf{h}^H)\right\} \\
&= \mathrm{tr}((\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2\mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\}) \\
&\overset{(d)}{=} \mathrm{tr}((\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2\mathbf{\Phi}) \\
&\overset{def}{=} \mathrm{tr}(\tilde{\mathbf{A}}\mathbf{\Phi})
\end{aligned} \tag{D.4}$$

in step (a), we substitute (D.2) into (4.26), step (b) follows from the fact that trace of a scalar is equal to the scalar itself, step (c) utilizes the property that the vectors (matrices) in a trace of a product can be switched, and step (d) is according to (B.4).

Similar to (B.7), $\tilde{\mathbf{A}}$ can be rewritten as

$$\begin{aligned}
\tilde{\mathbf{A}} &= (\mathbf{Q}_m\mathbf{Q}_m^H - \mathbf{I}_M)^2 \\
&= (\mathbf{I}_M - \mathbf{Q}_m\mathbf{Q}_m^H)^2 \\
&= (\mathbf{Q}\underbrace{\left(\mathbf{I}_M - \begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{m\times(M-m)} \\ \mathbf{O}_{(M-m)\times m} & \mathbf{O}_{(M-m)} \end{bmatrix}\right)}_{\mathbf{D}_m}\mathbf{Q}^H)^2 \\
&= \mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H \\
&= \mathbf{Q}\mathbf{D}_m^2\mathbf{Q}^H \\
&\overset{(a)}{=} \mathbf{Q}\mathbf{D}_m\mathbf{Q}^H
\end{aligned} \tag{D.5}$$

where step (a) uses the idempotent property of $\mathbf{D}_m$ in (B.8).

Accordingly, substitute (D.5) into (D.4) yield

$$\begin{aligned}
b(\hat{\mathbf{h}}) &= \mathrm{tr}(\mathbf{A}\boldsymbol{\Phi}) \\
&= \mathrm{tr}(\mathbf{Q}\mathbf{D}_m\mathbf{Q}^H\boldsymbol{\Phi}) \\
&= \mathrm{tr}(\mathbf{D}_m\mathbf{Q}^H\boldsymbol{\Phi}\mathbf{Q}) \\
&\overset{def}{=} \mathrm{tr}(\mathbf{D}_m\tilde{\mathbf{B}}),
\end{aligned} \tag{D.6}$$

thus proves *Theorem 4.5.1*. It is worth mentioning that $\tilde{\mathbf{B}}$ is the positive semi-definite bias matrix of the SSFC-2 estimator.

Appendix E

On Remark 17

If the mean AoA is not 0°, $\mathbf{\Phi} \in \mathbb{C}^{M\times M}$ and, according to (C.1), can be regarded as a phase-rotated version of a real matrix, $\bar{\mathbf{\Phi}} \in \mathbb{R}^{M\times M}$.

As DFT$\{e^{j\omega_0 n}x[n]\} = X(e^{j(\omega-\omega_0)})$, where $X(e^{j\omega})$ =DFT$\{x[n]\}$, performing two-dimensional (2D) DCT-2 on the phase-rotated matrix $\mathbf{\Phi}$, the $\eta-$support also shifted to some medium frequency indices. In other words, suppose that $\mathbf{x}$ is a length-M complex vector, and after transformed by the DCT-2,

$$1 - \frac{\sum_{\ell=m_r-1}^{\hat{m}^\star} x_l}{\sum_{\ell=1}^{M} x_l} \ll 1 \tag{E.1}$$

with $\hat{m}^\star$ being defined as the index of first entry smaller than η on $\mathbf{x}$, and all entries whose index larger than $\hat{m}^\star$ are also smaller than η, m_r being defined as the index of first entry larger than η on $\mathbf{x}$, and $\hat{m}^\star - m_r \ll M$.

Consequently, $\tilde{\mathbf{B}}$ can be decomposed into the following two steps:

Step 1 Rotating the phase of a real matrix $\bar{\mathbf{\Phi}}$ to obtain the phase-rotated complex matrix, $\mathbf{\Phi}$.

