# Pilot Reuse for Massive MIMO Transmission over Spatially Correlated Rayleigh Fading Channels

Li You, *Student Member, IEEE,* Xiqi Gao, *Fellow, IEEE,* Xiang-Gen Xia, *Fellow, IEEE,*
Ni Ma, *Senior Member, IEEE,* and Yan Peng

*Abstract*—We propose pilot reuse (PR) in single cell for massive multiuser multiple-input multiple-output (MIMO) transmission to reduce the pilot overhead. For spatially correlated Rayleigh fading channels, we establish a relationship between channel spatial correlations and channel power angle spectrum when the base station antenna number tends to infinity. With this channel model, we show that sum mean square error (MSE) of channel estimation can be minimized provided that channel angle of arrival intervals of the user terminals reusing the pilots are non-overlapping, which shows feasibility of PR over spatially correlated massive MIMO channels with constrained channel angular spreads. Regarding that channel estimation performance might degrade due to PR, we also develop the closed-form robust multiuser uplink receiver and downlink precoder that minimize sum MSE of signal detection, and reveal a duality between them. Subsequently, we investigate pilot scheduling, which determines the PR pattern, under two minimum MSE related criteria, and propose a low complexity pilot scheduling algorithm which relies on the channel statistics only. Simulation results show that the proposed PR scheme provides significant performance gains over the conventional orthogonal training scheme in terms of net spectral efficiency.

*Index Terms*—Pilot reuse, massive MIMO, multiuser MIMO, pilot scheduling, robust transmission.

## I. INTRODUCTION

**M**ASSIVE multiple-input multiple-output (MIMO) transmission employs a large number of antennas at the base station (BS) to serve a relatively smaller number of user terminals (UTs) simultaneously [2]. With the potential large gains in spectral efficiency and energy efficiency, massive MIMO is a promising technology that the next generation of wireless systems may incorporate, and has received tremendous research interest recently [3], [4].

Channel state information (CSI) at the BS plays an important role in massive MIMO transmission, and in realistic

L. You and X. Q. Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, P. R. China (e-mail: liyou@seu.edu.cn; xqgao@seu.edu.cn).

X.-G. Xia is with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA (e-mail: xxia@ee.udel.edu).

N. Ma and Y. Peng are with the Huawei Technologies Co., Ltd., Shenzhen 518129, P. R. China (e-mail: ni.ma@huawei.com; yan.peng@huawei.com).

systems it is typically obtained with assistance of the periodically inserted pilot signals [5]. In time-division duplex (TDD) massive MIMO transmission, CSI at the BS can be obtained from uplink (UL) training via leveraging the channel reciprocity [2], [6]. For the conventional orthogonal training (OT) scheme [6], the pilot overhead is proportional to the number of the UT antennas. As the UT antenna number grows, the heavy pilot overhead decreases the system efficiency greatly and can become the system bottleneck.

In order to reduce the pilot overhead, we propose pilot reuse (PR) in single cell for massive MIMO transmission in this paper. The motivation stems from that, in realistic outdoor wireless propagation environments where BS is located at an elevated position, the scattering around the BS is usually limited, and the MIMO channels are not spatially isotropic [7], [8], i.e., most of the channel power lies in a finite number of spatial directions compared with the whole massive MIMO channel dimension. For UTs with channels lying in almost orthogonal spatial directions, PR is feasible and beneficial.

In the proposed PR scheme, massive MIMO transmission consists of the following phases: statistical CSI acquisition for pilot scheduling, UL training for channel estimation, UL data transmission, and downlink (DL) data transmission. The pilot scheduler at the BS determines the PR pattern, and allocates the available pilot signals to the UTs. Due to the slow-varying nature of the long term channel statistics, it is reasonable to exploit the statistical CSI at the BS to perform pilot scheduling. With the resulting PR pattern, the UTs transmit the respective assigned pilot signals periodically to enable the BS to obtain the channel estimates. The channel estimation performance might degrade due to PR, thus it is natural to design the UL and DL data transmissions robust to the channel estimation error.

In this work, we consider the spatially correlated Rayleigh fading channels, and show that when the BS antenna number tends to infinity, eigenvectors of the channel covariance matrix are determined by the BS array response vectors, while eigenvalues depend on the channel power angle spectrum (PAS), which reveals a relationship between channel spatial correlations and channel power distribution in the angular domain. For this channel model, we show that sum mean square error of channel estimation (MSE-CE) can be minimized, provided that channel angle of arrival (AoA) intervals of the UTs reusing the pilots are non-overlapping, which shows feasibility of PR over spatially correlated massive MIMO channels with constrained channel angular spreads (ASs). Regarding that channel estimation performance might degrade due to PR, we investigate

robust data transmissions for both UL and DL with channel estimation error due to PR taken into account. The closed-form robust multiuser UL receiver and DL precoder which are applicable to arbitrary PR pattern based on the minimum MSE of signal detection (MMSE-SD) criterion are developed, and an interesting MMSE duality between them is revealed. Subsequently, we study pilot scheduling under two MMSE related criteria, and in both cases the designs are formulated as combinatorial optimization problems. We show that both criteria can be optimized provided that channel AoA intervals of the UTs reusing the pilots are non-overlapping, and propose a low complexity pilot scheduling algorithm (called the statistical greedy pilot scheduling [SGPS] algorithm) motivated by the channel AoA non-overlapping condition. Simulation results show that the proposed PR scheme provides significant performance gains over the conventional OT scheme in terms of net spectral efficiency.

*Related Works and Our Contributions:* Most of the previous works assumed pilot reuse among cells for massive MIMO transmission, where UTs in the same cell use orthogonal pilots, while the same set of orthogonal pilots is reused among cells [2], [9], [10]. It has been shown that pilot contamination [10] caused by inter-cell pilot reuse can degrade the performance of massive MIMO transmission. In order to mitigate pilot contamination, several approaches including, e.g., coordinated channel estimation [11], time-shifted pilot allocation [12], eigenvalue decomposition based blind channel estimation [13], cooperative pilot contamination precoding [14], and distributed MMSE precoding [10] were proposed, respectively. In contrast to these existing works where pilot overhead was simply set to be fixed, our work focuses on reducing the pilot overhead, and in the meanwhile, balancing the tradeoff between the pilot overhead and the pilot interference.[1] Specifically, we propose PR among UTs, where the required number of orthogonal pilots can be much smaller than the number of UTs in a cell, by allowing that different UTs in a cell share the same pilot. We investigate pilot scheduling, channel estimation, and robust UL and DL transmissions under the systematic PR framework, which, has not been thoroughly addressed in the literature yet. Compared with existing works which dealt with inter-cell pilot contamination, our work is more general in the sense that we deal with both the pilot interference and pilot overhead under the PR framework. In this paper, our analysis focuses on the single-cell scenario for the sake of clarity. Regarding the multi-cell scenario where pilot contamination persists [15], performances of the proposed PR scheme will depend on the system configurations. For example, if statistical CSI coordination among cells is possible, i.e., channel covariance matrices can be exchanged among cells, pilot scheduling and pilot interference (as well as pilot contamination) mitigation can be performed jointly via exploiting statistical CSI of the intra-cell as well as inter-cell UTs.

*Notations:* We use $\bar{\jmath} = \sqrt{-1}$ to denote the imaginary unit. Upper (lower) case boldface letters are used to denote matrices (column vectors). $\mathbf{I}_N$ denotes the $N \times N$ dimensional identity

matrix, and the subscript is omitted for brevity in some cases where it is clear. $\mathbf{0}$ denotes the all-zero vector (matrix). The superscripts $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ denote the conjugated-transpose, transpose, and conjugate operations, respectively. The operator $\mathrm{diag}\{\mathbf{x}\}$ denotes the diagonal matrix with $\mathbf{x}$ along its main diagonal, and $\mathrm{tr}\{\cdot\}$ denotes the matrix trace operation. We employ $[\mathbf{a}]_i$ and $[\mathbf{A}]_{i,j}$ to denote the $i^{\text{th}}$ element of the vector $\mathbf{a}$, and the $(i,j)^{\text{th}}$ element of the matrix $\mathbf{A}$, respectively, where the element indices start from 1. $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^H \mathbf{a}}$ denotes the $\ell_2$-norm of $\mathbf{a}$, and $\|\mathbf{X}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}\{\mathbf{X}^H \mathbf{X}\}}$ denotes the Frobenius norm of $\mathbf{X}$. $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^H \mathbf{b}$ denotes the inner product between $\mathbf{a}$ and $\mathbf{b}$. $\mathbf{A} \succ \mathbf{0}$ ($\mathbf{A} \succeq \mathbf{0}$) denotes that $\mathbf{A}$ is Hermitian positive definite (semi-definite), and $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$. $\mathbb{C}^{M \times N}$ denotes the $M \times N$ dimensional complex vector space. $\mathbb{E}\{\cdot\}$ denotes the expectation operation. $\mathcal{CN}(\mathbf{a}, \mathbf{B})$ denotes the circular symmetric complex Gaussian distribution with mean $\mathbf{a}$ and covariance $\mathbf{B}$. $\delta(\cdot)$ denotes the Dirac delta function. $\lfloor x \rfloor$ denotes the largest integer that is not greater than $x$. The notation $\triangleq$ is used for definitions, and $\sim$ means "be distributed as". The superscripts "p", "u", and "d" stand for the expressions related to pilot, UL data, and DL data, respectively.

*Outline:* The rest of the paper is organized as follows. In Section II, we investigate the massive MIMO channel model, and establish the relationship between channel spatial correlations and channel PAS. In Section III, we present PR for UL channel estimation and show how PR affects the channel estimation performance. We also provide a condition under which the MSE-CE can be minimized. In Section IV, we develop the robust multiuser UL receiver and DL precoder under the MMSE-SD criterion, and reveal a MMSE duality between them. In Section V, we study pilot scheduling and propose a low complexity pilot scheduling algorithm that relies on the channel statistics only. Simulation results are provided in Section VI and the paper is concluded in Section VII.

## II. MASSIVE MIMO CHANNEL MODEL

We consider massive MIMO transmission in the TDD mode in single-cell scenario, where the BS with $M$ antennas serves $K(\ll M)$ single-antenna UTs over frequency-flat fading channels on a narrow-band sub-carrier. We assume that channels vary in time according to the block fading model, where channel states stay constant over the coherence block with a length of $T$ symbols, and evolve from block to block in an independent and identically distributed manner according to some ergodic process.

With the ray-tracing based approach [7], [16], [17], the UL channel between the $M$ antennas at the BS and the antenna of the $k^{\text{th}}$ UT can be modeled as

$$\mathbf{g}_k = \int_{\mathcal{A}} \mathbf{v}(\theta) g_k(\theta) \, \mathrm{d}\theta = \int_{\theta^{\min}}^{\theta^{\max}} \mathbf{v}(\theta) g_k(\theta) \, \mathrm{d}\theta \qquad (1)$$

where $g_k(\theta)$ and $\mathbf{v}(\theta) \in \mathbb{C}^{M \times 1}$ are the complex channel gain function and the BS array response vector corresponding to the incidence angle $\theta$, respectively. We assume that $\|\mathbf{v}(\theta)\|_2 = \sqrt{M}$ for power normalization. We assume that the channel

---

[1] Pilot interference caused by intra-cell PR is defined in (13), and can be seen as intra-cell pilot contamination.

power seen at the BS is constrained to lie in the angle interval $\mathcal{A} = [\theta^{\min}, \theta^{\max}]$, which can be achieved via placing directional antennas at the BS, and thus no power is received at the BS for incidence angle $\theta \notin \mathcal{A}$.

