Joint Spatial Division and Multiplexing in Massive MIMO: A Neighbor-Based Approach

Yunchao Song, *Member, IEEE*, Chen Liu, *Member, IEEE*, Yiliang Liu, *Student Member, IEEE*, Nan Cheng, *Member, IEEE*, Yongming Huang, *Senior Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*— In this paper, we propose a joint spatial division and multiplexing (JSDM) beamforming based on a neighbor scheme for frequency division duplex (FDD) massive multi-input multi-output (MIMO) systems. The neighbor-based JSDM (N-JSDM) can fully utilize signal space, leading to higher spectral efficiency over the conventional JSDMs. The reason is that for the neighbor scheme, neighbors and non-neighbors are classified adaptively by the angles of departure (AoD), and the prebeamformer is designed to mitigate the non-neighbors' interference by the statistical channel state information. The effective channel matrix after the prebeamformer then becomes a band matrix, from which the downlink training length (DTL) and the channel feedback length are much smaller than the number of antennas. Moreover, an optimal prebeamformer which is proved to be able to achieve the same system capacity as the full CSI system is proposed, followed by a suboptimal prebeamformer with constrained DTL, and a DFT-based prebeamformer. On the other hand, the neighbors' interference is mitigated using the banded channel state information. Simulation results validate the good performance of the proposed N-JSDM.**

***Index Terms*— Massive MIMO, beamforming, joint spatial division and multiplexing, band matrix.**

Manuscript received December 3, 2019; revised April 24, 2020 and July 4, 2020; accepted July 12, 2020. Date of publication July 29, 2020; date of current version November 11, 2020. This work was supported in part by the Natural Science Foundation of China under Grant 61771257 and Grant 61871238, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190732, in part by the Natural Science Foundation of Jiangsu Higher Education Institutions under Grant 18KJB510027, in part by the 2018 Jiangsu Provincial Government Scholarship Program, in part by the Scientific Research Staring Foundation for Introduced Talents in Nanjing University of Posts and Telecommunications (NUPTSF) under Grant NY217125, and in part by the Natural Sciences and Engineering Research Council (NSERC), Canada. The associate editor coordinating the review of this article and approving it for publication was J. Zhang. *(Corresponding author: Chen Liu.)*

Yunchao Song and Chen Liu are with the College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: songyc@njupt.edu.cn; liuch@njupt.edu.cn).

Yiliang Liu is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: liuyiliang@hit.edu.cn).

Nan Cheng is with the State Key Laboratory of ISN, Xidian University, Xi'an 710071, China, and also with the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: dr.nan.cheng@ieee.org).

Yongming Huang is with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: huangym@seu.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2020.3011101

I. Introduction

MULTI-INPUT multi-output (MIMO) systems have shown a great potential to improve the spectral efficiency [1]–[3]. To meet the requirement of ultra-high data rate in future moblie communications [4]–[6], massive MIMO, i.e., the MIMO technology employing large scale antenna array at the base station (BS), is adopted to further enhance system performance [7]–[11]. It is interesting to note that with the instantaneous channel state information at the transmitter (CSIT) available, the performance gain of the massive MIMO can be maximized. In time division duplex (TDD) massive MIMO systems, the instantaneous CSIT can be obtained through uplink channel estimation due to channel reciprocity [12], [13]. Hence, TDD massive MIMO attracts a lot of attention in academia and industry. However, the frequency division duplex (FDD) is adopted in most current cellular systems and has more mature industrial products and market share. Thus it is important and necessary to develop massive MIMO systems under FDD mode. Since the uplink and downlink channels are not reciprocal in FDD [14], [15], it is usually expensive for FDD systems to obtain the instantaneous CSIT, due to the large downlink training length (DTL) and CSIT uplink feedback.

Certain works have been dedicated to reducing the DTL and channel feedback in FDD massive MIMO. Similar to the TDD mode, extrapolating the downlink instantaneous CSI from uplink pilot estimates can completely remove channel feedback [16], although it brings in the estimation error. Many works consider the spatially-correlated MIMO channels and exploit the CSI structure to reduce the DTL as well as the channel feedback [17]–[21]. Specifically, the compressive sensing based techniques [17], [18] are used to exploit the channel sparsity. By exploiting the low-rank channel covariance matrices (CCMs), the DTL and channel feedback are shown to be reduced significantly [19], [20]. Using the low-rank channel matrix caused by correlation among users, the joint channel estimation for all users is proposed [21]. Some other works exploit the correlation in both time and spatial domain [22], [23]. Utilizing the prior information, i.e., the long-term channel statistics, an open-loop and closed-loop training framework [22] is proposed to reduce the DTL. A trellis-code based quantization codebooks can reduce the cost of both CSI quantization and feedback [23]. The above works design efficient estimation and feedback techniques to obtain the whole instantaneous CSIT. Recently, considering

the CCMs can be obtained from the uplink estimation, a 3D beamforming utilizing only CCMs is proposed to remove the DTL and channel feedback, although there exists performance loss compared to the instantaneous CSIT case [24]. Furthermore, aided by the CCMs, a two-stage beamforming scheme, namely joint spatial division and multiplexing (JSDM) [14], is proposed. The JSDM divides users into several groups and transmits signals to each group separately, and thus the low dimensional effective CSI in each group is considered, which can significantly reduce the DTL and channel feedback. It is shown that the JSDM incurs no loss of optimality with respect to the instantaneous CSIT case under a "tall unitary" condition [14], where the "tall unitary" condition means that the covariance eigenspaces of different groups are orthogonal to each other. This interesting result has aroused a lot of interests, and has prompted a series of studies, such as the beamformer designs [25]–[27] and the user grouping schemes [15], [28].

There are two stages in the JSDM beamforming. In the first stage, a prebeamformer is designed to mitigate the inter-group interference using statistical CSI (i.e., CCMs). In the second stage, a precoding scheme is used to manage the intra-group interference using the instantaneous CSI of each group. Many works have been attempted to enhance the JSDM. In [25], Kim *et al.* designed the prebeamformer by maximizing the signal-to-leakage-and-noise ratio (SLNR), leading to a higher spectral efficiency. The JSDM beamforming designing a prebeamformer and a multi-user precoder sequentially by minimum mean squared error (MMSE) criterion is proposed to further improve the spectral efficiency [26]. Moreover, to alleviate the high hardware cost caused by a large number of radio frequency (RF) chains, Liu *et al.* used the phase shifters in the prebeamformer [27]. Chen *et al.* studied a multi-cell case where the inter-cell interference is canceled by the prebeamformer and the intra-cell multiplexing is achieved in the second stage [29]. Considering that most works focus on one-ring channel, the extension of JSDM to the multi-cluster scenario in millimeter wave is proposed [30]. These works assume that users are distributed in a way that users in the same group have the same CCMs.

In real-world scenarios, users are randomly distributed. As a result, to implement the JSDM beamforming, some grouping ways have been proposed to make the users in the same group have similar CCMs while the users in different groups have approximately orthogonal CCMs. For example, two kinds of grouping schemes are proposed in [15]. One is the K-means clustering algorithm, and the other is a heuristic grouping algorithm which performs relatively worse than K-means way. K-means method uses the chordal distance between the covariance eigen-spaces to reflect the distance between groups, and then generates the groups by comparing the possible grouping realizations. A recent work [28] adopts a hierarchical clustering algorithm which considers both target number of clusters and chordal distance threshold to achieve a better grouping performance. It should be noted that, due to the randomly distributed users, there always exist overlapped signal space between different groups, while the overlapped signal space cannot be used to transmit signals. It follows that the "tall unitary" condition is not satisfied, and there exists performance loss compared to the instantaneous CSIT case.

In this paper, we consider a single-cell FDD massive MIMO system, where the users are randomly distributed, and propose a neighbor-based JSDM (N-JSDM) to improve the spectral efficiency. The N-JSDM avoids the user grouping problem by adopting the neighbor scheme to fully utilize the signal space. Hence, it has a higher spectral efficiency than the conventional JSDMs. The proposed N-JSDM is a two-stage scheme. In the first stage (the prebeamforming stage), the neighbors and non-neighbors of each user is defined based on the users' angles of departure (AoD), such that the CCM of a user is approximately orthogonal to that of its non-neighbors. Then, a prebeamformer using the statistical CSI is designed to cancel the interference from non-neighbors. Moreover, the effective channel matrix after the prebeamformer becomes banded. It follows that the CSIT feedback length and the DTL can be significantly reduced, when compared to the number of antennas. On the other hand, the remaining interference is mitigated in the second stage. Simulations validate that the N-JSDM outperforms the conventional JSDMs in terms of the spectral efficiency. The contributions of this paper are summarized as follows.

- A new JSDM based on the neighbor scheme is proposed. Under the neighbor scheme, all signal space can be utilized such that the spectral efficiency can be improved compared with the conventional JSDMs. Moreover, the effective channel matrix in N-JSDM is a band matrix. Thus, similar to the conventional JSDMs, the CSIT feedback length is much smaller than the number of antennas. The pilot matrix is designed to have the circular structure, yielding that the DTL is also much smaller than the number of antennas.
- Three prebeamformer designs are proposed in the N-JSDM. The first one is an optimal prebeamformer which can achieve the same system capacity as the system with full CSI. In the second prebeamformer design, the DTL is constrained such that the N-JSDM has a smaller DTL. Finally, when the number of antennas becomes infinity, a low-complexity prebeamformer design based on the discrete Fourier transform (DFT) matrix is proposed.

The remainder of this paper is organized as follows. Section II describes the system model along with the JSDM beamforming. In Section III, we formulate the detail of the N-JSDM. The optimal and suboptimal prebeamformer designs are presented in Section IV, followed by the DFT-based prebeamformer design when the number of antennas approaches infinity. Simulation results are presented in Section V, and we conclude this paper in Section VI.

Notations: Bold uppercase letters denote matrices, and bold lowercase letters denote column vectors. $\mathbf{a}_i$ and $\mathbf{a}^i$ denote the i-th column and the i-th row of the matrix $\mathbf{A}$ respectively. The superscripts $(\cdot)^H$ and $(\cdot)^\dagger$ indicate the matrix conjugate-transpose and pseudo inverse operations, respectively. $span\,(\mathbf{A})$ denotes the space spaned by $\mathbf{A}$. The complex number set is represented by $\mathbb{C}$, and $\iota = \sqrt{-1}$. $\mathcal{K} = \{1, 2, \cdots, K\}$.

II. Preliminary

A. System Model

We consider a single-cell FDD massive MIMO system with one base station (BS) serving K single-antenna users. The BS is equipped with a unitary linear antenna (ULA) array of M elements. In the downlink, the BS applies an $M \times K\,(M \geq K)$ precoder $\mathbf{V}$ to transmit symbols. Let $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ denote the downlink channel vector from the BS to user k, and $\mathbf{y} = [y_1, y_2, \cdots, y_K]^H \in \mathbb{C}^{K\times 1}$ is the received signal with y_k being user k's signal, then

$$\mathbf{y} = \mathbf{H}^H\mathbf{V}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2 \cdots, \mathbf{h}_K]$, $\mathbf{s} = [s_1, s_2, \cdots, s_K]^H \in \mathbb{C}^{K\times 1}$ is the transmitted signal with $E\left(\mathbf{s}\mathbf{s}^H\right) = \mathbf{I}$, and $\mathbf{n} = [n_1, n_2, \cdots, n_K]^H \in \mathbb{C}^{K\times 1}$ is the Gaussian noise with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

In this paper, we consider a one-ring channel [31], [32], where user k has an azimuth center angle θ_k and an angular spread (AS) Δ, and θ_k is randomly distributed. Here, users are defined such that $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_K$. The (m, n)th element of user k's the CCM $\mathbf{R}_k$ is [31], [32]

$$[\mathbf{R}_k]_{m,n} = \frac{1}{2\Delta}\int_{\theta_k-\Delta}^{\theta_k+\Delta} e^{-\iota 2\pi\frac{D}{\lambda_c}(m-n)\sin\theta}d\theta \tag{2}$$

where λ_c is the wave length, D is the spacing between two antenna elements, and D is set to $\frac{\lambda_c}{2}$. Using the Karhunen-Loeve representation, user k's channel vector is represented as $\mathbf{h}_k = \mathbf{R}_k^{1/2}\mathbf{z}_k$ where $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the small-scale fading.

