FDD Massive MIMO Based on Efficient Downlink Channel Reconstruction

Yu Han, *Student Member, IEEE,* Qi Liu, Chao-Kai Wen, *Member, IEEE,* Shi Jin, *Senior Member, IEEE* and Kai-Kit Wong, *Fellow, IEEE,*

Abstract

This paper focuses on frequency division duplex (FDD) massive multiple-input multiple-output (MIMO) systems and proposes a transceiver design that fully exploits the downlink spatial multiplexing gain with only a small amount of overhead. The bottleneck lies in the acquisition of downlink channel state information (CSI), which occurs when large scale antenna array is employed in FDD transmission systems. Fortunately, the spatial reciprocity between uplink and downlink inspires us to reconstruct the downlink channel based on the frequency-independent parameters (downtilts, azimuths and delays) that can be derived in the uplink. We first extract these parameters through an enhanced Newtonized orthogonal matching pursuit (e-NOMP) algorithm which is proposed in this paper to fit the massive MIMO orthogonal frequency division multiplexing (OFDM) system. After formulating the requirement to achieve an acceptable estimation error rate, we propose a low-cost downlink training scheme to estimate the downlink gains of each user channel. This scheme saves the training time resource by introducing a predefined spatial angle grid which corresponds to a beam set and by minimizing the number of selected beams which is equal to the number of OFDM symbols used for downlink training. Having obtained the reconstructed multiuser channel, the BS can maximize the spatial multiplexing gain by serving all the users simultaneously without causing severe interference. Numerical results verify the precision of the e-NOMP algorithm, and demonstrate that sum-rate performance of the reconstruction-based transceiver design approximates that of using perfect CSI.

Index Terms

FDD massive MIMO, downlink channel reconstruction, multiuser transmission.

I. Introduction

Massive multiple input multiple output (MIMO) has been recognized as one of the key enablers of the fifth generation and future mobile communications networks [1]. Large scale antenna arrays are equipped at the base stations (BSs) to fully exploit the spatial degrees of freedom, providing huge room for spatial division multiplexing [2]–[4]. Multiple users can be served by the BS on a same time-frequency resource block, and the spatial multiplexing dimension can be further expanded by scaling up the antenna array at the BS. The large array is usually structured in planar or circular topology, thereby exploits both the horizontal and the vertical space, contributing to three dimensional (3D) MIMO techniques. A beam formed by the array can flexibly target at any direction in the 3D space according to practical requirements [5]–[7]. We can also design beam weights to produce a set of spatially orthogonal beams using the large scale array and transmit multiple data streams on these beams without causing interference [8]. These advantages promise the high sum-rate performance of multiuser massive MIMO systems.

One essential foundation to gain these advantages is the acquisition of channel state information (CSI). Due to the lack of uplink-downlink reciprocity in frequency division duplex (FDD) systems, downlink training and feedback is a typical solution of downlink channel estimation. In the fourth generation mobile communication era and before, the number of BS antennas is relatively small and downlink CSI can be easily acquired by sending orthogonal downlink pilots, applying linear channel estimation at user side and finally feeding the estimates back to the BS. While in massive MIMO systems, using completely orthogonal downlink pilots and sending back high dimensional complex channel matrix are not applicable in practical systems. Obtaining downlink CSI at BS side becomes a bottleneck in FDD massive MIMO system. Researchers have started to search for new solutions to obtain downlink CSI and design corresponding transmission schemes.

A. *Related work*

Some methods follow the traditional approach by transmitting downlink pilots and sending back the estimates to the BS. For instance, compressed sensing is introduced to estimate the sparse channel through a small amount of downlink measurements [9], [10], but comprehensive signal processing is conducted at the user side, raising an exorbitant requirement on the capability of user equipment. For another instance, time-correlation of the wireless channel can be utilized

in both downlink training and feedback phases. [11] suggested that user can estimate downlink CSI based on the currently received downlink pilots as well as the estimated CSI that is obtained at previous moment. Before feeding back the estimates, [12] and [13] proposed to quantize the channel based on previous results within the coherence time by using a trellis-extended codebook and an angle of departure-adaptive subspace codebook, respectively. These methods rely heavily on the accuracy of the initial estimates.

During recent years, the spatial reciprocity between uplink and downlink has attracted increasing attentions. There has been existing work that suggests to acquire downlink CSI by using the information obtained from uplink. The existing work generally aims to obtain two categories of downlink CSI. The first category is partial CSI, such as the spatial information or the reduced dimensional channel. For example, only the angles of propagation paths are estimated during the training phase. Or instead, channel sparsity in beamspace is utilized and spatial angle estimation is translated to searching for the non-zero elements in the beamspace channel. In this condition, only spatial directions or beam indices are known at BS and user scheduling is required to avoid spatial overlapping among different users [8], [14]. The second category is full CSI, which describes the full-dimensional channel and contains the complete information in the propagation environment. Full CSI is usually obtained by channel estimation or reconstruction scheme. With full CSI, the BS is able to serve more users simultaneously by employing linear precoding scheme and eliminate the interference perfectly. The BS also can conduct a comprehensive user scheduling scheme to maximize the sum-rate performance and to fully exploit the spatial multiplexing gain.

The majority of latest work focuses on the acquisition of full CSI [15]–[19]. Facing to the clustering channel that covers a continuous angular region, [16]–[18] suggested to estimate channel based on the downlink channel covariance matrix which describes the angular domain energy distribution and can be derived from its uplink version. [18] further selected users by maximizing the downlink multiplexing gain based on the uplink channel covariance matrices. When it comes to the limited scattering channel where multiple distinct paths exist, authors of [19] proposed a unified transmission strategy with the aid of the spatial basis expansion model. Based on channel sparsity, the reduced-dimensional angular-domain downlink channel from each user was first estimated and then transformed back to the full-dimensional antenna domain. With the obtained full CSI, the BS conducted greedy scheduling before downlink multiuser

transmission to achieve sub-optimal sum-rate performance. All of the above are based on a beam or angular domain which uniformly over-samples the space. Instead, [20], [21] proposed to detect the frequency-independent parameters of each path from the uplink and reconstruct the downlink channel using these parameters as well as limited amount of downlink training overhead. They also demonstrated the effectiveness of the proposed scheme in over-the-air tests. However, the schemes introduced in [20], [21] are designed for single-antenna or small-scale antenna array systems and are not applicable in massive MIMO scenario.

B. Contributions

Based on the spatial reciprocity in FDD systems, this paper proposes an FDD massive MIMO transceiver design based on an efficient multiuser downlink channel reconstruction scheme when large scale antenna array is equipped at the BS and orthogonal frequency division multiplexing (OFDM) is applied. To acquire downlink CSI, an enhanced Newtonized orthogonal matching pursuit (e-NOMP) algorithm is proposed to extract the frequency-independent parameters during the uplink sounding phase. Then, an efficient downlink training scheme is proposed to estimate the downlink gains for multiple users based on the uplink-estimated frequency-independent parameters. Using the reconstructed channels, the BS can design a multiuser transmission scheme with strong interference cancellation capability. The major contributions are as follows:

1) An e-NOMP algorithm to extract frequency-independent parameters in a 3D MIMO wide-band channel. Uniform planar array (UPA) is employed at the BS to exploit both the horizontal and the vertical planes. In this condition, the frequency-independent parameters in the wireless channel include the downtilt, azimuth and delay of each propagation path. We introduce an e-NOMP algorithm that can detect the component paths from their noised mixture and extract the downtilt, azimuth and delay of each path. A codebook whose structure fits well with the model of a component path is first generated. Then, the Newton step is adjusted to refine the downtilt, azimuth and delay simultaneously. By the end of the e-NOMP algorithm, the BS obtains all the frequency-independent parameters from all the user channels.
2) An efficient multiuser downlink channel reconstruction scheme. To avoid the one-path-by-one-path training mode which costs extremely long training time, we first study the main factor that determines the error rate of the estimated downlink gains and formulate

the requirement to achieve an acceptable error rate. Theoretical analysis tells us that the downlink training beamforming can be flexibly designed if the formulated requirement is satisfied. Therefore, we first introduce a spatial angle grid and project the estimated spatial angles on the sampled angles. Each sampled angle corresponds to a downlink training beam. The spatial angle grid translates the beamformed user-dedicated downlink pilots to cell-common pilots, and meanwhile the beam selection strategy promises a relatively high receiving SNR at user side. Then, a beam scheduling scheme is proposed to minimize the number of selected beams and reduce the downlink training overhead. We remove the beams with the least contribution first and keep the beams that are needed by most users. By means of the previously formulated requirement to achieve an acceptable error rate, extra beams are removed from the selected beam set as long as the requirement still can be satisfied. This beam scheduling scheme promises the low-cost advantage of the proposed multiuser downlink channel reconstruction scheme.

3) Design and evaluation of the reconstruction-based transmission scheme. If the number of users is far less than the number of BS antennas, the BS can serve all the users simultaneously because the inter-user interference can be well eliminated using the precisely reconstructed multiuser channel. We apply zero-forcing (ZF) precoding in the downlink and evaluate the multiuser sum-rate performance by studying signal-to-interference-noise ratio (SINR) at a user. We find that the toleration of estimation error will impact the target signal power and introduce interference, further driving down the SINR. Numerical results show that the decrease of tolerate error rate contributes to the increase of sum-rate, but meanwhile calls for much more downlink training overhead. Therefore, we should make a balance between performance and overhead, and find a proper metric to design our scheme.

