Virtual Angular-Domain Channel Estimation for FDD Based Massive MIMO Systems with Partial Orthogonal Pilot Design

Peiyao Zhao, *Student Member, IEEE*, Ke Ma, Zhaocheng Wang, *Senior Member, IEEE*, Sheng Chen, *Fellow, IEEE*

Abstract—This paper proposes an virtual angular-domain channel estimation scheme for massive multiple-input multiple-output systems operating in frequency division duplex (FDD) mode. Different from the conventional scheme where orthogonal pilots are transmitted on different antennas, we propose to transfer the channel estimation problem to the virtual angular domain and utilize the channel sparsity to reduce the training and feedback overhead. An orthogonal matching pursuit with Gram-Schmidt orthogonalization algorithm is proposed to construct the unitary transformation between the spatial domain and the virtual angular domain, which achieves higher sparsity than the existing approaches. Furthermore, we propose to estimate the downlink (DL) dominant angular set, which captures most of the channel power with only a few elements, by utilizing the directional reciprocity of FDD systems, where a calibration algorithm is introduced to handle the different wavelengths of uplink and DL transmissions. Based on the estimated dominant sets, we introduce a partial orthogonal criterion for virtual angular-domain pilot design and further propose two pilot assignment algorithms which minimize pilot overhead and pilot-reuse interference, respectively. Theoretical analyses on pilot overhead and the mean square error (MSE) performance are also presented. Simulation results demonstrate that our proposed virtual angular-domain channel estimation scheme provides excellent MSE performance with much reduced pilot overhead and, consequently, enjoys much larger per-user achievable rate in comparison to the conventional schemes.

Index Terms—Massive MIMO, FDD, channel estimation, pilot design, sparsity, virtual angular domain

I. INTRODUCTION

Massive multi-input multi-output (MIMO) technology [1] has attracted great attention from both academia and industry due to its high spectrum and energy efficiencies [2], achievable by deploying a large number of antennas at the base station (BS). To fully utilize the benefits of massive MIMO, knowledge of downlink (DL) channel state information (CSI) must be available at the BS for the operations including user equipment (UE) scheduling and precoding [3]. Massive MIMO systems with time division duplex (TDD) mode have been extensively investigated [1]–[6], where the DL CSI can be easily acquired from the uplink (UL) channel by exploiting

the channel reciprocity. Nevertheless, TDD massive MIMO systems suffer from performance degradation caused by pilot contamination [5] and calibration error of radio frequency (RF) chains [6]. While pilot contamination can be mitigated or even eliminated by careful pilot designs [7]–[9], calibration error of RF chains will render the DL channel and the UL channel nonreciprocal. Moreover, massive MIMO is expected to be compatible with the current cellular systems, where frequency division duplex (FDD) protocol dominates [4]. Thus exploring an effective DL channel estimation approach is essential for FDD based massive MIMO systems.

The conventional approach to estimate DL CSI is based on orthogonal pilots [10], [11]. However, the pilot and feedback overhead, which scales linearly with the number of antennas at the BS [11], is overwhelmingly high in the context of massive MIMO. Recently, many overhead reduction methods have been proposed for FDD massive MIMO systems. One solution is to exploit the low-rank property of the channel covariance matrix (CCM) [12]–[14]. For example, open-loop and closed-loop strategies were proposed in [12] to reduce the DL channel estimation overhead where the spatial correlation and the temporal correlation are utilized. The work [13] proposed a joint spatial division multiplexing scheme where the dimension of the effective channel is reduced via a pre-beamforming matrix obtained from the CCM. Moreover, the studies [15]–[17] have demonstrated the feasibility of extracting the channel statistics from the outdated-channel-state-information by utilizing time-domain channel coherence. However, acquiring the CCM is challenging in FDD massive MIMO systems, and the overhead of accurate CCM feedback can be excessively high.

Fortunately, massive MIMO channels exhibit sparsity in virtual angular domain due to the small angle spread observed at BS [18]. The experimental results of [19] show that the DL channel reconstructed with only 4 paths captures 85% of the total power on average under both line-of-sight (LoS) and non-line-of-sight outdoor environments, where the BS is equipped with a 8×8 uniform planar array. Utilizing the virtual angular domain channel sparsity, various compressive sensing (CS) based algorithms have been proposed [20]–[24]. Specifically, a non-orthogonal DL pilot design and a distributed sparsity adaptive matching pursuit algorithm were proposed in [21] by exploiting the common sparsity across different subcarriers. Weighted l_1-minimization frameworks were adopted in [22] to enhance the performance of sparse reconstructions over the greedy-based algorithms. The work [23] proposed a closed-loop resource adaptation scheme to handle the uncertainty of channel sparsity level, and the study [24] combined CS

technique and the conventional least squares (LS) algorithm by assuming slow-varying channel statistics. Nevertheless, the CS based algorithms rely heavily on the randomness of non-orthogonal pilots for sparsity recovery, which makes them difficult to be applied in the orthogonal pilot case, and the underlying requirement of perfect measurement feedback from UE to BS is impractical. The imperfection of channel sparsity due to power leakage also degrades the performance.

Previous studies and experiments have shown that the UL channel and the DL channel have similar directional and power properties in FDD systems [25]–[27]. In WINNER II channel models [26], the parameters of each multipath component from both the UL and DL channels, including angle of departures (AoDs), angle of arrivals (AoAs) and path gains, are modeled as the same. With extensive field experiments, the work [19] also reveals that the dominant paths of the DL channel can be obtained from the estimated UL AoAs. Thus, the DL dominant angular set, which is defined as the set that contains the indexes of several largest elements of the virtual angular-domain DL channel, can be estimated from the UL channel with high reliability. Based on the knowledge of dominant angular sets, a user grouping based and a threshold-based DL training methods were proposed in [28] and [29], respectively. However, the method of [28] relies on the assumption of non-overlapping AoAs for different UE groups, while the method of [29] assumes that some common scatterers are shared by all the UE. These assumptions can be unrealistic in practical systems since UE are randomly located.

In this paper, we propose a novel virtual angular-domain channel estimation scheme with partial orthogonal pilot design for FDD massive MIMO systems. Specifically, the DL dominant angular sets are firstly estimated from the UL channel by utilizing the directional reciprocity [25], [26], where an orthogonal matching pursuit with Gram-Schmidt orthogonalization (OMP-GSO) algorithm is proposed for sparsity enhancement, and a calibration algorithm is introduced to handle the different wavelengths of the UL/DL channels. Based on the estimated DL dominant angular sets, virtual angular-domain pilot signals are designed with a partial orthogonal criterion and transmitted to UE. UE then perform DL channel estimation in virtual angular domain and feed the measurements back to the BS for channel reconstruction. Owing to the sparsity of virtual angular-domain channels, the required training and feedback overheads for our approach increase linearly with the sparsity level rather than the number of antennas at the BS. Compared to the approaches of [28], [29], our proposed scheme can handle arbitrary DL dominant angular sets without any assumption on UE locations. If the CCM is available, our proposed scheme can be easily combined with the knowledge of CCM. Specifically, the optimal unitary transformation between the spatial-domain and the virtual angular-domain channels can be obtain from the singular value decomposition (SVD) of the CCM. Simulation results demonstrate that the proposed scheme enjoys reliable channel estimation performance with low training overhead and, consequently, achieves much larger DL throughput in comparison to other existing schemes. The main contributions of our paper are summarized as follows.

- We propose an OMP-GSO algorithm to construct the uni-

tary transformation between the spatial-domain channel and the virtual angular-domain channel, which achieves better sparsity in comparison with the previous methods [28], [30], especially in multi-user scenarios. Consequently, pilot and feedback overhead can be further reduced.

- We propose a partial orthogonal criterion for multi-user angular-domain pilot design, which removes impractical assumptions on UE locations, such as non-overlapping AoAs [28] and common scatterers [29]. Furthermore, we propose two pilot assignment algorithms which minimize the pilot overhead and the pilot-reuse interference, respectively. Theoretical analysis on pilot overhead and the mean square error (MSE) are also conducted.

The rest of the paper is organized as follows. Section II introduces the system model. In Section III, we present the DL dominant angular sets estimation, including the OMP-GSO algorithm and UL/DL calibration. In Section IV, we investigate the virtual angular-domain channel estimation with partial orthogonal pilot design. Numerical simulations are conducted to demonstrate the superior performance of the proposed scheme in Section V, and Section VI concludes the paper.

II. SYSTEM MODEL

As with all the existing works [12]–[14], [20]–[25], [28]–[30], we consider a single cell multi-user massive MIMO system operating in FDD mode, where K single-antenna UE are served simultaneously by the BS equipped with M antennas[1]. The DL transmission is divided into three stages: pilot training, CSI feedback and data transmission.

A. Downlink Training Model

At the pilot training stage, the BS broadcasts the pilot signal matrix $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{s}_1^{\mathrm{T}} \cdots \boldsymbol{s}_M^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M \times \tau}$ for channel estimation, where $\boldsymbol{s}_m \in \mathbb{C}^{1 \times \tau}$ is the length-τ pilot sequence transmitted on the mth antenna element for $1 \leq m \leq M$. The received signal sequence $\boldsymbol{y}_k \in \mathbb{C}^{1 \times \tau}$ at the kth UE for $1 \leq k \leq K$ can be expressed as

$$\boldsymbol{y}_k = \left(\boldsymbol{h}_k^{\mathrm{DL}} \right)^{\mathrm{T}} \boldsymbol{\Phi} + \boldsymbol{n}_k, \tag{1}$$

where $\boldsymbol{h}_k^{\mathrm{DL}} \in \mathbb{C}^{M \times 1}$ is the DL channel vector between the BS and the kth UE, and $\boldsymbol{n}_k \in \mathbb{C}^{1 \times \tau}$ is the complex additive white Gaussian noise (AWGN) sequence with the covariance matrix $\mathsf{E}\{\boldsymbol{n}_k^{\mathrm{H}} \boldsymbol{n}_k\} = \sigma_n^2 \boldsymbol{I}_\tau$. Here, $\mathsf{E}\{\cdot\}$ is the expectation operator and $\boldsymbol{I}_\tau$ is the $\tau \times \tau$ identity matrix.

In the conventional channel estimation scheme, orthogonal pilots are transmitted on different antennas, i.e., $\boldsymbol{s}_i \boldsymbol{s}_j^{\mathrm{H}} = \delta_{i,j}$,

---

[1]For the multi-cell senario, two issues need to be resolved, namely, UL pilot contamination and DL inter-cell interference. UL pilot contamination can be largely mitigated or even removed completely using the schemes originally designed for TDD systems [31]–[33]. There are two solutions dealing with downlink inter-cell interference in DL channel estimation based on frequency division multiplexing (FDM) and time division multiplexing (TDM). In the FDM based method, the neighboring BSs transmit their DL pilots on orthogonal frequency resources to avoid interference, but the pilot overhead becomes very large. In the TDM based method, the pilots of the adjacent cells are transmitted on different time slots to alleviate the interference.

where $\delta_{i,j} = 1$ if $i = j$; otherwise $\delta_{i,j} = 0$. Utilizing the LS algorithm, the DL channel of the kth UE is estimated according to

$$\widehat{\boldsymbol{h}}_k^{\mathrm{DL}} = \left(\boldsymbol{\Phi}^*\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Phi}^*\boldsymbol{y}_k^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{\Phi}^*$ is the conjugate of $\boldsymbol{\Phi}$. The estimated channel $\widehat{\boldsymbol{h}}_k^{\mathrm{DL}}$ is then fed back to the BS through quantization codebooks, such as the precoding matrix index (PMI) in LTE [10]. However, the orthogonality of pilots requires $\tau \geq M$, which imposes huge training and feedback overhead when the number of antennas at the BS M is large.