B. JSDM Beamforming

In the conventional JSDMs, by taking the similarity of their CCMs into account, users are partitioned into G groups with the g-th$(g = 1, 2, \cdots, G)$ group containing K_g users, and $\sum_{g=1}^{G} K_g = K$. There are two stages in the conventional JSDMs, one involving a prebeamforming $\mathbf{B}$ and the other involving a beamforming $\mathbf{W}$. The precoding matrix is written as $\mathbf{V} = \mathbf{B}\mathbf{W}$. In the first stage, each group's prebeamforming matrix is designed to reduce the inter-group interference according to the CCMs since the CCMs are easy to obtain. Let $\mathbf{C}_g(g = 1, 2, \cdots, G)$ denote the CCM of the g-th group. The prebeamforming matrix $\mathbf{B}_g$ of the g-th group is designed satisfying $span(\mathbf{B}_g) \in span^{\perp}(\mathbf{U}_j)$ for all $j \neq g$, where $span^{\perp}(\cdot)$ denotes the orthogonal space of $span(\cdot)$, $\mathbf{U}_j$ denotes the eigenvectors of $\mathbf{C}_j$ corresponding to the r_j dominnant eigenvalues, and r_j is determined by system parameters [15], [28]. As a result, $\mathbf{B}_g^H\mathbf{C}_j^{1/2} \approx \mathbf{0}\ (j \neq g)$, yielding that $\mathbf{H}_j^H\mathbf{B}_g \approx \mathbf{0}\ (j \neq g)$, where $\mathbf{H}_j$ denotes the channel matrix of the users in the j-th group. After the prebeamformer,

$$\mathbf{H}^H\mathbf{B} = \begin{pmatrix} \mathbf{H}_1^H\mathbf{B}_1 & \mathbf{H}_1^H\mathbf{B}_2 & \cdots & \mathbf{H}_G^H\mathbf{B}_G \\ \mathbf{H}_2^H\mathbf{B}_1 & \mathbf{H}_2^H\mathbf{B}_2 & \cdots & \mathbf{H}_G^H\mathbf{B}_G \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{H}_G^H\mathbf{B}_1 & \mathbf{H}_G^H\mathbf{B}_2 & \cdots & \mathbf{H}_G^H\mathbf{B}_G \end{pmatrix} \tag{3}$$

where $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_G]$. Since $\mathbf{H}_j^H\mathbf{B}_g \approx \mathbf{0}\ (j \neq g)$, the effective channel matrix $\mathbf{H}^H\mathbf{B}$ becomes a block diagonal matrix. Hence, to obtain the instantaneous CSI $\mathbf{H}^H\mathbf{B}$, a more practical approach would be to estimate and feedback only G diagonal matrices, i.e., $\mathbf{H}_g^H\mathbf{B}_g, g = 1, 2, \cdots, G$, which will reduce the overhead of both the downlink training and the uplink CSI feedback. In the second stage, $\mathbf{W}$ is designed as $\mathbf{W} = diag(\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_G)$ so as to mitigate the intra-group interference. The received signal of the g-th group's users is represented as

$$\mathbf{y}_g = \mathbf{H}_g^H\mathbf{B}_g\mathbf{W}_g\mathbf{s}_g + \sum_{g'\neq g}\mathbf{H}_g^H\mathbf{B}_{g'}\mathbf{W}_{g'}\mathbf{s}_{g'} + \mathbf{n}_g \tag{4}$$

where $\mathbf{n}_g \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$. By assuming the perfect effective CSI $\mathbf{H}_g^H\mathbf{B}_g\,(g = 1, 2, \cdots, G)$, we can use the zero-forcing (ZF) precoding [33], [34] to calculate $\mathbf{W}_g$ based on $\mathbf{H}_g^H\mathbf{B}_g$.

The conventional JSDMs partition users into G groups, and when the users are distributed randomly, there always exist common space between the signal spaces of adjacent groups, where the signal space of the g-th group is denoted by $span(\mathbf{H}_g)$. To mitigate the inter-group interference, $span(\mathbf{B}_g)$ is orthogonal to all the signal space $span(\mathbf{H}_j), j \neq g$. It follows that $span(\mathbf{B}_g)$ is orthogonal to all the overlapped signal space, and hence $\bigcup_{g=1,2,\cdots,G} span(\mathbf{B}_g)$ (i.e., $span(\mathbf{B})$) is orthogonal to all the overlapped signal space. As a result, $span(\mathbf{H}) \nsubseteq span(\mathbf{B})$, yielding that the dimension of the utilized signal space $span\left(\mathbf{B}^H\mathbf{H}\right)$ is smaller than that of the full signal space $span(\mathbf{H})$. Consequently, there exists spectral efficiency loss between the JSDM and the instantaneous case in (1).

III. Formulation of N-JSDM

In this section, the N-JSDM using the neighbor scheme instead of the grouping scheme is proposed. We first give the definitions of "neighbor" and "non-neighbor". For user k, the non-neighbors are defined as the users whose azimuth center angles are far away from that of user k, from which the index set of user k's non-neighbors is defined as

$$\bar{\Omega}_k = \left\{j \middle| \, |\theta_k - \theta_j| > \tau\right\} \tag{5}$$

where τ is called neighbor angular spread (NAS). Then the index set of user k and its neighbors is $\Omega_k = \left\{j \middle| \, |\theta_k - \theta_j| \leq \tau\right\}$. Since $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_K$, the elements in Ω_k are the consecutive numbers, and the index set Ω_k can be written as $\Omega_k = \{k_l, \cdots, k, \cdots, k_u\}$. It is straightforward that the larger τ is, the more neighbors user k has.

Our N-JSDM is a two-stage scheme. In the first stage (the prebeamforming stage), we design the prebeamformer based on the CCMs to reduce the non-neighbors' interference, and the effective channel matrix becomes banded. The neighbors' interference is canceled in the second stage. The proposed N-JSDM is based on the fact that, for each user, there exist many non-neighbors whose channel paths' AoDs are far away from this user's channel paths' AoDs. When the number of antennas becomes large, the CCMs of a user is approximately orthogonal to the CCMs of its' nonneighbors [35], and consequently it is easy to cancel the interference from

non-neighbors, which is implemented in the prebeamforming stage.

In the first stage of the proposed N-JSDM, we design each user's prebeamforming matrix $\mathbf{B}_k\,(k=1,2,\cdots,K)$ to mitigate the interference from non-neighbors, which is different from the conventional JSDMs where the prebeamforming matrix is designed for each group. It is worth noting that the number of columns of each $\mathbf{B}_k$ is small and not fixed, which depends on the designing ways. Given the neighbors and non-neighbors, to cancel the interference from non-neighbors, we use the CCMs $\mathbf{R}_k\,(k=1,2,\cdots,K)$ to design $\mathbf{B}_i$ to satisfy, for each k,

$$\mathbf{h}_k^H\mathbf{B}_i=\mathbf{0},\quad i\in\bar{\Omega}_k \tag{6}$$

the detail of which is shown in the next section. The effective channel matrix after the prebeamforming stage is given as

$$\mathbf{H}^H\mathbf{B}=\begin{pmatrix}\mathbf{h}_1^H\mathbf{B}_1 & \mathbf{h}_1^H\mathbf{B}_2 & \cdots & \mathbf{h}_1^H\mathbf{B}_K\\ \mathbf{h}_2^H\mathbf{B}_1 & \mathbf{h}_2^H\mathbf{B}_2 & \cdots & \mathbf{h}_2^H\mathbf{B}_K\\ \cdots & \cdots & \cdots & \cdots\\ \mathbf{h}_K^H\mathbf{B}_1 & \mathbf{h}_K^H\mathbf{B}_2 & \cdots & \mathbf{h}_K^H\mathbf{B}_K\end{pmatrix} \tag{7}$$

where the whole prebeamforming matrix is represented by $\mathbf{B}=[\mathbf{B}_1,\mathbf{B}_2,\cdots,\mathbf{B}_K]$. From (6), the k-th row of $\mathbf{H}^H\mathbf{B}$ can be written as

$$\begin{aligned}\mathbf{h}_k^H\mathbf{B}&=\begin{pmatrix}\mathbf{h}_k^H\mathbf{B}_1, & \mathbf{h}_k^H\mathbf{B}_2, & \cdots, & \mathbf{h}_k^H\mathbf{B}_K\end{pmatrix}\\&=\begin{pmatrix}\cdots, & 0, & 0, & \mathbf{h}_k^H\mathbf{B}_{\Omega_k}, & 0, & 0, & \cdots\end{pmatrix}\end{aligned} \tag{8}$$

where $\mathbf{B}_{\Omega_k}=[\mathbf{B}_{k_l},\cdots,\mathbf{B}_{k_u}]$ is the aggregate prebeamforming matrix for user k and its neighbors. It is easy to observe that, to obtain the instantaneous effective CSI $\mathbf{H}^H\mathbf{B}$, we only need to consider the non-zero entries. For user j and user k, if $k<j$, we have $\theta_k\le\theta_j$, leading to $k_l\le j_l$ and $k_u\le j_u$. As a consequence, $\mathbf{H}^H\mathbf{B}$ is a band matrix.

Remark: Subject to the constraint that the interference from non-neighbors is canceled in the prebeamforming stage, i.e., the constraint in (6), the signal space $span\left(\mathbf{B}^H\mathbf{H}\right)$ in our N-JSDM can be designed to be equivalent to the full signal space $span\,(\mathbf{H})$, which is proved in the next section. Thus our N-JSDM can provide higher spectral efficiency than the conventional JSDMs. Moreover, the effective channel matrix $\mathbf{H}^H\mathbf{B}$ is a band matrix, and consequently, user k only need to feedback $\mathbf{h}_k^H\mathbf{B}_{\Omega_k}$ to the BS, resulting in a feedback length of d_k, where d_k denotes the number of elements of $\mathbf{h}_k^H\mathbf{B}_{\Omega_k}$, i.e., the number of columns of $\mathbf{B}_{\Omega_k}$. By setting proper $\mathbf{B}$, the channel feedback can be reduced.

Next, we show that the banded structure of the effective channel matrix is also beneficial in simplifying the downlink training. Letting $\mathbf{X}$ be the pilot matrix, the received signal is given as

$$\mathbf{Y}^H=\mathbf{H}^H\mathbf{B}\mathbf{X}+\mathbf{N} \tag{9}$$

where $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_K]$ with $\mathbf{y}_k$ being the received signal of user k. $\mathbf{N}$ is the noise matrix with the i.i.d. zero mean and δ^2 variance entries. Consequently, the received signal of user k is given by

$$\mathbf{y}_k^H=\mathbf{h}_k^H\mathbf{B}\mathbf{X}+\mathbf{n}^k \tag{10}$$

where $\mathbf{n}^k$ is the k-th row of $\mathbf{N}$.

Define $\bar{\mathbf{h}}_k=\mathbf{h}_k^H\mathbf{B}_{\Omega_k}$. According to (8), to acquire user k's effective channel $\mathbf{h}_k^H\mathbf{B}$, only $\bar{\mathbf{h}}_k$ needs to be estimated. Letting $\bar{\mathbf{X}}^k$ denote the matrix composed by the rows of $\mathbf{X}$ corresponding to $\bar{\mathbf{h}}_k$, and by removing the zero terms in (10), we have

$$\mathbf{y}_k^H=\bar{\mathbf{h}}_k^H\bar{\mathbf{X}}^k+\mathbf{n}^k. \tag{11}$$

During the estimation of $\bar{\mathbf{h}}_k$, orthogonal pilot matrix is preferable due to good performance and easy implementation, i.e., the rows in $\bar{\mathbf{X}}^k$ should be orthogonal to each other. Note that d_k is the number of non-zero values in the k-th row in $\mathbf{H}^H\mathbf{B}$. Let the pilot matrix $\mathbf{X}\in\mathbb{C}^{r\times\mathcal{L}}$ having the following circular formulation

$$\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_N,\mathbf{x}_1,\mathbf{x}_2\cdots,\mathbf{x}_N,\cdots]^H \tag{12}$$

where r is the same as the number of columns of $\mathbf{B}$, N is a constant with $N\ge\max_k d_k$, $\mathcal{L}\,(\mathcal{L}\ge N)$ is the DTL, and $\mathbf{x}_i\in\mathbb{C}^{\mathcal{L}\times 1}$ are orthogonal to each other. Then it is easy to verify that all $\bar{\mathbf{X}}^k$ are unitary matrices, and the minimum DTL is $L=\max_k d_k$. In this work, we do not focus on the channel estimation techniques [36], and similar to the conventional JSDMs [15], [25]–[28], the perfect effective CSI $\bar{\mathbf{h}}_k^H$ for each k is assumed available at the BS.

