Channel Estimation for Multicell Multiuser Massive MIMO Uplink Over Rician Fading Channels

Liang Wu, *Member, IEEE*, Zaichen Zhang, *Senior Member, IEEE*, Jian Dang, Jiangzhou Wang, *Fellow, IEEE*, Huaping Liu, *Senior Member, IEEE*, and Yongpeng Wu, *Member, IEEE*

***Abstract*—Pilot contamination (PC) is a major problem in massive multiple-input multiple-output (MIMO) systems. This paper proposes a novel channel estimation scheme for such a system in Rician fading channels. First, the possible angle of arrivals (AOAs) of users served by a base station (BS) are derived by exploiting the channel statistical information, assuming a traditional pilot structure, where the pilots for the same-cell users are orthogonal but are identical for the same-indexed users from different cells. Although with this pilot structure the channel state information (CSI) derived contains CSI from other-cell users caused by PC, the line-of-sight (LOS) component of the desired user is PC-free when the number of antennas equipped at the BS is large. Then, based on the AOAs and the contaminated CSI, the LOS component of each user served by a BS is estimated, and data are detected by using the derived LOS components. Finally, the decoded data are used to update the CSI estimate via an iterative process. The achievable spectral efficiency of the proposed scheme is analyzed in detail, and simulation results are presented to compare the performance of the proposed scheme with that of three existing schemes.**

***Index Terms*—Massive MIMO, multicell, multiuser, pilot contamination (PC), Rician fading, channel estimation.**

I. Introduction

Massive multiple-input multiple-output (MIMO) system is a promising candidate for the fifth generation (5G) wireless mobile communications because of its potential to achieve high spectral and energy efficiencies [1]–[3]. In massive MIMO systems, the base station (BS) is equipped with a large number of antennas while the users typically can have one or only a few antennas, shifting hardware cost and computational complexity to the BS. While the increased number of BS antennas could bring additional advantages over the traditional MIMO systems [4]–[6], such as simpler detection algorithms to achieve a good performance [7], [8], new challenges arise as well. Downlink channel state information (CSI) acquisition at the BS is a challenging problem [9]. The performance of massive MIMO systems depends on the quality of the CSI acquired by the BS. Most of the theoretical performances of massive MIMO assume perfect CSI at the BS, which is unrealistic in practice [10]. In frequency division duplex (FDD) systems [11], users estimate the downlink channels using the pilots transmitted by the BS and then send the downlink CSI back to the BS. Because of the large number of BS antennas, pilot overhead and feedback overhead will be very large [12]. In time division duplex (TDD) systems, by exploiting the reciprocity of the uplink and downlink channels, the downlink CSI can be obtained by using the pilots transmitted by the users. However, because the channel coherence time could be very short for high-mobility users using higher radio frequencies (e.g., milimeter wave (mm-wave)), reducing the uplink pilot overhead is critical.

One of the popular methods to reduce uplink pilot overhead is to reuse pilots in adjacent cells, that is, the pilots of users within the same cells are orthogonal, but the pilots of different cells are the same. This causes pilot contamination (PC) [13]–[15], and the system performance could be severely affected. The spectral efficiency of massive MIMO when PC is taken into consideration is analyzed in [2], [3]. Yin *et al.* propose a coordinated approach to reduce the effect of PC by employing the second-order statistical information of the user channels [16]. A game theoretic approach to reuse the pilots for channel estimation is proposed in [17]. This scheme could achieve the same performance as the optimal pilot assignment scheme. Based on the analytic expression of the error variance of the channel estimator, Wang *et al.* develop a criterion for optimal non-orthogonal pilot signal design [18]. Farhang *et al.* exploits the inherent blind equalization property of the CMT waveform to address the PC problem in cosine modulated multitone (CMT) based massive MIMO networks [19]. A precoding scheme for downlink transmission in multicell TDD systems based on estimated CSI is proposed in [14]. The effect of PC on the physical channel models is studied in [20] and a pilot reuse strategy to reduce the pilot overhead in spatially correlated Rayleigh fading channels is proposed in [21]. A time-shifted pilot-based scheme is proposed to reduce the effect of PC by rearranging the uplink pilot transmission order for different cells, which shows that interference can be decreased significantly [10]. In [22], eigenvalue decomposition of the sample covariance matrix of the received signal is proposed to enable blind channel estimation. Another blind channel estimation algorithm is proposed in [23], which is based on spectral decomposition of the matrix formed by the received signal vectors collected within one coherence time interval of the channel. These two blind channel estimation schemes rely highly on the distinction of the eigenvalues. So far, Rayleigh fading is assumed in these works.

L. Wu, Z. Zhang and J. Dang are with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: wuliang@seu.edu.cn; zczhang@seu.edu.cn; newwanda@seu.edu.cn).

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

H. Liu is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon 97331, USA (e-mail: hliu@eecs.oregonstate.edu).

Y. Wu is with Institute for Communications Engineering, Technical University of Munich, Theresienstrasse 90, D-80333 Munich, Germany (email:yongpeng.wu2016@gmail.com)

Rician fading fits a much broader range of scenarios than Rayleigh since a line-of-sight (LOS) often exists between the transmitter and the receiver. For example, in mm-wave massive MIMO communications, the LOS component dominates the channel [24]; in small-cell networks, an LOS path often exists; in MIMO vehicular networks, where a moving vehicle communicates with either another vehicle or with the roadside, the typical channel is Rician [25]. In [26], the achievable uplink rate of multicell massive MIMO systems is analyzed assuming that the LOS component and Rician K-factor of all users served by a BS are perfectly known at both the transmitter and receiver, and an LOS path does not exist between the BS and other-cell users. In [27], a beamforming scheme and a power-scaling law for single-cell massive MIMO systems are investigated, also assuming that both the transmitter and receiver know the LOS components of all users. Li *et al.* investigate a 3-dimensional downlink beamforming algorithm for single-cell multiuser systems over Rician fading channels, and channel statistical information of each user is assumed known at the BS [28]. In [29], precoding design criteria are proposed for large-scale MIMO systems with finite alphabet inputs over Rician fading channels.

This paper deals with uplink transmissions of multiuser multicell massive MIMO systems in Rician fading channels, with a focus on developing a novel PC-resistant channel estimation scheme. In this scheme, the traditional pilot structure is employed, that is the pilots are orthogonal for all users of the same cell but are common for different cells, and the estimated CSI suffers from PC. We will first derive the possible LOS angles of arrivals (AOAs) by using the statistical information of the channels. The LOS components of the users served by a BS are then estimated by using the possible AOAs and the contaminated CSI. With the LOS component obtained, data are detected and are finally employed to update the channel estimates via an iterative process. A distinction of the work in this paper from most of the existing literature on the same topic is that the scheme is built upon a more realistic assumption: neither the transmitter nor the receiver knows the exact LOS component, the Rician K-factor, or the large-scale fading coefficients. For the proposed channel estimation scheme, a proper receiver is also developed, and the achievable spectral efficiency is analyzed. To assess the effectiveness of the proposed scheme, the achievable spectral efficiency of the proposed scheme is compared with three schemes: 1) the traditional pilot-reuse (PR) scheme, 2) the time-shifted pilot scheme, and 3) the no-PC scheme, in which the pilots of all users are mutually orthogonal. The results show that the proposed scheme achieves the highest spectral efficiency when the number of antennas equipped at the BS is large, due to its effectiveness in combating PC.

The main contributions of this work are summarized as follows.

1) A novel channel estimation method that works with common pilot structures. We exploit the property that the LOS component is not affected by PC when the number of BS antennas is large. Thus, we first propose a method that uses the channel statistical information to obtain the LOS-component accurately. The estimated LOS component is then used for channel estimation, minimizing PC effects.
2) Channel estimation algorithms. With the estimated LOS components, we develop two suitable channel estimation algorithms: LOS-component-based algorithm and data-aided iterative algorithm.
3) Rigorous analysis. The achievable spectral efficiency, power scaling, and the effect of the Rician K-factor of the proposed scheme are analyzed in detail, showing that the transmit power of each user can be reduced proportional to $1/M$ (M is the number of BS antennas).