The rest of this paper is organized as follows. Section II describes the massive MIMO-OFDM system working in FDD transmission mode, introduces the uplink and downlink channel models based on the frequency-independent parameters and briefly overview the proposed downlink channel reconstruction and multiuser transmission scheme. Section III presents the e-NOMP algorithm, highlighting the new codebook and the updated Newton step designed for massive MIMO-OFDM scenario. Section IV gives a detailed description about the theoretical rationale and the working principle of the proposed low-cost downlink training strategy for the multiuser

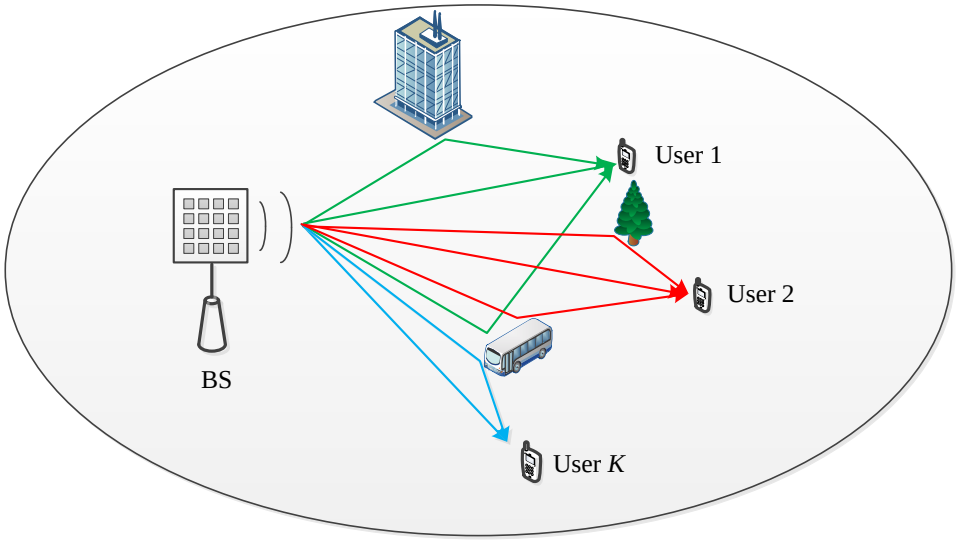

Fig. 1. The BS is equipped with UPA and serves multiple users simultaneously. Buildings, trees and cars are all scatters in the wireless channel.

system. Section V applies the reconstruction results to the downlink multiuser transmission scheme and analyzes the sum-rate performance. Section VI presents the numerical results of the developed NOMP algorithm and the proposed downlink channel reconstruction scheme. Section VII concludes the paper.

Notations—We denote the matrices and vectors by uppercase and lowercase boldface letters, respectively, while use the superscripts $(\cdot)^{\dagger}$, $(\cdot)^{H}$, and $(\cdot)^{T}$ as taking the pseudo-inverse, conjugate-transpose, and transpose, respectively. We also denote $[\mathbf{A}]_{i,:}$ and $[\mathbf{A}]_{:,j}$ as the ith row and the jth column of matrix $\mathbf{A}$, and $[\mathbf{A}]_{i,j}$ as the (i,j)th entry of $\mathbf{A}$. $\mathcal{R}\{\cdot\}$ represents taking the real component of a complex number, while $\mathbb{E}\{\cdot\}$ represents taking the expectation with respect to the random variables inside the brackets. $|\cdot|$ and $\|\cdot\|$ means taking the absolute value and modulus operations, and $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ means rounding a decimal number to its nearest lower and higher integers, respectively.

II. System Model

A. Channel Model

A single cell of a massive MIMO system is considered to function in FDD transmission mode and OFDM modulation is applied. We denote the uplink and downlink carrier frequencies as f^{ul} and f^{dl}, respectively, and assume that each of the uplink and downlink frequency bands

has N sub-carriers with spacing $\triangle f$. The BS simultaneously serves K users who are randomly distributed in the cell. The BS is equipped with a uniform planar array (UPA) and each user has a single antenna. This UPA contains $M = M_h M_v$ antenna elements, including M_h elements in each row and M_v elements in each column. The distance between two horizontal or vertical elements is $d = \lambda/2$, where λ is the carrier wavelength. As shown in Fig. 1, scatters exist in the space, and the user channel is composed of multiple propagation paths. The wireless signal can arrive at the user side along the line-of-sight path, or be reflected by these scatters. Different user channels may share a common scatter and are spatially overlapped with each other.

For user k, when down-converted to the baseband, its uplink multipath channel is expressed as

$$\mathbf{h}_k^{\mathrm{ul}} = \sum_{l=1}^{L_k} g_{k,l}^{\mathrm{ul}} \mathbf{a}(\theta_{k,l}, \phi_{k,l}) \otimes \mathbf{p}(\tau_{k,l}), \tag{1}$$

where L_k is the number of propagation paths of user k, $\otimes$ denotes taking Kronecker product, $\mathbf{a}(\theta, \phi)$ is the is the steering vector of UPA and can be expressed as

$$\begin{aligned} &\mathbf{a}(\theta, \phi) = \mathbf{a}_v(\theta) \otimes \mathbf{a}_h(\theta, \phi), \\ &\mathbf{a}_v(\theta) = \left[1, e^{j2\pi \frac{d}{\lambda} \sin\theta}, \ldots, e^{j2\pi (M_v - 1) \frac{d}{\lambda} \sin\theta}\right]^T, \\ &\mathbf{a}_h(\theta, \phi) = \left[1, e^{j2\pi \frac{d}{\lambda} \cos\theta \sin\phi}, \ldots, e^{j2\pi (M_h - 1) \frac{d}{\lambda} \cos\theta \sin\phi}\right]^T, \end{aligned} \tag{2}$$

where $\theta_{k,l} \in [-\pi/2, \pi/2)$ and $\phi_{k,l} \in [-\pi/2, \pi/2)$ are the downtilt and the azimuth of the lth propagation path of user k,

$$\mathbf{p}(\tau) = \left[1, e^{j2\pi \triangle f \tau}, \ldots, e^{j2\pi (N-1) \triangle f \tau}\right]^T \tag{3}$$

is the delay vector on the OFDM sub-carriers, and $\tau_{k,l}$ is the delay of the lth propagation path of user k.

Based on the frequency-independent feature of the delays and angles, in the downlink, the baseband channel of user k can be modeled as

$$\mathbf{h}_k^{\mathrm{dl}} = \sum_{l=1}^{L_k} g_{k,l}^{\mathrm{dl}} \mathbf{a}^T(\theta_{k,l}, \phi_{k,l}) \otimes \mathbf{p}^T(\tau_{k,l}) e^{j2\pi (f^{\mathrm{dl}} - f^{\mathrm{ul}}) \tau_{k,l}}, \tag{4}$$

where $g_{k,l}^{\mathrm{dl}}$ is the downlink complex gain of the lth propagation path in the kth user's channel.

B. Channel Reconstruct Based Downlink Multiuser Transmission

The spatial reciprocity reflected through these frequency-independent parameters provides the possibility of reconstructing downlink channel by utilizing the uplink derived delays and angles. However, since different carrier frequencies are applied in uplink and downlink, wireless signal experiences different phase shifts when reflection happens during propagation. Thus, $g_{k,l}^{\mathrm{dl}} \neq g_{k,l}^{\mathrm{ul}}$, demonstrating that $g_{k,l}^{\mathrm{ul}}$ obtained from uplink can not be applied in the downlink channel model, and we must estimate $g_{k,l}^{\mathrm{dl}}$ specially from downlink. Once the gains, delays and angles are known by the BS, the downlink channel can be reconstructed at the transmitter side.

Based on this, we design an FDD massive MIMO transceiver via downlink channel reconstruction. We first extract the delays and angles from uplink, and then estimate the downlink gains through downlink training. After reconstructing the downlink channels based on the frequency-independent parameters and the downlink estimated gains fed back from the users, the BS designs interference-eliminable precoders based on the full CSI and transmits data to all users on a same time-frequency resource block.

In this paper, the employment of large scale antenna array and the exploitation of vertical space bring new challenges in the reconstruction of downlink channel. On the one hand, three types of spatial parameters are frequency-independent and to be detected in this system, including downtilts, azimuths and delays. On the other hand, we should reconstruct downlink channels for multiple users with an acceptable amount of overhead to promise that enough time resources are remained for data transmission. These two key problems will be solved one by one in the following sections.

III. Extract Delays and Angles in the Uplink

In this section, we focus on the first challenge mentioned above and investigate how to obtain the frequency-independent parameters, including the delays, azimuths and downtilts of each user channel for this massive MIMO-OFDM system. Frequency-independence of these parameters inspires us to extract them in the uplink.

A. Uplink Sounding RS Model

During the uplink sounding phase, each user sends sounding RSs to the BS. Sounding RSs from different users are frequency-separated or time-separated, and the BS knows the time-

frequency resource positions of the RS from each user. Assume that all-1 type sounding RSs are applied and the sounding RS from user k occupy the kth OFDM symbol in the uplink slot. The received sounding RS at BS from user k is expressed as

$$\mathbf{y}_k^{\mathrm{ul}} = \sum_{l=1}^{L_k} g_{k,l}^{\mathrm{ul}} \mathbf{a}(\theta_{k,l}, \phi_{k,l}) \otimes \mathbf{p}(\tau_{k,l}) + \mathbf{z}_k^{\mathrm{ul}}, \tag{5}$$

where $\mathbf{z}_k^{\mathrm{ul}} \in \mathbb{C}^{MN\times 1}$ is the additive noise vector on all subcarriers of OFDM symbol k and on all antenna elements, and each element of $\mathbf{z}_k^{\mathrm{ul}}$ is i.i.d. with zero mean and unit variance. We aim to extract $\{\tau_{k,l}, \theta_{k,l}, \phi_{k,l}\}_{l=1,\ldots,L_k}$ from the noised mixture $\mathbf{y}_k^{\mathrm{ul}}$.

This frequency estimation problem can be solved by utilizing the NOMP algorithm which detects frequencies from the noised mixture of multiple sinusoids. However, both the original NOMP algorithm in [22] and the extended NOMP algorithm proposed in [21] do not cover the case when three types of frequencies are to be extracted. Hence, in this paper, we further propose an e-NOMP algorithm to suit the massive MIMO-OFDM system.

B. e-NOMP for Massive MIMO-OFDM System

E-NOMP is an iteration-based algorithm which extracts a new component within each iteration. To precisely match the practical component, e-NOMP refines the e-OMP results through e-Newton steps. When the algorithm terminates, the number of extracted components is exactly the number of practical components if each component is accurately estimated.

By the end of the ith iteration of the e-NOMP algorithm, the ith component path will be removed from the noised mixture. If the parameters are precisely estimated, the ith component path will be completely eliminated and the residual of noised mixture will be minimized. The e-OMP step and the e-Newton step in the ith iteration are designed base on this target. Now we give a detailed description of the e-OMP step and the e-Newton step in the ith iteration of the e-NOMP algorithm.