Step 2 Using the two-dimensional DCT-2 (or discrete polynomial transform, applied subsequently to rows and columns of the phase-rotated matrix $\mathbf{W}(\phi)\bar{\mathbf{\Phi}}\mathbf{W}^H(\phi)$. Owning to the energy compaction property of DCT-2 (or discrete polynomial transform) [34, Ch. 4] [35, Ch. 2], the resulting frequency domain coefficients larger than a fixed value, $\eta = 1$, are highly concentrated; however, they are mostly

concentrated within some medium frequency indices due to the frequency translation property. More precisely,

$$1 - \frac{\sum_{\ell=m_r-1}^{\hat{m}^\star}[\tilde{\mathbf{B}}]_{ll}}{\sum_{\ell=1}^{M}[\tilde{\mathbf{B}}]_{ll}} = 1 - \frac{\sum_{\ell=m_r-1}^{\hat{m}^\star}[\tilde{\mathbf{B}}]_{ll}}{M} \ll 1 \tag{E.2}$$

with $\hat{m}^\star - m_r \ll M$.

Note that when the polynomial basis is used, we can treat it as a kind of discrete polynomial transform which also has energy compaction property as mentioned in Appendix C, and the above interpretation is still valid. Therefore, it would lead to the same result as the DCT-2 basis case.

Appendix F

Proof of Lemma 4.5.3

Since

$$\begin{aligned} & \operatorname{tr}(\mathbf{D}_m\tilde{\mathbf{B}}) - \operatorname{tr}(\mathbf{D}_m\mathbf{B}) \\ = \; & \operatorname{tr}(\mathbf{D}_m\mathbf{Q}^H\mathbf{\Phi}\mathbf{Q}) - \operatorname{tr}(\mathbf{D}_m\mathbf{Q}^H\mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi)\mathbf{Q}) \\ = \; & \operatorname{tr}(\mathbf{D}_m\mathbf{Q}^H(\mathbf{\Phi} - \mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi))\mathbf{Q}) \\ \overset{(a)}{=} \; & \operatorname{tr}(\mathbf{D}_m^2\mathbf{Q}^H(\mathbf{\Phi} - \mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi))\mathbf{Q}) \\ \overset{(b)}{=} \; & \operatorname{tr}(\mathbf{D}_m\mathbf{Q}^H(\mathbf{\Phi} - \mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi))\mathbf{Q}\mathbf{D}_m) \end{aligned} \tag{F.1}$$

where step (a) utilizes the idempotent property of $\mathbf{D}_m$ in (B.8), and step (b) follows from the fact that trace of product of matrices is commutative.

We need the following two lemmas from [32, Appendix 1].

Lemma F.0.1. *A square matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$ is positive semi-definite if and only if the principal minors are all nonnegative.*

Lemma F.0.2. *A square matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$ is positive semi-definite if and only if it can be written as*

$$\mathbf{A} = \mathbf{Z}\mathbf{Z}^H \tag{F.2}$$

where $\mathbf{Z} \in \mathbb{C}^{M\times M}$ may not be full rank.

It is clear that the principal minors are all zero, hence nonnegative, due to the all-zero diagonal terms of $(\mathbf{\Phi} - \mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi))$. $(\mathbf{\Phi} - \mathbf{W}(\phi)^H\mathbf{\Phi}\mathbf{W}(\phi))$ is thus positive

semi-definite by *Lemma F.0.1*. Applying *Lemma F.0.2* to $(\mathbf{\Phi} - \mathbf{W}(\phi)^H \mathbf{\Phi} \mathbf{W}(\phi))$, we obtain

$$\mathbf{\Phi} - \mathbf{W}(\phi)^H \mathbf{\Phi} \mathbf{W}(\phi) = \mathbf{Z}\mathbf{Z}^H \tag{F.3}$$

which, when substituting into (F.1), gives

$$\mathrm{tr}(\mathbf{D}_m \tilde{\mathbf{B}}) - \mathrm{tr}(\mathbf{D}_m \mathbf{B}) = \mathrm{tr}(\mathbf{D}_m \mathbf{Q}^H \mathbf{Z}\mathbf{Z}^H \mathbf{Q}\mathbf{D}_m) = \|\mathbf{Z}^H \mathbf{Q}\mathbf{D}_m\|_F^2 \geq 0 \tag{F.4}$$

and completes the proof.

