Pilot Decontamination in Noncooperative Massive MIMO Cellular Networks Based on Spatial Filter

Zijun Gong, Fan Jiang, and Cheng Li
Faculty of Engineering and Applied Science
Memorial University of Newfoundland, St. John's, A1B 3X5, Canada
Email: {zijun.gong, fjiang, licheng}@mun.ca

Abstract

Pilot contamination has been known as one of the most challenging issues in massive multiple input multiple output (MIMO) systems. To be specific, every user will experience interferences from adjacent cell users who employ the same pilot sequence. For cell-edge users, pilot contamination is particularly detrimental, because their signals might be overwhelmed by interference from adjacent cells. In this paper, we propose a pilot decontamination method based on spatial filter, which exploits the spatial sparsity of massive MIMO channels. In massive MIMO communication protocols, there are generally four phases: pilot transmission, processing, uplink data transmission, downlink data transmission. In the first phase, the base station (BS) receives both desired signal and pilot contamination. In the second one, all users in the target cell stay silent for one symbol period and the BS only receives interference from adjacent cells. Then, fast Fourier transform can be employed to analyze the spatial spectrums of received signals in these two phases. Due to the spatial sparsity of massive MIMO channels, it is possible to identify pilot contamination components by comparing these two spectrums on different spatial signatures (or angles of arrival). Then, the next step is to construct a spatial filter and eliminate pilot contaminations. Both theoretical analysis and simulation results support the effectiveness of the proposed method. Besides, the complexity of the proposed method is comparable to that of the traditional channel estimator based on matched filter.

Index Terms

Spatial filter, massive MIMO, channel estimation, pilot contamination, cell-edge users.

I. INTRODUCTION

Massive multiple input multiple output (MIMO) has been a very hot research topic in recent years for its great potentials in improving spectral and energy efficiencies [1]. As one of the most important techniques for the fifth generation (5G) cellular networks, it aims to serve tens of single-antenna users with hundreds of antennas at base station (BS). The increase of antenna number at BSs brings many advantages. First, energy efficiency can be improved because antenna array at BSs can concentrate energy on users through beamforming [2]. Besides, spatial division multiple access (SDMA) is employed over orthogonal frequency division multiple access (OFDMA), which allows the same time-frequency resources to be reused by all users in the same cell, leading to much higher spectral efficiency. Moreover, low-complexity algorithms (e.g., matched filter) can be employed for precoding and decoding in massive MIMO [3], [4], because channels of different users are asymptotically orthogonal when BS antenna number is sufficiently large. Although the large antenna arrays at BSs will induce high computational complexity, researchers are working on iterative algorithms to maintain it on an acceptable level [5]–[7]. Last but not least, cheap power amplifiers working at milli-Watt level can be employed, because of the great power gain of BS antenna array [8], [9]. In spite of these great advantages, the practical application of massive MIMO is still facing some challenges and one of them is pilot contamination.

Similar to the traditional multi-user MIMO, channel estimation is indispensable for massive MIMO. Generally, orthogonal pilot sequences are assigned to users in the same cell and intra-cell interference can be totally eliminated. However, the length, and thus the number of pilot sequences are limited by coherence time and bandwidth, which leads to the unavoidable reuse of pilots in adjacent cells. Therefore, users sharing the same pilot will interfere with each other in the process of channel estimation and this phenomenon is referred to as pilot contamination, which puts a fundamental limit on the capacity of massive MIMO systems [1]. For cell-edge users, this problem is particularly detrimental, because their signal strength is comparable to their peers in adjacent cells. To break this barrier, many possible solutions have been proposed [10]–[22], and they will be briefly reviewed in the following paragraphs.

In our previous work [23], the pilot decontamination strategies are divided into four categories. The first choice is to alleviate pilot contamination by creating more orthogonal pilot sequences.

This can be done in time domain by lengthen pilot sequences, or in frequency domain by reducing the frequency reuse factor, as suggested by [1]. In [24], the authors employ a pilot length seven times the number of users per cell, so as to guarantee pilot orthogonality among users in adjacent cells. These methods can suppress pilot contamination to great extent, because adjacent cell users are major interfering sources. However, spectral efficiency will decrease due to the increased length of pilots in time or frequency domain. To overcome this deficiency, the authors of [13] proposed a time-shifted pilot scheme. In this method, when users of a specific cell are transmitting pilot sequences, all the adjacent cell users are at the phase of downlink data transmission or processing. By doing this, pilot contamination is no longer an issue, but inter-cell interference might be stronger, because cell-edge users will experience strong interferences from BSs in adjacent cells during pilot transmission phase. In this case, channel estimation will be conducted in low signal to interference and noise ratio (SINR). For the second type of decontamination strategies, the basic idea is to identify the subspace of desired signals by utilizing the statistics of channel state information (CSI) and received signals. For example, eigenvalue decomposition (EVD) is employed for channel estimation in [2], because the authors proved that every channel vector is an eigenvector of the covariance matrix of received signals when BS antenna number is sufficiently large. In [17], this idea is further developed, which proposes to obtain the subspace of channel vectors through singular value decomposition (SVD) of the received signal matrix. Then, pilot contamination can be eliminated by projecting the received signal onto this subspace. However, the efficacy of these methods is based on the assumption that desired signals are always stronger than pilot contaminations, which can not be guaranteed for cell-edge users. When the received signals are sparse in space, there is another possible solution for pilot contamination proposed in [11] and [25]. In [11], the authors show that MMSE estimator can completely eliminate pilot contamination, given that angle spreads of desired and interfering users do not overlap. Therefore, the same pilot sequence should be assigned to those users with minimum overlap in angular domain. In [25], the authors take both angular and power domain discriminations for pilot decontamination, based on the assumption that desired signals are generally stronger than pilot contaminations. In the third type of decontamination methods, both data and pilot are employed for channel estimation, as proposed in [16] and [21]. Noticing that data is generally longer than pilot, even when users are sharing the same pilot sequence, their data are still asymptotically orthogonal. Therefore, decoded data, in spite of being partially

correct, can be used to suppress pilot contamination and improve channel estimation accuracy. The last strategy is referred to as pilot contamination precoding (PCP) [19], [20], [26]. In all the other three types of decontamination methods, different cells work almost independently, while PCP is dependent on multi-cell cooperation. The basic idea is to add a precoding and decoding layer among adjacent cells, by doing which pilot contamination can be totally eliminated, given that BS antenna number approaches infinity. The price of PCP algorithms is the overhead of CSI exchange among BSs, which might be a problem for 5G, considering the large user population, high data rate and low latency demands.

In many researches above mentioned, desired signals are assumed to be stronger than pilot contamination. However, it is very difficult to guarantee this condition for cell-edge users in practical scenarios. Taking this into consideration, we proposed an innovative pilot decontamination method by employing the spatial sparsity of signals in massive MIMO systems. As we know, there are generally four phases in massive MIMO communication protocols: *pilot transmission*, *processing*, *uplink data transmission*, and *downlink data transmissions*. When a cell is in *pilot transmission* phase, the BS will receive signals from both desired and interfering users. Then, by employing fast fourier transformation (FFT), the BS can obtain the energy distribution of received signal in space. Then, during *processing* phase, users in the target cell will stay silent and the BS can monitor the signals from interfering users in adjacent cells. In this case, the BS can obtain the energy distribution of interfering users in space. By comparing these two distributions, the BS can identify the direction of arrival (DOA) of pilot contamination and construct the corresponding subspace. The last step is to project the channel estimation of matched filter (MF) onto its complementary subspace to eliminate pilot contamination and improve channel estimation accuracy. The fundamental idea behind this new method is that wireless channels are sparse in space domain, i.e., most energy of the desired signals concentrates on a small number of paths (or directions, equivalently) [11], [15], [27]–[30]. This is especially true for massive MIMO, because compared with the huge quantity of antennas at BS, the number of significant paths is much smaller. In [11] and [15], the authors even assume that the DOA spread of every terminal is limited to a narrow angle. Moreover, the propagation model of millimeter wave, being viewed as the perfect match for massive MIMO [31], tends to be line of sight (LOS) or near-LOS [32], which means the wireless channel for millimeter wave will be even sparser. Part of our work has been published in [33], where we presented the basic idea and

some preliminary numerical results. In this paper, the major contributions include the following aspects.

- First, we systematically introduced the concepts of spatial spectrum and spatial sampling. We prove that more than 80 percent energy concentrates on the mainlobe for the signal coming from an arbitrary direction. Therefore, received signals are sparse in space.
- Second, we show that the probability that the mainlobes of desired signal and pilot contamination overlap in space is inversely proportional to the number of BS antennas. For massive MIMO, this probability is negligible. As a result, it is possible to isolate pilot contamination from desired signals in space.
- Third, we propose an algorithm to identify the mainlobes of pilot contamination. The BS receives signal in both *pilot transmission* and *processing* phases. Desired signal only exists in *pilot transmission* phase, while pilot contamination exists in both phases. Therefore, on a specific direction, the ratio of the received signal strength in these two phases is a random variable, and its distribution depends on whether pilot contamination exists or not on this direction. We analyze the conditional distributions of the ratio in different scenarios and employ maximum a posterior (MAP) algorithm to identify pilot contamination components.

The remaining parts of this paper is organized as follows. In Section II, system model of massive MIMO systems and the pilot contamination issue are presented. In Section III, we introduced the concepts of spatial spectrum and spatial sampling. In Section IV, the proposed method is presented and analyzed in detail. The last two sections are simulations and conclusions, respectively.

Notations: throughout the paper, the upper case bold letters represent matrixes while the lower case bold font denotes column vectors. $\mathrm{Tr}\{\mathbf{A}\}$ means the trace of an arbitrary matrix $\mathbf{A}$. $E\{\cdot\}$ represents the expectation of any random variable, vector or matrix. $V\{\cdot\}$ gives the covariace matrix of a random vector or variance of a random variable. $\mathbf{A}^T$ and $\mathbf{A}^H$ indicate the transpose and Hermite transpose of matrix $\mathbf{A}$. $|\cdot|$ is the norm of the target vector. $Pr(A)$ is the probability that event A happens. For an arbitrary vector $\mathbf{a}$, $\mathbf{a}[n]$ indicates the n-th component. $\Im\{a\}$ and $\Re\{a\}$ represent the imaginary and real parts of a complex number a.