The remainder of the paper is organized as follows. In Section II, the massive MIMO Rician fading channel model is presented. The proposed LOS component derivation scheme is presented in Section III. In Section IV, we develop the LOS component based channel estimation and the data-aided iterative channel estimation. Effects of the number of BS antennas and the Rician K-factor on the spectral efficiency is analyze din Section V. Simulation results are provided in Section VI to validate the proposed scheme, and the paper is concluded in Section VII.

II. Channel Model

Consider a network with L cells, where each BS has M linear antennas to serve K users $(K < M)$, each with one antenna, using a frequency reuse factor of 1. As in [26], it is assumed that there exists an LOS component between a BS and the users it serves, and no LOS components exist between a BS and the users of other cells. This is a reasonable assumption because the users of other cells are far away from a specific BS, and thus the probability that there exists an LOS component between the BS and the users of other cells is low. Therefore, the $M \times 1$ channel vector from the k-th user of the l-th cell to the i-th BS is expressed as [30]

$$\mathbf{h}_{(i),(l,k)} = \begin{cases} g_{(i),(i,k)} \left(\sqrt{\frac{1}{\kappa_{(i),(i,k)}+1}} \mathbf{c}_{(i),(i,k)} \right. \\ \qquad \left. + \sqrt{\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}} \bar{\mathbf{c}}_{(i),(i,k)} \right), & l = i, \\ g_{(i),(l,k)} \mathbf{c}_{(i),(l,k)}, & l \neq i, \end{cases} \tag{1}$$

where $g_{(i),(l,k)}$ is the large scale fading coefficient from the k-th user of the l-th cell to the i-th BS, $\kappa_{(i),(i,k)}$ denotes the Rician K-factor of the channel from the k-th user of the i-th cell to the i-th BS, $\mathbf{c}_{(i),(l,k)}$ is related to the non-LOS (NLOS) component and its elements are circularly symmetric complex Gaussian random variables with zero mean and unit variance, that is, $\mathbf{c}_{(i),(l,k)} \sim CN(0, \mathbf{I}_M)$, where $\mathbf{I}_M$ denotes the identity matrix of rank M and $CN(\cdot)$ denotes the complex normal distribution, and $\bar{\mathbf{c}}_{(i),(i,k)}$ is related to the LOS component from the k-th user of the i-th cell to the i-th BS. As commonly used in massive MIMO systems, it is assumed that uniform linear arrays (ULAs) are deployed at the BS. Therefore, $\bar{\mathbf{c}}_{(i),(i,k)}$ is expressed as [30]

$$\begin{aligned} \bar{\mathbf{c}}_{(i),(i,k)} &= \sqrt{M} \boldsymbol{\alpha}(\theta_{(i),(i,k)}) \\ &= [1 \quad e^{-j2\pi d \cos(\theta_{(i),(i,k)})/\lambda} \\ &\qquad \cdots \quad e^{-(M-1)j2\pi d \cos(\theta_{(i),(i,k)})/\lambda}]^T, \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha}(\cdot)$ stands for unit steering vector, $\theta_{(i),(i,k)}$ is the angle of arrival (AOA) of the LOS component from the k-th user of the i-th cell to the i-th BS, λ is the carrier wavelength, d is the distance between the adjacent antennas of the BS, and $(\cdot)^T$ stands for transpose.

The received signal of the i-th BS is

$$\begin{aligned}\mathbf{y}_{(i)} &= \sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{h}_{(i),(l,k)}\sqrt{p_t}x_{(l,k)} + \mathbf{n}_{(i)} \\ &= \sum_{l=1}^{L}\sqrt{p_t}\mathbf{H}_{(i),(l)}\mathbf{x}_{(l)} + \mathbf{n}_{(i)},\end{aligned} \tag{3}$$

where p_t is the average transmit power, $x_{(l,k)}$ is the signal transmitted by the k-th user in the l-th cell, assumed to be a random variable with zero mean and unit variance, $\mathbf{H}_{(i),(l)} = [\mathbf{h}_{(i),(l,1)} \cdots \mathbf{h}_{(i),(l,K)}]$, $\mathbf{x}_{(l)} = [x_{(l,1)}, \cdots, x_{(l,K)}]^T$, $\mathbf{n}_{(i)}$ is the noise component, whose elements are complex Gaussian random variables with zero mean and unit variance, that is, $\sigma^2_{n_{(i)}} = 1$.

In this paper, the channel statistical information including channel mean information (CMI) and channel variance information (CVI) is assumed to be known as shown in (4) at the top of next page, where $\mathrm{E}[\cdot]$ and $\mathrm{cov}(\cdot)$ stand for mean and covariance of a random vector, respectively.

III. LOS COMPONENT DERIVATION

A. AOA Estimation

In the Rician fading channel, the LOS component is related to AOA. We resort to the well-established MUSIC algorithm [31] for AOA estimation based on CVI. The procedure is as follows. Solve the eigen system

$$\mathbf{R}_{(i),(i)}\mathbf{U}_{(i)} = \mathbf{U}_{(i)}\boldsymbol{\Lambda}_{(i)}, \tag{5}$$

where $\mathbf{U}_{(i)}$ is a unitary matrix, $\boldsymbol{\Lambda} = \mathrm{diag}\{a_1, a_2, \ldots, a_M\}$, $a_1 \geq a_2 \geq \cdots \geq a_M$ are the eigenvalues of $\mathbf{R}_{(i),(i)}$, and determine the noise subspace expressed as

$$\mathbf{V}_{i,n} = [\mathbf{U}_{(i)}]_{(:,K+1:\mathrm{end})}, \tag{6}$$

where $[\cdot]_{(:,v:\mathrm{end})}$ denotes extracting the v-th column through the last column of a matrix.

Define

$$P(\phi) = \frac{\boldsymbol{\alpha}(\phi)^H\boldsymbol{\alpha}(\phi)}{\boldsymbol{\alpha}(\phi)^H\mathbf{V}_{i,n}\mathbf{V}_{i,n}{}^H\boldsymbol{\alpha}(\phi)}. \tag{7}$$

Find K peaks $(P(\phi_{i,1}),\ P(\phi_{i,2}),\ \ldots\ ,P(\phi_{i,K}))$, which are the local maxima of $P(\phi)$, and the corresponding angles $\phi_{i,1},\ \phi_{i,2},\ \ldots\ ,\phi_{i,K}$ are the estimates of the AOAs.

Note that the AOAs can be estimated by employing the MUSIC algorithm, so that the BS knows the AOAs from all users in the region but does not know which AOA corresponds to which particular user. This assumption is reasonable, because AOA estimation is based on the statistical properties of the channel as expressed in (4), which is assumed to be known.

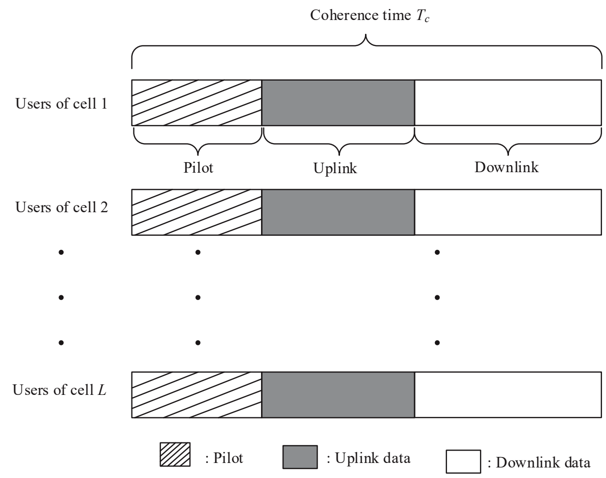

Fig. 1. Transmission frame structures of the users in different cells.

B. Traditional pilot structure

The traditional pilot structure is employed in the proposed scheme, that is the pilots of the same-cell users are orthogonal, but the same set are reused in other cells. Therefore, the estimated channel suffers from inter-cell interference. The transmission frame structures of the users in different cells are illustrated in Fig. 1. Assume that the lengths of the coherence interval is equal to T_c. The resulting pilot overhead is $K < T_c$.