Appendix G

Proof of Lemma 5.1.2

We focus on RRH-ℓ and omit the superscript ℓ of M_ℓ and K_ℓ henceforth. Let $\mathcal{S}$ be the current antenna set, and $M_F = |\mathcal{S}|$.

(Capacity of MF precoder)

Without loss of generality, we assume equal power allocation and denote by P, $\mathbf{h}_\ell^H$, $\mathbf{W}_\mathcal{S} = \mathbf{G}_\mathcal{S}^H$, and $\mathbf{x}$, the total power, ℓth row of $\mathbf{H}_\mathcal{S}$, precoding matrix and transmitted signal, respectively. We then have

$$\sqrt{p_k} = \sqrt{\frac{P}{\mathrm{tr}(\mathbf{W}_\mathcal{S}^H \mathbf{W}_\mathcal{S})}} \xrightarrow{a.s.} \sqrt{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}}$$

and the (composite) received signal from all K serving users is given by

$$\mathbf{y} = \mathbf{G}_\mathcal{S}\mathbf{W}_\mathcal{S}\mathbf{P}_\mathcal{S}\mathbf{x} + \mathbf{n} \xrightarrow{a.s.} \sqrt{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}} \begin{bmatrix} \sqrt{\beta_1}\mathbf{h}_1^H \\ \vdots \\ \sqrt{\beta_K}\mathbf{h}_K^H \end{bmatrix} \times (x_1\sqrt{\beta_1}\mathbf{h}_1 + \cdots + x_K\sqrt{\beta_K}\mathbf{h}_K) + \mathbf{n}$$

The signal received by the ith user is

$$\begin{aligned} y_i \xrightarrow{a.s.} \quad & \sqrt{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}}\beta_i \mathbf{h}_i^H (x_i \mathbf{h}_i) + \sqrt{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}}\sqrt{\beta_i}\mathbf{h}_i^H \\ & \times \left(x_1\sqrt{\beta_1}\mathbf{h}_1 + \cdots + x_{i-1}\sqrt{\beta_{i-1}}\mathbf{h}_{i-1} + x_{i+1}\sqrt{\beta_{i+1}}\mathbf{h}_{i+1} + \cdots + x_K\sqrt{\beta_K}\mathbf{h}_K \right) + n_i \end{aligned}$$

with the corresponding SINR given on the top of next page.

$$\begin{aligned} \mathrm{SINR}_{i,MF} \quad & \xrightarrow{a.s.} \quad \frac{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}\beta_i^2 \|\mathbf{h}_i\|^4 |x_i|^2}{\left|\sqrt{\frac{P}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}}\sqrt{\beta_i}\mathbf{h}_i^H \left(\sum_{j=1,j\neq i}^K x_j\sqrt{\beta_j}\mathbf{h}_j\right)\right|^2 + 1} \\ & \xrightarrow{a.s.} \quad \frac{\frac{PM\beta_i^2}{\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}}{\frac{P\beta_i}{M\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})}\left|\mathbf{h}_i^H \left(\sum_{j=1,j\neq i}^K x_j\sqrt{\beta_j}\mathbf{h}_j\right)\right|^2 + 1} \end{aligned} \tag{G.1}$$

Thus the capacity is given by

$$C = \sum_i \log(1 + \text{SINR}_{i,MF}) \tag{G.2}$$

Theoretically, the $\mathbf{h}_i^H \mathbf{h}_j, \forall j \neq i$ terms in (G.1), which are the non-diagonal terms of $\mathbf{H}\mathbf{H}^H$, approach to zero as $\frac{M}{K} \to \infty$ in according with Lemma 3.3.1. However, our numerical experiment indicates that even if $\frac{M}{K} = 1000000$, and the there is no spatial correlation, the non-diagonal terms of $\mathbf{H}\mathbf{H}^H$ are still large (about several hundreds to several thousand). These large non-diagonal terms result in strong interference in (G.1), thus, $\text{SINR}_{i,MF}$ becomes much smaller than expected.

$$\text{SINR}_{i,MF} \ll \frac{PM\beta_i^2}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}})} = \frac{PM}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}})/\beta_i^2} \tag{G.3}$$

From the above derivation and simulation results, we conclude that in practice, MF precoder is not a good precoder in massive MIMO system.