We assume that the channel phases are uniformly distributed, thus $\mathbb{E}\{\mathbf{g}_k\} = \mathbf{0}$. We assume that channels with different incidence angles are uncorrelated, i.e., $\mathbb{E}\{g_k(\theta) g_k^*(\theta')\} = \beta_k \mathsf{S}_k(\theta) \delta(\theta - \theta')$ where $\beta_k$ represents the large scale fading, and $\mathsf{S}_k(\theta)$ represents the channel PAS which models the channel power distribution in the angular domain [18]. Then from (1), the channel covariance matrix (BS spatial correlation matrix) is given by

$$\mathbf{R}_k = \mathbb{E}\{\mathbf{g}_k \mathbf{g}_k^H\} = \beta_k \int_{\theta^{\min}}^{\theta^{\max}} \mathbf{v}(\theta) \mathbf{v}^H(\theta) \mathsf{S}_k(\theta) \, \mathrm{d}\theta. \quad (2)$$

We assume that $\int_{-\infty}^{\infty} \mathsf{S}_k(\theta) \, \mathrm{d}\theta = 1$, and channel power normalization should be satisfied as

$$\mathrm{tr}\{\mathbf{R}_k\} = \beta_k M \int_{\theta^{\min}}^{\theta^{\max}} \mathsf{S}_k(\theta) \, \mathrm{d}\theta. \quad (3)$$

A specific property of the massive antenna array is its high resolution to the channels in the angular domain [19], and we introduce an assumption about it in the following.

*Assumption 1:* [20] Array response vectors corresponding to distinct angles are asymptotically orthogonal when the BS antenna number tends to infinity, i.e., for $\forall \zeta, \vartheta \in \mathcal{A}$,

$$\lim_{M \to \infty} \frac{1}{M} \langle \mathbf{v}(\zeta), \mathbf{v}(\vartheta) \rangle = \delta(\zeta - \vartheta). \quad (4)$$

Note that Assumption 1 is valid for uniform linear array (ULA) as one shall see in Remark 1. Based on this assumption, we can obtain the following result on massive MIMO channel covariance matrix.

*Lemma 1:* Let

$$\mathbf{V} = \frac{1}{\sqrt{M}} [\mathbf{v}(\vartheta(\psi_0)), \mathbf{v}(\vartheta(\psi_1)), \ldots, \mathbf{v}(\vartheta(\psi_{M-1}))] \quad (5)$$

$$[\mathbf{r}_k]_m = \beta_k M \cdot \mathsf{S}_k(\vartheta(\psi_{m-1})) [\vartheta(\psi_m) - \vartheta(\psi_{m-1})],$$
$$\text{for} \quad m = 1, 2, \ldots, M \quad (6)$$

where $\psi_{m'} = m'/M$ for $m' = 0, 1, \ldots, M$, and $\theta = \vartheta(\psi)$ over the support $[0, 1]$ is a strictly increasing continuous function[2] that satisfies $\vartheta(0) = \theta^{\min}$ and $\vartheta(1) = \theta^{\max}$. Then under Assumption 1, matrices $\mathbf{V}^H \mathbf{V}$ and $\mathbf{R}_k$ tend to be the identity matrix and $\mathbf{V}\mathrm{diag}\{\mathbf{r}_k\}\mathbf{V}^H$, respectively, when $M \to \infty$, in the sense that, for fixed positive integers $i$ and $j$,

$$\lim_{M \to \infty} [\mathbf{V}^H \mathbf{V} - \mathbf{I}_M]_{i,j} = 0 \quad (7)$$

$$\lim_{M \to \infty} [\mathbf{R}_k - \mathbf{V}\mathrm{diag}\{\mathbf{r}_k\}\mathbf{V}^H]_{i,j} = 0. \quad (8)$$

*Proof:* See Appendix A. ∎

[2]The function $\vartheta(\psi)$ can be interpreted as a mapping from the space domain to the physical angle domain, and it indeed depends on the BS array structure. We assume the function $\vartheta(\psi)$ to be strictly increasing and continuous over the support to guarantee that the function is a one-to-one mapping.

The result in Lemma 1 indicates that, when the BS antenna number $M$ is sufficiently large, the channel covariance matrix $\mathbf{R}_k$ can be well approximated by

$$\mathbf{R}_k \approx \mathbf{V}\mathrm{diag}\{\mathbf{r}_k\}\mathbf{V}^H. \quad (9)$$

Note that the matrix $\mathbf{V}$ tends to be unitary when $M$ is sufficiently large. This establishes a relationship between channel spatial correlations and channel power distribution in the angular domain. Specifically, for massive MIMO channels, eigenvector matrices of the channel covariance matrices for different UTs tend to be the same, and are determined by the BS array response vectors, while eigenvalues depend on the respective channel PASs.

*Remark 1:* When BS is equipped with the ULA, and the $M$ antennas are spaced with a half wavelength distance, the array response vector can be represented as [7]

$$\mathbf{v}(\theta) = \Big[1, \exp(-\bar{\jmath}\pi \sin(\theta)), \ldots,$$
$$\exp(-\bar{\jmath}\pi(M-1)\sin(\theta))\Big]^T. \quad (10)$$

We assume that the AoA interval equals $\mathcal{A} = [-\pi/2, \pi/2]$, and it is not hard to show (4) in Assumption 1, i.e., Assumption 1 is valid in this case. Let $\theta = \vartheta(\psi) = \arcsin(2\psi - 1)$, then $\vartheta(\psi_{m'}) = \arcsin\left(\frac{2m'}{M} - 1\right)$ for $m' = 0, 1, \ldots, M$, and elements of $\mathbf{V}$ reduce to $[\mathbf{V}]_{i,j} = \frac{1}{\sqrt{M}} \exp\left(-\bar{\jmath}2\pi \frac{(i-1)(j-1-M/2)}{M}\right)$ for $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, M$. This indicates that, for the ULA case, when $M$ is sufficiently large, eigenvector matrix of the channel covariance matrix can be well approximated by the unitary discrete Fourier transform (DFT) matrix (up to some matrix elementary operations). Similar channel covariance matrix decomposition for the ULA case was derived in [11] and [21], however, the result in Lemma 1 applies to the more general BS array configurations. In addition, the relationship between eigenvalues of the channel covariance matrix and channel PAS is established in Lemma 1. ∎

The channel model proposed in Lemma 1, as well as the ULA case in Remark 1, is based on Assumption 1, where angular resolution of the antenna array is assumed to tend to infinity when the antenna number grows to infinity. It is well known that angular resolution of an array is proportional to the array size [7], [8]. Thus, the channel model in Lemma 1 is applicable for arbitrary antenna array configuration with sufficiently large array size. However, in practical wireless communication systems, the antenna array size is always finite. Nevertheless, for a fixed array size, a fairly large number of antennas can still be accommodated at the BS if wireless transmission is performed over higher carrier frequency in, e.g., millimeter wave massive MIMO systems [22]. For example, considering the case where the antenna number equals 128 and the carrier frequency equals 30 GHz, which lies in the millimeter wave spectrum, the size of the ULA with half wavelength spacing considered in Remark 1 is just 0.64 m. Note that the channel model in Lemma 1 for the ULA case has been shown as a good approximation with finite but large

number of antennas [11], [21]. For these reasons, the proposed channel model is of great importance from both practical and theoretical perspectives.

In this paper, we employ the widely accepted assumption that channels are wide-sense stationary [8], thus channel covariance matrices can be obtained by the BS. However, stationarity of the realistic wireless channels can only be satisfied in a local manner, i.e., channel covariance matrices also vary over time. Thus, it requires that the channel covariance matrices being periodically estimated at the BS. Estimation of massive MIMO channel covariance matrices is rather challenging and resource-consuming [23]. However, from the result given by Lemma 1, only the eigenvalues rather than the whole massive MIMO channel covariance matrices need to be estimated, thus the number of parameters to estimate can be significantly reduced. In addition, the channel covariance matrices vary much less frequently than the instantaneous CSI, and thus can be estimated via averaging over time. Furthermore, the channel covariance matrices have been shown to stay constant over a wide frequency interval [24], and thus can be estimated via averaging over frequency in practical wideband systems. Therefore, there will be enough time-frequency resources to estimate the channel covariance matrices, and the estimation accuracy can be guaranteed in practice. In the rest of the paper, we will assume that the channel covariance matrices of all the UTs are known by the BS.

We assume that the channel elements to be jointly Gaussian from the law of large numbers, i.e., $\mathbf{g}_k \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_k\right)$. We assume that channels of different UTs are mutually statistically independent, and denote the UL channels of all the UTs as $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$.

## III. PR FOR UL CHANNEL TRAINING

In this section, we present PR for UL channel training, and investigate how PR affects the channel estimation performance. Our following analysis applies to arbitrary PR pattern, while how to form the PR pattern exploiting the statistical CSI will be discussed in Section V.

We denote the UT set as $\mathcal{K} = \{1, 2, \ldots, K\}$ where $k \in \mathcal{K}$ is the UT index. We assume that the UL training interval length equals $\tau(< K)$,[3] and all the UTs transmit the respective pilot sequences in the length of $\tau$ simultaneously during the training interval.[4] Note that the maximum number of available orthogonal sequences is equal to the sequence length, and we assume that the available orthogonal pilot sequence number equals $\tau$ for simplicity. We denote the available orthogonal pilot set as $\mathcal{T} = \{1, 2, \ldots, \tau\}$, and the $\pi^{\text{th}}$ pilot sequence as $\mathbf{x}_\pi \in \mathbb{C}^{\tau \times 1}$ where $\pi \in \mathcal{T}$ is the orthogonal pilot index. Different pilot sequences are assumed to satisfy the orthogonal

---

[3]It should be noted that the results obtained in the following are applicable for arbitrary $\tau$, and $\tau$ can be either set to be a fixed number, or determined dynamically by the BS. One example of how to dynamically determine $\tau$ based on the net spectral efficiency maximization criterion will be discussed in Section VI-C.

[4]For the case that one UT sends pilot signals during several particular channel uses while remains silent during other channel uses, the pilot signals can be seen as a specific pilot sequence with several non-zero entries at the corresponding channel uses.

condition that $\mathbf{x}_\pi^H \mathbf{x}_{\pi'} = \tau \sigma_{\text{x}}^{\text{p}} \cdot \delta\left(\pi - \pi'\right)$ where $\sigma_{\text{x}}^{\text{p}}$ is the pilot signal transmit power.

We denote an arbitrary PR pattern with UT set $\mathcal{K}$ and pilot set $\mathcal{T}$ as $\mathcal{P}(\mathcal{K}, \mathcal{T}) = \{(k, \pi_k) : k \in \mathcal{K}, \pi_k \in \mathcal{T}\}$ where $(k, \pi_k) \in \mathcal{P}(\mathcal{K}, \mathcal{T})$ denotes that the $\pi_k^{\text{th}}$ pilot sequence $\mathbf{x}_{\pi_k}$ is allocated to the $k^{\text{th}}$ UT. We use $\mathcal{K}_\pi = \{k : \pi_k = \pi\}$ to denote the set of the UTs using the $\pi^{\text{th}}$ pilot sequence.

