Kent Academic Repository

Full text document (pdf)

Citation for published version

Zhang, Yu and Wang, Dongming and Wang, Jiangzhou and You, Xiaohu (2016) Channel Estimation for Massive MIMO-OFDM Systems by Tracking the Joint Angle-Delay Subspace. IEEE Access . (In press)

DOI

http://doi.org/10.1109/ACCESS.2016.2634025

Link to record in KAR

http://kar.kent.ac.uk/59632/

Document Version

Author's Accepted Manuscript

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

Channel Estimation for Massive MIMO-OFDM Systems by Tracking the Joint Angle-Delay Subspace

Yu Zhang, Dongming Wang, *Member, IEEE,* Jiangzhou Wang, *Senior Member, IEEE,* and Xiaohu You, *Fellow, IEEE*

***Abstract*—In this paper, we propose joint angle-delay subspace based channel estimation in single cell for broadband massive multiple-input and multiple-output (MIMO) systems employing orthogonal frequency division multiplexing (OFDM) modulation. Based on a parametric channel model, we present a new concept of the joint angle-delay subspace which can be tracked by the low-complexity low-rank adaptive filtering (LORAF) algorithm. Then, we investigate an interference-free transmission condition that the joint angle-delay subspaces of the users reusing the same pilots are non-overlapping. Since the channel statistics are usually unknown, we develop a robust minimum mean square error (MMSE) estimator under the worst precondition of pilot decontamination, considering that the joint angle-delay subspaces of the interfering users fully overlap. Furthermore, motivated by the interference-free transmission criteria, we present a novel low-complexity greedy pilot scheduling algorithm to avoid the problem of initial value sensitivity. Simulation results show that the joint angle-delay subspace can be estimated effectively, and the proposed pilot reuse scheme combined with robust MMSE channel estimation offers significant performance gains.**

***Index Terms*—Joint angle-delay subspace, massive MIMO-OFDM, pilot decontamination, pilot scheduling, robust channel estimation.**

I. Introduction

Massive multiple-input and multiple-output (MIMO) systems [1], [2], where each base station (BS) employs a very large antenna array to simultaneously serve many users in the same time-frequency resource, are currently considered as a promising future cellular network technology. Such systems can provide huge improvements in spectral efficiency, energy efficiency and simplification of the multiple access layer [3], which makes it one of key candidate technologies for the next generation wireless communication [4]–[6]. Orthogonal frequency division multiplexing (OFDM) is an enabler for achieving high data rate transmission over mobile radio channels [7], [8]. Owing to its prominent capability to combat frequency-selective fading and gain high spectral efficiency, OFDM with massive MIMO (also called massive MIMO-OFDM) is an attractive scheme of achieving broadband massive MIMO transmission. However, the effectiveness of massive MIMO-OFDM is severely affected by the accuracy of channel estimation [9], [10].

For conventional MIMO-OFDM systems, pilot-aided channel estimation is generally adopted, where nonparametric approach can be realised by exploiting the time and frequency correlation of the channel frequency response [11]. Without any constraints about the channel, the dimension of the estimation problem can be quite large. However, the radio channel in wireless communication is often characterized by a few dominant paths, typically two to six [12]. Moreover, the high-speed data transmission potentially results in a sparse multipath channel. When the channel correlation matrix is constructed based on the parametric channel model, the signal subspace dimension of the correlation matrix can be effectively reduced [12]. Although the orthogonal pilot approaches [11]–[13] can eliminate pilot interference, the pilot overhead issue has not been taken into account, which is one of important issues in massive MIMO-OFDM.

For the traditional orthogonal training scheme [4], the overhead related to channel training is proportional to the number of users per cell, and is independent of the number of antennas per BS when exploiting channel reciprocity [14], [15] in time division duplex (TDD) mode [16]. When increasing the number of users served by the BS, orthogonal pilots for multiple users should be reused due to a limited number of orthogonal pilots. However, this pilot reuse severely degrades the channel estimation performance due to pilot contamination (i.e. co-pilot interference). Thus, pilot contamination causes a major bottleneck in achieving the promising performance gain of massive MIMO-OFDM [17].

Recently, significant research works have been presented to tackle the pilot contamination problem. For a multi-cell scenario, a Bayesian estimator was adopted to suppress the interference from users when the channel covariance matrices satisfy a certain non-overlapping condition on their dominant subspace [18]. Since channels of different users tend to be pairwise orthogonal in a large-scale antenna system, blind discrimination is doable between the desired signals and interferences based on the prior known average channel gains [19]. [20] proposed a novel robust channel estimation algorithm by exploiting path diversity in both angle and power domains.

This work was supported in part by the National Basic Research Program of China (973 Program 2013CB336600); the Natural Science Foundation of China (NSFC) under grants 61271205, 61501113, 61521061, and 61372100; National High Technology Research and Development Program of China (863 Program 2014AA01A704); the Natural Science Foundation of Jiangsu Province under grant BK20150630; the Hong Kong, Macao and Taiwan Science & Technology Cooperation Program of China (2014DFT10290).

Y. Zhang, D. Wang and X. You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: yuzhang@seu.edu.cn; wangdm@seu.edu.cn; xhyu@seu.edu.cn).

J. Wang is with the School of Engineering and Digital Arts, University of Kent, Canterbury CT2 7NT, U.K. (e-mail: j.z.wang@kent.ac.uk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

In contrast to previous works that considered the channel feature in space or power domain, the inherent sparse nature of channel pulse response in time domain was exploited for broadband wireless transmission [21]. It has been concluded that the power delay profile (PDP) of the desired channel can be extracted from the contaminated channel estimate in time domain with an artificial randomized pilot assignment method. Furthermore, [22] considered channel sparse patterns in angle-delay domain, and showed that the best channel acquisition performance in massive MIMO-OFDM systems can be obtained when the user channel power distributions in angle-delay domain are made non-overlapping with the proposed adjustable phase shift pilots in a single-cell scenario.

In this paper, we consider a parametric channel model where the channel frequency response is estimated using an P-path channel model, and propose a joint angle-delay subspace based channel estimation scheme for massive MIMO-OFDM systems. In [22], the angle-delay domain channel power matrix was estimated, which contained not only multipath angles and delays but also powers. However, this paper only requires the joint angle-delay subspace. Subspace based high-resolution methods play a significant role in sensor array processing and spectral analysis. Several early works have successfully applied them to channel estimation in wireless communication systems [23] [24]. Low-complexity subspace tracking algorithms [25] [26] have been proposed to track the signal subspace recursively, due to the computation burden of eigen value decomposition (EVD) and unsuitability for adaptive processing of conventional approaches, such as multiple signal classification (MUSIC) method [27] and estimation of signal parameters via rotational invariance techniques (ESPRIT) estimator [28].

The main contributions of this paper are three-fold. First of all, based on a sparse physical channel model, we extend the concept of delay subspace presented in [29] to joint angle-delay subspace for massive MIMO-OFDM systems, and establish a relationship between the space-frequency domain channel correlation matrix (SFCCM) and joint angle-delay subspace matrix (JADSM). Then, we utilize a subspace tracking algorithm called low-rank adaptive filter (LORAF) to adaptively estimate the number of prominent paths and JADSMs.

The second contribution of this paper is the development of a robust minimum mean square error (MMSE) channel estimator under the worst precondition of pilot decontamination in a single-cell multi-user scenario. Based on the sparse channel structure, the SFCCM can be constructed by the JADSM and PDP. However, compared with the JADSM acquisition, the power of each path is relatively difficult to track. We show that the proposed robust channel estimator is insensitive to the channel statistics and can significantly improve the performance of massive MIMO-OFDM systems.

The third contribution of this paper is the presentation of a modified greedy pilot reuse (PR) scheme exploiting the joint angle-delay subspace. We prove that the mean square error of channel estimation (MSE-CE) can be minimized provided that the joint angle-delay subspace for different users reusing the same pilots can be made non-overlapping with proper pilot scheduling. Simulation results show that the proposed PR method provides significant performance gains over random scheduling approach and conventional greedy scheduling approach.

The rest of this paper is organized as follows. In Section II, the massive MIMO-OFDM channel model is described. In Section III, the concept of the joint angle-delay subspace and its adaptive estimation are presented. In Section IV, the PR for uplink training is analysed. In Section V, a modified greed pilot scheduling algorithm is proposed, which is based on the joint angle-delay subspace. Simulation results are given in Section VI. Finally, Section VII draws the conclusions.

The notations adopted in this paper are as follows. We use upper (lower) boldface to denote matrices (column vectors). Specifically, we adopt $\mathbf{I}_N$ to denote the $N \times N$ dimensional identity matrix, and $\mathbf{0}_N$ to denote the $N \times N$ dimensional all-zero matrix. The notation $\triangleq$ is used for definitions. The superscripts $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^*$ stand for the conjugate transpose, transpose and conjugate respectively. $\mathbb{E}\{\cdot\}$ denotes the expectation, $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, and $\mathrm{diag}\{\mathbf{x}\}$ denotes the diagonal matrix with $\mathbf{x}$ along its main diagonal. The symbols $\otimes$ and $\circledast$ denote the Kronecker product and Khatri-Rao product of two matrices, respectively. The operator $\mathrm{vec}\{\cdot\}$ stacks the columns of a matrix into a tall vector, and $\mathrm{tr}\{\cdot\}$ stands for the matrix trace operation. $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes the circular symmetric complex Gaussian distribution with mean μ and variance σ^2. $\delta(\cdot)$ denotes the Dirac delta function. $\mathbb{C}^{M\times N}$ denote the $M \times N$ dimensional complex vector space. $\mathbb{N}$ and $\mathbb{N}^+$ denote the set of integers and positive integers, respectively. The notation $\backslash$ denotes the set subtraction operation.

II. MASSIVE MIMO-OFDM CHANNEL MODEL

Consider a single-cell TDD broadband massive MIMO wireless system consisting of one BS equipped with M antennas and K single-antenna users. The user set is denoted as $\mathcal{K} = \{1, 2, \cdots, k, \cdots, K\}$ where $k \in \mathcal{K}$ is the user index. It is assumed that the BS is equipped with one-dimensional uniform linear array (ULA) and its antennas are separated by one-half wavelength. Then the array steering vector corresponding to the arrival angle θ with respect to the perpendicular to the array is given by

$$\mathbf{v}(\theta) = \begin{bmatrix} 1 \\ \exp\left(-j\pi \sin(\theta)\right) \\ \vdots \\ \exp\left(-j\pi (M-1) \sin(\theta)\right) \end{bmatrix} \in \mathbb{C}^{M\times 1}. \quad (1)$$

Consider OFDM modulation with N_c subcarriers, performed via the N_c points inverse discrete Fourier transform (DFT) operation, appended with a guard interval (also called cyclic prefix) of length $N_{\mathrm{g}}(\leq N_{\mathrm{c}})$ samples. $T_{\mathrm{sym}} = (N_{\mathrm{c}} + N_{\mathrm{g}}) T_{\mathrm{s}}$ and $T_{\mathrm{c}} = N_{\mathrm{c}} T_{\mathrm{s}}$ denote the OFDM symbol duration with and without the guard interval respectively, where T_{s} is the system sampling interval.