Let $K \times 1$ vectors $\boldsymbol{\eta}_k (k = 1, \cdots, K)$ be the set of pilots that are orthogonal for the K users of the same cell, which satisfy

$$\boldsymbol{\eta}_k^H\boldsymbol{\eta}_j = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}. \tag{8}$$

The received signal of the i-th BS during the pilot period is expressed as

$$\mathbf{B}_i = \sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{h}_{(i),(l,k)}\sqrt{p_u}\sqrt{K}\boldsymbol{\eta}_k^H + \mathbf{N}_i, \tag{9}$$

where p_u denotes the average transmit power of the uplink pilot symbols, and $\mathbf{N}_i$ is an $M \times K$ white noise matrix whose elements are Gaussian random variables with zero mean and unit variance. With a least-squares (LS) estimator for this pilot structure, the estimated CSI (contaminated) from the k-th user of i-th cell to the i-th BS can be expressed as [2]

$$\tilde{\mathbf{h}}_{(i),(i,k)} = \sum_{l=1}^{L}\mathbf{h}_{(i),(l,k)} + \frac{\mathbf{N}_i\boldsymbol{\eta}_k}{\sqrt{Kp_u}}. \tag{10}$$

The LOS components will be determined by using the property of the contaminated CSI and the possible AOAs in the section next.

C. Derivation of LOS components

Note that $\bar{\mathbf{c}}_{(i),(i,k)}$ appears in the contaminated CSI $\tilde{\mathbf{h}}_{(i),(i,k)}$ expressed in (10). We develop the following procedure to derive the LOS component corresponding to a specific served user using the contaminated CSI $\tilde{\mathbf{h}}_{(i),(i,k)}$) and the possible AOAs:

$$\begin{cases} \mathbf{q}_{(i)} = \mathrm{E}[\sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{h}_{(i),(l,k)}] = \sum_{k=1}^{K} g_{(i),(i,k)}\sqrt{\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}}\bar{\mathbf{c}}_{(i),(i,k)} \\ \mathbf{R}_{(i),(i)} = \mathrm{cov}(\sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{h}_{(i),(l,k)}) = \sum_{k=1}^{K} g^2_{(i),(i,k)}\left(\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}\bar{\mathbf{c}}_{(i),(i,k)}\bar{\mathbf{c}}^H_{(i),(i,k)} + \frac{1}{\kappa_{(i),(i,k)}+1}\mathbf{I}_M\right) + \sum_{l=1,l\neq i}^{L}\sum_{k=1}^{K} g^2_{(i),(l,k)}\mathbf{I}_M \end{cases} \quad (4)$$

Step 1) Form the steering vector as $\boldsymbol{\alpha}(\phi_{i,v})$ from the estimated AOAs ($\phi_{i,v}, v = 1, \cdots, K,$).

Step 2) Project $\tilde{\mathbf{h}}_{(i),(i,k)}$ onto $\boldsymbol{\alpha}(\phi_{i,v}), v = 1, \cdots, K$, as

$$\begin{aligned} m_{(i),(:,k),v} &= \boldsymbol{\alpha}(\phi_{i,v})^H \tilde{\mathbf{h}}_{(i),(i,k)} \\ &= \boldsymbol{\alpha}(\phi_{i,v})^H (\sum_{l=1}^{L}\mathbf{h}_{(i),(l,k)} + \frac{\mathbf{N}_i \boldsymbol{\eta}_k}{\sqrt{K p_u}}) \\ &= \sum_{l=1}^{L} m_{(i),(l,k),v} + \boldsymbol{\alpha}(\phi_{i,v})^H \frac{\mathbf{N}_i \boldsymbol{\eta}_k}{\sqrt{K p_u}}, \end{aligned} \quad (11)$$

where $m_{(i),(l,k),v} = \boldsymbol{\alpha}(\phi_{i,v})^H \mathbf{h}_{(i),(l,k)}$.

Step 3) Decision:

Theorem 1: Define

$$v_{\text{opt},(i,k)} = \arg\max_{v} \left\{ \Phi = m_{(i),(:,k),v} \right\}. \quad (12)$$

The angle of the k-th user in the i-th cell is $\phi_{i,v_{\text{opt},(i,k)}}$.

The proof of *Theorem* 1 is provided in Appendix A.

To ensure the derived angles of the served users do not overlap with one another, the procedure to determine the angles of the K users served by a BS is revised and listed in Table I.

TABLE I
THE PROCEDURE TO DETERMINE THE ANGLES OF THE K USERS SERVED BY THE i-TH BS.

Step 1:	From the estimated AOAs ($\phi_{i,v}, v = 1, \cdots, K,$), form the steering matrix as $\boldsymbol{\Omega} = [\boldsymbol{\alpha}(\phi_{i,1}), \boldsymbol{\alpha}(\phi_{i,2}), \cdots, \boldsymbol{\alpha}(\phi_{i,K})]$
Step 2:	Calculate $\boldsymbol{\Xi}_{(i),(i)} = \boldsymbol{\Omega}^H [\tilde{\mathbf{h}}_{(i),(i,1)}, \tilde{\mathbf{h}}_{(i),(i,2)}, \cdots, \tilde{\mathbf{h}}_{(i),(i,K)}]$
Step 3:	**For** $k = 1 : K$ Search the maximal value of $\boldsymbol{\Xi}_{(i),(i)}$, and the index of the maximal value is ($r_{\max}$, $c_{\max}$); The angle for the $c_{\max}$-th user of the i-th cell is $\phi_{i,r_{\max}}$; Set $[\boldsymbol{\Xi}_{(i),(i)}]_{(r_{\max},:)} = -\infty$ and $[\boldsymbol{\Xi}_{(i),(i)}]_{(:,c_{\max})} = -\infty$; where $[\cdot]_{(a,:)}$ denotes the a-th row of a matrix, and $[\cdot]_{(:,b)}$ denotes the b-th column of a matrix. **end**

With the estimated AOA, the unit steering vector of the k-th user of the i-th cell is $\boldsymbol{\alpha}(\phi_{i,v_{\text{opt},(i,k)}})$. The weighting coefficient for the LOS component of the channel between the k-th user of the i-th cell and the i-th BS can be derived given the CMI as

$$\beta_{(i),(i,k)} = \boldsymbol{\alpha}(\phi_{i,v_{\text{opt},(i,k)}})^H \mathbf{q}_{(i)}. \quad (13)$$

The estimated LOS component of the channel from the k-th user of the i-th cell to the i-th BS is expressed as

$$\mathbf{f}_{(i),(i,k)} = \beta_{(i),(i,k)} \boldsymbol{\alpha}(\phi_{i,v_{\text{opt},(i,k)}}). \quad (14)$$

Define the estimated LOS component matrix of the i-th BS as

$$\mathbf{F}_{(i),(i)} = [\mathbf{f}_{(i),(i,1)}, \mathbf{f}_{(i),(i,2)}, \cdots, \mathbf{f}_{(i),(i,K)}]. \quad (15)$$

The principle of the proposed LOS-component-derivation scheme is summarized as follows. In the Rician fading channel, the LOS component exists, and it is related to the AOA. The AOAs of the NLOS components are uniformly distributed over $[0, 2\pi)$. As the number of the antennas at the BS increases, the angular resolution increases, and the LOS AOAs of different users will not overlap with each other, which means that the LOS components of different users will not contaminate one another even when the same pilot is used. The possible LOS AOAs (the LOS AOAs of all served users) can be estimated according to CVI, but the BS does not know which AOA corresponds to a specific user. The LOS component of the k-th user of the i-th cell contains $\tilde{\mathbf{h}}_{(i),(i,k)}$ as shown (10). Therefore, we project $\tilde{\mathbf{h}}_{(i),(i,k)}$ onto the steering vectors corresponding to the possible LOS AOAs, as shown in (11). If the projection coefficient Φ is maximized as shown in (12), we can determine the LOS AOA of the k-th user in the i-th cell. The weighting coefficient corresponding the LOS component can be derived according to CMI as shown in (13).

Note that even when the AOAs of different users are the same, the proposed LOS component derivation scheme can can still work well. If the derived AOAs of the served users overlap with each other, the derived LOS component of a specific served user will be interfered by other users, and the performance of the proposed scheme will be degraded. The proposed AOA-derivation algorithm as shown in Table I ensures that the derived angles of the served users DO not overlap with one another. Besides, when users are uniformly distributed, the probability that the AOAs of different users are the same will be low.

IV. PILOT-CONTAMINATION-RESISTANT CHANNEL ESTIMATION

With the LOS component determined using the scheme presented in Sec. III, we propose a PC-resistant channel estimation scheme next.