With the PR pattern $\mathcal{P}(\mathcal{K}, \mathcal{T})$, the UTs transmit their assigned pilots periodically to enable the BS to estimate the channels. During the UL training phase of each coherence block, the received pilot signals at the BS can be written as

$$\mathbf{Y} = \mathbf{G}\mathbf{X} + \mathbf{N} \in \mathbb{C}^{M \times \tau} \tag{11}$$

where $\mathbf{G}$ is the UL channel matrix, $\mathbf{X} = [\mathbf{x}_{\pi_1}, \mathbf{x}_{\pi_2}, \ldots, \mathbf{x}_{\pi_K}]^T \in \mathbb{C}^{K \times \tau}$ is the UL pilot signal matrix, $\mathbf{N}$ is the independent additive Gaussian noise matrix with elements distributed as independently and identically $\mathcal{CN}\left(0, \sigma_{\text{z}}^{\text{p}}\right)$, and $\sigma_{\text{z}}^{\text{p}}$ is the noise power during the training phase. After decorrelation and power normalization of the received signals [2], the BS can obtain the channel observation of all the UTs. Specifically, for the $k^{\text{th}}$ UT in a given coherence block, the BS obtains the UL channel observation as

$$\mathbf{y}_{\pi_k}^{\text{p}} = \frac{1}{\sigma_{\text{x}}^{\text{p}} \tau} \mathbf{Y}\mathbf{x}_{\pi_k}^* = \frac{1}{\sigma_{\text{x}}^{\text{p}} \tau} \left(\sum_{\ell=1}^K \mathbf{g}_\ell \mathbf{x}_{\pi_\ell}^T + \mathbf{N}\right) \mathbf{x}_{\pi_k}^*$$

$$= \sum_{\ell=1}^K \mathbf{g}_\ell \cdot \delta\left(\pi_\ell - \pi_k\right) + \frac{1}{\sigma_{\text{x}}^{\text{p}} \tau} \mathbf{N}\mathbf{x}_{\pi_k}^*$$

$$= \sum_{\ell \in \mathcal{K}_{\pi_k}} \mathbf{g}_\ell + \frac{1}{\sigma_{\text{x}}^{\text{p}} \tau} \mathbf{N}\mathbf{x}_{\pi_k}^*. \tag{12}$$

With the property of unitary transformation, it is not hard to show that the noise term $\frac{1}{\sigma_{\text{x}}^{\text{p}} \tau} \mathbf{N}\mathbf{x}_{\pi_k}^*$ in (12) is still Gaussian with elements distributed as independently and identically $\mathcal{CN}\left(0, \frac{\sigma_{\text{z}}^{\text{p}}}{\sigma_{\text{x}}^{\text{p}} \tau}\right)$. Let $\rho^{\text{p}} = \sigma_{\text{x}}^{\text{p}}/\sigma_{\text{z}}^{\text{p}}$ be the UL channel training signal-to-noise ratio (SNR), then (12) can be rewritten as

$$\mathbf{y}_{\pi_k}^{\text{p}} = \sum_{\ell \in \mathcal{K}_{\pi_k}} \mathbf{g}_\ell + \frac{1}{\sqrt{\rho^{\text{p}} \tau}} \mathbf{n}_{\pi_k}^{\text{p}}$$

$$= \mathbf{g}_k + \underbrace{\sum_{\ell \in \mathcal{K}_{\pi_k} \backslash \{k\}} \mathbf{g}_\ell}_{\text{pilot interference}} + \underbrace{\frac{1}{\sqrt{\rho^{\text{p}} \tau}} \mathbf{n}_{\pi_k}^{\text{p}}}_{\text{pilot noise}} \tag{13}$$

where "\" denotes the set subtraction operation, and $\mathbf{n}_{\pi_k}^{\text{p}} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_M\right)$ is the normalized additive noise. Note that $\mathcal{K}_{\pi_k}$ represents the set of the UTs using the same pilot as the $k^{\text{th}}$ UT, and the BS has to estimate the channels of all the UTs reusing the $\pi_k^{\text{th}}$ pilot based on the observation $\mathbf{y}_{\pi_k}^{\text{p}}$. The MMSE estimate of the channel $\mathbf{g}_k$ based on the channel observation $\mathbf{y}_{\pi_k}^{\text{p}}$ is given by

$$\hat{\mathbf{g}}_k = \mathbf{R}_k \mathbf{C}_{\pi_k}^{-1} \mathbf{y}_{\pi_k}^{\text{p}} \tag{14}$$

where

$$\mathbf{C}_{\pi_k} \triangleq \sum_{\ell \in \mathcal{K}_{\pi_k}} \mathbf{R}_\ell + \frac{1}{\rho^{\text{p}} \tau} \mathbf{I}. \tag{15}$$

From the orthogonality principle of MMSE estimation [25],

channel estimation error $\tilde{\mathbf{g}}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k$ is independent of $\hat{\mathbf{g}}_k$, and the covariance of $\tilde{\mathbf{g}}_k$ is

$$\mathbf{R}_{\tilde{\mathbf{g}}_k} = \mathbf{R}_k - \mathbf{R}_k \mathbf{C}_{\pi_k}^{-1} \mathbf{R}_k. \tag{16}$$

Note that $\hat{\mathbf{g}}_k$ and $\mathbf{R}_{\tilde{\mathbf{g}}_k}$ are also mean and covariance of $\mathbf{g}_k$ conditioned on $\mathbf{y}_{\pi_k}^{\mathrm{p}}$, respectively [25].

The estimation error covariance is an important measure of the estimation performance, and we define the MSE-CE as

$$\epsilon^{\mathrm{p}} \triangleq \sum_{k=1}^{K} \mathrm{tr}\left\{ \mathbf{R}_{\tilde{\mathbf{g}}_k} \right\}. \tag{17}$$

Before we proceed, we first define the orthogonality between two arbitrary Hermitian positive semi-definite matrices using the angle $(0 \le \theta \le \pi/2)$ between them as

$$\theta(\mathbf{A}, \mathbf{B}) \triangleq \arccos \frac{\mathrm{tr}\left\{ \mathbf{A}^H \mathbf{B} \right\}}{\|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{B}\|_{\mathrm{F}}} = \arccos \frac{\mathrm{tr}\left\{ \mathbf{A}\mathbf{B} \right\}}{\|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{B}\|_{\mathrm{F}}},$$
$$\text{for} \quad \mathbf{A}, \mathbf{B} \succeq \mathbf{0}. \tag{18}$$

Then we present a condition under which the MSE-CE defined in (17) can be minimized in the following theorem.

*Theorem 1:* The minimum value of the MSE-CE $\epsilon^{\mathrm{p}}$ is given by

$$\varepsilon^{\mathrm{p}} = \sum_{k=1}^{K} \mathrm{tr}\left\{ \mathbf{R}_k - \mathbf{R}_k \left( \mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau} \mathbf{I} \right)^{-1} \mathbf{R}_k \right\} \tag{19}$$

and the minimum is achieved under the condition that, for $\forall i, j \in \mathcal{K}$ and $i \ne j$,

$$\theta(\mathbf{R}_i, \mathbf{R}_j) = \frac{\pi}{2}, \qquad \text{when} \quad \pi_i = \pi_j. \tag{20}$$

*Proof:* See Appendix B. ■

In the MSE-CE metric defined in (17), correlations between the channel estimation errors seen by different UTs are not taken into account. Actually, the correlations between the channel estimation errors of the UT $i$ and the UT $j$ $(j \ne i)$ can be obtained as

$$\mathbb{E}\left\{ \tilde{\mathbf{g}}_i \tilde{\mathbf{g}}_j^H \right\} = \mathbb{E}\left\{ \left( \mathbf{g}_i - \mathbf{R}_i \mathbf{C}_{\pi_i}^{-1} \mathbf{y}_{\pi_i}^{\mathrm{p}} \right) \left( \mathbf{g}_j - \mathbf{R}_j \mathbf{C}_{\pi_j}^{-1} \mathbf{y}_{\pi_j}^{\mathrm{p}} \right)^H \right\}$$
$$= -\mathbf{R}_i \mathbf{C}_{\pi_i}^{-1} \mathbf{R}_j \cdot \delta(\pi_i - \pi_j) \tag{21}$$

which indicates that channel estimation errors of the UTs with orthogonal pilots are independent, while those of the UTs reusing the pilots are correlated. However, if the condition given in Theorem 1 is satisfied, then $-\mathbf{R}_i \mathbf{C}_{\pi_i}^{-1} \mathbf{R}_j = \mathbf{0}$, i.e., channel estimation errors seen by different UTs will be uncorrelated no matter whether they reuse the pilots or not, and the condition given in Theorem 1 is still optimal.

To obtain clear insights of Theorem 1, we consider the asymptotic antenna number case, and the following corollary can be readily obtained from Lemma 1.

*Corollary 1:* When the BS antenna number $M \to \infty$, the MSE-CE $\epsilon^{\mathrm{p}}$ can be minimized provided that, for $\forall i, j \in \mathcal{K}$ and $i \ne j$,

$$\langle \mathbf{r}_i, \mathbf{r}_j \rangle = 0, \qquad \text{when} \quad \pi_i = \pi_j \tag{22}$$

where $\mathbf{r}_i$ for fixed positive integer $i$ is given in Lemma 1. ■

The result in Corollary 1 indicates that the MSE-CE $\epsilon^{\mathrm{p}}$ can be minimized if the UTs reusing the pilots have non-overlapping channel AoA intervals. This result is very intuitive, as in such cases, the channels of different UTs are strictly separated in the angular domain, and the pilot interference does not take into effect. Moreover, in the high SNR regime where the training SNR $\rho^{\mathrm{p}} \to \infty$, the pilot noise vanishes, and then the MSE-CE $\varepsilon^{\mathrm{p}} \to 0$, which implies that channel estimations tend to be perfect.

Although the conditions in Theorem 1 and Corollary 1 are desirable, they cannot always be well satisfied. However, in realistic outdoor propagation environments where the BS is located at an elevated position, channel AS seen by the BS is usually small [8], [26], which indicates that most of the channel power is concentrated in a narrow angle interval, and the channel power outside this angle interval is very small. For UTs located geographically apart in different spatial directions, the overlaps of their channel power in the angular domain might be neglected, and thus PR becomes feasible in such spatially correlated massive MIMO channels.