II. Channel Model and Pilot Contamination

In massive MIMO systems, hundreds of antennas are installed at BS to serve tens of users. Generally, the antenna array at BS can be in various forms, i.e., linear, rectangular or even cylindrical. However, in our case, to maximize the angular domain resolution, we assume that M antennas are linearly placed at the BS and the distance between any two adjacent antenna elements is equal to half the wavelength of the carrier [34]. To begin with, suppose that there is only one path between user and BS, and the angle of arrival of the desired signal is denoted by θ. Then, the system model will be given as: [34]

$$\mathbf{y} = x\sqrt{\rho}\mathbf{e}[\omega] + \mathbf{n}, \tag{1}$$

where $\omega = \pi\cos\theta$, x is the transmitted symbol, ρ is the path-loss coefficient, and $\mathbf{n}$ is white noise. $\mathbf{e}[\omega]$ is referred to as the spatial signature of this signal and is given by [34]:

$$\mathbf{e}[\omega] := \frac{1}{\sqrt{M}}\begin{bmatrix} 1 \\ \exp(-j\omega) \\ \vdots \\ \exp(-j(M-1)\omega) \end{bmatrix}. \tag{2}$$

When there are more than one pathes between user and BS, the channel model will be

$$\mathbf{h} = \sum_{b=1}^{B}\sqrt{\rho_b}\mathbf{e}[\omega_b]e^{j\phi_b}, \tag{3}$$

where ϕ_b and $\mathbf{e}[\omega_b]$ denote the random phase delay and spatial signature of the b-th path, respectively. ρ_b is the path-loss coefficient given by $\rho_b = \frac{s_b}{d^\gamma}$, in which γ is the path-loss exponent and d represents the distance between the terminal and the BS. s_b is a log-normal random variable, i.e., $10\lg s_b \sim \mathcal{N}(0, \sigma_{shad}^2)$[1] [1]. This model is widely adopted in massive MIMO related literatures [11], [15], [22], [35].

Assume the user number per cell is K, and pilot length is τ. To avoid intra-cell interference, the users in the same cell should be assigned with orthogonal pilot sequences, which demands $\tau \geq K$. Suppose we have L cells in the area of interest, and K orthogonal pilot sequences are

[1]In this paper, $\lg(\cdot)$ and $\ln(\cdot)$ represent base-10 and natural logarithms, respectively.

fully reused in all cells. Then, the received pilot signal at the l-th BS will be

$$\mathbf{Y}^{(l)} = \mathbf{H}_l^{(l)}\mathbf{P}^T + \sum_{l' \neq l} \mathbf{H}_{l'}^{(l)}\mathbf{P}^T + \mathbf{N}, \tag{4}$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_K]$, and $\mathbf{p}_k$ is the k-th pilot sequence. The elements in $\mathbf{N} \in \mathbb{C}^{M \times \tau}$ are i.i.d. zero-mean complex Gaussian noise, with a variance of σ_n^2. Therefore, we have $E\{\mathbf{N}\mathbf{N}^H\} = \tau\sigma_n^2 \cdot \mathbf{I}_M$. $\mathbf{H}_{l'}^{(l)} \in \mathbb{C}^{M \times K}$ is the channel matrix between the users in cell l' and the BS in cell l. With matched filter, channel estimation of the l-th cell will be

$$\hat{\mathbf{H}}_{mf}^{(l)} = \mathbf{Y}^{(l)}\mathbf{P}^*/\tau = \mathbf{H}_l^{(l)} + \sum_{l' \neq l} \mathbf{H}_{l'}^{(l)} + \mathbf{N}\mathbf{P}^*/\tau, \tag{5}$$

where we implicitly use the fact that $\mathbf{P}^T\mathbf{P}^* = \tau \cdot \mathbf{I}_K$, due to the orthogonality of different pilot sequences. For the k-th user in the l-th cell, the MF channel estimation will be

$$\hat{\mathbf{h}}_{mf}^{(l,k)} = \mathbf{h}_{l,k}^{(l)} + \sum_{l' \neq l} \mathbf{h}_{l',k}^{(l)} + \mathbf{n}_k, \tag{6}$$

where $\mathbf{n}_k = \mathbf{N}\mathbf{P}^*/\tau$. As we can observe in Eqn. (6), the channel estimation based on MF consists of three parts. The first and last parts are desired channel information and noise, respectively, while the middle part contains interference from other cells due to pilot reuse, i.e., pilot contamination. Based on the channel model in Equation (3), the received signal strength decreases fast with distance, which leads to the following observations.

(a). First, the most significant pilot contamination must come from adjacent cells. Considering typical hexagonal cells, for the k-th user in the l-th cell, there are only six neighbor cells and every cell only contains one interfering user. Therefore, pilot contamination for a specific target user is sparse, compared with the huge number of antennas at BS.

(b). For users close to the BS, pilot contamination is not a big issue because their signal will be much stronger than interference. However, cell-edge users are vulnerable to pilot contamination, because their signals may be overwhelmed by interfering users.

Motivated by these observations, we propose an innovative pilot decontamination method, aiming to improve the SINR of cell-edge users. Before presenting the proposed algorithm, we need to first briefly review the concepts of spatial spectrum and sampling in the following section.

III. Spatial Spectrum and Sampling

A. The Concept of Spatial Spectrum

Assume the received signal at the BS is given as $\mathbf{y} = x\sqrt{\rho_s}\mathbf{e}[\omega_{\rm s}] + \mathbf{n}$, where $\mathbf{e}[\omega_{\rm s}]$ is the spatial signature of received signal, and $\mathbf{n}$ represents noise. Then, we can define the *spatial spectrum* of the received signal as

$$Y(\omega) = \mathbf{e}[\omega]^H\mathbf{y} = x\sqrt{\rho_s}\mathbf{e}[\omega]^H\mathbf{e}[\omega_{\rm s}] + \mathbf{e}[\omega]^H\mathbf{n}. \tag{7}$$

From another perspective,

$$Y(\omega) = \mathbf{e}[\omega]^H\mathbf{y} = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{y}[m]e^{jm\omega} = \text{IDTFT}\,[\,\mathbf{y}\,], \tag{8}$$

where IDTFT is short for inverse discrete time fourier transform. From Eqn. (7), we can see that it is important to figure out how the envelope of $\mathbf{e}[\omega]^H\mathbf{e}[\omega_{\rm s}]$ varies with ω.

Define $f(\Delta\omega) := \mathbf{e}[\omega_1]^H\mathbf{e}[\omega_2]$ $(\forall\, \omega_1,\, \omega_2 \in [0, 2\pi])$, where $\Delta\omega = \omega_2 - \omega_1$, and we can obtain

$$f(\Delta\omega) = \frac{\sin\frac{M\Delta\omega}{2}}{M\sin\frac{\Delta\omega}{2}} \cdot e^{-j(M-1)\Delta\omega/2}. \tag{9}$$

Figure 1 shows numerical results of $|f(\Delta\omega)|$ with different M.

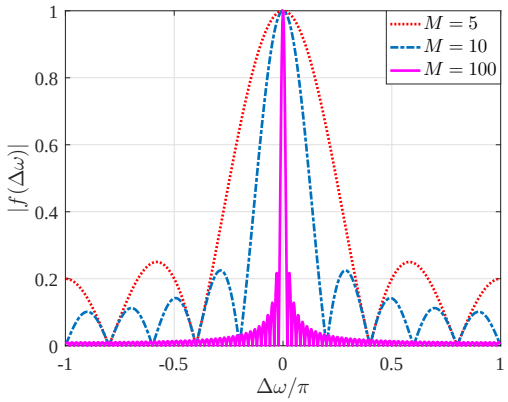

Fig. 1. The main lobe of $|f(\Delta\omega)|$ becomes narrower as M increases.

As shown in Fig. 1, the main lobe of $|f(\Delta\omega)|$ is centered at 0, with an width of $4\pi/M$. For massive MIMO, M is generally very large, and most energy concentrates on a small angle spread. Therefore, interference and desired signals are generally well separated in space (i.e., the

main lobes of signal and interference do not overlap). In this case, spatial filter can be constructed to filter out the main lobe of interference while maintain most energy of desired signals.

B. The Concept of Spatial Sampling

Analogous to the time domain signal processing, we can employ IDFT (Inverse Discrete Fourier Transform)[2] instead of IDTFT to analyze the spectrum of a target sequence, which is more suitable for the modern digital signal processors. To be specific, the discrete *spatial spectrum* of $\mathbf{y}$ is

$$\mathbf{y}_\omega = \text{IDFT}\left[\,\mathbf{y}\,\right] = \mathbf{F}_M\mathbf{y}, \tag{10}$$

where $\mathbf{F}_M$ is given as $\mathbf{F}_M := \left[\,\mathbf{e}\left[\,0\,\right], \mathbf{e}\left[\omega_0\right], \cdots, \mathbf{e}\left[(M-1)\omega_0\right]\,\right]^H$ and $\omega_0 = 2\pi/M$. As a result, $\mathbf{y}_\omega$ is the sampled version of $Y(\omega)$ in Equation (8). An arbitrary spatial vector $\mathbf{e}[\omega]$ $(\omega \in [0, 2\pi))$ can be decomposed as $\mathbf{e}[\omega] = \sum_{m=0}^{M-1} \alpha_m \mathbf{e}[m\omega_0]$, where α_m's and their absolute values are given by

$$\alpha_m = \mathbf{e}[m\omega_0]^H\mathbf{e}[\omega] \text{ and } |\alpha_m| = \frac{1}{M}\left|\frac{\sin\frac{M}{2}(\omega - m\omega_0)}{\sin\frac{1}{2}(\omega - m\omega_0)}\right|. \tag{11}$$

There exists an integer $l \in [0, M-1]$, which guarantees $\omega/\omega_0 \in [l,\, l+1)$. Therefore, we can write ω as $\omega = (l+\beta)\omega_0$ $(0 \le \beta < 1)$. Based on these definitions, we present the following theorem.

