Spectral Efficient TSB Scheme with User Scheduling for FDD Massive MIMO Systems

Tianbao Gao, Chen Liu, *Member, IEEE,* Yunchao Song, *Member, IEEE,* Zhisheng Yin, *Member, IEEE,* Huibin Liang, Nan Cheng, *Member, IEEE*

***Abstract*—This paper proposes a two-stage beamforming (TSB) scheme with user scheduling for FDD massive MIMO. The developed TSB scheme designs the analog pre-beamformer and schedules the users using statistical channel state information (S-CSI), reducing the overhead of the pilot and the feedback. Particularly, in the one-ring local scattering channel model, the pre-beamformer design and user scheduling problem is formulated as a 0-1 quadratic constrained quadratic programming (QCQP), which is further linearized to a mixed integer linear programming (MILP). In the multiple scattering clusters channel model, we design the pre-beamformer and schedule the users based on graph theory, where the chromatic number of the equivalent matrix represents the minimum number of orthogonal pilots. Then, we propose an iterative beam selection and user scheduling (I-BSUS) scheme that approximates the minimum pilot constraint by the maximum vertex degree. Moreover, the net spectrum efficiency (NSE) is improved using a multi-user digital precoder, which depends on the effective instantaneous CSI (EI-CSI). Simulation results validate the superiority of the proposed scheme in enhancing the NSE over the existing schemes.**

***Index Terms*—Massive MIMO, FDD, statistical CSI, beam selection, user scheduling.**

I. Introduction

FOR the sixth generation (6G) and beyond communication networks, various technologies have emerged, such as virtual reality, augmented reality, reconfigurable intelligent surface (RIS), integrated sensing and communications (ISAC), space-air-ground integration network (SAGIN), network-assisted full-duplex (FD), and channel prediction [1]–[6]. These emerging technologies promise higher spectral and energy efficiency, improved reliability and low latency, and massive connectivity, revolutionizing how we communicate and interact with the world around us. These technologies are advancing wireless communication from massive multiple-input multiple-output (MIMO) setups that require many antennas to transmit and receive signals to RIS-based manipulation of electromagnetic waves. Although all 6G is closely related to massive MIMO technology, fulfilling the requirements requires the instantaneous channel state information (I-CSI) to be perfectly known at the transmitter. Obtaining downlink I-CSI involves downlink pilot training and uplink feedback, with the overhead for downlink I-CSI estimation being proportional to the number of antennas at the base station (BS). Nevertheless, this imposes a high overhead of pilots and feedback in massive MIMO systems due to the hundreds or thousands of antennas in the BS. Furthermore, the delayed or expired I-CSI may degrade the system performance, making it crucial to overcome the challenges in acquiring I-CSI to promote massive MIMO technology in next-generation communication systems [7], [8].

In a time division duplex (TDD) system [9]–[11], the uplink and downlink channels operate on the same frequency and have reciprocity. The downlink I-CSI is obtained from the estimated uplink I-CSI. The number of pilots required for uplink I-CSI estimation is equal to the number of user antennas, which is much smaller than the number of antennas in the BS. In frequency division duplex (FDD) systems, channel reciprocity is no longer applicable, which makes it challenging to estimate the downlink I-CSI. Despite its lack of channel reciprocity, the FDD system is still widely studied [12], [13] due to its appealing delay sensitivity and performance in symmetrical services. However, the overhead associated with pilots and feedback for downlink communication in massive MIMO FDD systems is excessive, reducing spectral efficiency. To address this issue, various approaches [14]–[16] have been employed, aiming to reduce the overhead associated with downlink I-CSI estimation, leading to improved spectral efficiency and system performance in FDD systems.

In order to exploit the sparsity of spatially correlated channels, compressed sensing (CS) [17], [18] and deep learning (DL) [19]–[21] aided channel estimation is a significant approach to dealing with the challenges of I-CSI estimation. In [17] the authors developed a CS-based channel estimation scheme that utilizes a conventional least squares approach and CS technique simultaneously, and [18] enhanced the performance of traditional CS by introducing a closed-loop compressive CSI estimation framework with 1-bit feedback. Deep learning-based schemes have received extensive attention due to their ability to solve nonlinear problems effectively and having low complexity. Typical examples are the novel off-grid model [19], deep learning CS channel estimation [20], and a two-tier based channel estimation [21]. The aforementioned methods for reconstructing the full-dimensional I-CSI present challenges, including computational complexity, pilot and feedback overhead, and the potential for channel estimation

This work was supported in part by the Natural Science Foundation of China under Grant 61771257, 62101282, and 62371249. *(Corresponding authors: Yunchao Song.)*

Tianbao Gao, Chen Liu, Yunchao Song, and Huibin Liang are with the College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China (e-mail: 2018020231, liuch, songyc, 2021020305@njupt.edu.cn).

Zhisheng Yin, and Nan Cheng are with State Key Lab. of ISN and School of Telecommunications Engineering, Xidian University, Xi'an, 710071, China e-mail: (zsyin@xidian.edu.cn, dr.nan.cheng@ieee.org).

errors. These challenges can have a significant impact on the performance of wireless communication systems.

To address these issues due to directly estimating the full-dimensional I-CSI, [22] proved that the angle reciprocity of FDD systems is useful to reduce the overhead of pilots and feedback. The angle reciprocity is applicable in a wide range of frequencies. The uplink statistical CSI (S-CSI) is obtained through the long-term estimation, and the downlink S-CSI, which is constant within the coherence time, can be easily extrapolated from the uplink S-CSI by using the angle reciprocity of the uplink and downlink channel. Therefore, exploiting the S-CSI, [22] proposed a two-stage beamforming scheme named joint spatial division and multiplexing (JSDM), which can achieve significant savings in both pilots and I-CSI feedback. Subsequent works aimed to enhance the performance of JSDM regarding user grouping [23], [24] and pre-beamformers [25], [26]. However, user grouping schemes cause signal space overlap between groups, resulting in a loss of signal space during pre-beamforming. Instead of a user grouping scheme, [27] introduced a neighbor-based JSDM (N-JSDM), which fully utilizes the signal space to enhance performance. Besides, when the number of antennas in the BS is large, the downlink covariance matrix of users can be approximated by the columns of the discreet Fourier transformation (DFT) matrix. This approximation enables the BS to design the pre-beamforming matrix through codebook-based beam selection, which can reduce the computational and implementation complexity [28].

The radio frequency (RF) chains required by the system in [22]–[27] equal the number of columns of the pre-beamformer. However, the number of RF chains in a system is often limited, and exceeding/falling short of the required number results in hardware overhead waste/heavy. Hence, to prevent significant performance loss, the number of RF chains must be equal to or greater than the number of users served simultaneously by the BS. When the number of RF chains is more than twice the number of service users, hybrid precoding can maintain the performance of full digital precoding while simultaneously limiting the number of users BS serves. Additionally, as the number of users in the system increases, the spectrum efficiency of the system declines due to severe interference between the users. Therefore, carefully selecting users with favorable channel conditions is vital in mitigating interference and improving system capacity. Given that in the extended research of JSDM [23], user scheduling is considered, user scheduling is indispensable in DFT codebook-based beam selection.

This paper proposes a two-stage (TSB) scheme that comprises analog pre-beamforming and multi-user digital precoding to reduce the overhead of pilots and feedback significantly. The pre-beamformer is designed through codebook-based beam selection, which is pilot-limited since the pilot length affects the system's net spectral efficiency (NSE). In addition to designing the pre-beamformer, we consider the user scheduling problem to ensure good system performance. Since the number of scheduling users is limited by the number of RF chains in the system, the pre-beamformer design involves beam selection and user scheduling (BSUS). Thus, this paper proposes BSUS schemes under a one-ring local scattering model and a multiple scattering cluster model to sparsify the channel and schedule users adaptively. The main contributions of this work are summarized as follows:

1) We analyze the sparsity characteristics of effective I-CSI (EI-CSI) using DFT codebook-based pre-beamforming design. A criterion is established that neglect channel degrees of freedom with low energy. Subsequently, we propose BSUS schemes for one-ring local scattering model and the multiple scattering clusters model, respectively.
2) In the one-ring local scattering model, the designed BSUS scheme aims to maximize the received energy while satisfying both the restricted RF chains and pilot constraints. The BSUS scheme is formulated as a 0-1 quadratic constrained quadratic programming (QCQP) problem, which is challenging to find the optimal solution. To address this issue, we linearize the 0-1 QCQP problem into a mixed-integer linear programming (MILP) problem with the same optimal value as the original 0-1 QCQP problem. MILP problems can use efficient MILP solvers to find the optimal solution.
3) We design the BSUS scheme based on graph theory for the multiple scattering clusters model. The minimum number of orthogonal pilots (NOP) equals the chromatic number of the graph generated by the equivalent I-CSI (EI-CSI). To approximate the chromatic number of the graph, we use the maximum vertex degree, which serves as an upper bound for the chromatic number. Finally, we propose an iterative BSUS (I-BSUS) scheme to find the minimum pilots, where the pilot constraint is represented by the maximum vertex degree.
4) We propose a circular pilot matrix and an iterative pilot allocation (IPA) scheme for the EI-CSI structure of the two channel models to assist the sparse EI-CSI estimation. The IPA scheme is not related to the previously mentioned I-BSUS scheme. Although the I-BSUS scheme can determine whether the pilot exceeds the limit, it does not provide the pilot matrix design. The IPA scheme based on the non-zero random distribution characteristics of the EI-CSI.