1) e-OMP Step: At the beginning, the residual of noised mixture is expressed as

$$\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i) = \mathbf{y}_k^{\mathrm{ul}} - \sum_{l=1}^{i-1} \hat{g}_{k,l}^{\mathrm{ul}} \mathbf{a}(\hat{\theta}_{k,l}, \hat{\phi}_{k,l}) \otimes \mathbf{p}(\hat{\tau}_{k,l}), \tag{6}$$

where $\hat{g}_{k,l}^{\mathrm{ul}}$, $\hat{\theta}_{k,l}$, $\hat{\phi}_{k,l}$ and $\hat{\tau}_{k,l}$ are the gain, downtilt, azimuth and delay estimated in the lth iteration. In the e-OMP step, we exhaustively search a pre-defined codebook and find the optimal codeword

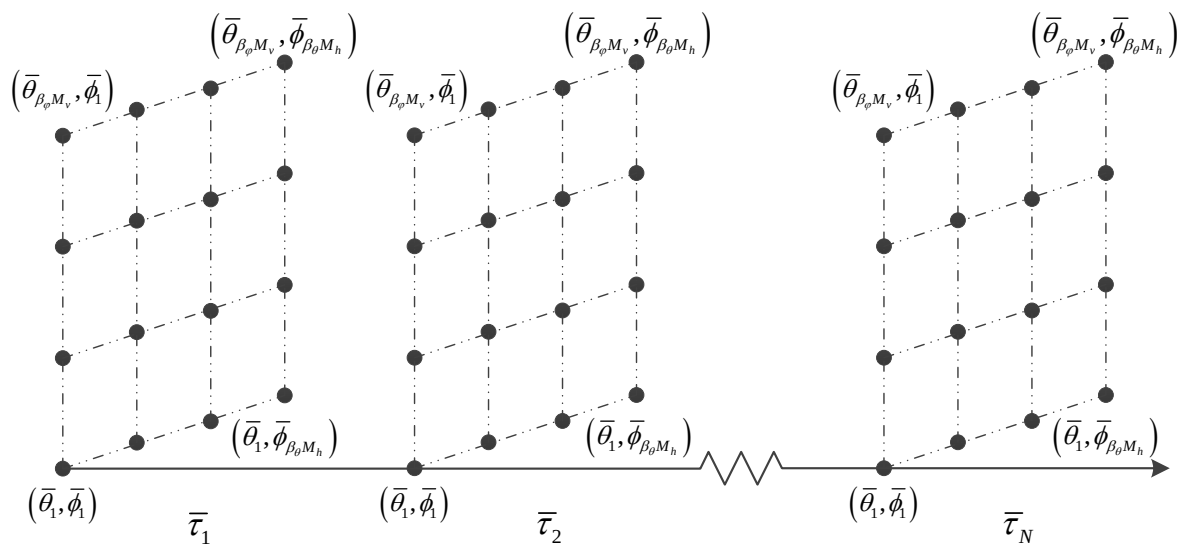

Fig. 2. Codebook used in the e-NOMP algorithm. Each vertical plane represents a lower dimensional sub-codebook that covers the sampled downtilts and azimuths on a sampled delay.

that best matches $\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)$.

To fit the massive MIMO-OFDM scenario, here, we design the codebook according to the structure of a component path in (5). A codeword is expressed as

$$\mathbf{c}(\bar{\theta}, \bar{\phi}, \bar{\tau}) = \mathbf{a}(\bar{\theta}, \bar{\phi}) \otimes \mathbf{p}(\bar{\tau}), \tag{7}$$

where $\bar{\theta} \in [-\pi/2, \pi/2)$, $\bar{\phi} \in [-\pi/2, \pi/2)$ and $\bar{\tau} \in [0, 1/\triangle f)$ are the downtilt, azimuth and delay that $\mathbf{c}$ represents. The codebook covers the 3D space and the delay domain that are concerned. Thus, we sample the angles and delays uniformly as

$$\begin{aligned}
&\bar{\theta} \in \{\bar{\theta}_1 = -\frac{\pi}{2}, \bar{\theta}_2 = -\frac{\pi}{2} + \frac{\pi}{\beta_\theta M_v}, \ldots, \bar{\theta}_{\beta_\theta M_v} = -\frac{\pi}{2} + \frac{(\beta_\theta M_v - 1)\pi}{\beta_\theta M_v}\}, \\
&\bar{\phi} \in \{\bar{\phi}_1 = -\frac{\pi}{2}, \bar{\phi}_2 = -\frac{\pi}{2} + \frac{\pi}{\beta_\phi M_h}, \ldots, \bar{\phi}_{\beta_\phi M_h} = -\frac{\pi}{2} + \frac{(\beta_\phi M_h - 1)\pi}{\beta_\phi M_h}\}, \\
&\bar{\tau} \in \{\bar{\tau}_1 = 0, \bar{\tau}_2 = \frac{1}{\beta_\tau N \triangle f}, \ldots, \bar{\tau}_{\beta_\tau N} = \frac{\beta_\tau N - 1}{\beta_\tau N \triangle f}\},
\end{aligned} \tag{8}$$

where β_θ, β_ϕ and β_τ are the over-sampling rates of the downtilt, azimuth and delay, respectively. As shown in Fig. 2, the black circles are the codewords. Each codeword points to a sampled spatial direction and covers a sampled delay. The e-OMP step selects the codeword that has the

maximum projection power from $\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)$, that is,

$$\mathbf{c}(\hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i}) = \arg\max_{(\bar{\theta},\bar{\phi},\bar{\tau})} \frac{|\mathbf{c}^H(\bar{\theta}, \bar{\phi}, \bar{\tau})\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)|^2}{\|\mathbf{c}(\bar{\theta}, \bar{\phi}, \bar{\tau})\|^2}, \tag{9}$$

where $\hat{\theta}_{k,i}$, $\hat{\phi}_{k,i}$ and $\hat{\tau}_{k,i}$ are the coarsely estimated downtilt, azimuth and delay of the ith component path. Then, the gain of the ith component path is calculated by

$$\hat{g}_{k,i}^{\mathrm{ul}} = \frac{\mathbf{c}^H(\hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i})\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)}{\|\mathbf{c}(\hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i})\|^2}. \tag{10}$$

It can be observed that the total number of codewords equals $\beta_\theta M_v \times \beta_\phi M_h \times \beta_\tau N$. Increasing the over-sampling rates helps improve the matching between $\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)$ and $\mathbf{c}(\hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i})$. However, it also multiplies the searching time and severely depresses the efficiency, especially when M_v, M_h and N are large. Therefore, the over-sampling rates are usually small when we design codebook for massive MIMO systems.

2) e-Newton Step: Before removing the ith component path from the noised mixture at the final stage of e-OMP, the e-Newton step is applied to tackle the off-grid effect and adjust the estimates towards the real values. The goal of minimizing the residual of noised mixture can be translated to maximizing

$$S(g^{\mathrm{ul}}, \theta, \phi, \tau) = 2\Re\left\{\mathbf{y}_{\mathrm{r}}^{\mathrm{ul}H}(i) g^{\mathrm{ul}} \mathbf{c}(\theta, \phi, \tau)\right\} - |g^{\mathrm{ul}}|^2 \|\mathbf{c}(\theta, \phi, \tau)\|^2. \tag{11}$$

The e-Newton step is designed to refine the downtilt, azimuth and delay simultaneously by

$$\begin{bmatrix} \hat{\theta}'_{k,i} \\ \hat{\phi}'_{k,i} \\ \hat{\tau}'_{k,i} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{k,i} \\ \hat{\phi}_{k,i} \\ \hat{\tau}_{k,i} \end{bmatrix} - \ddot{\mathbf{S}}\left(\hat{g}_{k,l}^{\mathrm{ul}}, \hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i}\right)^{-1} \dot{\mathbf{S}}\left(\hat{g}_{k,l}^{\mathrm{ul}}, \hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i}\right), \tag{12}$$

where

$$\dot{\mathbf{S}}\left(g^{\mathrm{ul}}, \theta, \phi, \tau\right) = \begin{bmatrix} \frac{\partial S}{\partial \theta} \\ \frac{\partial S}{\partial \phi} \\ \frac{\partial S}{\partial \tau} \end{bmatrix}, \ddot{\mathbf{S}}\left(g^{\mathrm{ul}}, \theta, \phi, \tau\right) = \begin{bmatrix} \frac{\partial^2 S}{\partial \theta^2} & \frac{\partial^2 S}{\partial \theta \partial \phi} & \frac{\partial^2 S}{\partial \theta \partial \tau} \\ \frac{\partial^2 S}{\partial \phi \partial \theta} & \frac{\partial^2 S}{\partial \phi^2} & \frac{\partial^2 S}{\partial \phi \partial \tau} \\ \frac{\partial^2 S}{\partial \tau \partial \theta} & \frac{\partial^2 S}{\partial \tau \partial \phi} & \frac{\partial^2 S}{\partial \tau^2} \end{bmatrix}. \tag{13}$$

In (11), we regard $\mathbf{y}_{\mathrm{r}}^{\mathrm{ul}}$ and g^{ul} as constant, and the derivation of S is transformed to the derivation of the codeword $\mathbf{c}$. We take the partial derivatives of S versus θ as the examples. The first-order

Algorithm 1 Working Steps of An Iteration in e-NOMP

Step 1: New detection. Coarsely estimate downtilt, azimuth and delay of a component path by searching for the best-matched codeword in the codebook.
Step 2: Single Newton Refinement. Apply Newton refinement to the newly detected component path.
Step 3: Cyclic Newton Refinement. Cyclicly apply Newton refinement to all the detected component paths one by one.
Step 4: Gains Update. Update the gains of all the detected component paths by utilizing the refined parameters.

partial derivative is calculated as

$$\frac{\partial S}{\partial \theta} = 2\Re\left\{\mathbf{y}_{\mathrm{r}}^{\mathrm{ul}H} g^{\mathrm{ul}} \frac{\partial \mathbf{c}}{\partial \theta} - \left|g^{\mathrm{ul}}\right|^2 \mathbf{c}^H \frac{\partial \mathbf{c}}{\partial \theta}\right\}. \tag{14}$$

The second-order partial derivative and the cross partial derivative are

$$\frac{\partial^2 S}{\partial \theta^2} = 2\Re\left\{\left(\mathbf{y}_{\mathrm{r}}^{\mathrm{ul}H} g^{\mathrm{ul}} - \left|g^{\mathrm{ul}}\right|^2 \mathbf{c}^H\right) \frac{\partial^2 \mathbf{c}}{\partial \theta^2}\right\} - 2\left|g^{\mathrm{ul}}\right|^2 \|\frac{\partial \mathbf{c}}{\partial \theta}\|^2 \tag{15}$$

and

$$\frac{\partial^2 S}{\partial \theta \partial \phi} = 2\Re\left\{\left(\mathbf{y}_{\mathrm{r}}^{\mathrm{ul}H} g^{\mathrm{ul}} - \left|g^{\mathrm{ul}}\right|^2 \mathbf{c}^H\right) \frac{\partial^2 \mathbf{c}}{\partial \theta \partial \phi} - \left|g^{\mathrm{ul}}\right|^2 \frac{\partial \mathbf{c}^H}{\partial \phi} \frac{\partial \mathbf{c}}{\partial \theta}\right\} \tag{16}$$

respectively. Other derivatives can be derived in a similar way. The gain of the ith component path will be updated through (10) by replacing $(\hat{\theta}_{k,i}, \hat{\phi}_{k,i}, \hat{\tau}_{k,i})$ with $(\hat{\theta}'_{k,i}, \hat{\phi}'_{k,i}, \hat{\tau}'_{k,i})$. Then, the refined ith component path will be removed from $\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}(i)$.