## IV. ROBUST UL/DL DATA TRANSMISSIONS

In the previous section, we showed feasibility of PR for massive MIMO transmission, and presented UL channel training with PR. In each coherence block, the BS obtains the channel estimates of all the UTs after UL channel training. The conventional data transmission design in massive MIMO treats the channel estimates as the real channels. However, with PR, the channel estimation performance will degrade in most cases, thus a robust data transmission design with channel estimation errors taken into account is of paramount importance in the considered PR based massive MIMO transmission. There are two main approaches to design a wireless system robust to the channel uncertainty: the worst-case approach and the statistical approach. In the worst-case approach, the channel uncertainty is modeled as being within a given set around the channel estimate, and a worst-case transmission performance can be guaranteed [27]. In the statistical approach, the channel uncertainty is modeled using the channel statistics, such as the mean and the covariance, and a statistical average performance can be guaranteed [28]. In this work, we employ the statistical approach to model the channel uncertainty. Specifically, in each coherence block, based on the received pilot signals, the CSI uncertainty at the BS can be modeled statistically using its conditional distribution, i.e., the conditional mean (the MMSE channel estimate) and the conditional covariance (the covariance of the channel estimation error). Note that the channel estimation error covariance $\mathbf{R}_{\tilde{\mathbf{g}}_k}$ given in (16) depends on the PR pattern $\mathcal{P}(\mathcal{K}, \mathcal{T})$ and the channel covariance, and thus can be known by the BS. In the following, we will develop robust data transmissions for UL and DL, respectively, under the MMSE-SD criterion.

### A. Robust UL Data Transmission

During the UL data transmission phase, the signal received at the BS at a channel use in the given coherence block can be expressed as

$$\mathbf{y}^{\mathrm{u}} = \mathbf{G}\mathbf{a}^{\mathrm{u}} + \frac{1}{\sqrt{\rho^{\mathrm{u}}}} \mathbf{n}^{\mathrm{u}} = \left( \hat{\mathbf{G}} + \tilde{\mathbf{G}} \right) \mathbf{a}^{\mathrm{u}} + \frac{1}{\sqrt{\rho^{\mathrm{u}}}} \mathbf{n}^{\mathrm{u}} \tag{23}$$

where $\hat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \ldots, \hat{\mathbf{g}}_K]$ is the channel estimate, $\tilde{\mathbf{G}} = [\tilde{\mathbf{g}}_1, \ldots, \tilde{\mathbf{g}}_K]$ is the channel estimation error, $\mathbf{a}^{\mathrm{u}} \in \mathbb{C}^{K \times 1}$ with mean $\mathbf{0}$ and covariance $\mathbf{I}_K$ denotes the UL data signal vector where $[\mathbf{a}^{\mathrm{u}}]_k$ is the signal sent by the $k^{\mathrm{th}}$ UT, $\mathbf{n}^{\mathrm{u}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the independent additive noise, and $\rho^{\mathrm{u}}$ is the UL data transmission SNR per UT.

We consider the linear receiver at the BS

$$\hat{\mathbf{a}}^{\mathrm{u}} = \mathbf{W}^T \mathbf{y}^{\mathrm{u}} \qquad (24)$$

and then MSE-SD of the UL transmission in the given coherence block can be defined as

$$\epsilon^{\mathrm{u}} \triangleq \mathbb{E}\left\{ \|\hat{\mathbf{a}}^{\mathrm{u}} - \mathbf{a}^{\mathrm{u}}\|_2^2 \right\} \qquad (25)$$

where the expectation is with respect to $\mathbf{a}^{\mathrm{u}}$, $\mathbf{n}^{\mathrm{u}}$, and the channel estimation error $\tilde{\mathbf{G}}$.

Finding the optimal UL receiver based on the MMSE-SD criterion can be formulated as

$$\min_{\mathbf{W}} \quad \epsilon^{\mathrm{u}} \qquad (26)$$

and we present the solution in the following theorem.

*Theorem 2:* The optimal solution to the problem (26) is given by

$$\mathbf{W}^{\mathrm{opt}} = \left[ \left( \hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I} \right)^{-1} \hat{\mathbf{G}} \right]^* \qquad (27)$$

and the corresponding MSE-SD is given by

$$\epsilon^{\mathrm{u,min}} = \mathrm{tr}\left\{ \left( \mathbf{I} + \hat{\mathbf{G}}^H \left( \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I} \right)^{-1} \hat{\mathbf{G}} \right)^{-1} \right\}. \qquad (28)$$

*Proof:* See Appendix C. ∎

For the conventional receiver with channel estimates assumed to be accurate, the impact of the channel estimation error is omitted. While for our robust MMSE receiver design, the channel estimation error due to PR is taken into account. Specifically, the expectation in (25) accounts for the channel estimation error $\tilde{\mathbf{G}}$, which leads to our robust MMSE receiver. Note that the robust MMSE receiver given in (27) exhibits a similar structure to the conventional receiver. When $\sum_{n=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_n} \to \mathbf{0}$, the robust MMSE receiver in (27) reduces to the conventional receiver

$$\mathbf{W}^{\mathrm{con}} = \left[ \left( \hat{\mathbf{G}}\hat{\mathbf{G}}^H + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I} \right)^{-1} \hat{\mathbf{G}} \right]^*. \qquad (29)$$

### B. Robust DL Data Transmission

During the DL data transmission phase, the signal received at the UTs at a channel use in the given coherence block can be expressed as

$$\mathbf{y}^{\mathrm{d}} = \mathbf{G}^T \mathbf{B} \mathbf{a}^{\mathrm{d}} + \frac{1}{\sqrt{\rho^{\mathrm{d}}}}\mathbf{n}^{\mathrm{d}} = \left( \hat{\mathbf{G}} + \tilde{\mathbf{G}} \right)^T \mathbf{B} \mathbf{a}^{\mathrm{d}} + \frac{1}{\sqrt{\rho^{\mathrm{d}}}}\mathbf{n}^{\mathrm{d}} \qquad (30)$$

where the DL channel $\mathbf{G}^T$ is the transpose of the UL channel due to the channel reciprocity of the TDD systems [2], $\mathbf{a}^{\mathrm{d}} \in$

$\mathbb{C}^{K \times 1}$ with mean $\mathbf{0}$ and covariance $\mathbf{I}_K$ denotes the DL data signal vector where $[\mathbf{a}^{\mathrm{d}}]_k$ is the signal for the $k^{\mathrm{th}}$ UT, $\mathbf{n}^{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the independent additive noise, $\rho^{\mathrm{d}}$ is the average DL data transmission SNR per UT, and $\mathbf{B}$ is the DL linear precoding matrix which satisfies the power constraint

$$\mathrm{tr}\left\{ \mathbf{B}\mathbf{B}^H \right\} \leq K. \qquad (31)$$

Then MSE-SD of the DL transmission in the given coherence block can be defined as

$$\epsilon^{\mathrm{d}} \triangleq \mathbb{E}\left\{ \|\alpha\mathbf{y}^{\mathrm{d}} - \mathbf{a}^{\mathrm{d}}\|_2^2 \right\} \qquad (32)$$

where the expectation is with respect to $\mathbf{a}^{\mathrm{d}}$, $\mathbf{n}^{\mathrm{d}}$, and $\tilde{\mathbf{G}}$, and $\alpha$ is a real scalar parameter corresponding to the potential power scaling performed at the UTs.[5]

Finding the optimal DL precoder based on the MMSE-SD criterion can be formulated as

$$\begin{aligned} \min_{\mathbf{B},\alpha} \quad & \epsilon^{\mathrm{d}} \\ \text{subject to} \quad & \mathrm{tr}\left\{ \mathbf{B}\mathbf{B}^H \right\} \leq K \end{aligned} \qquad (33)$$

and we present the solution in the following theorem.

*Theorem 3:* The optimal solution to the problem (33) is given by

$$\mathbf{B}^{\mathrm{opt}} = \frac{1}{\gamma^{\mathrm{opt}}} \left[ \left( \hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I} \right)^{-1} \hat{\mathbf{G}} \right]^* \qquad (34)$$

$$\alpha^{\mathrm{opt}} = \gamma^{\mathrm{opt}} \qquad (35)$$

where $\gamma^{\mathrm{opt}}$ is chosen to satisfy the power normalization constraint $\mathrm{tr}\left\{ \mathbf{B}^{\mathrm{opt}}(\mathbf{B}^{\mathrm{opt}})^H \right\} = K$, its value is given by

$$\gamma^{\mathrm{opt}} = \sqrt{ \frac{ \mathrm{tr}\left\{ \hat{\mathbf{G}}^H \left( \hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I} \right)^{-2} \hat{\mathbf{G}} \right\} }{ K } } \qquad (36)$$

and the corresponding MSE-SD is given by

$$\epsilon^{\mathrm{d,min}} = \mathrm{tr}\left\{ \left( \mathbf{I} + \hat{\mathbf{G}}^H \left( \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I} \right)^{-1} \hat{\mathbf{G}} \right)^{-1} \right\}. \qquad (37)$$

*Proof:* See Appendix D. ∎

From Theorem 3, we can observe that, similarly to the UL case, the robust MMSE precoder in (34) also embraces the structure of the conventional precoder.

### C. UL-DL Duality

From the results in Theorem 2 and Theorem 3, we can readily obtain the following UL-DL MMSE duality.

*Corollary 2:* In each TDD coherence block, if $\rho^{\mathrm{u}} = \rho^{\mathrm{d}}$, then $\mathbf{B}^{\mathrm{opt}} = \mathbf{W}^{\mathrm{opt}}/\gamma^{\mathrm{opt}}$, and $\epsilon^{\mathrm{u,min}} = \epsilon^{\mathrm{d,min}}$. ∎

The result in Corollary 2 indicates that, in the same TDD coherence block, if the UL data transmission SNR equals the DL data transmission SNR, then the robust DL precoder in

---

[5]Note that $\alpha$ is a real scalar, and the overhead for the UTs to obtain it can be neglected.

(34) can be achieved by the robust UL receiver in (27) with proper power normalization, and the complexity of computing the robust MMSE DL precoder can be reduced. In addition, if the robust MMSE receiver and robust MMSE precoder are used for data transmissions, then the same MMSE-SD can be achieved in both the UL and DL in each TDD coherence block. Note that similar UL-DL duality based on the perfect CSI assumption was provided in the literature such as [29] and [30], however, our result in Corollary 2 is established on the pilot-assisted CSI acquisition assumption.

In this section, we have investigated the robust UL and DL data transmissions with channel estimation error due to PR taken into account. It will be seen in the following section that, PR based massive MIMO transmission, which combines pilot scheduling and robust data transmission, can achieve the MMSE-SD optimality for both the UL and DL.

## V. PILOT SCHEDULING

Up to now, we have investigated channel training and data transmission of the massive MIMO transmission with PR, and the obtained results are applicable to arbitrary PR pattern. In this section, we study pilot scheduling which exploits the long term statistical CSI to allocate the available pilot signals to the UTs, and we focus on two MMSE related criteria.

### A. MMSE-CE Criterion

CSI is critical to massive MIMO transmission, and it is natural to design the pilot scheduler based on the MMSE-CE criterion, which leads to the following problem

$$\min_{\mathcal{P}(\mathcal{K},\mathcal{T})} \quad \epsilon^{\text{p}} \tag{38}$$

where $\epsilon^{\text{p}}$ is defined in (17).

The pilot scheduling problem in (38) is combinatorial, and the optimal PR pattern $\mathcal{P}(\mathcal{K},\mathcal{T})$ can be found through exhaustive search (ES). The complexity of the ES in (38), in terms of the (complex) scalar multiplication number which dominates the computational complexity, is briefly evaluated as follows. Recalling (17), the scalar multiplication number required in evaluation of the objective function in (38) is $\mathcal{O}(M^3 K)$. Thus, the computational complexity of running ES under the MMSE-CE criterion is $\mathcal{O}(\tau^K M^3 K)$.