Theorem 1: When M is infinitely large, more than 80% energy concentrates on $\mathbf{e}[l\omega_0]$ and $\mathbf{e}[(l+1)\omega_0]$. To be specific, $\lim\limits_{M\to\infty} |\alpha_l|^2 + |\alpha_{l+1}|^2 \ge 8/\pi^2$ and the equality holds when $\beta = 0.5$.

Proof: When $\beta = 0$, we have $\omega = l\omega_0$, leading to $|\alpha_l|^2 + |\alpha_{l+1}|^2 = |f(0)|^2 + |f(\omega_0)|^2 = 1$. When $\beta \neq 0$, the limit of $|\alpha_l|^2 + |\alpha_{l+1}|^2$ can be derived as follows.

$$\lim_{M\to\infty} |\alpha_l|^2 + |\alpha_{l+1}|^2 = \lim_{M\to\infty} |f(\beta\omega_0)|^2 + |f(\omega_0 - \beta\omega_0)|^2 = \frac{\sin^2\beta\pi}{\pi^2}\left[\frac{1}{\beta^2} + \frac{1}{(1-\beta)^2}\right].$$

Define $R(\beta) = \frac{\sin^2\beta\pi}{\pi^2}\left[\frac{1}{\beta^2} + \frac{1}{(1-\beta)^2}\right]$ $(\beta \in (0,1))$ and numerical results show that $R(\beta)$ is a concave function. Noticing that $\left.\frac{dR(\beta)}{d\beta}\right|_{\beta=0.5} = 0$, we can conclude that $R(\beta)$ has a minimum at $\beta = 0.5$, and $R(0.5) = 8/\pi^2$. ■

Theorem 1 shows that most energy of received signal from a specific direction concentrates on the two samples in the main lobe. As a result, if we use $\alpha_l\mathbf{e}[l\omega_0] + \alpha_{l+1}\mathbf{e}[(l+1)\omega_0]$ to

[2]It should be noted that we can employ fast Fourier transform (FFT) to reduce computational complexity.

approximate $\mathbf{e}[\omega]$, the residual error will be smaller than 20%. Generally, if we take the closest $2R$ components to reestablish $\mathbf{e}[\omega]$, residual error can be defined as

$$Res[\,R\,] = 1 - \sum_{r=1}^{R} |\alpha_{mod_M(l+1-r)}|^2 + |\alpha_{mod_M(l+r)}|^2, \tag{12}$$

where $mod_M(k) = \text{mod}(k, M)$. Then, the residual error is upper bounded by

$$Res[\,R] \leq \frac{\pi^2}{12} + \frac{1}{4R^2} - \sum_{r=1}^{R} \frac{1}{2r^2}, \tag{13}$$

and this bound is justified in Appendix A.

These analyses show that the discrete spatial spectrum of the desired signal is *sparse*, and most energy of the desired signals and interference concentrates on narrow angle spreads. As a result, the intuitive idea is to identify the main lobes of pilot contamination, and construct a spatial filter to eliminate them.

IV. Spatial Filter Based Channel Estimation Method

As has been discussed, pilot contamination mainly emerges from adjacent cells, and the received signal is generally sparse in space. Besides, cell-edge users are most vulnerable to pilot contamination, because their signals are not necessarily stronger than interferences. Based on these observations, we proposed an innovative pilot decontamination algorithm for massive MIMO users.

The first step is to divide all cells into three groups, G_1, G_2 and G_3, and make sure that adjacent cells belong to different groups. Then, we slightly modify the widely employed TDD communication protocol by shifting the processing phase of different groups, as illustrated in Figure 2.

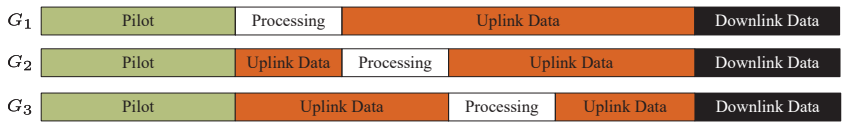

Fig. 2. Shifted pilot schedule.

As we can see in Figure 2, when one group is in the processing phase, the other two groups

are transmitting uplink data. Without loss of generality, we assume the l-th cell is the target cell, and it belongs to G_1. Then, by employing the MF channel estimator, we can obtain the channel estimation of the k-th user in the l-th cell as

$$\hat{\mathbf{h}}_{mf}^{(l,k)} = \mathbf{Y}^{(l)}\mathbf{p}_k^*/\tau = \mathbf{h}_{l,k}^{(l)} + \sum_{c_{l'}\notin G_1} \mathbf{h}_{l',k}^{(l)} + \sum_{c_{l'}\in G_1, l'\neq l} \mathbf{h}_{l',k}^{(l)} + \mathbf{n}_k, \tag{14}$$

where c_l denotes the l-th cell, and $\mathbf{n}_k = \mathbf{N}\mathbf{p}_k^*/\tau$. $\mathbf{h}_{l',k}^{(l)}$ represents the channel vector from the k-th user in the l'-th cell to the l-th BS. In Equation (14), the first and last parts are the desired CSI and white noise, respectively. The second part contains pilot contamination from the k-th users in G_2 and G_3. It should be noted that only six of them are located in adjacent cells for typical hexagonal cells, which means there are at most six strong pilot contamination components. The third part contains pilot contamination from users in G_1, which are all weak because they are at least of three cell radius away from c_l. Overall, there are potentially seven strong components in $\hat{\mathbf{h}}_{mf}^{(1,k)}$, and it is very sparse compared with the number of BS antennas.

During the *processing phase* of cells in G_1, all users in c_l are silent, and the BS can receive signals from users in active cells ($c_{l'} \in G_2 \cup G_3$ or $c_{l'} \notin G_1$) as

$$\mathbf{y}_{proc}^{(l)} = \sum_{c_{l'}\notin G_1} \mathbf{h}_{l',k}^{(l)} s_{l',k} + \sum_{k'\neq k} \sum_{c_{l'}\notin G_1} \mathbf{h}_{l',k'}^{(l)} s_{l',k'} + \mathbf{n}_{proc}, \tag{15}$$

where $s_{l,k'}$ is the transmitted symbol of the k'-th user in the l-th cell during the *processing phase* of cells in G_1. In Equation (15), the first part denotes signals from the k-th users in cells belong to G_2 and G_3, and contains at most six strong components. The second part denotes general inter-cell interference from active cells, which does not exist in Equation (14), because none of these interfering users in this part is using the k-th pilot sequence.

Based on the above discussions, we can see that $\hat{\mathbf{h}}_{mf}^{(l,k)}$ is sparse in space, and the major components include both desired signals and pilot contaminations. On the other hand, $\mathbf{y}_{proc}^{(l)}$ is composed of pilot contaminations and general inter-cell interference. The spatial spectrums of these two vectors will overlap on those spatial signatures dominated by pilot contaminations, as shown in Figure 3. As a result, the intuitive idea is to eliminate polite contamination by identifying their spatial signatures and construct the complementary subspace. To achieve this goal, the first step is to identify the major spatial signatures in $\hat{\mathbf{h}}_{mf}^{(l,k)}$, where pilot contamination

$$\hat{\mathbf{h}}_{mf}^{(l,k)} = \sum_{b=1}^{B} a_b^{(l,k)} \cdot \mathbf{e}[\omega_b^{(l,k)}] + \sum_{i=1}^{6} \sum_{b=1}^{B} a_b^{(l_i,k)} \cdot \mathbf{e}[\omega_b^{(l_i,k)}] + \sum_{l' \notin \{l,\{l_i\}_{i=1}^{6}\}} \sum_{b=1}^{B} a_b^{(l,k)} \cdot \mathbf{e}[\omega_b^{(l,k)}] + \mathbf{n}_k, \quad (16)$$

from adjacent cells must exist. By doing this, we only need to focus on a small number of spatial signatures when we try to identify pilot contamination in future steps, which will cut down complexity and processing time.

A. Identification of the Spatial Signatures of Desired Signals and Pilot Contaminations

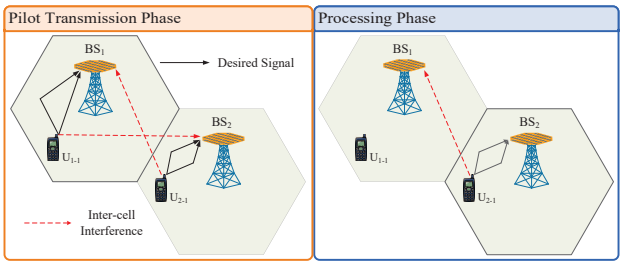

Fig. 3. Pilot contamination identification.

As we have mentioned, the third part in Equation (14) is negligible. Besides, there are only six strong components in the second part. Without loss of generality, assume the indexes of these six adjacent cells are $l_1, l_2, \cdots, l_6$, and Equation (14) can be reorganized as

where $a_b^{(l,k)} = \sqrt{\rho_b^{(k,l)}} e^{j\phi_b^{(k,l)}}$ denotes the coefficient of the b-th path from the k-th user in the l-th cell. $\rho_b^{(k,l)}$ and $e^{j\phi_b^{(k,l)}}$ are the path-loss coefficient and random phase, respectively.