Let $\hat{\bar{\mathbf{H}}}$ denote the estimation of the effective channel matrix $\mathbf{H}^H\mathbf{B}$. In the second stage of our N-JSDM, to mitigate the interference from neighbors, we use ZF precoding [34] to calculate $\mathbf{W}$ based on $\hat{\bar{\mathbf{H}}}$, i.e., $\mathbf{W}$ is given by

$$\mathbf{W}=\left(\hat{\bar{\mathbf{H}}}\right)^{\dagger}\mathbf{\Gamma} \tag{13}$$

where $\mathbf{\Gamma}=diag\,(\gamma_1,\gamma_2,\cdots,\gamma_K)$ is a diagonal matrix used to normalize each column in $\mathbf{W}$. As a consequence, the signal-to-interference-plus-noise ratio (SINR) of user k is given by

$$SINR_k=\frac{\left|\mathbf{h}_k^H\mathbf{B}\mathbf{w}_k\right|^2}{\sum_{k'\neq k}\left|\mathbf{h}_k^H\mathbf{B}\mathbf{w}_{k'}\right|^2+\sigma^2} \tag{14}$$

where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}$. The SINR can be used to obtain the spectral efficiency [37]. Also it is an important factor that reflects the characteristics of the wireless interference accurately [38], [39]. For example, in [38], Yu *et al*. designed a protocol based on SINR interference model for multi-hop wireless networks, which achieves a good performance in realistic environments.

IV. Prebeamformer Design

In this section, three prebeamformer designs based on the CCMs are given. As shown in Section III, to reduce DTL and CSIT feedback, the interference from non-neighbors is canceled in the prebeamforming stage, leading to the effective channel matrix to be banded. Moreover, to mitigate the non-neighbors' interference of user k, the prebeamforming matrix $\mathbf{B}$ should be designed satisfying (6). It is interesting to know that if user i is user k's neighbor, user k is also user i's neighbor, and thus (6) is equivalent to the problem of designing $\mathbf{B}_k$ to satisfy

$$\mathbf{h}_i^H\mathbf{B}_k=\mathbf{0}, i\in\bar{\Omega}_k \tag{15}$$

for each k. Since $\mathbf{h}_i = \mathbf{R}_i^{1/2}\mathbf{z}_i$ and $\mathbf{R}_i^{1/2}$ is a Hermitian matrix, (15) is equivalent to

$$\mathbf{z}_i^H \mathbf{R}_i^{1/2} \mathbf{B}_k = \mathbf{0}, i \in \bar{\Omega}_k. \tag{16}$$

In the prebeamforming stage, only the CCMs $\mathbf{R}_k$ is available at the BS. Without the knowledge of $\mathbf{z}_i$, Eq. (16) is equivalent to the following formulation,

$$\mathbf{R}_i^{1/2} \mathbf{B}_k = \mathbf{0}, i \in \bar{\Omega}_k \tag{17}$$

which means $span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\mathbf{R}_i^{1/2}\right)$ for each $i \in \bar{\Omega}_k$. We now give a lemma to analyze the constraints in (17).

Lemma 1: For any matrices $\mathbf{A}_i, i = 1, 2, \cdots, N$, define the space $\mathcal{S}$ as $\mathcal{S} = \bigcap_i span^{\perp}\left(\mathbf{A}_i\right)$, then

$$\mathcal{S} = span^{\perp}\left(\sum_{i=1}^{N} \mathbf{A}_i \mathbf{A}_i^H\right) \tag{18}$$

Proof: See Appendix A. ∎

Lemma 1 transforms multiple constraints in (17) into one constraint which is easy to analyze. From Lemma 1, subject to (17), $\mathbf{B}_k$ should satisfy $span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\sum_{i\in\bar{\Omega}_k} \mathbf{R}_i^{1/2}\left(\mathbf{R}_i^{1/2}\right)^H\right) = span^{\perp}\left(\sum_{i\in\bar{\Omega}_k} \mathbf{R}_i\right)$. Letting $\bar{\mathbf{R}}_k = \sum_{i\in\bar{\Omega}_k} \mathbf{R}_i$, we have

$$span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right). \tag{19}$$

The condition in (19) guarantees that the effective channel matrix is a band matrix. However, it cannot guarantee the DTL minimization or the spectral efficiency maximization. As shown above, the DTL in our N-JSDM is related to the column numbers of $\mathbf{B}_k, k = 1, 2, \cdots, K$. Then we hope the column number of each $\mathbf{B}_k$ is designed to be small, such that the DTL is small. However, the column numbers cannot be too small, otherwise the dimension of $span\left(\mathbf{B}^H\mathbf{H}\right)$ will be small, yielding that the signal space $span\left(\mathbf{B}^H\mathbf{H}\right)$ is much smaller than $span\left(\mathbf{H}\right)$, and consequently the capacity of our N-JSDM will be decreased. There exists a tradeoff between the capacity and the DTL in our N-JSDM. Next, we propose three kinds of prebeamformer designs.

A. Capacity-Approaching Prebeamformer

Given the banded effective channel matrix $\mathbf{H}^H\mathbf{B}$, and by the duality between multiple-access channel and broadcast channel [14], [40], the capacity of our N-JSDM is formulated as

$$\mathcal{C}^{sum}\left(\mathbf{H}, \mathbf{B}\right) = \max_{\mathbf{S}\succeq 0, Tr(\mathbf{S})\leq P} \log\left|\mathbf{I} + \mathbf{B}^H\mathbf{H}\mathbf{S}\mathbf{H}^H\mathbf{B}\right| \tag{20}$$

where $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_K]$, $\mathbf{S}$ denotes the input covariance matrix, and P is the power of the transmit signal. When designing $\mathbf{B}$, the second precoding matrix $\mathbf{W}$ is not considered.

In our N-JSDM, the matrix $\mathbf{B}$ is set to be a unitary matrix, i.e., $\mathbf{B}^H\mathbf{B} = \mathbf{I}$, such that our N-JSDM can achieve the same capacity as the multi-user system in (1), as shown in Theorem 2, when the number of antennas approaches infinity. With $\mathbf{B}$ a unitary matrix, each user's prebeamforming matrix $\mathbf{B}_k$ is orthogonal to both its neighbors and non-neighbors.

Recalling (19), the problem of maximizing $\mathcal{C}^{sum}$ is given by

$$\mathcal{P}_1: \max_{\mathbf{B}^H\mathbf{B}=\mathbf{I}} \mathcal{C}^{sum}\left(\mathbf{H}, \mathbf{B}\right) \tag{21}$$

$$s.t.\ span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right), \quad k = 1, 2, \cdots, K. \tag{21a}$$

We first give a lemma that is useful in solving $\mathcal{P}_1$.

Lemma 2: Denote the singular value decomposition (SVD) of a Hermitian matrix $\mathbf{F}$ by $\mathbf{F} = \left[\mathbf{U}_{\mathbf{F}_1}, \mathbf{U}_{\mathbf{F}_2}\right]\mathbf{\Lambda}_{\mathbf{F}}\left[\mathbf{U}_{\mathbf{F}_1}, \mathbf{U}_{\mathbf{F}_2}\right]^H$, where $\mathbf{U}_{\mathbf{F}_1}$ is composed by the eigenvectors of $\mathbf{F}$ corresponding to the non-zero eigenvalues. For a unitary matrix $\mathbf{Q}$, if $span\left(\mathbf{U}_{\mathbf{F}_1}\right) \subseteq span\left(\mathbf{Q}\right)$, the non-zero singular values of $\mathbf{Q}^H\mathbf{F}\mathbf{Q}$ are the same as that of $\mathbf{F}$.

Proof: See Appendix B. ∎

Based on Lemma 2, we can derive the following theorem.

Theorem 1: The capacity in $\mathcal{P}_1$ can be maximized when

$$\mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{H}\right) \subseteq span\left(\mathbf{B}\right) \tag{22}$$

where $\mathcal{S}_{\mathbf{R}} = \bigcup_{k=1}^{K} span^{\perp}\left(\bar{\mathbf{R}}_k\right)$.

Proof: See Appendix C. ∎

Remark: Theorem 1 gives a necessary and sufficient condition on $\mathbf{B}$ to achieve the maximal capacity. As a result, to maximize the capacity in (17), $\mathbf{B}$ should be designed satisfying (22). However, in the prebeamforming stage, only the statistical CSI is available at the BS. Letting $\mathbf{R} = \sum_{k=1}^{K} \mathbf{R}_i$, $span\left(\mathbf{H}\right)$ can be any subspace of $span\left(\mathbf{R}\right)$. Hence, to satisfy (22) in Theorem 1, $\mathbf{B}$ should be designed as

$$\mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{R}\right) \subseteq span\left(\mathbf{B}\right). \tag{23}$$

It is important to see what capacity our N-JSDM achieves. To exhibit the conclusion, the following assumption is needed, which is exact when the number of antennas approaches infinity [14].

Orthogonal assumption (OA): In one-ring channel, if the channel paths of two users are non-overlapped, the CCMs of the two users are orthogonal to each other.

Theorem 2: Under OA, if each user's NAS is set to be no smaller than 2Δ, the signal space $span\left(\mathbf{B}^H\mathbf{H}\right)$ is equivalent to $span\left(\mathbf{H}\right)$, and our N-JSDM can achieve the same sum capacity of the corresponding MU-MIMO downlink system in (1).

Proof: See Appendix D. ∎

Theorem 2 indicates that, under OA, the N-JSDM can fully utilize the signal space $span\left(\mathbf{H}\right)$ and achieve the same capacity as (1) if NAS is no smaller than 2Δ. Thus, to fully utilize the signal space, the NAS should be no smaller than 2Δ. It is worth noting that the number of antennas in reality is finite, from which the OA is not exact, and our scheme has spectral efficiency loss compared with the system capacity of (1), even if the NAS is no smaller than 2Δ.

Next, we design $\mathbf{B}$ to satisfy (23), such that $\mathcal{C}^{sum}$ in (20) is maximized. Actually, there are many solutions satisfying (23), and we wish to find one with minimum L so as to reduce the DTL from the preceding subsection. Since $L = \max d_k$,

the prebeamformer design problem is thus formulated as

$$\mathcal{P}_2: \min_{\mathbf{B}^H\mathbf{B}=\mathbf{I}} \max_k d_k \tag{24}$$

$$s.t.\ \mathcal{S}_{\mathbf{R}} \cap span(\mathbf{R}) \subseteq span(\mathbf{B}) \tag{24a}$$

$$span(\mathbf{B}_k) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right). \tag{24b}$$

It is difficult to obtain the optimal solution of $\mathcal{P}_2$. Next we give a greedy way to obtain one solution that yields a small L. We design each prebeamforming matrix $\mathbf{B}_k$ iteratively, and the number of the columns of $\mathbf{B}_k$ is minimized in the k-th iteration. Consequently, when the iteration process is finished, L will be small.

Before we give the iteration process, we give a lemma which divides the space $\mathcal{S}_{\mathbf{R}} \cap span(\mathbf{R})$ into K subspaces, such that $\mathcal{P}_2$ can be solved iteratively.