It is assumed that the channels remain constant during one OFDM symbol, and evolve from symbol to symbol. Suppose that there exist P physical paths between the BS and any user,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

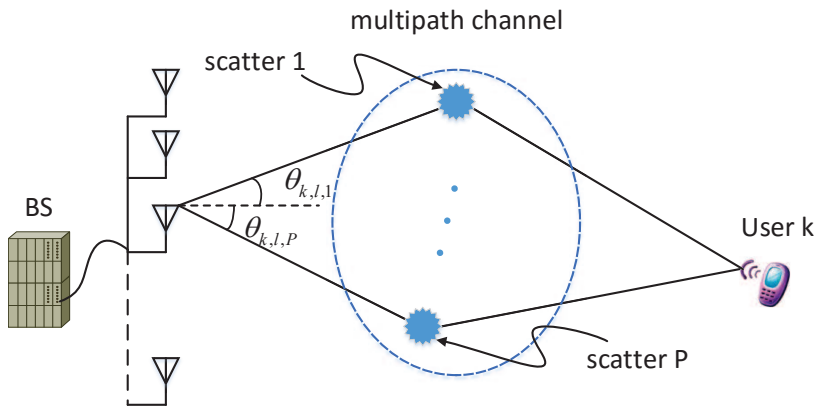


Fig. 1. Uplink multipath channel from user k with single antenna to its BS with M antennas.

and the pth path of user k over the lth OFDM symbol has a complex-valued attenuation of $\alpha_{k,l,p}$, an arrival angle of $\theta_{k,l,p}$ and a propagation delay of $\tau_{k,l,p}$. Using a physical channel model approach, the uplink channel frequency response vector between user k and the BS over the OFDM symbol l and the subcarrier n is given by

$$\mathbf{h}_{k,l,n} = \sum_{p=1}^{P} \alpha_{k,l,p} \mathbf{v}\left(\theta_{k,l,p}\right) \exp\left(-j2\pi \frac{n}{T_{\mathrm{c}}} \tau_{k,l,p}\right) \tag{2}$$

where $\mathbf{v}\left(\theta\right)$ is given by (1), illustrated in Fig.1.

The channel matrix of user k at the OFDM symbol l over all subcarriers is written as

$$\begin{aligned}\mathbf{H}_{k,l} &= \left[\mathbf{h}_{k,l,0}\mathbf{h}_{k,l,1}\cdots\mathbf{h}_{k,l,N_{\mathrm{c}}-1}\right] \\ &= \left[\mathbf{v}\left(\theta_{k,l,1}\right)\cdots\mathbf{v}\left(\theta_{k,l,P}\right)\right]\begin{bmatrix}\alpha_{k,l,1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \alpha_{k,l,P}\end{bmatrix}\begin{bmatrix}\mathbf{w}(\tau_{k,l,1})^T \\ \vdots \\ \mathbf{w}(\tau_{k,l,P})^T\end{bmatrix} \\ &\triangleq \mathbf{V}\left(\boldsymbol{\theta}_{k,l}\right)\mathrm{diag}\left\{\boldsymbol{\alpha}_{k,l}\right\}\mathbf{W}(\boldsymbol{\tau}_{k,l})^T\end{aligned} \tag{3}$$

which is referred as the space-frequency domain channel response matrix (SFCRM), where $\boldsymbol{\alpha}_{k,l}=\left[\alpha_{k,l,1}\ \cdots\ \alpha_{k,l,P}\right]^T$, $\boldsymbol{\theta}_{k,l} = \left[\theta_{k,l,1}\ \cdots\ \theta_{k,l,P}\right]^T$ and $\boldsymbol{\tau}_{k,l} = \left[\tau_{k,l,1}\ \cdots\ \tau_{k,l,P}\right]^T$ denote multipath amplitudes, angle of arrivals (AOAs) and delays, respectively. The delay response vector is denoted as

$$\mathbf{w}\left(\tau\right) = \begin{bmatrix}1 \\ \exp\left(-j\frac{2\pi}{N_{\mathrm{c}}}\frac{\tau}{T_{\mathrm{s}}}1\right) \\ \vdots \\ \exp\left(-j\frac{2\pi}{N_{\mathrm{c}}}\frac{\tau}{T_{\mathrm{s}}}\left(N_{\mathrm{c}}-1\right)\right)\end{bmatrix} \in \mathbb{C}^{N_{\mathrm{c}}\times 1}. \tag{4}$$

Define the vector $\mathbf{u}\left(\theta,\tau\right)$ as the space-time vector for a single path of unit amplitude arriving at the angle θ with a delay τ, i.e.

$$\mathbf{u}\left(\theta,\tau\right) \triangleq \mathbf{w}\left(\tau\right) \otimes \mathbf{v}\left(\theta\right). \tag{5}$$

According to (3) and (5), we have

$$\begin{aligned}\mathbf{h}_{k,l} &= \mathrm{vec}\left\{\mathbf{H}_{k,l}\right\} \\ &= \left[\mathbf{u}\left(\theta_{k,l,1},\tau_{k,l,1}\right)\cdots\mathbf{u}\left(\theta_{k,l,P},\tau_{k,l,P}\right)\right]\boldsymbol{\alpha}_{k,l} \\ &= \mathbf{U}\left(\boldsymbol{\theta}_{k,l},\boldsymbol{\tau}_{k,l}\right)\boldsymbol{\alpha}_{k,l} \\ &= \left[\mathbf{W}\left(\boldsymbol{\tau}_{k,l}\right) \circledast \mathbf{V}\left(\boldsymbol{\theta}_{k,l}\right)\right]\boldsymbol{\alpha}_{k,l}\end{aligned} \tag{6}$$

where $\mathbf{U}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)$ is called the space-time manifold matrix (STMM) which is parameterized by the set of AOAs and path delays [30].

As explained in Section I, the multipath channel exhibits a sparse nature in massive MIMO-OFDM systems. With tolerable energy leakage of adjacent channel taps, it is assumed that normalized delays $\tau_{k,l,p}/T_{\mathrm{s}}$ are integers as in [31]. Relative to the rapid variation of the amplitudes $\boldsymbol{\alpha}_{k,l}$, the angles $\boldsymbol{\theta}_{k,l}$ and delays $\boldsymbol{\tau}_{k,l}$ should vary slowly. In fact, the angles and delays can be considered constant in L OFDM symbols provided that their variations in LT_{sym} are much smaller than the temporal resolution of the system [29]. For convenience, it is assumed $\mathbf{U}\left(\boldsymbol{\theta}_{k,l},\boldsymbol{\tau}_{k,l}\right) \approx \mathbf{U}\left(\boldsymbol{\theta}_{k},\boldsymbol{\tau}_{k}\right)$.

Due to the motion of users, it is assumed that the amplitudes $\alpha_{k,l,p}$ are wide-sense stationary (WSS) narrow-band complex Gaussian processes with the so-called Jake's power spectrum , and different path gains are uncorrelated with each other [32]. Hence, the autocorrelation function of $\alpha_{k,l,p}$ is given by

$$\mathbb{E}\left\{\alpha_{k,l,p}\alpha^*_{k,l+\Delta_l,p}\right\} = J_0\left(2\pi\upsilon_k\Delta_l T_{\mathrm{sym}}\right)\sigma^2_{k,p} \tag{7}$$

where $J_0\left(\cdot\right)$ is the zeroth order Bessel function of the first kind, υ_k is the maximum Doppler frequency spread of user k and $\sigma^2_{k,p}$ is the average power of the pth path of user k. For simple analysis, it is assumed that all the users are of equal distance from the BS and uniformly distributed in a circle, leading to $\sum_{p=1}^{P}\sigma^2_{k,p} = \sigma^2_{\mathrm{h}}$, and $\sigma^2_{k,1} \geq \sigma^2_{k,2} \geq \cdots \geq \sigma^2_{k,P}$, for $\forall k \in \mathcal{K}$.

III. JOINT ANGLE-DELAY SUBSPACE ESTIMATION FOR SINGLE USER

Based on the sparse massive MIMO-OFDM channel model presented in the previous section, we propose to track the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

joint angle-delay subspace by a subspace tracking algorithm. In this section, we first investigate joint angle-delay subspace estimation for single user, while its important role in PR will be investigated in the next section.

A. Concept of Joint Angle-Delay Subspace

Since the channel corresponding to each user has the same statistics, we temporarily drop the subscript k from $\mathbf{H}_{k,l}$ in this section. The techniques presented here employ a comb pilot pattern. The number of pilot subcarriers is denoted as N_0. For simplicity, we set $N_{\mathrm{c}} = N_0 \times \triangle N$, where $\triangle N$ is the number of subcarriers contained by the coherent bandwidth. During the uplink training phase, namely, the lth OFDM symbol, the received pilot signals at the BS can be written as

$$\bar{\mathbf{Y}}_l = \bar{\mathbf{H}}_l \bar{\mathbf{X}} + \bar{\mathbf{Z}}_l \in \mathbb{C}^{M \times N_0} \tag{8}$$

where $\bar{\mathbf{X}} = \mathrm{diag}\{\mathbf{x}\} \in \mathbb{C}^{N_0 \times N_0}$ denotes the frequency pilot signal from the user and $\bar{\mathbf{Z}}_l$ is the additive white Gaussian noise (AWGN) matrix with elements identically and independently distributed (i.i.d.) as $\mathcal{CN}(0, \sigma_{\mathrm{z}}^2)$. More specifically, $\bar{\mathbf{X}}$ satisfies $\bar{\mathbf{X}}\bar{\mathbf{X}}^H = \sigma_{\mathrm{x}}^2 \mathbf{I}_{N_0}$, where σ_{x}^2 is the pilot signal transmit power,

After decorrelation and power normalization of $\bar{\mathbf{Y}}_l$, the BS can obtain an observation of the uplink channel $\bar{\mathbf{H}}_l$, given by

$$\bar{\mathbf{Y}}_l^{\mathrm{dec}} = \bar{\mathbf{H}}_l + \frac{1}{\sigma_{\mathrm{x}}^2} \bar{\mathbf{Z}}_l \bar{\mathbf{X}}^H. \tag{9}$$

Using the unitary transformation property, the second term in (9) exhibits a Gaussian distribution with i.i.d. elements distributed as $\mathcal{CN}(0, \sigma_{\mathrm{z}}^2/\sigma_{\mathrm{x}}^2)$, and (9) can be simplified as

$$\bar{\mathbf{Y}}_l^{\mathrm{dec}} = \bar{\mathbf{H}}_l + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{Z}}_{\mathrm{iid}} \tag{10}$$

where $\rho_{\mathrm{tr}} \triangleq \sigma_{\mathrm{x}}^2/\sigma_{\mathrm{z}}^2$ is the transmit signal-to-noise ratio (SNR) during the pilot segment, and $\bar{\mathbf{Z}}_{\mathrm{iid}} \in \mathbb{C}^{M \times N_0}$ is the normalized AWGN matrix with i.i.d. elements distributed as $\mathcal{CN}(0,1)$.

Using the vectorization operation, (10) can be represented as

$$\begin{aligned} \bar{\mathbf{y}}_l^{\mathrm{dec}} &= \mathrm{vec}\left\{\bar{\mathbf{Y}}_l^{\mathrm{dec}}\right\} \\ &= \bar{\mathbf{h}}_l + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}} \\ &= \bar{\mathbf{U}}(\boldsymbol{\theta}, \boldsymbol{\tau}) \bar{\boldsymbol{\alpha}}_l + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}} \end{aligned} \tag{11}$$

where $\bar{\mathbf{U}}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \bar{\mathbf{W}}(\boldsymbol{\tau}) \circledast \mathbf{V}(\boldsymbol{\theta})$ and $\bar{\mathbf{z}}_{\mathrm{iid}} \in \mathbb{C}^{MN_0 \times 1}$ is the normalized AWGN vector with i.i.d elements distributed as $\mathcal{CN}(0,1)$. The sample delay matrix of $\bar{\mathbf{W}}(\boldsymbol{\tau})$ is a equi-spaced extracted Fourier matrix, represented as

$$\bar{\mathbf{W}}(\mathbf{q}) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\frac{2\pi}{N_{\mathrm{c}}} q_1 \triangle N} & \cdots & \mathrm{e}^{-j\frac{2\pi}{N_{\mathrm{c}}} q_P \triangle N} \\ \vdots & \vdots & \vdots \\ e^{-j\frac{2\pi}{N_{\mathrm{c}}} q_1 (N_{\mathrm{c}} - \triangle N)} & \cdots & \mathrm{e}^{-j\frac{2\pi}{N_{\mathrm{c}}} q_P (N_{\mathrm{c}} - \triangle N)} \end{bmatrix} \in \mathbb{C}^{N_0 \times \mathrm{P}} \tag{12}$$

where $\mathbf{q} = \left[q_1, q_2, \cdots, q_P\right]^T \in \mathbb{N}^{P \times 1}$.