To identify the spatial signatures of desired signals and pilot contaminations, we need to analyze the spatial spectrum of $\hat{\mathbf{h}}_{mf}^{(l,k)}$ through IDFT as $\hat{\mathbf{h}}_{\omega}^{(l,k)} = \mathbf{F}_M \hat{\mathbf{h}}_{mf}^{(l,k)}$, which indicates the energy distribution of $\hat{\mathbf{h}}_{mf}^{(l,k)}$ on different directions and the m-th component of $\hat{\mathbf{h}}_{mf}^{(l,k)}$ is given by

$$\begin{aligned} \hat{\mathbf{h}}_{\omega}^{(l,k)}[m] = &\sum_{b=1}^{B} a_b^{(l,k)} \cdot f\left(\omega_b^{(l,k)} - m\omega_0\right) + \sum_{i=1}^{6} \sum_{b=1}^{B} a_b^{(l_i,k)} \cdot f\left(\omega_b^{(l_i,k)} - m\omega_0\right) + \\ &\sum_{l' \notin \{l,\{l_i\}_{i=1}^{6}\}} \sum_{b=1}^{B} a_b^{(l',k)} \cdot f\left(\omega_b^{(l',k)} - m\omega_0\right) + \mathbf{n}_k[m]. \end{aligned} \quad (17)$$

For $l' \notin \{l, \{l_i\}_{i=1}^{6}\}$, $\left|a_b^{(l',k)}\right|$ is negligible because users in these cells are at least three cell-radius away from the target BS. According to the *Central limit theorem*, the third part in Equation (17) can be viewed as Gaussian noise. Therefore, we define $\tilde{\mathbf{n}}_k[m]$ to replace the last two parts:

$$\tilde{\mathbf{n}}_k[m] = \sum_{l' \notin \{l_i\}_{i=1}^{6}} \sum_{b=1}^{B} a_b^{(l',k)} \cdot f\left(\omega_b^{(l',k)} - m\omega_0\right) + \mathbf{n}_k[m]. \tag{18}$$

As we have mentioned, most of the received energy concentrates on a few directions; therefore, the received signal will be much stronger than noise on those directions. Motivated by this observations, we define the following metric to identify the spatial signatures and desired signals:

$$\lambda_m^{(l,k)} = \left|\hat{\mathbf{h}}_\omega^{(l,k)}[m]\right|^2. \tag{19}$$

In this equation, $\lambda_m^{(l,k)}$ indicates the received energy of the l-th BS on the m-th spatial signature (or direction). When $\hat{\mathbf{h}}_\omega^{(l,k)}$ does not a strong component on $m\omega_0$ or the signal is substantially weaker than noise, λ is generally small and follows exponential distribution. On the other hand, λ will be much larger if $\hat{\mathbf{h}}_\omega^{(l,k)}$ has a strong component on $m\omega_0$, and the probability density function (PDF) of λ is very close to normal distribution. The empirical distribution of $\lambda_m^{(l,k)}$ is shown in Fig. 4. In

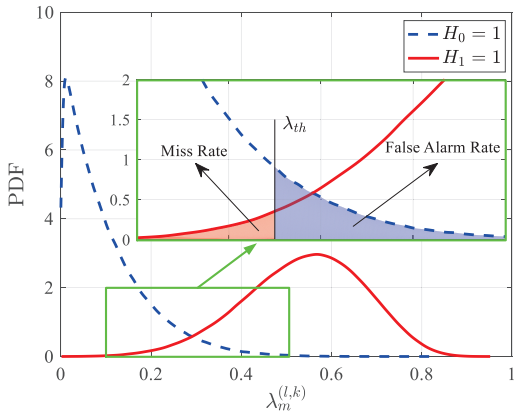

Fig. 4. Probability density function of $\lambda_m^{(l,k)}$ when signal is ten times stronger than noise on $m\omega_0$.

Figure 4, $H_0 = 1$ means that neither desired signal nor pilot contamination exist on $m\omega_0$, while $H_1 = 1$ represents the opposite situation. Given that $H_0 = 1$ on $m\omega_0$, the value of $\lambda_m^{(l,k)}$ is very small. On the contrary, when $H_1 = 1$, the received signal strength on $m\omega_0$ will be much larger.

As a result, we can set a threshold to identify these strong components and and their spatial signatures. Here, we just assume the threshold has been obtained as λ_{th}, and further details on threshold selection will be discussed in next section.

Then, the spatial signatures of the desired signals or pilot contamination of the k-th user in the l-the cell will be $\mathbf{U}_{S,P}^{(l,k)} = [\,\mathbf{e}\,[m_1\omega_0]\,, \mathbf{e}\,[m_2\omega_0]\,, \cdots\,, \mathbf{e}\,[m_Q\omega_0]\,]$, where m_q satisfies $\lambda_{m_q}^{(l,k)} > \lambda_{th}$.

B. Selection of Threshold λ_{th}

Let $f_0(\lambda_m^{(l,k)})$ and $f_1(\lambda_m^{(l,k)})$ denote the PDF of $\lambda_m^{(l,k)}$ in $H_0 = 1$ and $H_1 = 1$ scenarios. As is discussed in Appendix B, $f_0(\lambda_m^{(l,k)})$ is exponential distribution, while $f_1(\lambda_m^{(l,k)})$ is very close to normal distribution. Based on signal detection theory, the false alarm rate and the miss rate can be defined as:

$$R_{FA} = \Pr\left(\lambda_m^{(l,k)} > \lambda_{th} | H_0 = 1\right),\ R_M = \Pr\left(\lambda_m^{(l,k)} \leq \lambda_{th} | H_1 = 1\right), \tag{20}$$

and they are demonstrated in Fig. 4. To minimize the false alarm and miss rates, λ_{th} should be chosen as ,

$$\lambda_{th} = \left[3SNR_o + 1 - \sqrt{2SNR_o \cdot (4SNR_o + 2 - \ln(4\pi SNR_o))}\right] \cdot \sigma_o^2. \tag{21}$$

so that $f_0(\lambda_{th}) = f_1(\lambda_{th})$. σ_o^2 is the variance of $\tilde{\mathbf{n}}_k[m]$, while SNR_o indicates the ratio of signal (or pilot contamination) strength to σ_o^2. They can be found in Appendix B. The false alarm and miss rates are exclusively dependent on SNR_o, as shown in Equation (22a) and (22b), respectively. The proof can be found in Appendix B. Besides, we can prove that both R_M and R_{FA} decrease with SNR_o. Due to space limit, the proof is not included.

$$R_M = \frac{1}{2} erfc\left(-\sqrt{SNR_o} + \sqrt{\left(2SNR_o + 1 - \frac{1}{2}\ln(4\pi SNR_o)\right)}\right) \tag{22a}$$

$$R_{FA} = \exp\left(-3SNR_o - 1 + \sqrt{2SNR_o \cdot (4SNR_o + 2 - \ln(4\pi SNR_o))}\right) \tag{22b}$$

In practical scenarios, both SNR_o and σ_o are unknown. Therefore, we need to estimate them from the MF channel estimations, i.e., $\hat{\mathbf{h}}_{mf}^{(l,k)}$. As we have mentioned, desired signals and pilot contamination only exist on a small fraction of the components in $\hat{\mathbf{h}}_{\omega}^{(l,k)}$. As a result, we can sort the components in $\hat{\mathbf{h}}_{\omega}^{(l,k)}$ based on their strength, and choose the weakest ones of them to

estimate σ_o^2. On the other hand, the strongest ones can be employed to estimate signal (or pilot contamination) strength. Then, SNR_o can be estimated as the ratio of signal strength to σ_o^2.

C. The Isolation of Pilot Contamination in $\mathbf{y}_{proc}$

During the *processing* phase, the received signal is given by Equation (15), which is composed of signals from all cells in G_2 and G_3. However, only six of them are adjacent to the target BS and generate strong interferences. Therefore, we can rewrite Equation (15) as (23),

$$\mathbf{y}_{proc}^{(l)} = \sum_{i=1}^{6} \mathbf{h}_{l_i,k}^{(l)} s_{l_i,k} + \sum_{k' \neq k} \sum_{i=1}^{6} \mathbf{h}_{l_i,k'}^{(l)} s_{l_i,k'} + \sum_{c_{l'} \notin \left\{ G_1 \bigcup \{l_i\}_{i=1}^{6} \right\}} \sum_{k'} \mathbf{h}_{l',k'}^{(l)} s_{l',k'} + \mathbf{n}_{proc}, \tag{23}$$

and the spatial spectrum of $\mathbf{y}_{proc}^{(l)}$ can be obtained through IDFT as $\mathbf{y}_{\omega}^{(l)} = \mathbf{F}_M \mathbf{y}_{proc}^{(l)}$. The m-th element will be

$$\mathbf{y}_{\omega}^{(l)}[m] = \sum_{i=1}^{6} \sum_{b=1}^{B} a_b^{(l_i,k)} \cdot f\left(\omega_b^{(l_i,k)} - m\omega_0\right) s_{l_i,k} + \tilde{\mathbf{n}}_{proc}[m], \tag{24}$$

where $\tilde{\mathbf{n}}_{proc}$ indicates the IDFT of the last three parts in Eqn. (23) combined. Due to the large number of users, $\tilde{\mathbf{n}}_{proc}$ can be treated as Gaussian noise.

The strong components in $\mathbf{y}_{\omega}^{(l)}$ falls in two categories: pilot contamination and general inter-cell interferences. It should be noted that both these two categories come from adjacent cell users that are located close to the edges of c_l. Our major objective is to identify the spatial signatures of pilot contaminations through the following metric:

$$\phi_m^{(l,k)} = \frac{|\mathbf{y}_{\omega}^{(l)}[m]|^2}{|\hat{\mathbf{h}}_{\omega}^{(l,k)}[m]|^2}, \tag{25}$$

where $\mathbf{y}_{\omega}^{(l)}[m]$ is given by Equation (24). It should be noted that we only need to compute $\phi_m^{(l,k)}$'s for $m = m_q \, (q = 1, 2, \cdots, Q)$, because spatial signatures of pilot contamination of the k-th user in the l-th cell must be among the columns of $\mathbf{U}_{S,P}^{(l,k)}$.

Depending on whether desired signal or pilot contamination exists on $m\omega_0$, the conditional PDF of $\phi_m^{(l,k)}$ can be represented as $f_{ds}\left(\phi_m^{(l,k)}\right)$ or $f_{pc}\left(\phi_m^{(l,k)}\right)$, respectively. The closed form representations of these two functions will be discussed in next section. Then, we can identify pilot contamination through maximum a posteriori probability (MAP) estimation. As a result,

the subspace of pilot contamination for the k-th user in the l-th cell can be constructed as

$$\mathbf{U}_P^{(l,k)} = [\,\mathbf{e}\,[m_1'\omega_0]\,,\mathbf{e}\,[m_2'\omega_0]\,,\cdots,\mathbf{e}\,[m_P'\omega_0]\,]\,, \tag{26}$$

where m_p' satisfies $\lambda_{m_p'}^{(l,k)} > \lambda_{th}$ and $f_{pc}\left(\phi_{m_p'}^{(l,k)}\right) > f_{ds}\left(\phi_{m_p'}^{(l,k)}\right)$. The last step is to project the MF channel estimation onto the null space of $\mathbf{U}_P^{(l,k)}$ to eliminate strong pilot contaminations as

$$\hat{\mathbf{h}}_{sf}^{(l,k)} = (\mathbf{I} - \mathbf{U}_P^{(l,k)}\mathbf{U}_P^{(l,k)^H})\hat{\mathbf{h}}_{mf}^{(l,k)}, \tag{27}$$

where $\hat{\mathbf{h}}_{sf}^{(l,k)}$ is the new channel estimation of the k-th user in c_l based on *spatial filter*. The proposed algorithm is briefly summarized in Algorithm 1.