B. FDD Channel Model

Previous studies and experiments have shown that the UL and DL channels of a FDD system share similar directivity [19], [25], [26]. Specifically, we adopt a geometric narrow-band channel model with P paths, where the UL and DL channels have the same small-scale parameters, including the AoA/AoD and gain of each path [26]. However, due to different UL and DL wavelengths, the transmitted UL and DL signals experience different phase changes after propagating the same distance, which is modeled as independent random phases. We assume perfect power control at the BS such that the effects of large scale fading and shadow fading are taken care of. By omitting the UE subscript index, the UL channel $\boldsymbol{h}^{\mathrm{UL}} \in \mathbb{C}^{M\times 1}$ and DL channel $\boldsymbol{h}^{\mathrm{DL}} \in \mathbb{C}^{M\times 1}$ can be represented respectively by

$$\boldsymbol{h}^{\mathrm{UL}} = \sqrt{\frac{M}{P}} \sum_{p=1}^{P} \alpha_p e^{\mathrm{j}\psi_p^{\mathrm{UL}}} \boldsymbol{a}^{\mathrm{UL}}(\phi_p, \theta_p), \tag{3}$$

$$\boldsymbol{h}^{\mathrm{DL}} = \sqrt{\frac{M}{P}} \sum_{p=1}^{P} \alpha_p e^{\mathrm{j}\psi_p^{\mathrm{DL}}} \boldsymbol{a}^{\mathrm{DL}}(\phi_p, \theta_p), \tag{4}$$

where the complex gain of the pth path α_p follows the complex Gaussian distribution with zero mean and unity power, denoted by $\alpha_p \sim \mathcal{CN}(0,1)$, ψ_p^{UL} and ψ_p^{DL} are the independent random phases of the pth path in the UL and DL channels, respectively, which are both uniformly and randomly distributed in $[0,\ 2\pi]$, while ϕ_p and θ_p denote the horizontal and vertical AoA/AoD of the pth path in UL/DL, and $\boldsymbol{a}^{\mathrm{UL}}(\phi_p, \theta_p)/\boldsymbol{a}^{\mathrm{DL}}(\phi_p, \theta_p) \in \mathbb{C}^{M\times 1}$ is the steering vector at the pth path direction of UL/DL. We assume that ϕ_p and θ_p follow the uniform distributions in the ranges of $[\phi_{\mathrm{c}}-\Delta\phi,\ \phi_{\mathrm{c}}+\Delta\phi]$ and $[\theta_{\mathrm{c}}-\Delta\theta,\ \theta_{\mathrm{c}}+\Delta\theta]$, respectively, where $\Delta\phi$ and $\Delta\theta$ are the horizontal and vertical angle spreads, while ϕ_{c} and θ_{c} denote the respective central directions.

For a uniform linear array (ULA), the steering vector is independent of the vertical angle θ, and it can be written as

$$\boldsymbol{a}_{\mathrm{ULA}}(\phi) = \frac{1}{\sqrt{M}} \left[1\ e^{\mathrm{j}\frac{2\pi}{\lambda}d\sin\phi} \cdots e^{\mathrm{j}\frac{2\pi}{\lambda}d(M-1)\sin\phi} \right]^{\mathrm{T}}, \tag{5}$$

where λ is the wavelength, d is the antenna spacing, and we have omitted the superscript for UL/DL. Note that the wavelengths λ^{UL} and λ^{DL} for the UL and DL channels are different. Without loss of generalization, the antenna spacing at the BS is set to $d = \lambda^{\mathrm{UL}}/2$.
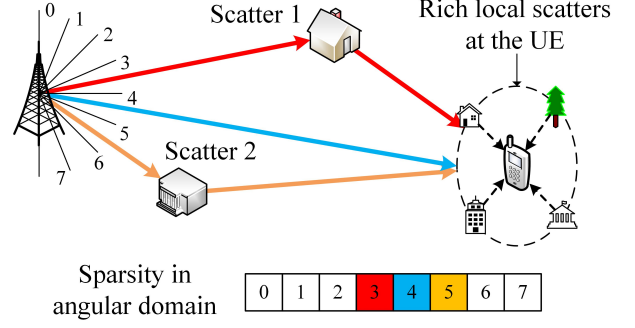

Fig. 1. Illustration of the virtual angular-domain channel sparsity.

C. Virtual Angular-Domain Channel Sparsity

Since the BS is usually elevated high with few scatterers around, the angle spread observed at the BS is small [21], [28]. Intuitively, most of the paths come from a few angular intervals nearby the LoS direction at the BS side, as illustrated in Fig. 1. Consequently, in angular domain, the channel coefficients within a small angular range of the LoS are dominant, while the other channel coefficients outside this range tends to be zeros when the antenna number M is large, which indicates the angular-domain sparsity of massive MIMO channels. Specifically, the spatial-domain channel $\boldsymbol{h}$ can be transformed into the angular-domain channel $\widetilde{\boldsymbol{h}}$ via an unitary transform matrix $\boldsymbol{A}$ as $\widetilde{\boldsymbol{h}} = \boldsymbol{A}\boldsymbol{h}$, which exhibits sparsity. In another words, $\boldsymbol{h}$ can be expressed as

$$\boldsymbol{h} = \boldsymbol{A}^{\mathrm{H}}\widetilde{\boldsymbol{h}} = \sum_{i=1}^{M} \left[\widetilde{\boldsymbol{h}}\right]_i \boldsymbol{b}_i, \tag{6}$$

where $\left[\widetilde{\boldsymbol{h}}\right]_i$ is the ith element of $\widetilde{\boldsymbol{h}}$, $\boldsymbol{b}_i = \left(\left[\boldsymbol{A}\right]_{[i:]}\right)^{\mathrm{H}} \in \mathbb{C}^{M\times 1}$ is the ith orthogonal basis, and $\left[\boldsymbol{A}\right]_{[i:]}$ is the ith row of $\boldsymbol{A}$. Due to the sparsity of the virtual angular-domain channel, $\boldsymbol{h}$ can be approximated by the expansion of N orthogonal bases, where $N \ll M$, as

$$\boldsymbol{h} \approx \sum_{i\in\Omega} \left[\widetilde{\boldsymbol{h}}\right]_i \boldsymbol{b}_i. \tag{7}$$

In (7), Ω denotes the dominant angular set, which is defined as the set that contains the indexes of the N largest elements of the virtual angular-domain channel $\widetilde{\boldsymbol{h}}$, i.e.,

$$\left|\left[\widetilde{\boldsymbol{h}}\right]_i\right| \geq \left|\left[\widetilde{\boldsymbol{h}}\right]_j\right|,\ \forall i\in\Omega, \forall j\notin\Omega. \tag{8}$$

Consequently, the DL training model (1) can be rewritten as

$$\boldsymbol{y}_k = \left(\boldsymbol{A}^{\mathrm{H}}\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\right)^{\mathrm{T}}\boldsymbol{\Phi} + \boldsymbol{n}_k = \left(\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\right)^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}} + \boldsymbol{n}_k, \tag{9}$$

where $\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}$ is the the kth UE's virtual angular-domain DL channel, and $\widetilde{\boldsymbol{\Phi}} = \boldsymbol{A}^*\boldsymbol{\Phi}$ defines the virtual angular-domain pilot signal matrix. Similar to antenna ports, we can define the angular port n, $1 \leq n \leq N$, as the virtual port that transmits the virtual angular-domain pilot signal $\left[\widetilde{\boldsymbol{\Phi}}\right]_{[n:]}$. By exploiting the sparsity of $\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}$ and the dominant angular set information Ω_k^{DL}, the virtual angular-domain pilot signal matrix $\widetilde{\boldsymbol{\Phi}}$ can be carefully designed for pilot reduction.

Note that the sparsity of the virtual angular-domain channel relies on the unitary transform matrix $\boldsymbol{A}$. The optimal design

of $\boldsymbol{A}$ can be obtained by performing the SVD on the CCM [13], but the CCM is difficult to obtain in FDD massive MIMO systems [34]. Benefiting from the structure of steering vectors, the discrete Fourier transform (DFT) matrices are widely adopted in previous works [21], [29], [30]. Specifically, for an ULA with M antennas, $\boldsymbol{A} = \boldsymbol{F}_M$, where $\boldsymbol{F}_M$ is the size $M \times M$ DFT matrix whose ith row jth column element is given by $\left[\boldsymbol{F}_M\right]_{i,j} = e^{-\mathrm{j}\frac{2\pi}{M}ij}$. However, due to the mismatch between continuous AoAs/AoDs and discrete sampling of DFT, power leakage usually occurs, leading to degradation of virtual angular-domain sparsity.

III. DOMINANT ANGULAR SETS ESTIMATION

This section is devoted to how to estimate the DL dominant angular sets from the UL channel by utilizing the directional reciprocity of FDD systems. Specifically, we propose the OMP-GSO algorithm to construct the unitary transform matrix $\boldsymbol{A}$, which achieves remarkable sparsity enhancement in comparison to the existing methods. We also introduce a calibration method to handle the different wavelengths of the UL and DL channels.

A. Unitary Transform Matrix Construction

To provide insights on how to construct unitary transform matrix $\boldsymbol{A}$, we introduce two metrics for measuring the sparsity of the virtual angular domain channel $\widetilde{\boldsymbol{h}}$ achieved with $\boldsymbol{A}$, called the η-sparsity and N-power. In the following, $0 \leq \eta \leq 1$, $N \in \mathbb{N}$ and $1 \leq N \leq M$, and $\Omega \subseteq \{1, 2, \cdots, M\}$.

Definition 1. The η-sparsity of vector $\boldsymbol{v} \in \mathbb{C}^{M \times 1}$ is defined as the minimum size of the index set Ω that contains more than the η portion of the total power $\|\boldsymbol{v}\|^2$, denoted by N_η. Specifically,

$$N_\eta = \min |\Omega| \text{ s.t. } \frac{\sum_{i \in \Omega} |[\boldsymbol{v}]_i|^2}{\|\boldsymbol{v}\|^2} \geq \eta. \tag{10}$$

Definition 2. The N-power of vector $\boldsymbol{v} \in \mathbb{C}^{M \times 1}$ is defined as the percentage of the total power $\|\boldsymbol{v}\|^2$ captured by the largest N elements of $\boldsymbol{v}$, denoted by η_N. Specifically,

$$\eta_N = \max \frac{\sum_{i \in \Omega} |[\boldsymbol{v}]_i|^2}{\|\boldsymbol{v}\|^2}, \text{ s.t. } |\Omega| = N. \tag{11}$$

It is clear that smaller η-sparsity and larger N-power indicate better sparsity.

To enhance the sparsity of the virtual angular-domain channel, the work [28] proposed a spatial rotation strategy by modifying the unitary transform matrix as $\boldsymbol{A} = \boldsymbol{F}_M \boldsymbol{\Psi}(\varphi)$, where

$$\boldsymbol{\Psi}(\varphi) = \mathrm{diag}\left\{1, e^{\mathrm{j}\varphi}, \cdots, e^{\mathrm{j}(M-1)\varphi}\right\} \tag{12}$$

is a diagonal matrix with $\varphi \in \left(-\frac{\pi}{M}, \frac{\pi}{M}\right]$ as the spatial rotation parameter. As a result, the original orthogonal bases $\boldsymbol{b}_i = \left([\boldsymbol{F}_M]_{[i:]}\right)^{\mathrm{H}}$ will be rotated simultaneously with the same angle φ while maintaining the orthogonality. However, its performance gain over the conventional DFT method is very limited since the different paths tend to require different optimal spatial rotation parameters. Moreover, this method hardly works in the multi-user scenario because it is very

difficult if not impossible to find a common rotation parameter for randomly located UEs.

The underlying reason is that the orthogonal condition of the bases restricts the searching to a one-dimension space φ. To construct a unitary matrix that achieves better sparsity, we should remove the orthogonal constraints on bases to minimize the approximation error and calculate the unitary matrix by the Gram-Schmidt orthogonalization (GSO).

Lemma 1. Assume that the channel $\boldsymbol{h}$ can be approximated by the N non-orthogonal but linearly independent bases $\{\boldsymbol{b}_i\}_{i=1}^N$, that is,

$$\boldsymbol{h} = \sum_{i=1}^N \alpha_i \boldsymbol{b}_i + \boldsymbol{h}_e, \tag{13}$$

with the residual error $\boldsymbol{h}_e$. By performing the GSO on $\{\boldsymbol{b}_i\}_{i=1}^N$, a set of the normalized orthogonal bases $\{\boldsymbol{b}_i^{\mathrm{o}}\}_{i=1}^N$ are obtained, and $\boldsymbol{h}$ can be approximated by the expansion of the orthogonal bases $\{\boldsymbol{b}_i^{\mathrm{o}}\}_{i=1}^N$ as

$$\boldsymbol{h} = \sum_{i=1}^N \alpha_i^{\mathrm{o}} \boldsymbol{b}_i^{\mathrm{o}} + \boldsymbol{h}_e^{\mathrm{o}}, \tag{14}$$

with the residual error $\boldsymbol{h}_e^{\mathrm{o}}$, where $\alpha_i^{\mathrm{o}} = \left(\boldsymbol{b}_i^{\mathrm{o}}\right)^{\mathrm{H}} \boldsymbol{h}$. Then we have:

1) $\|\boldsymbol{h}_e^{\mathrm{o}}\|^2 \leq \|\boldsymbol{h}_e\|^2$.
2) The sparsity of the virtual angular-domain channel $\widetilde{\boldsymbol{h}}^{\mathrm{o}}$ obtained via the base set $\{\boldsymbol{b}_i^{\mathrm{o}}\}_{i=1}^N$ cannot be lower than the sparsity of the virtual angular-domain channel $\widetilde{\boldsymbol{h}}$ obtained via the base set $\{\boldsymbol{b}_i\}_{i=1}^N$.