Lemma 3: Denote $\mathcal{S}_k = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span(\mathbf{R}_k), k = 1, 2, \cdots, K$. Under OA, we have

$$\mathcal{S}_{\mathbf{R}} \cap span(\mathbf{R}) = \bigcup_k \mathcal{S}_k. \tag{25}$$

Proof: See Appendix E. ∎

It can be concluded from Lemma 3 that, if we design the prebeamforming matrix $\mathbf{B}_k$ satisfying, in the k-th iteration,

$$\mathcal{S}_k \subseteq span([\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_k]), span(\mathbf{B}_k) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \tag{26}$$

then the constraints in $\mathcal{P}_2$ will be satisfied when the iterations are completed. Moreover, as shown above, the number of $\mathbf{B}_k$'s columns is minimized in the k-th iteration. Denoting by b_k the number of the columns of $\mathbf{B}_k$, the problem in the k-th iteration is given as

$$\mathcal{P}_3: \min_{\mathbf{B}_k} b_k \tag{27}$$

$$s.t.\ \mathcal{S}_k \subseteq \bigcup_{j=1}^{k} span(\mathbf{B}_j) \tag{27a}$$

$$span(\mathbf{B}_k) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \tag{27b}$$

$$\mathbf{G}_k^H \mathbf{G}_k = \mathbf{I} \tag{27c}$$

where $\mathbf{G}_k = [\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_k]$. To guarantee (27a), i.e., $\mathcal{S}_k \subseteq \bigcup_{j=1}^{k} span(\mathbf{B}_j)$, $\mathbf{B}_k$ should be designed satisfying $\mathcal{S}_k - \bigcup_{j=1}^{k-1} span(\mathbf{B}_j) \subseteq span(\mathbf{B}_k)$, where $\mathcal{S}_1 - \mathcal{S}_2$ means subtracting the common subspace of $\mathcal{S}_1$ and $\mathcal{S}_2$ from $\mathcal{S}_1$. Therefore, $\mathcal{P}_3$ is equivalent to

$$\mathcal{P}_4: \min_{\mathbf{B}_k} b_k \tag{28}$$

$$s.t.\ \mathcal{S}_k - \bigcup_{j=1}^{k-1} span(\mathbf{B}_j) \subseteq span(\mathbf{B}_k) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \tag{28a}$$

$$\mathbf{G}_k^H \mathbf{G}_k = \mathbf{I}. \tag{28b}$$

It is easy to verify that, subject to (28a), the matrix $\mathbf{B}_k$ that achieves the smallest b_k is the orthogonal basis of the space $\mathcal{S}_k - \bigcup_{j=1}^{k-1} span(\mathbf{B}_j)$, i.e.,

$$span(\mathbf{B}_k) = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span(\mathbf{R}_k) - \bigcup_{j=1}^{k-1} span(\mathbf{B}_j) \tag{29}$$

from which $\mathbf{B}_k$ is orthogonal to $\mathbf{B}_j (j < k)$ and $\mathbf{G}_k^H \mathbf{G}_k = \mathbf{I}$. Accordingly, (28b) is also satisfied, and the optimal solution of $\mathcal{P}_4$ is the matrix $\mathbf{B}_k$ in (29).

In the following, we give the process to obtain $\mathbf{B}_k$ in (29). It follows from (29) that $span(\mathbf{B}_k) \subseteq span(\mathbf{R}_k)$ in the k-th iteration. Similarly, in the l-th$(l \leq k)$ iteration, we also have $span(\mathbf{B}_l) \subseteq span(\mathbf{R}_l)$. If user l is not user k's neighbor, $span(\mathbf{R}_l) \subseteq span\left(\bar{\mathbf{R}}_k\right)$ when the number of antennas tends to infinity. Thus, $span(\mathbf{B}_l) \subseteq span(\mathbf{R}_l) \subseteq span\left(\bar{\mathbf{R}}_k\right)$ yielding $span(\mathbf{B}_l)$ is not in the space $span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span(\mathbf{R}_k)$. As a consequence, (29) is equivalent to

$$\begin{aligned} span(\mathbf{B}_k) &= span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span(\mathbf{R}_k) - \bigcup_{j \in \Omega_k, j<k} span(\mathbf{B}_j) \\ &= span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span(\mathbf{R}_k) \cap span^{\perp}(\mathbf{B}_{\Phi_k}) \end{aligned} \tag{30}$$

where $\Phi_k = \{j | j \in \Omega_k, j < k\} = \{k_l, k_l + 1, \cdots, k-1\}$ (i.e., the set of user k's neighbors whose indexes are smaller than k), and $\mathbf{B}_{\Phi_k} = [\mathbf{B}_{k_l}, \mathbf{B}_{k_l+1}, \cdots, \mathbf{B}_{k-1}]$. From (30), $span(\mathbf{B}_k)$ is orthogonal to both $span(\mathbf{B}_{\Phi_k})$ and $span\left(\bar{\mathbf{R}}_k\right)$. Thus $span(\mathbf{B}_k)$ is orthogonal to $span(\mathbf{B}_{\Phi_k})$, and $\mathbf{B}_k$ can be represented as

$$\mathbf{B}_k = \bar{\mathbf{B}}_{\Phi_k} \mathbf{M} \tag{31}$$

where $\bar{\mathbf{B}}_{\Phi_k}$ is the orthogonal basis of the space $span^{\perp}(\mathbf{B}_{\Phi_k})$, and $\mathbf{M}$ is an arbitrary unitary matrix whose row number is the same as the column number of $\bar{\mathbf{B}}_{\Phi_k}$. Moreover, $span(\mathbf{B}_k)$ is orthogonal to $span\left(\bar{\mathbf{R}}_k\right)$, leading to

$$\mathbf{B}_k^H \cdot \bar{\mathbf{R}}_k = \mathbf{0} \Rightarrow \mathbf{M}^H \bar{\mathbf{B}}_{\Phi_k}^H \cdot \bar{\mathbf{R}}_k = \mathbf{0}. \tag{32}$$

Denote by $\mathbf{U_P}$ the matrix composed by the eigenvectors of $\bar{\mathbf{B}}_{\Phi_k}^H \cdot \bar{\mathbf{R}}_k$ corresponding to the zero eigenvalues. It is easy to know that the solution of (32) can be represented by $\mathbf{M} = \mathbf{U_P N}$, where $\mathbf{N}$ is an arbitrary unitary matrix.

Remark: Note that the AoD range of user k's non-neighbors does not include $(\theta_k - \tau + \Delta, \theta_k + \tau - \Delta)$. As a result, many of the eigenvalues of $\bar{\mathbf{R}}_k$ are close to zero when the number of antennas approaches infinity. Since $\bar{\mathbf{B}}_{\Phi_k}^H \bar{\mathbf{B}}_{\Phi_k} = \mathbf{I}$, according to Poincare separation theorem [41], it is easy to verify $\bar{\lambda}_{M-m+i} \leq \lambda_i \leq \bar{\lambda}_i$, where m is the number of columns of $\bar{\mathbf{B}}_{\Phi_k}$, λ_i and $\bar{\lambda}_i$ are the i-th largest eigenvalues of $\bar{\mathbf{B}}_{\Phi_k}^H \bar{\mathbf{R}}_k$ and $\bar{\mathbf{R}}_k$, respectively. Thus, many of the eigenvalues of $\bar{\mathbf{B}}_{\Phi_k}^H \bar{\mathbf{R}}_k$ may be close to zero. If $\mathbf{U_P}$ is composed by the eigenvectors corresponding to the exact zero eigenvalues, the dimension of $\mathbf{U_P}$ will be very small, resulting in small dimensions of $\mathbf{M}$ and $\mathbf{B}_k$, which will reduce the available signal space. Thus, we let $\mathbf{U_P}$ be the matrix composed by the eigenvectors of $\bar{\mathbf{B}}_{\Phi_k}^H \bar{\mathbf{R}}_k$ corresponding to the eigenvalues smaller than ε, where ε is a constant. As a consequence, the valid signal space will be larger, as well as the spectral efficiency. From above, the dimension of a orthogonal space is calculated by the eigenvalues and ε, which is different from that in the conventional JSDMs.

Substituting $\mathbf{M} = \mathbf{U_P N}$ into (31), we can get

$$\mathbf{B}_k = \bar{\mathbf{B}}_{\Phi_k} \mathbf{U_P N} \tag{33}$$

and $span(\mathbf{B}_k)$ is orthogonal to both $span(\mathbf{B}_{\Phi_k})$ and $span\left(\bar{\mathbf{R}}_k\right)$. $\bar{\mathbf{B}}_{\Phi_k} \mathbf{U_P}$ in (33) is the orthogonal basis of

Algorithm 1 The Optimal Prebeamformer Design

Data: $\mathbf{R}_k\,(k=1,2,\cdots,K)$, NAS
Result: $\mathbf{B}$
for $k=1:K$ **do**
  Obtain user k's neighbors set $\Omega_k=\{k_l,\cdots,k_u\}$ and non-neighbors set $\bar{\Omega}_k=\mathcal{K}-\Omega_k$;
  Calculate $\mathbf{B}_{\Phi_k}=[\mathbf{B}_{k_l},\cdots,\mathbf{B}_{k-1}]$;
  Calculate the orthogonal basis of $span^{\perp}\left(\mathbf{B}_{\Phi_k}\right)$ as $\bar{\mathbf{B}}_{\Phi_k}$;
  Calculate $\bar{\mathbf{R}}_k=\sum_{i\in\bar{\Omega}_k}\mathbf{R}_i$;
  Calculate the orthogonal basis of $\bar{\mathbf{B}}_{\Phi_k}^H\bar{\mathbf{R}}_k$ as $\mathbf{U}_{\mathbf{P}}$;
  Calculate $\mathbf{Z}_k=\mathbf{B}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H\mathbf{B}_{\Phi_k}^H\mathbf{R}_k$;
  Calculate the dominant eigenvectors of $\mathbf{Z}_k$ as $\mathbf{B}_k$;
end
Procedure End

$span^{\perp}\left(\bar{\mathbf{R}}_k\right)\cap span^{\perp}\left(\mathbf{B}_{\Phi_k}\right)$, leading to the projection matrix of $span^{\perp}\left(\bar{\mathbf{R}}_k\right)\cap span^{\perp}\left(\mathbf{B}_{\Phi_k}\right)$ being $\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H$. Thus, one basis of $span\left(\mathbf{B}_k\right)$ in (30) can be written as $\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H\mathbf{R}_k$, and consequently, we have $span\left(\mathbf{B}_k\right)=span\left(\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H\mathbf{R}_k\right)$, yielding that the orthogonal basis $\mathbf{B}_k$ is the dominant eigenvectors of $\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H\mathbf{R}_k$. Here we define the dominant eigenvectors of one matrix as the eigenvectors corresponding to the values that are larger than v_m/κ, where v_m denotes the maximal eigenvalue, and κ is a constant.

The detail of the optimal N-JSDM is given in Algorithm 1.

Next, we briefly discuss L (the minimal DTL) in the optimal N-JSDM. From Section III, we know that L is the largest one among the numbers of the columns in $\mathbf{B}_{\Omega_k}\,(k=1,2,\cdots,K)$. By (30), $span\left(\mathbf{B}_k\right)\subseteq span\left(\mathbf{R}_k\right)$, leading to

$$span\left(\mathbf{B}_{\Omega_k}\right)=\bigcup_{l\in\Omega_k}span\left(\mathbf{B}_l\right)\subseteq\bigcup_{l\in\Omega_k}span\left(\mathbf{R}_l\right). \quad (34)$$

Thus, the number of columns of $\mathbf{B}_{\Omega_k}$ is smaller than the dimension of $\bigcup_{l\in\Omega_k}span\left(\mathbf{R}_l\right)$, which is affected by the CCMs of user k and its neighbors.

B. Prebeamformer With Constrained DTL

In the previous subsection, an optimal prebeamformer design is proposed to achieve the same capacity as the full CSI system. However, this optimal design has a large DTL. We propose a prebeamformer design with constrained DTL in this subsection, i.e., the size of each user's prebeamforming matrix is constrained. Specifically, the number of columns of $[\mathbf{B}_1,\cdots,\mathbf{B}_k]$ is set to $\lfloor g\cdot k\rfloor$ where g is a constant and $\lfloor\cdot\rfloor$ is the round down operation. Thus the number of columns of $\mathbf{B}_k$ is $\lfloor g\cdot k\rfloor-\lfloor g\cdot(k-1)\rfloor$. The DTL will be small if g is small. Similar to the conventional JSDMs, the parameter g is designed according to the number of users, i.e., $g=\frac{M}{K}$ [28].