### B. MMSE-SD Criterion

MSE-SD is an important performance measure of the data transmission, and in the sequel we study the pilot scheduler design regarding the MSE-SD metric. Due to the UL-DL MMSE-SD duality in each coherence block presented in Corollary 2, we assume that $\rho^{\text{u}} = \rho^{\text{d}}$ for simplicity, and denote that $\rho^{\text{u}} = \rho^{\text{d}} = \rho^{\text{t}}$ and $\epsilon^{\text{u,min}} = \epsilon^{\text{d,min}} = \epsilon^{\text{t,min}}$, where the superscript "t" stands for expression related to data transmission. We consider pilot scheduling under the MMSE-SD criterion, which can be formulated as

$$\min_{\mathcal{P}(\mathcal{K},\mathcal{T})} \quad \mathbb{E}\left\{\epsilon^{\text{t,min}}\right\}$$
$$= \mathbb{E}\left\{\text{tr}\left\{\left(\mathbf{I} + \hat{\mathbf{G}}^H \left(\mathbf{R}^{\text{t,n,eff}}\right)^{-1} \hat{\mathbf{G}}\right)^{-1}\right\}\right\} \tag{39}$$

where the expectation is with respect to the channel fading and the noise distributions, and the effective noise covariance matrix is defined as

$$\mathbf{R}^{\text{t,n,eff}} \triangleq \sum_{k=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\text{t}}}\mathbf{I}. \tag{40}$$

The objective function in (39) is the average of the MMSE-SD that can be achieved by the robust MMSE receiver and robust MMSE precoder in each coherence block, and it depends on the statistics of the channel fading and pilot noise distributions. It should be noted that here we still use the term MMSE-SD for brevity, however the meaning of it differs from that when we consider the designs of the receiver and precoder in the previous section.

Due to the difficulty in obtaining the closed-form expression of the objective function $\mathbb{E}\left\{\epsilon^{\text{t,min}}\right\}$ in (39), we first present a lower bound of it in the following lemma.

*Lemma 2:* The average MSE-SD $\mathbb{E}\left\{\epsilon^{\text{t,min}}\right\}$ is lower bounded by

$$\mathbb{E}\left\{\epsilon^{\text{t,min}}\right\} \geq \epsilon^{\text{t,alb}} = \text{tr}\left\{\left(\mathbf{I}_K + \mathbf{\Omega}\right)^{-1}\right\} \tag{41}$$

where for fixed positive integers $i$ and $j$,

$$[\mathbf{\Omega}]_{i,j} = \text{tr}\left\{\mathbf{C}_{\pi_i}^{-1}\mathbf{R}_i \left(\mathbf{R}^{\text{t,n,eff}}\right)^{-1} \mathbf{R}_j\right\} \cdot \delta\left(\pi_i - \pi_j\right). \tag{42}$$

*Proof:* See Appendix E. ∎

It will be seen in Section VI-A that, the lower bound presented in Lemma 2 is tight over a wide SNR region. By replacing the objective function $\mathbb{E}\left\{\epsilon^{\text{t,min}}\right\}$ with its lower bound presented in Lemma 2, the pilot scheduling problem (39) can be simplified as

$$\min_{\mathcal{P}(\mathcal{K},\mathcal{T})} \quad \epsilon^{\text{t,alb}}. \tag{43}$$

The pilot scheduling problem in (43) is also combinatorial. The optimal PR pattern $\mathcal{P}(\mathcal{K},\mathcal{T})$ can be found through ES. Note that the scalar multiplication number required in evaluation of the objective function in (43) is $\mathcal{O}(M^3 K^2)$, thus, the computational complexity of running ES under the MMSE-SD criterion is $\mathcal{O}(\tau^K M^3 K^2)$.

Before we proceed, we present a condition under which $\epsilon^{\text{t,alb}}$ can be minimized in the following theorem.

*Theorem 4:* The minimum value of the lower bound average MSE-SD $\epsilon^{\text{t,alb}}$ is given by

$$\varepsilon^{\text{t}} = \sum_{i=1}^{K} \frac{1}{1 + [\boldsymbol{\omega}]_i} \tag{44}$$

where $[\boldsymbol{\omega}]_i$ for fixed positive integer $i$ is given by (45), shown at the top of the next page, and the minimum is achieved under the condition that, for $\forall i, j \in \mathcal{K}$ and $i \neq j$,

$$\theta\left(\mathbf{R}_i, \mathbf{R}_j\right) = \frac{\pi}{2}, \qquad \text{when} \quad \pi_i = \pi_j. \tag{46}$$

*Proof:* See Appendix F. ∎

Recalling Lemma 1, we can readily obtain the following corollary.

*Corollary 3:* When the BS antenna number $M \to \infty$, the lower bound average MSE-SD $\epsilon^{\text{t,alb}}$ can be minimized provided

$$[\boldsymbol{\omega}]_i = \mathrm{tr}\left\{\left(\mathbf{R}_i + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_i\left[\sum_{k=1}^{K}\left(\mathbf{R}_k - \mathbf{R}_k\left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_k\right) + \frac{1}{\rho^{\mathrm{t}}}\mathbf{I}\right]^{-1}\mathbf{R}_i\right\} \qquad (45)$$

that, for $\forall i, j \in \mathcal{K}$ and $i \neq j$,

$$\langle \mathbf{r}_i, \mathbf{r}_j \rangle = 0, \qquad \text{when} \quad \pi_i = \pi_j \qquad (47)$$

where $\mathbf{r}_i$ for fixed positive integer $i$ is given in Lemma 1. ∎

Interestingly, conditions for optimal data transmission obtained in Theorem 4 and Corollary 3 are the same as those for optimal channel training obtained in Theorem 1 and Corollary 1. The intuitive interpretation lies in that, for massive MIMO transmission, if channels of the UTs reusing the pilots can be rigorously spatially separated, then not only the pilot interference but also the transmission data interference vanishes. Furthermore, in the high SNR regime where both the training SNR $\rho^{\mathrm{p}}$ and transmission SNR $\rho^{\mathrm{t}}$ tend to infinity, the remaining additive noise vanishes, and the average MSE-SD $\varepsilon^{\mathrm{t}} \to 0$. This result shows that the PR based transmission scheme, which combines pilot scheduling and robust data transmission, can achieve the MMSE-SD optimality.

### C. SGPS Algorithm

In the above subsections, we have investigated pilot scheduling under two MMSE related criteria. In both cases, the designs are formulated as combinatorial optimization problems, and the optimal PR patterns can be formed through ES. However, due to the exponential complexity, ES becomes hard to implement in practice as the UT number grows.

In this subsection, we propose a low complexity pilot scheduling algorithm called the statistical greedy pilot scheduling (SGPS) algorithm which is motivated by the conditions for optimal channel estimation and data transmission given in Theorem 1 and Theorem 4, and the main idea is that channel covariance matrices of the UTs reusing the pilots should be as orthogonal as possible. Detailed description of the SGPS algorithm is summarized in Algorithm 1. Coordinated pilot allocation algorithm for mitigating the inter-cell pilot contamination using similar idea was proposed in [11], however, the above SGPS algorithm is dedicated for the single-cell scenario.

We evaluate the complexity of the SGPS algorithm as follows. In the process of the SGPS algorithm, no more than $\sum_{m=1}^{K-1} m(K-m) = (K-1)K(K+1)/6$ orthogonality calculations defined in (18) are needed. Note that the scalar multiplication number needed in each orthogonality calculation is $\mathcal{O}(M^2)$, thus, the computational complexity of running the SGPS algorithm is $\mathcal{O}(M^2 K^3)$. In above subsections, we have shown that the ES complexity under the MMSE-CE and MMSE-SD criteria are $\mathcal{O}(\tau^K M^3 K)$ and $\mathcal{O}(\tau^K M^3 K^2)$, respectively. This indicates that the SGPS algorithm gives a significant computational complexity reduction compared with ES. Meanwhile, simulation results in Section VI-B will show that performances of the low complexity SGPS algorithm can closely approach those of ES.

---

**Algorithm 1** Statistical Greedy Pilot Scheduling (SGPS) Algorithm

---

**Input:** The UT set $\mathcal{K} = \{1, 2, \ldots, K\}$ and the channel covariance information $\mathbf{R}_k (k \in \mathcal{K})$, the orthogonal pilot set $\mathcal{T}$ with the pilot length $\tau (1 < \tau < K)$

**Output:** PR pattern $\mathcal{P}(\mathcal{K}, \mathcal{T}) = \{(k, \pi_k) : k \in \mathcal{K}, \pi_k \in \mathcal{T}\}$

1: Initialize the unscheduled UT set $\mathcal{K}^{\mathrm{un}} = \mathcal{K}$, the unused pilot set $\mathcal{T}^{\mathrm{un}} = \mathcal{T}$

  **Step 1)** Schedule the UTs with "similar" channel covariance matrices and assign them with orthogonal pilots

2: $m_1 = 1$, $\pi_1 = 1$, $\mathcal{K}_1 = \{1\}$, $\mathcal{K}^{\mathrm{un}} \leftarrow \mathcal{K}^{\mathrm{un}} \setminus \{1\}$, $\mathcal{T}^{\mathrm{un}} \leftarrow \mathcal{T}^{\mathrm{un}} \setminus \{1\}$

3: **while** $\mathcal{T}^{\mathrm{un}} \neq \varnothing$ **do**

4:   For the pilot $t \in \mathcal{T}^{\mathrm{un}}$, select the UT $m_t = \arg\max_{\ell \in \mathcal{K}^{\mathrm{un}}} \sum_{j \in \mathcal{T} \setminus \mathcal{T}^{\mathrm{un}}} \cos\theta\left(\mathbf{R}_\ell, \mathbf{R}_{m_j}\right)$

5:   Assign the pilot $t$ to the UT $m_t$, $\pi_{m_t} = t$, $\mathcal{K}_t = \{m_t\}$

6:   Update $\mathcal{K}^{\mathrm{un}} \leftarrow \mathcal{K}^{\mathrm{un}} \setminus \{m_t\}$, $\mathcal{T}^{\mathrm{un}} \leftarrow \mathcal{T}^{\mathrm{un}} \setminus \{t\}$

7: **end while**

  **Step 2)** Each unscheduled UT is assigned with the "best" pilot so that the channel covariance matrices of the UTs reusing the pilots are as orthogonal as possible

8: **while** $\mathcal{K}^{\mathrm{un}} \neq \varnothing$ **do**

9:   For the UT $k \in \mathcal{K}^{\mathrm{un}}$, select the pilot $n_k = \arg\min_{q \in \mathcal{T}} \sum_{s \in \mathcal{K}_q} \cos\theta\left(\mathbf{R}_k, \mathbf{R}_s\right)$

10:   Assign the pilot $n_k$ to the UT $k$, $\pi_k = n_k$, $\mathcal{K}_{n_k} \leftarrow \mathcal{K}_{n_k} \cup \{k\}$

11:   Update $\mathcal{K}^{\mathrm{un}} \leftarrow \mathcal{K}^{\mathrm{un}} \setminus \{k\}$