Proof: Denote $\mathcal{V} = \mathrm{span}\left(\{\boldsymbol{b}_i\}_{i=1}^N\right)$ as the subspace spanned by the bases $\{\boldsymbol{b}_i\}_{i=1}^N$. Since the GSO procedure does not change this subspace, we also have $\mathcal{V} = \mathrm{span}\left(\{\boldsymbol{b}_i^{\mathrm{o}}\}_{i=1}^N\right)$. By projecting $\boldsymbol{h}_e$ onto the subspace $\mathcal{V}$, we obtain the projected vector $\boldsymbol{h}_e^{\parallel} \in \mathcal{V}$ and the error vector $\boldsymbol{h}_e^{\perp} \perp \mathcal{V}$ with $\boldsymbol{h}_e = \boldsymbol{h}_e^{\parallel} + \boldsymbol{h}_e^{\perp}$. Thus (13) can be rewritten as

$$\boldsymbol{h} = \left(\sum_{i=1}^N \alpha_i \boldsymbol{b}_i + \boldsymbol{h}_e^{\parallel}\right) + \boldsymbol{h}_e^{\perp}. \tag{15}$$

Since the GSO procedure guarantees that $\boldsymbol{h}_e^o \perp \mathcal{V}$, we have

$$\begin{aligned} \|\boldsymbol{h}_e\|^2 &= \left(\boldsymbol{h}_e^{\parallel} + \boldsymbol{h}_e^{\perp}\right)^{\mathrm{H}} \left(\boldsymbol{h}_e^{\parallel} + \boldsymbol{h}_e^{\perp}\right) \\ &= \left\|\boldsymbol{h}_e^{\parallel}\right\|^2 + \left\|\boldsymbol{h}_e^{\perp}\right\|^2 \geq \left\|\boldsymbol{h}_e^{\perp}\right\|^2 = \left\|\boldsymbol{h}_e^{\mathrm{o}}\right\|^2. \end{aligned} \tag{16}$$

The η-sparsity metrics associated with the non-orthogonal base set $\{\boldsymbol{b}_i\}_{i=1}^N$ and the orthogonal base set $\{\boldsymbol{b}_i^{\mathrm{o}}\}_{i=1}^N$ are calculated as $\eta_N = 1 - \frac{\|\boldsymbol{h}_e\|^2}{\|\boldsymbol{h}\|^2}$ and $\eta_N^{\mathrm{o}} = 1 - \frac{\|\boldsymbol{h}_e^{\mathrm{o}}\|^2}{\|\boldsymbol{h}\|^2}$, respectively. Since $\|\boldsymbol{h}_e^{\mathrm{o}}\|^2 \leq \|\boldsymbol{h}_e\|^2$, we have $\eta_N^{\mathrm{o}} \geq \eta_N$, and this completes the proof. ∎

Inspired by Lemma 1, we start by constructing a set of non-orthogonal but linearly independent bases using the orthogonal matching pursuit (OMP) algorithm. Then, we perform the GSO procedure to obtain the unitary matrix $\boldsymbol{A}$. The proposed algorithm is therefore called the OMP-GSO. Specifically, we adopt the uniformly sampled steering vectors as the dictionary $\mathcal{D} = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_{Q-1}\}$, where Q is the dictionary size. Note that the sampling is performed in $\sin \varphi$ rather than in the

Algorithm 1 OMP-GSO algorithm

Input: Dictionary $\mathcal{D} = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_{Q-1}\}$, channels of K UE $\boldsymbol{h}_k \in \mathbb{C}^{M \times 1}$, $1 \leq k \leq K$, number of iterations N_{iter}, $1 \leq N_{\text{iter}} \leq M$.

Output: Unitary transform matrix $\boldsymbol{A}$.

1: Initialize residuals $\boldsymbol{r}_k^{(0)} = \boldsymbol{h}_k$, $1 \leq k \leq K$.
2: Initialize the matrix of chosen bases $\boldsymbol{B}$ as an empty matrix.
3: **for** $i = 1$ **to** N_{iter} **do**
4:    Find $\boldsymbol{b}_i \in \mathcal{D}$ with the maximum inner product $\left| \boldsymbol{b}_i^{\text{H}} \boldsymbol{r}_k^{(i-1)} \right|$:

$$\boldsymbol{b}_i = \arg\max_{\boldsymbol{d}} \left\{ \max_{1 \leq k \leq K} \left| \boldsymbol{d}^{\text{H}} \boldsymbol{r}_k^{(i-1)} \right| \right\}, \text{ s.t. } \boldsymbol{d} \in \mathcal{D}.$$

5:    Update the matrix of chosen bases as $\boldsymbol{B} = [\boldsymbol{B} \ \boldsymbol{b}_i]$.
6:    Update the spanned coefficients $\boldsymbol{s}_k^{(i)} = \left(\boldsymbol{B}^{\text{H}} \boldsymbol{B}\right)^{-1} \boldsymbol{B}^{\text{H}} \boldsymbol{h}_k$, $1 \leq k \leq K$.
7:    Update residuals $\boldsymbol{r}_k^{(i)} = \boldsymbol{h}_k - \boldsymbol{B} \boldsymbol{s}_k^{(i)}$, $1 \leq k \leq K$.
8: **end for**
9: Randomly generate matrix $\boldsymbol{C} \in \mathbb{C}^{M \times (M - N_{\text{iter}})}$, and obtain $\boldsymbol{B}_{\text{f}} = [\boldsymbol{B} \ \boldsymbol{C}]$ such that $\boldsymbol{B}_{\text{f}}$ is a full-rank square matrix with size $M \times M$.
10: Perform the GSO procedure on the columns of $\boldsymbol{B}_{\text{f}}$ to obtain the normalized orthogonal bases which constitute the unitary transform matrix $\boldsymbol{A}$.

physical angle φ. Thus the qth vector $\boldsymbol{d}_q$, $0 \leq q \leq Q - 1$, is expressed as

$$\boldsymbol{d}_q = \left[1 \ e^{\text{j} \frac{2\pi}{Q} q} \cdots e^{\text{j} \frac{2\pi}{Q} q(M-1)} \right]^{\text{T}}. \quad (17)$$

The detailed OMP-GSO algorithm is presented in Algorithm 1, where the criterion adopted in step 4 of Algorithm 1 is determined empirically.

In the step of finding the basis with the maximum inner product (line 4), we consider the residuals of all the UE, which ensures that every UE's spatial feature can be captured. Compared to the spatial rotation method [28], which relies only on a one-dimensional search space, the proposed OMP-GSO enjoys much better performance owing to a much higher-dimensional search space. We use an example with $M = 128$ and $P = 20$ to compare the achievable sparsity performance of the conventional DFT method [21], [30], the spatial rotation method [28] and our proposed OMP-GSO method in the two scenarios of $K = 1$ and $K = 4$. We set the dictionary size to $Q = 8M$ and the number of iterations to $N_{\text{iter}} = M$ for our OMP-GSO algorithm.

Fig. 2 depicts the η-sparsity performance of these three methods as the functions of the angle spread $\Delta\phi$. As with all the other simulation results presented in this paper, the results of Fig. 2 are obtained by averaging over a sufficiently large number of channel realizations. In each realization, the central AoAs of the K UE are independently generated from the uniform distribution in the range of $[0, \ 2\pi]$. From Fig. 2, it is seen that the sparsity performance gain of the spatial rotation strategy over the DFT method almost vanishes in the multi-user scenario of $K = 4$. This is not surprising, since the spatial
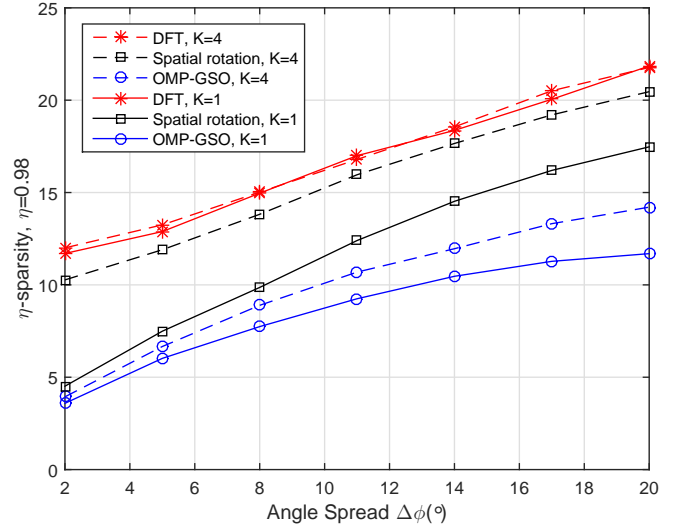

Fig. 2. Comparison of the η-sparsity performance of three methods as the functions of angle spread $\Delta\phi$, given $\eta = 0.98$.

rotation method is unable to meet the requirement that different UE need different optimal spatial rotation parameters. By contrast, the proposed OMP-GSO algorithm attains significant performance gain in terms of achieved sparsity over the other two methods, particularly when the angle spread is large. This demonstrates that the proposed scheme is capable of capturing the channel's spatial feature much more effectively.

The computational complexity of Algorithm 1 comprises the following two parts.

1) OMP: According to [35], the complexity of the OMP is on the order of $\mathsf{O}\left(2 N_{\text{iter}} M Q + 3 N_{\text{iter}}^2 M\right)$ for the sparse reconstruction of a single vector. In Algorithm 1, OMP operations are applied to all the K UEs. Further considering that $N_{\text{iter}} \leq M \leq Q$, the computational complexity involved in the OMP is on the order of $\mathsf{O}\left(N_{\text{iter}} M Q K\right)$.
2) GSO: The complexity of the GSO procedure for M-dimensional space is $\mathsf{O}(M^3)$.

Thus the computational complexity of the proposed OMP-GSO algorithm is $\mathsf{O}\left(M^3 + N_{\text{iter}} M Q K\right)$. As a comparison, the CCM-based methods require $\mathsf{O}(M^3)$ complexity for SVD, which is on the same order of our proposed scheme. On the other hand, the spatial rotation strategy [28] involves $\mathsf{O}(M^2)$ complexity by performing a one-dimensional search on the rotation parameter. However, acquiring the CCM is challenging in FDD massive MIMO systems, while the performance of the spatial rotation strategy degrades significantly in multi-user scenarios with randomly located UEs, as demonstrated in Fig. 2, since it is difficult to find a common rotation parameter. Noting that the number of antennas M and the number of UE K are fixed, the complexity of Algorithm 1 is controlled by the number of iterations N_{iter} and the dictionary size Q. By adjusting N_{iter} and Q, different tradeoffs between performance and complexity can be achieved for our proposed method.

Fig. 3 shows the performance of Algorithm 1 as a function of N_{iter} for three cases of K. It is seen that the algorithm converges to the sparsity level of $N_\eta \approx 8$ with $N_{\text{iter}} \approx 10$ iterations for the single-UE case, and it converges to $N_\eta \approx 9$
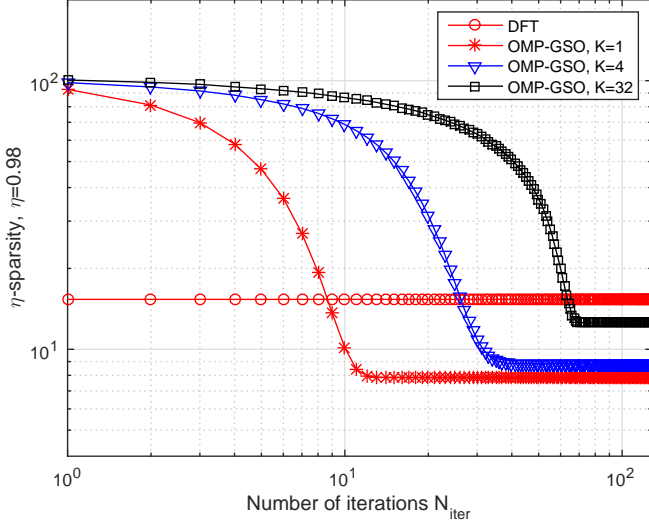
Fig. 3. Performance of the OMP-GSO algorithm as a function of N_{iter}. $\Delta\phi = 8°$ and the other parameters are the same as in Fig. 2.

with $N_{\text{iter}} \approx 40$ iterations for the case of $K = 4$, while the algorithm converges to $N_\eta \approx 16$ with $N_{\text{iter}} \approx 70$ iterations for the case of $K = 32$. Empirically, it seems that the number of iterations for Algorithm 1 can be set to $N_{\text{iter}} \ll M$. In particular, if K times the sparsity level is much smaller than M, setting N_{iter} to K times the sparsity level is sufficient to ensure the convergence of the algorithm.