The remainder of the paper is organized as follows. Section (II) presents the system model. Section (III) describes the spatially correlated sparse channel based on an established codebook and details the beam selection and user scheduling process with a limited pilot in one-ring local scattering and multiple scattering clusters models. Section (IV) describes the estimation of the sparse EI-CSI by pilots and feedback and presents the performance of multi-user digital precoding, and Section (V) presents the simulation results. Finally, Section (VI) concludes this paper.

Notation: Lower-case boldface letter $\boldsymbol{a}$ is a column vector, and upper-case boldface letter $\boldsymbol{A}$ is a matrix. Non-bold letters a, A are scalar values. $\boldsymbol{a}_i$ is the i-th column of matrix $\boldsymbol{A}$, and a_i is the i-th value of vector $\boldsymbol{a}$. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{\dagger}$ denote transpose, conjugate transpose, and pseudo-inverse, respectively. $\|\boldsymbol{A}\|_{\mathrm{F}}$ and $|\boldsymbol{A}|$ are the Frobenius norm and determinant of $\boldsymbol{A}$. The complex number field is represented by $\mathbb{C}$, and

the expectation of a complex number is denoted by $\mathbb{E}\{\cdot\}$. $\boldsymbol{I}_N$ denotes a $N \times N$ identity matrix, and $\mathcal{CN}(\boldsymbol{a}, \boldsymbol{A})$ is a complex Gaussian vector with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{A}$. $card(A)$ represents the number of elements in set A. $col(\boldsymbol{A})$ represents the number of columns of $\boldsymbol{A}$.

II. System Model

Consider a single-cell FDD massive MIMO system, where a central BS serves K single-antenna users. The BS has a uniform linear array (ULA) with $M \gg K$ antennas and N_{RF} RF chains. The spacing between adjacent antennas of ULA is set to $\lambda/2$, where λ is the carrier wavelength. The downlink transmission is also considered.

Fig. 1 illustrates a two-stage beamforming structure comprising an analog pre-beamforming matrix $\boldsymbol{F}$ and a multi-user digital precoding matrix $\boldsymbol{W}$. The first stage (analog pre-beamformer), which only relies on S-CSI $\boldsymbol{R}$, is implemented only once in the entire coherence time. The analog pre-beamformer design aims to sparsify the channel matrix, thereby reducing the overhead of downlink orthogonal pilots and uplink I-CSI feedback. Typically, the analog pre-beamforming matrix $\boldsymbol{F}$ is implemented using the phase shifters which satisfy the magnitude constraint $\left|[\boldsymbol{F}]_{i,j}\right| = \frac{1}{\sqrt{M}}$ [29].

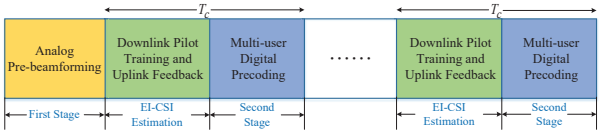


Fig. 1: Two-Stage Beamforming Architecture.

Taking advantage of the sparsity of the spatially correlated channel, we design an analog pre-beamforming matrix $\boldsymbol{F}$ to sparsify the EI-CSI $\bar{\boldsymbol{H}}$ as

$$\bar{\boldsymbol{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}, \tag{1}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]$ is the channel between users and the BS, and $\boldsymbol{h}_k$ is the spatial channel between user k and the BS. The downlink pilots are used to estimate the sparse EI-CSI, and only the non-zero values of the EI-CSI must be fed back to the BS. The number of orthogonal pilots in the EI-CSI estimation is far less than the number of antennas M.

Based on the estimated EI-CSI, the second stage (multi-user digital precoding) is designed for data symbol transmission. We use linear precoding techniques to design the digital precoder $\boldsymbol{W}$. The received signal of user k is written as

$$y_k = \bar{\boldsymbol{h}}_k \boldsymbol{W}\boldsymbol{s} + n_k, \tag{2}$$

where $\bar{\boldsymbol{h}}_k$ is the sparse channel vector between user k and the BS. The overall EI-CSI is formulated as $\bar{\boldsymbol{H}} = \left[\bar{\boldsymbol{h}}_1^T, \cdots, \bar{\boldsymbol{h}}_K^T\right]^T$. For a total transmitted power P at BS, the overall transmitted signal $\boldsymbol{s}$ satisfies $\mathbb{E}\left\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\right\} = \frac{P}{K}\boldsymbol{I}_K$, $n_k \sim \mathcal{CN}\left(0, \sigma^2\right)$ is additive white Gaussian noise.

Assuming no line of sight channel, we have $\boldsymbol{h}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_k)$, where $\boldsymbol{R}_k = \mathbb{E}\left[\boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}}\right]$ is the channel covariance matrix of user k. Using the Karhunen-Loeve representation, the channel vector $\boldsymbol{h}_k$ can be written as

$$\boldsymbol{h}_k = \boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{w}_k, \tag{3}$$

where the entries of $\boldsymbol{w}_k$ are i.i.d. complex Gaussian variables with zero-mean and unit-variance.

Fig. 2(a) depicts the widely used one-ring local scattering model (macro cell or suburban statistical channel features model). The paths of user k are distributed in a scattering ring of radius r, where θ_k and Δ_k are the center angle and angular spread (AS), respectively. The channel covariance matrix $\boldsymbol{R}_{si,k}$ is expressed as

$$\boldsymbol{R}_{si,k} = \int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma(d\theta)\, \boldsymbol{a}(\theta)\, \boldsymbol{a}(\theta)^{\mathrm{H}}, \tag{4}$$

where $\gamma(d\theta)$ is the channel power azimuth spectrum (PAS) and $\int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma(d\theta) = 1$, $\boldsymbol{a}(\theta) = \left[1, \cdots, e^{\jmath\pi(M-1)\sin(\theta)}\right]^T$.

Fig. 2(b) illustrates the multiple scattering clusters model for millimeter-wave channels. There are C_k scattering clusters between user k and the BS, and the channel covariance matrix $\boldsymbol{R}_{mu,k}$ is expressed as

$$\boldsymbol{R}_{mu,k} = \frac{1}{C_k}\sum_{c=1}^{C_k} \boldsymbol{R}_c, \tag{5}$$

where $\boldsymbol{R}_c$ is similar to Eq. (4).

III. Spatially Correlated Channel Sparsification by Beam Selection and User Scheduling

This section presents the BSUS schemes with a restricted number of pilots N for the one-ring local scattering and multiple scattering clusters models. Specifically, the BSUS schemes account for the problem of user scheduling when the total number of users in the cell significantly exceeds the number of users the system can serve simultaneously.

By exploiting the sparsity of the spatially correlated channel matrix $\boldsymbol{H}$, an analog pre-beamforming matrix $\boldsymbol{F}$ is designed to sparsify the EI-CSI $\bar{\boldsymbol{H}}$, which can effectively reduce the overhead of pilots and feedback in downlink channel estimation. To make the correlated channel sparse, we have

$$\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{f}_l = 0, \tag{6}$$

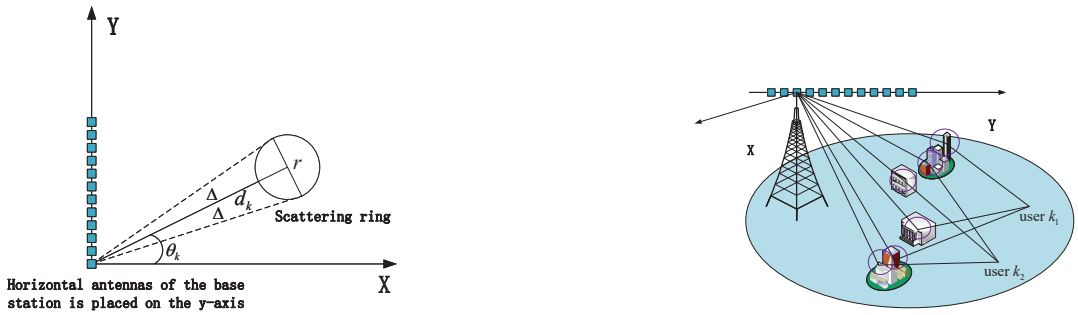
where $\boldsymbol{f}_l$ is the l-th column of the analog pre-beamforming matrix $\boldsymbol{F}$. By substituting $\boldsymbol{h}_k = \boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{w}_k$ into Eq. (6) and noticing that the random vector $\boldsymbol{w}_k$ is unknown to the transmitter, $\boldsymbol{f}_l$ should satisfy

$$\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{f}_l = \boldsymbol{0}. \tag{7}$$

Thus, $\boldsymbol{F}$ only depends on the S-CSI $\boldsymbol{R}$, which is constant in a coherence time and assumed to be known at the BS [22], [27], [28]. To simplify the pre-beamformer design, $\boldsymbol{F}$ is designed using the DFT codebook, i.e., $\boldsymbol{F}$ is composed of the columns of the DFT matrix. The DFT orthogonal codebook is given by

$$\mathcal{S} = \left\{\boldsymbol{e}_m : \boldsymbol{e}\left(\frac{2m}{M} - 1\right), m \in \mathcal{M}\right\}, \tag{8}$$

where $\mathcal{M} = \{1, 2, \cdots, M\}$ denotes the beam indices, and $\boldsymbol{e}(x) = \frac{1}{\sqrt{M}}\left[1, \cdots, e^{-\jmath\pi(M-1)x}\right]^T$, where $x = \sin(\theta)$. $\mathcal{S}$ forms an orthogonal basis for the received signal space. We set $\mathcal{S}$ as the DFT codebook and select the beam vectors to form the pre-beamforming matrix $\boldsymbol{F}$.