Algorithm 1 Channel estimation for the k-th user in the l-th cell based on spatial filter.

Input:

Received pilot, $\mathbf{Y}^{(l)}$;
Received signal on processing phase, $\mathbf{y}_{proc}^{(l)}$;
The k-th pilot sequence, $\mathbf{p}_k$.

Output:

Channel estimation

1: MF channel estimation $\hat{\mathbf{h}}_{mf}^{(l,k)}$;
2: Compute the spectrum of $\hat{\mathbf{h}}_{mf}^{(l,k)}$;
3: Identify the spatial signatures of desired signals and pilot contamination through $\lambda_m^{(l,k)}$;
4: Compute the spectrum of $\mathbf{y}_{proc}^{(l)}$;
5: Subspace construction of pilot contamination through $\phi_m^{(l,k)}$;
6: Pilot contamination elimination by subspace projection $\hat{\mathbf{h}}_{sf}^{(l,k)} = (\mathbf{I} - \mathbf{U}_P^{(l,k)}\mathbf{U}_P^{(l,k)^H})\hat{\mathbf{h}}_{mf}^{(l,k)}$;
7: **return** $\hat{\mathbf{h}}_{sf}^{(l,k)}$;

D. Conditional PDF of $\phi_m^{(l,k)}$

As we have mentioned, $\phi_m^{(l,k)}$ tends to be smaller when pilot contamination exists on $m\omega_0$. On the other hand, $\phi_m^{(l,k)}$ will be larger when desired signals come from $m\omega_0$. Therefore, $\phi_m^{(l,k)}$ follows different distributions in these two scenarios. To employ the MAP estimator, we need to first analyze the statistical properties of $\phi_m^{(l,k)}$ in different situations.

We will first analyze the distribution of $\phi_m^{(l,k)}$ when pilot contamination exists on $m\omega_0$. Without loss of generality, we assume pilot contamination on $m\omega_0$ comes from the b-th path of the

k-th user in the l_i-th cell. In other words, $\omega_b^{(l_i,k)}$ is close to $m\omega_0$, there will be strong pilot contamination components in both $\mathbf{y}_\omega^{(l)}[m]$ and $\hat{\mathbf{h}}_\omega^{(l,k)}[m]$, presented as

$$\begin{cases} \hat{\mathbf{h}}_\omega^{(l,k)}[m] = a_b^{(l_i,k)} \cdot f(\omega_b^{(l_i,k)} - m\omega_0) + \tilde{\mathbf{n}}_k[m] \\ \mathbf{y}_\omega^{(l)}[m] = a_b^{(l_i,k)} \cdot f(\omega_b^{(l_i,k)} - m\omega_0) s_{l_i,k} + \tilde{\mathbf{n}}_{proc}[m], \end{cases} \tag{28}$$

where $\tilde{\mathbf{n}}_{proc} \sim N(\mathbf{0}, \sigma_{proc}^2 \cdot \mathbf{I}_M)$. In this case,

$$\phi_m^{(l,k)} = \frac{\left|r_{pc} s_{l_i,k} + \tilde{\mathbf{n}}_{proc}[m]\right|^2}{\left|r_{pc} + \tilde{\mathbf{n}}_k[m]\right|^2}, \tag{29}$$

where $r_{pc} = a_b^{(l_i,k)} \cdot f(\omega_b^{(l_i,k)} - m\omega_0)$. $|\tilde{\mathbf{n}}_k[m]|^2$ can be neglected because it's much weaker than r_{pc}. Therefore, we have

$$\phi_m^{(l,k)} 2|r_{pc}|^2/\sigma_{proc}^2 = \left|\frac{\Re\{r_{pc}\} + \Re\left\{\tilde{\mathbf{n}}_{proc}[m] s_{l_i,k}^*\right\}}{\sigma_{proc}/\sqrt{2}}\right|^2 + \left|\frac{\Im\{r_{pc}\} + \Im\left\{\tilde{\mathbf{n}}_{proc}[m] s_{l_i,k}^*\right\}}{\sigma_{proc}/\sqrt{2}}\right|^2. \tag{30}$$

By noticing that the right hand side of Equation (30) follows noncentral chi-square distribution, we can obtain the approximate PDF of $\phi_m^{(l,k)}$ as

$$f_{pc}(\phi_m^{(l,k)}) = \frac{|r_{pc}|^2}{\sigma_{proc}^2} \exp\left(-\frac{|r_{pc}|^2}{\sigma_{proc}^2}\left(\phi_m^{(l,k)} + 1\right)\right) \cdot I_0\left(\frac{2|r_{pc}|^2}{\sigma_{proc}^2}\sqrt{\phi_m^{(l,k)}}\right), \tag{31}$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first kind given by

$$I_v(y) = (y/2)^v \sum_{j=0}^{\infty} \frac{(y^2/4)^j}{j!\Gamma(v+j+1)}. \tag{32}$$

This result is justified in Appendix D.

On the other hand, when a specific path of desired signals is close to $m\omega_0$, there will be strong signal component in $\hat{\mathbf{h}}_\omega^{(l,k)}[m]$. Without loss of generality, assume the index of that path is b and we have the following approximation:

$$\begin{cases} \hat{\mathbf{h}}_\omega^{(l,k)}[m] = a_b^{(l,k)} \cdot f(\omega_b^{(l,k)} - m\omega_0) + \tilde{\mathbf{n}}_k[m] \\ \mathbf{y}_\omega^{(l)}[m] = \tilde{\mathbf{n}}_{proc}[m]. \end{cases} \tag{33}$$

In this case,

$$\phi_m^{(l,k)} = \frac{\left|\tilde{\mathbf{n}}_{proc}[m]\right|^2}{\left|r_{ds} + \tilde{\mathbf{n}}_k[m]\right|^2} \approx \frac{\left|\tilde{\mathbf{n}}_{proc}[m]\right|^2}{\left|r_{ds}\right|^2}. \tag{34}$$

where $r_{ds} = a_b^{(l,k)} \cdot f(\omega_b^{(l,k)} - m\omega_0)$. $|\tilde{\mathbf{n}}_k[m]|^2$ can be neglected because it's much weaker than r_{ds}. Therefore, the PDF of $\phi_m^{(l,k)}$ can be approximated by exponential distribution, given by

$$f_{ds}\left(\phi_m^{(l,k)}\right) = \frac{|r_{ds}|^2}{\sigma_{proc}^2} \exp\left(-\frac{|r_{ds}|^2}{\sigma_{proc}^2} \cdot \phi_m^{(l,k)}\right). \tag{35}$$

In practical scenarios, $|r_{pc}|^2$ and $|r_{ds}|^2$ can be approximated by $|\hat{\mathbf{h}}_\omega^{(l,k)}[m]|^2$, because they are much stronger than $\tilde{\mathbf{n}}_k[m]$. Besides, $|\tilde{\mathbf{n}}_{proc}[m]|^2$ can be estimated from those components in $\mathbf{y}_\omega^{(l)}$, where neither desired signal nor pilot contamination exist.

E. The Probability that the Main Lobes of Desired Signal and Pilot Contamination Overlap

In previous discussions, we assume the main lobes of desired signals and pilot contamination from adjacent cells do not overlap. As a matter of fact, it is possible for the desired signals to overlap with pilot contaminations from adjacent cells in space, and we will analyze the probability in this section. To simplify notations, we define the following symbols:

$$DS_m = \begin{cases} 1, & \exists b : \left|\omega_b^{(l,k)} - m\omega_0\right| < \omega_0, \\ 0, & \text{otherwise;} \end{cases} \quad \text{and } PC_m = \begin{cases} 1, & \exists b, i : \left|\omega_b^{(l_i,k)} - m\omega_0\right| < \omega_0. \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Given that a strong component of desired signals exists on $m\omega_0$, the probability that strong pilot contamination coexists is $Pr(PC_m = 1|DS_m = 1)$. On the other hand, given that a strong component of pilot contamination on $m\omega_0$, the probability that strong desired signal coexists is $Pr(DS_m = 1|PC_m = 1)$. Then, we have the following theorem.

Theorem 2: When the number of BS antennas is very large, both $Pr(PC_m = 1|DS_m = 1)$ and $Pr(DS_m = 1|PC_m = 1)$ are inversely proportional to BS antenna number. To be specific, we have

$$\begin{aligned} Pr(PC_m = 1|DS_m = 1) &= 12B/M + o(1/M), \\ Pr(DS_m = 1|PC_m = 1) &= 2B/M + o(1/M), \end{aligned} \tag{37}$$

where $o(1/M)$ indicates a component much smaller than $1/M$.

The proof of Theorem 2 can be found in Appendix C. For massive MIMO, M is generally very large, and it is safe for us to assume that desired signals do not overlap with pilot contaminations in space.

F. Computational Complexity

Due to the large number of antennas at BSs, it is important to maintain the computational complexity of channel estimators at an acceptable level. The proposed channel estimation algorithm contains three parts: the MF channel estimation, pilot contamination identification and SF channel estimation. The complexity of MF channel estimator is $\mathcal{O}\{M\tau^2\}$. The second part contains two steps: the first step is to find both desired signals and pilot contaminations from MF channel estimation, while the second step is to pick out the pilot contaminations from the components identified in step one. For every user, the complexity of spatial spectrum analysis is $\mathcal{O}\{M\log_2 M\}$ through FFT. Considering K users per cell and the spatial spectrum analysis of $\mathbf{y}_{proc}^{(l)}$, the overall complexity of the second part is $\mathcal{O}\{(K+1)M\log_2 M\}$. In the third part, we can reorganize Equation (27) as $\hat{\mathbf{h}}_{sf}^{(l,k)} = \hat{\mathbf{h}}_{mf}^{(l,k)} - \mathbf{U}_P^{(l,k)}\left(\mathbf{U}_P^{(l,k)^H}\hat{\mathbf{h}}_{mf}^{(l,k)}\right)$. The complexity of $\mathbf{U}_P^{(l,k)^H}\hat{\mathbf{h}}_{mf}^{(l,k)}$ is $\mathcal{O}\{MP\}$, where P is the number of detected pilot contamination components in space. Considering K users per cell, the complexity of the third part will be $\mathcal{O}\{KMP\}$.