Fig. 4 depicts the impacts of the dictionary size Q on the performance of Algorithm 1 for two cases of K. With $Q/M = 1$, the OMP-GSO degenerates to the DFT method since the dictionary $\mathcal{D}$ becomes exactly the DFT matrix. The results of Fig. 4 show that $Q/M = 8$ is generally sufficient for the OMP-GSO to converge to near optimal performance.
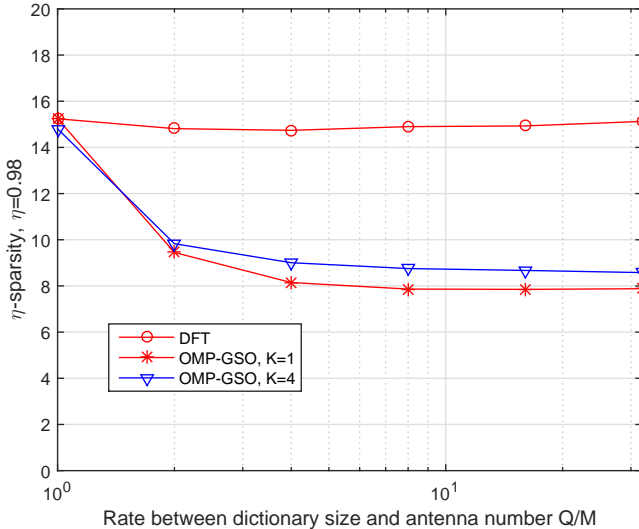

Fig. 4. Performance of the OMP-GSO algorithm as a function of the dictionary size Q. $\Delta\phi = 8°$ and the other parameters are the same as in Fig. 2

B. Downlink Dominant Angular Sets Estimation

In the UL training stage, all the UE simultaneously transmit their UL pilots, and the received signal $\boldsymbol{Y}^{\text{UL}} \in \mathbb{C}^{M \times \tau^{\text{UL}}}$ at the BS is expressed as

$$\boldsymbol{Y}^{\text{UL}} = \sum_{k=1}^{K} \boldsymbol{h}_k^{\text{UL}} \boldsymbol{s}_k^{\text{UL}} + \boldsymbol{n}^{\text{UL}}, \qquad (18)$$

where $\boldsymbol{h}_k^{\text{UL}} \in \mathbb{C}^{M \times 1}$ and $\boldsymbol{s}_k^{\text{UL}} \in \mathbb{C}^{1 \times \tau^{\text{UL}}}$ are the UL channel and orthogonal pilot of the kth UE, respectively, and τ^{UL} is the pilot length, while $\boldsymbol{n}^{\text{UL}} \in \mathbb{C}^{M \times \tau^{\text{UL}}}$ is the UL AWGN matrix. The required pilot length τ^{UL} for UL training is $\tau^{\text{UL}} \geq K$. The UL training problem has been extensively discussed in [31]–[33], and the LS estimate of $\boldsymbol{h}_k^{\text{UL}}$ is given by

$$\widehat{\boldsymbol{h}}_k^{\text{UL}} = \frac{\boldsymbol{Y}^{\text{UL}} \big(\boldsymbol{s}_k^{\text{UL}}\big)^{\text{H}}}{\boldsymbol{s}_k^{\text{UL}} \big(\boldsymbol{s}_k^{\text{UL}}\big)^{\text{H}}}. \qquad (19)$$

From the estimated $\widehat{\boldsymbol{h}}_k^{\text{UL}}$, $1 \leq k \leq K$, the unitary transform matrix $\boldsymbol{A}^{\text{UL}}$ for the UL channels can be constructed using the OMP-GSO algorithm, and the UL dominant angular set Ω_k^{UL} of the kth UE contains the indexes of the N largest elements of $\widehat{\bar{\boldsymbol{h}}}_k^{\text{UL}} = \boldsymbol{A}^{\text{UL}} \widehat{\boldsymbol{h}}_k^{\text{UL}}$.

We can construct the DL unitary transform matrix $\boldsymbol{A}^{\text{DL}}$ and to estimate the DL dominant angular sets Ω_k^{DL} by utilizing directional reciprocity. The unitary transform matrices and dominant angular sets of the UL and DL channels tend to be the same due to directional reciprocity. But simply setting $\boldsymbol{A}^{\text{DL}} = \boldsymbol{A}^{\text{UL}}$ is inadvisable, since the steering vectors of the UL and DL channels have different wavelengths. We propose a calibration algorithm to handle the effects of different UL and DL wavelengths. Considering the UL steering vector of a ULA $\boldsymbol{a}_{\text{ULA}}^{\text{UL}}(\phi)$ for the AoA direction ϕ, its matched basis in the dictionary $\mathcal{D}$ is $\boldsymbol{d}_q$. From (5) and (17), we have $e^{j \frac{2\pi d}{\lambda^{\text{UL}}} \sin \phi} = e^{j \frac{2\pi}{Q} q}$, which can be expressed as

$$\sin \phi = \begin{cases} \frac{q \lambda^{\text{UL}}}{Q d}, & 0 \leq q \leq \frac{Q}{2}, \\ \frac{(q-Q) \lambda^{\text{UL}}}{Q d}, & \frac{Q}{2} < q \leq Q - 1. \end{cases} \qquad (20)$$

Thus the corresponding DL steering vector $\boldsymbol{a}_{\text{ULA}}^{\text{DL}}(\phi)$ for the AoD direction ϕ can be obtained as $\boldsymbol{a}_{\text{ULA}}^{\text{DL}}(\phi) = \boldsymbol{d}_q^{\text{c}}$, where $\boldsymbol{d}_q^{\text{c}}$ is the calibrated basis expressed as

$$\boldsymbol{d}_q^{\text{c}} = \begin{cases} \left[1\ e^{j \frac{2\pi \lambda^{\text{UL}}}{Q \lambda^{\text{DL}}} q} \cdots e^{j \frac{2\pi \lambda^{\text{UL}}}{Q \lambda^{\text{DL}}} q(M-1)} \right], \\ \qquad\qquad 0 \leq q \leq \frac{Q}{2}, \\ \left[1\ e^{j \frac{2\pi \lambda^{\text{UL}}}{Q \lambda^{\text{DL}}} (q-Q)} \cdots e^{j \frac{2\pi \lambda^{\text{UL}}}{Q \lambda^{\text{DL}}} (q-Q)(M-1)} \right], \\ \qquad\qquad \frac{Q}{2} < q \leq Q - 1. \end{cases} \qquad (21)$$

We notice that the UL and DL steering vectors $\boldsymbol{a}_{\text{ULA}}^{\text{UL}}(\phi)$ and $\boldsymbol{a}_{\text{ULA}}^{\text{DL}}(\phi)$ still correspond to the same basis index q. We propose to estimate the DL dominant angular sets using Algorithm 2.

Remark 1: The proposed calibration algorithm focuses on enhancing the sparsity of the DL virtual angular-domain channels, and this is very different from the previous methods [36], [37], which calibrate the long term statistical information in FDD systems. Moreover, the proposed calibration algorithm is fundamentally different from the index calibration algorithms of [28], [30]. In these two previous works, the UL and DL channels share the same unitary transform, and only the

Algorithm 2 DL dominant sets estimation

Input: Estimates of UL channels $\widehat{\boldsymbol{h}}_k^{\mathrm{UL}}$, $1 \leq k \leq K$, dictionary $\mathcal{D} = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_{Q-1}\}$, and size of dominant angular set N.

Output: Unitary transform matrix for DL channels $\boldsymbol{A}^{\mathrm{DL}}$ and DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$.

1: Call Algorithm 1 with inputs $\widehat{\boldsymbol{h}}_k^{\mathrm{UL}}$, $1 \leq k \leq K$, and $\mathcal{D}$ to obtain UL unitary transform matrix $\boldsymbol{A}^{\mathrm{UL}}$, and record selected non-orthogonal bases $\boldsymbol{B} \in \mathbb{C}^{M \times N_{\mathrm{iter}}}$.

2: Calculate virtual angular-domain UL channels $\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{UL}} = \boldsymbol{A}^{\mathrm{UL}}\widehat{\boldsymbol{h}}_k^{\mathrm{UL}}$ and obtain UL dominant angular sets Ω_k^{UL} by selecting N largest elements of $\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{UL}}$ for $1 \leq k \leq K$.

3: Calibrate bases $\boldsymbol{b}_i$, $1 \leq i \leq N_{\mathrm{iter}}$, of $\boldsymbol{B}$ using (21) to obtain calibrated bases $\boldsymbol{b}_i^{\mathrm{c}}$, $1 \leq i \leq N_{\mathrm{iter}}$, and hence calibrated matrix $\boldsymbol{B}^{\mathrm{c}} = \left[\boldsymbol{b}_1^{\mathrm{c}} \cdots \boldsymbol{b}_{N_{\mathrm{iter}}}^{\mathrm{c}}\right]$.

4: Randomly generate matrix $\boldsymbol{C}$ of size $M \times (M - N_{\mathrm{iter}})$ and obtain $\boldsymbol{B}_{\mathrm{f}}^{\mathrm{c}} = \left[\boldsymbol{B}^{\mathrm{c}}\ \boldsymbol{C}\right]$ such that $\boldsymbol{B}_{\mathrm{f}}^{\mathrm{c}}$ is a full-rank square matrix of size $M \times M$.

5: Perform GSO procedure on columns of $\boldsymbol{B}_{\mathrm{f}}^{\mathrm{c}}$ to get normalized orthogonal bases, which constitute $\boldsymbol{A}^{\mathrm{DL}}$, and DL dominant angular sets are given by $\Omega_k^{\mathrm{DL}} = \Omega_k^{\mathrm{UL}}$, $1 \leq k \leq K$.

dominant angular sets are calibrated. Our proposed scheme calibrates the underlying bases to better capture the sparsity features of the DL channels.

IV. VIRTUAL ANGULAR-DOMAIN DL CHANNEL ESTIMATION

We first consider the single-user case. Then we introduce the partial orthogonal criterion for pilot design in the multi-user scenario, and propose two pilot assignment algorithms, which minimize the pilot overhead and the pilot-reuse interference, respectively.

A. Single-user Scenario

The virtual angular-domain DL channel coefficients indexed by the DL dominant angular set captures the most power of the channel, namely, $\sum_{i \in \Omega^{\mathrm{DL}}} \left|\left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_i\right|^2 / \sum_{i=1}^{M} \left|\left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_i\right| \geq \eta$, where η is close to 1. Denote $\Omega^{\mathrm{DL}} = \{p_1^{\mathrm{DL}}, \cdots, p_N^{\mathrm{DL}}\}$, where p_n^{DL} is the nth element of Ω^{DL}. Due to the sparsity of virtual angular-domain channel, the size of Ω^{DL} $N = |\Omega^{\mathrm{DL}}| \ll M$. Hence the pilot overhead τ^{DL} required to estimate the coefficients of $\widetilde{\boldsymbol{h}}^{\mathrm{DL}}$ indexed by Ω^{DL} is very small, and the orthogonal pilot design only requires $\tau^{\mathrm{DL}} = N$ rather than $\tau^{\mathrm{DL}} = M$. We can design the virtual angular-domain pilot signal $\widetilde{\boldsymbol{\Phi}} \in \mathbb{C}^{M \times \tau^{\mathrm{DL}}}$ as

$$\left[\widetilde{\boldsymbol{\Phi}}\right]_{\Omega^{\mathrm{DL}}} = \left[\boldsymbol{s}_1^{\mathrm{T}} \cdots \boldsymbol{s}_N^{\mathrm{T}}\right]^{\mathrm{T}} \text{ and } \left[\widetilde{\boldsymbol{\Phi}}\right]_{\bar{\Omega}^{\mathrm{DL}}} = \boldsymbol{0}_{(M-N) \times \tau^{\mathrm{DL}}}, \quad (22)$$

where $\{\boldsymbol{s}_i \in \mathbb{C}^{1 \times \tau^{\mathrm{DL}}}\}_{i=1}^N$ is the set of the length-τ^{DL} orthogonal pilots, $\bar{\Omega}^{\mathrm{DL}} = \{1, \cdots, M\} \setminus \Omega^{\mathrm{DL}}$ is the complementary set of Ω^{DL}, and $\boldsymbol{0}_{l_1 \times l_2}$ is the zero matrix of size $l_1 \times l_2$, while $\left[\widetilde{\boldsymbol{\Phi}}\right]_{\Omega^{\mathrm{DL}}}$ and $\left[\widetilde{\boldsymbol{\Phi}}\right]_{\bar{\Omega}^{\mathrm{DL}}}$ indicate the sub-matrices obtained by extracting the rows of $\widetilde{\boldsymbol{\Phi}}$ indexed by Ω^{DL} and $\bar{\Omega}^{\mathrm{DL}}$,

respectively. The BS can transmit the designed orthogonal pilots on the corresponding angular ports. Combining the DL training model (9) and the virtual angular-domain pilot design (22), the received signal at the UE $\boldsymbol{y}$ can be obtained as

$$\boldsymbol{y} = \left(\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right)^{\mathrm{T}} \widetilde{\boldsymbol{\Phi}} + \boldsymbol{n} = \sum_{n=1}^{N} \left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_{p_n^{\mathrm{DL}}} \boldsymbol{s}_n + \boldsymbol{n}. \quad (23)$$

The virtual angular-domain channel coefficients corresponding to each pilot are then obtained using LS estimation as

$$\frac{\boldsymbol{y}\boldsymbol{s}_n^{\mathrm{H}}}{\boldsymbol{s}_n\boldsymbol{s}_n^{\mathrm{H}}} = \left[\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}\right]_{p_n^{\mathrm{DL}}} = \left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_{p_n^{\mathrm{DL}}} + \frac{\boldsymbol{n}\boldsymbol{s}_n^{\mathrm{H}}}{\tau^{\mathrm{DL}}}, \; 1 \leq n \leq N. \quad (24)$$

Note that the recent advances in FDD massive MIMO DL channel feedback, such as [38], [39], are not applicable here because the UE do not have the CCM. Moreover, due to the unavailability of both the unitary transform matrix $\boldsymbol{A}^{\mathrm{DL}}$ and the dominant angular set Ω^{DL} at the UE, the UE quantizes and feeds back $\left\{\frac{\boldsymbol{y}\boldsymbol{s}_1^{\mathrm{H}}}{\boldsymbol{s}_1\boldsymbol{s}_1^{\mathrm{H}}}, \cdots, \frac{\boldsymbol{y}\boldsymbol{s}_N^{\mathrm{H}}}{\boldsymbol{s}_N\boldsymbol{s}_N^{\mathrm{H}}}\right\}$, which forms a N-dimensional vector, to the BS with the pre-designed codebooks, such as Grassmannian codebooks [40] or random vector quantization (RVQ) codebooks [41]. Note that the feedback overhead is relatively small and it increases linearly with N, which is much smaller than the antenna number M, due to the angular-domain sparsity.