A specific property of the massive antenna array is its high resolution to the channels in angle domain, and we introduce an lemma [33] about it in the following.

Lemma 1: Array response vectors corresponding to distinct angles are asymptotically orthogonal when the number of BS antennas tends to infinity, i.e., for $\forall \varphi, \eta \in [-\pi, \pi]$,

$$\lim_{M \to \infty} \frac{1}{M} \mathbf{v}^H(\varphi) \mathbf{v}(\eta) = \delta(\varphi - \eta). \tag{13}$$

Note that Lemma 1 is valid for uniform linear array. For each user, we can obtain the following result on massive MIMO-OFDM channels.

Proposition 1: For large M, the sample space-time manifold matrix (SSTMM) $\bar{\mathbf{U}}(\boldsymbol{\theta}, \mathbf{q})$ satisfies

$$\frac{1}{N_{\mathrm{u}}} \bar{\mathbf{U}}^H(\boldsymbol{\theta}, \mathbf{q}) \bar{\mathbf{U}}(\boldsymbol{\theta}, \mathbf{q}) = \mathbf{I}_P \tag{14}$$

where $N_{\mathrm{u}} = M \times N_0$ is the number of rows in $\bar{\mathbf{U}}(\boldsymbol{\theta}, \mathbf{q})$.

Proof: See Appendix A. ■

Proposition 1 demonstrates that for massive MIMO-OFDM channels, different column vectors of the SSTMM are approximately mutually orthogonal. Moreover, each column vector of the SSTMM corresponds to a different physical path composed of a specific AOA and delay, which can be resolved in the massive MIMO-OFDM with a sufficiently large antenna array aperture. Then, we refer to $\mathbf{Q} = \frac{1}{\sqrt{N_{\mathrm{u}}}} \bar{\mathbf{U}}(\boldsymbol{\theta}, \mathbf{q})$ as the JADSM, whose column vectors contain a set of orthonormal basis, which spans the P-dimensional joint angle-delay subspace. Therefore, the JADSM contains all the information on multi-path delays and AOAs of the user.

B. Adaptive Estimation of JADSM

In this subsection we investigate the estimation of the JADSM $\mathbf{Q}$. The correlation matrix of the observation vector in (11) is given by

$$\begin{aligned} \mathbf{R}_{\bar{\mathbf{y}}} &= \mathbb{E}\left\{\bar{\mathbf{y}}_l^{\mathrm{dec}} \left(\bar{\mathbf{y}}_l^{\mathrm{dec}}\right)^H\right\} \\ &= \mathbb{E}\left\{\left(\sqrt{N_{\mathrm{u}}} \mathbf{Q} \bar{\boldsymbol{\alpha}}_l + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}}\right)\left(\sqrt{N_{\mathrm{u}}} \mathbf{Q} \bar{\boldsymbol{\alpha}}_l + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}}\right)^H\right\} \\ &= N_{\mathrm{u}} \mathbf{Q} \mathbb{E}\left\{\bar{\boldsymbol{\alpha}}_l \bar{\boldsymbol{\alpha}}_l^H\right\} \mathbf{Q}^H + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}} \\ &= \underbrace{\mathbf{Q}\, \mathrm{diag}\left\{N_{\mathrm{u}}\sigma_1^2 \quad \cdots \quad N_{\mathrm{u}}\sigma_P^2\right\} \mathbf{Q}^H}_{\triangleq \mathbf{R}_{\bar{\mathbf{h}}}} + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}} \end{aligned} \tag{15}$$

where the JADSM $\mathbf{Q}$ is the truncated eigen matrix (i.e. signal subspace) of the space-frequency domain channel correlation matrix (SFCCM) $\mathbf{R}_{\bar{\mathbf{h}}}$ of the channel $\bar{\mathbf{h}}_l$, since $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_P$ according to Proposition 1. Each eigen value of $\mathbf{R}_{\bar{\mathbf{h}}}$ is proportional to the average power of the corresponding path, which indicates that the SFCCM of any user can be determined by its JADSM and uncontaminated PDP. It is worth noting that this explanation is similar to that in [22], except that our SFCCM only contains the prominent paths with almost the whole channel power, as in [34].

The subspace basis, i.e., the JADSM $\mathbf{Q}$ in this application, can be generally estimated as the span of $\bar{r}$ largest eigenvectors of sample correlation matrix $\widehat{\mathbf{R}}_{\bar{\mathbf{h}}} = \sum_{l=1}^{L} \bar{\mathbf{y}}_l^{\mathrm{dec}} \left(\bar{\mathbf{y}}_l^{\mathrm{dec}}\right)^H$. The rank $\bar{r}$ can be estimated from the obtained sample correlation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

matrix by using the traditional minimum description length (MDL) criterion. However, this approach has no capability to effectively separate the signal from the superimposed noise and leads to the high-complexity eigenvalue decomposition (EVD).

Fortunately, a fast subspace tracking algorithm, known as LORAF, was proposed by suitably projecting the observed data into the signal subspace instead of the complete data space, which has already been applied to the delay subspace tracking for OFDM systems [29]. Next, this algorithm is modified for the application discussed here.

This is essentially a low-rank recursive least squares (RLS) adaptive filtering process. It is assumed that the orthogonal basis of the joint angle-delay subspace is made up of $\bar{r}$ maximal eigenvectors of the correlation matrix $\mathbf{\Phi}_l = \sum_{i=1}^{l} \mu^{l-i} \bar{\mathbf{y}}_i^{\mathrm{dec}} \left(\bar{\mathbf{y}}_i^{\mathrm{dec}}\right)^H$, where μ is the forgetting factor. In order to track dominant eigenvectors and eigenvalues, simultaneous orthogonal iteration [35] based on a skinny QR decomposition is adopted as

$$\mathbf{A}_l = \mathbf{\Phi}_l \mathbf{B}_{l-1} \tag{16}$$

$$\mathbf{A}_l = \mathbf{B}_l \mathbf{R}_l \ \ (\text{QR decomposition}). \tag{17}$$

Due to its slowly time-varying characteristic, the correlation matrix $\mathbf{\Phi}_l$ is continuously updated according to

$$\mathbf{\Phi}_l = \mu \mathbf{\Phi}_{l-1} + (1-\mu)\, \bar{\mathbf{y}}_l^{\mathrm{dec}} \left(\bar{\mathbf{y}}_l^{\mathrm{dec}}\right)^H. \tag{18}$$

Substituting (18) into (16), we obtain the updated

$$\mathbf{A}_l = \mu \mathbf{A}_{l-1} \mathbf{\Theta}_{l-1} + (1-\mu)\, \bar{\mathbf{y}}_l^{\mathrm{dec}} \mathbf{b}_l^H, \tag{19}$$

where the matrix $\mathbf{\Theta}_l = \mathbf{B}_{l-1}^H \mathbf{B}_l$ is the cosines of angles between subsequent subspaces, and $\mathbf{b}_l = \mathbf{B}_{l-1}^H \bar{\mathbf{y}}_l^{\mathrm{dec}}$. Furthermore, the subspace trackers are operated with a predetermined rank $\bar{r}_{\max}$, which should be sufficiently large in any case. In each iteration, the estimation of the data power p_n, which is the sum of the noise and signal power, is compatible with the forgetting rule used in the subspace trackers. Subsequently, the estimated eigenvalues are compared with an estimated noise floor level $\widehat{\sigma}^2$ multiplied by a certain factor β (a thorough analysis of the optimal multiplicative coefficient β can be found in [36]), so as to determine the corresponding eigenvectors used for signal reconstruction. This subspace tracking algorithm [26] is summarized in Algorithm 1, where the notation $\mathrm{card}\{\cdot\}$ denotes the cardinality of the set under parentheses and the operator $[\cdot]_{\mathrm{col}(1:k)}$ selects the first k columns of its argument.

Since the angles and delays are stationary over long time intervals, unlike the instantaneous channel fading [30], there will be enough resources to obtain an estimate of the JADSM with guaranteed accuracy in practice. Therefore, it is assumed that the JADSMs of all the users are known by the BS in the rest of the paper.

Algorithm 1 Subspace Tracking Algorithm (LORAF)

Initialize:

$0 \le \mu \le 1$; $p_0 = 0$; β; $\bar{r}_{\max}$;

$\mathbf{\Theta}_0 = \mathbf{I}_{\bar{r}_{\max}}$; $\mathbf{A}_0 = \mathbf{0}$; $\mathbf{B}_0 = \begin{bmatrix} \mathbf{I}_{\bar{r}_{\max}} \\ \mathbf{0} \end{bmatrix}_{N_{\mathrm{u}} \times \bar{r}_{\max}}$.

Iteration:

for symbol $l = 1 : L$ **do**

1. Subspace tracking:

$\mathbf{b}_l = \mathbf{B}_{l-1}^H \bar{\mathbf{y}}_l^{\mathrm{dec}}$

$\mathbf{A}_l = \mu \mathbf{A}_{l-1} \mathbf{\Theta}_{l-1} + (1-\mu)\, \bar{\mathbf{y}}_l^{\mathrm{dec}} \mathbf{b}_l^H$

$\mathbf{A}_l = \mathbf{B}_l \mathbf{R}_l$ (skinny QR decomposition)

$\mathbf{\Theta}_l = \mathbf{B}_{l-1}^H \mathbf{B}_l$

2. Adaptive rank estimation:

$\widehat{\lambda}_i = [\mathbf{R}_l]_{ii} \quad i = 1, 2, \cdots, \bar{r}_{\max}$

$p_n = \mu p_{n-1} + \frac{(1-\mu)}{N_{\mathrm{u}}} tr\left\{\bar{\mathbf{y}}_l^{\mathrm{dec}} \left(\bar{\mathbf{y}}_l^{\mathrm{dec}}\right)^H\right\}$

$\widehat{\sigma}^2 = \frac{N_{\mathrm{u}}}{N_{\mathrm{u}} - \bar{r}_{\max}} p_n - \frac{1}{N_{\mathrm{u}} - \bar{r}_{\max}} tr\left\{\mathbf{R}_l \mathbf{\Theta}_l\right\}$

$\widehat{\bar{r}}_l = \mathrm{card}\left\{\widehat{\lambda}_i : \widehat{\lambda}_i > \beta \cdot \widehat{\sigma}^2\right\}$

3. Basis updating:

$\widehat{\mathbf{Q}}_l = [\mathbf{B}_l]_{\mathrm{col}(1:\widehat{\bar{r}}_l)}$

end for

IV. PILOT REUSE FOR UPLINK CHANNEL TRAINING

In this section, we present PR for uplink channel training, and investigate how PR affects the channel estimation performance. Our following analysis can be applied to arbitrary PR pattern, while the pattern design by exploiting the JADSM will be discussed in the next section.

It is assumed that all the users are synchronized. For ease of exposition, the worst situation is considered, i.e., a unique pilot signal reused by all the users. During the uplink pilot segment, all the users transmit the scheduled pilots simultaneously, and the space-frequency domain signal received at the BS can be represented as

$$\bar{\mathbf{Y}}_l = \sum_{k'=1}^{K} \bar{\mathbf{H}}_{k',l} \bar{\mathbf{X}}_{k'} + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{Z}}_l \tag{20}$$

where the frequency pilot signal satisfies $\mathbf{X}_{k'} \mathbf{X}_k^H = \sigma_{\mathrm{x}}^2 \mathbf{I}_{N_0}$. After decorrelation and power normalization, the channel observation of all the users can be obtained by the BS. Specifically, for user k in a given coherence block, the uplink channel observation can be written as

$$\bar{\mathbf{y}}_{k,l} = \bar{\mathbf{h}}_{k,l} + \underbrace{\sum_{k' \neq k} \bar{\mathbf{h}}_{k',l}}_{\text{pilot interference}} + \underbrace{\frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}}}_{\text{pilot noise}} \tag{21}$$

via employing the vector operator identity $\mathrm{vec}\{\mathbf{A} + \mathbf{B}\} = \mathrm{vec}\{\mathbf{A}\} + \mathrm{vec}\{\mathbf{B}\}$.