12: **end while**

---

### VI. NUMERICAL RESULTS

In this section, we present numerical simulations to evaluate performances of the proposed PR based massive MIMO transmission. We assume that the BS is equipped with the 128-antenna ULA, and the antennas are spaced with a half wavelength distance. We set the AoA interval as $\mathcal{A} = [-\pi/2, \pi/2]$. We consider the typical outdoor wireless propagation environments where the channel PAS can be modeled as the truncated Laplacian distribution [18], [31] given by (48), shown at the top of the next page, where $\sigma_k$ and $\theta_k$ represent the AS and the mean AoA of the $k^{\mathrm{th}}$ UT's channel, respectively. We assume that channel ASs are the same for all the UTs so that $\sigma_k = \sigma$ ($\forall k$). We assume that all the UTs are of equal distance from the BS, and set the large scale fading coefficients as $\beta_k = 1$ ($\forall k$). We assume that the UTs uniformly locate in a $120°$ sector, i.e., the mean channel AoA $\theta_k$ is uniformly distributed in the angle interval $[-\pi/3, \pi/3]$ in radian. The channel covariance matrices of the UTs are generated according to the model given by Remark 1, and we impose the constraint in (3) for channel power normalization. We assume that the

$$S_k^{\text{Lap}}(\theta) = \frac{1}{\sqrt{2}\sigma_k \left(1 - \exp\left(-\sqrt{2}\pi/\sigma_k\right)\right)} \cdot \exp\left(\frac{-\sqrt{2}\,|\theta - \theta_k|}{\sigma_k}\right), \qquad \text{for } \theta \in [\theta_k - \pi, \theta_k + \pi] \qquad (48)$$

channel training SNR and the data transmission SNR are equal such that $\rho^{\text{p}} = \rho^{\text{u}} = \rho^{\text{d}} = \rho$.

### A. Performance of Robust Transmission

In this subsection, we employ the average MSE-SD metric to evaluate performances of the robust receiver and precoder developed in Section IV. Due to the UL-DL MMSE duality given in Corollary 2, we only consider the UL transmission case for brevity.

We compare performances of the robust MMSE receiver given in (27) with those of the conventional receiver given in (29). We consider the case with $K = 10$, $\sigma = 10°$, and the mean channel AoAs of the UTs from UT 1 to UT 10 are

$$[0.6592, 0.8499, -0.7812, 0.8658, 0.2772,$$
$$- 0.8429, -0.4639, 0.0982, 0.9582, 0.9737]$$

in radian. We assume that the pilot length equals $\tau = 5$, and consider two PR patterns. Specifically, the pilot indices that the UTs use from UT 1 to UT 10 are $[1, 1, 2, 2, 3, 3, 4, 4, 5, 5]$ and $[1, 2, 3, 4, 3, 5, 4, 5, 5, 3]$ for PR patterns A and B, respectively.[6] In Fig. 1, we plot the average MSE-SD performances of the robust MMSE receiver (using the true and the estimated channel covariance matrices that are obtained via averaging over 100 samples, respectively) and the conventional receiver. The lower bound of the average MSE-SD achieved by the robust MMSE receiver given in Lemma 2 is also shown. We can have the following observations: 1) the average MSE-SD performance loss using the estimated channel covariance matrices compared with true channel covariance matrices can be almost neglected; 2) the robust MMSE receiver outperforms the conventional receiver, especially in high SNR regime where pilot interference dominates; 3) compared with the robust MMSE receiver, the conventional receiver is quite sensitive to the channel estimation error, and increasing the SNR may result in additional MSE-SD for the conventional receiver; 4) the closed-form lower bound of the average MSE-SD given in Lemma 2 is tight over a wide SNR region for different PR patterns; 5) pilot scheduling is crucial to the data transmission performance.

### B. Performance of SGPS Algorithm

In this subsection, we evaluate performances of the SGPS algorithm, and compare them with those of ES. In Fig. 2 and Fig. 3, we plot the MSE-CE metric in (17) and the average MSE-SD metric in (41) versus the pilot length for different values of SNR with $K = 10$ and $\sigma = 10°$, respectively. It can be observed that, in both cases the performances of the

[6]PR pattern B is determined by the SGPS algorithm proposed in Section V-C, and we arbitrarily set a PR pattern as pattern A for comparison. We employ such setting to exemplify that pilot scheduling is crucial to the transmission performance.



Fig. 1. Comparison of the average MSE-SD performances between the robust MMSE receiver (using the true and the estimated channel covariance matrices that are obtained via averaging over 100 samples, respectively) and the conventional receiver. Results are shown versus the SNR for two specific PR patterns with $K = 10$, $\tau = 5$ and $\sigma = 10°$. The lower bound of the average MSE-SD achieved by the robust MMSE receiver is also depicted.



Fig. 2. Comparison of the MSE-CE performances between the SGPS algorithm and ES. Results are shown versus the pilot length for different values of SNR $\rho$ with $K = 10$ and $\sigma = 10°$.

SGPS algorithm closely approach those of ES over a wide SNR region for different values of pilot length.

### C. Net Spectral Efficiency Comparison

In this subsection, we compare the *net* spectral efficiency performance between the proposed PR scheme and the conventional OT scheme. The net spectral efficiency is given by

$$R^{\text{net}} = \left(1 - \frac{\tau}{T}\right) R^{\text{ach}} \qquad (49)$$

Fig. 3. Comparison of the average MSE-SD performances between the SGPS algorithm and ES. Results are shown versus the pilot length for different values of SNR $\rho$ with $K = 10$ and $\sigma = 10°$.



Fig. 4. Comparison of the UL spectral efficiency performances between the PR scheme and the OT scheme. Results are shown versus the coherence block length $T$ for different values of AS $\sigma$ with $K = 10$ and $\rho = 20$ dB.

and the achievable rate $R^{\mathrm{ach}}$ can be set as the UL sum achievable rate $R^{\mathrm{u,sum}}$, or the DL sum achievable rate $R^{\mathrm{d,sum}}$, or the weighted summation of $R^{\mathrm{u,sum}}$ and $R^{\mathrm{d,sum}}$. The UL sum achievable rate $R^{\mathrm{u,sum}}$ [9], [32] is given by (50), shown at the top of the next page, where $\mathbf{w}_k$ is the $k^{\mathrm{th}}$ column of the UL receiver matrix $\mathbf{W}$ given in (27). The DL sum achievable rate $R^{\mathrm{d,sum}}$ [9], [10] is given by (51), shown at the top of the next page, where $\mathbf{b}_k$ is the $k^{\mathrm{th}}$ column of the DL precoding matrix $\mathbf{B}$ given in (34), and $\alpha$ is the power scaling performed at the UTs given in (35).

For the PR scheme, we consider a dynamic pilot length strategy. Specifically, for a given UT set, the achievable rates in (50) and (51) can be obtained for arbitrary pilot length $\tau(< K)$ with pilot scheduling performed by the SGPS algorithm, and then the optimal pilot length and the net spectral efficiency can be obtained. While for the OT scheme, the pilot length $\tau$ is set as $\tau = K$ if $K \leq T/2$, or $\tau = \lfloor T/2 \rfloor$ if $K > T/2$ where only $\lfloor T/2 \rfloor$ UTs are serviced simultaneously [6].

The UL net spectral efficiency performances of the PR scheme and the OT scheme are compared in Fig. 4 and Fig. 5, while the DL net spectral efficiency performances are compared in Fig. 6 and Fig. 7, with $K = 10$. It can be observed that, the proposed PR scheme shows performance gains over the conventional OT scheme in terms of the net spectral efficiency, and the gains become larger as the channel AS becomes smaller. Moreover, in the high SNR regime where the pilot interference dominates, and in the small coherence block length regime where the pilot overhead dominates, the proposed PR scheme provides significant performance gains. Specifically, for the case with $K = 10$, $\sigma = 2°$, $\rho = 20$ dB and $T = 20$, the proposed PR scheme provides approximately 35 bits/s/Hz net spectral efficiency gains over the conventional OT scheme for both the UL and DL data transmissions.

## VII. Conclusion

In this paper, we proposed pilot reuse (PR) in single cell for massive MIMO transmission to reduce the pilot overhead. We



Fig. 5. Comparison of the UL spectral efficiency performances between the PR scheme and the OT scheme. Results are shown versus the SNR for different values of AS $\sigma$ with $K = 10$ and $T = 20$.

exploited the fact that, in realistic outdoor wireless propagation environments where the BS is located at an elevated position, most of the channel power lies in a limited number of spatial directions compared with the whole massive MIMO channel dimension, and thus PR among UTs of spatial localization becomes feasible and beneficial. We first established the relationship between the channel covariance matrix and the channel PAS for the massive MIMO channels. Then we showed that, if channel AoA intervals of the UTs reusing the pilots are non-overlapping, then MSE of the channel estimation can be minimized. We also developed the robust multiuser UL receiver and DL precoder with the channel estimation error due to PR taken into account, and revealed the UL-DL MMSE duality between them. Moreover, we presented pilot scheduling under two MMSE related criteria, and proposed a low complexity pilot scheduling algorithm motivated by the channel AoA non-overlapping condition. The simulation

$$R^{\text{u,sum}} = \sum_{k=1}^{K} \mathbb{E} \left\{ \log_2 \left( 1 + \frac{\left| \mathbf{w}_k^T \hat{\mathbf{g}}_k \right|^2}{\mathbf{w}_k^T \left( \sum_{m \neq k} \hat{\mathbf{g}}_m \hat{\mathbf{g}}_m^H + \sum_{n=1}^{K} \mathbf{R}_{\tilde{\mathbf{g}}_n} + \frac{1}{\rho} \mathbf{I} \right) \mathbf{w}_k^* } \right) \right\} \tag{50}$$

$$R^{\text{d,sum}} = \sum_{k=1}^{K} \log_2 \left( 1 + \frac{\left| \mathbb{E} \left\{ \alpha \mathbf{g}_k^T \mathbf{b}_k \right\} \right|^2}{\sum_{m=1}^{K} \mathbb{E} \left\{ \alpha^2 \left| \mathbf{g}_k^T \mathbf{b}_m \right|^2 \right\} - \left| \mathbb{E} \left\{ \alpha \mathbf{g}_k^T \mathbf{b}_k \right\} \right|^2 + \frac{1}{\rho} \mathbb{E} \left\{ \alpha^2 \right\} } \right) \tag{51}$$



Fig. 6. Comparison of the DL spectral efficiency performances between the PR scheme and the OT scheme. Results are shown versus the coherence block length $T$ for different values of AS $\sigma$ with $K = 10$ and $\rho = 20$ dB.



Fig. 7. Comparison of the DL spectral efficiency performances between the PR scheme and the OT scheme. Results are shown versus the SNR for different values of AS $\sigma$ with $K = 10$ and $T = 20$.

results showed significant performance gains of the proposed PR scheme over the conventional OT scheme in terms of net spectral efficiency.

## APPENDIX A
## PROOF OF LEMMA 1

From the definition of $\mathbf{V}$ in (5), we have

$$\lim_{M \to \infty} \left[ \mathbf{V}^H \mathbf{V} - \mathbf{I}_M \right]_{i,j}$$

$$= \lim_{M \to \infty} \frac{1}{M} \left( \mathbf{v} \left( \vartheta \left( \psi_{i-1} \right) \right) \right)^H \mathbf{v} \left( \vartheta \left( \psi_{j-1} \right) \right) - \delta \left( i - j \right)$$

$$\overset{(a)}{=} \delta \left( i - j \right) - \delta \left( i - j \right) = 0 \tag{52}$$

where (a) follows from Assumption 1 and $\vartheta(\psi)$ is a strictly increasing function. This concludes the proof of (7).