Alg. 1 briefly summarizes the working steps of an iteration in the e-NOMP algorithm for massive MIMO-OFDM system. Stopping criterion of the e-NOMP iterations is designed based on the required false alarm rate P_{fa} [22]. The algorithm terminates when

$$\|\mathcal{F}\{\mathbf{y}_{\mathrm{r},k}^{\mathrm{ul}}\}\|_\infty^2 > -\log(1-(1-P_{\mathrm{fa}})^{\frac{1}{MN}}), \tag{17}$$

where $\mathcal{F}\{\cdot\}$ represents taking Fourier transformation, and $\|\cdot\|_\infty$ denotes the infinite norm. Finally, the BS obtains the estimated frequency-independent parameters of all the user channels, which are denoted by $\{\hat{\tau}_{k,l}, \hat{\theta}_{k,l}, \hat{\phi}_{k,l}\}$, where $l = 1, \ldots, \hat{L}_k$, $k = 1, \ldots, K$.

IV. Efficient Downlink Channel Reconstruction

This section aims to solve the second problem mentioned in *Section II*, that is, how to estimate the downlink gains for multiple users by using limited overhead. Here, we formulate the requirement on a successful estimation and propose an efficient downlink channel reconstruction scheme which utilizes a small amount of overhead to estimate the downlink gains of each user channel.

A. Beamforming for Downlink Training

To reconstruct the downlink channel, we need to estimate the downlink gains, which will cost downlink resources for pilots and uplink resources for the feedback of gains. Since the BS has known the majority of spatial parameters (angles and delays) to reconstruct the downlink channel, the amount of required resources for downlink training and feedback is small.

We first analyze the requirements for successful downlink training. For a certain user in the system, its real spatial parameters are $\{\tau_l, \theta_l, \phi_l\}_{l=1,\ldots,L}$ and their estimates are $\{\hat{\tau}_l, \hat{\theta}_l, \hat{\phi}_l\}_{l=1,\ldots,\hat{L}}$. During the downlink gain refinement phase, the BS transmits downlink pilots in J successive OFDM symbols and alters the beamforming weight every OFDM symbol.

Suppose comb-type all-1 pilots are used and these pilots are sparsely and uniformly inserted in the downlink frequency band. The received pilots on the jth OFDM symbol are

$$\mathbf{y}(j) = \sum_{l=0}^{L-1} \sqrt{P} g_l^{\mathrm{dl}} \mathbf{p}_p(\tau_l) \mathbf{a}^T(\theta_l, \phi_l) \mathbf{b}_j + \mathbf{z}(j), \tag{18}$$

where P is the transmit power,

$$\mathbf{p}_p(\tau) = \left[e^{j2\pi(f^{\mathrm{dl}} - f^{\mathrm{ul}} + n_1 \triangle f)\tau}, \ldots, e^{j2\pi(f^{\mathrm{dl}} - f^{\mathrm{ul}} + n_{N_p} \triangle f)\triangle f \tau} \right]^T \tag{19}$$

describes the delay on N_p downlink subcarriers that are occupied by downlink pilots, $\mathbf{b}_j \in \mathbb{C}^{M\times 1}$ represents the beamforming weight that is used on the jth OFDM symbol, and $\mathbf{z}(k)$ is the noise vector on the jth OFDM symbol whose elements are i.i.d. with zero mean and unit variance. Since L, θ_l, ϕ_l and τ_l have been estimated in the uplink and will be sent to this user, in the following derivation we replace them by $\hat{L}$, $\hat{\theta}_l$, $\hat{\phi}_l$ and $\hat{\tau}_l$. We denote

$$\Theta^{l,j} = \mathbf{a}^T(\hat{\theta}_l, \hat{\phi}_l) \mathbf{b}_j \tag{20}$$

to simplify the expressions. By stacking all the received pilots into a big vector, we obtain

$$\mathbf{y} = \sqrt{P}\mathbf{A}\mathbf{g}^{\mathrm{dl}} + \mathbf{z}, \tag{21}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(J) \end{bmatrix}, \mathbf{g}^{\mathrm{dl}} = \begin{bmatrix} g_1^{\mathrm{dl}} \\ \vdots \\ g_{\hat{L}}^{\mathrm{dl}} \end{bmatrix}, \mathbf{z} = \begin{bmatrix} \mathbf{z}(1) \\ \vdots \\ \mathbf{z}(J) \end{bmatrix} \tag{22}$$

are the stacked received pilot, downlink gain and noise vectors, respectively, and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}(1,1) & \cdots & \mathbf{A}(1,\hat{L}) \\ \vdots & & \vdots \\ \mathbf{A}(J,1) & \cdots & \mathbf{A}(J,\hat{L}) \end{bmatrix} \tag{23}$$

is the coefficient matrix with submatrix

$$\mathbf{A}(j,l) = e^{j2\pi(f^{\mathrm{dl}}-f^{\mathrm{ul}})\hat{\tau}_l}\Theta^{l,j}\mathbf{p}(\hat{\tau}_l). \tag{24}$$

Then, the gains can be LS estimated by

$$\hat{\mathbf{g}}^{\mathrm{dl}} = \frac{1}{\sqrt{P}}\mathbf{A}^{\dagger}\mathbf{y} = \frac{1}{\sqrt{P}}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{y}. \tag{25}$$

Obviously, the first requirement for the successful estimation of $\mathbf{g}^{\mathrm{dl}}$ is that $\mathbf{A}^H\mathbf{A}$ is invertible. Thus, $\mathbf{A}$ must have full column-rank, that is, $\mathrm{rank}(\mathbf{A}) = \hat{L}$. An implied condition for full column-rank is that $N_pJ \geq \hat{L}$, which means the number of downlink pilots should be no less than the number of downlink gains to be estimated. We apply singular-value-decomposition on the coefficient matrix $\mathbf{A}$ by $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{C}^{N_pJ\times N_pJ}$ and $\mathbf{V} \in \mathbb{C}^{\hat{L}\times\hat{L}}$ are unitary matrices, and $\mathbf{\Lambda} \in \mathbb{C}^{N_pJ\times\hat{L}}$ satisfies

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_{\hat{L}} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, \tag{26}$$

where $\lambda_1 \geq \ldots \geq \lambda_{\hat{L}}$ are the singular values. To make sure that $\mathbf{A}^H\mathbf{A} = \mathbf{U}\mathbf{\Lambda}^H\mathbf{\Lambda}\mathbf{V}^H$ has inverse,

the smallest singular value must hold that $|\lambda_{\hat{L}}|^2 > 0$.

We rewrite (21) by

$$\underline{\mathbf{y}} = \sqrt{P}\mathbf{\Lambda}\underline{\mathbf{g}}^{\mathrm{dl}} + \underline{\mathbf{z}}, \tag{27}$$

where $\underline{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$, $\underline{\mathbf{g}}^{\mathrm{dl}} = \mathbf{V}^H\mathbf{g}^{\mathrm{dl}}$ and $\underline{\mathbf{z}} = \mathbf{U}^H\mathbf{z}$. Then, the LS estimation of $\underline{\mathbf{g}}^{\mathrm{dl}}$ can be derived by

$$\hat{\underline{\mathbf{g}}}^{\mathrm{dl}} = \underline{g}^{\mathrm{dl}} + \frac{1}{\sqrt{P}}(\mathbf{\Lambda}^H\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^H\underline{\mathbf{z}}. \tag{28}$$

To be specific, for the lth estimated path,

$$\hat{\underline{g}}_l^{\mathrm{dl}} = \underline{g}_l^{\mathrm{dl}} + \frac{\lambda_l^*}{\sqrt{P}|\lambda_l|^2}\underline{z}_l. \tag{29}$$

The normalized mean square error (NMSE) of the estimated downlink gains is defined as

$$\mathrm{NMSE} = \frac{\mathbb{E}\{\|\hat{\underline{\mathbf{g}}}^{\mathrm{dl}} - \underline{\mathbf{g}}^{\mathrm{dl}}\|^2\}}{\mathbb{E}\{\|\underline{\mathbf{g}}^{\mathrm{dl}}\|^2\}} = \frac{1}{P\|\underline{\mathbf{g}}^{\mathrm{dl}}\|^2}\sum_{l=1}^{\hat{L}}\frac{\mathbb{E}\{|\underline{z}_l|^2\}}{|\lambda_l|^2}. \tag{30}$$

The unitary character of $\mathbf{U}$ determines that the new noise vector $\underline{\mathbf{z}}$ holds the same statistic characters with $\mathbf{z}$. Thus, the elements of $\underline{\mathbf{z}}$ are also i.i.d. with zero mean and unit variance, that is, $\mathbb{E}\{|\underline{z}_l|^2\} = 1$. Besides, according to [16], the spatial power difference between the uplink and downlink is small. We use $\|\mathbf{g}^{\mathrm{ul}}\|^2$ to approximates $\|\underline{\mathbf{g}}^{\mathrm{dl}}\|^2$. Then, the NMSE of the gains can be calculated as

$$\mathrm{NMSE} = \frac{1}{P\|\mathbf{g}^{\mathrm{ul}}\|^2}\sum_{l=1}^{\hat{L}}\frac{1}{|\lambda_l|^2}. \tag{31}$$

Definition 1: Estimation of the downlink gains of a user is successful when the practical NMSE is below an acceptable error rate δ, that is,

$$\frac{1}{P\|\mathbf{g}^{\mathrm{ul}}\|^2}\sum_{l=1}^{\hat{L}}\frac{1}{|\lambda_l|^2} < \delta \tag{32}$$

where $0 < \delta < 1$ is the acceptable error rate.

Thus, from (31) we can formulate the requirement for a successful estimation as

$$\sum_{l=1}^{\hat{L}}\frac{1}{|\lambda_l|^2} < \delta P\|\mathbf{g}^{\mathrm{ul}}\|^2. \tag{33}$$

In the following subsection, we will base on (33) and design downlink training beamforming

vectors $\mathbf{b}_j, j = 1, \ldots, J$ to achieve successful estimations.