Generally, we have $\log_2 M < \tau = K$, while P should be smaller than 10. Therefore, the total complexity of these three parts combined will be $\mathcal{O}\{MK^2\}$, almost identical to that of MF estimator.

V. Performance Evaluation

In this section, we will conduct numerical simulations to compare the performance of the proposed algorithm with existing ones, in terms of normalized channel estimation error, achieved signal to interference and noise ratio (SINR) and achievable rate.

In the Monte Carlo simulations, we employ the channel model in Equation (3) with $B = 3$. The simulation parameters are almost identical to those employed in Table [1], and they have been specified in I. Same to [1], we consider interferences from all cells whose distance to the target cell is less than eight cell-diameters. Therefore, there are totally 199 cells in our simulations. As show in Table I, we assume a coherence time of 0.5 millisecond, equivalent to seven OFDM symbols in LTE systems, while the frequency smooth interval is equal to 14 sub-carriers. Therefore, the channel can be viewed as constant for 98 time-frequency resource blocks. The simulation results are presented in Figure 5, 6 and 7. Apart from the proposed channel estimation based on spatial filter, we also present the simulation results for data-aided, MF and SVD-based channel estimators.

TABLE I
SIMULATION PARAMETERS

Path Loss Exponent	3.8
Cell Radius	500 m
Cell-Hole Radius	50 m
User Number / Cell	10
Antenna Number / BS	400
Pilot Length (τ)	10
Frequency Smooth Interval	14 Carriers
Coherence Time	0.5 ms
Standard Deviation of Shadowing	8 dB

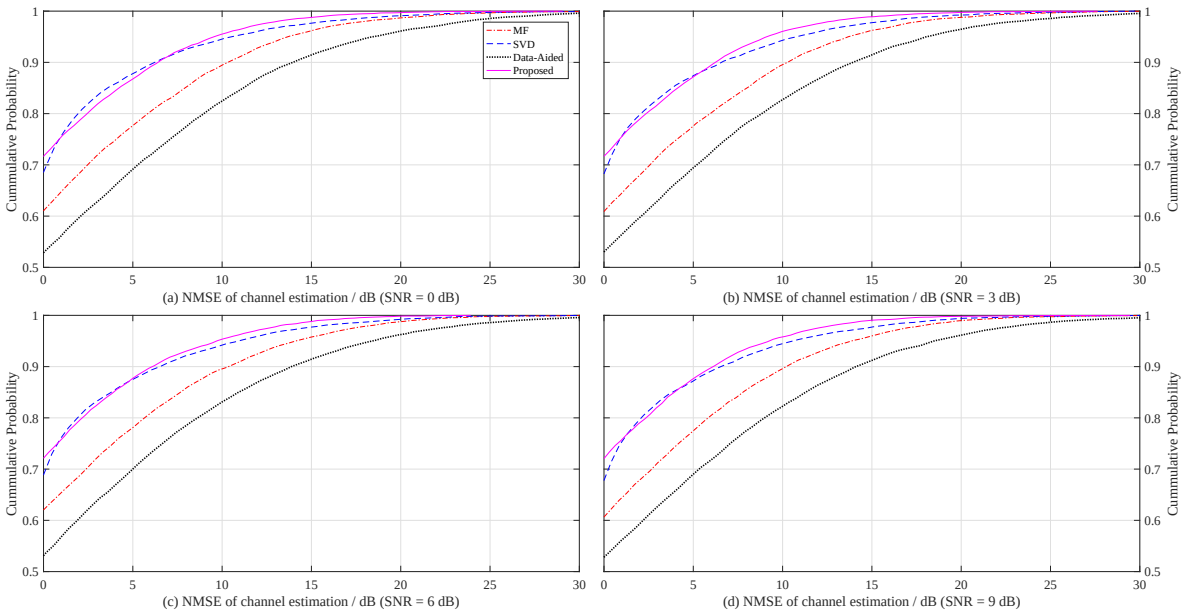

Fig. 5. This figure shows the normalized channel estimation error of different channel estimators when cell-edge user has an SNR of 0, 3, 6, or 9 dB.

In Figure 5, we can see that the data-aided channel estimator shows the worst performance. As a matter of fact, by employing both pilot and data for channel estimation, data-aided channel estimator can suppress pilot contamination to some extent. However, intra-cell interference will be inevitable because the data sequences of different users in the same cell are not totally orthogonal. Because intra-cell interference is generally stronger than inter-cell interference, the data-aided channel estimator does not necessarily improve the overall performance. The SVD-based and the proposed estimators have similar performance in terms of NMSE of channel

estimations. However, for cell-edge users, the proposed estimator has a slight advantage. For example, 95 percent user will experience an NMSE less than 9.5 dB if the proposed estimator is employed; while that value for the SVD-based estimator is 10.5 dB. Besides, we can observe that both the proposed channel estimator and the SVD-based one have a significant gain over the MF channel estimator. For example, only around five percent users will experience an NMSE larger than 10 dB for the proposed and SVD-based algorithm. As comparison, this number is doubled for the channel estimators based on MF.

From Figure 5, we also notice that the performance of channel estimators in different SNR are almost consistent. This is not surprising because the large antenna arrays at BSs can provide significant gains and boost actual SNR. In this case, system performance will be mainly dependent on pilot contamination because noise is negligible. Similar results can be observed in Figure 6 and 7.

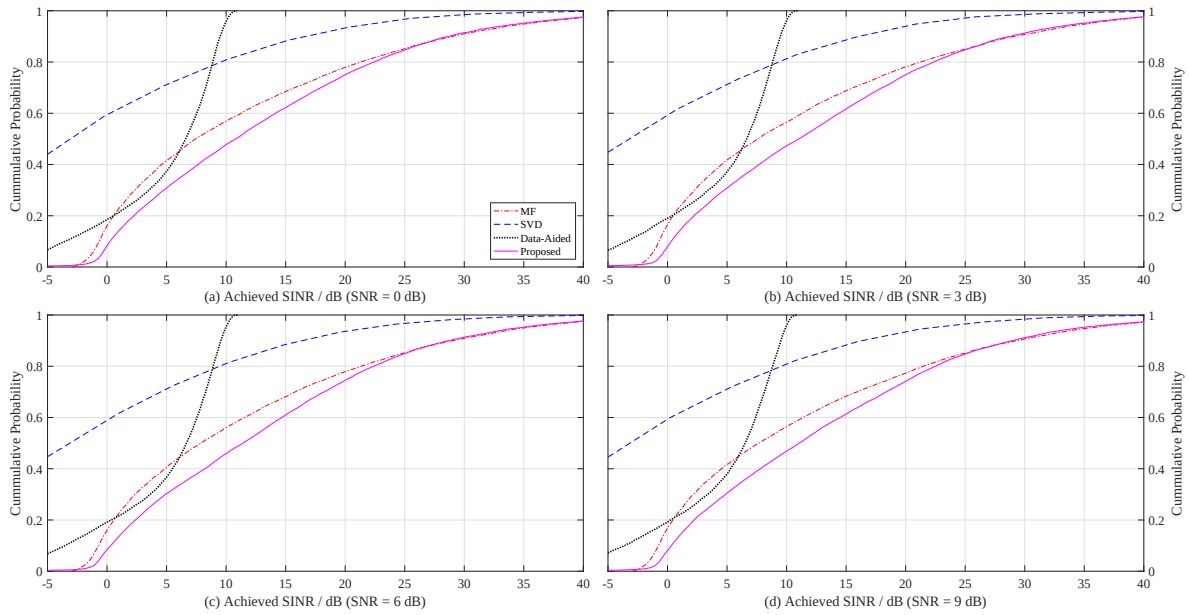

Fig. 6. This figure shows the achieved SINR of different channel estimators, given that zero-forcing is employed for data detection. In different sub-figures, cell-edge users have an SNR of 0, 3, 6, and 9 dB, respectively.

In Figure 6, the achieved SINR is simulated based on zero-forcing symbol detector, which is reported to have similar performance with more complicated detectors, due to the asymptotical orthogonality of different users' channel vectors in the same cell [1]. As we can see, the proposed method outperforms the current ones in terms of achieved SINR. For example, when SNR equals

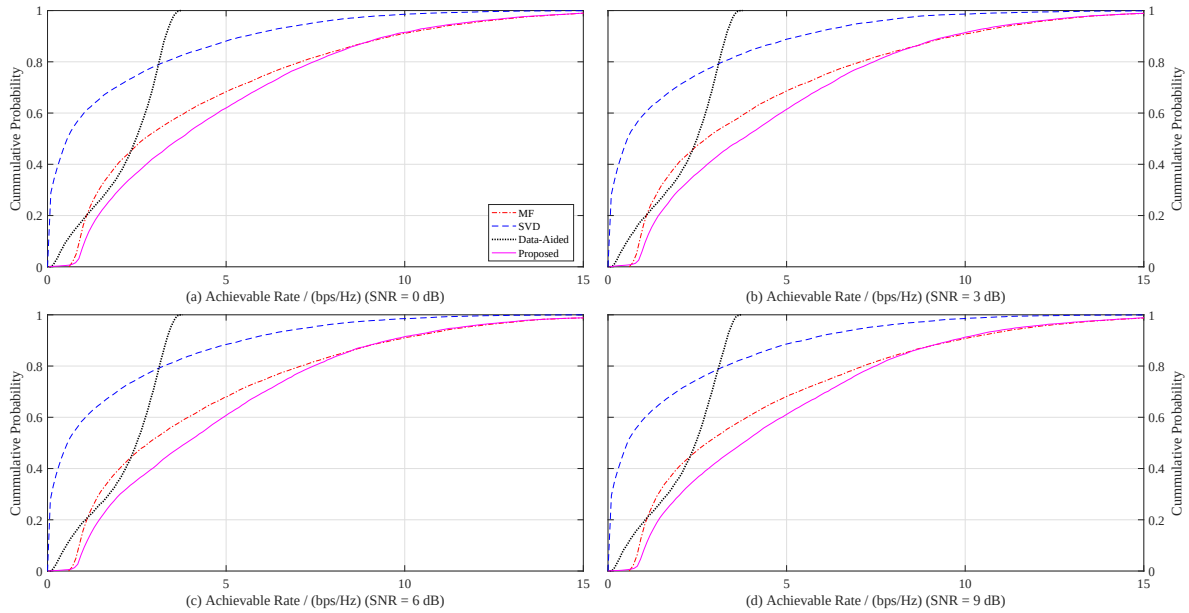

Fig. 7. This figure shows the achievable rates of different channel estimators, given that zero-forcing is employed for data detection. In different sub-figures, cell-edge users have an SNR of 0, 3, 6, and 9 dB, respectively.