The proof of (8) can be obtained as

$$\lim_{M \to \infty} \left[ \mathbf{R}_k - \mathbf{V} \text{diag} \left\{ \mathbf{r}_k \right\} \mathbf{V}^H \right]_{i,j}$$

$$\overset{(a)}{=} \lim_{M \to \infty} \left[ \mathbf{R}_k \right]_{i,j} -$$

$$\lim_{M \to \infty} \left[ \frac{1}{M} \sum_{m=1}^{M} \left[ \mathbf{r}_k \right]_m \mathbf{v} \left( \vartheta \left( \psi_{m-1} \right) \right) \mathbf{v}^H \left( \vartheta \left( \psi_{m-1} \right) \right) \right]_{i,j}$$

$$\overset{(b)}{=} \lim_{M \to \infty} \left[ \mathbf{R}_k \right]_{i,j} - \beta_k \lim_{M \to \infty} \sum_{m=1}^{M} \left[ \mathbf{v} \left( \vartheta \left( \psi_{m-1} \right) \right) \right]_i$$

$$\cdot \left[ \mathbf{v} \left( \vartheta \left( \psi_{m-1} \right) \right) \right]_j^* \mathsf{S}_k \left( \vartheta \left( \psi_{m-1} \right) \right) \left[ \vartheta \left( \psi_m \right) - \vartheta \left( \psi_{m-1} \right) \right]$$

$$\overset{(c)}{=} \beta_k \int_{\theta^{\min}}^{\theta^{\max}} \left[ \mathbf{v} \left( \theta \right) \right]_i \left[ \mathbf{v} \left( \theta \right) \right]_j^* \mathsf{S}_k \left( \theta \right) d\theta$$

$$- \beta_k \int_{\vartheta(\psi_0)}^{\vartheta(\psi_M)} \left[ \mathbf{v} \left( \vartheta \left( \psi \right) \right) \right]_i \left[ \mathbf{v} \left( \vartheta \left( \psi \right) \right) \right]_j^* \mathsf{S}_k \left( \vartheta \left( \psi \right) \right) d\vartheta \left( \psi \right)$$

$$\overset{(d)}{=} \beta_k \int_{\theta^{\min}}^{\theta^{\max}} \left[ \mathbf{v} \left( \theta \right) \right]_i \left[ \mathbf{v} \left( \theta \right) \right]_j^* \mathsf{S}_k \left( \theta \right) d\theta$$

$$- \beta_k \int_{\theta^{\min}}^{\theta^{\max}} \left[ \mathbf{v} \left( \theta \right) \right]_i \left[ \mathbf{v} \left( \theta \right) \right]_j^* \mathsf{S}_k \left( \theta \right) d\theta = 0 \tag{53}$$

where (a) follows from (5), (b) follows from (6), (c) follows from (2) and the integral definition, (d) follows from that $\vartheta \left( \psi_0 \right) = \vartheta \left( 0 \right) = \theta^{\min}$ and $\vartheta \left( \psi_M \right) = \vartheta \left( 1 \right) = \theta^{\max}$.

## APPENDIX B
## PROOF OF THEOREM 1

We start by presenting a lemma that is required in the following proof.

*Lemma 3:* For $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$, $\theta \left( \mathbf{A}, \mathbf{B} \right) = \pi / 2$ is equivalent to $\mathbf{AB} = \mathbf{0}$.

*Proof:* Recalling (18), $\theta(\mathbf{A}, \mathbf{B}) = \pi/2$ is equivalent to $\mathrm{tr}\{\mathbf{AB}\} = 0$. Furthermore, $\mathrm{tr}\{\mathbf{AB}\} = 0$ is equivalent to $\mathbf{AB} = \mathbf{0}$ for $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$ [33, Prop. 4.26]. This concludes the proof. ∎

Now we proceed with the proof of the theorem. Due to the positive semi-definiteness of the covariance matrix, we can obtain

$$\mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{g}}_k}\} = \mathrm{tr}\{\mathbf{R}_k - \mathbf{R}_k \mathbf{C}_{\pi_k}^{-1}\mathbf{R}_k\}$$

$$= \mathrm{tr}\left\{\mathbf{R}_k - \mathbf{R}_k \left(\sum_{\ell \in \mathcal{K}_{\pi_k}} \mathbf{R}_\ell + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1} \mathbf{R}_k\right\}$$

$$\geq \mathrm{tr}\left\{\mathbf{R}_k - \mathbf{R}_k \left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1} \mathbf{R}_k\right\}. \quad (54)$$

From Lemma 3 which states that $\theta(\mathbf{R}_i, \mathbf{R}_j) = \pi/2$ is equivalent to $\mathbf{R}_i \mathbf{R}_j = \mathbf{0}$, we can obtain

$$\mathbf{C}_{\pi_k}\mathbf{R}_k = \left(\sum_{\ell \in \mathcal{K}_{\pi_k}} \mathbf{R}_\ell + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)\mathbf{R}_k$$

$$= \left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)\mathbf{R}_k = \mathbf{R}_k\left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right) \quad (55)$$

which indicates that

$$\mathbf{C}_{\pi_k}^{-1}\mathbf{R}_k = \mathbf{R}_k\left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1} = \left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_k. \quad (56)$$

Substituting (56) into (16), we can obtain

$$\mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{g}}_k}\} = \mathrm{tr}\left\{\mathbf{R}_k - \mathbf{R}_k\left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_k\right\} \quad (57)$$

which achieves the minimum in (54). This concludes the proof. ∎

## APPENDIX C
## PROOF OF THEOREM 2

The MSE-SD defined in (25) can be simplified as

$$\epsilon^{\mathrm{u}} = \mathrm{tr}\left\{\mathbf{W}^T\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)\mathbf{W}^*\right.$$

$$\left. + \mathbf{I} - \mathbf{W}^T\hat{\mathbf{G}} - \hat{\mathbf{G}}^H\mathbf{W}^*\right\}. \quad (58)$$

Note that $\epsilon^{\mathrm{u}}$ is convex with respect to $\mathbf{W}$.

By setting the derivative of $\epsilon^{\mathrm{u}}$ with respect to $\mathbf{W}^*$ [34] to zero,

$$\frac{\partial}{\partial \mathbf{W}^*}\epsilon^{\mathrm{u}} = \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)^T\mathbf{W} - \hat{\mathbf{G}}^*$$

$$\overset{\text{(a)}}{=} \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)^*\mathbf{W} - \hat{\mathbf{G}}^* = \mathbf{0} \quad (59)$$

where (a) follows from $\mathbf{A}^T = \mathbf{A}^*$ for the Hermitian matrix

$\mathbf{A}$, we can obtain

$$\mathbf{W}^{\mathrm{opt}} = \left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right]^*. \quad (60)$$

Substituting (60) into (58), we can obtain the corresponding MSE-SD as

$$\epsilon^{\mathrm{u,min}} = \mathrm{tr}\left\{\mathbf{I} - \hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right\}$$

$$\overset{\text{(a)}}{=} \mathrm{tr}\left\{\left(\mathbf{I} + \hat{\mathbf{G}}^H\left(\sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{u}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right)^{-1}\right\} \quad (61)$$

where (a) follows from the Woodbury matrix inversion identity [33, Prop. 15.3]. This concludes the proof. ∎

## APPENDIX D
## PROOF OF THEOREM 3

We start by simplifying the MSE-SD defined in (32) as

$$\epsilon^{\mathrm{d}} = \mathbb{E}\left\{\left\|\alpha\left(\mathbf{G}^T\mathbf{B}\mathbf{a}^{\mathrm{d}} + \frac{1}{\sqrt{\rho^{\mathrm{d}}}}\mathbf{n}^{\mathrm{d}}\right) - \mathbf{a}^{\mathrm{d}}\right\|_2^2\right\}$$

$$= \mathbb{E}\left\{\|(\alpha\mathbf{G}^T\mathbf{B} - \mathbf{I})\mathbf{a}^{\mathrm{d}}\|_2^2\right\} + \frac{\alpha^2 K}{\rho^{\mathrm{d}}}$$

$$= \mathrm{tr}\left\{\alpha^2\mathbf{B}^H\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^*\right)\mathbf{B}\right.$$

$$\left. - \alpha\hat{\mathbf{G}}^T\mathbf{B} - \alpha\mathbf{B}^H\hat{\mathbf{G}}^*\right\} + \left(\frac{\alpha^2}{\rho^{\mathrm{d}}} + 1\right)K. \quad (62)$$

The simplified objective function in (62) is non-convex with respect to $(\mathbf{B}, \alpha)$. We first show that there exists a global minimum for the problem (33) in the following lemma.

*Lemma 4:* For the problem (33), there exists a global optimal solution.

*Proof:* The problem in (33) is equivalent to

$$\min_{\mathbf{B}} \quad \min_{\alpha(\mathbf{B})} \quad \epsilon^{\mathrm{d}}(\mathbf{B}, \alpha)$$
$$\text{subject to} \quad \mathrm{tr}\{\mathbf{B}\mathbf{B}^H\} \leq K \quad (63)$$

and the optimal $\alpha$ for the inner unconstrained optimization problem can be readily obtained as

$$\alpha = \frac{\mathrm{tr}\left\{\hat{\mathbf{G}}^T\mathbf{B} + \mathbf{B}^H\hat{\mathbf{G}}^*\right\}}{2\left[K/\rho^{\mathrm{d}} + \mathrm{tr}\left\{\mathbf{B}^H\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^*\right)\mathbf{B}\right\}\right]}. \quad (64)$$

Then the problem (63) is equivalent to

$$\min_{\mathbf{B}} \quad \epsilon^{\mathrm{d}}(\mathbf{B})$$

$$= K - \frac{\left[\mathrm{tr}\left\{\hat{\mathbf{G}}^T\mathbf{B} + \mathbf{B}^H\hat{\mathbf{G}}^*\right\}\right]^2}{4\left[K/\rho^{\mathrm{d}} + \mathrm{tr}\left\{\mathbf{B}^H\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^*\right)\mathbf{B}\right\}\right]}$$

$$\text{subject to} \quad \mathrm{tr}\{\mathbf{B}\mathbf{B}^H\} \leq K. \quad (65)$$

The feasible set of (65) given by $\{\mathbf{B} : \mathrm{tr}\{\mathbf{B}\mathbf{B}^H\} \leq K\}$ is compact (closed and bounded), and the objective function

of (65) is continuous over the feasible set. Thus, according to Weierstrass extreme value theorem [35, Appx. E], there exists a global minimum for the problem (65), and so does the equivalent problem (33). ∎

Lemma 4 shows that there exists a global optimum for the problem (33). Note that the global optimal solution should satisfy the Karush-Kuhn-Tucker (KKT) necessary conditions [36]. In the following we will seek out all the solutions that satisfy the KKT conditions and identify the optimal solution among them.