As discussed in [29], via projecting $\bar{\mathbf{h}}_{k,l}$ onto the joint angle-delay subspace, the subspace projecting (SP) estimation of the channel $\bar{\mathbf{h}}_{k,l}$ can be given by

$$\widehat{\bar{\mathbf{h}}}_{k,l}^{\mathrm{sp}} = \mathbf{Q}_k \mathbf{Q}_k^H \bar{\mathbf{y}}_{k,l}. \tag{22}$$

For further analysis, we first define a metric to measure the degree of the spatial-frequency orthogonality between two arbitrary JADSMs.

Definition 1: Assuming $\mathbf{A}$, $\mathbf{B} \in \mathbb{C}^{N \times P}$, and their columns originate from the same orthonormal basis $\left\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_N\right\}$, the degree of orthogonality between them can be defined as

$$d_{\mathrm{orth}}(\mathbf{A}, \mathbf{B}) \triangleq \left\|\mathbf{A}^H \mathbf{B}\right\|_{\mathrm{F}}^2 \tag{23}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

where $d_{\rm orth} \in [0,\ P]$, $d(\mathbf{A},\mathbf{B}) = 0$ when $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, and $d(\mathbf{A},\mathbf{B}) = P$ if $\mathbf{A}$ and $\mathbf{B}$ have identical column vectors. Due to the definition of the orthogonality measure in (23) and (A.2), if $d_{\rm orth}(\mathbf{Q}_k,\mathbf{Q}_{k'}) = P$, we have $\mathbf{Q}_k = \mathbf{Q}_{k'}\mathbf{\Pi}$, which indicates that the multipath angles and delays of user k are consistent with that of user k'. Note that $\mathbf{\Pi}$ is a permutation matrix,

Then, three theorems for channel estimation are given as follows:

Theorem 1: For an ULA BS, under the condition that $d_{\rm orth}(\mathbf{Q}_k,\mathbf{Q}_{k'}) = 0$ for $\forall k' \neq k$, we have

$$\lim_{M\to\infty} \widehat{\bar{\mathbf{h}}}_{k,l}^{\rm sp} = \widehat{\bar{\mathbf{h}}}_{k,l}^{\rm no\ int}, \tag{24}$$

where the superscript *no int* refers to the "no interference case".

Proof: See Appendix B. ■

Define the SFCCM $\mathbf{R}_k = {\rm E}\left\{\bar{\mathbf{h}}_{k,l}\bar{\mathbf{h}}_{k,l}^H\right\}$, the MMSE estimate of the channel $\bar{\mathbf{h}}_{k,l}$ based on the channel observation $\bar{\mathbf{y}}_{k,l}$ is given by [37]

$$\widehat{\bar{\mathbf{h}}}_{k,l}^{\rm mmse} = \mathbf{R}_k\left(\sum_{k'\in\mathcal{K}}\mathbf{R}_{k'} + \frac{1}{\rho_{\rm tr}}\mathbf{I}_{N_{\rm u}}\right)^{-1}\bar{\mathbf{y}}_{k,l}. \tag{25}$$

Let $\widetilde{\bar{\mathbf{h}}}_{k,l}^{\rm mmse} = \bar{\mathbf{h}}_{k,l} - \widehat{\bar{\mathbf{h}}}_{k,l}^{\rm mmse}$ be the angle-delay domain channel estimation error of user k. From the orthogonality principle of MMSE estimation [37], $\widetilde{\bar{\mathbf{h}}}_{k,l}^{\rm mmse}$ is independent of $\widehat{\bar{\mathbf{h}}}_{k,l}^{\rm mmse}$. Thus the corresponding covariance matrix of $\widetilde{\bar{\mathbf{h}}}_{k,l}^{\rm mmse}$ is given by

$$\mathbf{C}_k = \mathbf{R}_k - \mathbf{R}_k\left(\sum_{k'\in\mathcal{K}}\mathbf{R}_{k'} + \frac{1}{\rho_{\rm tr}}\mathbf{I}_{N_{\rm u}}\right)^{-1}\mathbf{R}_k. \tag{26}$$

The estimation error covariance is an important measure of the estimation performance. The MSE-CE is defined as

$$\mathcal{M}_k = {\rm tr}\left\{\mathbf{C}_k\right\}. \tag{27}$$

Based on the MMSE estimation, Theorem 2 is presented.

Theorem 2: The range of the MSE-CE $\mathcal{M}_k$ is given by

$$\mathcal{M}_k^{\rm min} \le \mathcal{M}_k \le \mathcal{M}_k^{\rm max} \tag{28}$$

where

$$\mathcal{M}_k^{\rm min} = N_{\rm u}\sigma_{\rm h}^2 - \sum_{p=1}^{P}\frac{\sigma_{k,p}^4}{\sigma_{k,p}^2 + \frac{1}{\rho_{\rm tr}}}, \tag{29}$$

$$\mathcal{M}_k^{\rm max} = N_{\rm u}\sigma_{\rm h}^2 - \sum_{p=1}^{P}\frac{\sigma_{k,p}^4}{\sum\limits_{k'=1}^{K}\sigma_{k',\vartheta_{k',p}}^2 + \frac{1}{\rho_{\rm tr}}}, \tag{30}$$

and $\{\vartheta_{k',1},\cdots,\vartheta_{k',p},\cdots,\vartheta_{k',P}\}$ is a permutation of $\{1,2,\cdots,P\}$. For $\forall k,k' \in \mathcal{K}$ and $k \neq k'$, the minimum value is achieved under the condition that $d_{\rm orth}(\mathbf{Q}_k,\mathbf{Q}_{k'}) = 0$, while $d_{\rm orth}(\mathbf{Q}_k,\mathbf{Q}_{k'}) = P$ conversely.

Proof: See Appendix C. ■

Theorem 2 shows that the MSE-CE can be minimized when $d_{\rm orth}(\mathbf{Q}_k,\mathbf{Q}_{k'}) = 0$, which is also equivalent to that the channels of any two users transmitting the same pilots have different paths, i.e., for $\forall k,k' \in \mathcal{K}$, $(\theta_{k,p} - \theta_{k',p'})(q_{k,p} - q_{k',p'}) \neq 0$. At this point, the joint angle-delay subspace of the users reusing the same pilots are non-overlapping, which can be referred to as the interference-free transmission criteria. This conclusion is consistent with that in [22], i.e., the equivalent channel power distributions in angle-delay domain for different users are non-overlapping. Therefore, it should be particularly noted that delay domain can be exploited to provide significant performance improvement when the channels of different interference users have overlapping bins in angle domain. Meanwhile, Theorem 2 also implies that the MSE-CE gets its maximum when the joint angle-delay subspaces of the users reusing the same pilots fully overlap, which is the worst precondition of pilot decontamination.

Obviously, the MMSE estimator provides more superior estimation performance to the SP estimator, though with higher complexity. However, in mobile wireless links, the channel statistics depend on the particular environments, for example, indoor and outdoor, urban and suburban, and change with time. From (15), multipath powers are required to constitute the SFCCM $\mathbf{R}_k$ with the JADSM known. Due to the extreme difficulty of power tracking for different path, it is not robust to design a channel estimator that exactly matches the channel statistics.

Nevertheless, it has been shown in [34] and [38] that the channel estimator designed for uniform channel power delay profile is robust to the correlation matrix mismatch in OFDM systems for single user. Inspired by this exploration, we will investigate the robust MMSE estimator in a multi-user scenario.

Theorem 3: A robust MMSE estimator under the worst precondition of pilot decontamination is given by

$$\widehat{\bar{\mathbf{h}}}_{k,l}^{\rm robust} = \mathbf{Q}_k\mathbf{Q}_k^H\left(\sum_{k'\in\mathcal{K}}\mathbf{Q}_{k'}\mathbf{Q}_{k'}^H + \frac{P}{\rho_{\rm tr}N_{\rm u}\sigma_{\rm h}^2}\mathbf{I}_{N_{\rm u}}\right)^{-1}\bar{\mathbf{y}}_{k,l} \tag{31}$$

as well as the robust MSE-CE

$$\mathcal{M}_k^{\rm robust} = N_{\rm u}\sigma_{\rm h}^2 - \frac{N_{\rm u}\sigma_{\rm h}^2}{P} \cdot {\rm tr}\left\{\mathbf{Q}_k\mathbf{Q}_k^H\left(\sum_{k'\in\mathcal{K}}\mathbf{Q}_{k'}\mathbf{Q}_{k'}^H + \frac{P}{\rho_{\rm tr}N_{\rm u}\sigma_{\rm h}^2}\mathbf{I}_{N_{\rm u}}\right)^{-1}\right\}. \tag{32}$$

Proof: See Appendix D. ■

Although (31) and (32) are derived under the worst precondition of pilot decontamination, it will be shown in Section VI-B how robustness can be maintained in general interference-dominated cases.

V. JOINT ANGLE-DELAY SUBSPACE BASED PILOT SCHEDULING

So far, we have analyzed the joint angle-delay subspace acquisition and channel training of massive MIMO-OFDM transmission with pilot reuse.

The available orthogonal pilot set is denoted as $\mathcal{P} = \{1,2,\cdots,T\}$, and the πth pilot matrix as $\mathbf{X}_\pi \in \mathbb{C}^{N_0\times N_0}$. We also denote an pilot reuse pattern with user set and pilot set as $\mathcal{H}(\mathcal{K},\mathcal{P}) = \{(k,\pi_k) : k\in\mathcal{K}, \pi_k\in\mathcal{P}\}$. If interfering

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

users are partitioned into T index sets $\{\mathcal{K}_t\}$, with $K_t = |\mathcal{K}_t|$ users in the tth group, then the pilot scheduling problem can be described as follows

$$\begin{aligned} &\mathcal{H}_{\text{opt}}(\mathcal{K},\mathcal{P}) = \underset{\mathcal{H}(\mathcal{K},\mathcal{P})}{\arg\min} \sum\nolimits_{t\in\mathcal{P}} \sum\nolimits_{k\in\mathcal{K}_t} \mathcal{M}_k \\ &s.t. \quad K_t \in \mathbb{N}^+ \; and \; \sum\nolimits_{t\in\mathcal{P}} K_t = K. \end{aligned} \tag{33}$$

Algorithm 2 Modified Greedy Pilot Scheduling (MGPS)

Input: The user set $\mathcal{K}$, the orthogonal pilot set $\mathcal{P}$ and the JADSMs $\{\mathbf{Q}_k\}$.

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

Step 1 Assign the users with "similar" JADSMs with orthogonal pilots. Set the initial user set $\mathcal{K}^{\text{int}} = \mathcal{K}$.

while $\mathcal{K}^{\text{int}} \neq \emptyset$ **do**

Initialize the unscheduled user set $\mathcal{K}^{\text{un}} = \mathcal{K}$ and the unused pilot set $\mathcal{P}^{\text{un}} = \mathcal{P}$. Select the first user as the initial solution, i.e., $m_1 = \mathcal{K}^{\text{int}}(1)$, $\pi_1 = 1$, $\mathcal{K}^{\text{un}} \leftarrow \mathcal{K}^{\text{un}} \backslash \{m_1\}$ and $\mathcal{P}^{\text{un}} \leftarrow \mathcal{P}^{\text{un}} \backslash \{1\}$.

  while $\mathcal{P}^{\text{un}} \neq \emptyset$ **do**

    for the pilot $t \in \mathcal{P}^{\text{un}}$ **do**

      $m_t = \underset{i\in\mathcal{K}^{\text{un}}}{\arg\max} \sum_{j\in\mathcal{P}\backslash\mathcal{P}^{\text{un}}} d\left(\mathbf{Q}_i, \mathbf{Q}_{m_j}\right)$;

      $\pi_{m_t} = t$;

      $\mathcal{K}^{\text{un}} \leftarrow \mathcal{K}^{\text{un}} \backslash \{m_t\}$ and $\mathcal{P}^{\text{un}} \leftarrow \mathcal{P}^{\text{un}} \backslash \{t\}$.

    end for

  end while

Calculate the robust MSE-CE in (33) and $\mathcal{K}^{\text{int}} \leftarrow \mathcal{K}^{\text{int}} \backslash \{1, m_2\}$.

end while

Find $\mathcal{K}^{\text{un}}$ corresponding to the minimum robust MSE-CE.