Lemma 3 shows that the capacity of our N-JSDM is the same as that of (1) if $span\left(\mathbf{R}\right)\subseteq span\left(\mathbf{B}\right)$. However, with the number of columns of $\mathbf{B}$ constrained, the space $span\left(\mathbf{B}\right)$ cannot cover the space $span\left(\mathbf{R}\right)$. Hence, the spectral efficiency of our N-JSDM with constrained DTL will be decreased compared with the optimal designs. Our goal is to design $span\left(\mathbf{B}\right)$ to approach $span\left(\mathbf{R}\right)$, such that the spectral efficiency of the constrained N-JSDM is maximized.

We use the chordal distance [15] to describe the difference between two spaces. The chordal distance between spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is given by

$$D_{chor}\left(\mathcal{S}_1,\mathcal{S}_2\right)=\left\|\mathbf{U}_{\mathcal{S}_1}\mathbf{U}_{\mathcal{S}_1}^H-\mathbf{U}_{\mathcal{S}_2}\mathbf{U}_{\mathcal{S}_2}^H\right\|_F \quad (35)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and $\mathbf{U}_{\mathcal{S}_1}$ and $\mathbf{U}_{\mathcal{S}_2}$ are the orthogonal basis of $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. It is easy to know that, the smaller the chordal distance is, the closer $\mathcal{S}_1$ and $\mathcal{S}_2$ will be. Specifically, given the dimension of the two spaces (e.g., N_1 and N_2), if the chordal distance between the two spaces is 0, these two spaces are the same. When the chordal distance is N_1+N_2 (the maximal value), these two spaces are orthogonal.

In order to design $span\left(\mathbf{B}\right)$ approaching $span\left(\mathbf{R}\right)$, the chordal distance between $span\left(\mathbf{B}\right)$ and $span\left(\mathbf{R}\right)$ should be minimized. Subject to the constraint that non-neighbors' interference is mitigated, i.e., the condition in (19), the problem of designing $\mathbf{B}$ is given by

$$\mathcal{P}_5:\ \min_{\mathbf{B}^H\mathbf{B}=\mathbf{I}}D_{chor}\left(span\left(\mathbf{R}\right),span\left(\mathbf{B}\right)\right) \quad (36)$$

$$s.t.\ span\left(\mathbf{B}_k\right)\subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right),k=1,2,\cdots,K. \quad (36a)$$

We develop a greedy algorithm to solve $\mathcal{P}_5$. In the algorithm, we first divide the space $span\left(\mathbf{R}\right)$ into K subspaces, i.e., $\bar{\mathcal{S}}_k=span\left(\mathbf{R}_k\right),k=1,2,\cdots,K$. Then we design the prebeamforming matrix $\mathbf{B}_k$ iteratively, such that the chordal distance between $\bigcup_{j=1}^{k}span\left(\mathbf{B}_k\right)$ and $\bar{\mathcal{S}}_k$ is minimized in the k iteration, and thus $D_{chor}\left(span\left(\mathbf{R}\right),span\left(\mathbf{B}\right)\right)$ is small when the iterations complete.

In the k-th iteration, the problem of designing $\mathbf{B}_k$ is given by

$$\mathcal{P}_6:\ \min_{\mathbf{B}_k}D_{chor}\left(span\left(\mathbf{G}_k\right),\bar{\mathcal{S}}_k\right) \quad (37)$$

$$s.t.\ span\left(\mathbf{B}_k\right)\subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \quad (37a)$$

$$\mathbf{G}_k^H\mathbf{G}_k=\mathbf{I} \quad (37b)$$

where $\mathbf{G}_k=[\mathbf{B}_1,\mathbf{B}_2,\cdots,\mathbf{B}_k]$.

Let $\mathbf{U}_{\bar{\mathcal{S}}_k}$ denote the orthogonal basis of $\bar{\mathcal{S}}_k$. Since $\mathbf{G}_k$ is the orthogonal basis of $span\left(\mathbf{G}_k\right)$,

$$D_{chor}\left(span\left(\mathbf{G}_k\right),\bar{\mathcal{S}}_k\right)=\left\|\mathbf{G}_k\mathbf{G}_k^H-\mathbf{U}_{\bar{\mathcal{S}}_k}\mathbf{U}_{\bar{\mathcal{S}}_k}^H\right\|_F. \quad (38)$$

Observing that

$$\begin{aligned}\mathbf{G}_k\mathbf{G}_k^H&=[\mathbf{B}_1,\mathbf{B}_2,\cdots,\mathbf{B}_k]\,[\mathbf{B}_1,\mathbf{B}_2,\cdots,\mathbf{B}_k]^H\\&=\mathbf{G}_{k-1}\mathbf{G}_{k-1}^H+\mathbf{B}_k\mathbf{B}_k^H\end{aligned} \quad (39)$$

and substituting (39) into (38), we have

$$\begin{aligned}&D_{chor}\left(span\left(\mathbf{G}_k\right),\bar{\mathcal{S}}_k\right)\\&\quad=\left\|\mathbf{G}_{k-1}\mathbf{G}_{k-1}^H+\mathbf{B}_k\mathbf{B}_k^H-\mathbf{U}_{\bar{\mathcal{S}}_k}\mathbf{U}_{\bar{\mathcal{S}}_k}^H\right\|_F.\end{aligned} \quad (40)$$

Denoting $\mathbf{T} = \mathbf{G}_{k-1}\mathbf{G}_{k-1}^H - \mathbf{U}_{\bar{\mathcal{S}}_k}\mathbf{U}_{\bar{\mathcal{S}}_k}^H$, we have

$$\begin{aligned} &D_{chor}\left(span\left(\mathbf{G}_k\right), \bar{\mathcal{S}}_k\right) \\ &= \left\|\mathbf{B}_k\mathbf{B}_k^H + \mathbf{T}\right\|_F \\ &= Tr\left(\left(\mathbf{B}_k\mathbf{B}_k^H + \mathbf{T}\right)\left(\mathbf{B}_k\mathbf{B}_k^H + \mathbf{T}\right)^H\right) \\ &= Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{B}_k\mathbf{B}_k^H + 2\mathbf{B}_k\mathbf{B}_k^H\mathbf{T} + \mathbf{T}\mathbf{T}^H\right) \\ &= Tr\left(\mathbf{B}_k\mathbf{B}_k^H + 2\mathbf{B}_k\mathbf{B}_k^H\mathbf{T} + \mathbf{T}\mathbf{T}^H\right) \\ &= Tr\left(\mathbf{B}_k\mathbf{B}_k^H\right) + 2Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\right) + Tr\left(\mathbf{T}\mathbf{T}^H\right). \end{aligned} \quad (41)$$

Moreover, $Tr\left(\mathbf{B}_k\mathbf{B}_k^H\right)$ and $Tr\left(\mathbf{T}\mathbf{T}^H\right)$ are constants. Then $\mathcal{P}_6$ can be re-written as

$$\mathcal{P}_7: \min_{\mathbf{B}_k^H\mathbf{B}_k=\mathbf{I}} Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\right) \quad (42)$$

$$s.t.\ span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \quad (42a)$$

$$\mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}. \quad (42b)$$

By $Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\right) = Tr\left(\mathbf{B}_k^H\mathbf{T}\mathbf{B}_k\right)$, we have

$$Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\right) = Tr\left(\mathbf{B}_k^H\left(\mathbf{G}_{k-1}\mathbf{G}_{k-1}^H - \mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H\right)\mathbf{B}_k\right). \quad (43)$$

Note that $\mathbf{G}_k = [\mathbf{G}_{k-1}, \mathbf{B}_k]$ is a unitary matrix, i.e., $\mathbf{B}_k^H\mathbf{G}_{k-1} = \mathbf{0}$, and then (43) is equivalent to

$$Tr\left(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\right) = -Tr\left(\mathbf{B}_k^H\left(\mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H\right)\mathbf{B}_k\right). \quad (44)$$

Hence, $\mathcal{P}_7$ can be written as

$$\mathcal{P}_8: \max_{\mathbf{B}_k^H\mathbf{B}_k=\mathbf{I}} Tr\left(\mathbf{B}_k^H\left(\mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H\right)\mathbf{B}_k\right) \quad (45)$$

$$s.t.\ span\left(\mathbf{B}_k\right) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right) \quad (45a)$$

$$\mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}. \quad (45b)$$

Considering the objective function in $\mathcal{P}_8$, it is easy to verify that $span\left(\mathbf{B}_k\right) \subseteq span\left(\mathbf{U}_{\mathcal{S}_k}\right)$, since the orthogonal space of $span\left(\mathbf{U}_{\mathcal{S}_k}\right)$ has no effect on the trace. As a result, $span\left(\mathbf{B}_k\right) \subseteq span\left(\mathbf{U}_{\mathcal{S}_k}\right) \subseteq span\left(\mathbf{R}_k\right)$ in the k-th iteration. Similarly, in the j-th($j < k$) iteration, we have $span\left(\mathbf{B}_j\right) \subseteq span\left(\mathbf{R}_j\right)$. If user $j(j < k)$ is not user k's neighbor, $span\left(\mathbf{B}_j\right) \subseteq span\left(\mathbf{R}_j\right) \subseteq span\left(\bar{\mathbf{R}}_k\right)$ when the number of antennas approaches infinity, from which $span\left(\mathbf{B}_k\right)$ is orthogonal to $span\left(\mathbf{B}_j\right)$. Thus to make $\mathbf{G}_k$ a unitary matrix (the second constraint in $\mathcal{P}_8$), $span\left(\mathbf{B}_k\right)$ should be orthogonal to $span\left(\mathbf{B}_{\Phi_k}\right)$, where $\mathbf{B}_{\Phi_k} = [\mathbf{B}_{k_l}, \mathbf{B}_{k_l+1}, \cdots, \mathbf{B}_{k-1}]$ as defined in the previous subsection. Moreover, from the first constraint in $\mathcal{P}_8$, $span\left(\mathbf{B}_k\right)$ is orthogonal to $span^{\perp}\left(\bar{\mathbf{R}}_k\right)$. From the above discussion, the constraints in $\mathcal{P}_8$ is equivalent to that $span\left(\mathbf{B}_k\right)$ should be orthogonal to both $span^{\perp}\left(\bar{\mathbf{R}}_k\right)$ and $span\left(\mathbf{B}_{\Phi_k}\right)$. This constraint is the same as that in (30), leading to

$$\mathbf{B}_k = \bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{N} \quad (46)$$

from (33), where $\mathbf{U}_{\mathbf{P}}$ is the matrix composed by the eigenvectors of $\bar{\mathbf{B}}_{\Phi_k}^H\bar{\mathbf{R}}_k$ corresponding to the eigenvalues smaller than ε. $\mathbf{N}$ is an arbitrary unitary matrix having $\lfloor g\cdot k\rfloor - \lfloor g\cdot(k-1)\rfloor$ columns, since the number of columns of $\mathbf{N}$ is the same as that of $\mathbf{B}_k$. As a consequence, $\mathcal{P}_8$ can be equivalently transferred to the following problem,

$$\max_{\mathbf{N}^H\mathbf{N}=\mathbf{I}} Tr\left(\mathbf{N}^H\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H\left(\mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H\right)\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{N}\right). \quad (47)$$

It is easy to know that the solution of $\mathbf{N}$ in (47) is the matrix composed by the eigenvectors of $\mathbf{U}_{\mathbf{P}}^H\bar{\mathbf{B}}_{\Phi_k}^H\left(\mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H\right)\bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}$ corresponding to the $\lfloor g\cdot k\rfloor - \lfloor g\cdot(k-1)\rfloor$ largest eigenvalues. Once we get $\mathbf{N}$, we can use (46) to get $\mathbf{B}_k$.

The detail of achieving $\mathbf{B}$ is given in Algorithm 2.