Since the BS knows Ω^{DL}, it can map $\frac{\boldsymbol{y}\boldsymbol{s}_n^{\mathrm{H}}}{\boldsymbol{s}_n\boldsymbol{s}_n^{\mathrm{H}}}$ to $\left[\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}\right]_{p_n^{\mathrm{DL}}}$ and forms $\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}$ with $\left[\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}\right]_i = 0, \forall i \in \bar{\Omega}^{\mathrm{DL}}$. The DL channel is then reconstructed as $\widehat{\boldsymbol{h}}^{\mathrm{DL}} = \left(\boldsymbol{A}^{\mathrm{DL}}\right)^{\mathrm{H}}\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}$. Clearly, the feedback overhead only grows linearly with N, and this is a dramatic reduction compared to the conventional schemes that directly quantize the DL channel $\widehat{\boldsymbol{h}}^{\mathrm{DL}}$ at the UE side [42], [43].

By neglecting feedback losses, the MSE of the estimated DL channel $\widehat{\boldsymbol{h}}^{\mathrm{DL}}$ is given by

$$\begin{aligned}
\mathrm{MSE} &= \mathsf{E}\left\{\left\|\boldsymbol{h}^{\mathrm{DL}} - \widehat{\boldsymbol{h}}^{\mathrm{DL}}\right\|^2\right\} = \mathsf{E}\left\{\left\|\widetilde{\boldsymbol{h}}^{\mathrm{DL}} - \widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}\right\|^2\right\} \\
&= \sum_{i \in \Omega^{\mathrm{DL}}} \mathsf{E}\left\{\left|\left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_i - \left[\widehat{\widetilde{\boldsymbol{h}}}^{\mathrm{DL}}\right]_i\right|^2\right\} + \sum_{i \in \bar{\Omega}^{\mathrm{DL}}} \left|\left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_i\right|^2 \\
&= \sigma_n^2 + \sum_{i \in \bar{\Omega}^{\mathrm{DL}}} \left|\left[\widetilde{\boldsymbol{h}}^{\mathrm{DL}}\right]_i\right|^2, \quad (25)
\end{aligned}$$

where we have used $\tau^{\mathrm{DL}} = N$. The second term in the righthand side of (25) is due to the approximation by discarding the coefficients outside the DL dominant angular set. Since these coefficients only capture a very small part of the channel power, 0.02 for $\eta = 0.98$, the proposed virtual angular-domain channel estimation scheme achieves excellent MSE performance with a much reduced pilot overhead and feedback overhead.

B. Multi-user Scenario

Due to the randomness of UE locations, different UE have different AoAs at the BS, leading to different DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$. We assume that the sizes of these DL dominant angular sets are the same, denoted by N. Then the DL dominant angular set of the kth UE is given

as $\Omega_k^{\mathrm{DL}} = \{p_{1,k}^{\mathrm{DL}}, \cdots, p_{N,k}^{\mathrm{DL}}\}$. A straightforward design is to transmit the orthogonal pilots on the angular ports indexed by the union set $\Omega^{\mathrm{DL}} = \bigcup_{k=1}^K \Omega_k^{\mathrm{DL}}$, known as the complete orthogonal pilot design (COPD) [30]. Since Ω_k^{DL} is a subset of Ω^{DL}, all the coefficients indexed by Ω_k^{DL} can be estimated without interference. However, the required pilot overhead is $\tau^{\mathrm{DL}} = N_p = |\Omega^{\mathrm{DL}}|$ which can be very large. In the worst case where all the Ω_k^{DL}, $1 \leq k \leq K$, do not intersect, a total of $N_p = \min\{KN, M\}$ orthogonal pilots are required.

We notice that parts of the angular ports can reuse the same pilot to reduce pilot overhead with a slight degradation in channel estimation performance. Specifically, we only guarantee that the pilots transmitted on the angular ports indexed by Ω_k^{DL} are orthogonal. Mathematically, this 'partial orthogonal' pilot design only ensures

$$[\widetilde{\boldsymbol{\Phi}}]_{\Omega_k^{\mathrm{DL}}}\big([\widetilde{\boldsymbol{\Phi}}]_{\Omega_k^{\mathrm{DL}}}\big)^{\mathrm{H}} = \tau^{\mathrm{DL}}\boldsymbol{I}_{N \times N}, \ 1 \leq k \leq K. \quad (26)$$

But if two indexes i and j are not included in the same dominant angular set Ω_k^{DL}, the ith row and the jth row of $\widetilde{\boldsymbol{\Phi}}$ can reuse the same pilot. To ensure (26), it is necessary that the number of orthogonal pilots $N_p \geq N$. Fig. 5 compares the conventional orthogonal pilot design [10], the COPD [30] and our partial orthogonal pilot design (POPD) using an example of $M = 8$, $K = 2$, $N = 3$, $\Omega_1^{\mathrm{DL}} = \{1, 2, 3\}$ and $\Omega_2^{\mathrm{DL}} = \{3, 4, 5\}$. The numbers of the required orthogonal pilots are $N_p = M = 8$, $N_p = |\Omega^{\mathrm{DL}}| = 5$ and $N_p = N = 3$ for these three designs, respectively.

Similar to the single-user case, the kth UE only needs to estimate the virtual angular-domain channel coefficients $[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_{p_{n,k}^{\mathrm{DL}}}$ indexed by Ω_k^{DL}. Recall the DL training model

$$\boldsymbol{y}_k = (\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}})^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}} + \boldsymbol{n} = \sum_{i=1}^M [\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i[\widetilde{\boldsymbol{\Phi}}]_i + \boldsymbol{n}, \ 1 \leq k \leq K. \quad (27)$$

The LS estimate of $[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_{p_{n,k}^{\mathrm{DL}}}$ is given by

$$\begin{aligned}
\big[\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{DL}}\big]_{p_{n,k}^{\mathrm{DL}}} &= \frac{\boldsymbol{y}_k\big([\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big)^{\mathrm{H}}}{[\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big([\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big)^{\mathrm{H}}} = \big[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\big]_{p_{n,k}^{\mathrm{DL}}} \\
&+ \sum_{j \neq p_{n,k}^{\mathrm{DL}}} \big[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\big]_j \frac{[\widetilde{\boldsymbol{\Phi}}]_j\big([\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big)^{\mathrm{H}}}{\tau^{\mathrm{DL}}} + \frac{\boldsymbol{n}\big([\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big)^{\mathrm{H}}}{\tau^{\mathrm{DL}}}. \quad (28)
\end{aligned}$$

It is noticed from (28) that due to pilot reuse at different angular ports, interference is introduced when estimating the desired coefficient. However, the interference can only come from the angular ports indexed by $\bar{\Omega}_k^{\mathrm{DL}} = \{1, \cdots, M\} \setminus \Omega_k^{\mathrm{DL}}$ that reuse $[\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}$. Since the coefficients $[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_j$, $j \in \bar{\Omega}_k^{\mathrm{DL}}$, only account for a very small part of the channel power, the interference introduced by pilot-reuse is limited. By denoting the set of angular ports that reuse the pilot $[\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}$ as $\mathcal{I}_{k,n} = \{j | j \in \bar{\Omega}_k^{\mathrm{DL}}, \text{ s.t. } [\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}} = [\widetilde{\boldsymbol{\Phi}}]_j\}$, (28) can be rewritten as

$$\big[\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{DL}}\big]_{p_{n,k}^{\mathrm{DL}}} = \big[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\big]_{p_{n,k}^{\mathrm{DL}}} + \sum_{j \in \mathcal{I}_{k,n}} \big[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}\big]_j + \boldsymbol{n}\big([\widetilde{\boldsymbol{\Phi}}]_{p_{n,k}^{\mathrm{DL}}}\big)^{\mathrm{H}}/\tau^{\mathrm{DL}}. \quad (29)$$
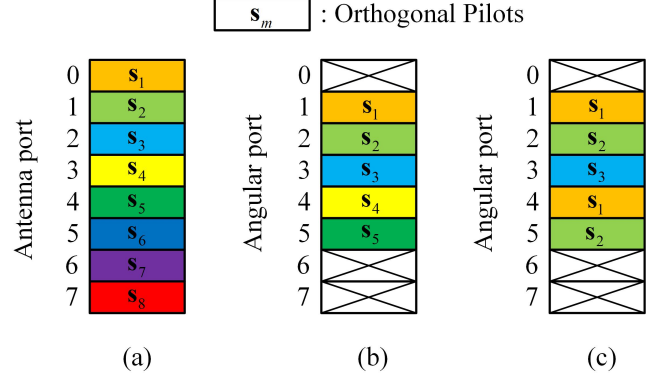

Fig. 5. (a) Conventional orthogonal pilot design, (b) virtual angular-domain complete orthogonal pilot design, (c) virtual angular-domain partial orthogonal pilot design, where $M = 8$, $K = 2$, $N = 3$, $\Omega_1^{\mathrm{DL}} = \{1, 2, 3\}$ and $\Omega_2^{\mathrm{DL}} = \{3, 4, 5\}$.

By neglecting feedback losses, the MSE of the kth UE's channel estimate $\widehat{\boldsymbol{h}}_k^{\mathrm{DL}}$ is upper bounded by

$$\mathrm{MSE}_k \leq \frac{N\sigma_n^2}{\tau^{\mathrm{DL}}} + \big(\max_{1 \leq n \leq N} |\mathcal{I}_{k,n}| + 1\big) \sum_{i \in \bar{\Omega}_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2. \quad (30)$$

where $\mathcal{I}_k = \bigcup_{n=1}^N \mathcal{I}_{k,n}$. The proof of this MSE upper bound is given in the Appendix.

Remark 2: Similar to the single-user scenario, each UE only knows the orthogonal pilot set $\{\boldsymbol{s}_n\}_{n=1}^{N_p}$. Thus the kth UE can provide the LS estimates of the N_p coefficients $\frac{\boldsymbol{y}_k\boldsymbol{s}_n^{\mathrm{H}}}{\boldsymbol{s}_n\boldsymbol{s}_n^{\mathrm{H}}}$, $1 \leq n \leq N_p$. These N_p coefficients are quantized and fed back to the BS. Since the BS knows Ω_k^{DL}, it can map N coefficients of $\{\frac{\boldsymbol{y}_k\boldsymbol{s}_n^{\mathrm{H}}}{\boldsymbol{s}_n\boldsymbol{s}_n^{\mathrm{H}}}\}_{n=1}^{N_p}$ to $\{[\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{DL}}]_{p_{n,k}^{\mathrm{DL}}}\}_{n=1}^N$, and discards the rest of them. Clearly, the feedback overhead grows linearly with N_p.

C. Partial Orthogonal Pilot Design

Given the DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$, a set of N_p orthogonal pilots are required to satisfy the partial orthogonal criterion. We introduce a novel graph representation for the pilot assignment process and propose two pilot assignment algorithms, called minimum overhead pilot assignment and minimum interference pilot assignment.