(Capacity of ZF precoder)

Similar to the MF case, we assume the system employs equal power allocation with total power P. With the precoding matrix $\tilde{\mathbf{W}}_{\mathcal{S}} = \mathbf{G}_{\mathcal{S}}^{\dagger} = \mathbf{G}_{\mathcal{S}}^H(\mathbf{G}_{\mathcal{S}}\mathbf{G}_{\mathcal{S}}^H)^{-1}$, we have

$$\sqrt{p_k} = \sqrt{\frac{P}{\text{tr}(\tilde{\mathbf{W}}_{\mathcal{S}}^H \tilde{\mathbf{W}}_{\mathcal{S}})}} \xrightarrow{a.s.} \sqrt{\frac{PM}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}}^{-1})}}$$

so that the (composite) received vector become

$$\begin{aligned} \mathbf{y} &= \mathbf{G}_S \tilde{\mathbf{W}}_S \mathbf{P}_S \mathbf{x} + \mathbf{n} \\ &\xrightarrow{a.s.} \sqrt{\frac{PM}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}}^{-1})}} \mathbf{x} + \mathbf{n} \end{aligned} \tag{G.4}$$

whose ith component represents the received signal of the ith user

$$y_i \xrightarrow{a.s.} \sqrt{\frac{PM}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}}^{-1})}} x_i + n_i \tag{G.5}$$

The corresponding SINR renders the asymptote

$$\text{SINR}_{i,ZF} \xrightarrow{a.s.} \frac{PM}{\text{tr}(\mathbf{D}_{\boldsymbol{\beta}}^{-1})} \tag{G.6}$$

The resulting capacity has a similar expression as (G.2).

As the LSFC for all K serving users are approximately equal due to the fact that an RRH only serves the nearby MSs, we obtain from (G.3) and (G.6)

$$\mathrm{SINR}_{i,MF} \overset{\text{a.s.}}{\ll} \frac{PM}{\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}})/\beta_i^2} \approx \frac{PM}{K/\beta_i} \tag{G.7}$$

$$\mathrm{SINR}_{i,ZF} \xrightarrow{a.s.} \frac{PM}{\mathrm{tr}(\mathbf{D}_{\boldsymbol{\beta}}^{-1})} \approx \frac{PM}{K/\beta_i}, \tag{G.8}$$

which causes

$$\mathrm{SINR}_{i,MF} \overset{\text{a.s.}}{\ll} \mathrm{SINR}_{i,ZF} \tag{G.9}$$

This means ZF precoding can remove all interference and offer an SINR much higher than that achievable by using an MF precoder even in a practical massive MIMO system where $\frac{M}{K} < \infty$.

Bibliography

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Proces. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] S. Payami and F. Tufvesson, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. ACSSC*, Nov. 2012.

[3] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] A. Liu and V. Lau, "Joint power and antenna selection optimization in large distributed MIMO networks," Tech. Rep., 2012.

[5] L. Rong, X. Su, J. Zeng, Y. Kuang, and J. Li, "Large scale MIMO transmission technology in the architecture of cloud base-station," in *Proc. IEEE GLOBECOM Workshops*, pp. 255–260, Dec. 2012.

[6] H. Q. Ngo, M. Matthaiou and E. G. Larsson, "Performance analysis of large scale MU-MIMO with optimal linear receivers," *2012 Swedish Communication Technologies Workshop (Swe-CTW)*, pp.59-64, 24-26 Oct. 2012.