(a) (b)

Fig. 2: (a) One-ring Local Scattering Model, (b) Multiple Scattering Clusters Model.

To obtain the optimal EI-CSI $\bar{\boldsymbol{H}}$, we analyze its sparse structure characteristics before designing $\boldsymbol{F}$. In order to satisfy Eq. (7), we evaluate the value of $\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{e}_m$. Moreover, to observe the sparse characteristics more clearly, we examine the properties of function $f = \left|\boldsymbol{e}\left(x_1\right)^{\mathrm{H}}\boldsymbol{e}\left(x_2\right)\right|$, where $\boldsymbol{e}\left(x_1\right)^{\mathrm{H}} = \boldsymbol{a}\left(\theta\right)^{\mathrm{H}}$ represents a part of the S-CSI $\boldsymbol{R}$. The sparse structure characteristics of EI-CSI follow the sparse structure characteristics of function f. Note that for any x_1 and x_2, f is a periodic function with a period of $\frac{2}{M}$, and its peak value gradually decreases with the increase of $|x_1 - x_2|$. As shown in Fig. 3, f has the maximum value when $|x_1 - x_2| = 0$ and equals zero when $|x_1 - x_2| = \frac{2}{M} \cdot g$, where g is any integer. Additionally, f is much smaller than the maximum peak 1 with the increase of $|x_1 - x_2|$. Indeed, f is considered to be 0 when $f \leq \epsilon$, where ϵ (less than noise power) is a small value.

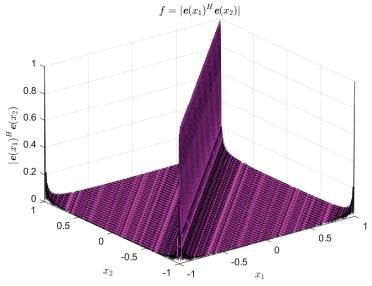


Fig. 3: Properties of function $f = \left|\boldsymbol{e}\left(x_1\right)^{\mathrm{H}}\boldsymbol{e}\left(x_2\right)\right|$, $M = 128$.

The periodicity of function f follows from the periodicity of DFT codeword $\boldsymbol{e}\left(x\right)$ and the value of function f depends only on the difference of $|x_1 - x_2|$. Signals that arrive along paths with angular separation larger than $2/M$ are considered approximately orthogonal. Fig. 3 reveals that when the difference of $|x_1 - x_2|$ is large, the value of f decreases rapidly and tends to zero. By designing pre-beamforming based on a DFT codebook, the EI-CSI $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$ is sparse.

To sparsify the channel $\boldsymbol{H}$ shown in Eq. (7), the elements of $\boldsymbol{H}$ satisfying

$$\left\|\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{e}_m\right\|^2 \leq \epsilon, \tag{9}$$

are assumed to be zeros. This accounts for the negative impact of the DFT matrix, which neglects certain channel degrees of freedom that contribute a relatively low proportion of the total channel energy. This effectively reduces the overhead of downlink channel estimation and uplink feedback, significantly increasing the received energy.

A. Problem Formulation

Similar to [28], we use the received energy maximization as the design criterion to improve spectral efficiency. Therefore, the pre-beamformer design must meet two constraints: the number of beams $col\left(\boldsymbol{F}\right)$ should be less than the number of RF chains N_{RF}, and the NOP μ must be less than the restricted number of pilots N. Given the DFT orthogonal codebook $\mathcal{S}$, we formulate the pre-beamformer design problem as the following optimized model [30]:

$$\begin{aligned}
(\mathcal{P}_1) \quad & \max_{\boldsymbol{F}} \mathbb{E}\left\{\left\|\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right\|_{\mathrm{F}}^2\right\} && \text{(10a)}\\
& s.t.\ col\left(\boldsymbol{F}\right) \leq N_{RF} && \text{(10b)}\\
& \mu \leq N && \text{(10c)}\\
& \boldsymbol{f}_l \in \mathcal{S}. && \text{(10d)}
\end{aligned}$$

The I-CSI $\boldsymbol{H}$ is unknown at the BS in the analog pre-beamforming stage prior to the I-CSI estimation. In order to address this issue, $\boldsymbol{H}$ is replaced with the S-CSI $\boldsymbol{R}$ using the following transformation. For the convenience of symbol representation, we assume that the pre-beamforming matrix $\boldsymbol{F}$ has L columns and then

$$\begin{aligned}
\mathbb{E}\left\{\left\|\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right\|_{\mathrm{F}}^2\right\} &= \mathbb{E}\left\{tr\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right)^{\mathrm{H}}\right)\right\}\\
&= \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\left\{\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{h}_k\right\}\\
&= \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\left\{\boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{w}_k\right\}\\
&= \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\left\{tr\left(\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{H}}\right)\right\}\\
&= \sum_{k=1}^{K}\sum_{l=1}^{L}tr\left(\mathbb{E}\left\{\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{H}}\right\}\right)\\
&\overset{(a)}{=} \sum_{k=1}^{K}\sum_{l=1}^{L}tr\left(\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{R}_k\right),
\end{aligned} \tag{11}$$

where (a) can be obtained from $\mathbb{E}\{\boldsymbol{AB}\} = \mathbb{E}\{\boldsymbol{A}\}\mathbb{E}\{\boldsymbol{B}\}$ for independent $\boldsymbol{A}$ and $\boldsymbol{B}$. Therefore, the design of the pre-beamforming matrix only relies on S-CSI $\boldsymbol{R}$.

B. One-ring Local Scattering Model

Based on the beam $\boldsymbol{e}_m$ definition presented in the previous section, we can infer the sparse structure characteristics of EI-CSI in advance. Specifically, EI-CSI is sparse, with any element whose value is less than ϵ set to zero. This accounts for the negative impact of the DFT codebook, which ignores channel degrees of freedom with very low channel energy. Consequently, there is no need to use pilots to estimate these zero elements, reducing the overhead of downlink pilots and uplink feedback.

Let matrix $\boldsymbol{C}$ denote the weights between beams in $\mathcal{S}$ and the S-CSI of user k, i.e. $tr\left(\boldsymbol{e}_m\boldsymbol{e}_m^{\mathrm{H}}\boldsymbol{R}_{si,k}\right)$. The element $c_{k,m}$ is expressed as

$$c_{k,m} = \begin{cases} \left\|\boldsymbol{R}_{si,k}^{\frac{1}{2}}\boldsymbol{e}_m\right\|^2, & \left\|\boldsymbol{R}_{si,k}^{\frac{1}{2}}\boldsymbol{e}_m\right\|^2 > \epsilon; \\ 0, & \left\|\boldsymbol{R}_{si,k}^{\frac{1}{2}}\boldsymbol{e}_m\right\|^2 \le \epsilon. \end{cases} \tag{12}$$

For the BSUS scheme, we employ two variable vectors $\boldsymbol{b} = [b_1, \cdots, b_M]^{\mathrm{T}}$ and $\boldsymbol{u} = [u_1, \cdots, u_K]^{\mathrm{T}}$ to describe the state of beam selection and user scheduling, respectively. The definition of b_m and u_k are given as

$$b_m = \begin{cases} 1, & \text{beam } m \text{ is selected;} \\ 0, & \text{beam } m \text{ is not selected,} \end{cases} \tag{13}$$

$$u_k = \begin{cases} 1, & \text{user } k \text{ is scheduled;} \\ 0, & \text{user } k \text{ is not scheduled,} \end{cases} \tag{14}$$

where $m \in \mathcal{M}$ and $k \in \mathcal{K}$. Finally, the beams b_m that $b_m = 1$ are constructed as the pre-beamforming matrix, and the users u_k that have $u_k = 1$ are scheduled.

The received energy in Eq. (11) is represented by the beam selection state vector $\boldsymbol{b}$, the user scheduling state vector $\boldsymbol{u}$, and the weight matrix $\boldsymbol{C}$, which are expressed as

$$\sum_{k=1}^{K}\sum_{l=1}^{L} tr\left(\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{H}}\boldsymbol{R}_k\right) = \sum_{k=1}^{K}\sum_{m=1}^{M} u_k b_m c_{k,m} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{b}. \tag{15}$$

The first constraint in Eq. (10b) means that the number of selected beams $col\,(\boldsymbol{F})$ is limited by the number of RF chains. The selected beams are represented by the beam selection state vector $\boldsymbol{b}$ with $b_m = 1$, and Eq. (10b) can be formulated as

$$\sum_{m=1}^{M} b_m \le N_{RF}. \tag{16}$$

The second constraint in Eq. (10c) shows that the number of pilots N limits the NOP μ. In the BSUS scheme, the angles of users are first arranged in ascending order, i.e., $\theta_1 \le \theta_2 \le \cdots \le \theta_K$. Then, the EI-CSI $\bar{\boldsymbol{H}}$ is a banded matrix in which all non-zero elements fall in a banded region centered on the main diagonal. The NOP μ equals the bandwidth of $\bar{\boldsymbol{H}}$, i.e., the maximum number of non-zero values in each row. To express the limited NOP μ presented in Eq. (10c), we define a Boolean matrix $\bar{\boldsymbol{C}}$ and the (k,m)-th element of the matrix $\bar{\boldsymbol{C}}$ is given as

$$\bar{c}_{k,m} = \begin{cases} 1, & c_{k,m} \neq 0; \\ 0, & \text{else.} \end{cases} \tag{17}$$