B. Spatial Angle Grid

If there is only one user in the cell, all the downlink pilots can be dedicated to this user and beamformed to the uplink estimated spatial angles. In this condition, $J = \hat{L}$, $\mathbf{b}_j = \mathbf{a}^*(\hat{\theta}_l, \hat{\phi}_l)/\sqrt{M}$, and $\Theta^{j,j}$ is equal to its maximum value $\sqrt{M}$. From (31), we can see that the estimation error is inversely proportional to $|\lambda_1|^2, \ldots, |\lambda_{\hat{L}}|^2$. If $\Theta^{j,j} \gg \Theta^{l,j}$ for $l \neq j$, $|\lambda_1|, \ldots, |\lambda_{\hat{L}}|$ approaches $\sqrt{M}$. Then, we can obtain a precise estimate with extremely small error.

However, in the multiuser system that is concerned, it has a high probability that different user channels have entirely different spatial angles. A simple extension to the multiuser scenario is to perform the single user downlink gain estimation process for each user one by one. This method will be extremely time consuming if the total number of the paths to be estimated are large. More than that is exceeding the correlation time of the channel, which will destroy the effectiveness of the reconstruction results.

To make a balance between precision and efficiency, we introduce a spatial angle grid and project the original angles to the angles on the grid. The downlink pilots are beamformed to the angles on the grid, instead of on the estimated angles. The spatial angle grid covers the whole space and samples the spatial directions uniformly. We have defined a codebook with similar usage in e-NOMP algorithm. As shown Fig. 2, each vertical plane represents a lower dimensional sub-codebook that uniformly samples the whole 3D space. Therefore, we can reuse this sub-codebook as the spatial angle grid. Each spatial angle on the grid corresponds to a downlink training beam. Here, we set $\beta_\theta = \beta_\phi = 1$ to restrict the quantity of downlink training beams.

For the ith grid point, $i = 1, \ldots, M_v M_h$, we denote $(\bar{\theta}, \bar{\phi})_i = (\bar{\theta}_{i_v}, \bar{\phi}_{i_h})$, where

$$\bar{\theta}_{i_v} = \frac{\pi}{M_v}(i_v - \frac{M_v}{2} - 1), \bar{\phi}_{i_h} = \frac{\pi}{M_h}(i_h - \frac{M_h}{2} - 1), \tag{34}$$

and the corresponding indices of the sampled downtilt and azimuth angles are

$$i_v = \lceil \frac{i}{M_h} \rceil, i_h = i - M_h(i_v - 1), \tag{35}$$

respectively. Each grid point corresponds to a beamforming vector $\mathbf{a}^*(\bar{\theta}, \bar{\phi})/\sqrt{M}$.

Each estimated pair of downtilt and azimuth angles is rounded to its "nearest" spatial grid point which has the largest projection power from the original estimated angle pair. For the lth estimated path of user k, the projection power on grid point $(\bar{\theta}, \bar{\phi})$ is defined as

$$P_{\mathrm{e}\to\mathrm{g}}^{k,l}(\bar{\theta}, \bar{\phi}) = \frac{1}{M}|\mathbf{a}^T(\hat{\theta}_{k,l}, \hat{\phi}_{k,l})\mathbf{a}^*(\bar{\theta}, \bar{\phi})|^2. \tag{36}$$

To enhance the receiving power at the user side, we calculate the projection power on all the grid points, select the one with the maximum value

$$(\bar{\theta}^{k,l}, \bar{\phi}^{k,l}) = \underset{\substack{\bar{\theta} = \bar{\theta}_1, \cdots, \bar{\theta}_{M_v} \\ \bar{\phi} = \bar{\phi}_1, \cdots, \bar{\phi}_{M_h}}}{\arg\max} \quad P_{\mathrm{e}\to\mathrm{g}}^{k,l}(\bar{\theta}, \bar{\phi}) \tag{37}$$

and mark this grid point.

The main advantages of using a spatial angle grid are: (1) The estimated angle pair can be projected to a grid point and the projection power is large enough for the successful estimation of downlink gains. (2) The downlink dedicated pilots are translated to downlink common pilots, and all the users can receive these common pilots simultaneously without causing interference.

C. DL Training Strategy

Having known the projected grid points from K users, the BS records all the marked grid points $(\bar{\theta}, \bar{\phi})_{i_1}, \ldots, (\bar{\theta}, \bar{\phi})_{i_b}$ and includes them in the selected grid angle set Φ. Then, b OFDM symbols will be occupied by the downlink pilots for the estimation of the downlink gains of each user. In a massive MIMO system, M and K are always large, then b is large as well. To ensure that enough time is remained for data transmission within the correlation period, we hope to reduce the overhead for training.

To satisfy the requirement for a successful estimation of the gains, we project each estimated angle pair to its nearest grid point to promise the high receiving SNR. If the downlink pilot transmit power is large, the resulted $\sum_{l=1}^{\hat{L}} 1/|\lambda_l|^2$ is much lower than the threshold $\delta|\underline{\mathbf{g}}^{\mathrm{ul}}|^2$. Hence, the nearest grid points are the optimal but not necessary solutions. According to [19], when the number of antennas is large but not unlimited, the estimated angle pair has significant projection power on more than one grid angle. If the projection power on other grid points is large enough and the resulted singular values of $\mathbf{A}$ satisfy (33), the suboptimal options can work as well. This gives the possibility to cut down grid points in Φ.

1) Share A Common Grid Point: Suppose only one user exists in the cell. A special case is that the propagation paths are spatially-closed with each other. Even if these paths are projected on different optimal grid points, it may happen that they share a common suboptimal grid point. When BS beamforms the downlink pilots on the direction of suboptimal grid point, the coefficient matrix $\mathbf{A}$ becomes

$$\mathbf{A} = \left[\Theta^{1,1}\mathbf{p}_p(\hat{\tau}_1), \Theta^{2,1}\mathbf{p}_p(\hat{\tau}_2), \ldots, \Theta^{\hat{L},1}\mathbf{p}_p(\hat{\tau}_{\hat{L}})\right]. \tag{38}$$

Firstly, $\mathbf{p}_p(\hat{\tau}_l)$ is a vector with completely different elements, and $\hat{\tau}_l$ is not equal to each other for $l = 1, \ldots, \hat{L}$. If $\mathbf{A}$ satisfies (33), then the downlink gains can still be successfully estimated, showing that multiple optimal grid points can be replaced by this single suboptimal grid point. In this condition, $|\Phi|$ equals 1 rather than $\hat{L}$, saving training overhead by $\hat{L}-1$ OFDM symbols.

2) Utilize Grid Points Selected by Other Users: In the multiuser system that is concerned, Φ contains a large number of optimal grid points selected by different users. For the l_1th estimated path of user k_1, it has a high possibility that one of the path's suboptimal grid point is selected as an optimal grid point by the l_2th estimated path of user k_2. Since user k_1 is able to receive the broadcasted downlink pilots that are beamformed to the optimal grid point selected by user k_2, the suboptimal grid point can be utilized by user k_1 as well. Besides, receiving useless downlink pilots does not impact the success of estimation. When $\mathbf{A}$ already satisfies (33), the newly added low-SNR measurements makes negligible contribution to the LS estimation, further making little difference to the estimation results. Thus, each user can receive all the downlink pilots, whether they are useful or useless.

3) Beam Scheduling Scheme: Based on these findings, we now exclude the redundant grid points in Φ and only remain the useful grid points shared by multiple users. For each user, if the derived $\mathbf{A}$ satisfies (33) and no more grid points can be excluded in Φ, then we obtain the final set of selected grid points. Each selected grid point will be transformed to a downlink beam. To save the downlink training overhead, we write the target of the beam scheduling scheme as

$$\arg\min |\Phi|, \quad s.t. \quad (33) \quad \text{holds} \quad \text{for} \quad k = 1, \ldots, K. \tag{39}$$

If we exhaustively search all the possible grid angle subsets of Φ to find the optimum subset, the scheduling will be extremely time-consuming. Thus, we follow a similar approach as the Greedy method but use it in a converse way. The redundant grid angles will be found out and

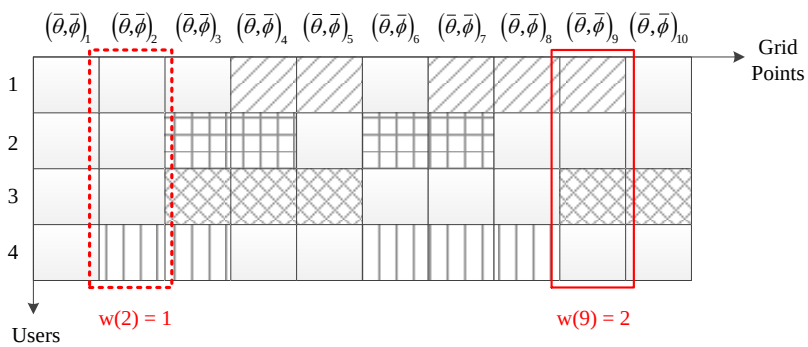

Fig. 3. Weights of grid points. Grid point $(\bar{\theta},\bar{\phi})_2$ has the first minimum weight and will be abandoned first.

excluded from Φ one by one. Here, we introduce the concept of "weight" for each grid point. For grid point $(\bar{\theta},\bar{\phi})_i$, if it is chosen by $\bar{K}$ users, then its weight is calculated as $w(i) = \bar{K}$. As shown in Fig. 3, the weights of grid points $(\bar{\theta},\bar{\phi})_2$ and $(\bar{\theta},\bar{\phi})_9$ are 1 and 2, respectively. We remain the grid points that are shared by most users and abandon the grid points that are chosen by only one user. Thus, we compare $w(i^1),\ldots,w(i^b)$ and find the first smallest weight $w(i^j)$. Then, grid point $(\bar{\theta},\bar{\phi})_{i^j}$ will be excluded in Φ first if needed. In Fig. 3, grid point $(\bar{\theta},\bar{\phi})_2$ has the first smallest weight and it will be abandoned first.

We keep excluding grid points whose weights are small, till no more grid points can be excluded in Φ. Working principle of the beam scheduling scheme is summarized in Alg 2. We denote the final beam scheduling results as follows:

- Selected grid points with indices $i^1,\ldots,i^{T_p}$.