Step 2 Each unscheduled user is assigned with the best pilot so that the JADSMs of the users reusing the same pilots are as orthogonal as possible.

while $\mathcal{K}^{\text{un}} \neq \emptyset$ **do**

  for user $k \in \mathcal{K}^{\text{un}}$ **do**

    $n_k = \underset{q\in\mathcal{P}}{\arg\min} \sum_{s\in\mathcal{K}_q} d\left(\mathbf{Q}_k, \mathbf{Q}_s\right)$;

    $\pi_k = n_k$ and $\mathcal{K}_{n_k} \leftarrow \mathcal{K}_{n_k} \cup \{k\}$;

    $\mathcal{K}^{\text{un}} \leftarrow \mathcal{K}^{\text{un}} \backslash \{k\}$.

  end for

end while

The solution to this optimization is computationally complex and involves not only a brute-force search over every possible grouping, but also the calculation of the MSE-CE for each user. The complexity of classical exhausted search in (33), in terms of the (complex) scalar multiplication number which dominates the computational complexity, is briefly evaluated as follows. Recalling (27), the scalar multiplication number required in the evaluation of the objective function in (33) is $\mathcal{O}\left(N_{\text{u}}^3 T\right)$. Thus, the computational complexity of running exhausted search is $\mathcal{O}\left(T^{K+1} N_{\text{u}}^3\right)$.

Fortunately, user clustering method is an effective approach to solve this problem, which is motivated by the conditions for optimal channel estimation in Theorem 1 and Theorem 2. Similar difficulty for the single-cell scenario was addressed in [39], with the statistical greedy pilot scheduling (SGPS) algorithm. But, the SGPS algorithm is sensitive to its initialization. Therefore, we propose a modified greedy pilot scheduling (MGPS) algorithm based on the SGPS algorithm. Detailed description of our algorithm is summarized in Algorithm 2.

We evaluate the complexity of the MGPS algorithm as follows. In the process of the MGPS algorithm, $\begin{pmatrix} K \\ 2 \end{pmatrix} = \frac{K(K-1)}{2}$ orthogonality computations defined in (23) are needed, which can be performed before Step 1 in the Algorithm 2. Note that the scalar multiplication number needed in each orthogonality computation is $\mathcal{O}\left(P^2 N_{\text{u}}\right)$, thus the computational complexity of the preparation work is $\mathcal{O}\left(K^2 P^2 N_{\text{u}}\right)$. In addition, in order to obtain good initialization for greedy pilot scheduling, the objective function in (33) are evaluated $\frac{K}{2}$ times, which requires $\mathcal{O}\left(K N_{\text{u}}^3 T\right)$ computation. These results indicate that our MGPS algorithm gives a substantial computational complexity reduction compared to the exhausted search method. Meanwhile, the additional overhead in the calculation of the robust MSE-CE in (33) compared with the SGPS algorithm brings significant performance improvements, as shown in Section VI-B.

TABLE I
SYSTEM PARAMETERS

Parameter	Value
System bandwidth	3 MHz
Sampling duration T_{s}	0.26 μs
Subcarrier spacing	15 kHz
Subcarrier number N_{c}	256
Guard interval N_{g}	16
Symbol length T_{sym}	70.8 μs
Pilot subcarrier number N_0	16

TABLE II
BASIC SIMULATION PARAMETERS

Parameter	Value
Cell radius	1 km
Cell edge SNR ρ_{eg}	20 dB
Number of users	42
Distance from a user to its BS	800 m
Path loss exponent γ	3
Antenna Spacing	$\lambda/2$
Number of paths	6
RMS delay spread ς	0.77 μs
Doppler frequency υ	400 Hz

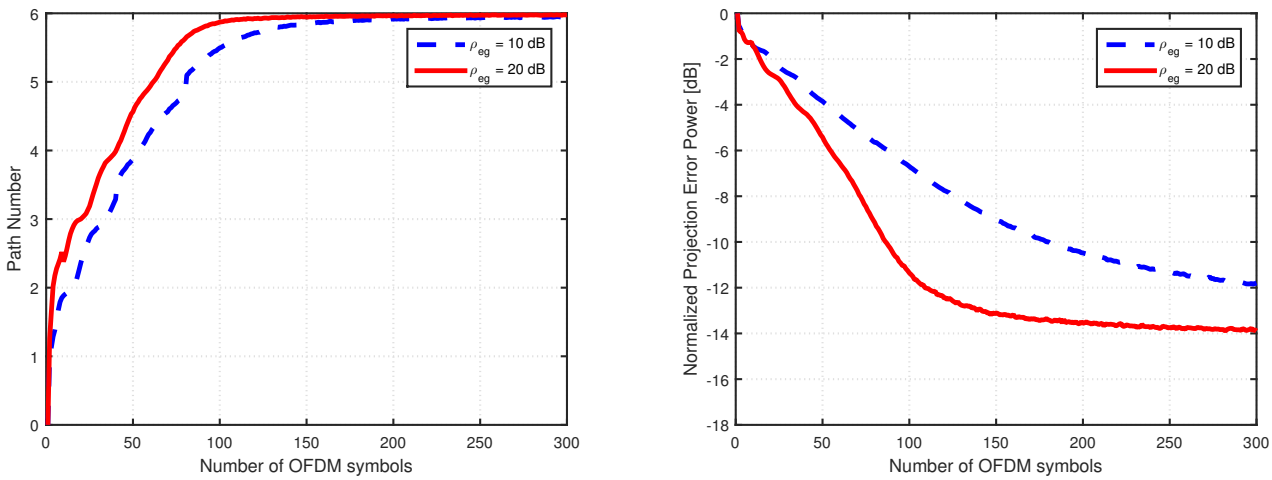
VI. NUMERICAL RESULTS

In this section, we present numerical simulation results to evaluate the performance of the proposed channel estimation for massive MIMO-OFDM systems in multipath Rayleigh fading channel. The major system parameters are summarized in Table I.

Consider channels with 6 taps in time domain, which exhibit an exponential PDP

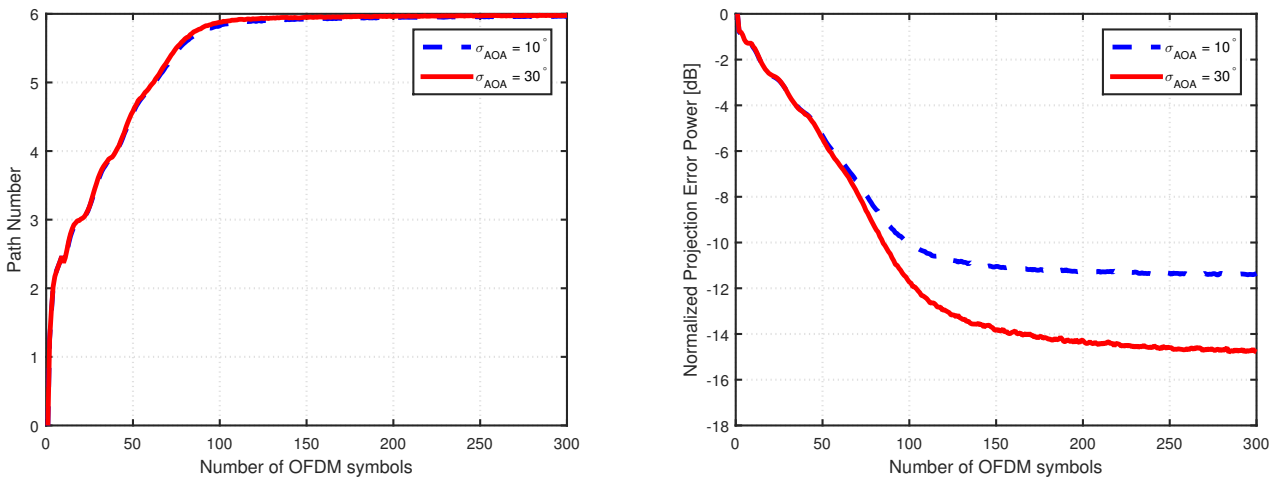
$$S_k^{\text{del}}(\tau) \propto \exp\left(-\tau/\varsigma_k\right), \text{ for } \tau \in [0, N_g T_s]. \tag{34}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access



(a) Adaptive estimate of path number versus number of OFDM symbols. (b) Normalized projection error power versus number of OFDM symbols.

Fig. 2. Performance comparison of subspace tracking between $\rho_{\rm eg}$ = 10 dB and $\rho_{\rm eg}$ = 20 dB with $M = 10$ and $\sigma_{\rm AOA} = 20°$.



(a) Adaptive estimate of path number versus number of OFDM symbols. (b) Normalized projection error power versus number of OFDM symbols.

Fig. 3. Performance comparison of subspace tracking between $\sigma_{\rm AOA} = 10°$ and $\sigma_{\rm AOA} = 30°$ with $M = 10$ and $\rho_{\rm eg}$ = 20 dB.

where ς_k denotes the root mean square (RMS) delay spread of user k. Note that the average channel power of user k is given by

$$\eta_k = \frac{c}{d_k^{\gamma}}, \tag{35}$$

where c is a constant, dependent on the prescribed average SNR at cell edge, γ is the path loss exponent, and d_k is the geographical distance. Some basic parameters are given in Table II.

The AOA distribution considered here is Gaussian distribution. For the channel coefficients generated according to (2), the AOAs of all P paths are i.i.d. Gaussian random variables with mean $\bar{\theta}_k$ and standard deviation $\sigma_{\rm AOA}$. It is assumed that all the desired channels and interference channels have the same standard deviation of AOA.

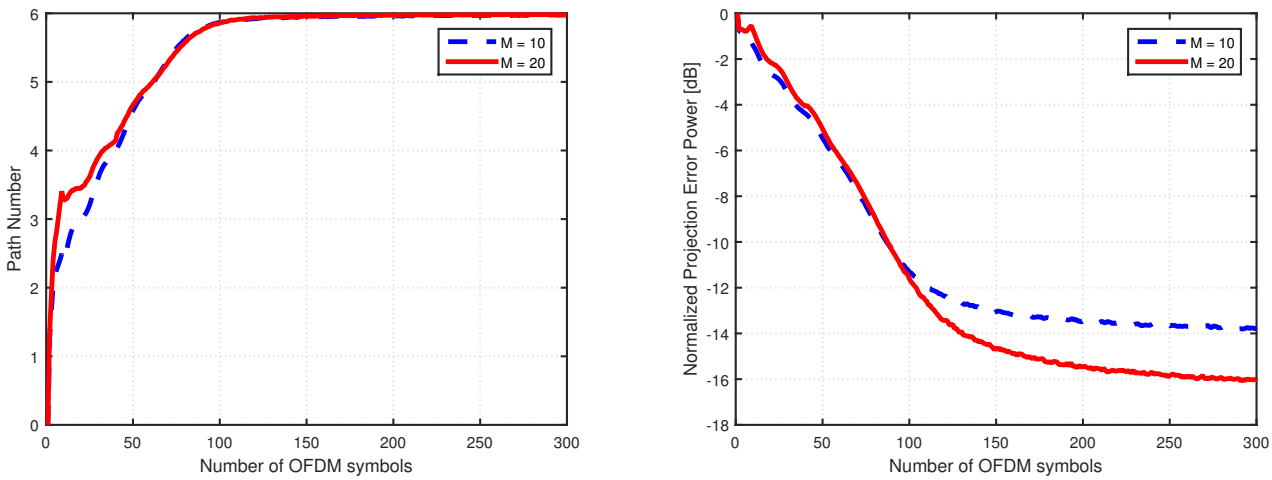
A. Performance of the LORAF Algorithm Applied in Joint Angle-Delay Subspace Tracking

In this subsection, we use the *normalized projection error power* [40] over the pilot subcarriers to evaluate the accuracy of joint angle-delay subspace tracking algorithm developed in Section III-B. To define this quantity let us first introduce the normalized projection error power at OFDM symbol l as

$$e_{\rm np}(l) \triangleq \mathbb{E}\left\{ \frac{\left\| \widehat{\mathbf{Q}}_l \widehat{\mathbf{Q}}_l^H - \mathbf{Q}\mathbf{Q}^H \right\|_{\rm F}^2}{\left\| \mathbf{Q}\mathbf{Q}^H \right\|_{\rm F}^2} \right\}. \tag{36}$$

The parameters of the subspace tracking algorithm (Algorithm 1) have been set to be $\mu = 0.999$, $\bar{r}_{\max} = 8$ and $\beta = 1$. Notice that the parameters have been selected without any attempt of optimization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access



(a) Adaptive estimate of path number versus number of OFDM symbols.