[7] "Spatial channel model for multiple input multiple output (MIMO) simulations," 3GPP TR 25.996 V11.0.0, Sep. 2012. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25996.htm

[8] C.-X. Wang, X. Hong, H. Wu and W. Xu, "Spatial-temporal correlation properties of the 3GPP spatial channel model and the Kronecker MIMO channel model," *EURASIP J. Wireless Commun. and Netw.*, 2007.

[9] 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.

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

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

[12] Y.-C. Chen and Y. T. Su, "MIMO channel estimation in correlated Fading Environments," *IEEE Trans. Commun.*, vol. 9, no. 3, pp. 1108–1119, Mar. 2010.

[13] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE ISIT*, pp. 2184–2188, Jul. 2009.

[14] F. Fernandes, A. Ashikhmin and T. L. Marzetta, "Inter-Cell Interference in Noncooperative TDD Large Scale Antenna Systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp.192–201, Feb. 2013.

[15] "Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects," 3GPP TR 36.814 V9.0.0, Mar. 2010. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm

[16] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[17] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 3226–3238, Sep. 2012.

[18] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[19] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," *IEEE VTC 2004*, vol. 6, pp. 4265–4269, Sep. 2004.

[20] J. Schäfer and K. Strimmer, "A Shrinkage Approach to Large Scale Covariance Matrix Estimation and Implications for Functional Genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, Article 32, 2005.

[21] M. Gharavi-Alkhansari and A. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

[22] C. Artigue and P. Loubaton, "On the precoder design of flat fading MIMO systems equipped with MMSE receivers: a large-system approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4138–4155, Jul. 2011.

[23] R. Vaze, "Sub-modularity and antenna selection in MIMO systems", *IEEE Commun. Lett.*, vol. 16, no. 9, pp. 1446–1449, Sep. 2012.

[24] S. Huang, H. Yin, H. Li, and V. C. M. Leung, "Decremental user selection for large-scale multi-user MIMO downlink with zero-forcing beamforming," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 480–483, Oct. 2012.

[25] "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Confor-

mance testing," 3GPP TS 36.521-1 V11.0.1, Mar. 2013. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36521-1.htm

[26] C. J. Zarowski, "The MDL criterion for rank determination via effective singular values," *IEEE Trans. Signal Process.*, vol.46, no.6, pp.1741–1744, Jun 1998.

[27] J. J. Blanz, "Method and apparatus for reduced rank channel estimation in a communications system," U.S. Patent 6,907,270, issued June 14, 2005.

[28] C.-C. Cheng, Y.-C. Chen, Y. T. Su, and H. Sari, "Model-based channel estimation and codeword selection for correlated MIMO channels," in *Proc. IEEE SPAWC*, pp. 540–544, Jun. 2012.

[29] 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.

[30] K. J. Horadam, *Hadamard Matrices and Their Applications*, Princeton, NJ: Princeton University Press, 2007.

[31] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, New York, NY, USA: Cambridge University Press, 2011.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[33] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Nov. 2008. [Online]. Available: `http://orion.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

[34] K. R. Rao and P. C. Yip, *The Transform and Data Compression Handbook*, CRC Press, Inc. Boca Raton, FL, USA, 2000.

[35] P. R. Haddad, A. N. Akansu, *Multiresolution Signal Decomposition, Second Edition: Transforms, Subbands, and Wavelets*, Academic Press, Oct. 2000.

[36] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing: Third Edition*, Pearson, 2010.

[37] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, J. Wiley, New York, 2002.

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

[39] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Series in Statistics, New York, NY, USA, 2009.

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

[41] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, 23, pp. 90–93, 1974.

[42] P.-H. Lin and S.-H. Tsai, "Performance analysis and algorithm designs for transmit antenna selection in linearly precoded multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1698–1708, May 2012.

[43] D. I. Klaus, "Spatial Channel Characteristics for Adaptive Antenna Downlink Transmission," Ph.D. dissertation, Vienna University of Technology, Vienna, Austria, 2002.