Algorithm 2 The Constrained Prebeamformer Design

Data: $\mathbf{R}_k\ (k = 1, 2, \cdots, K)$, NAS
Result: $\mathbf{B}$
for $k = 1 : K$ **do**
  Obtain user k's neighbors set $\Omega_k = \{k_l, \cdots, k_u\}$ and non-neighbors set $\bar{\Omega}_k = \mathcal{K} - \Omega_k$;
  Calculate $\mathbf{B}_{\Phi_k} = [\mathbf{B}_{k_l}, \cdots, \mathbf{B}_{k-1}]$;
  Calculate the orthogonal basis of $span^{\perp}\left(\mathbf{B}_{\Phi_k}\right)$ as $\bar{\mathbf{B}}_{\Phi_k}$;
  Calculate the dominant eigenvectors of $\mathbf{R}_k$ as $\mathbf{U}_{\mathcal{S}_k}$;
  Calculate $\bar{\mathbf{R}}_k = \sum_{i\in\bar{\Omega}_k}\mathbf{R}_i$;
  Calculate the orthogonal basis of $\bar{\mathbf{B}}_{\Phi_k}^H\bar{\mathbf{R}}_k$ as $\mathbf{U}_{\mathbf{P}}$;
  Use (47) to get the matrix $\mathbf{N}$, and $\mathbf{B}_k = \bar{\mathbf{B}}_{\Phi_k}\mathbf{U}_{\mathbf{P}}\mathbf{N}$;
end
Procedure End

C. The DFT Prebeamformer in Large Antenna Systems

In this section, we focus on the system where the number of antennas approaches infinity. When the number of antennas becomes infinity, the eigenvectors of the CCMs can be approximated by a subset of the columns of the DFT matrix [15], [37], where the (m, n)-th element of the DFT matrix $\mathbf{F} \in \mathbb{C}^{M\times M}$ is given as

$$[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{M}}e^{\frac{-\iota 2\pi}{M}(m-1)(n-1)}. \quad (48)$$

We write $\mathbf{F} = \left[\mathbf{f}_{-\frac{M}{2}+1}, \mathbf{f}_{-\frac{M}{2}+2}, \cdots, \mathbf{f}_{-1}, \mathbf{f}_0, \mathbf{f}_1, \cdots, \mathbf{f}_{\frac{M}{2}}\right]$. Denoting the azimuth center angle and the AS of user k by θ_k and Δ, respectively, the eigenvectors of $\mathbf{R}_k$ is given by

$$\mathbf{U}_{\mathbf{R}_k} = \left[\mathbf{f}_{\Theta_k}\right], \Theta_k = \left\{\bar{k}_l, \bar{k}_l + 1\cdots, \bar{k}_u\right\} \quad (49)$$

where $\bar{k}_l = \left\lfloor\frac{MD\sin(\theta_k-\Delta)}{2}\right\rfloor$, $\bar{k}_u = \left\lceil\frac{MD\sin(\theta_k+\Delta)}{2}\right\rceil$, and $[\mathbf{f}_{\Theta_k}]$ denotes the matrix whose columns are $\mathbf{f}_i, i \in \Theta_k$ with the index of $\mathbf{f}_i$ in an ascending order.

We use the DFT matrix to form the prebeamforming matrix in our N-JSDM, which is similar to the method in subsection IV. A. First, we divide $\mathcal{S}_{total} = \mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{A}\right)$ into K subspaces, with the k-th subspace

$$\mathcal{S}_k = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_k\right). \quad (50)$$

It is clear that $\mathcal{S}_{total} = \bigcup_{k=1}^{K}\mathcal{S}_k$. From above, the eigenvectors of $\mathbf{R}_k$ are $[\mathbf{f}_{\Theta_k}]$, yielding that the orthogonal basis of

$span\left(\bar{\mathbf{R}}_k\right)$ is $\mathbf{U}_{\bar{\mathbf{R}}_k} = \left[\mathbf{f}_{\bigcup_{j\in\bar{\Omega}_k}\Theta_j}\right]$. Consequently, the orthogonal basis of $\mathcal{S}_k$ is

$$\mathbf{U}_{\mathcal{S}_k} = \left[\mathbf{f}_{i\in\bar{\Theta}_k}\right], \quad \bar{\Theta}_k = \Theta_k - \bigcup_{j\in\bar{\Omega}_k}\Theta_j. \tag{51}$$

From section IV. A, we know that the prebeamforming matrix $\mathbf{B}_{\mathrm{DFT},k}$ for user k is designed satisfying $span\left(\mathbf{B}_{\mathrm{DFT},k}\right) = \mathcal{S}_k - \bigcup_{j<k}\mathcal{S}_j$. It follows that the prebeamforming matrix $\mathbf{B}_{\mathrm{DFT}} = [\mathbf{B}_{\mathrm{DFT},1}, \mathbf{B}_{\mathrm{DFT},2}, \cdots, \mathbf{B}_{\mathrm{DFT},K}]$ is

$$\mathbf{B}_{\mathrm{DFT}} = \left[\mathbf{f}_{\bigcup_{k=1}^{K}\bar{\Theta}_k}\right]. \tag{52}$$

With the prebeamforming matrix designed as (52), the effective channel matrix becomes a band matrix, and also the capacity can be maximized. Moreover, the number of non-zero entries of the effective channel vector of user k is the number of columns of $\mathbf{B}_{\Omega_k}$, which is the cardinal number of set $\bigcup_{k\in\Omega_k}\bar{\Theta}_k$, and thus the DTL of the DFT scheme is $\max_k\bigcup_{k\in\Omega_k}\bar{\Theta}_k$.

Remark: From Theorem 2, with the NAS τ no smaller than 2Δ, the capacity of our N-JSDM is the same as that of (1). In this case, the paths of user k are non-overlapped with the paths of user k's non-neighbors, yielding

$$\bar{\Theta}_k = \Theta_k - \bigcup_{j\in\bar{\Omega}_k}\Theta_j = \Theta_k. \tag{53}$$

Combining (52) with (53), we have

$$\mathbf{B}_{\mathrm{DFT}} = \left[\mathbf{f}_{\bigcup_{k=1}^{K}\Theta_k}\right] \tag{54}$$

and the DTL is $\max_k\bigcup_{j\in\Omega_k}\Theta_j$. In the conventional JSDMs, each group has overlapped signal space with its adjacent groups. To mitigate the inter-group interference, the overlapped signal space is not available in transmission. Hence, the spectral efficiency of the conventional JSDMs will be reduced compared with the capacity in (1). However, our scheme utilizes all signal space to transmit information, which outperforms the conventional JSDMs. Moreover, with proper setting of NAS, our N-JSDM can obtain the same capacity as that of the system in (1) when the number of the antennas becomes infinity.

Up to this point, we have discussed three prebeamformer designs, i.e., the optimal, the constrained and the DFT prebeamformer designs. The DFT prebeamformer has the lowest computational complexity. Moreover, the DFT prebemaformer also has the lowest hardware cost, since it can be implemented by the phase shifters due to the DFT based prebeamforming matrix. However, the number of antennas in reality is not infinite, and the inter-user interference from the non-neighbors cannot be well canceled by the DFT prebeamformer, leading to a lower spectral efficiency than the optimal and constrained prebeamformers. Comparing with the optimal and the constrained prebeamformers, the optimal one can provide a larger spectral efficiency than the constrained one, but at the cost of a larger DTL. When the ratio between the DTL and the number of symbols T_c in a coherence block is small (that is, in the case of small NAS or large T_c), the affection of the DTL is not obvious, and then the optimal prebeamfomer will provide a higher effective spectral efficiency. Otherwise, the DTL has an important affection, and a lower DTL in the constrained prebeamfomer will yield a higher effective spectral efficiency. Moreover, the constrained prebeamformer design has a smaller feedback length than the optimal design.

V. Simulation Results

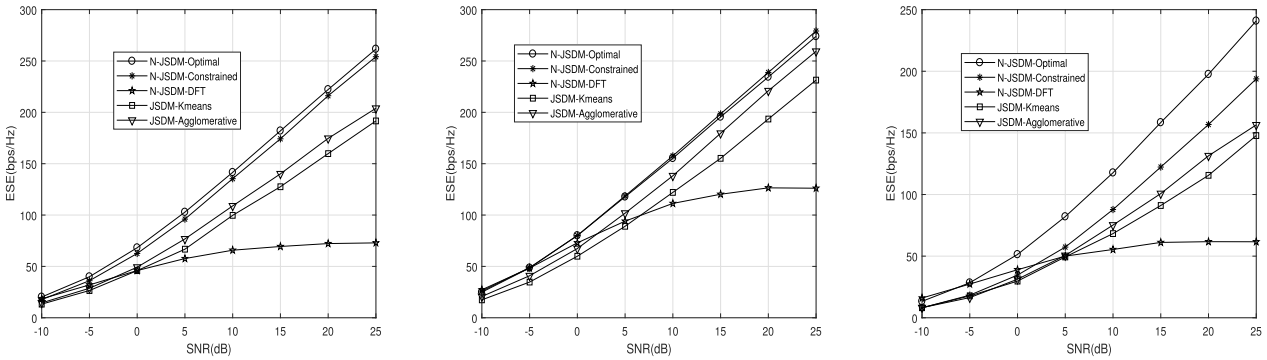
In this section, we demonstrate the efficiency of our proposed N-JSDM schemes through simulation results. For all simulations, we assume a single-cell massive MIMO system where the BS is equipped with M antennas serving K single-antenna users. In this paper, M is set to 64. The azimuth center angles of all users are uniformly distributed in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$. The threshold of eigenvalues of orthogonal basis $\mathbf{U_P}$ defined in Section IV. A, i.e., ε, is set to 10^{-2}. κ is set to 10. We compare our optimal, constrained, and DFT-based schemes with the conventional JSDMs with K-means [15] and agglomerative [28] groupings, and the full precoding scheme [7]. The full precoding scheme is not efficient due to the requirements of large DTL and uplink feedback. To be fair, the mean angular spread of all groups in the conventional JSDMs is set to be equal to the NAS, i.e., τ, and thus the number of groups is $\left\lfloor\frac{\pi}{3\tau}\right\rceil$, where $\lfloor\cdot\rceil$ denotes the round operation. In the conventional JSDMs, the dimension of the effective channel seen by the g-th group is $\lfloor K_g M/K\rfloor$ [28], where K_g is the number of users in the g-th group.

Considering the pilot overhead, the effective spectral efficiency (ESE) of user k in a coherence block with T_c symbols is given by

$$R_k = \left(1 - \frac{DTL_k}{T_c}\right)\log\left(1 + SINR_k\right) \tag{55}$$

where DTL_k denotes the DTL of user k. In LTE systems, a resource block consists of 168 complex symbols [20], [42], that is, 14 OFDM symbols in two time slots multiplied by 12 subcarriers. Hence, similar to [20], T_c is set to 100. In the conventional JSDMs, each group estimates the effective channel matrix separately, and the minimum DTLs for users in the same group are the same. As shown in Section III, the minimum DTLs of all users in our N-JSDMs are $max_k d_k$. To better exhibit the ESE, we consider the case where the pilot has the smallest DTL, i.e., the DTL equals to the minimum DTL.

Fig. 1 shows the ESEs of all JSDMs under different SNRs. The AS is set to 5°. The numbers of users in Figs. 1(a) and 1(b) are 32, while the number of users served in Fig. 1(c) is 40. The NASs in Figs. 1(a) and 1(c) are 10°, and the NAS in Fig. 1(b) is 15°. The ESEs of all the precoding schemes increase with the increasing SNR. More importantly, both our optimal N-JSDM and constrained N-JSDM can achieve higher ESEs than the conventional JSDMs. Also, the DFT-based N-JSDM achieves good performance at low SNR regions, while has worse spectral efficiency performance than other JSDMs when the SNR becomes large. This is because, the DFT algorithm enlarges the energy of both the received signal and the interference, and the inter-user interference of the DFT algorithm is proportional to the transmission power. When the transmission power is at low level, the noise is dominant over the inter-user interference,



(a) K = 32, Δ = 5°, NAS= 10°. (b) K = 32, Δ = 5°, NAS= 15°. (c) K = 40, Δ = 5°, NAS= 10°.