A. Channel estimation by using the LOS component

The estimated LOS component matrix can be viewed as a form of estimated channel, which is accurate when the Rician K-factor is large. Therefore, the estimated channel is expressed as

$$\hat{\mathbf{H}}_{(i),(i)} = \mathbf{F}_{(i),(i)}. \quad (16)$$

Define

$$\mathbf{W} = \hat{\mathbf{H}}_{(i),(i)} - \mathbf{H}_{(i),(i)}. \tag{17}$$

If the LOS components are estimated correctly, then the k-th column of $\mathbf{W}$ are random variables with zero mean and variances $\frac{g^2_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}$. The received signal can be rewritten as

$$\mathbf{y}_{(i)} = \sqrt{p_t}\hat{\mathbf{H}}_{(i),(i)}\mathbf{x}_{(i)} - \sqrt{p_t}\mathbf{W}\mathbf{x}_{(i)} + \sum_{l=1,l\neq i}^{L} \sqrt{p_t}\mathbf{H}_{(i),(l)}\mathbf{x}_{(l)} + \mathbf{n}_{(i)}. \tag{18}$$

In massive MIMO systems, zero-forcing (ZF) detection and minimum mean square error (MMSE) detection achieve almost the same performance at moderate signal-to-noise ratio (SNR) values [3]. Therefore, the ZF based detection algorithm is employed in the proposed scheme because of its simplicity. With the estimated LOS component matrix $\mathbf{F}_{(i),(i)}$, the estimate of $\mathbf{x}_{(i)}$ is expressed as [34]

$$\hat{\mathbf{x}}_{(i)} = \left(\hat{\mathbf{H}}^H_{(i),(i)}\hat{\mathbf{H}}_{(i),(i)}\right)^{-1}\hat{\mathbf{H}}^H_{(i),(i)}\mathbf{y}_{(i)}/\sqrt{p_t}. \tag{19}$$

By defining

$$\mathbf{d}_{(i,k)} = \left[\hat{\mathbf{H}}_{(i),(i)}(\hat{\mathbf{H}}^H_{(i),(i)}\hat{\mathbf{H}}_{(i),(i)})^{-1}\right]_{(:,k)}, \tag{20}$$

where $[\cdot]_{(:,k)}$ denotes extracting the k-th column of a matrix, the estimate of $x_{(i,k)}$ can be re-written as (21) at the top of next page, where $n_{(i,k)} = \mathbf{d}_{(i,k)}{}^H\mathbf{n}_{(i)}/\sqrt{p_t}$ and has a variance equal to $\sigma^2_{n_{(i,k)}}$.

The signal-to-interference-plus-noise ratio (SINR) is expressed as (22) at the top of next page.

When the LOS-component-based channel estimation and the ZF receiver are employed in the uplink transmission, the achievable spectral efficiency of the k-th user in the i-th cell is

$$R^{\text{ZF}}_{(i,k)} = \frac{T_c - K}{T_c}\gamma \text{E}\left[\log_2(1 + \text{SINR}^{\text{ZF}}_{(i,k)})\right], \tag{23}$$

where $T_c - K$ is the effective transmission interval, and $\gamma \in (0,1)$ is the portion of the effective transmission interval for the uplink data transmission. By using Jensen's inequality [3], the lower bound of the achievable spectral efficiency can be expressed as (24) at the top of next page. Note that in (24), $\hat{\mathbf{H}}_{(i),(i)}$ is based on the LOS component matrix and is a slowly changing parameter.

B. Data-aided iterative channel estimation

The quality of the channel estimate can be improved by exploiting the initial detected data. This process is discussed next. In the τ ($\tau \geq KL$) intervals the received signals can be expressed as

$$\mathbf{Y}_{(i)} = \sqrt{p_t}\mathbf{H}_{(i),(i)}\mathbf{X}_{(i)} + \sum_{l=1,l\neq i}^{L} \sqrt{p_t}\mathbf{H}_{(i),(l)}\mathbf{X}_{(l)} + \mathbf{D}_{(i)}, \tag{25}$$

where $\mathbf{X}_{(i)}$ is the $K \times \tau$ data matrix, and $\mathbf{D}_{(i)}$ is the $K \times \tau$ noise matrix, and the elements of $\mathbf{D}_{(i)}$ are zero mean and unit variance complex Gaussian random variables. The data $\mathbf{X}_{(i)}$ can be estimated by using (19). Let $\hat{\mathbf{X}}_{(i)}$ denote the estimate of $\mathbf{X}_{(i)}$ and define

$$\mathbf{Z}_{(i)} = \hat{\mathbf{X}}_{(i)} - \mathbf{X}_{(i)}. \tag{26}$$

Each column of $\mathbf{Z}_{(i)}$ is a Gaussian random vector with zero mean and covariance matrix

$$\text{cov}(\mathbf{Z}_{(i)}) = \left(\sum_{k=1}^{K}\frac{p_t g^2_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1} + \sum_{l=1,l\neq i}^{L}\sum_{k=1}^{K} p_t g^2_{(i),(l,k)} + 1\right) \cdot \left(\hat{\mathbf{H}}^H_{(i),(i)}\hat{\mathbf{H}}_{(i),(i)}\right)^{-1}. \tag{27}$$

Based on the decoded data, the NLOS component of $\mathbf{H}_{(i),(i)}$, $\mathbf{H}^{\text{NLOS}}_{(i),(i)}$, is estimated as

$$\begin{aligned}\hat{\mathbf{H}}^{\text{NLOS}}_{(i),(i)} &= (\mathbf{Y}_{(i)} - \mathbf{F}_{(i),(i)}\sqrt{p_t}\hat{\mathbf{X}}_{(i)})\hat{\mathbf{X}}^H_{(i)}/(\tau\sqrt{p_t}) \\ &= \mathbf{H}^{\text{NLOS}}_{(i),(i)} - \mathbf{H}_{(i),(i)}\mathbf{Z}_{(i)}\hat{\mathbf{X}}^H_{(i)}/\tau \\ &\quad + \sum_{l=1,l\neq i}^{L}\mathbf{H}_{(i),(l)}\mathbf{X}_{(l)}\hat{\mathbf{X}}^H_{(i)})/\tau + \mathbf{D}_{(i)}\mathbf{X}^H_{(i)}/(\tau\sqrt{p_t}) \\ &\approx \mathbf{H}^{\text{NLOS}}_{(i),(i)} - \mathbf{H}_{(i),(i)}\mathbf{Z}_{(i)}\hat{\mathbf{X}}^H_{(i)}/\tau \end{aligned} \tag{28}$$

where the approximation becomes more accurate as τ increases. The estimated channel after employing the data detected is expressed as

$$\hat{\mathbf{H}}^f_{(i),(i)} = \mathbf{F}_{(i),(i)} + \hat{\mathbf{H}}^{\text{NLOS}}_{(i),(i)}. \tag{29}$$

Eq. (28) shows that if more accurate decoded data $\hat{\mathbf{X}}_{(i)}$ will result in a more accurate updated NLOS component estimate $\hat{\mathbf{H}}^{\text{NLOS}}_{(i),(i)}$. Therefore, the NLOS channel can be updated iteratively. The proposed data-aided iterative channel estimation algorithm is shown in Table II.

TABLE II
PROPOSED DATA-AIDED ITERATIVE CHANNEL ESTIMATION ALGORITHM.

1) **Initialization**: Obtain the estimate of $\mathbf{H}_{(i),(i)}$, denoted by $\hat{\mathbf{H}}_{(i),(i)}$ based on (16).
2) **Iteration**:
   i) Carry out data detection according to (19) and (25), and derive the estimate of $\mathbf{X}_{(i)}$, denoted by $\hat{\mathbf{X}}_{(i),(i)}$;
   ii) Update the estimate of the NLOS channel component $\hat{\mathbf{H}}^{\text{NLOS}}_{(i),(i)}$ by using (28);
   iii) Update the estimated channel by using (29). If $\hat{\mathbf{H}}^f_{(i),(i)}$ is the same as its previous estimate, or the iteration has reached a pre-determined limit, then go to 3); otherwise go to i).
3) Derive the estimate of the channel matrix using (29).