(b) Normalized projection error power versus number of OFDM symbols.

Fig. 4. Performance comparison of subspace tracking between $M = 10$ and $M = 20$ with $\sigma_{\mathrm{AOA}} = 20^\circ$ and ρ_{eg} = 20 dB.

Fig.2 shows the performance comparison of the LORAF algorithm applied in joint angle-delay subspace tracking with two different values of cell edge SNR. It can be seen from Fig.2a that the path number estimate converges after 100 OFDM symbols. From Fig.2b, it can be seen that a bigger ρ_{eg} results in faster convergence of the LORAF algorithm, leading to a more accurate estimation of the JADSM.

Fig.3 depicts the performance comparison of the LORAF algorithm applied in joint angle-delay subspace tracking with two different values of AOA spread. It can be seen from Fig.3a that the path number estimate using the LORAF algorithm is insensitive to the AOA spread. It can be seen from Fig.3b that larger AOA spread provides better orthogonality between columns of the JADSM with finite antennas in the BS, bringing a little faster decline rate of the projection error power. Since the BS is generally high on top of a tower with few scatters around it, the angle dispersion becomes small [41]. However, with tens of antennas deployed in the BS in massive MIMO, the high accuracy of the JADSM acquisition can be provided, as illustrated in Fig.4.

B. Performance of Pilot Scheduling Algorithms Combined with Channel Estimators

In this subsection, we evaluate the performance of three pilot scheduling algorithms, namely the random pilot scheduling (RPS) method, SGPS method, and MGPS method, combined with three channel estimation approaches containing the SP, MMSE and robust MMSE estimator.

Fig.5 plots the normalized MSE-CE metric in (33) versus the number of antennas in the BS. The AOAs of different users have the same Gaussian distribution with $\sigma_{\mathrm{AOA}} = 20^\circ$, yielding serious angle overlap between desired and interfering multipaths. However, with an increasing number of BS antennas, better orthogonality between JADSMs of different users leads to a more superior channel estimation performance. Obviously, the RPS approach has the poorest performance no matter what channel estimators are used. It is seen that the interference-suppressed performance of the SGPS and MGPS method is much better than that of the RPS method, especially with the MMSE and robust MMSE estimator. Moreover, the proposed MGPS method notably outperforms the SGPS method and its MSE-CE gap between the MMSE and robust MMSE estimator becomes smaller with more antennas.

Fig.6 shows the impact of standard deviation σ_{AOA} of Gaussian AOAs on the mentioned pilot scheduling algorithms. It is seen that the channel estimation in delay-angle domain is insensitive to the angle spread. Although the AOAs of different users may have more overlaps with a bigger σ_{AOA}, both the MMSE and robust MMSE channel estimation provide steady performance owing to the separation of interfering channels in delay domain. In particular, the proposed MGPS method combined with the robust MMSE estimator achieves considerable performance gains even for a large AOA spread.

Fig.7 shows the normalized MSE-CE versus the number of orthogonal pilots. It is seen that the orthogonal pilot number has a significant impact on different pilot scheduling schemes. As the pilot number increases, less users are assigned to any group, leading to less pilot interference. Besides, the MGPS method can have better performance gains with a relatively smaller orthogonal pilot number than the SGPS method, when MMSE or robust MMSE channel estimation is performed.

Fig.8 shows the impact of cell edge SNR ρ_{eg} on different channel estimators. Since the pilot interference dominates the noise in our scenario, the SP estimator has poor performance even in the high ρ_{eg} regime, as seen from Fig.8. Nonetheless, the other two channel estimators achieve better performance with higher ρ_{eg}, by adding an identity matrix multiplied by a ρ_{eg} related factor before inverting, which is similar to the MMSE multi-user precoding [42]. Furthermore, with the MMSE or robust MMSE estimator, the MSE-CE of the proposed MGPS method decreases at a faster pace than that of the SGPS method as the ρ_{eg} increases.

In general, the proposed robust MMSE estimator has a close performance to the ideal MMSE estimator, as shown in Fig.5-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

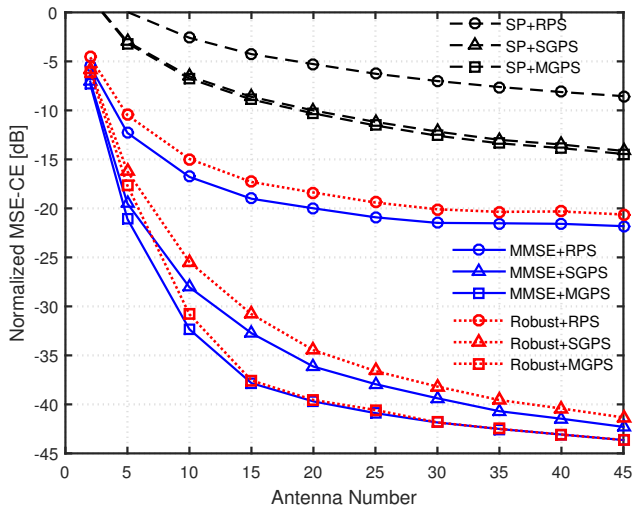


Fig. 5. Normalized MSE-CE versus antenna number deployed in the BS with 6 orthogonal pilots, $\sigma_{\mathrm{AOA}} = 20^{\circ}$ and ρ_{eg} = 20 dB.

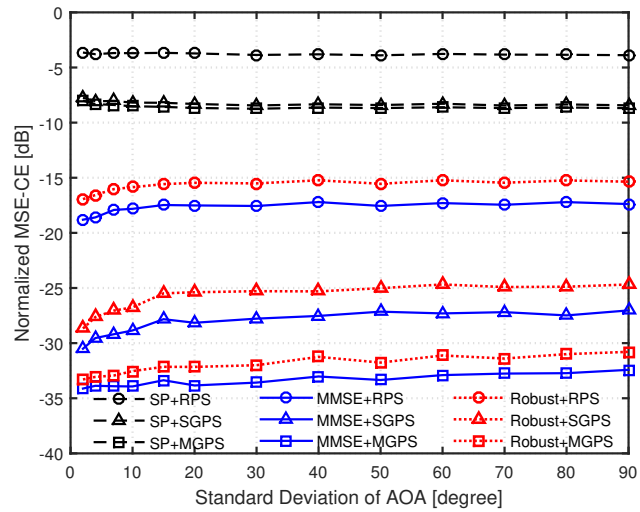


Fig. 6. Normalized MSE-CE versus standard deviation of Gaussian distributed AOAs with $M = 10$, ρ_{eg} = 20 dB and 6 orthogonal pilots.

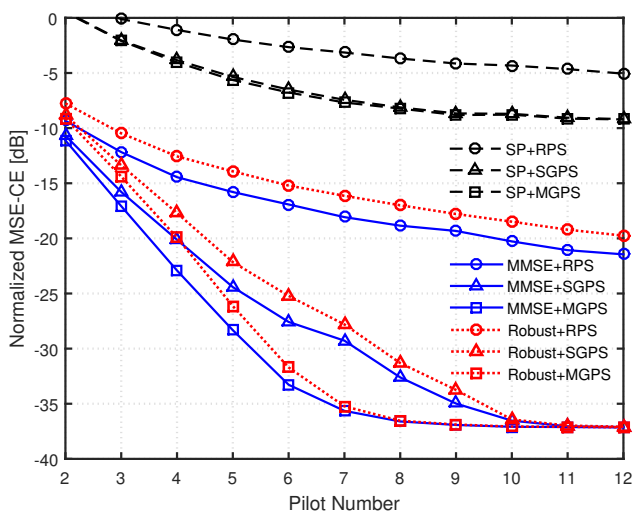


Fig. 7. Normalized MSE-CE versus number of orthogonal pilots with $M = 10$ and $\sigma_{\mathrm{AOA}} = 20^{\circ}$ and ρ_{eg} = 20 dB.

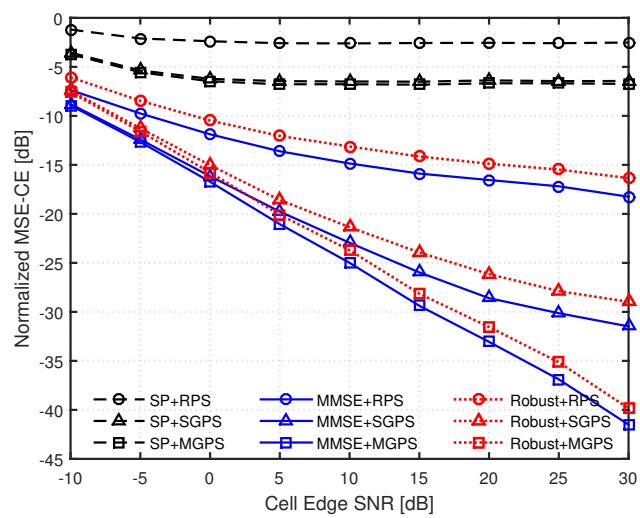


Fig. 8. Normalized MSE-CE versus cell edge SNR with $M = 10$, $\sigma_{\mathrm{AOA}} = 20^{\circ}$ and 6 orthogonal pilots.

8. In addition, the MGPS method relatively has an excellent capability of pilot scheduling.

VII. Conclusion

In this paper, we have proposed a novel joint angle-delay subspace based channel estimation framework for massive MIMO-OFDM systems. By employing the parametric channel modeling, we first derived the relationship between the SFCCM and JADSM, and applied the low-complexity LORAF algorithm to JADSM estimation. Then, we investigated channel estimation for massive MIMO-OFDM with the JADSMs of users, and provided an interference-free transmission condition that the JADSMs of users transmitting the same pilots are non-overlapping. Furthermore, we developed an efficient pilot scheduling strategy relying on the proposed robust MMSE estimation. The performance has been evaluated in terms of the normalized MSE-CE both theoretically and numerically, demonstrating the significant benefits of the proposed PR scheme and channel estimation algorithms.

Appendix A
Proof of Proposition 1

We refer to the pth column vector of $\bar{\mathbf{U}}(\theta, \mathbf{q})$ as $\bar{\mathbf{u}}_p = \bar{\mathbf{w}}(q_p) \otimes \mathbf{v}(\theta_p)$. From Lemma 1, $\mathbf{v}^H(\theta_p)\mathbf{v}(\theta_{p'}) = 0$ for $\theta_p \neq \theta_{p'}$. From (12), it can be shown that

$$\begin{aligned}\bar{\mathbf{w}}^H(q_p)\bar{\mathbf{w}}(q_{p'}) &= 1 + \exp\left(-j\frac{2\pi}{N_{\mathrm{c}}}(q_p - q_{p'})\Delta N\right) + \cdots \\ &\quad + \exp\left(-j\frac{2\pi}{N_{\mathrm{c}}}(q_p - q_{p'})(N_{\mathrm{c}} - \Delta N)\right) \\ &= \frac{1 - \exp(-j2\pi(q_p - q_{p'}))}{1 - \exp\left(-j\frac{2\pi}{N_{\mathrm{c}}}(q_p - q_{p'})\Delta N\right)} = 0\end{aligned} \tag{A.1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

where $q_p - q_{p'} \in \mathbb{N} \backslash \{0\}$.