0 dB, only 7 percent users will experience an SINR less than 0 dB for the proposed algorithm. However, this number will be doubled for the other algorithms. These users are experiencing low SINR because they are located at cell edges, leading to attenuated signal strength and strong pilot contaminations. As we can observe from Figure 6, the SVD based channel estimator has the worst performance, which seems to be contradictory to the results in Figure 5. As a matter of fact, SVD based algorithm assumes that desired signals are always stronger than pilot contamination. However, pilot contamination can be stronger than desired signals for cell edge users. As a result, the CSI of some users may totally get lost and these users will experience very small SINR. However, the NMSE of their channel estimation results is not necessarily large. For example, when the channel state information of a specific user is totally lost, the NMSE of channel estimation will be around 0 dB, which is not very bad. However, the achieved SINR (in dB) will be infinitely small. Therefore, small NMSE does not necessarily lead to large achieved SINR. By removing the assumption that desired signals are always stronger than pilot contaminations, the proposed algorithm achieves a noticeable gain, because it can effectively identify the spatial signatures of pilot contaminations and eliminate them. As a matter of fact,

when pilot contamination is stronger, the proposed algorithm has a better chance to find it. To be specific, two PDFs of $\phi_m^{(l,k)}$ can be better isolated when pilot contamination is strong on $m\omega_0$. In Figure 7, we can observe similar results, because achievable rate increase monotonically with achieved SINR.

VI. Conclusions

In this paper, we proposed an innovative channel estimation algorithm for massive MIMO systems. The basic idea is to identify the spatial signatures of pilot contaminations from adjacent cells and eliminate them by constructing a spatial filter. The stronger the pilot contamination is, the better chance we have to identify it. This algorithm is especially helpful for cell-edge users, because their signals can be overwhelmed by pilot contaminations from adjacent cells. Compared with the currently available pilot decontamination methods, the proposed method has two major advantages: first, no apriori statistical information is required; second, we do not need to assume that desired signals are always stronger than pilot contaminations. As a matter of fact, the probability that the second condition cannot be fulfilled for cell-edge users is too large to be ignored. This algorithm is evaluated through simulations in channel estimation error, achievable SINR and outage probability. Both simulation results and theoretical analyses are promising.

Appendix A
Proof of the Upper Bound of Residual Error

This part is the proof of the upper bound of the residual error in Eqn (12).

To begin with, we have

$$|\alpha_{mod_M(l+1-r)}|^2 = \frac{1}{M^2}\frac{\sin^2(\beta-1+r)\pi}{\sin^2(\beta-1+r)\pi/M} \leq \frac{1}{M^2}\frac{1}{\sin^2(\beta-1+r)\pi/M} \leq \frac{1}{4(\beta-1+r)^2},$$

where we use the inequality $1 \geq \sin^2 x \geq \frac{4}{\pi^2}x^2$ $(|x| \leq \pi/2)$. Similarly, we have

$$|\alpha_{mod_M(l+r)}|^2 \leq \frac{1}{4(r-\beta)^2}$$

Then, define $\beta_r = |\alpha_{mod_M(l+1-r)}|^2 + |\alpha_{mod_M(l+r)}|^2$ and $f(\beta) = \frac{1}{(r-\beta)^2} + \frac{1}{(r-1+\beta)^2}$.

$\because \frac{\partial f(\beta)}{\partial \beta} < 0, \ \therefore f(\beta) < f(0) = \frac{1}{r^2} + \frac{1}{(r-1)^2}$,

which leads to $\beta_r \leq \frac{1}{4}(\frac{1}{r^2} + \frac{1}{(r-1)^2})$.

When $M = 2F$,

$$Res[\,R\,] = 1 - \sum_{r=1}^{R} \beta_r = \sum_{r=R+1}^{F} \beta_r, \tag{38}$$

When M is very large, we can obtain the following inequality:

$$Res[\,R\,] = \lim_{M\to\infty} \sum_{r=R+1}^{F} \beta_r \leq \frac{1}{4}\left[\lim_{M\to\infty} \sum_{r=R+1}^{F} \frac{1}{r^2} + \lim_{M\to\infty} \sum_{r=R+1}^{F} \frac{1}{(r-1)^2}\right] = \frac{\pi^2}{12} + \frac{1}{4R^2} - \sum_{r=1}^{R} \frac{1}{2r^2}, \tag{39}$$

in which we employed the well acknowledged limit $\lim_{M\to\infty} \sum_{r=1}^{M} 1/r^2 = \frac{\pi^2}{6}$.

When $M = 2F+1$, we have $Res[\,R\,] = \sum_{r=R+1}^{F} \beta_r + |\alpha_{mod_M(l-F)}|^2$, where $|\alpha_{mod_M(l-F)}|^2 \leq 1/F^2$. When M approaches infinity, this component is negligible and the result is identical to that of $M = 2F$.

As a result, regardless of whether M is odd or even, if we take the closest $2R$ components to reestablish $\mathbf{e}\,[\omega]$, the upper bound of residual error can be given as

$$Res[\,R\,] \leq \frac{\pi^2}{12} + \frac{1}{4R^2} - \sum_{r=1}^{R} \frac{1}{2r^2}. \tag{40}$$

APPENDIX B
JUSTIFICATION OF THE DISTRIBUTION OF $\lambda_m^{(l,k)}$

$$\hat{\mathbf{h}}_{\omega}^{(l,k)}[m] = \underbrace{\sum_{b=1}^{B} a_b^{(l,k)} \cdot f(\omega_b^{(l,k)} - m\omega_0) + \sum_{i=1}^{6}\sum_{b=1}^{B} a_b^{(l_i,k)} \cdot f(\omega_b^{(l_i,k)} - m\omega_0)}_{\mathbf{h}_{S,P}^{(l,k)}[m]} + \tilde{\mathbf{n}}_k[m]. \tag{41}$$

By defining $\mathbf{h}_{S,P}^{(l,k)}[m]$ in Equation (41), Equation (19) can be simplified as

$$\lambda_m^{(l,k)} = \left|\mathbf{h}_{S,P}^{(l,k)}[m] + \tilde{\mathbf{n}}_k[m]\right|^2. \tag{42}$$

When $DS_m = 0$ and $PC_m = 0$, $\left|f(\omega_b^{(l_i,k)} - m\omega_0)\right|$'s and $\left|f(\omega_b^{(l,k)} - m\omega_0)\right|$'s will be very small because $\left|\omega_b^{(l_i,k)} - m\omega_0\right| > \omega_0$ and $\left|\omega_b^{(l,k)} - m\omega_0\right| > \omega_0$ hold for any given i and b. Therefore, the absolute value of $\mathbf{h}_{S,P}^{(l,k)}[m]$ will be very small in this case, and we have $\lambda_m^{(l,k)} \approx |\tilde{\mathbf{n}}_k[m]|^2$. Then,

we can conclude that $\lambda_m^{(l,k)}$ follows exponential distribution as

$$f_0\left(\lambda_m^{(l,k)}\right) = \frac{1}{\sigma_0^2} \cdot \exp\left(-\lambda_m^{(l,k)}/\sigma_0^2\right), \tag{43}$$

where $\sigma_0^2 = E\left\{\left|\tilde{\mathbf{n}}_k[m]\right|^2\right\}$.

On the other hand, when $DS_m = 1$ or $PC_m = 1$, $\left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|$ cannot be neglected, and $\lambda_m^{(l,k)}$ can be approximated by

$$\lambda_m^{(l,k)} = \left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2 + 2\Re\left\{\mathbf{h}_{S,P}^{(l,k)}[m] \cdot \tilde{\mathbf{n}}_k[m]^*\right\} + \left|\tilde{\mathbf{n}}_k[m]\right|^2. \tag{44}$$

Given that signal is much stronger than noise, $\left|\tilde{\mathbf{n}}_k[m]\right|^2$ can be neglected and we can obtain Equation (45).

$$\lambda_m^{(l,k)} \approx \left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2 + 2\left(\Re\left\{\mathbf{h}_{S,P}^{(l,k)}[m]\right\} \cdot \Re\{\tilde{\mathbf{n}}_k[m]\} + \Im\left\{\mathbf{h}_{S,P}^{(l,k)}[m]\right\} \cdot \Im\{\tilde{\mathbf{n}}_k[m]\}\right). \tag{45}$$

In Equation (45), the first part is a constant for a given realization of the channel, while the second part follows zero-mean Gaussian distribution. Therefore, the overall PDF of $\lambda_m^{(l,k)}$ is given in Equation (46)

$$f_1\left(\lambda_m^{(l,k)}\right) = \frac{1}{\sqrt{4\pi\left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2\sigma_o^2}} \cdot \exp\left(-\frac{\left(\lambda_m^{(l,k)} - \left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2 - \sigma_o^2\right)^2}{4\left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2\sigma_o^2}\right) \tag{46}$$

In our case, λ_{th} is chosen to minimize the sum of false alarm and miss rates, which means $f_0(\lambda_{th}) = f_1(\lambda_{th})$. Define

$$SNR_o = \left|\mathbf{h}_{S,P}^{(l,k)}[m]\right|^2/\sigma_o^2, \tag{47}$$

and λ_{th} can be given as Equation (21).