The Lagrangian associated with the problem (33) is

$$\mathcal{L}\left(\mathbf{B}, \alpha, \lambda\right) = \epsilon^{\mathrm{d}} + \lambda\left(\mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\} - K\right) \quad (66)$$

where $\epsilon^{\mathrm{d}}$ is given in (62), and $\lambda$ is the Lagrange multiplier associated with the inequality constraint.

The KKT necessary conditions for the problem (33) can be obtained as [34]

$$\frac{\partial}{\partial \mathbf{B}^*}\mathcal{L}\left(\mathbf{B}, \alpha, \lambda\right) = \alpha^2 \left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^*\right)\mathbf{B}$$
$$- \alpha\hat{\mathbf{G}}^* + \lambda\mathbf{B} = \mathbf{0} \quad (67)$$

$$\frac{\partial}{\partial \alpha}\mathcal{L}\left(\mathbf{B}, \alpha, \lambda\right) = 2\alpha\mathrm{tr}\left\{\mathbf{B}^H\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^*\right)\mathbf{B}\right\}$$
$$- \mathrm{tr}\left\{\hat{\mathbf{G}}^T\mathbf{B} + \mathbf{B}^H\hat{\mathbf{G}}^*\right\} + \frac{2\alpha K}{\rho^{\mathrm{d}}} = 0 \quad (68)$$

$$\lambda \geq 0, \quad \mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\} \leq K \quad (69)$$

$$\lambda\left(\mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\} - K\right) = 0. \quad (70)$$

An obvious solution that satisfies the above KKT conditions is $(\alpha = 0, \mathbf{B} = \mathbf{0}, \lambda = 0)$, and the corresponding $\epsilon^{\mathrm{d}}$ equals $K$. For the case with $\alpha \neq 0$, (67) is equivalent to

$$\hat{\mathbf{G}}^* = \alpha\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^* + \frac{\lambda}{\alpha^2}\mathbf{I}\right)\mathbf{B} \quad (71)$$

which leads to

$$\hat{\mathbf{G}}^T\mathbf{B} = \mathbf{B}^H\hat{\mathbf{G}}^*$$
$$= \alpha\mathbf{B}^H\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^* + \frac{\lambda}{\alpha^2}\mathbf{I}\right)\mathbf{B}. \quad (72)$$

Combining (72) with (68) yields

$$\frac{\alpha^2 K}{\rho^{\mathrm{d}}} = \lambda\mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\}. \quad (73)$$

Substituting (73) into (70), we can obtain

$$\lambda = \frac{\alpha^2}{\rho^{\mathrm{d}}} > 0, \quad \mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\} = K. \quad (74)$$

Substituting (74) into (71) yields

$$\mathbf{B} = \frac{1}{\alpha}\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^* + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}^*$$
$$= \frac{1}{\alpha}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right]^* \quad (75)$$

where $\alpha$ is chosen to satisfy the constraint $\mathrm{tr}\left\{\mathbf{B}\mathbf{B}^H\right\} = K$.

Substituting (75) into (62), we can obtain the corresponding MSE-SD as

$$\epsilon^{\mathrm{d}} = \mathrm{tr}\left\{\mathbf{I} - \hat{\mathbf{G}}^T\left(\hat{\mathbf{G}}^*\hat{\mathbf{G}}^T + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k}^* + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}^*\right\}$$
$$= \mathrm{tr}\left\{\mathbf{I} - \hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H + \sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right\} \quad (76a)$$

$$= \mathrm{tr}\left\{\left(\mathbf{I} + \hat{\mathbf{G}}^H\left(\sum_{k=1}^{K}\mathbf{R}_{\tilde{\mathbf{g}}_k} + \frac{1}{\rho^{\mathrm{d}}}\mathbf{I}\right)^{-1}\hat{\mathbf{G}}\right)^{-1}\right\} \quad (76b)$$

where (76a) follows from the trace identity $\mathrm{tr}\left\{\mathbf{A}\right\} = \mathrm{tr}\left\{\mathbf{A}^T\right\}$ [34, Eq. (2.95)] and $\mathbf{R}_{\tilde{\mathbf{g}}_k}$ is Hermitian, (76b) follows from the Woodbury matrix inversion identity [33, Prop. 15.3]. Note that the MSE-SD in (76a) is smaller than $K$ that previously obtained from the solution $(\alpha = 0, \mathbf{B} = \mathbf{0}, \lambda = 0)$. Therefore, we obtain that the precoder given by (75) is optimal. This concludes the proof. ∎

## APPENDIX E
### PROOF OF LEMMA 2

Via invoking the matrix-valued Jensen's inequality which states that $\mathbb{E}\left\{\mathbf{A}^{-1}\right\} \succeq \left(\mathbb{E}\left\{\mathbf{A}\right\}\right)^{-1}$ for $\mathbf{A} \succ \mathbf{0}$ [33, Prop. 21.64], we can obtain

$$\mathbb{E}\left\{\epsilon^{\mathrm{t,min}}\right\} \geq \mathrm{tr}\left\{\left(\mathbf{I}_K + \mathbb{E}\left\{\hat{\mathbf{G}}^H\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\hat{\mathbf{G}}\right\}\right)^{-1}\right\}$$
$$= \mathrm{tr}\left\{\left(\mathbf{I}_K + \mathbf{\Omega}\right)^{-1}\right\} \quad (77)$$

and $\mathbf{\Omega}$ satisfies that

$$[\mathbf{\Omega}]_{i,j} = \left[\mathbb{E}\left\{\hat{\mathbf{G}}^H\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\hat{\mathbf{G}}\right\}\right]_{i,j}$$
$$= \mathbb{E}\left\{\hat{\mathbf{g}}_i^H\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\hat{\mathbf{g}}_j\right\}$$
$$\overset{(a)}{=} \mathbb{E}\left\{\left(\mathbf{y}_{\pi_i}^{\mathrm{p}}\right)^H\mathbf{C}_{\pi_i}^{-1}\mathbf{R}_i\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\mathbf{R}_j\mathbf{C}_{\pi_j}^{-1}\mathbf{y}_{\pi_j}^{\mathrm{p}}\right\}$$
$$\overset{(b)}{=} \mathrm{tr}\left\{\mathbf{C}_{\pi_i}^{-1}\mathbf{R}_i\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\mathbf{R}_j\right\}\delta\left(\pi_i - \pi_j\right) \quad (78)$$

where (a) follows from (14), and (b) follows from (13). This concludes the proof. ∎

## APPENDIX F
### PROOF OF THEOREM 4

Via invoking the Schwartz inequality as in [37, Lemma 1], we can obtain

$$\epsilon^{\mathrm{t,alb}} = \mathrm{tr}\left\{\left(\mathbf{I}_K + \mathbf{\Omega}\right)^{-1}\right\} \geq \sum_{i=1}^{K}\frac{1}{1 + [\mathbf{\Omega}]_{i,i}} \quad (79)$$

where the equality is attained if and only if $\mathbf{\Omega}$ is diagonal.

Recalling Lemma 3 which states that $\mathbf{R}_i\mathbf{R}_j = \mathbf{0}$ is equivalent to $\theta\left(\mathbf{R}_i, \mathbf{R}_j\right) = \pi/2$, we only have to show that if $\mathbf{R}_i\mathbf{R}_j = \mathbf{0}$ for $\forall i \neq j$ and $\pi_i = \pi_j$, then $\mathbf{\Omega}$ is diagonal, i.e., $[\mathbf{\Omega}]_{i,j} = 0$ for $\forall i \neq j$ and $\pi_i = \pi_j$.

For $\forall i \neq j$ and $\pi_i = \pi_j$, if $\mathbf{R}_i\mathbf{R}_j = \mathbf{0}$, then

$$
\begin{aligned}
[\mathbf{\Omega}]_{i,j} &= \mathrm{tr}\left\{ \mathbf{C}_{\pi_i}^{-1}\mathbf{R}_i\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\mathbf{R}_j \right\} \\
&\stackrel{(a)}{=} \mathrm{tr}\left\{ \mathbf{R}_i\left(\mathbf{R}_i + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1}\mathbf{R}_j \right\} \\
&= \mathrm{tr}\left\{ \mathbf{R}_j\mathbf{R}_i\left(\mathbf{R}_i + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\left(\mathbf{R}^{\mathrm{t,n,eff}}\right)^{-1} \right\} = 0 \quad (80)
\end{aligned}
$$

where (a) follows from (56).

Furthermore, if $\mathbf{R}_i\mathbf{R}_j = \mathbf{0}$ for $\forall i \neq j$ and $\pi_i = \pi_j$, then diagonal elements of $\mathbf{\Omega}$ reduces to

$$
\begin{aligned}
[\mathbf{\Omega}]_{i,i} = \mathrm{tr}\bigg\{ &\left(\mathbf{R}_i + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_i \\
\cdot &\left[\sum_{k=1}^{K}\left(\mathbf{R}_k - \mathbf{R}_k\left(\mathbf{R}_k + \frac{1}{\rho^{\mathrm{p}}\tau}\mathbf{I}\right)^{-1}\mathbf{R}_k\right) + \frac{1}{\rho^{\mathrm{t}}}\mathbf{I}\right]^{-1}\mathbf{R}_i \bigg\}
\end{aligned}
$$
$$(81)$$

via invoking (56). This concludes the proof.

## REFERENCES

[1] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Massive MIMO transmission with pilot reuse in single cell," in *Proc. IEEE Int. Conf. Communications (ICC)*, Sydney, Australia, 2014, pp. 4794–4799.

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

[3] 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 Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[6] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Annu. Asilomar Conf. Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, 2006, pp. 359–363.

[7] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY: Cambridge University Press, 2005.

[8] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*, 2nd ed. Oxford, UK: Academic Press, 2013.

[9] 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.

[10] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD Systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[11] 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.

[12] 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.

[13] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 3249–3252.

[14] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, 2012, pp. 1137–1141.

[15] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, June 2013.

[16] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wide-band correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.

[17] G. Barriac and U. Madhow, "Characterizing outage rates for space-time communication over wideband channels," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2198–2208, Dec. 2004.

[18] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 437–447, Mar. 2000.

[19] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. Annu. Asilomar Conf. Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, 2012, pp. 295–299.

[20] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 469–477, Feb. 1995.

[21] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing–The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[22] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sept. 2014.

[23] T. L. Marzetta, G. H. Tucci, and S. H. Simon, "A random matrix-theoretic approach to handling singular covariance estimates," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6256–6271, Sept. 2011.

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

[25] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall, 2000.

[26] R. Vaughan and J. B. Andersen, *Channels, Propagation and Antennas for Mobile Communications*. Milton Keynes, UK: Institution of Electrical Engineers, 2003.

[27] A. Pascual-Iserte, D. P. Palomar, A. I. Pérez-Neira, and M. Á. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[28] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[29] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[30] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: Duality and sum-MSE minimization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5436–5446, Nov. 2007.

[31] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Singapore: John Wiley & Sons (Asia) Pte Ltd, 2010.

[32] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[33] G. A. F. Seber, *A Matrix Handbook for Statisticians*. Hoboken, NJ: John Wiley & Sons, Inc., 2008.

[34] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. New York, NY: Cambridge University Press, 2011.

[35] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY: Cambridge University Press, 2012.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[37] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2138–2145, Sept. 2004.