Since any two different paths must be subject to $(\theta_p - \theta_{p'})(q_p - q_{p'}) \neq 0$, we have

$$\begin{aligned}\bar{\mathbf{u}}_p^H \bar{\mathbf{u}}_{p'} &= \left(\bar{\mathbf{w}}(q_p) \otimes \mathbf{v}(\theta_p)\right)^H \left(\bar{\mathbf{w}}(q_{p'}) \otimes \mathbf{v}(\theta_{p'})\right) \\ &\overset{(a)}{=} \left(\bar{\mathbf{w}}^H(q_p)\bar{\mathbf{w}}(q_{p'})\right) \otimes \left(\mathbf{v}^H(\theta_p)\mathbf{v}(\theta_{p'})\right) = 0\end{aligned} \tag{A.2}$$

where (a) is derived from the Kronecker product identities $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$ and $(\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) = (\mathbf{AB}) \otimes (\mathbf{CD})$. With $\bar{\mathbf{w}}^H(q_p)\bar{\mathbf{w}}(q_p) = N_{\mathrm{p}}$ and $\mathbf{v}^H(\theta_p)\mathbf{v}(\theta_p) = M$, we have $\bar{\mathbf{u}}_p^H \bar{\mathbf{u}}_p = MN_{\mathrm{p}}$. This concludes the proof.

Appendix B
Proof of Theorem 1

Combining the subspace tracking based channel estimate (22) and the channel observation (21) at the BS, we obtain

$$\widehat{\bar{\mathbf{h}}}_{k,l}^{\mathrm{sp}} = \mathbf{Q}_k \mathbf{Q}_k^H \left(\bar{\mathbf{h}}_{k,l} + \sum_{k' \neq k} \bar{\mathbf{h}}_{k',l} + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \bar{\mathbf{z}}_{\mathrm{iid}}\right). \tag{B.1}$$

Since the singular value decomposition (SVD) of $\bar{\mathbf{h}}_{k',l}$ is $\mathbf{Q}_{k'} \mathbf{\Lambda}_{k'} \mathbf{P}_{k'}^H$ and $\mathbf{Q}_k^H \mathbf{Q}_{k'} = 0$ for $\forall k \neq k'$, for large M,

$$\begin{aligned}\widehat{\bar{\mathbf{h}}}_{k,l}^{\mathrm{sp}} &\approx \bar{\mathbf{h}}_{k,l} + \sum_{k' \neq k} \mathbf{Q}_k \left(\mathbf{Q}_k^H \mathbf{Q}_{k'}\right) \mathbf{\Lambda}_{k'} \mathbf{P}_{k'}^H \bar{\alpha}_{k',l} \\ &\quad + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \mathbf{Q}_k \mathbf{Q}_k^H \bar{\mathbf{z}}_{\mathrm{iid}} \\ &\approx \bar{\mathbf{h}}_{k,l} + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \mathbf{Q}_k \mathbf{Q}_k^H \bar{\mathbf{z}}_{\mathrm{iid}}.\end{aligned} \tag{B.2}$$

This concludes the proof.

Appendix C
Proof of Theorem 2

Let $\mathbf{Q}_k^{\perp}$ be the orthogonal complement matrix of $\mathbf{Q}_k$, column space of which is orthogonal to that of $\mathbf{Q}_k$. Then, due to the column orthogonality of $\mathbf{Q}_k$, we construct an unitary orthogonal matrix $\mathbf{E}_k = \left[\mathbf{Q}_k \; \mathbf{Q}_k^{\perp}\right]$, with each column vector corresponding to a different path, which has the same form as $\frac{1}{\sqrt{N_{\mathrm{u}}}} \bar{\mathbf{u}}_p$ in (5).

Thus, we have

$$\mathbf{E}_k = \mathbf{E}_{k'} \mathbf{\Pi}_{k'} \tag{C.1}$$

where the permutation matrix $\mathbf{\Pi}_{k'}$ is an unitary orthogonal matrix and $\mathbf{\Pi}_{k'}^2 = \mathbf{I}_{N_{\mathrm{u}}}$. From (C.1), the correlation matrix $\mathbf{R}_{k'}$ can be transformed into

$$\begin{aligned}\mathbf{R}_{k'} &= \mathbf{Q}_{k'} \mathbf{\Sigma}_{k'} \mathbf{Q}_{k'}^H \\ &= \mathbf{E}_{k'} \begin{bmatrix} \mathbf{\Sigma}_{k'} & \\ & \mathbf{0}_{N_{\mathrm{u}}-P} \end{bmatrix} \mathbf{E}_{k'}^H \\ &= \mathbf{E}_k \mathbf{\Pi}_{k'} \begin{bmatrix} \mathbf{\Sigma}_{k'} & \\ & \mathbf{0}_{N_{\mathrm{u}}-P} \end{bmatrix} \mathbf{\Pi}_{k'} \mathbf{E}_k^H \\ &= \mathbf{E}_k \mathbf{\Sigma}_{k'}^{\mathrm{pm}} \mathbf{E}_k^H\end{aligned} \tag{C.2}$$

where $\mathbf{\Sigma}_{k'} = \mathrm{diag}\left\{N_{\mathrm{u}}\sigma_{k',1}^2 \cdots N_{\mathrm{u}}\sigma_{k',P}^2\right\}$, and $\mathbf{\Sigma}_{k'}^{\mathrm{pm}}$ is still a diagonal matrix with its diagonal elements permuted.

To simplify the MSE expression, we use the fact that $\mathrm{tr}\{\mathbf{A} \pm \mathbf{B}\} = \mathrm{tr}\{\mathbf{A}\} \pm \mathrm{tr}\{\mathbf{B}\}$ and $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$. Substituting (C.2) into (27), we have

$$\begin{aligned}\mathcal{M}_k &= \mathrm{tr}\{\mathbf{R}_k\} \\ &\quad - \mathrm{tr}\left\{\mathbf{R}_k \left(\sum_{k' \in \mathcal{K}} \mathbf{E}_k \mathbf{\Sigma}_{k'}^{\mathrm{pm}} \mathbf{E}_k^H + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}}\right)^{-1} \mathbf{R}_k\right\} \\ &= \mathrm{tr}\left\{\begin{bmatrix} \mathbf{\Sigma}_k & \\ & \mathbf{0}_{N_{\mathrm{u}}-P} \end{bmatrix}\right\} \\ &\quad - \mathrm{tr}\left\{\begin{bmatrix} \mathbf{\Sigma}_k^2 & \\ & \mathbf{0}_{N_{\mathrm{u}}-P} \end{bmatrix} \left(\sum_{k' \in \mathcal{K}} \mathbf{\Sigma}_{k'}^{\mathrm{pm}} + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}}\right)^{-1}\right\} \\ &\triangleq N_{\mathrm{u}}\sigma_{\mathrm{h}}^2 - \mathcal{S}_k.\end{aligned} \tag{C.3}$$

From (C.3), the maximum value of $\mathcal{S}_k$ equals

$$\mathcal{S}_k^{\max} = \sum_{p=1}^{P} \frac{N_{\mathrm{u}}^2 \sigma_{k,p}^4}{N_{\mathrm{u}}\sigma_{k,p}^2 + \frac{1}{\rho_{\mathrm{tr}}}}, \tag{C.4}$$

when the positions of the diagonal elements of $\mathbf{\Sigma}_{k'}^{\mathrm{pm}}$ are non-overlapping with that of $\mathbf{\Sigma}_k^{\mathrm{pm}}$, for $k \neq k'$, $d_{\mathrm{orth}}(\mathbf{Q}_k, \mathbf{Q}_{k'}) = 0$. On the contrary, when $d_{\mathrm{orth}}(\mathbf{Q}_k, \mathbf{Q}_{k'}) = P$, the positions of the diagonal elements of $\mathbf{\Sigma}_{k'}^{\mathrm{pm}}$ are all overlapped with that of $\mathbf{\Sigma}_k^{\mathrm{pm}}$, thus the minimum value of $\mathcal{S}_k$ is

$$\mathcal{S}_k^{\min} = \sum_{p=1}^{P} \frac{N_{\mathrm{u}}^2 \sigma_{k,p}^4}{\sum_{k'=1}^{K} N_{\mathrm{u}}\sigma_{k',\vartheta_{k',p}}^2 + \frac{1}{\rho_{\mathrm{tr}}}}, \tag{C.5}$$

where $\{\vartheta_{k',1}, \cdots, \vartheta_{k',p}, \cdots, \vartheta_{k',P}\}$ is a permutation of $\{1, 2, \cdots, P\}$ and its order is related to $\mathbf{Q}_k$ and $\mathbf{Q}_{k'}$. This concludes the proof.

Appendix D
Proof of Theorem3

In this appendix, we consider the worst precondition of pilot decontamination, and design a robust MMSE estimator for multi-users transmitting the same pilots. From Theorem 2, when $d(\mathbf{Q}_k, \mathbf{Q}_{k'}) = P$, the MSE-CE $\mathcal{M}_k$ achieves its maximum value. Similar to the derivation of (C.3), we can get

$$\begin{aligned}\widehat{\bar{\mathbf{h}}}_{k,l}^{\mathrm{worst}} &= \mathbf{E}_k \mathbf{\Sigma}_k^{\mathrm{pm}} \mathbf{E}_k^H \left(\sum_{k' \in \mathcal{K}} \mathbf{E}_k \mathbf{\Sigma}_{k'}^{\mathrm{pm}} \mathbf{E}_k^H + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}}\right)^{-1} \bar{\mathbf{y}}_{k,l} \\ &= \mathbf{E}_k \mathbf{\Sigma}_k^{\mathrm{pm}} \left(\sum_{k' \in \mathcal{K}} \mathbf{\Sigma}_{k'}^{\mathrm{pm}} + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_{N_{\mathrm{u}}}\right)^{-1} \mathbf{E}_k^H \bar{\mathbf{y}}_{k,l} \\ &= \left[\mathbf{Q}_k \; \mathbf{Q}_k^{\perp}\right] \begin{bmatrix} \mathbf{\Lambda}_k & \\ & \mathbf{0}_{N_{\mathrm{u}}-P} \end{bmatrix} \left[\mathbf{Q}_k \; \mathbf{Q}_k^{\perp}\right]^H \bar{\mathbf{y}}_{k,l} \\ &= \mathbf{Q}_k \mathbf{\Lambda}_k \mathbf{Q}_k^H \bar{\mathbf{y}}_{k,l}\end{aligned} \tag{D.1}$$

where the diagonal matrix $\mathbf{\Lambda}_k = \mathrm{diag}\left\{\lambda_1 \; \lambda_2 \cdots \lambda_P\right\}$ with its pth diagonal element $\lambda_p = \frac{N_{\mathrm{u}}\sigma_{k,p}^2}{N_{\mathrm{u}} \sum_{k'=1}^{K} \sigma_{k',\vartheta_{k',p}}^2 + 1/\rho_{\mathrm{tr}}}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

The estimation error $\boldsymbol{\epsilon}^{\text{worst}} = \bar{\mathbf{h}}_{k,l} - \widehat{\bar{\mathbf{h}}}_{k,l}^{\text{worst}}$ of (D.1) is

$$\boldsymbol{\epsilon}^{\text{worst}} = \mathbf{Q}_k \left(\mathbf{I}_P - \boldsymbol{\Lambda}_k\right) \mathbf{Q}_k^H \bar{\mathbf{h}}_{k,l} - \mathbf{Q}_k \boldsymbol{\Lambda}_k \mathbf{Q}_k^H \sum_{k' \neq k} \bar{\mathbf{h}}_{k',l} - \mathbf{Q}_k \boldsymbol{\Lambda}_k \mathbf{Q}_k^H \frac{1}{\sqrt{\rho_{\text{tr}}}} \bar{\mathbf{z}}_{\text{iid}} \tag{D.2}$$

and the MSE is

$$\mathcal{M}_k^{\text{worst}} = \text{tr}\left\{\mathbb{E}\left\{\boldsymbol{\epsilon}^{\text{worst}} \left(\boldsymbol{\epsilon}^{\text{worst}}\right)^H\right\}\right\}. \tag{D.3}$$

To simplify this expression, we will use the following trace identities that

- $\text{tr}\{\mathbf{AB}\} = \text{tr}\{\mathbf{BA}\}$;
- $\text{tr}\{\mathbf{A} \pm \mathbf{B}\} = \text{tr}\{\mathbf{A}\} \pm \text{tr}\{\mathbf{B}\}$;
- $\text{tr}\left\{\mathbf{DAD}^H\right\} = \sum_i a_{i,i} d_i^2$ when $\mathbf{D}$ is a diagonal matrix with the elements d_i on its diagonal and $\mathbf{A}$ has diagonal elements $a_{i,i}$.