The SINR and achievable spectral efficiency based on the iteratively estimated CSI and the ZF receiver can also be expressed as (22) and (23), respectively, where

$$\mathbf{d}_{(i,k)} = \left[\hat{\mathbf{H}}^f_{(i),(i)}\left((\hat{\mathbf{H}}^f_{(i),(i)})^H\hat{\mathbf{H}}^f_{(i),(i)}\right)^{-1}\right]_{(:,k)}. \tag{30}$$

A general lower-bound of the achievable rate with the iteratively estimated CSI is difficult derive analytically. We analyze two extreme cases:

Case I: The decoded data are completely wrong. In this case, the the estimated NLOS component $\hat{\mathbf{H}}^{\text{NLOS}}_{(i),(i)} \approx 0$, and

$$\hat{x}_{(i,k)} = \mathbf{d}_{(i,k)}{}^{H}\mathbf{y}_{(i)}/\sqrt{p_t}$$
$$= \mathbf{d}_{(i,k)}{}^{H}\mathbf{f}_{(i),(i,k)}x_{(i,k)} + \sum_{j=1,j\neq k}^{K}\mathbf{d}_{(i,k)}{}^{H}\mathbf{f}_{(i),(i,j)}x_{(i,j)} - \mathbf{d}_{(i,k)}{}^{H}\mathbf{W}\mathbf{x}_{(i)} + \sum_{l=1,l\neq i}^{L}\mathbf{d}_{(i,k)}{}^{H}\mathbf{H}_{(i),(l)}\mathbf{x}_{(l)} + \mathbf{d}_{(i,k)}{}^{H}\mathbf{n}_{(i)}/\sqrt{p_t}$$
$$= x_{(i,k)} - \mathbf{d}_{(i,k)}{}^{H}\mathbf{W}\mathbf{x}_{(i)} + \sum_{l=1,l\neq i}^{L}\mathbf{d}_{(i,k)}{}^{H}\mathbf{H}_{(i),(l)}\mathbf{x}_{(l)} + n_{(i,k)} \tag{21}$$

$$\mathrm{SINR}_{(i,k)}^{\mathrm{ZF}} = \frac{p_t}{\left(p_t\left\|\mathbf{d}_{(i,k)}{}^{H}\mathbf{W}\right\|^2 + \sum\limits_{l=1,l\neq i}^{L}\sum\limits_{k=1}^{K}p_t\|\mathbf{d}_{(i,k)}{}^{H}\mathbf{H}_{(i),(l)}\|^2 + \left\|\mathbf{d}_{(i,k)}\right\|^2\right)} \tag{22}$$

$$R_{(i,k)}^{\mathrm{ZF}} \geq \frac{T_c-K}{T_c}\gamma\log_2\left(1+E\left[\mathrm{SINR}_{(i,k)}^{\mathrm{ZF}}\right]\right)$$
$$= \frac{T_c-K}{T_c}\gamma\log_2\left(1+\frac{p_t}{\left(\sum\limits_{k=1}^{K}\frac{p_t g_{(i),(i,k)}^2}{\kappa_{(i),(i,k)}+1} + \sum\limits_{l=1,l\neq i}^{L}\sum\limits_{k=1}^{K}p_t g_{(i),(l,k)}^2 + 1\right)\left[\left(\hat{\mathbf{H}}_{(i),(i)}^{H}\hat{\mathbf{H}}_{(i),(i)}\right)^{-1}\right]_{k,k}}\right). \tag{24}$$

the lower bound of the achievable spectral efficiency are given by (24).

Case II: The decoded data are perfect, that is,

$$\hat{\mathbf{H}}_{(i),(i)}^{f} \approx \mathbf{H}_{(i),(i)}. \tag{31}$$

The lower bound of achievable spectral efficiency for this case is given by (32) at the top of next page.

V. Asymptotic analysis

In this section, we analyze the effects of the number of BS antennas and the Rician K-factor on the lower bound of the Ergodic achievable spectral efficiency per user when the LOS component based channel estimation is used. These results and conclusions drawn will be similar when the iteratively estimated channel is used.

By fixing the transmit power p_t and the number of BS antennas M, it can be shown that in the special case of $\kappa_{(i),(i,k)} \to \infty$ ($k = 1,\cdots,K$) the lower bound in (24) reduce to (33) at the top of next page. When $\kappa_{(i),(i,k)} \to 0$ ($k = 1,\cdots,K$), the lower bound is equal to 0, and the performance will degrade.

Lemma 1: When M is large, we have

$$\lim_{M\to\infty}\boldsymbol{\alpha}(\phi_n)^H\boldsymbol{\alpha}(\phi_m) = 0 \quad \text{if} \quad |\phi_n-\phi_m| \geq \theta_{\min}, \tag{34}$$

where $\theta_{\min}$ is the minimum distinguishable angle, which can be calculated by using the half power beamwidth (HPBW) [36].

The proof of *Lemma* 1 is provided in Appendix B.

Lemma 1 shows that the probability that the columns of $\hat{\mathbf{H}}_{(i),(i)}$ are mutually orthogonal will be high when the number of antenna M is large. Therefore, as $M\to\infty$, (24) can be rewritten as (35) given at the top of next page.

With a transmit power $p_t = \frac{E_t}{M}$ and when the Rician K-factors are fixed, we have

$$\lim_{M\to\infty}\left\{R_{(i,k)}^{\mathrm{ZF}}\right\}_{LB} = \frac{T_c-K}{T_c} \cdot\gamma\log_2\left(1+\frac{\kappa_{(i),(i,k)}g_{(i),(i,k)}^2}{\kappa_{(i),(i,k)}+1}E_t\right). \tag{36}$$

where E_t is fixed. Therefore, in multicell multiuser Rician fading channels, when the number of BS antennas M grows to infinity, the transmit power of each user can be reduced proportionally to $1/M$ when the proposed channel estimation scheme is applied.

VI. Simulation Results

Here is the configuration for the simulation: there are $L=7$ BSs and frequency reuse factor is 1, that is, a given cell will be interfered by its 6 neighboring cells. In each cell, there are $K=10$ users. All direct gains are normalized to 1, that is, $g_{(i),(i,k)} = 1$. The cross-gains are uniformly distributed over $(0,1]$, which means that there are interfering users at the cell edge as well as far from the cell. The users in the same cell are assumed to have the same Rician K-factor, and $\kappa_{(i),(l,k)} = 0$ ($l \neq i$, $k = 1,2,\cdots,K$), that is, fading cross different cells is modeled as Rayleigh [26]. The LOS AOAs of the users are uniformly distributed over $[-\pi/2,\pi/2]$. The spacing between two adjacent antennas is $d = \frac{1}{2}\lambda$, and $\gamma = 0.5$; that is, the uplink occupies the half of the effective transmission interval. The spectral efficiency is defined as

$$R_i^{\mathrm{ZF}} = \sum_{k=1}^{K}R_{(i,k)}^{\mathrm{ZF}}. \tag{37}$$

$$R_{(i,k)}^{\mathrm{ZF},f} \geq \frac{T_c - K}{T_c}\gamma\log_2\left(1+\frac{p_t}{\left(\sum\limits_{k=1}^{K}\frac{p_t g_{(i),(i,k)}^2}{\kappa_{(i),(i,k)}+1}+\sum\limits_{l=1,l\neq i}^{L}\sum\limits_{k=1}^{K}p_t g_{(i),(l,k)}^2+1\right)\mathrm{E}\left\{\left[\left(\mathbf{H}_{(i),(i)}^H\mathbf{H}_{(i),(i)}\right)^{-1}\right]_{k,k}\right\}}\right) \tag{32}$$

$$\lim_{\kappa_{(i),(i,k)}\to\infty}\left\{R_{(i,k)}^{\mathrm{ZF}}\right\}_{LB} = \frac{T_c - K}{T_c}\gamma\log_2\left(1+\frac{g_{(i),(i,k)}^2 M p_t}{\left(\sum\limits_{l=1,l\neq i}^{L}\sum\limits_{k=1}^{K}p_t g_{(i),(l,k)}^2+1\right)}\right) \tag{33}$$

$$R_{(i,k)}^{\mathrm{ZF}} \geq \frac{T_c - K}{T_c}\gamma\log_2\left(1+\frac{\frac{\kappa_{(i),(i,k)}g_{(i),(i,k)}^2}{\kappa_{(i),(i,k)}+1}Mp_t}{\left(\sum\limits_{k=1}^{K}\frac{p_t g_{(i),(i,k)}^2}{\kappa_{(i),(i,k)}+1}+\sum\limits_{l=1,l\neq i}^{L}\sum\limits_{k=1}^{K}p_t g_{(i),(l,k)}^2+1\right)}\right). \tag{35}$$