[44] Y. Oda, R. Tsuchihashi, K. Tsunekawa and M. Hata, "Measured Path Loss and Multipath Propagation Characteristics in UHF and Microwave Frequency Bands for Urban Mobile Communications," *Proc. IEEE 53rd Vehicular Technology Conference* (VTC'2001 Spring), pp. 337-341, Rhodes, Greece, 2001.

[45] Y. Yu and D. Gu, "Enhanced MU-MIMO Downlink Transmission in the FDD-Based Distributed Antennas System," *IEEE Commun. Lett.*, vol.16, no.1, pp.37–39, January 2012.

[46] J. Zhang, P. Soldati, Y. Liang, L. Zhang and K. Chen, "Pathloss determination of uplink power control for UL CoMP in heterogeneous network," 2012 IEEE *Globecom Workshops (GC Wkshps)*, pp. 250–254, 3–7 Dec. 2012.

作者簡歷

一、 關於作者

1988 年，九月，秋分的日子，誕生於台北萬華。
從小在熱鬧的都市裡長大，故對純樸的鄉下一直頗富好奇心。
1995 年，開始了長達十八年的學生生涯，
喜歡做學問、追根究柢，對於別人的疑惑，也竭盡所能、盡心回答，時常對所學的內容提出疑惑。也由於樂於主動吸收新知，培養了不錯的邏輯思辯能力。
2007 年，清大，首次離開家鄉台北，
期間體會到大自然與樹林之美，四年的美好的回憶至今仍點滴在心頭。
2011 年盛夏，進入交大電信所，拜蘇育德教授為師，
在老師的諄諄教誨下，除修習通信相關課程，
也常參加大大小小的研討會及演講，皆使得我對最新的通訊技術有更深的認識。與聯發科技產學合作不啻增進我對學界理論和業界實際需求的看法，也讓自己更加能夠努力克服各種挑戰。
2013 年初秋，在幾經波折後，終於完成了研究所兩年來的心血結晶，
儘管過程或許不大平順，常遇到困難，但隨著經驗的累積，
與同儕共同向上提昇，各方面能力都有顯著的進步，愈挫愈勇。
即將邁入 2014 年了，期許自己在未來也能秉持著做研究的態度，堅持到底。凡事沒有最好，永遠都要不斷地求新、求變、終身學習，才能持續保有熱情，將會是此生最受用的座右銘。

二、 學歷

高中	國立台北市立建國高級中學	2004/09-2007/06
大學	國立清華大學 電機工程學系	2007/09-2011/06
研究所	國立交通大學 電信工程研究所	2011/09-2013/08

三、專長技能

程式語言	Matlab 、 C/C++、Verilog
相關證照	Linux NCLA、Linux NCLP、Novell Data Center Technical Specialist、Novell Data Center Advanced Technical Specialist、TQC 中英打、TQC WinXP
得獎記錄	大學五學期書卷獎、碩士兩學期書卷獎、林公熊徵學田獎學金、電資院優秀學生獎學金、交大研究生獎學金、東光教育基金會獎學金、新北市政府優秀學生獎學金、宋羅姍英才獎學金、潘文淵文教基金會獎學金、新北市新莊區聯合獎助學金

四、已修畢之相關課程

專業課程類別	課程列表		
數理基礎課程	微積分	微分方程	線性代數
	機率	普通物理	普通化學
	材料科學導論	工程力學	偏微分方程與 複變函數
基礎專業課程(I)	訊號與系統	電磁學	電路學
	電子學	數位電路分析與 設計	邏輯設計
基礎專業課程 (II)	通訊概論	數位訊號處理概論	經濟學原理
	控制系統	類比電路設計入門	通訊系統
電腦&實驗課程	程式語言	計算機程式設計	計算機結構
	計算機網路概論	DSP 實驗	通訊系統實驗
	電子電路實驗	邏輯設計實驗	普通物理實驗
研究所專業課程	通訊之最佳化方法	數位通訊	數位訊號處理
	通訊之隨機程序	適應性訊號處理	計算機網路
	無線通訊之矩陣 理論	編碼理論	檢測與估計理論
旁聽過課程	消息理論	高等編碼理論	