The false alarm rate and miss rate will be

$$R_{FA} = \int_{\lambda_{th}}^{\infty} f_0(\lambda)d\lambda = \mathrm{e}^{-\lambda_{th}/\sigma_o^2}. \tag{48}$$

while the miss rate is

$$R_M = \int_{-\infty}^{\lambda_{th}} f_1(\lambda)d\lambda. \tag{49}$$

Replace λ with $t = \frac{\lambda-(SNR_o+1)\sigma_o^2}{2\sqrt{SNR_o\sigma_o^2}}$, and Equation (49) can be rewritten as:

$$R_M = \frac{1}{2} erfc\left(-\frac{\lambda_{th} - (SNR_o+1)\sigma_o^2}{2\sqrt{SNR_o\sigma_o^2}}\right), \tag{50}$$

where $erfc(x)$ is the complementary error function, defined as $erfc(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$. According to Equation (21), The false alarm and miss rates are exclusively dependent on SNR_o, as shown in Equation (22a) and (22b), respectively.

APPENDIX C
PROOF OF THEOREM 2

Given that a strong component of desired signals exists on $m\omega_0$, the probability that strong pilot contamination coexists is bounded by

$$Pr(PC_m = 1|DS_m = 1) = 1 - \left(1 - \frac{2}{M}\right)^{6B}, \tag{51}$$

and the right part can be expanded as Equation (52).

$$\begin{aligned} 1 - \left(1 - \frac{2}{M}\right)^{6B} &= \frac{12B}{M} + \sum_{i=1}^{3B}\left[\binom{6B}{2i+1}\left(\frac{2}{M}\right)^{2i+1} - \binom{6B}{2i}\left(\frac{2}{M}\right)^{2i}\right] - \left(\frac{2}{M}\right)^{6B} \\ &= \frac{12B}{M} + \sum_{i=1}^{3B}\binom{6B}{2i}\left(\frac{2}{M}\right)^{2i}\left(\frac{2(6B-2i-1)}{(2i+1)M} - 1\right) - \left(\frac{2}{M}\right)^{6B}. \end{aligned} \tag{52}$$

In the last line of Equation (52), $\frac{2(6B-2i-1)}{(2i+1)M} - 1$ decreases with i monotonically for $1 \le i \le 3B-1$, which gives us

$$\frac{2}{(6B-1)M} \le \frac{2(2i+1)}{(6B-2i-1)M} \le \frac{4B-2}{M} \ll 1. \tag{53}$$

Therefore, we can decide that

$$Pr(PC_m = 1|DS_m = 1) < \frac{12B}{M}, \tag{54}$$

which means $Pr(PC_m = 1|DS_m = 1)$ decreases with the number of BS antennas at least linearly. When BS antennas is very large, this probability approaches 0.

To find the lower bound of $Pr(PC_m = 1|DS_m = 1)$, we can expand the right part of Equation

(51) as Equation (55).

$$1-\left(1-\frac{2}{M}\right)^{6B}=\frac{12B}{M}-\binom{6B}{2}\left(\frac{2}{M}\right)^{2}+\sum_{i=2}^{3B}\left[\binom{6B}{2i-1}\left(\frac{2}{M}\right)^{2i-1}-\binom{6B}{2i}\left(\frac{2}{M}\right)^{2i}\right]$$
$$=\frac{12B}{M}\left(1-\frac{6B-1}{M}\right)+\sum_{i=2}^{3B}\binom{6B}{2i-1}\left(\frac{2}{M}\right)^{2i-1}\left(1-\frac{6B-2i+1}{Mi}\right). \tag{55}$$

In the last line of Equation (52), $\frac{6B-2i+1}{Mi}$ increases with i monotonically for $2\leq i\leq 3B$, which gives us

$$\frac{1}{3MB}\leq\frac{6B-2i+1}{Mi}\leq\frac{6B-3}{2M}\ll 1. \tag{56}$$

Therefore, we can obtain $Pr(PC_m=1|DS_m=1)>\frac{12B}{M}\left(1-\frac{6B-1}{M}\right)$.

On the other hand, given that a strong component of pilot contamination on $m\omega_0$, the probability that strong desired signal coexists is

$$Pr(PC_m=1|DS_m=1)=1-\left(1-\frac{2}{M}\right)^{B}. \tag{57}$$

Similar to the above analysis, $Pr(PC_m=1|DS_m=1)$ is bounded by bounded by

$$\frac{2B}{M}\left(1-\frac{B-1}{M}\right)<Pr(DS_m=1|PC_m=1)<\frac{2B}{M}. \tag{58}$$

Based on these discussions, Theorem 2 is proved.

APPENDIX D
CONDITIONAL PDF OF $\phi_m^{(l,k)}$

Assume X_c's $(c=1,2,\cdots,C)$ are C independent and normally distributed random variables, and $z=\sum_{c=1}^{C}X_c^2$. Then, z follows noncentral chi-square distribution, and the PDF is given as

$$f_Z(z;k,\mu)=\frac{1}{2}\exp\left(-\frac{z+\mu}{2}\right)\left(\frac{z}{\mu}\right)^{k/4-1/2}I_{k/2-1}(\sqrt{\mu z}) \tag{59}$$

Based on Equation (29), define a random variable y given in Equation (60).

$$y=\phi_m^{(l,k)}2|r_{pc}|^2/\sigma_{proc}^2=\left|\frac{\Re\{r_{pc}\}+\Re\left\{\tilde{\mathbf{n}}_{proc}[m]s_{l_i,k}^*\right\}}{\sigma_{proc}/\sqrt{2}}\right|^2+\left|\frac{\Im\{r_{pc}\}+\Im\left\{\tilde{\mathbf{n}}_{proc}[m]s_{l_i,k}^*\right\}}{\sigma_{proc}/\sqrt{2}}\right|^2 \tag{60}$$

The PDF of $\tilde{\mathbf{n}}_{proc}[m]s_{l_i,k}$ is identical to $\tilde{\mathbf{n}}_{proc}[m]$. Therefore, the right hand side of Equation (60) follows non-central chi-square distribution given by

$$f_Y(y) = \frac{1}{2}\exp\left(-\frac{y+\eta}{2}\right)I_0\left(\sqrt{\eta y}\right) \tag{61}$$

where $\eta = 2|r_{pc}|^2/\sigma^2_{proc}$, and $I_0(\cdot)$ is given in Equation (32). The CDF of y is

$$F_Y(y) = Pr(Y < y) = \int_{-\infty}^{y} f_Y(v)dv. \tag{62}$$

The CDF of $\phi_m^{(l,k)}$ will be

$$F_\Phi\left(\phi_m^{(l,k)}\right) = Pr\left(\Phi < \phi_m^{(l,k)}\right) = Pr(Y/\eta < \phi_m^{(l,k)}) = Pr(Y < \eta\phi_m^{(l,k)}) = F_Y(\eta\phi_m^{(l,k)}), \tag{63}$$

and the PDF can be obtained as

$$f_{pc}(\phi_m^{(l,k)}) = \frac{\partial F_\Phi\left(\phi_m^{(l,k)}\right)}{\partial \phi_m^{(l,k)}} = \frac{\partial F_Y\left(\eta\phi_m^{(l,k)}\right)}{\partial \phi_m^{(l,k)}} = \eta f_Y(\eta\phi_m^{(l,k)}). \tag{64}$$

REFERENCES

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

[2] H. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, Feb. 2013.

[3] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

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

[5] F. Jiang, C. Li, Z. Gong, and R. Su, "Stair matrix and its applications to massive MIMO uplink data detection," *IEEE Transactions on Communications*, to appear in 2018.

[6] F. Jiang, C. Li, and Z. Gong, "Block Gauss-Seidel method based detection in vehicle-to-infrastructure massive MIMO uplink," in *2017 IEEE GLOBECOM*, Signapore, Dec. 2017, pp. 1–5.

[7] ——, "A low complexity soft-output data detection scheme based on Jacobi method for massive MIMO uplink transmission," in *2017 IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–5.

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

[9] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.

[10] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *IEEE International Symposium on Information Theory (ISIT) 2009*, Seoul, South Korea, June 2009, pp. 2184–2188.

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

[12] ——, "Decontaminating pilots in massive MIMO systems," in *2013 IEEE International Conference on Communications (ICC)*, Budapest, Hungary, June 2013, pp. 3170–3175.

[13] F. Fernandes, A. Ashikhmin, and T. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[14] K. Appaiah, A. Ashikhmin, and T. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *2010 IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010, pp. 1–5.

[15] C. Wen, S. Jin, K. Wong, J. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1356–1368, March 2015.

[16] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 525–536, Jan. 2016.

[17] R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[18] H. Ngo and E. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 3249–3252.

[19] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*, Cambridge, MA, USA, July 2012, pp. 1137–1141.

[20] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[21] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3111–3124, June 2014.

[22] H. Yin, L. Cottatellucci, D. Gesbert, R. Müller, and G. He, "Pilot decontamination using combined angular and amplitude based projections in massive MIMO systems," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Stockholm, Sweden, June 2015, pp. 216–220.

[23] Z. Gong, C. Li, and F. Jiang, "Pilot contamination mitigation strategies in massive MIMO systems," *IET Communications*, vol. 11, no. 16, pp. 2403–2409, Nov. 2017.

[24] X. Vu, T. Vu, and T. Quek, "Successive pilot contamination elimination in multiantenna multicell networks," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 617–620, Dec. 2014.

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

[26] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for interference reduction in large scale antenna systems," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Oct. 2013, pp. 226–232.

[27] M. Masood, L. Afify, and T. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 104–118, Jan. 2015.

[28] F. Wan, W. Zhu, and M. Swamy, "Semiblind sparse channel estimation for MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2569–2582, July 2011.

[29] M. Ozdemir and H. Arslan, "Toward real-time adaptive low-rank LMMSE channel estimation of MIMO-OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2675–2678, Oct. 2006.

[30] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Transactions on Communications*, vol. 60, no. 12, Dec. 2012.

[31] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[32] T. Rappaport, S. Sun, and R. Mayzus *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[33] Z. Gong, C. Li, and F. Jiang, "Pilot decontamination for cell-edge users in multi-cell massive MIMO based on spatial filter," in *2018 IEEE International Conference on Communications (ICC)*, Kansas, MO, US, May 2018, pp. 1–5.

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

[35] 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 Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 759–772, Oct. 2014.