The graph representation for the pilot assignment process is constructed as follows. Define the indication function $C(i, j)$ of two different angular ports $i \neq j$ as

$$C(i, j) = \begin{cases} 1, & i, j \in \Omega_k^{\mathrm{DL}}, \\ 0, & \text{otherwise}. \end{cases} \quad (31)$$

Clearly, if $C(i, j) = 1$, angular ports i and j must transmit orthogonal pilots, while if $C(i, j) = 0$, they can reuse the same pilot. First, we construct an initial graph containing M vertices, where each vertex corresponds to an angular port. If $C(i, j) = 1$, vertices i and j are connected by an edge $e_{i,j}$. We remove the vertices without any neighbors, which are indexed by $\bar{\Omega} = \{1, \cdots, M\} \setminus \Omega^{\mathrm{DL}}$. Then we iteratively merge two vertices i and j that are not connected. The merged vertex inherits all the neighbors of its two 'father' vertices and the graph is updated after every merging operation. After all

Algorithm 3 Minimum overhead pilot assignment

Input: DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$.

Output: Angular domain pilot signal $\widetilde{\Phi}$.

1: Construct initial graph $\mathcal{G}$ with M vertices, where vertices i and j are connected with edge $e_{i,j}$ if $i, j \in \Omega_k^{\mathrm{DL}}$ for any k, and then remove the vertices without any neighbors.

2: Initial $S_i = \{i\}$, $i \in \mathcal{G}$, to record the vertices merging process.

3: **while** $\mathcal{G}$ is not a complete graph **do**

4:     Find disconnected vertex pair $\{i, j\}$ in $\mathcal{G}$ with maximum number of common neighbors:

$$\{i, j\} = \arg \max_{i,j} |\mathcal{N}(i) \bigcap \mathcal{N}(j)|, i, j \in \mathcal{G}, e_{i,j} \notin \mathcal{E}(\mathcal{G}),$$

    where $\mathcal{N}(i)$ is the set of neighbor vertices of i, and $\mathcal{E}(\mathcal{G})$ is the set of edges of $\mathcal{G}$.

5:     Merge vertex j into i and record the merging operation as $S_i = S_i \bigcup S_j$, $S_j = \varnothing$. The merged vertex i inherits all the neighbors of its two father vertices.

6: **end while**

7: Assign orthogonal pilots to the vertices of the obtained complete graph $\mathcal{G}$. Assuming that pilot s_i is assigned to vertex i, $i \in \mathcal{G}$, then all the original vertices which are merged into i reuse pilot s_i, i.e., $[\widetilde{\Phi}]_j = s_i$, $\forall j \in S_i$.

8: Set other rows of $\widetilde{\Phi}$ to the zero sequence.

Fig. 6. An example of minimum overhead pilot assignment, where $M = 8$, $K = 3$, $N = 3$, $\Omega_1^{\mathrm{DL}} = \{1, 2, 3\}$, $\Omega_2^{\mathrm{DL}} = \{1, 3, 4\}$ and $\Omega_3^{\mathrm{DL}} = \{1, 5, 6\}$.

the possible vertices have been merged, a complete graph is obtained, where all the vertices of the graph are connected with each other. The required number of the orthogonal pilots N_p is exactly the number of vertices in this complete graph.

Remark 3: In this graph-based pilot assignment process, vertices denote angular ports, edges indicate the orthogonal constraints between angular ports and vertex pair merge operation represents pilot reuse. From the aforementioned graph representation for the pilot assignment process, it is easy to see that the number of the required orthogonal pilots N_p that satisfy the partial orthogonal criterion is upper bounded by

$$N_p \leq \min\left\{M, \sqrt{K}N\right\}. \tag{32}$$

The above bound is not tight, particularly when the size of the dominant angular set N is large. This is due to the following two reasons. 1) Different dominant angular sets produce reduplicated constraints (31) if they have intersections. Thus the number of edges in the initial graph is actually far less than $\frac{1}{2}N(N-1)K$. 2) In the vertices merging operations, the edges between the merged two vertices and their common neighbors are merged. For example, if vertices i and j are to be merged and they share the same neighbor vertex k, then the edges $e_{i,k}$ and $e_{j,k}$ will be merged and the number of edges in the obtain graph will decrease by 1. Both these two reasons lead to fewer edges in the complete graph and, consequently, the number of the required orthogonal pilots N_p is far less than the above loose upper bound. The exact iterative procedure depends on how one chooses vertex pair to merge. We now discuss one such iterative procedure.

1) Minimum overhead pilot assignment: We propose a pilot assignment algorithm which minimizes the pilot overhead
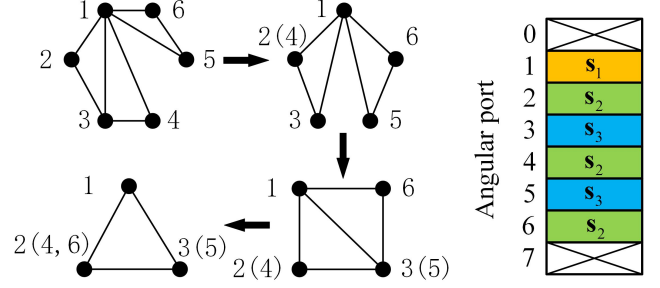
N_p. Obviously, to minimize the pilot overhead, we need to maximize pilot reuse. Note that the number of edges in the initial graph is constant, and the number of edges in the final complete graph is $\frac{1}{2}N_p(N_p-1)$. Thus to achieve the minimum N_p, we need to merge the edges as much as possible in the vertices merging operations. We propose to choose the vertex pair that has the maximum number of common neighbors to merge in each step. The detailed procedure is listed in Algorithm 3.

We use an example with $M = 8$, $K = 3$, $N = 3$, $\Omega_1^{\mathrm{DL}} = \{1, 2, 3\}$, $\Omega_2^{\mathrm{DL}} = \{1, 3, 4\}$ and $\Omega_3^{\mathrm{DL}} = \{1, 5, 6\}$ to illustrate the minimum pilot overhead assignment, Algorithm 3, in Fig. 6. It is noticed that in the first step, vertices 2 and 4 are merged since they have the maximum number of common neighbors in the current graph. This process is repeated until a complete graph is obtained, which yields $N_p = 3$ orthogonal pilots required for this example.

Next we conduct a simulation to evaluate the number of required orthogonal pilots N_p found by performing the proposed minimum overhead pilot assignment for different dominant angular set size N and UE number K, under the simulated system with the parameters $M = 128$, $Q = 8M$, $P = 20$, $d = \lambda^{\mathrm{UL}}/2$, $\lambda^{\mathrm{UL}} = 0.95\lambda^{\mathrm{DL}}$, and the randomly located UE. The results are shown in Fig. 7, where each point in Fig. 7 is obtained by selecting the maximum value in 10^4 realizations. It can be seen that given K, the number of required orthogonal pilots N_p approximately grows linearly with N. Also N_p
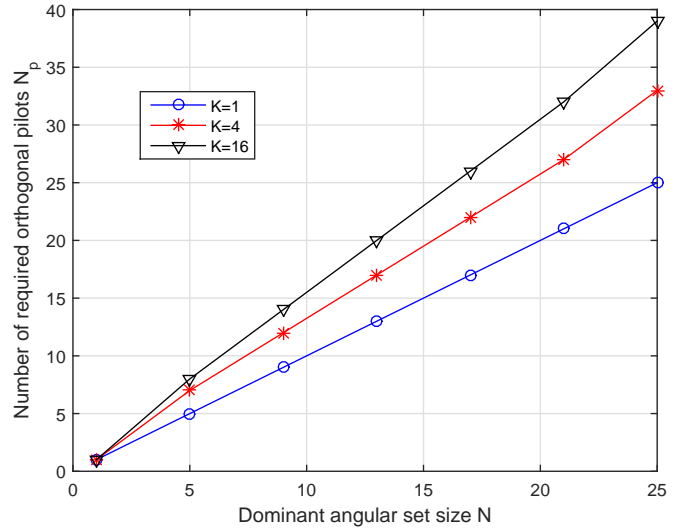

Fig. 7. Number of required orthogonal pilots N_p found by Algorithm 3 for different N and K.

increases slightly with the increase of the UE number K. For the multi-user senario, it is worth noticing that with the proposed minimum pilot overhead assignment algorithm, the number of required orthogonal pilots $N_p \ll \sqrt{K}N$, and in particular, N_p is much smaller than the number of orthogonal pilots required by the COPD [30].

2) Minimum interference pilot assignment: In practice, the number of available orthogonal pilots is often given, e.g., determined by the angle spread. For such systems, it is highly beneficial to minimize pilot-reuse interference utilizing the available pilots. Thus we propose another pilot assignment algorithm to minimize the interference for the given DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$, and the number of available orthogonal pilots N_p. Intuitively, to minimize the pilot-reuse interference, we should best utilize the available orthogonal pilots. Specifically, we start by constructing a complete graph using the minimum pilot overhead assignment algorithm. Then we iteratively split the vertex that involves the maximum number of pilot reuses until all the available pilots are utilized. The detailed procedure of this minimum interference pilot assignment is presented in Algorithm 4.

Algorithm 4 Minimum interference pilot assignment

Input: DL dominant angular sets Ω_k^{DL}, $1 \leq k \leq K$; number of available orthogonal pilots N_p.

Output: Virtual angular-domain pilot signal $\widetilde{\boldsymbol{\Phi}}$.

1: Call Algorithm 3 to perform minimum pilot overhead assignment and to obtain complete graph $\mathcal{G}$ as well as S_i, $i \in \mathcal{G}$, that record the vertices merging process.

2: **while** Vertex number of $\mathcal{G}$ is less than N_p **do**

3:    Find vertex i in $\mathcal{G}$ with maximum number of pilot reuses, i.e.,
$$i = \arg\max_{j \in \mathcal{G}} |S_j|.$$

4:    Split vertex i into two new vertices $p, q \in S_i$ and distribute the elements of S_i to S_p and S_q equally and non-repeatedly, which can be expressed as
$$S_p \cup S_q = S_i, S_p \cap S_q = \varnothing, ||S_p| - |S_q|| = 0 \text{ or } 1.$$

5: **end while**

6: Assign orthogonal pilots to vertices of $\mathcal{G}$. Assuming that pilot $\boldsymbol{s}_i$ is assigned to vertex i, $i \in \mathcal{G}$, then all original vertices merged into vertex i reuse pilot $\boldsymbol{s}_i$, i.e., $[\widetilde{\boldsymbol{\Phi}}]_j = \boldsymbol{s}_i$, $\forall j \in S_i$.

7: Set other rows of $\widetilde{\boldsymbol{\Phi}}$ to the zero sequence.

An example of the proposed minimum interference pilot assignment using Algorithm 4 is illustrated in Fig. 8. After firstly performing the minimum pilot overhead assignment, we obtain $S_1 = \{1\}$, $S_2 = \{2, 4, 6\}$ and $S_3 = \{3, 5\}$ that record the merging process. Then vertex 2 is split in the first step since it has the maximum pilot reuses in the current graph. The vertex split process is repeated until all the available N_p pilots are utilized.

3) Discussions: It is important to decide an appropriate number of orthogonal pilots when designing a wireless system. For a massive MIMO system, the number of antennas M at
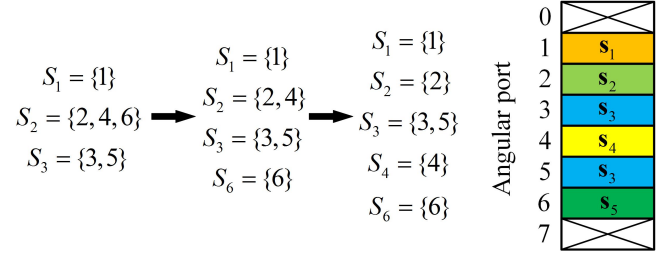

Fig. 8. An example of minimum interference pilot assignment, where $M = 8$, $K = 3$, $N = 3$, $\Omega_1^{\mathrm{DL}} = \{1, 2, 3\}$, $\Omega_2^{\mathrm{DL}} = \{1, 3, 4\}$, and $\Omega_2^{\mathrm{DL}} = \{1, 5, 6\}$, while $N_p = 5$.

the BS is very large. Thus the number of orthogonal pilots N_p depends on the number of UE K and the dominant angular set size N. The maximum number of supported UE is a given system design parameter, and the size of dominant angular set N is decided by the channel angle spread. Based on the planned operating circumstance, we can first measure the angle spread to decide an appropriate size of the dominant angular set N. Then we can determine the minimum number of the required orthogonal pilots by Monte-Carlo simulation using the minimum pilot overhead assignment algorithm. Finally, the number of orthogonal pilots for the system is set to be slightly larger than the minimum required number to provide sufficient margin for safety. The actually assignment of the orthogonal pilots to the angular ports can be achieved with the minimum interference pilot assignment algorithm.