Fig. 1. Comparison of the ESEs under different SNRs.

and due to the large received power of signal, the DFT algorithm achieves a large SINR, which results in a high ESE. When the transmission power becomes large, the inter-user interference also becomes large, and thus the ESE of DFT based N-JSDM will not increase with the increasing SNR. It can be seen from Figs. 1(a) and 1(b) that the optimal N-JSDM provides a higher ESE than the constrained N-JSDM when the NAS is small, while provides a lower ESE than the constrained N-JSDM when the NAS increases. This is because the DTL is low in small NAS, and the affection of the DTL on the ESE is not obvious, and then the ESE is mainly affected by the spectral efficiency. Thus the optimal N-JSDM can provide a higher ESE. However, when the NAS increases, the affection of the DTL becomes important, and consequently, the constrained prebeamfomer can provide a higher ESE due to a lower DTL. We show the relationship between ESE and SNR under different numbers of users in Figs. 1(a) and 1(c), respectively. When the number of users increases, the inter-user interference is increased, and more energy of signal will be used to cancel the inter-user interference. Thus, the energy of desired signals will be degraded, leading to a decreased ESE. Since the DFT-based N-JSDM is mainly designed for the case that the number of antennas approaches infinity, the performance of the DFT-based N-JSDM is not shown in the following.

Fig. 2 shows the ESEs of our proposed N-JSDMs under different NASs, where the number of users is 32, the AS is set to 5°, and the SNR is 20 dB. The N-JSDMs achieve higher ESEs than the conventional JSDMs, while the full precoding with perfect CSI achieves the smallest ESE. The ESE of the full digital precoding remains unchanged, since the full digital precoding is unrelated to the NAS. It is shown that the ESEs of all JSDMs first increase and then decrease with the increasing NAS. With the increasing of NAS, the utilized signal space in our N-JSDMs increases, yielding that the spectral efficiency increases. However, the DTL will also be increased. For the conventional JSDMs, with the increasing of the NAS, the number of groups decreases, and the overlapped signal space between different groups decreases. Thus, both the utilized signal space and the spectral efficiency increase with the increasing NAS. Similar to the N-JSDMs, the DTLs of the conventional JSDMs will be increased. When the NAS

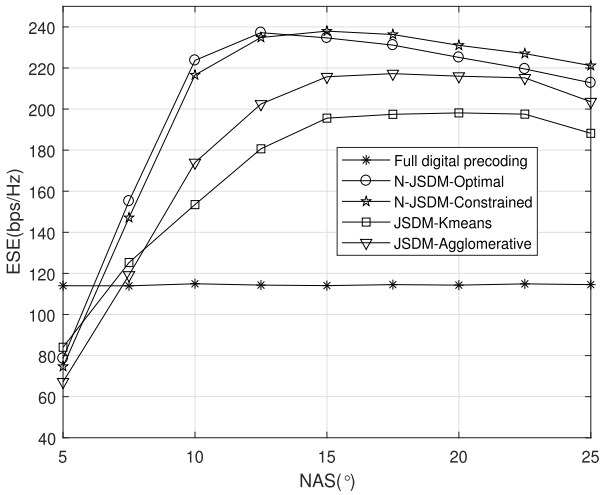


Fig. 2. Comparison of the ESEs under different NASs. $K = 32$, $\Delta = 5^\circ$ and SNR $=$ 20dB.

is small, the ESEs of JSDMs are mainly affected by the spectral efficiency as shown above, and with the increasing NAS, the spectral efficiencies will be increased. Thus, when the NAS is small, the ESEs of all JSDMs increase with the increasing NAS. However, when the NAS is large, the utilized signal space will not increase a lot by increasing the NAS, and then the spectral efficiencies will not be increased obviously. Moreover, increasing the NAS will increase the DTL. Consequently, the ESEs of all JSDMs decrease with the increasing NAS. Similar to Fig. 1, the optimal N-JSDM provides a higher ESE than the constrained N-JSDM when the NAS is small, while provides a lower ESE than the constrained N-JSDM when the NAS becomes large.

Fig. 3 depicts the ESEs of our proposed N-JSDMs under different numbers of users. The AS of each user is set to 5°, the NAS is set to 10°, and the SNR is 20 dB. We can observe some special phenomenons as follows. With the increasing number of users, the ESEs of all the schemes first increase, and then decrease when the number of users is large. This is because, when the number of users becomes large, the inter-user interference will be large, and we should uses a part of

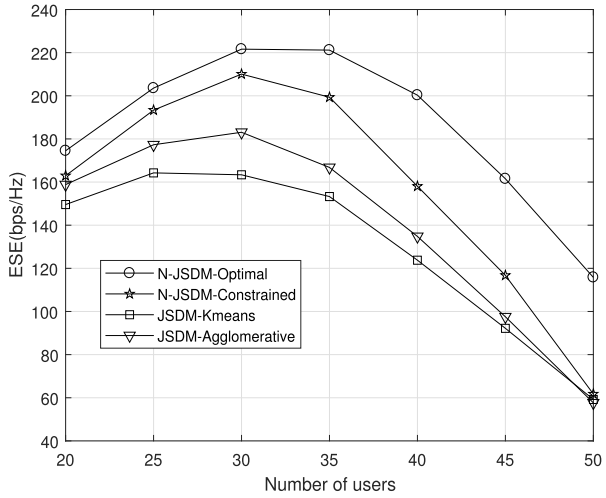


Fig. 3. Comparison of the ESEs under different numbers of users. $\Delta = 5^\circ$, NAS $= 10^\circ$ and SNR $=$ 20dB.

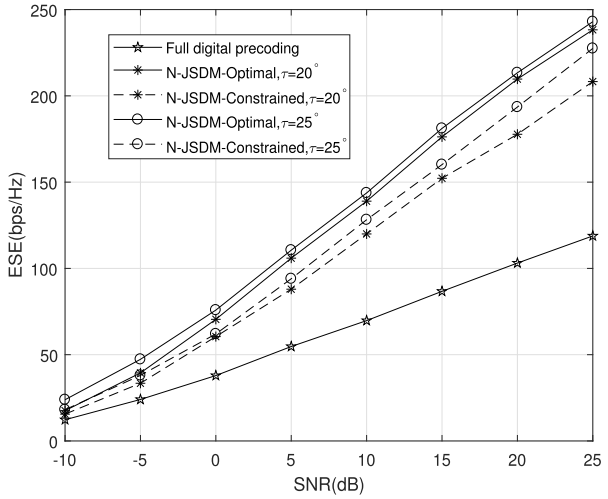
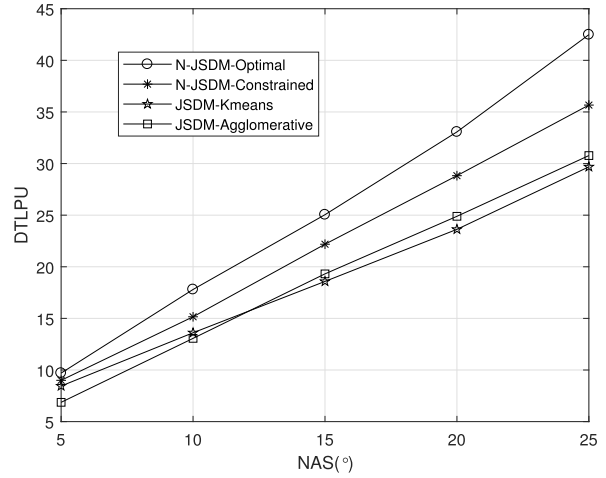
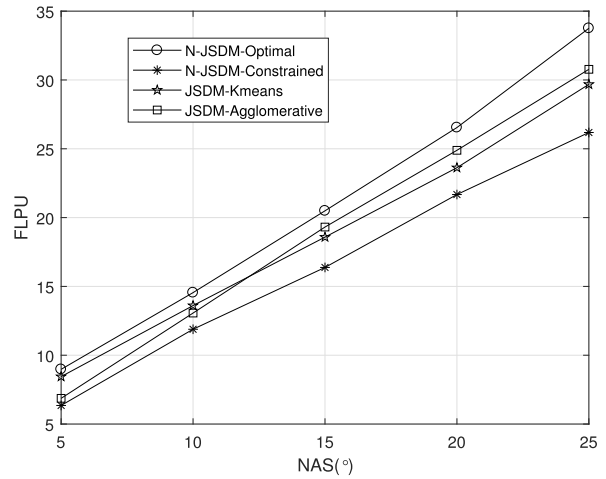


Fig. 5. Comparison of the ESEs under different SNRs. $K = 32$ and $\Delta = 5^\circ$.



(a) DTLPU versus NAS.



(b) FLPU versus NAS.

Fig. 4. Comparisons of the DTL and the feedback length. $K = 32$ and $\Delta = 5^\circ$.

degrees of freedom to cancel the inter-user interference, and thus, the energy of desired signals will be degraded.

Fig. 4 shows the DTL and the channel feedback length of our N-JSDM. Fig. 4(a) shows the DTL per user (DTLPU) of our N-JSDMs and the conventional JSDMs versus the NAS, where the DTLPU in the conventional JSDMs is $\frac{\sum_{k=1}^{K} \bar{d}_k}{K}$, and $\bar{d}_k$ is the DTL of user k. The DTLPUs of all schemes increase with the increasing NAS. The DTLs of our N-JSDMs are larger than that of JSDMs, and the optimal N-JSDM has the largest DTL. Fig. 4(b) shows the curves of feedback length per user (FLPU) versus NAS. In the N-JSDMs, the feedback length of user k is d_k, and the FLPU is given by $\frac{\sum_{k=1}^{K} d_k}{K}$, which is smaller than the DTL of the N-JSDMs, i.e., $max_k d_k$. The FLPU in the conventional JSDMs is $\frac{\sum_{k=1}^{K} \bar{d}_k}{K}$, which is the same as the DTL of the conventional JSDMs. The optimal N-JSDM has a larger FLPU than the JSDMs. Our constrained N-JSDM has the smallest FLPU. This is because, in the conventional JSDMs, some groups will have large numbers of users, and these unbalanced user numbers will enlarge the FLPU.

The above simulations have validated the good performance of our N-JSDMs in the one-ring channel model. In fact, our N-JSDM is also suitable for other channel models such as the Laplacian channel model. In the following, we give the simulation of our N-JSDMs in the Laplacian channel model [20], [43]. Supposing the azimuth center angle of user k is θ_k, the (m, n)th element of user k's CCM $\mathbf{R}_k$ in the Laplacian channel model is [20]

$$[\mathbf{R}_k]_{m,n} = \frac{1}{\sqrt{2}\bar{\Delta}} \int_{\theta_k - \pi}^{\theta_k + \pi} e^{-\frac{\sqrt{2}}{\bar{\Delta}}|\theta - \theta_k| - \iota 2\pi \frac{D}{\lambda_c}(m-n)\sin\theta} d\theta \tag{56}$$

where λ_c and D has been defined in Section II, and $\bar{\Delta}$ is the AS in the Laplacian channel model. Fig. 5 shows the ESEs of our N-JSDMs and the full digital precoding versus the SNR. The AS $\bar{\Delta}$ is set to 5°. The number of users is 32, and the NASs are set to 20° and 25°. Our N-JSDMs can achieve much higher ESEs than the full precoding. The optimal N-JSDM has a higher ESE than the constrained N-JSDM. It can be seen that the optimal and constrained N-JSDMs with NAS 25° have higher ESEs than that with NAS 20°, respectively. This is because, with the same AS in one-ring and Laplacian channel models (i.e., $\bar{\Delta} = \Delta$), the dominant energy of one user's signal

in the Laplacian model lies in a larger AoD range than that in the one-ring channel model. As a result, when designing the prebeamformer to cancel the interference from non-neighbors, a larger NAS can help to increase the utilized signal space and the ESE.

VI. Conclusion

In this paper, we have proposed an N-JSDM beamforming in FDD massive MIMO, which provides a higher spectral efficiency than the conventional JSDMs. Moreover, the channel feedback length and the DTL of the proposed N-JSDM are much smaller than the number of antennas, due to the resulting banded effective channel matrix. Three efficient prebeamformer designs have been given to obtain the largest system capacity, to reduce the DTL, and to simplify the system when the number of antenna tends to infinity, respectively. Simulations have been provided to validate the good performance of our proposed N-JSDM. In our future work, we will provide efficient user scheduling algorithms to further improve the spectral efficiency when the number of users is large.