The number of non-zero values in the k-th row is $\sum_{m=1}^{M} u_k b_m \bar{c}_{k,m}$, and the NOP μ can be formulated as

$$\mu = \max_{k\in\mathcal{K}} \sum_{m=1}^{M} u_k b_m \bar{c}_{k,m}. \tag{18}$$

The constraint in Eq. (10c) $\mu \le N$ can be rewritten as

$$\sum_{m=1}^{M} u_k b_m \bar{c}_{k,m} \le N, \forall k \in \mathcal{K}. \tag{19}$$

Based on the above description, the optimization problem in Eq. (10a)-(10d) can be reformulated into a 0-1 integer programming setup as presented below:

$$(\mathcal{P}_2) \quad \max_{\boldsymbol{u},\boldsymbol{b}} \sum_{k=1}^{K}\sum_{m=1}^{M} u_k b_m c_{k,m} \tag{20a}$$

$$s.t. \sum_{m=1}^{M} b_m \le N_{RF} \tag{20b}$$

$$\sum_{m=1}^{M} u_k b_m \bar{c}_{k,m} \le N, \forall k \in \mathcal{K} \tag{20c}$$

$$b_m = \{0,1\}, \forall m \in \mathcal{M} \tag{20d}$$

$$u_k = \{0,1\}, \forall k \in \mathcal{K}. \tag{20e}$$

The $\mathcal{P}_2$ is a 0-1 QCQP problem. Let $\boldsymbol{x} = \left[\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}\right]^{\mathrm{T}}$, we rewrite the QCQP problem in the general form:

$$(\mathcal{P}_3) \quad \min_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} \tag{21a}$$

$$s.t. \ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \le N_{RF} \tag{21b}$$

$$\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{Q}}_k\boldsymbol{x} \le N, \forall k \in \mathcal{K} \tag{21c}$$

$$x_i = \{0,1\}, \forall i = 1, \cdots, (K+M), \tag{21d}$$

where

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{0}_{K\times K} & -\boldsymbol{C} \\ \boldsymbol{0}_{M\times K} & \boldsymbol{0}_{M\times M} \end{bmatrix}, \quad \bar{\boldsymbol{Q}}_k = \begin{bmatrix} \boldsymbol{0}_{K\times K} & \bar{\boldsymbol{C}}_k \\ \boldsymbol{0}_{M\times K} & \boldsymbol{0}_{M\times M} \end{bmatrix}. \tag{22}$$

$\boldsymbol{0}_{K\times K}$, $\boldsymbol{0}_{M\times K}$, $\boldsymbol{0}_{M\times M}$ are $K\times K$, $M\times K$, $M\times M$ dimensional zero matrices, respectively. $\bar{\boldsymbol{C}}_k$ is constructed according to $\bar{\boldsymbol{C}}$, and its k-th row is the same as $\bar{\boldsymbol{C}}$ while the other elements are zero. $\boldsymbol{a} = [\underbrace{0, \cdots, 0}_{K}, \underbrace{1, \cdots, 1}_{M}]^{\mathrm{T}}$.

Finding the optimal solution for the 0-1 QCQP problem is challenging. Therefore, we linearize the 0-1 QCQP problem in Eq. (21a)-(21d) as a MILP problem, which is easier to solve.

$$
\begin{aligned}
(\mathcal{P}_4) \quad & \min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{s},\boldsymbol{t}_k} \boldsymbol{r}^{\mathrm{T}}\boldsymbol{s} - A\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x} && \text{(23a)}\\
& s.t.\ \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{s} + A\boldsymbol{r} = \boldsymbol{0} && \text{(23b)}\\
& \quad \boldsymbol{z} \le 2A\left(\boldsymbol{r} - \boldsymbol{x}\right) && \text{(23c)}\\
& \quad \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \le N_{RF} && \text{(23d)}\\
& \quad \bar{\boldsymbol{Q}}_k\boldsymbol{x} - \boldsymbol{t}_k \le \boldsymbol{0}, \forall k \in \mathcal{K} && \text{(23e)}\\
& \quad \boldsymbol{r}^{\mathrm{T}}\boldsymbol{t}_k \le N, \forall k \in \mathcal{K} && \text{(23f)}\\
& \quad x_i = \{0,1\}, \forall i = 1,2,\cdots,(K+M) && \text{(23g)}\\
& \quad \boldsymbol{z} \ge \boldsymbol{0}, \boldsymbol{s} \ge \boldsymbol{0}, && \text{(23h)}
\end{aligned}
$$

where $A = \|\boldsymbol{Q}\|_\infty$, $\bar{A}_k = \|\bar{\boldsymbol{Q}}_k\|_\infty$ and $\boldsymbol{r}$ is an all-one vector.

Theorem 1. *$\boldsymbol{x}^*$ is the optimal solution to problem Eq.* (21a)-(21d) *if and only if there exist $\boldsymbol{z}^*$, $\boldsymbol{s}^*$, $\boldsymbol{t}_k^*$ such that $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is the optimal solution to Eq.* (23a)-(23h).

Proof. See Appendix A. □

The MILP problem presented in Eq. (23a)-(23h) can be solved using the existing optimization toolboxes. The optimal user scheduling $\boldsymbol{u}$ and beam selection $\boldsymbol{b}$ can be derived from $\boldsymbol{x}$.

C. Multiple Scattering Clusters Model

In the BSUS scheme of the multiple scattering clusters model, we also aim to maximize the received energy to improve the spectral efficiency while ensuring that the two constraints outlined in (10b) and (10c) are satisfied. Unlike the quadratic pilot constraint in the one-ring local scattering model, the NOP μ in the multiple scattering clusters model takes on a complex and nonlinear form due to the new structure of the EI-CSI, where the non-zero elements are randomly distributed. In this section, the BSUS scheme is designed using graph theory, where the chromatic number of EI-CSI represents the minimum NOP. Then, we propose an I-BSUS scheme that approximates the pilot constraint with the maximum vertex degree to find the minimum NOP.

The weight matrix $\boldsymbol{A}$, which represents the received energy of the S-CSI of user k to the beams in $\mathcal{S}$, is initially provided, and its element $a_{k,m}$ is expressed as

$$
a_{k,m} = \begin{cases} \left\| \boldsymbol{R}_{mu,k}^{\frac{1}{2}} \boldsymbol{e}_m \right\|^2, & \left\| \boldsymbol{R}_{mu,k}^{\frac{1}{2}} \boldsymbol{e}_m \right\|^2 > \epsilon; \\ 0, & \left\| \boldsymbol{R}_{mu,k}^{\frac{1}{2}} \boldsymbol{e}_m \right\|^2 \le \epsilon. \end{cases} \tag{24}
$$

Whether the beams and users are selected is described using the beam selection state variable b_m and the user scheduling state variable u_k, as defined in (13) and (14), respectively. The received energy is expressed as

$$
\max_{\boldsymbol{u},\boldsymbol{b}} \sum_{k=1}^{K} \sum_{m=1}^{M} u_k b_m a_{k,m}. \tag{25}
$$

Additionally, the BSUS scheme must satisfy two constraints: (1) the number of beams $col\left(\boldsymbol{F}\right)$ must be less than the number of RF chains N_{RF}, and (2) the NOP μ must be less than the minimum pilot constraint N. The first constraint can be expressed as

$$
\sum_{m=1}^{M} b_m \le N_{RF}, \tag{26}
$$

which is the same as in the one-ring local scattering model. In the multiple scattering clusters model, due to the random arrangement of non-zero elements in the EI-CSI $\bar{\boldsymbol{H}}$, it is not possible to directly calculate NOP using $\bar{\boldsymbol{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$ as described in (20c). For the second constraint, we use the chromatic number theory of the graph to calculate the minimum NOP in the multiple scattering clusters model, as described below.

Theorem 2. *The minimum NOP can be represented by the chromatic number of the corresponding graph of the equivalent matrix $\bar{\boldsymbol{H}} = \boldsymbol{H}^{H}\boldsymbol{F}$.*

Proof. $\mathcal{G}_1$: The minimum NOP is equivalent to the number of orthogonal columns in $\boldsymbol{X}$ that satisfy

$$
\boldsymbol{X}_{nz,k}^{\mathrm{H}} \boldsymbol{X}_{nz,k} = \boldsymbol{I}, k \in \mathcal{K}, \tag{27}
$$

where $\boldsymbol{X}_{nz,k}$ is selected by $\boldsymbol{X}$ and contains the index of the non-zero value of the k-th row of $\bar{\boldsymbol{H}}$.

$\mathcal{G}_2$: The chromatic number of the corresponding graph of $\bar{\boldsymbol{H}}$ equals the number of columns of $\boldsymbol{X}$ such that any two columns of $\boldsymbol{X}_{nz,k}$ are different.

Let $\boldsymbol{X}_1$ be the solution of $\mathcal{G}_1$, then any two columns in $\boldsymbol{X}_{1,nz,k}$ are different, so $\boldsymbol{X}_1$ is also the solution to $\mathcal{G}_2$. The minimum number of distinct columns in $\boldsymbol{X}_1$ is not less than the number of distinct columns of $\boldsymbol{X}$ (denoted as μ) satisfying $\mathcal{G}_2$.