V. SIMULATION STUDY

We conduct numerical simulations to evaluate the effectiveness of the proposed virtual angular-domain channel estimation scheme, labeled as ADCE. We consider a wideband FDD massive MIMO system adopting orthogonal frequency division multiplexing (OFDM), where the channel coefficients within a coherence time-frequency block are assumed to be the same. Moreover, the channel coefficients of each coherence time-frequency block are independently generated using the narrowband channel model given in Section II-B. In our simulated system, the BS is equipped with a large-scale ULA to serve a number of randomly located UEs simultaneously, where the AoAs of UE are independently generated from uniform distribution in the range of $[0, 2\pi]$. The default parameters of this simulated system are listed in Table I. The antenna spacing is set to $d = \lambda^{\mathrm{UL}}/2$. A coherence time of $T_c \approx 4.8\,\mathrm{ms}$ is equivalent to $L_t = T_c B / (N_{\mathrm{FFT}} + N_g) \approx 21.4$ time slots. In

TABLE I
THE DEFAULT PARAMETERS OF THE SIMULATED FDD MASSIVE MIMO SYSTEM.

Number of BS antennas M	128
Number of UE K	8
UL carrier frequency f_{UL}	2.0 GHz
DL carrier frequency f_{DL}	1.9 GHz
Number of channel paths P	20
System bandwidth B	10 MHz
DL OFDM FFT size N_{FFT}	2048
Cyclic prefix length N_g	144
Angle spread $\Delta\phi$	4°
Maximum UE mobility velocity v	54 km/h
Maximum Doppler frequency f_d	90 Hz
Channel coherence time T_c	4.8 ms

the frequency domain, the adjacent $L_f = N_{\text{FFT}}/N_g$ subcarriers are correlated and, consequently, a coherence time-frequency block contains $J = L_t L_f \approx 304$ OFDM symbols. We set $N_{\text{iter}} = M$ and $Q = 8M$ in the OMP-GSO algorithm. For simplicity, the quantization effect of feedback is not considered. We define the system's signal-to-noise ratio (SNR) as $\text{SNR} = \frac{1}{\sigma_n^2}$. Three metrics are used in our performance comparison, and they are:

1) The normalized MSE (NMSE) of the DL channel estimation which is defined by

$$\text{NMSE} = \frac{1}{K} \sum_{k=1}^{K} \frac{\left\| \widehat{\boldsymbol{h}}_k - \boldsymbol{h}_k \right\|^2}{\left\| \boldsymbol{h}_k \right\|^2}. \tag{33}$$

2) The DL training overhead r, which is defined as the ratio of the DL training sequence length τ^{DL} and the coherence block size J, i.e., $r = \frac{\tau^{\text{DL}}}{J}$.

3) Per-user achievable rate, which is calculated according to

$$C = \frac{1-r}{K} \sum_{k=1}^{K} \log_2 \left(1 + \text{SINR}_k \right), \tag{34}$$

where SINR_k is the signal to interference-plus-noise ratio (SINR) of the kth UE. Assuming that the zero-forcing (ZF) precoding is applied at the BS, SINR_k can be obtained as

$$\text{SINR}_k = \frac{\left| \left(\boldsymbol{h}_k^{\text{DL}} \right)^{\text{T}} \boldsymbol{p}_k \right|^2}{\sum_{j \neq k} \left| \left(\boldsymbol{h}_k^{\text{DL}} \right)^{\text{T}} \boldsymbol{p}_j \right|^2 + \sigma_n^2}. \tag{35}$$

In (35), $\boldsymbol{p}_k$ is the precoding vector for the kth UE which is the kth column of the ZF precoding matrix $\boldsymbol{P}$, calculated according to $\boldsymbol{P} = \widehat{\boldsymbol{H}}^* \left(\widehat{\boldsymbol{H}}^{\text{T}} \widehat{\boldsymbol{H}}^* \right)^{-1}$ with $\widehat{\boldsymbol{H}} = \left[\widehat{\boldsymbol{h}}_1^{\text{DL}} \cdots \widehat{\boldsymbol{h}}_K^{\text{DL}} \right]$.

1) Comparison with conventional LS method: Fig. 9 depicts the NMSE performance of the conventional LS method and our ADCE scheme as the functions of SNR. For our ADCE scheme, the estimated UL channels are required to estimate the DL dominant angular sets and the associated size N, We assume that the SNR for UL channel estimation is $15\,\text{dB}$,

which is reasonable. For our ADCE scheme, we consider three different combinations of $(N, N_p) = (10, 15)$, $(20, 30)$ and $(30, 45)$. The DL angular pilot signal $\widetilde{\boldsymbol{\Phi}} \in \mathbb{C}^{M \times N_p}$ for the ADCE scheme is designed using the minimum interference pilot assignment algorithm based on the estimated Ω_k^{DL}, $1 \leq k \leq K$, and the given N_p. For the conventional LS method, we have the orthogonal pilot matrix $\widetilde{\boldsymbol{\Phi}} \in \mathbb{C}^{M \times M}$. For a fair comparison, we scale each length-M pilot sequence by a factor $\frac{N_p}{M}$ so that its power is the same as a length-N_p pilot sequence for the ADCE scheme. Thus in Fig. 9, the conventional LS curve labeled by $N_p = 15$ indicates that the power of each orthogonal pilot sequence is scaled by $\frac{N_p}{M} = \frac{15}{128}$, and so on. It can be seen from Fig. 9 that at the low SNR region, our ADCE scheme achieves lower NMSE since it exploits the virtual angular-domain channel sparsity and simply sets the coefficients outside the dominant angular sets to zeros. Owing to the same reason, the ADCE scheme exhibits a NMSE floor as the SNR increases. As expected, for the ADCE scheme, the NMSE is reduced with the increase of the number of orthogonal pilots N_p.

The DL training overhead for the conventional LS is $r = \frac{M}{J} \approx 42\%$, while the overheads for our ADCE are $r = \frac{N_p}{J} \approx 5\%$, 10% and 15%, respectively, given $N_p = 15$, 30 and 45.

The per-user rate performance of our ADCE scheme and the conventional LS method are compared in Fig. 10 as the functions of the number of orthogonal pilots N_p for the ADCE scheme, where the length-M pilot sequences for the conventional LS are scaled by $\frac{N_p}{M}$. It can be seen from Fig. 10 that our ADCE scheme attains much higher per-user rate in comparison to the conventional LS method. For example, given $N_p = 30$, the per-user rates of the ADCE scheme are 28% and 50% higher than those of the conventional LS method for SNR = $15\,\text{dB}$ and $0\,\text{dB}$, respectively. Also observe that for the ADCE scheme, as N_p increases, the achieved rate firstly increases and then reduces. The reason is that when N_p is small, increasing N_p significantly improves the accuracy of channel estimation and thus increases the achievable rate but
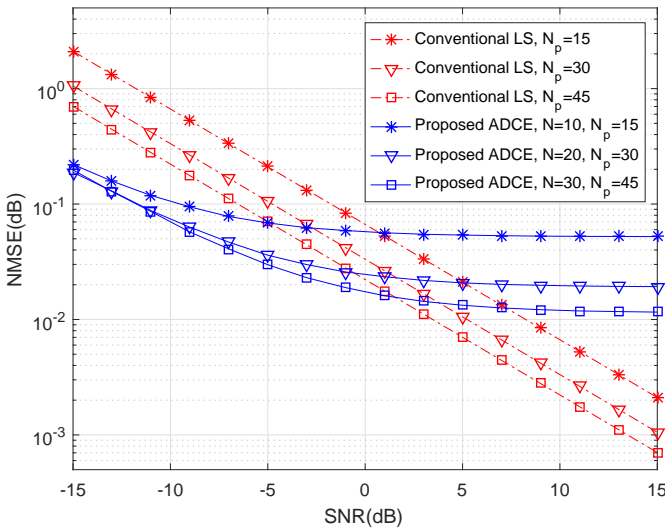

Fig. 9. NMSE performance as the functions of the system's SNR fo. the conventional LS method and the ADCE scheme with three different combinations of the dominant angular set size and the number of orthogonal pilots.
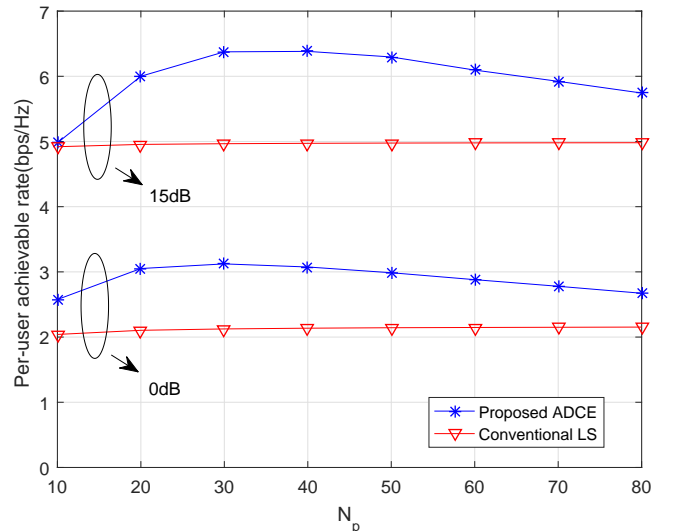

Fig. 10. Per-user achievable rate performance of the conventional LS method and the ADCE scheme with $N_p = 2N$ as the functions of the number of orthogonal pilot N_p, given two different system's SNRs.
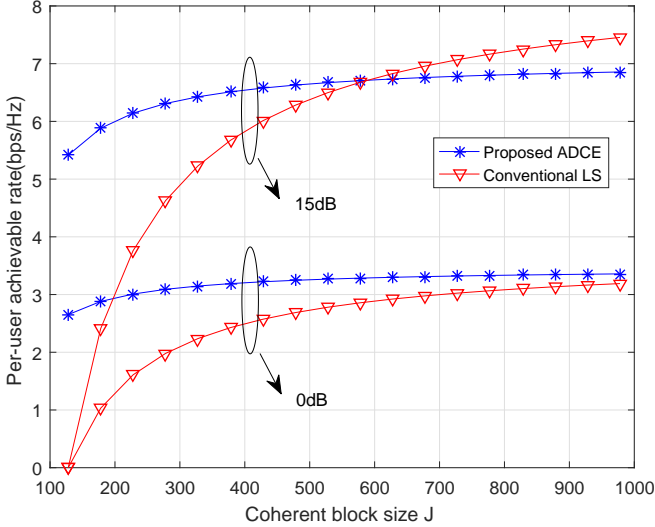
Fig. 11. Per-user achievable rate performance of the conventional LS method and the ADCE scheme with $N_p = 2N = 30$ as the functions of the coherent block size J, given two different system's SNRs.

for large N_p, pilot overhead produces large adverse effects on the achievable rate.

Fig. 11 shows the per-user achievable rates of the conventional LS method and our ADCE scheme with $N_p = 2N = 30$, for different coherent block sizes J. As expected when $J = 128$, data transmission is impossible for the conventional LS method, and it achieves zero rate. Given SNR $= 0$ dB our ADCE scheme outperforms the conventional LS method. Given SNR $= 15$ dB, our ADCE scheme also attains higher rate than the conventional LS method for the coherent block size up to $J = 600$, while for very large J, the conventional LS method outperforms our scheme because the effect of the training overhead to the achievable rate becomes minor. Note that for practical systems that supports the mobility of UE, the coherent block size is not very large, and our scheme enjoys much better performance than the conventional LS method.

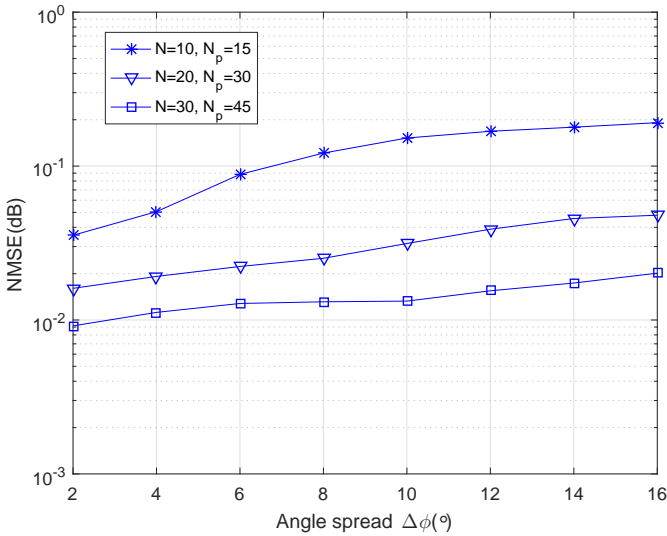Fig. 12 investigates the impact of the angle spread on the

Fig. 12. NMSE performance of the ADCE scheme as a function of the angle spread $\Delta\phi$, given SNR $= 20$ dB.

NMSE performance of the ADCE scheme, given SNR $= 20$ dB. Again the SNR for UL channel estimation is 10 dB. Note that $\Delta\phi$ has no influence on the conventional LS method because it does not utilize any spatial property. As expected, given a combination of (N, N_p), the NMSE increases with $\Delta\phi$.