Before the start of the downlink pilot transmission phase, each selected user is informed with its delays, angles, and the grid angles for the downlink pilots. Then, the BS broadcasts cell-common pilots in successive T_p OFDM symbols. Each user receives all the pilots and calculates their downlink gains according to (25). The estimated downlink gains are sent back to the BS. Utilizing the uplink-estimated downtilts, azimuths and delays and the downlink-estimated gains, the BS reconstructs the downlink channel of user k as

$$\hat{\mathbf{h}}_k^{\mathrm{dl}} = \sum_{l=1}^{\hat{L}_k} \hat{g}_{k,l}^{\mathrm{dl}} \mathbf{a}^T(\hat{\theta}_{k,l},\hat{\phi}_{k,l}) \otimes \mathbf{p}^T(\hat{\tau}_{k,l}) e^{j2\pi(f^{\mathrm{dl}}-f^{\mathrm{ul}})\hat{\tau}_{k,l}}. \tag{40}$$

Finally, the BS obtains $\hat{\mathbf{h}}_1^{\mathrm{dl}},\ldots,\hat{\mathbf{h}}_K^{\mathrm{dl}}$.

Algorithm 2 Beam Scheduling Strategy

Require: Selected grid point set Φ
Initialize: $\Phi = \{i^1, \ldots, i^b\}$
1: Calculate weights of grid points in Φ
2: Reorder Φ in increasing weight
for $j = 1, \ldots, b$ **do**
  1) Set $\Phi_{\text{temp}} = [\Phi \backslash i^j]$, flag = 1
  2) **for** $k = 1, \ldots, K$ **do**
    a) **if** (33) does not hold
    b) set flag = 0, **break**,
    c) **end if**
  3) **end for**
  4) **if** flag = 1, set $\Phi = \Phi_{\text{temp}}$
  5) **else if** flag = 0, set $\Phi = \Phi$
  6) **end if**
end for
Output: Φ

V. Downlink Multiuser Transmission

With the downlink reconstructed channels of all the users, the BS is able to make user scheduling and design transmission schemes. It is worth noticing that the BS has obtained full CSI. Therefore, if the number of users is much smaller than the number of BS antennas, user scheduling is not necessary and these users can be simultaneously served.

During the downlink transmission phase, the BS sends data streams to all the K users simultaneously. To overcome interference, the BS adopts ZF precoding before transmitting downlink data. The ZF precoder on downlink subcarrier n is expressed as

$$\mathbf{W}(n) = \hat{\mathbf{H}}^{\dagger}(n)\mathbf{\Lambda}(n), \tag{41}$$

where $\hat{\mathbf{H}}(n) \in \mathbb{C}^{K\times M}$ is the multiuser channel matrix on downlink subcarrier n satisfying $[\hat{\mathbf{H}}(n)]_{k,:} = \hat{\mathbf{h}}_k^{\text{dl}}(n)$, and $\mathbf{\Lambda}(n) = diag\{\alpha_1, \ldots, \alpha_K\}$ is the normalization matrix. Here, we choose uniform power allocation strategy and set

$$\alpha_k = \frac{1}{\sqrt{K}\| \left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k} \|}. \tag{42}$$

For user k, its received data on subcarrier n comprises both the target data and the inter-user

interference, that is,

$$\begin{aligned} r_k(n) &= \sqrt{P}\mathbf{h}_k^{\mathrm{dl}}(n)\left[\mathbf{W}(n)\right]_{:,k} d_k(n) + \sum_{j\neq s}\sqrt{P}\mathbf{h}_k^{\mathrm{dl}}(n)\left[\mathbf{W}(n)\right]_{:,k} d_j(n) + z_k(n) \\ &= \sqrt{P}\alpha_k\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k} d_k(n) + \sum_{j\neq s}\sqrt{P}\alpha_j\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k} d_j(n) + z_k(n), \end{aligned} \tag{43}$$

where $r_k(n)$ is the received data by user k on subcarrier n, P is the total downlink transmit power, $d_k(n)$ is the transmit data with unit power, and $z_k(n)$ is the noise with unit variance.

The average multiuser sum-rate of this system is calculated as

$$R = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\log_2(1+\mathrm{SINR}_k(n)), \tag{44}$$

where $\mathrm{SINR}_k(n)$ is the signal-to-interference-noise ratio (SINR) at user k on subcarrier n and can be expressed as

$$\mathrm{SINR}_k(n) = \frac{P\alpha_k^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}|^2}{\sum\limits_{j\neq k}P\alpha_j^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,j}|^2 + |z_k(n)|^2}. \tag{45}$$

It can be observed that the multiuser sum-rate increases in proportion to SINR. Thus, we evaluate the achievable rate of a single user k by analyzing its SINR. To simplify the expressions, we neglect the subcarrier index n in the following derivations. *Theorem 1* provides the approximation of the expectation of SINR_k when the acceptable error rate equals δ.

Theorem 1: The expectation of the SINR at user k approximates

$$\mathbb{E}\{\mathrm{SINR}_k\} \approx \frac{S_k}{I_k+1}, \tag{46}$$

where

$$\begin{aligned} S_k &= \frac{P(1+\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}{K(\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2+\delta\sum_{j=1}^{K}\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{j,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}, \\ I_k &= \sum_{j\neq k}\frac{P(\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,j}|^2)}{K(\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2+\delta\sum_{i=1}^{K}\|\left[\mathbf{H}^{\dagger}\right]_{:,i}\|^2\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{i,m}\left[\mathbf{H}^{\dagger}\right]_{m,j}|^2)} \end{aligned} \tag{47}$$

can be viewed as the expected signal power and the expected interference power, respectively.

Proof: The proof of *Theorem 1* can be found in Appendix. ■

From *Theorem 1*, we can see that

$$S_k = \frac{P(1+\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}{K\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2(1+\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)+K\delta X}, \tag{48}$$

where $X = \sum_{j\neq k}\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{j,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2 > 0$. Then, it holds that

$$S_k < \frac{P(1+\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}{K\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2(1+\delta\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)} = \frac{P}{K\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2}, \tag{49}$$

which is the expected signal power when $\delta = 0$. The expected signal power degrades if error is acceptable during the LS estimation. When the the value of δ becomes large, the expected signal power decreases.

As to the interference item, it can be found that only if $\delta = 0$ and the reconstructed downlink multiuser channel are precise, the interference can be completely eliminated. Otherwise, $I_k > 0$, showing that the interference exists. After rewriting the expression of I_k by

$$I_k = \sum_{j\neq k}\frac{P(\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,j}|^2)}{K(\delta^{-1}\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2+\sum_{i=1}^{K}\|\left[\mathbf{H}^{\dagger}\right]_{:,i}\|^2\sum_{m=1}^{M}|\left[\mathbf{H}\right]_{i,m}\left[\mathbf{H}^{\dagger}\right]_{m,j}|^2)}, \tag{50}$$

we can easily see that the interference increases in proportion to δ.

In summary, when δ increases, signal power degrades and interference becomes more severe. Then, SINR decreases, which further results in a significant degradation of multiuser sum-rate. To prevent the slash of multiuser sum-rate performance, the value of δ should be small enough. However, the amount of downlink training resource will not be efficiently reduced when the NMSE requirement is too strict. Therefore, we should choose a proper δ to balance between the rate performance and the resources economization.

VI. Numerical Results

In this section, we evaluate the performance of the downlink channel reconstruction-based FDD massive MIMO transceiver design. There are 16 antenna elements in a row of the UPA at the BS and 8 elements in a column. The uplink/downlink frequency band is comprised of 256 sub-carriers and the subcarrier space is 75 kHz. The carrier separation between uplink and downlink is 300 MHz. In the downlink training phase, the pilots are inserted in every 4 subcarriers. We set $P_{\text{fa}} = 10^{-2}$ and $\beta_\tau = 1$, $\beta_\theta = \beta_\phi = 2$ when conducting the developed

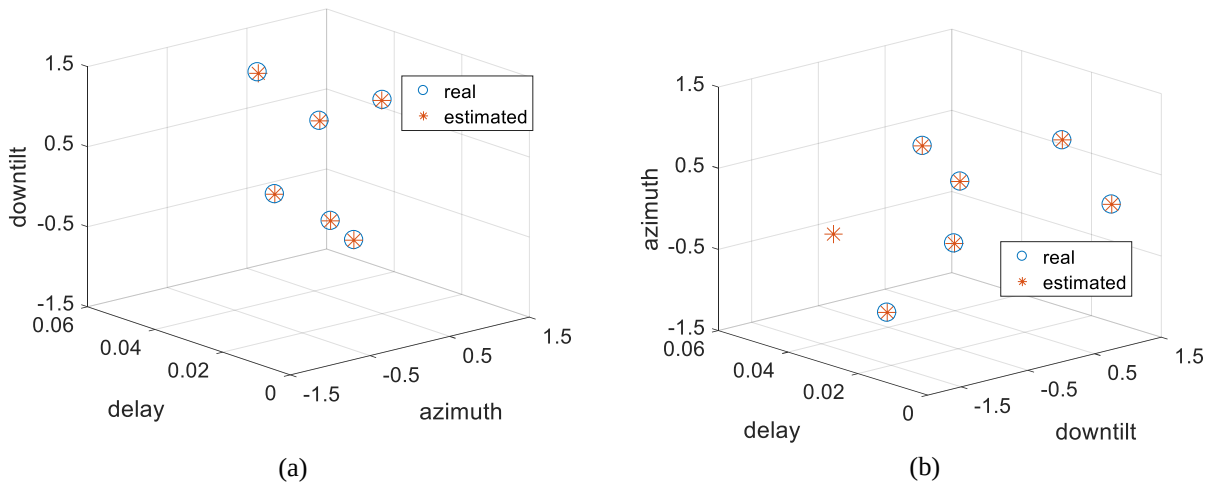

Fig. 4. Results of two implementations of the e-NOMP algorithm. The real frequency-independent parameters are illustrated by circles and their estimates are stars.

NOMP algorithm. Each user channel contains 5 propagation paths. The angles and the delays of the paths from different users are set independently. For each user, the delays are randomly distributed in $[0, 1/\triangle f)$. The downtilts and azimuths of each user are randomly distributed in $[-\pi/2, \pi/2)$. Power attenuation happens during the propagation of a wireless signal and the total attenuation for each user channel is randomly set within $[0, -10]$ dB. Both the uplink and the downlink transmit SNR are equal to 10 dB.