Furthermore, we can find a solution $\boldsymbol{X}_2$ that satisfies $\mathcal{G}_2$, and $\boldsymbol{X}_2$ is expressed as

$$
\boldsymbol{X}_2 = \left[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_\mu\right], \tag{28}
$$

where $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_\mu$ are the μ columns of the unitary matrix $\boldsymbol{V}$. Any two columns of $\boldsymbol{X}_{2,nz,k}$ are orthogonal, so $\boldsymbol{X}_2$ is also the solution of $\mathcal{G}_1$, and its minimum number of columns is μ.

The minimum NOP is equal to the chromatic number of the graph corresponding to $\bar{\boldsymbol{H}}$. □

The steps to determine the minimum NOP from the equivalent matrix $\bar{\boldsymbol{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$ are as follows:

Step 1: Let $\bar{\boldsymbol{A}}$ denote the sparse feature of the weight matrix $\boldsymbol{A}$, and the element of $\bar{\boldsymbol{A}}$ is defined as

$$
\left[\bar{A}\right]_{k,m} = \begin{cases} 1, & a_{k,m} \ne 0; \\ 0, & a_{k,m} = 0. \end{cases} \tag{29}
$$

The sparse feature matrix $\boldsymbol{G}$ with BSUS is given by

$$
\boldsymbol{G} = \bar{\boldsymbol{A}}\left(\mathcal{U}, \mathcal{B}\right). \tag{30}
$$

where $\mathcal{U} = \{k | u_k = 1, k \in \mathcal{K}\}$ denotes the indices of scheduling users, and $\mathcal{B} = \{m | b_m = 1, m \in \mathcal{M}\}$ denotes the indices of selected beams.

Step 2: The adjacency matrix $\bar{\boldsymbol{G}}$ can be established as

$$
\left[\bar{\boldsymbol{G}}\right]_{i,j} = \begin{cases} 0, & i = j \text{ or } \left[\hat{\boldsymbol{G}}\right]_{i,j} = 0; \\ 1, & \text{else}, \end{cases} \tag{31}
$$

where $\hat{\boldsymbol{G}} = \boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}$ is established according to $\boldsymbol{G}$.

Step 3: The chromatic number of the graph generated by the equivalent matrix $\bar{\boldsymbol{H}}$ equals the adjacent matrix $\bar{\boldsymbol{G}}$, and solving the chromatic number of the graph is considered to be an NP-hard problem. To determine the chromatic number of the graph corresponding to the equivalent matrix $\bar{\boldsymbol{H}}$, we use the relationship between maximum vertex degree and chromatic number

$$\mathcal{X}\left(\bar{\boldsymbol{G}}\right) \leq \Delta\left(\bar{\boldsymbol{G}}\right) + 1, \tag{32}$$

where $\mathcal{X}\left(\bar{\boldsymbol{G}}\right)$ is the chromatic number of $\bar{\boldsymbol{G}}$, and $\Delta\left(\bar{\boldsymbol{G}}\right)$ denotes the maximum vertex degree of $\bar{\boldsymbol{G}}$. The minimum NOP is set as the upper bound of the chromatic number.

The second constraint (Eq. (20c)) can be expressed as

$$\Delta\left(\bar{\boldsymbol{G}}\right) + 1 \leq N. \tag{33}$$

Although we use the maximum vertex degree to represent the pilot constraint, representing the maximum vertex degree of $\bar{\boldsymbol{G}}$ is complex and cannot be expressed as a simple linear or quadratic inequality constraint. This complexity makes the optimization problem difficult to solve. To address this issue, we propose an I-BSUS scheme to find the minimum pilot. In the proposed I-BSUS scheme, we iteratively schedule users and select beams without increasing the orthogonal pilots. Specifically, let $\mathcal{A} = \mathcal{M} - \mathcal{B}$ denote the remaining unselected beams, where $\mathcal{B}$ contains the first $l-1$ selected beams. Let $\mathcal{C} = \mathcal{K} - \mathcal{U}$ denote the remaining unscheduling users, where $\mathcal{U}$ contains the first $l-1$ scheduling users.

In the l-th iteration, the adjacency matrix $\bar{\boldsymbol{G}}_l$ with the beams in $\mathcal{A}$ and the users in $\mathcal{C}$ are calculated. Beams, users, and beam-user pairs with the same maximum vertex degree as $\bar{\boldsymbol{G}}_{l-1}$ are denoted by $\mathcal{D}$. The index of optimal user and beam is selected as

$$\{k^*, m^*\} = \arg \max_{\{k',m'\}\in\mathcal{D}} \sum_{m=1}^{M} b_m a_{k',m} + \sum_{k=1}^{K} u_k a_{k,m'} - a_{k',m'}. \tag{34}$$

If $\mathcal{D}$ is an empty set, i.e., no beam, the user, and beam-user pair have a non-adjacent relationship with the beams in $\mathcal{B}$ or the users in $\mathcal{U}$. The optimal beam and user will be selected as

$$\{k^*, m^*\} = \arg \max_{k'\in\mathcal{A},m'\in\mathcal{C}} \sum_{m=1}^{M} b_m a_{k',m} + \sum_{k=1}^{K} u_k a_{k,m'} - a_{k',m'}. \tag{35}$$

The details of the I-BSUS scheme are shown in Algorithm 1.

Algorithm 1: The I-BSUS scheme

Input: $A;\bar{A};\mathcal{M};\mathcal{K}$.
Output: $\mathcal{B};\mathcal{U}$.
Initialization $\mathcal{B} = \varnothing$ and $\mathcal{A} = \mathcal{M} - \mathcal{B}$.
Initialization $\mathcal{U} = \varnothing$ and $\mathcal{C} = \mathcal{K} - \mathcal{U}$.
while $l \leq N_{RF}$ *and* NOP $< N$ **do**
  Repeat
  Establish the new $\mathcal{B} = \mathcal{B} + m, m \in \mathcal{A}$ and $\mathcal{U} = \mathcal{U} + k, k \in \mathcal{C}$.
  Establish the sparse feature matrix $\boldsymbol{G}$ as described in Eq. (30).
  Compute the adjacency matrix $\bar{\boldsymbol{G}}$ as described in Eq. (31).
  The beams, users, and beam-user pairs with the same maximum vertex degree as $\bar{\boldsymbol{G}}_{l-1}$ are denoted by $\mathcal{D}$.
  The index of optimal user and beam is selected as

$$\{k^*, m^*\} = \arg \max_{\{k',m'\}\in\mathcal{D}} \sum_{m=1}^{M} b_m a_{k',m} + \sum_{k=1}^{K} u_k a_{k,m'} - a_{k',m'}. \tag{36}$$

  if $\mathcal{D} = \varnothing$ **then**
    The index of optimal user and beam is selected as

$$\{k^*, m^*\} = \arg \max_{k'\in\mathcal{A},m'\in\mathcal{C}} \sum_{m=1}^{M} b_m a_{k',m} + \sum_{k=1}^{K} u_k a_{k,m'} - a_{k',m'}. \tag{37}$$

  end
end
Procedure End

IV. Effective Instantaneous CSI Estimation and Performance with Baseband Digital Precoding

The EI-CSI $\bar{\boldsymbol{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$ at the BS is obtained by downlink pilots and uplink feedback. As shown in Fig. 4, suppose the coherence time $T_c = T_p + T_f + T_d$, where T_p, T_f, and T_d are the numbers of symbols occupied by downlink pilot training, uplink feedback, and downlink data transmission, respectively. After pre-beamforming, the EI-CSI $\bar{\boldsymbol{H}}$ is a sparse matrix. Estimating $\bar{\boldsymbol{H}}$ requires fewer pilots T_p and lower uplink feedback T_f than estimating $\boldsymbol{H}$, which requires more symbols for downlink data transmission, thereby improving the performance of NSE. This paper provides pilot matrix design schemes according to the different structures of $\bar{\boldsymbol{H}}$ under the one-ring local scattering and multiple scattering clusters models.

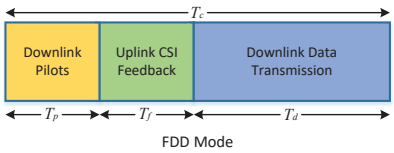


Fig. 4: Duration of a time slot in FDD mode.

A. Design of Pilot Matrix

Given the pilot matrix $\boldsymbol{X}$, the received signal of user k is expressed as

$$\boldsymbol{y}_{p,k}^{\mathrm{H}} = \boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}\boldsymbol{X} + \boldsymbol{n}_k^{\mathrm{H}}, \tag{38}$$

where $\boldsymbol{h}_k$ is the channel between user k and the BS, and $\boldsymbol{n}_k \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$ is additive white Gaussian noise.

1) Pilot Matrix of One-ring Local Scattering Model: Similar to [27], if the center angle $\theta_k, k \in \mathcal{K}$ are sorted in an ascending order, where $\mathcal{K} = \{1, 2, \cdots, K\}$ is the indices of users, then $\bar{\boldsymbol{H}}$ has a banded structure. Therefore, the pilot matrix $\boldsymbol{X}$ is designed to have the same structure as in [27] to estimated the EI-CSI $\bar{\boldsymbol{H}}$. To achieve good performance and for simplicity of implementation, $\boldsymbol{X}$ has the following circular formulation

$$\boldsymbol{X} = \left[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_\mu, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_\mu, \cdots\right]^{\mathrm{H}}, \tag{39}$$

where $\boldsymbol{x}_i, i = 1, 2, \cdots, \mu$ are orthogonal to each other, and μ is NOP that depends on the sparse structure of $\bar{\boldsymbol{H}}$. We design the NOP in Section (III) to avoid the loss of system spectrum efficiency.