2) Sparsity enhancement evaluation: The OMP-GSO algorithm is a key contribution of our work, and it is used to construct the unitary transform matrix for our ADCE scheme. In Section III-A, we have already demonstrated the sparsity enhancement performance of our OMP-GSO algorithm over the conventional DFT method and the spatial rotation strategy [28]. Here, we further evaluate its performance gain over the other two methods. Specifically, we use the OMP-GS, conventional DFT and spatial rotation methods to construct the unitary transform matrix for our ADCE scheme. The corresponding NMSE performance of the three methods are compared in Fig. 13 as the functions of the system's SNR, where the perfect UL channels are assumed, and $N = 30$ and $N_p = 45$. It can be seen that in the single-user case, the spatial rotation strategy attains significant performance gain over the conventional DFT approach, and our OMP-GSO algorithm outperforms the spatial rotation strategy considerably. For the multi-user case, our OMP-GSO algorithm also achieves the best performance, although the performance differences between the three methods become small.
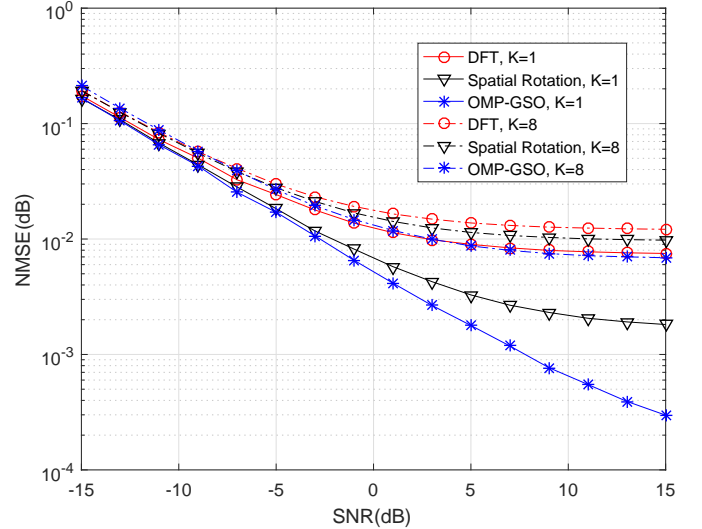

Fig. 13. NMSE performance comparison between three different unitary transform construction methods for two different values of K, given the perfect UL channels, and $N = 30$ and $N_p = 45$.

So far we realistically consider that all the UE are randomly located. In some special situations, K/G UE have similar AoAs and they can be grouped together to shared the same pilots. This is the UE grouping strategy considered in [28] for the spatial rotation method. For such a special senario where the UE can be naturally divided into G groups, we can also apply this UE grouping strategy to our OMP-GSO algorithm. There are two benefits of doing UE grouping if the UE are naturally grouped: 1. the UE inside the same group tend to contribute similar bases in the OMP-GSO algorithm and thus higher sparsity is achieved, and 2. the UE inside
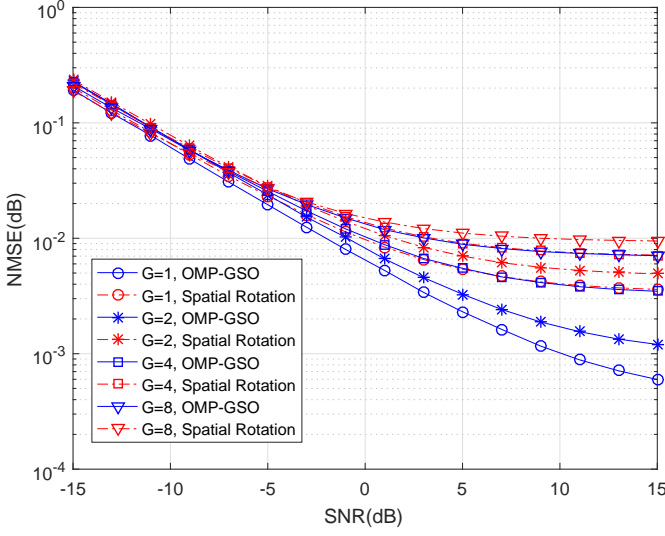
Fig. 14. NMSE performance comparison between two different unitary transform construction methods under different UE groupings, given $K = 8$. Other system parameters are the same as in Fig. 13.

the same group can simply share the same pilots without introducing pilot-reuse interference. Fig. 14 compares the NMSE performance achieved by the spatial rotation strategy and our OMP-GSO algorithm under different UE groupings. In Fig. 14, $G = 1$ indicates that all the $K = 8$ UE have very similar AoAs and are grouped together as one group, while $G = 8$ means that the $K = 8$ UE are randomly located and have very different AoAs, thus forming '8 groups'. As expected, the OMP-GSO algorithm outperforms the spatial rotation method. Observe that when more UE have similar AoAs and, therefore, can be divided into less groups, the performance gap between the two methods become larger, indicating that our OMP-GSO algorithm is more effective in exploiting the beneficial efforts of UE grouping.

VI. CONCLUSIONS

A virtual angular-domain channel estimation scheme with partial orthogonal pilot design for FDD massive MIMO systems has been proposed by exploiting the virtual angular-domain channel sparsity, which significantly reduces the training and feedback overhead. Specifically, utilizing the directional reciprocity of UL and DL channels, we have proposed to estimate the DL dominant angular sets from the UL channels with a novel OMP-GSO algorithm to construct unitary transform, and to handle the problem of different UL and DL wavelengths with an index calibration algorithm. Furthermore, we have introduced the partial orthogonal criterion for virtual angular-domain pilot design and have proposed two pilot assignment algorithms to minimize the pilot overhead and pilot-reuse interference, respectively. Simulation results have demonstrated that our proposed scheme can provide excellent MSE performance with much reduced training overhead and, consequently, achieves much higher per-user achievable rate compared to the conventional LS method. The sparsity enhancement and the performance gain of the OMP-GSO algorithm over the conventional DFT and spatial rotation methods have also be demonstrated extensively.

APPENDIX

The MSE of the kth UE's channel estimate $\widehat{\boldsymbol{h}}_k^{\mathrm{DL}}$ can be represented as

$$
\begin{aligned}
\mathrm{MSE}_k &= \mathsf{E}\big\{\big\|\boldsymbol{h}_k^{\mathrm{DL}} - \widehat{\boldsymbol{h}}_k^{\mathrm{DL}}\big|^2\big\} = \mathsf{E}\big\{\big\|\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}} - \widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{DL}}\big\|^2\big\} \\
&= \sum_{n=1}^N \mathsf{E}\big\{\big|\,[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_{p_{n,k}^{\mathrm{DL}}} - [\widehat{\widetilde{\boldsymbol{h}}}_k^{\mathrm{DL}}]_{p_{n,k}^{\mathrm{DL}}}\big|^2\big\} + \sum_{i\in\bar\Omega_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2 \\
&= \frac{N\sigma_n^2}{\tau^{\mathrm{DL}}} + \sum_{n=1}^N \bigg|\sum_{j\in\mathcal{I}_{k,n}} [\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_j\bigg|^2 + \sum_{i\in\bar\Omega_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2 \\
&\le \frac{N\sigma_n^2}{\tau^{\mathrm{DL}}} + \sum_{n=1}^N \bigg(|\mathcal{I}_{k,n}|\sum_{j\in\mathcal{I}_{k,n}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_j\big|^2\bigg) + \sum_{i\in\bar\Omega_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2 \\
&\le \frac{N\sigma_n^2}{\tau^{\mathrm{DL}}} + \max_{1\le n\le N}|\mathcal{I}_{k,n}|\bigg(\sum_{n=1}^N\sum_{j\in\mathcal{I}_{k,n}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_j\big|^2\bigg) \\
&\quad + \sum_{i\in\bar\Omega_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2 = \frac{N\sigma_n^2}{\tau^{\mathrm{DL}}} + \max_{1\le n\le N}|\mathcal{I}_{k,n}|\sum_{j\in\mathcal{I}_k} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_j\big|^2 \\
&\quad + \sum_{i\in\bar\Omega_k^{\mathrm{DL}}} \big|[\widetilde{\boldsymbol{h}}_k^{\mathrm{DL}}]_i\big|^2,
\end{aligned}
\tag{36}
$$

where $\mathcal{I}_k = \bigcup_{n=1}^N \mathcal{I}_{k,n}$, and the first inequality is due to the Cauchy-Schwarz inequality. Since $\mathcal{I}_{k,n_1}\bigcap\mathcal{I}_{k,n_2} = \emptyset$, $\forall n_1 \ne n_2$, $\mathcal{I}_k \subset \bar\Omega_k^{\mathrm{DL}}$, we obtain the upper bound of MSE_k (30).

REFERENCES

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

[2] 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.

[3] L. Lu, *et al.*, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

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

[5] 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.

[6] M. Petermann, *et al.*, "Multi-user pre-processing in multi-antenna OFDM TDD systems with non-reciprocal transceivers," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3781–3793, Sep. 2013.

[7] Z. Wang, P. Zhao, C. Qian, and S. Chen, "Location-aware channel estimation enhanced TDD based massive MIMO," *IEEE Access*, vol. 4, pp. 7828–7840, 2016.

[8] X. Guo, *et al.*, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[9] J. Zhang, *et al.*, "Pilot contamination elimination for large-scale multiple antenna OFDM systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[10] Y.-H. Nam, *et al.*, "Evolution of reference signals for LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 132–138, Feb. 2012.

[11] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[12] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

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

[14] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.

[15] K. N. Le, "Distributions of even-degree-of-freedom generalized-Rician fading for opportunistic-relay selection employing MRC under outdated-CSI conditions," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6441–6455, Sep. 2019.

[16] K. N. Le, "Opportunistic outdated relays under even-degree-of-freedom generalized-Rician fading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2472–2487, Mar. 2019.

[17] K. N. Le, *OFDM and OFDMA with Linear Diversity for Future Wireless Communications*. BenTham Science Publishing, USA, eISBN: 978-1-60805-188-5 (on line), 2011.

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

[19] X. Zhang, L. Zhong and A. Sabharwal, "Directional training for FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5183–5197, Aug. 2018.

[20] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[21] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[22] C.-C. Tseng, J.-Y. Wu, and T.-S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block l_1-minimization," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1055–1067, Mar. 2016.

[23] A. Liu, F. Zhu, and V. K. N. Lau, "Closed-loop autonomous pilot and compressive CSIT feedback resource adaptation in multi-user FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 173–183, Jan. 2017.

[24] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.

[25] S. Imtiaz, G. S. Dahman, F. Rusek, and F. Tufvesson, "On the directional reciprocity of uplink and downlink channels in frequency division duplex systems," in *Proc. PIMRC 2014* (Washington, DC, USA), Sep. 2-5, 2014, pp. 172–176.

[26] P. Kyösti, *et al.*, *WINNER II Channel Models*. IST-4-027756 WINNER II D1.1.2 V1.2, Sep. 2007.

[27] *Technical Specification Group Radio Access Network; Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*. 3GPP,TR 25.996, V9.0.0, Dec. 2009.

[28] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Techno.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[29] X. Xiong, X. Wang, X. Gao, and X. You, "Beam-domain channel estimation for FDD massive MIMO systems with optimal thresholds," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4669–4682, Jul. 2017.

[30] P. Zhao, Z. Wang, and C. Sun, "Angular domain pilot design and channel estimation for FDD massive MIMO networks," in *Proc. ICC 2017* (Paris, France), May 21-25, 2017, pp. 1–6.

[31] J. Zhang, *et al.*, "Pilot contamination elimination for large-scale multiple antenna aided OFDM systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[32] X. Guo, *et al.*, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[33] Z. wang, P. Zhao, C. Qian, and S. Chen, "Location-aware channel estimation enhanced TDD based massive MIMO," *IEEE Access*, vol. 4, pp. 7828–7840, Nov. 2016.

[34] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[35] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[36] Y. C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems,"*IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

[37] M. Jordan, X. Gong, and G. Ascheid, "Conversion of the spatio-temporal correlation from uplink to downlink in FDD systems," in *Proc. WCNC 2009* (Budapest, Hungary),Apr. 5-8, 2009, pp. 1–6.

[38] B. Clerckx, G. Kim, and S. Kim, "MU-MIMO with channel statistics-based codebooks in spatially correlated channels," in *Proc. GLOBECOM 2008* (New Orleans, LA, USA), Nov. 30-Dec. 4, 2008, pp. 1–5.

[39] W. Shen, *et al.*, "On the performance of channel-statistics-based codebook for massive MIMO channel feedback", *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7553–7557, Aug. 2017.

[40] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-advanced to FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2007–2016, Apr. 2015.

[41] J. Joung, E. Kurniawan, and S. Sun, "Channel correlation modeling and its application to massive MIMO channel feedback reduction," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 3787–3797, May. 2017.

[42] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[43] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.