A. Evaluate the e-NOMP algorithm

We first evaluate the estimation precision of the e-NOMP algorithm by checking if the values of extracted frequency-independent parameters are equal to their real values. Fig. 4 examines two implementations of the e-NOMP algorithm when both the channel attenuation and the transmit power equals 0 dB and displays the results in a 3D coordinate system. The coordinate of each point in the 3D coordinate system is composed by the delay, azimuth and downtilt. The blue circles are the real frequency-independent parameters and the red stars are their estimates. We can observe from Fig. 4(a) that the estimates coincide with the real values and the number of extracted paths is exactly the real number of paths. It demonstrates that the e-NOMP algorithm can precisely detect each path from the mixture. While in Fig. 4(b), an extra and fake path is detected, reflecting that false alarm may happen during the implementation of e-NOMP. This is because that the estimated parameters are not accurate enough, then the component paths can

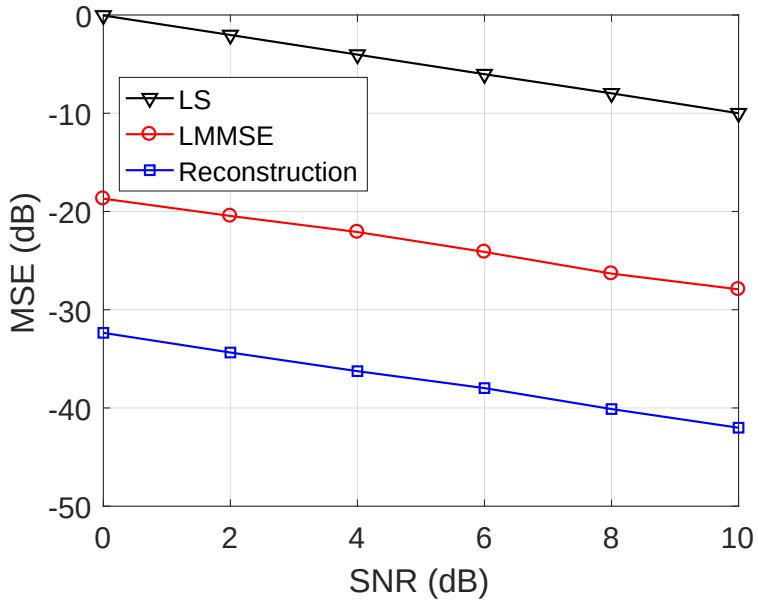

Fig. 5. MSE performance of the e-NOMP algorithm. The e-NOMP-reconstructed uplink channel is even more accurate than the LMMSE-estimated uplink channel.

not be completely eliminated from the mixture. The integration of the residual components will result in a new fake component path, which is falsely detected by the algorithm.

Even though the occurrence of false alarm is inevitable, the e-NOMP algorithm can still provide a globally precise reconstruction result. We examine the global precision of the e-NOMP-based uplink channel reconstruction by evaluating the MSE which is calculated as

$$\mathrm{MSE} = \frac{\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|^2\}}{\mathbb{E}\{\|\mathbf{z}\|^2\}}. \tag{51}$$

The classical LS and linear minimum mean square error (LMMSE) channel estimation methods are introduced as benchmarks. Fig. 5 compares the MSE performances of the estimated or reconstructed uplink channels. Here, the wireless channel attenuation is set to 0 dB, and SNR equals the transmit power. As expected, the LS estimated channel has the worst precision and LMMSE improves the performance by a large margin. While the e-NOMP-based uplink channel reconstruction further brings down the MSE significantly. When SNR equals 0 dB, e-NOMP is able to achieve almost 10^{-3} MSE, and the value continuously drops with the increase of SNR. These results strongly demonstrate the high global-precision of the e-NOMP algorithm.

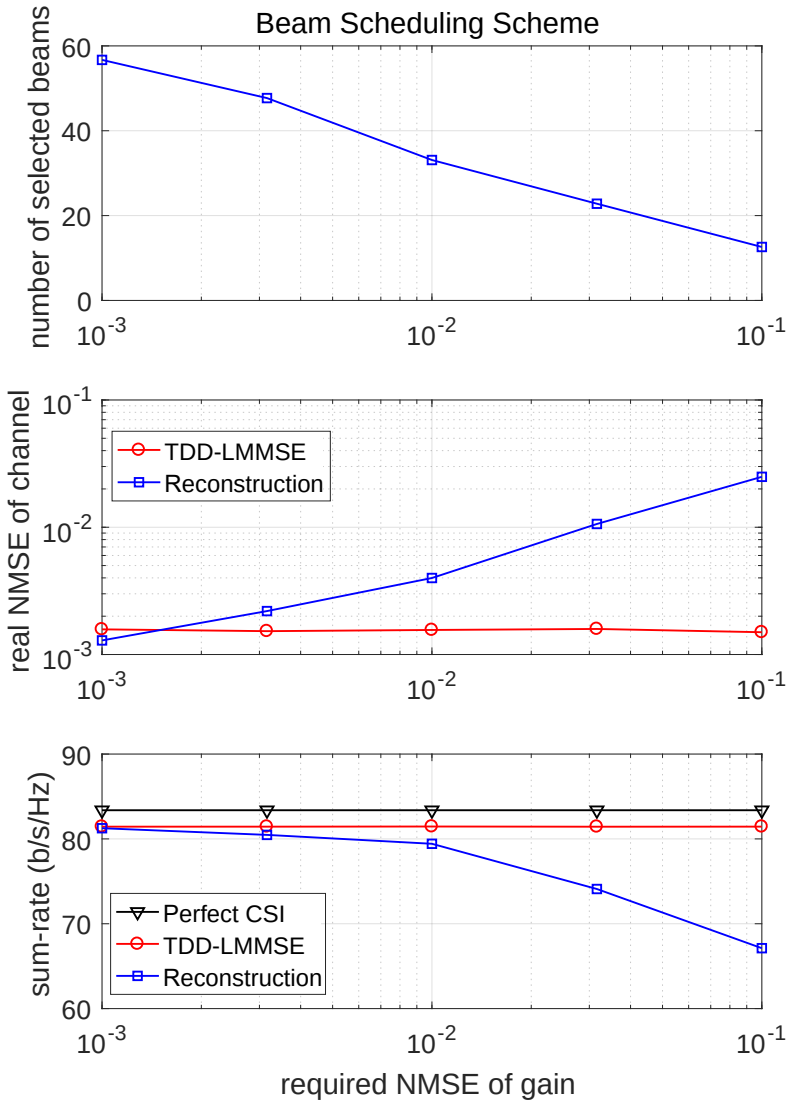

Fig. 6. Evaluation of the MSE and multiuser sum-rate performance of the proposed transceiver design by setting $K = 10$.

B. Evaluate the Reconstruction Based Transceiver

To compare the performance of the proposed downlink channel reconstruction based FDD massive MIMO transceiver design, we introduce LMMSE channel estimation as a benchmark again and assume that TDD mode is adopted and the downlink channel is derived from the LMMSE estimated uplink channel. We also evaluate the case when perfect downlink CSI is known at the BS.

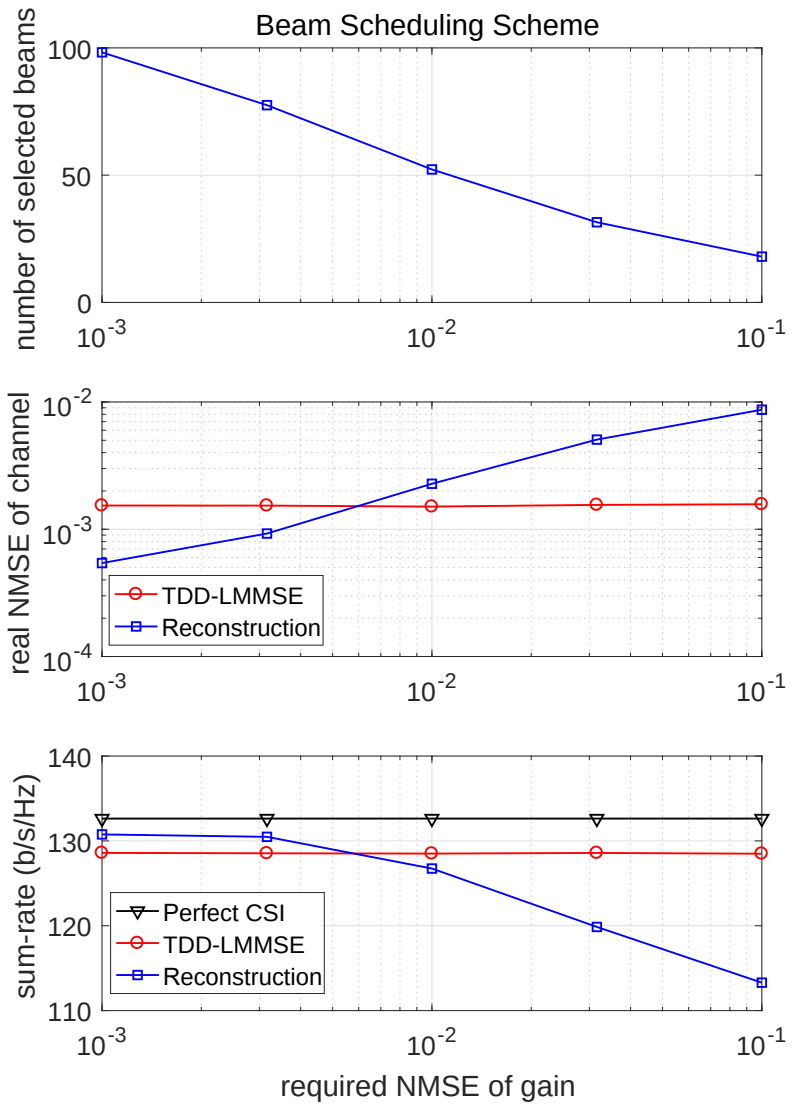

Fig. 7. Evaluation of the MSE and multiuser sum-rate performance of the proposed transceiver design by setting $K = 20$.

Fig. 6 evaluate the performance of the transceiver design when we have different requirements on the NMSE of the estimated downlink gains. We first test the case when there are only a small number of users in the cell by setting $K = 10$. If we set $\delta = 10^{-3}$, there must be enough pilots to promise the high estimation precision of the downlink gains. Thus, seldom grid points are excluded in the selected grid point set. In this condition, error rate of the reconstructed downlink channel is even lower than the LMMSE estimation results and the achieved multiuser sum-rate

becomes higher. With the increase of δ, the high precision requirement is relaxed gradually. The number of remained grid points decreases as well. The reduction of training pilots results in the inaccuracy of the estimates, which further leads to the decrease of multiuser sum-rate. Fortunately, the sum-rate gap between the case of $\delta = 10^{-3}$ and that of $\delta = 10^{-2}$ is small and acceptable, while the number of OFDM symbols for downlink training is reduced almostly from 56 to 32. However, if we further release the error rate requirement and set $\delta = 10^{-1}$, the multiuser sum-rate performance will be significantly damaged, validating the theoretical analysis in *Section V*.