2) Pilot Matrix of Multiple Scattering Clusters Model: In the multiple scattering clusters model, the distribution of the non-zero values of the EI-CSI $\bar{\boldsymbol{H}}$ is irregular, imposing difficulty in the design of the pilot matrix. In Section (III), we present an I-BSUS scheme for obtaining the EI-CSI $\bar{\boldsymbol{H}}$. Based on the structure of the resulting $\bar{\boldsymbol{H}}$, we propose a simple IPA scheme for designing the pilot matrix $\boldsymbol{X}$.

The pilot matrix comprises μ orthogonal pilot vectors and is formulated as

$$\boldsymbol{X} = \left[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_\mu\right]^{\mathrm{H}}. \tag{40}$$

The main idea of the IPA scheme is to iteratively assign pilots to the indices of the elements of the k ($k \in \mathcal{K}$)-th row while ensuring that the indices of non-zero elements in each row of the EI-CSI $\bar{\boldsymbol{H}}$ are allocated to orthogonal pilots.

In the k-th iteration, we allocate pilots to the indices of non-zero elements in the k-th row of $\bar{\boldsymbol{H}}$. The indices of non-zero elements that need to be assigned pilots are contained in $\bar{\mathcal{I}}_k$,

$$\bar{\mathcal{I}}_k = \mathcal{I}_k \setminus \mathcal{I}_{sum}, \tag{41}$$

where $\mathcal{I}_{sum}$ is the set containing the indices that have assigned pilots in the previous $k-1$ iterations, and $\mathcal{I}_k$ contains the indices of the non-zero elements in the k-th row of $\bar{\boldsymbol{H}}$.

Let $\mathcal{F}_i$ be the indices of non-zero elements in the i-th column of $\bar{\boldsymbol{H}}$ for $i \in \mathcal{I}_k$. The pilots of the indices in $\mathcal{I}_{re}$ can be used for i, and $\mathcal{I}_{re}$ is expressed as

$$\mathcal{I}_{re} = \mathcal{I}_{sum} \setminus \bigcup_{k \in \mathcal{F}_i} \mathcal{I}_k. \tag{42}$$

If $\mathcal{I}_{re}$ is not empty, we assign pilots to index i using the pilots of the indices in $\mathcal{I}_{re}$. Otherwise, the new pilots are used. We update $\mathcal{I}_{sum}$ and $\bar{\mathcal{I}}_k$ and recompute $\mathcal{I}_{re}$ until all indices in $\bar{\mathcal{I}}_k$ are assigned pilots. The details of the IPA scheme are presented in Algorithm 2.

Algorithm 2: Iterative pilot allocation (IPA)

Input: $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$.
Output: $\boldsymbol{X}$.

```
Initialize I_sum = ∅.
Generate I_k from H^H F.
for k = 1 : K do
    Repeat
    Calculate Ī_k as shown in Eq. (41).
    for i ∈ Ī_k do
        Repeat
        Calculate I_re as shown in Eq. (42).
        if I_re ≠ ∅ then
            Assigned pilots to i using the pilots of the
              indices in I_re.
        else
            The new pilot is assigned to i.
        end
        Update I_sum and Ī_k.
    end
end
Procedure End
```

B. Performance with Baseband Digital Precoding

Similar to [23], [27], the non-zero entries of the EI-CSI $\bar{\boldsymbol{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}$ are assumed to be perfectly estimated and obtained. The precoder $\boldsymbol{W}$ only requires the knowledge of $\bar{\boldsymbol{H}}$, which is designed by the commonly used linear multi-user digital ZF precoder as

$$\begin{aligned}\boldsymbol{W} &= \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right)^{\dagger}\boldsymbol{\Gamma} \\ &= \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right)^{\mathrm{H}}\left(\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right)\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\right)^{\mathrm{H}}\right)^{-1}\boldsymbol{\Gamma} \\ &= \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{\Gamma},\end{aligned} \tag{43}$$

where $\boldsymbol{\Gamma} = diag(\gamma_1, \gamma_2, \cdots, \gamma_K)$ is a diagonal matrix used to normalize each column in $\boldsymbol{W}$.

Given the NOP μ, the NSE of user k is given by

$$\bar{R}_k = (1 - \mu/T_c)\, R_k, \tag{44}$$

where T_c is the coherence time and $R_k = \log_2(1 + \mathrm{SINR}_k)$. With the given beamforming matrix $\boldsymbol{F}$ and $\boldsymbol{W}$, SINR_k is expressed as

$$\mathrm{SINR}_k = \frac{\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}\boldsymbol{w}_k\right|^2 \frac{P}{K}}{\sum_{i \neq k}\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}\boldsymbol{w}_i\right|^2 \frac{P}{K} + \sigma^2}, \tag{45}$$

where $\boldsymbol{w}_k$ is the k-th column of $\boldsymbol{W}$.

The sum rate of the system is given by

$$R_{sum} = \left(1 - \frac{\mu}{T_c}\right)\sum_{k=1}^{K}\log_2\left(1 + \frac{\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}\boldsymbol{w}_k\right|^2 \frac{P}{K}}{\sum_{i \neq k}\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}\boldsymbol{w}_i\right|^2 \frac{P}{K} + \sigma^2}\right). \tag{46}$$

V. Simulations

This section evaluates the performance of the proposed MILP and I-BSUS schemes on various simulations and compares them with the active channel sparsification (ACS) and JSDM schemes [22], [28]. In the following trials, the number of antennas in the BS is $M = 128$. The central AODs of each

user are assumed to be uniformly distributed in the interval $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$. Under the assumption that $r \ll d_k$, the AS is set to $\Delta_{H,k} = \Delta_H = 5°$. Finally, the number of RF chains is set to $N_{RF} = 64$.

A. Simulation Results of One-ring Local Scattering Clusters Model

Fig. 5 depicts the NSE of the proposed MILP, ACS, and JSDM schemes at different SNRs. The number of users is set to 20 and 40, the restricted number of pilots is $N = 15$, and T_c used in Eq. (44) is set to 100 symbols [27]. Fig. 5 highlights that the proposed MILP scheme consistently has a higher NSE than ACS and JSDM, regardless of the number of users. This allows the MILP to reduce pilot length and inter-user interference through its efficient user scheduling scheme. As the number of users increases, the performance gains of the proposed MILP over the ACS become more pronounced, and the NSE of the JSDM scheme is better than ACS at high SNR. This is because the limited number of RF chains hinders the growth of NSE, and the proposed MILP scheme has a good BSUS strategy.

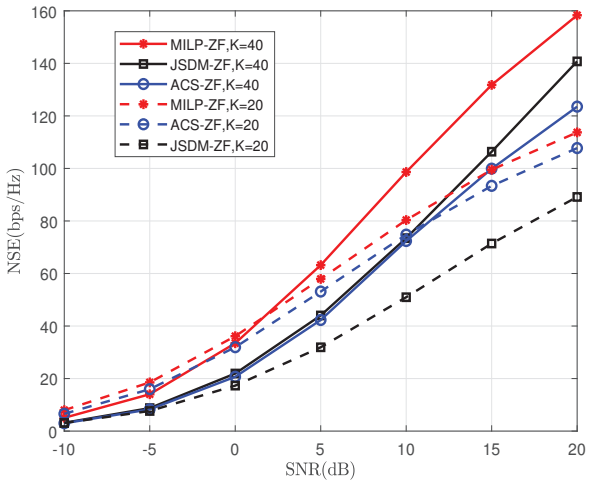


Fig. 5: Comparison of the NSE under different SNRs.

Fig. 6 depicts the NOP of the proposed MILP, ACS, and JSDM schemes under a varying pilot constraint N. The SNR is set to 20 dB. As the pilot constraint N increases, the number of NOP tends to stabilize for all three schemes: MILP, ACS, and JSDM. Notably, the ACS scheme consistently achieves lower NOP than the MILP scheme, particularly when the number of users K is large. Meanwhile, despite having more pilots than the ACS scheme, the MILP scheme exhibits higher NSE values when the NOP stabilizes. Furthermore, the MILP scheme exhibits nearly identical NOP values across different numbers of users. In contrast, JSDM's NOP increases with the number of beams as the number of users increases, as it does not involve user scheduling.

Fig. 7 illustrates the relationship between NSE versus the number of users K, and the SNR is set to 20 dB. T_c is set to 100 symbols. The proposed MILP scheme generally achieves

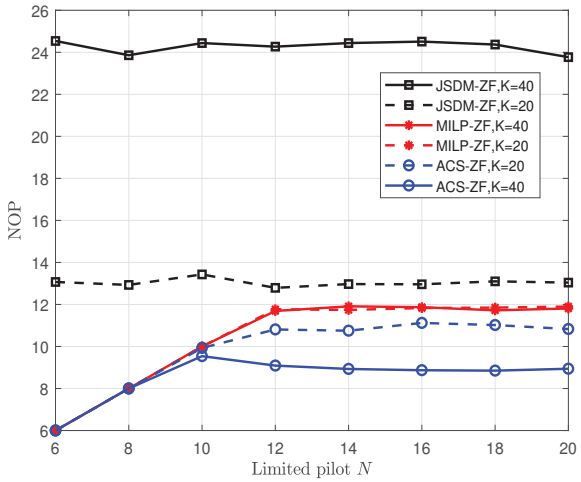


Fig. 6: Comparison of the NOP under different N, SNR=20 dB.

higher NSE than the ACS scheme, especially for a large number of users K. However, the NSE of all three schemes initially increases and decreases as K increases. This is due to the growing user interference, the energy required to transmit signals, and the increasing number of pilots that decrease the NSE. The ACS scheme outperforms JSDM due to its user scheduling scheme, which reduces inter-user interference. Moreover, the limited number of RF chains at the BS hinders the improvement of ESE, especially as the number of users increases, and this effect is particularly significant in the ACS scheme.