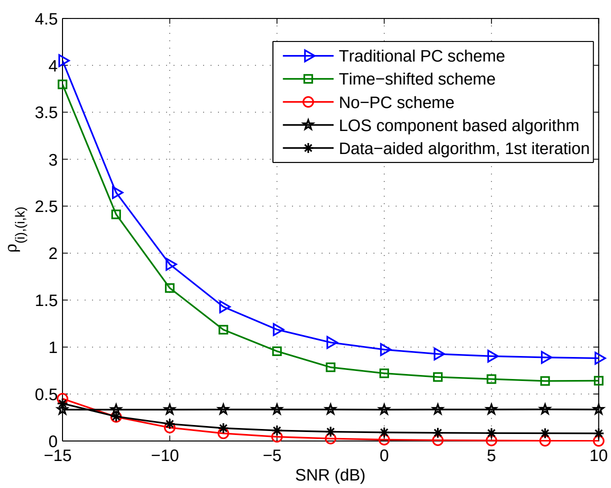

Fig. 2. NMSE of channel estimation versus SNR, when $M = 200$ and $\kappa_{(i),(i,k)} = 3$dB.

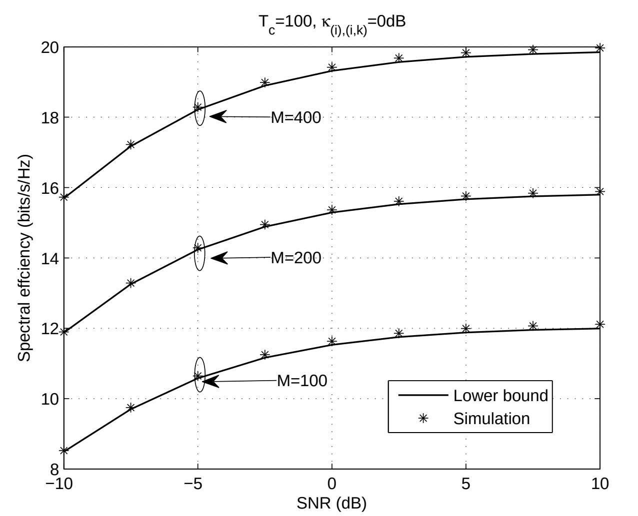

Fig. 3. Numerically calculated and simulated lower bounds of the spectral efficiency with the LOS component based scheme ($T_c = 100$ and $\kappa_{(i),(i,k)} = 0$dB).

In the uplink transmissions, SNR is defined as $\mathrm{SNR} = p_t/\sigma_{n_{(i)}}^2 = p_t$, and $p_u = p_t$. The normalized mean square error (NMSE) is employed to evaluate the proposed channel estimation performance, and it is defined as

$$\begin{aligned}\rho_{(i),(i,k)} &= \frac{\mathrm{power}(\mathrm{Interference}+\mathrm{noise})}{\mathrm{power}(\mathrm{Desired\ component})}\\ &= \frac{\|\hat{\mathbf{h}}_{(i),(i,k)}-\mathbf{h}_{(i),(i,k)}\|_2^2}{\|\mathbf{h}_{(i),(i,k)}\|_2^2}.\end{aligned} \tag{38}$$

where $\hat{\mathbf{h}}_{(i),(i,k)}$ is the estimation of $\mathbf{h}_{(i),(i,k)}$.

When $M = 200$ and $\kappa_{(i),(i,k)} = 3$dB, Fig. 2 shows the NMSE of different schemes including, the proposed scheme, the traditional PR scheme, the no-PC scheme, and the time-shifted pilot-based scheme. In the traditional PR scheme, the pilots of different users in the same cell are orthogonal and the pilots are reused in other cells, the length of the pilot is K, and no other processing is applied. For the no-PC scheme, the pilots of all the users in all the cells are mutually orthogonal, resulting in a pilot overhead of KL, and the estimated channel will not suffer from PC. In the time-shifted pilot-based scheme [10], pilots are transmitted at non-overlapping times in each cell, and the pilot overhead is K. For example, when the uplink pilots are transmitted in a specific cell, and other cells are transmitting downlink data symbols. For the time-shifted scheme, the large scale fading coefficient from the l-th BS to the i-th BS $g_{i,l}$ is assumed to be uniformly distributed over $(0.1, 0.5)$, and it is time invariant. The cell group strategy is one cell per group, and the normalized transmit power of each BS is ten times of the normalized pilot power [8].

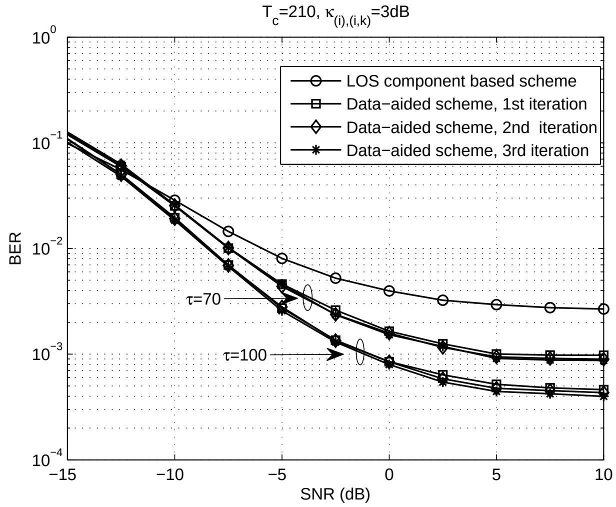

Fig. 4. Average BER versus SNR with the proposed channel estimation scheme ($T_c = 210$, $M = 100$, $\kappa_{(i),(i,k)} = 3$dB).

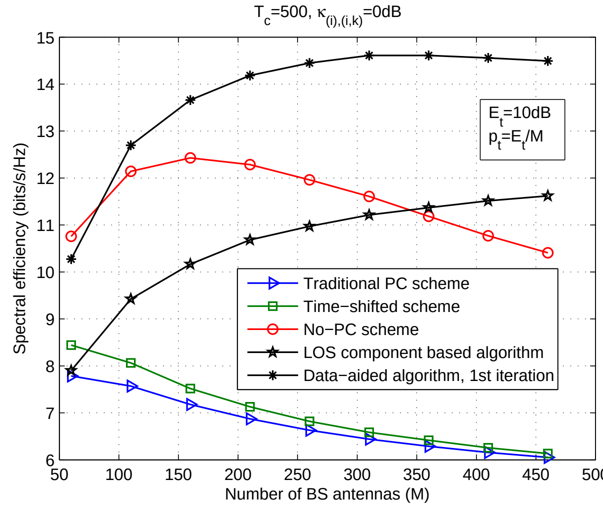

Fig. 6. Comparison of spectral efficiency versus the number of BS antennas M with various schemes (the reference transmit power is $E_t = 10$dB, $T_c = 500$, and $\kappa_{(i),(i,k)} = 0$dB).

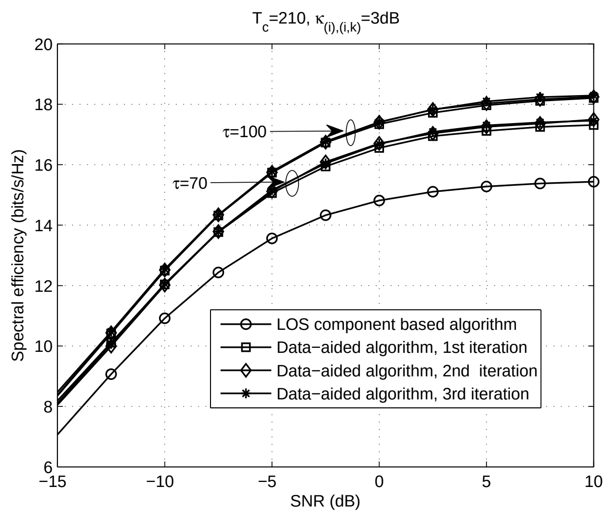

Fig. 5. Achievable spectral efficiency versus SNR with the proposed channel estimation scheme ($T_c = 210$, $M = 100$, $\kappa_{(i),(i,k)} = 3$dB).