Then, Fig. 7 displays the results when we increase the number of users in the cell by setting $K = 20$. Comparing with Fig. 6, we can find from the top sub-figure that if $\delta = 10^{-3}$ and seldom beams are removed, the number of selected beams will not double, even though the number of users doubles. This is because that with the increase of the user number, it has a higher probability that a grid point is marked simultaneously by multiple users. Sharing of a common training beam is always appreciated in downlink training process. By comparing the real NMSE performance, we can easily find that with more users, the number of selected beams increases, and the NMSE performance of the efficient reconstruction scheme improves significantly. Therefore, the proposed reconstruction based transceiver design performs better when increasing the users in the massive MIMO system.

VII. Conclusion

In this paper, we focused on the FDD massive MIMO system and proposed a reconstruction-based transceiver design by utilizing the spatial reciprocity between uplink and downlink. An e-NOMP algorithm was first introduced to extract the frequency-independent parameters in the uplink. Then, we formulated the requirement to achieve an acceptable NMSE of the estimated downlink gains and designed the downlink training scheme based on this requirement. A spatial angle grid was introduced and we proposed a beam scheduling scheme to further reduce the training overhead. The reconstructed multiuser channels enabled the downlink multiuser transmission using linear precoding methods, and theoretical analysis revealed the effect of the value of acceptable NMSE on the sum-rate performance. Numerical results proved that the frequency-parameters can be precisely estimated through the e-NOMP algorithm and near optimal sum-rate can be achieved by utilizing the downlink reconstruction results.

APPENDIX

According to (45), the SINR at user k satisfies

$$\mathbb{E}\{\mathrm{SINR}_k\} \approx \frac{\mathbb{E}\{P|\alpha_k|^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}|^2\}}{\sum_{j\neq k}\mathbb{E}\{P|\alpha_j|^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,j}|^2\}+\mathbb{E}\{|z_k|^2\}}. \tag{52}$$

Firstly, it is obvious that $\mathbb{E}\{|z_k|^2\}=1$. When it comes to the signal and the interference items, they can be approximated to, respectively,

$$\mathbb{E}\{P|\alpha_k|^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}|^2\} \approx \frac{P\mathbb{E}\{|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}|^2\}}{K\mathbb{E}\{\|\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}\|^2\}} \tag{53}$$

and

$$\mathbb{E}\{P|\alpha_j|^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,j}|^2\} \approx \frac{P\mathbb{E}\{|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,j}|^2\}}{K\mathbb{E}\{\|\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,j}\|^2\}}, \tag{54}$$

when (42) is applied. Since error exists in the reconstructed multiuser channel, we model the reconstructed multiuser channel by $\hat{\mathbf{H}}=\mathbf{H}+\mathbf{E}$, where $\mathbf{H}\in\mathbb{C}^{K\times M}$ is the real downlink channel matrix, $[\mathbf{H}]_{k,:}=\mathbf{h}_k^{\mathrm{dl}}$, $\mathbf{E}$ is the reconstruction error whose elements are i.i.d Gaussian with zero mean. According to (33), the NMSE of $\hat{\mathbf{g}}_k^{\mathrm{dl}}$ approximates δ, and the NMSE of $\hat{\mathbf{h}}_k^{\mathrm{dl}}$ approximates δ as well. Thus, we make the following approximation

$$\mathbb{E}\{|\left[\mathbf{E}\right]_{k,i}|^2\} \approx \delta|\left[\mathbf{H}\right]_{k,i}|^2. \tag{55}$$

Besides, $\hat{\mathbf{H}}^{\dagger}$ can be Taylor expanded by [23]

$$\hat{\mathbf{H}}^{\dagger}=(\mathbf{H}+\mathbf{E})^{\dagger}\approx\mathbf{H}^{\dagger}-\mathbf{H}^{\dagger}\mathbf{E}\mathbf{H}^{\dagger}, \tag{56}$$

then its kth column is expressed as

$$\left[\hat{\mathbf{H}}^{\dagger}\right]_{:,k}\approx\left[\mathbf{H}^{\dagger}\right]_{:,k}-\mathbf{H}^{\dagger}\mathbf{E}\left[\mathbf{H}^{\dagger}\right]_{:,k}. \tag{57}$$

Since

$$\mathbf{h}_k^{\mathrm{dl}}\left[\mathbf{H}^{\dagger}\right]_{:,j}=\begin{cases}1, & j=k,\\ 0, & j\neq k,\end{cases} \tag{58}$$

we can derive that

$$|\mathbf{h}_k^{\mathrm{dl}}\left[\hat{\mathbf{H}}^{\dagger}\right]_{:,j}|^2=\begin{cases}1-\left[\mathbf{H}^{\dagger}\right]_{:,k}^H\left[\mathbf{E}\right]_{k,:}^H-\left[\mathbf{E}\right]_{k,:}\left[\mathbf{H}^{\dagger}\right]_{:,k}+\left[\mathbf{H}^{\dagger}\right]_{:,k}^H\left[\mathbf{E}\right]_{k,:}^H\left[\mathbf{E}\right]_{k,:}\left[\mathbf{H}^{\dagger}\right]_{:,k}, & j=k\\ \left[\mathbf{H}^{\dagger}\right]_{:,j}^H\left[\mathbf{E}\right]_{k,:}^H\left[\mathbf{E}\right]_{k,:}\left[\mathbf{H}^{\dagger}\right]_{:,j}, & j\neq k\end{cases} \tag{59}$$

whose expectation approximates

$$\mathbb{E}\{|\mathbf{h}_k^{\mathrm{dl}}\left[\hat{\mathbf{H}}^{\dagger}\right]_{:,j}|^2\}\approx\begin{cases}1+\delta\sum_{m=1}^M|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2, & j=k\\ \delta\sum_{m=1}^M|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,j}|^2, & j\neq k\end{cases} \tag{60}$$

when applying (55) to (59). Besides,

$$\begin{aligned}\|\left[\hat{\mathbf{H}}^{\dagger}\right]_{:,k}\|^2\approx&\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2-\left[\mathbf{H}^{\dagger}\right]_{:,k}^H\mathbf{H}^{\dagger}\mathbf{E}\left[\mathbf{H}^{\dagger}\right]_{:,k}\\ &-\left[\mathbf{H}^{\dagger}\right]_{:,k}^H\mathbf{E}^H\mathbf{H}^{\dagger H}\left[\mathbf{H}^{\dagger}\right]_{:,k}+\left[\mathbf{H}^{\dagger}\right]_{:,k}^H\mathbf{E}^H\mathbf{H}^{\dagger H}\mathbf{H}^{\dagger}\mathbf{E}\left[\mathbf{H}^{\dagger}\right]_{:,k}\end{aligned} \tag{61}$$

and its expectation satisfies

$$\mathbb{E}\{\|\left[\hat{\mathbf{H}}^{\dagger}\right]_{:,k}\|^2\}\approx\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2+\delta\sum_{j=1}^K\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2\sum_{m=1}^M|\left[\mathbf{H}\right]_{j,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2. \tag{62}$$

By applying (60) and (62) into (53), we obtain

$$\begin{aligned}&\mathbb{E}\{P|\alpha_k|^2|\mathbf{h}_k^{\mathrm{dl}}(n)\left[\hat{\mathbf{H}}^{\dagger}(n)\right]_{:,k}|^2\}\\ \approx&\frac{P(1+\delta\sum_{m=1}^M|\left[\mathbf{H}\right]_{k,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}{K(\|\left[\mathbf{H}^{\dagger}\right]_{:,k}\|^2+\delta\sum_{j=1}^K\|\left[\mathbf{H}^{\dagger}\right]_{:,j}\|^2\sum_{m=1}^M|\left[\mathbf{H}\right]_{j,m}\left[\mathbf{H}^{\dagger}\right]_{m,k}|^2)}\end{aligned} \tag{63}$$

which is exactly S_k. Similarly, $\mathbb{E}\{I\}\approx I_k$. Therefore, (46) is obtained.

REFERENCES

[1] E. 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.

[2] A. Adhikary, J. Nam, J. 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.

[3] S. K. Mohammed, and E. G. Larsson, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 547-550, Oct. 2013.

[4] Y. Han, S. Jin, J. Zhang, J. Zhang, and K. K. Wong, "DFT-based hybrid beamforming multiuser systems: Rate analysis and beam selection" *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 514-528, Jun. 2018.

[5] H. Halbauer, S. Saur, J. Koppenborg, and C. Hoek, "3D Beamforming: Performance improvement for cellular networks," *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 37-56, 2013.

[6] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE Int. Conf. Commun.*, pp. 5865 - 5870, Jun. 2014.

[7] X. Li, S. Jin, X. Gao, and R. W. Heath, "Three-dimensional beamforming for large-scale FD-MIMO systems exploiting statistical channel state information," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 8992-9005, Nov. 2016.

[8] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170-2184, Jun. 2015.

[9] 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.

[10] 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.

[11] 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.

[12] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codes 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.

[13] W. Shen, L. Dai, B. Shim, Z. Wang, and R. W. Heath, "Channel feedback based on AoD-adaptive subspace codebook in FDD massive MIMO systems," *arXiv preprint arXiv:1704.00658*, 2017.

[14] Y. Han, H. Zhang, S. Jin, X. Li, R. Yu and Y. Zhang, "Investigation of transmission schemes for millimeter-wave massive MU-MIMO systems," *IEEE Syst. J.*, vol. 11, no. 1, pp. 72-83, Mar. 2017.

[15] U. Ugurlu, R. Wichman, C. B. Ribeiro and C. Wijting, "A multipath extraction-based CSI acquisition method for FDD cellular networks with massive antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2940-2953, Apr. 2016.

[16] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206-4218, Jun. 2018.

[17] S. Haghighatshoar, M. B. Khalilsarai, and G. Caire, "Multi-band covariance interpolation with applications in massive MIMO," *arXiv:1801.03714*, Jan. 2018.

[18] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *arXiv:1803.05754*, Mar. 2018.

[19] 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. Technol.*, vol. 66, no. 4, pp. 3170-3184, Apr. 2017.

[20] Y. Han, T.-H. Hsu, C.-K. Wen, K.-K. Wong, and S. Jin, "Efficient downlink channel reconstruction for FDD transmission systems," in *Proc. 27th WOCC*, Jun. 2018, pp. 1-5.

[21] Y. Han, T.-H. Hsu, C.-K. Wen, K.-K. Wong, and S. Jin, "Efficient downlink channel reconstruction for FDD multi-antenna systems," *arXiv:1805.07027*, May. 2018.

[22] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066-5081, Oct. 2016.

[23] C. Wang, E. K. S. Au, R. D. Murch, W.-H. Mow, R. S. Cheng, and V. K. N. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 805-810, Mar. 2007.