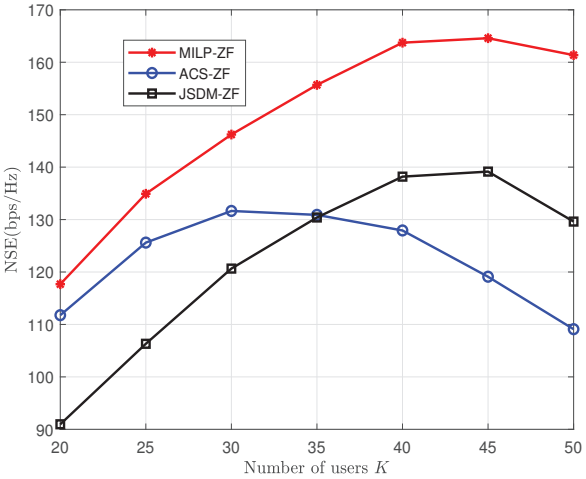


Fig. 7: Comparison of the NSE under different K.

B. Simulation Results of Multiple Scattering Clusters Model

Fig. 8 presents the NSE of the proposed I-BSUS and ACS schemes under varying SNR, with the number of users set to 50. The results infer that the proposed I-BSUS scheme exhibits a higher NSE than the ACS scheme, particularly when the

number of users is large. This is because the I-BSUS scheme uses fewer orthogonal pilots than the ACS scheme. In contrast, the limited number of RF chains impairs the NSE of the ACS scheme.

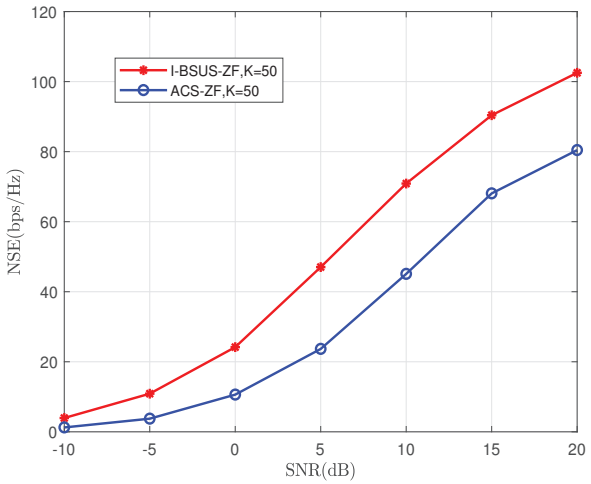


Fig. 8: Comparison of the NSE under different SNRs.

Fig. 9 illustrates the NOP of the proposed I-BSUS and ACS schemes under limited pilot constraint N, with the number of users set to 20 and 40. In Fig. 9, the NOP of the proposed I-BSUS and ACS schemes increases with the limited pilot constraint N and stabilizes at around $N > 25$. The NOP of the proposed I-BSUS scheme is consistently lower than that of the ACS scheme, regardless of the number of users K. Due to the user scheduling scheme, the NOP of the proposed I-BSUS scheme remains almost the same when $K = 20$ and $K = 40$, with the NOP of $K = 40$ being slightly smaller than that of $K = 20$.

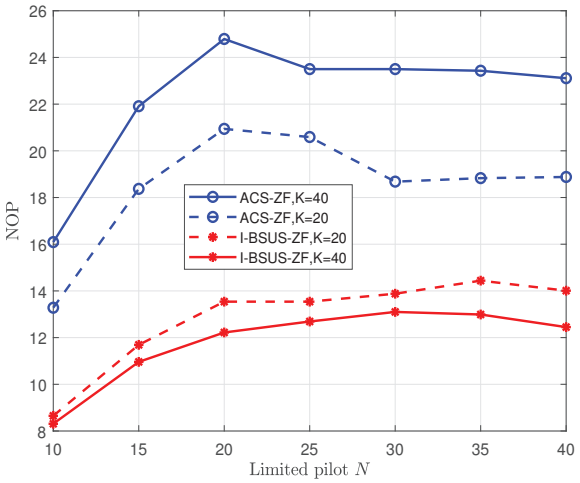


Fig. 9: Comparison of the NOP under different N.

Fig. 10 depicts the NSE of both schemes as a function of the number of users K, with the SNR set to 10 dB. The proposed I-BSUS scheme exhibits a higher NSE than the ACS scheme. As the number of users increases, the NSE of the proposed I-BSUS scheme remains relatively stable. Conversely, the NSE of the ACS scheme decreases as the number of users increases. The proposed I-BSUS scheme, with its smaller number of orthogonal pilots and superior user scheduling scheme, achieves the highest NSE.

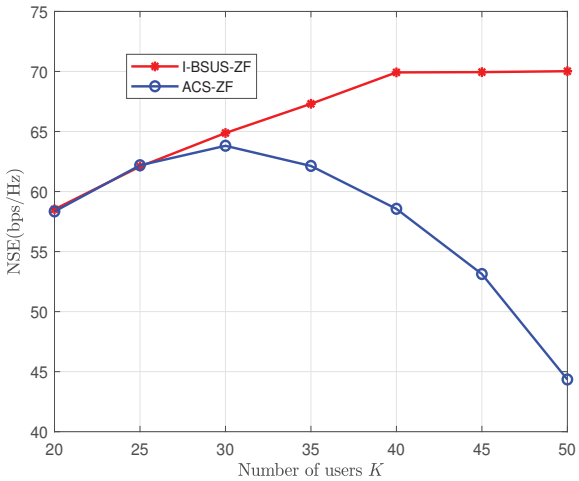


Fig. 10: Comparison of the NSE under different K.

VI. Conclusion

A TSB scheme is proposed to reduce the overhead of pilots and feedback for FDD massive MIMO systems. Additionally, two BSUS schemes that rely only on S-CSI are proposed that reduce pilots and feedback overhead by exploiting channel sparsity. Accurately representing the pilot constraint in the MILP problem enables the optimal BSUS scheme to be obtained. Indeed, the I-BSUS scheme obtains minimum pilots while ensuring system performance. Together, the optimal BSUS scheme and minimum pilots provide a better NSE improvement. The simulation results confirm that the proposed MILP and I-BSUS schemes outperform the ACS scheme regarding NSE. Future work will investigate the TSB scheme with user scheduling in three-dimensional massive MIMO systems.

Appendix
Proof of the Linearization of QCQP

Necessity: If $\boldsymbol{x}^*$ is the optimal solution of $\mathcal{P}_3$, $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is also the optimal solution of $\mathcal{P}_4$.

Given $\boldsymbol{x} \in \{0,1\}^{K+M}$ and $A = \|\boldsymbol{Q}\|_\infty$, it follows that $\boldsymbol{Q}\boldsymbol{x} + A\boldsymbol{r} \geq \boldsymbol{0}$. Therefore, there exist $\boldsymbol{z} \geq \boldsymbol{0}$, $\boldsymbol{s} \geq \boldsymbol{0}$ that satisfy the following equations:

$$\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{s} + A\boldsymbol{r} = \boldsymbol{0}, \tag{47}$$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z} = 0. \tag{48}$$

$\boldsymbol{z}^*$ and $\boldsymbol{s}^*$ can be selected for $\boldsymbol{x}^*$ to satisfy Eq. (47) and Eq. (48) while minimizing $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^*$ in $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}$.

It is worth noting that $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z} = 0$ is a quadratic constraint. From Eq. (47), we have

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{s} + A\boldsymbol{r} \\ &\leq \boldsymbol{Q}\boldsymbol{x} + A\boldsymbol{r} \leq 2A\boldsymbol{r}. \end{aligned} \tag{49}$$

This provides an upper bound for $\boldsymbol{z}$. From Eq. (48), it follows that when $\boldsymbol{x}_i = 0$, $\boldsymbol{z}_i \leq 2A = 2A\left(\boldsymbol{r} - \boldsymbol{x}\right)_i$, and when $\boldsymbol{x}_i = 1$, $\boldsymbol{z}_i = 0 = 2A\left(\boldsymbol{r} - \boldsymbol{x}\right)_i$. Therefore, $\boldsymbol{z} \leq 2A\left(\boldsymbol{r} - \boldsymbol{x}\right)$, and we can use the linear constraint $\boldsymbol{z} \leq 2A\left(\boldsymbol{r} - \boldsymbol{x}\right)$ instead of $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z} = 0$.

By substituting $\boldsymbol{x}^*$ into Eq. (47) and multiplying both sides by $\left(\boldsymbol{x}^*\right)^{\mathrm{T}}$, we obtain

$$\left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}^* - \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{z}^* - \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{s}^* + A\left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{r} = 0. \tag{50}$$

Since $\boldsymbol{0} \leq \boldsymbol{z} \leq 2A\left(\boldsymbol{r} - \boldsymbol{x}\right)$, there must be $\boldsymbol{z}_i = 0$ when $\boldsymbol{x}_i = 1$, and when $\boldsymbol{x}_i = 0$, we have $\boldsymbol{z}_i \leq 2A$, which implies $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z} = 0$. Eq. (50) can be reformulated as

$$\begin{aligned} \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}^* &= \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{s}^* - A\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}^* \\ &\overset{(b)}{=} \boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^* - A\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}^*, \end{aligned} \tag{51}$$

where (b) can be proved by Lemma 1. Thus, $(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*)$ is the optimal solution of problem $\mathcal{P}_4$.