Fig. 2 shows that the proposed scheme achieves lower NMSE than the traditional PR scheme and the time-shifted pilot-based scheme, which means that proposed scheme can efficiently reduce the pilot interference. The NMSE of the proposed LOS-component based algorithm is slightly affected by SNR. That is because the estimated channel of the LOS-component based algorithm is based on the estimated AOA, which is slightly affected by SNR. Besides, the proposed data-aided algorithm, where $\tau = 100$, achieves better NMSE performance than the LOS-component based algorithm.

Fig. 3 shows the simulated spectral efficiency and the proposed analytical lower bounds for the ZF receiver using the channel estimated via the LOS-component-based algorithm. Here $T_c = 100$, $\kappa_{(i),(i,k)} = 0$dB, and the number of BS antennas, M, is 100, 200, and 400, respectively. Very tight bounds are observed from this figure.

Fig. 4 shows the bit error rate (BER) performance of the proposed scheme with 4-ary quadrature amplitude modulation (QAM) assuming $T_c = 210$, $M = 100$, $\kappa_{(i),(i,k)} = 3$dB. It is found that the decoded-data-aided algorithm could significantly improve the BER performance, even with only one iteration. The BER performance improves as τ increases.

Fig. 5 shows the achievable spectral efficiency. Similar to BER performance, the achievable spectral efficiency with the data-aided algorithm is significantly higher than with the LOS-component-based algorithm in this condition.

Fig. 6 shows the spectral efficiency of the uplink transmission versus the number of BS antennas, M, for $p_t = E_t/M$. Other parameters are: $E_t = 10$dB, $T_c = 500$, and $\kappa_{(i),(i,k)} = 0$dB. The traditional PR scheme, the no-PC scheme, and the time-shifted pilot scheme are adopted for performance comparison. MMSE detection [34] is used in these three schemes. For the proposed data-aided algorithm, $\tau = 200$. It is observed that with $p_t = E_t/M$, the spectral efficiency of the proposed algorithm approaches a constant value as M increases, but decreases 0 with the three existing schemes. This shows that with the proposed algorithm, the transmit power of each user as E_t/M can be scaled down proportionally to $1/M$. The proposed data-aided algorithm achieves a higher spectral efficiency than the LOS-component-based algorithm.

Fig. 7 shows the spectral efficiencies of different schemes assuming $p_t = E_t/M$, $T_c = 100$, $E_t = 10$dB, and $\kappa_{(i),(i,k)} = 0$dB. The proposed data-aided algorithm cannot be applied in this case, because τ should be greater than $KL = 70$ and only 45 time intervals are used for uplink data transmission. The proposed LOS-component-based algorithm achieves the highest spectral efficiency among the four schemes compared when the number of BS antennas $M > 70$. The same power scaling law can be derived as in Fig. 7.

The achievable spectral efficiency versus the Rician K-factor ($\kappa_{(i),(i,k)}$) is plotted in Fig. 8, where $\mathrm{SNR} = 10$dB, $T_c = 210$, and $\tau = 100$. As $\kappa_{(i),(i,k)}$ increases, the achievable spectral efficiency of the proposed scheme increases, because

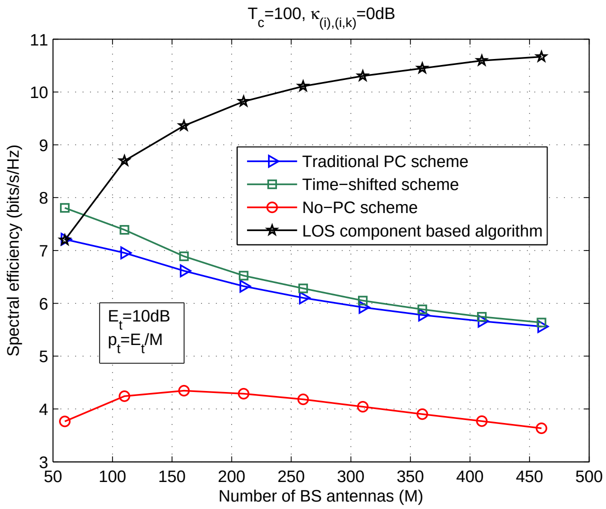

Fig. 7. Comparison of spectral efficiency versus the number of BS antennas M with various schemes (the reference transmit power is $E_t = 10$dB, $T_c = 100$, and $\kappa_{(i),(i,k)} = 0$dB).

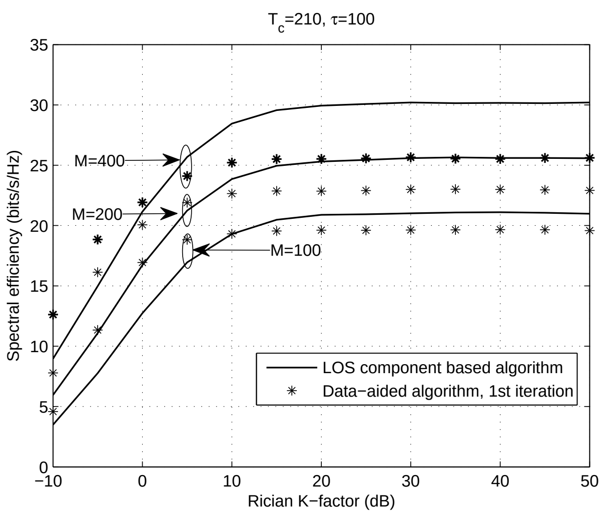

Fig. 8. Achievable spectral efficiency versus $\kappa_{(i),(i,k)}$ in the uplink data transmission (SNR=10dB, and $T_c = 210$).

as $\kappa_{(i),(i,k)}$ increases, more energy of the desired signals can be collected from the LOS component. When the Rician K-factor is low, the data-aided algorithm achieves a higher spectral efficiency than the LOS-component-based algorithm, because when the Rician K-factor is low, the LOS component will not dominate the channel, and the data-aided algorithm can improve the accuracy of the estimated channel. When the Rician K-factor is high, the LOS-component-based algorithm achieves a higher spectral efficiency, because when the Rician K-factor is high ($>$ 10dB), the LOS component dominates the channel, and the estimate of the LOS component is sufficiently accurate, while the data-aided algorithm might make the channel estimation less accurate due to the NLOS component estimation.

VII. Conclusion

In this paper, we have proposed channel estimation algorithms for multicell multiuser massive MIMO uplink in Rician fading channels based on the channel statistical information and the contaminated CSI. In coherence-time-limited systems, for example, vehicular networks, the achievable spectral efficiency of the proposed scheme is higher than those of the traditional PR scheme, the no-PC scheme, and the time-shifted pilot scheme. The achievable spectral efficiency and power scaling law of the proposed scheme are analyzed. The proposed data-aided algorithm achieves a higher spectral efficiency than the LOS-component-based algorithm in weak Rician channels, and the LOS-component-based algorithm achieves a higher spectral efficiency in strong Rician channels. Simulation results show that the proposed scheme could combat PC efficiently, and is a an excellent candidate technique for massive MIMO systems in Rician fading environments.

Appendix A
Proof of Theorem 1

The estimated AOAs are dominated by the LOS components of different users' channels. The NLOS components are uniformly distributed over $[0, 2\pi)$. Therefore, when M we have

$$\begin{aligned}
& m_{(i),(:,k),v} \\
&= \left\{ \sum_{l=1}^{L} \boldsymbol{\alpha}(\phi_{i,v})^H \mathbf{h}_{(i),(l,k)} + \boldsymbol{\alpha}(\phi_{i,v})^H \frac{\mathbf{N}_i \boldsymbol{\eta}_k}{\sqrt{K p_u}} \right\} \\
&= g_{(i),(i,k)} \sqrt{\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}} \boldsymbol{\alpha}(\phi_{i,v})^H \boldsymbol{\alpha}(\theta_{(i),(i,k)}) \\
&\quad + g_{(i),(i,k)} \sqrt{\frac{1}{\kappa_{(i),(i,k)}+1}} \boldsymbol{\alpha}(\phi_{i,v})^H \mathbf{c}_{(i),(i,k)} \\
&\quad + \sum_{l=1, l\neq i}^{L} g_{(i),(l,k)} \boldsymbol{\alpha}(\phi_{i,v})^H \mathbf{c}_{(i),(l,k)} + \boldsymbol{\alpha}(\phi_{i,v})^H \frac{\mathbf{N}_i \boldsymbol{\eta}_k}{\sqrt{K p_u}} \\
&\approx g_{(i),(i,k)} \sqrt{\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}} \boldsymbol{\alpha}(\phi_{i,v})^H \boldsymbol{\alpha}(\theta_{(i),(i,k)}) \\
&\leq g_{(i),(i,k)} \sqrt{\frac{\kappa_{(i),(i,k)}}{\kappa_{(i),(i,k)}+1}} \qquad \text{(A.1)}
\end{aligned}$$

where the approximation holds because $\mathbf{c}_{(i),(l,k)}$ and $\mathbf{N}_i \boldsymbol{\eta}_k$ are zero-mean Gaussian random vectors, and the projection to the steering vector $\boldsymbol{\alpha}(\phi_{i,v})$ will be very small. The last equality holds only when $\phi_{i,v}$ is equal to $\theta_{(i),(i,k)}$. Therefore, when Φ is maximized, the corresponding angle is $\theta_{(i),(i,k)}$.