Substituting (D.2) in (D.3), the MSE becomes

$$\begin{aligned}\mathcal{M}_k^{\text{worst}} &= \text{tr}\Big\{\mathbf{Q}_k \left(\mathbf{I}_P - \boldsymbol{\Lambda}_k\right) \mathbf{Q}_k^H \mathbf{R}_k \mathbf{Q}_k (\mathbf{I}_P - \boldsymbol{\Lambda}_k)^H \mathbf{Q}_k^H \\ &\quad + \mathbf{Q}_k \boldsymbol{\Lambda}_k \mathbf{Q}_k^H \sum_{k' \neq k} \mathbf{R}_{k'} \mathbf{Q}_k \boldsymbol{\Lambda}_k^H \mathbf{Q}_k^H \\ &\quad + \mathbf{Q}_k \boldsymbol{\Lambda}_k \mathbf{Q}_k^H \frac{1}{\rho_{\text{tr}}} \mathbf{I}_P \mathbf{Q}_k \boldsymbol{\Lambda}_k^H \mathbf{Q}_k^H \Big\} \\ &= N_{\text{u}} \sum_{p=1}^{P} \sigma_{k,p}^2 (1 - \lambda_p)^2 \\ &\quad + N_{\text{u}} \sum_{k' \neq k} \sum_{p=1}^{P} \sigma_{k',\vartheta_{k',p}}^2 {\lambda_p}^2 + \frac{P {\lambda_p}^2}{\rho_{\text{tr}}}\end{aligned} \tag{D.4}$$

where $\sum_{p=1}^{P} \sigma_{k',\vartheta_{k',p}}^2 = \sum_{p=1}^{P} \sigma_{k',p}^2 = \sigma_{\text{h}}^2$, for $\forall k' \in \mathcal{K}$. If $\lambda_1 = \lambda_2 = \cdots = \lambda_P$, the MSE $\mathcal{M}_k^{\text{worst}}$ will match different channel power delay profile, thus we get

$$\mathcal{M}_k^{\text{worst}} = N_{\text{u}} \sigma_{\text{h}}^2 (1 - \lambda_p)^2 + (K - 1) N_{\text{u}} \sigma_{\text{h}}^2 \lambda_p^2 + \frac{P {\lambda_p}^2}{\rho_{\text{tr}}}. \tag{D.5}$$

The minimum MSE can be obtained as

$$\frac{\partial}{\partial \lambda_p} \mathcal{M}_k^{\text{worst}} = 2\left(K N_{\text{u}} \sigma_{\text{h}}^2 + \frac{P}{\rho_{\text{tr}}}\right) \lambda_p - 2 N_{\text{u}} \sigma_{\text{h}}^2 = 0. \tag{D.6}$$

Then, we consider the uniform channel power delay profile and denote $\sigma_{k',p}^2 = \zeta$, for $\forall k' \in \mathcal{K}$ and $p = 1, 2 \cdots, P$. Substituting $\lambda_p = \frac{N_{\text{u}} \zeta}{K N_{\text{u}} \zeta + 1/\rho_{\text{tr}}}$ into (D.6) yields the optimum solution

$$\zeta_{opt} = \frac{\sigma_{\text{h}}^2}{P}. \tag{D.7}$$

Combining (D.7) with (25), the robust MMSE estimator can be written as

$$\begin{aligned}\widehat{\bar{\mathbf{h}}}_{k,l}^{\text{robust}} &= \mathbf{Q}_k \frac{N_{\text{u}} \sigma_{\text{h}}^2}{P} \mathbf{I}_P \mathbf{Q}_k^H \\ &\quad \cdot \left(\sum_{k' \in \mathcal{K}} \mathbf{Q}_{k'} \frac{N_{\text{u}} \sigma_{\text{h}}^2}{P} \mathbf{I}_P \mathbf{Q}_{k'}^H + \frac{1}{\rho_{\text{tr}}} \mathbf{I}_{N_{\text{u}}}\right)^{-1} \bar{\mathbf{y}}_{k,l} \\ &= \mathbf{Q}_k \mathbf{Q}_k^H \left(\sum_{k' \in \mathcal{K}} \mathbf{Q}_{k'} \mathbf{Q}_{k'}^H + \frac{P}{\rho_{\text{tr}} N_{\text{u}} \sigma_{\text{h}}^2} \mathbf{I}_{N_{\text{u}}}\right)^{-1} \bar{\mathbf{y}}_{k,l}\end{aligned} \tag{D.8}$$

with its robust MSE-CE in (D.9). This concludes the proof.

References

[1] Z. Ma, Z. Zhang, Z. Ding, P. Fan, and H. Li, "Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, no. 4, pp. 1–20, Feb. 2015.

[2] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, no. 8, pp. 1–18, Aug. 2016.

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

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

[5] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.

[6] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual coupling calibration for multiuser massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 606–619, Jan. 2016.

[7] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems–Part II: Joint chunk, power and bit allocation," *IEEE Tran. Commun.*, vol. 60, no. 2, pp. 499–509, Feb. 2012.

[8] H. Zhu, "Radio resource allocation for OFDMA systems in high speed environments," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 748–759, May 2012.

[9] R. Kudo, S. Armour, J. P. McGeehan, and M. Mizoguchi, "A channel state information feedback method for massive MIMO-OFDM," *J. Commun. Netw.*, vol. 15, no. 4, pp. 352–361, Aug. 2013.

[10] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[11] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[12] B. Yang, K. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[13] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Tran. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[14] H. Wei, D. Wang, J. Wang, and X. You, "Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding," *Sci. China Inf. Sci.*, vol. 59, no. 2, pp. 1–14, Feb. 2016.

[15] ——, "TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent," *Sci. China Inf. Sci.*, vol. 59, no. 10, pp. 1–10, Oct. 2016.

[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] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access

$$\begin{aligned}\mathcal{M}_k^{\text{robust}} &= \text{tr}\left\{\mathbf{Q}_k \frac{N_\text{u}\sigma_\text{h}^2}{P}\mathbf{I}_P\mathbf{Q}_k^H\right\} - \text{tr}\left\{\mathbf{Q}_k \frac{N_\text{u}\sigma_\text{h}^2}{P}\mathbf{I}_P\mathbf{Q}_k^H\left(\sum_{k'\in\mathcal{K}}\mathbf{Q}_{k'}\frac{N_\text{u}\sigma_\text{h}^2}{P}\mathbf{I}_P\mathbf{Q}_{k'}^H + \frac{1}{\rho_\text{tr}}\mathbf{I}_{N_\text{u}}\right)^{-1}\mathbf{Q}_k\frac{N_\text{u}\sigma_\text{h}^2}{P}\mathbf{I}_P\mathbf{Q}_k^H\right\} \\ &= N_\text{u}\sigma_\text{h}^2 - \frac{N_\text{u}\sigma_\text{h}^2}{P}\text{tr}\left\{\mathbf{Q}_k\mathbf{Q}_k^H\left(\sum_{k'\in\mathcal{K}}\mathbf{Q}_{k'}\mathbf{Q}_{k'}^H + \frac{P}{\rho_\text{tr}N_\text{u}\sigma_\text{h}^2}\mathbf{I}_{N_\text{u}}\right)^{-1}\right\}.\end{aligned} \tag{D.9}$$

[19] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[20] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.

[21] Z. Chen and C. Yang, "Pilot decontamination in massive MIMO systems: Exploiting channel sparsity with pilot assignment," in *Proc. IEEE GlobalSIP*, Atlanta, GA, USA, 2014, pp. 637–641.

[22] L. You, X. Gao, A. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.

[23] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 44, no. 8, pp. 1009–1020, Aug. 1996.

[24] B. Muquet, M. De Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[25] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[26] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.

[27] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[28] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Trans. Acous., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[29] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315–325, Jan. 2004.

[30] M. C. Vanderveen, A.-J. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.

[31] J.-J. Van De Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE VTC*, Chicago, IL, 1995, pp. 815–819.

[32] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. New York, NY, USA: IEEE Press, 1994.

[33] 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.

[34] O. Edfors, M. Sandell, J.-J. Van De Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[35] G. H. Golub and C. F. Van Loan, *Matrix Computations, Fourth Edition*. Baltimore, MD: JHU Press, 2013.

[36] A. Kavčić and B. Yang, "Adaptive rank estimation for spherical subspace trackers," *IEEE Tran. Signal Process.*, vol. 44, no. 6, pp. 1573–1579, Jun. 1996.

[37] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[38] Y. Li, L. J. Cimini Jr, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[39] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[40] X. G. Doukopoulos and G. V. Moustakides, "Fast and stable subspace tracking," *IEEE Tran. Signal Process.*, vol. 56, no. 4, pp. 1452–1465, Apr. 2008.

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

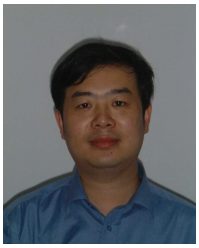
[42] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

YU ZHANG received the B.S. degree from Soochow University, Suzhou, China, in 2010 and the M.S. degree from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014. He is currently working toward the Ph.D. degree in the National Mobile Communications Research Laboratory at Southeast University, Nanjing, China. His current research interests focus on large-scale MIMO communications.

DONGMING WANG (M'06) received the B.S. degree from Chongqing University of Posts and Telecommunications, Chongqing, China, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, and the Ph.D. degree from the Southeast University, Nanjing, China, in 1999, 2002, and 2006, respectively. He joined the National Mobile Communications Research Laboratory, Southeast University, in 2006, where he has been an Associate Professor since 2010. His research interests include turbo detection, channel estimation, distributed antenna systems, and large-scale MIMO systems.

JIANGZHOU WANG (M'91-SM'94) is currently the Chair of Telecommunications and Head of Communications Research Group, School of Engineering and Digital Arts, University of Kent, Canterbury, U.K. He has authored over 200 papers in the international journals and conferences in the areas of wireless mobile communications and has written/edited three books. His research interests include wireless multiple access techniques, massive MIMO and small-cell technologies, device to device communications in cellular networks, distributed antenna systems, and cooperative communications. He is a Fellow of the IET and was an IEEE Distinguished Lecturer from January 2013 to December 2014. He serves/served as an Editor or Guest Editor for a number of international journals, such as the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was the Technical Program Chair of the IEEE WCNC2013 in Shanghai and the Executive Chair of the IEEE ICC2015 in London. He was the recipient of the Best Paper Award from the IEEE Globecom2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2634025, IEEE Access



XIAOHU YOU (F'12) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nanjing Institute of Technology, Nanjing, China, in 1982, 1985, and 1989, respectively. From 1987 to 1989, he was with Nanjing Institute of Technology as a Lecturer. From 1990 to present, he has been with the Southeast University, first as an Associate Professor and later as a Professor. His research interests include mobile communications, adaptive signal processing, and artificial neural networks with applications to communications and biomedical engineering. He is the Chief of the Technical Group of China 3G/B3G Mobile Communication R&D Project. He was the recipient of the Excellent Paper Prize from China Institute of Communications in 1987 and the Elite Outstanding Young Teacher Awards from the Southeast University, in 1990, 1991, and 1993, respectively, the 1989 Young Teacher Award of Fok Ying Tung Education Foundation, State Education Commission of China.