Lemma 1. *For any i, if $x_i^* = 0$, there must be $s_i^* = 0$.*

Proof. Suppose there exists i_0 such that $x_{i_0}^* = 0, s_{i_0}^* > 0$, and $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^*$ is the smallest. When $j = i_0$, we record $\bar{z}_j = z_{i_0}^* + s_{i_0}^*, \bar{s}_j = 0$. When $j \neq i_0$, we record $\bar{z}_j = z_{i_0}^*$ and $\bar{s}_j = s_j^*$.

Since $\bar{\boldsymbol{z}} + \bar{\boldsymbol{s}} = \boldsymbol{z}^* + \boldsymbol{s}^*$, it satisfies Eq. (47) and Eq. (48), but $\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{s}} < \boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^*$, which contradicts the assumption that $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^*$ is the smallest. Therefore, when $x_i^* = 0$, there must be $s_i^* = 0$. □

If $\boldsymbol{x}^*$ is the optimal solution of problem $\mathcal{P}_3$, we define $\boldsymbol{t}_k^* \in \mathbb{R}^{K+M}$ such that for $1 \leq i \leq (K+M)$,

$$\boldsymbol{t}_k^* = \begin{cases} \left(\bar{\boldsymbol{Q}}_k\boldsymbol{x}\right)_i, & x_i^* = 1; \\ 0, & x_i^* = 0. \end{cases} \tag{52}$$

Since $\bar{\boldsymbol{Q}}_k\boldsymbol{x}$ is non-negative, we have $\bar{\boldsymbol{Q}}_k\boldsymbol{x} - \boldsymbol{t}_k^* \leq \boldsymbol{0}$ and $\boldsymbol{t}_k^* \geq \bar{\boldsymbol{Q}}_k$. Using $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{t}_k^* = \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{t}_k^*$, we obtain

$$\begin{aligned} \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\bar{\boldsymbol{Q}}_k\boldsymbol{x}^* &= \sum_{x_i^*=1}\left(\bar{\boldsymbol{Q}}_k\boldsymbol{x}^*\right)_i \\ &= \sum_{x_i^*=1}\left(\boldsymbol{t}_k^*\right)_i \\ &= \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{t}_k^* \\ &= \boldsymbol{r}^{\mathrm{T}}\boldsymbol{t}_k^* \leq N. \end{aligned} \tag{53}$$

From the above proof, we can see that there exists a solution $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ that satisfies the constraint conditions of problem $\mathcal{P}_4$. Therefore, $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is the optimal solution of $\mathcal{P}_4$, and the optimal values of $\mathcal{P}_3$ and $\mathcal{P}_4$ are the same.

Sufficiency: If $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is the optimal solution of problem $\mathcal{P}_4$, then $(\boldsymbol{x}^*)$ must be the optimal solution of problem $\mathcal{P}_3$.

Suppose $\boldsymbol{x}^*$ is not the optimal solution of problem $\mathcal{P}_3$, and $\bar{\boldsymbol{x}}$ is the optimal solution of problem $\mathcal{P}_3$. Then, we have

$$\bar{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{Q}\bar{\boldsymbol{x}} < \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}^*. \tag{54}$$

At the same time, using the method of finding the optimal solution of problem $\mathcal{P}_3$ in the necessity proof, we obtain that $\bar{\boldsymbol{z}}$, $\bar{\boldsymbol{s}}$, $\bar{\boldsymbol{t}}_k$ satisfy the constraints, such that $\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{s}} - A\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{x}}$ is the smallest. Thus, we have

$$\bar{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{Q}\bar{\boldsymbol{x}} = \boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{s}} - A\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{x}}. \tag{55}$$

However, since $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is the optimal solution of the problem $\mathcal{P}_4$, it can be seen from the necessity proof that

$$\left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}^* = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^* - A\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}^*. \tag{56}$$

From $\bar{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{Q}\bar{\boldsymbol{x}} < \left(\boldsymbol{x}^*\right)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}^*$, we know that $\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{s}} - A\boldsymbol{r}^{\mathrm{T}}\bar{\boldsymbol{x}} < \boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}^* - A\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}^*$, which contradicts the fact that $\left(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{s}^*, \boldsymbol{t}_k^*\right)$ is the optimal solution of problem $\mathcal{P}_4$. Therefore, $\boldsymbol{x}^*$ must be the optimal solution to problem $\mathcal{P}_3$, and the optimal values of $\mathcal{P}_3$ and $\mathcal{P}_4$ are the same.

REFERENCES

[1] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[2] A. Ranjha and G. Kaddoum, "URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4618–4627, 2021.

[3] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1295–1315, 2021.

[4] B. Cao, J. Zhang, X. Liu, Z. Sun, W. Cao, R. M. Nowak, and Z. Lv, "Edge–cloud resource scheduling in space–air–ground-integrated networks for internet of vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5765–5772, 2022.

[5] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 363–367, 2021.

[6] L. Wang, G. Liu, J. Xue, and K.-K. Wong, "Channel prediction using ordinary differential equations for mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2111–2119, 2023.

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

[8] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[9] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2350–2361, 2013.

[10] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, "Uplink interference reduction in large-scale antenna systems," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2194–2206, 2017.

[11] D. Verenzuela, E. Björnson, X. Wang, M. Arnold, and S. ten Brink, "Massive-MIMO iterative channel estimation and decoding (MICED) in the uplink," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 854–870, 2020.

[12] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, 2015.

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

[14] Y. Gu and Y. D. Zhang, "Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2334–2346, 2019.

[15] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, 2014.
[16] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.
[17] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2852–2862, 2017.
[18] V. K. N. Lau, S. Cai, and A. Liu, "Closed-loop compressive CSIT estimation in FDD massive MIMO systems with 1 bit feedback," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2146–2155, 2016.
[19] J. Dai, A. Liu, and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2584–2599, 2018.
[20] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2838–2849, 2020.
[21] X. Zheng and V. Lau, "Federated online deep learning for CSIT and CSIR estimation of FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2253–2266, 2022.
[22] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
[23] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 876–890, 2014.
[24] X. Sun, X. Gao, G. Y. Li, and W. Han, "Agglomerative user clustering and downlink group scheduling for FDD massive MIMO systems," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.
[25] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2200–2211, 2015.
[26] Y. Jeon, C. Song, S.-R. Lee, S. Maeng, J. Jung, and I. Lee, "New beamforming designs for joint spatial division and multiplexing in large-scale MISO multi-user systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3029–3041, 2017.
[27] Y. Song, C. Liu, Y. Liu, N. Cheng, Y. Huang, and X. Shen, "Joint spatial division and multiplexing in massive MIMO: A neighbor-based approach," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7392–7406, 2020.
[28] M. Barzegar Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 121–135, 2019.
[29] Z. Zou, S. Zhao, G. Huang, and D. Tang, "Novel design of user scheduling and analog beam selection in downlink millimeter-wave communications," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4168–4178, 2022.
[30] S. Zou, J. Wu, H. Yu, W. Wang, L. Huang, W. Ni, and Y. Liu, "Efficiency-optimized 6G: A virtual network resource orchestration strategy by enhanced particle swarm optimization," *Digital Communications and Networks*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352864823001141

Tianbao Gao received the B.S. degree in Electronic Information Science and Technology from Xi'an Technological University, China, in 2013. He is currently working toward the Ph.D. degree with the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China. His research interests include massive MIMO systems.

Chen Liu (Member, IEEE) received the B.E. degree in Electrical and Information Engineering from the Nanjing Institute of Technology (Southeast University), China, in 1985, the M.S. degree in circuits and systems from Anhui University, China, in 1988, and the Ph.D. degree in signal and information processing from Southeast University, China, in 2005. In 1988, he joined the Nanjing University of Posts and Telecommunications (NUPT), where he has been a Professor since 2002. His current research interest includes massive MIMO systems.

Yunchao Song (Member, IEEE) received the B.E. degree in Electronic Science and Technology and the Ph.D. degree in Circuits and Systems from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2010 and 2016, respectively. Since 2017, he has been an Instructor with the College of Electronic and Optical Engineering, NUPT. He is currently a Visiting Scholar with the BBCR Lab, Department of Electronics and Communication Engineering (ECE), University of Waterloo, Canada. His research interest includes massive MIMO systems.

Zhisheng Yin (Member, IEEE) received the Ph.D. degree from the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China, in 2020. From September 2018 to September 2019, he visited with the BBCR Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently an Assistant Professor with State Key Lab of ISN, and with School of Telecommunications Engineering, Xidian University, Xi'an, China. His research interests include SAGINs, cybertwin, wireless communications, and physical layer security.

Huibin Liang received the B.S. degree in Optoelectronic Information Science and Engineering from Nanjing University of Posts and Telecommunications, China, in 2021. He is currently working toward the Ph.D. degree with the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include massive MIMO systems.

Nan Cheng (Member, IEEE) received the B.E. and M.S. degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2016. From 2017 to 2018, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON. He is currently a Professor with State Key Lab of ISN, and with School of Telecommunications Engineering, Xidian University, Xi'an, China. His current research interests include space-air-ground integrated system, Big Data in vehicular networks, and application of AI for vehicular networks.