Appendix B
Proof of Lemma 1

The HPBW can be viewed as the angular resolution, that is, two sources separated by angular distances equal to or greater than HPBW can be resolved [36]. The BS is equipped with M antennas, and the HPBW for the $d = \lambda/2$ spacing is [36]

$$\text{HPBW}(d = \lambda/2) \simeq \frac{1.06}{\sqrt{M-1}} \approx \frac{1}{\sqrt{M}} \qquad \text{(B.1)}$$

The angular resolution is $\theta_{\min} = \frac{1}{\sqrt{M}}$. It is assumed that

$$\begin{cases} \phi_m = \frac{1}{\sqrt{M}} m \\ \phi_n = \frac{1}{\sqrt{M}} n \end{cases} \tag{B.2}$$

where m and n are positive integers, and $m \neq n$. Therefore,

$$\begin{aligned} \boldsymbol{\alpha}(\phi_n)^H \boldsymbol{\alpha}(\phi_m) &= \frac{1}{M} \sum_{l=1}^{M-1} e^{j\pi l (\cos(\frac{n}{\sqrt{M}}) - \cos(\frac{m}{\sqrt{M}}))} \\ &= \frac{1}{M} \sum_{l=1}^{M-1} e^{-j2\pi l \sin(\frac{n-m}{\sqrt{M}}) \sin(\frac{n+m}{\sqrt{M}})} \end{aligned} \tag{B.3}$$

When M goes to infinity, it has

$$\begin{cases} \lim\limits_{M\to\infty} \sin(\frac{n-m}{\sqrt{M}}) = \frac{n-m}{\sqrt{M}} \\ \lim\limits_{M\to\infty} \sin(\frac{n+m}{\sqrt{M}}) = \frac{n+m}{\sqrt{M}} \end{cases} \tag{B.4}$$

Therefore, when $m \neq n$, we have

$$\lim_{M\to\infty} \boldsymbol{\alpha}(\phi_n)^H \boldsymbol{\alpha}(\phi_m) = \lim_{M\to\infty} \frac{1}{M} \sum_{l=1}^{M-1} e^{\frac{j2\pi l}{M}(n^2 - m^2)} = 0 \tag{B.5}$$

References

[1] F. Boccardi, R. W. Heath Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communi. Mag.*, vol. 52, no. 2, pp. 74-80, Feb. 2014.

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

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.* vol. 61, no. 4, pp. 1436-1449, Apr. 2013.

[4] H. Zhu, "Performance comparison between distributed antenna and microcellular systems," *IEEE J. Select. Areas Commun.*, vol. 29, no. 6, pp. 1151-1163, Jun. 2011.

[5] L Dai, "A Comparative study on uplink sum capacity with co-located and distributed antennas," *IEEE J. Select. Areas Commun.*, vol. 29, pp. 1200-1213, Jun. 2011.

[6] J. Wang, H. Zhu, and N. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Select. Areas Commun.*, vol. 30, pp. 675-683, May 2012.

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

[8] S. Jin, X. Wang, Z. Li, K. Wong, Y. Huang, and X. Tang, "On massive MIMO zero-forcing transceiver using time-shifted pilots," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 59-74, Jan. 2015.

[9] 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.

[10] 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.

[11] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - Part I: chunk allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734-2744, Sept. 2009.

[12] E. Bjornson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," Online Available: http://arxiv.org/abs/1503.06854, Aug. 2015.

[13] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," *Proc. Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, Illinois, Oct. 2013, pp. 293-298.

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

[15] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Theory,* vol. 62, pp. 3880-3890, Jul. 2016

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

[17] H. Ahmadi, A. Farhang, N. Marchetti, and A. MacKenzie, "A game theoretic approach for pilot contamination avoidance in massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 12-15, Feb. 2016.

[18] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, "On design of non-orthogonal pilot signals for a multi-cell massive MIMO system," *IEEE Wireless Commun. Lett.* , vol. 4, no. 2, pp.129-132, April 2015.

[19] A. Farhang, A. Aminjavaheri, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Pilot decontamination in CMT-based massive MIMO networks," in *Proc. 11th Int. Symp. Wireless Commun. Systems (WSWCS)*, 26-29 Aug. 2014, pp. 589-593.

[20] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, May 2011, pp. 3464-3467.

[21] 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.

[22] H. Q. Ngo, and E. G. Larsson, "EVD-Based channel estimations for multicell multiuser MIMO with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoustics, Speed and Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 3249-3252.

[23] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Select. Topics Signal Process.*, vol. 8, no. 5, 773-768, Sept. 2014.

[24] A. M. Sayeed and N. Behdad, "Continuous aperture phased MIMO: a new architecture for optimum line-of-sight links," in *Proc. IEEE Int. Symp. Ant. Propag. (APS)*, July 2011, pp. 293-296.

[25] M. Matthaiou, P. D. Kerret, G. K. Karagiannidis, and J. A. Nossek, "Mutual information statistics and beamforming performance analysis of optimized LoS MIMO systems," *IEEE Trans. Commun.*, vol. 58, no. 11, pp.180-191, Nov. 2010.

[26] Q. Zhang, S. Jin, Y. Huang, and H. Zhu, "Uplink rate analysis of multicell massive MIMO systems in Ricean fading," in *Proc. IEEE Globecom'14*, Dec. 2014, pp. 3279–3284.

[27] D. Yue and G. Li, "LOS-based conjugate beamforming and power-scaling law in massive-MIMO systems," Online Available: http://arxiv.org/abs/1404.1654, 2015.

[28] X. Li, S. Jin, H. A. Suraweera, J. Hou and X. Gao, "Statistical 3D beamforming for large-scale MIMO downlink systems over Rician fading channels," *IEEE Trans. Commun.*, vol. 64, pp. 1529-1543, Apr. 2016.

[29] Y. Wu, C. K. Wen, D. W. K. Ng, R. Schober, and A. Lozano, "Low complexity MIMO precoding for finite alphabet signaling," *IEEE Trans. Wireless Commun.,* in revision. Online Available: https://arxiv.org/abs/1606.03380

[30] S. Jin, X. Gao, and X. You, "On the ergodic capacity of rank-1 Ricean fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 502-517, Feb. 2007.

[31] A. J. Barabell, J. Capon, D. F. Delong, J. R. Johnson, and K. Senne "Performance comparison of superresolution array processing algorithms," *Tech. Rep. TST-72, Lincoln Lab., M.I.T.*, 1984.

[32] W. Tan, S. Jin, J. Wang, and M. Matthaiou, "Achievable sum-rate of multiuser massive MIMO downlink in Rican fading channels," in *Proc. IEEE ICC'15*, June 2015, pp. 3056-3061.

[33] K. Mullen, "A note on the tatio of two independent random varialbes," *The American Statisticians*, vol. 21, no. 3, pp. 30-31, Jun. 1967.

[34] D. Tse and P. Viswanath, "Fundamentals of wireless communication," *Cambridge University Press*, 2005.

[35] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.

[36] C. A. Balanis, "Antenna theory: analysis and design," 3rd ed. *JOHN WILEY & SONS, INC. Press*, 2005.

[37] J. Proakis, *Digital communications*, 4th ed. *New York: McGraw-Hill*, 2001.