Appendix

A. Proof of Lemma 1

Define $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N]$. It is easy to see that

$$\bigcap_i span^{\perp}(\mathbf{A}_i) = span^{\perp}(\mathbf{A}) \tag{57}$$

i.e., $\mathcal{S} = span^{\perp}(\mathbf{A})$. Denote the SVD of $\mathbf{A}$ as $\mathbf{A} = [\mathbf{U}_1, \mathbf{U}_2]\,\mathbf{\Lambda}\,[\mathbf{V}_1, \mathbf{V}_2]^H$, where $\mathbf{U}_1$ and $\mathbf{U}_2$ are the matrices composed by the left singular vectors corresponding to the non-zero and zero singular values, respectively. It follows from [44] and [45] that $\mathbf{S}^{\perp} = span(\mathbf{U}_1)$, and

$$\mathbf{S} = span(\mathbf{U}_2). \tag{58}$$

Using $\mathbf{A} = [\mathbf{U}_1, \mathbf{U}_2]\,\mathbf{\Lambda}\,[\mathbf{V}_1, \mathbf{V}_2]^H$, the SVD of $\sum_{i=1}^{N} \mathbf{A}_i\mathbf{A}_i^H$ is given by

$$\sum_{i=1}^{N} \mathbf{A}_i\mathbf{A}_i^H = \mathbf{A}\mathbf{A}^H = [\mathbf{U}_1, \mathbf{U}_2]\,\mathbf{\Lambda}^2\,[\mathbf{U}_1, \mathbf{U}_2]^H \tag{59}$$

leading to

$$span^{\perp}\left(\sum_{i=1}^{N} \mathbf{A}_i\mathbf{A}_i^H\right) = span(\mathbf{U}_2). \tag{60}$$

From (58) and (60), we can get Lemma 1.

B. Proof of Lemma 2

Since $\mathbf{U}_{\mathbf{F}_1}$ is composed by the eigenvectors of $\mathbf{F}$ corresponding to the non-zero eigenvalues, we have $\mathbf{U}_{\mathbf{F}_1}^H\mathbf{U}_{\mathbf{F}_1} = \mathbf{I}$. We write $\mathbf{U}_{\mathbf{F}_1} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r)$, and by $span(\mathbf{U}_{\mathbf{F}_1}) \subset span(\mathbf{Q})$, we can find a unit norm vector $\mathbf{p}_i$ such that $\mathbf{u}_j = \mathbf{Q}\mathbf{p}_j$, leading to $\mathbf{U}_{\mathbf{F}_1} = \mathbf{Q}\mathbf{P}$, where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_r]$. Thus, we have

$$\mathbf{P}^H\mathbf{P} = (\mathbf{Q}\mathbf{P})^H\,\mathbf{Q}\mathbf{P} = \mathbf{U}_{\mathbf{F}_1}^H\mathbf{U}_{\mathbf{F}_1} = \mathbf{I}. \tag{61}$$

Denote by $\mathbf{\Lambda}_{\mathbf{F}_1}$ the diagonal matrix whose diagonal values are composed by the nonzero singular values of $\mathbf{F}$. It is easy to see that $\mathbf{F}$ can be written as $\mathbf{F} = \mathbf{U}_{\mathbf{F}_1}\mathbf{\Lambda}_{\mathbf{F}_1}\mathbf{U}_{\mathbf{F}_1}^H$, and we have

$$\mathbf{Q}^H\mathbf{F}\mathbf{Q} = \mathbf{Q}^H\mathbf{U}_{\mathbf{F}_1}\mathbf{\Lambda}_{\mathbf{F}_1}\mathbf{U}_{\mathbf{F}_1}^H\mathbf{Q}. \tag{62}$$

From $\mathbf{U}_{\mathbf{F}_1} = \mathbf{Q}\mathbf{P}$ as shown above, we know $\mathbf{Q}^H\mathbf{U}_{\mathbf{F}_1} = \mathbf{P}$. Substituting $\mathbf{Q}^H\mathbf{U}_{\mathbf{F}_1} = \mathbf{P}$ into (62), we can get

$$\mathbf{Q}^H\mathbf{F}\mathbf{Q} = \mathbf{P}\mathbf{\Lambda}_{\mathbf{F}_1}\mathbf{P}^H. \tag{63}$$

Using (61), $\mathbf{P}^H\mathbf{P} = \mathbf{I}$, and according to (63), we can easily obtain that the singular values of the matrix $\mathbf{Q}^H\mathbf{F}\mathbf{Q}$ are the same as that of $\mathbf{F}$.

C. Proof of Theorem 1

Since $\mathbf{B}_k$ is an arbitrary unitary matrix satisfying $span(\mathbf{B}_k) \subseteq span^{\perp}(\bar{\mathbf{R}}_k)$, it follows from $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2 \cdots, \mathbf{B}_K]$ that $span(\mathbf{B}) \subseteq \mathcal{S}_{\mathbf{R}}$ where $\mathcal{S}_{\mathbf{R}} = \bigcup_{K=1}^{K} span^{\perp}(\bar{\mathbf{R}}_k)$. Actually, the prebeamforming matrix $\mathbf{B}$ can be any unitary matrix satisfying $span(\mathbf{B}) \subseteq \mathcal{S}_{\mathbf{R}}$. Hence, the constraints in $\mathcal{P}_1$ are equivalent to $span(\mathbf{B}) \subseteq \mathcal{S}_{\mathbf{R}}$, and consequently, the problem $\mathcal{P}_1$ can be represented as

$$\max_{\mathbf{B}^H\mathbf{B}=\mathbf{I}} \max_{\mathbf{S}\succeq 0, Tr(\mathbf{S})\le P} \log\left|\mathbf{I} + \mathbf{B}^H\mathbf{H}\mathbf{S}\mathbf{H}^H\mathbf{B}\right| \tag{64}$$

$$s.t.\ span(\mathbf{B}) \subseteq \mathcal{S}_{\mathbf{R}}. \tag{64a}$$

Denote the orthogonal basis of the space $\mathcal{S}_{\mathbf{R}}$ by $\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}$, and from $span(\mathbf{B}) \subseteq \mathcal{S}_{\mathbf{R}}$, the unitary matrix $\mathbf{B}$ can be represented as $\mathbf{B} = \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\mathbf{M}$, where $\mathbf{M}$ is an arbitrary unitary matrix. Substituting $\mathbf{B} = \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\mathbf{M}$ into (64), we have

$$\max_{\mathbf{M}^H\mathbf{M}=\mathbf{I}} \log\left|\mathbf{I}_M + \mathbf{M}^H\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\mathbf{M}\right| \tag{65}$$

where $\mathbf{F} = \mathbf{H}\mathbf{S}\mathbf{H}^H$. Note that the projection matrix of $span(\mathbf{U}_{\mathcal{S}_{\mathbf{R}}})$ is $\mathbf{P} = \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H$, and we write matrix $\mathbf{F}$ as $\mathbf{F} = \mathbf{P}\mathbf{F} + (\mathbf{I}-\mathbf{P})\mathbf{F}$. Then we have

$$\begin{aligned}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} &= \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H(\mathbf{P}\mathbf{F} + (\mathbf{I}-\mathbf{P})\mathbf{F})\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} \\ &= \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{P}\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} + \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H(\mathbf{I}-\mathbf{P})\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\end{aligned} \tag{66}$$

and

$$\begin{aligned}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H(\mathbf{I}-\mathbf{P})\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} &= \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} - \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} \\ &= \mathbf{0}\end{aligned} \tag{67}$$

from which $\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}} = \mathbf{U}_{\mathcal{S}_{\mathbf{R}}}^H\mathbf{P}\mathbf{F}\mathbf{U}_{\mathcal{S}_{\mathbf{R}}}$ follows. Thus, (65) can be written as

$$\max_{\mathbf{M}^H\mathbf{M}=\mathbf{I}} \log\left|\mathbf{I}_M + \mathbf{B}^H\mathbf{P}\mathbf{F}\mathbf{B}\right|. \tag{68}$$

Since $span(\mathbf{P}\mathbf{F})$ is equivalent to the space $\mathcal{S}_{\mathbf{R}} \cap span(\mathbf{F})$, it follows from $span(\mathbf{F}) \subseteq span(\mathbf{H})$ that, if $\mathbf{B}$ is designed satisfying

$$\mathcal{S}_{\mathbf{R}} \cap span(\mathbf{H}) \subseteq span(\mathbf{B}) \tag{69}$$

we have $\mathcal{S}_{\mathbf{R}} \cap span(\mathbf{F}) \subseteq span(\mathbf{B})$. Then from Lemma 2, $\mathbf{B}^H\mathbf{P}\mathbf{F}\mathbf{B}$ has the same singular values as $\mathbf{P}\mathbf{F}$. Hence, the spectral efficiency in $\mathcal{P}_1$ will be maximized, when the objective function in (68) is maximized.

D. Proof of Theorem 2

Note that user k's NAS is set to be no smaller than 2Δ. It is easy to verify that if user j is user k's neighbor, $|\theta_j - \theta_k| \leq 2\Delta$. Otherwise, $|\theta_j - \theta_k| > 2\Delta$. Since the AS of each user is Δ, user k is non-overlapped with its non-neighbors, yielding that the CCMs of user k's non-neighbors are orthogonal to that of user k under OA. Hence, $span\left(\mathbf{R}_k\right) \subseteq span^{\perp}\left(\bar{\mathbf{R}}_k\right)$, leading to

$$\bigcup span\left(\mathbf{R}_k\right) \subseteq \bigcup span^{\perp}\left(\bar{\mathbf{R}}_k\right). \tag{70}$$

By $span\left(\mathbf{h}_k\right) \subseteq span\left(\mathbf{R}_k\right)$, we have

$$span\left(\mathbf{H}\right) \subseteq \bigcup span\left(\mathbf{R}_k\right). \tag{71}$$

Combining (70) with (71), we can get

$$span\left(\mathbf{H}\right) \subseteq \bigcup span\left(\mathbf{R}_k\right) \subseteq \bigcup span^{\perp}\left(\bar{\mathbf{R}}_k\right) \tag{72}$$

from which $\mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{H}\right) = span\left(\mathbf{H}\right)$. Moreover, by $\mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{H}\right) \subseteq \mathbf{B}$ in Theorem 1,

$$span\left(\mathbf{H}\right) = \mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{H}\right) \subseteq span\left(\mathbf{B}\right). \tag{73}$$

Due to $span\left(\mathbf{HSH}^H\right) \subseteq span\left(\mathbf{H}\right)$, and from (73),

$$span\left(\mathbf{HSH}^H\right) \subseteq span\left(\mathbf{H}\right) \subseteq span\left(\mathbf{B}\right). \tag{74}$$

Applying Lemma 2 to (74), we can find that the singular values of the matrix $\mathbf{B}^H\mathbf{HSH}^H\mathbf{B}$ is the same as that of $\mathbf{HSH}^H$, leading to

$$\begin{aligned}&\max_{\mathbf{S}\succeq 0, Tr(\mathbf{S})\leq P} \log\left|\mathbf{I} + \mathbf{B}^H\mathbf{HSH}^H\mathbf{B}\right| \\ &\quad = \max_{\mathbf{S}\succeq 0, Tr(\mathbf{S})\leq P} \log\left|\mathbf{I} + \mathbf{HSH}^H\right|.\end{aligned} \tag{75}$$

E. Proof of Lemma 3

Let $\tilde{\mathcal{S}}_k = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}\right)$. Then $\mathcal{S}_{\mathbf{R}} \cap span\left(\mathbf{R}\right) = \bigcup_k \tilde{\mathcal{S}}_k$. Since $\mathcal{S}_k \subseteq \tilde{\mathcal{S}}_k$, we have

$$\bigcup_k \mathcal{S}_k \subseteq \bigcup_k \tilde{\mathcal{S}}_k. \tag{76}$$

If we can prove $\bigcup_k \mathcal{S}_k \supseteq \bigcup_k \tilde{\mathcal{S}}_k$, we have $\bigcup_k \tilde{\mathcal{S}}_k = \bigcup_k \mathcal{S}_k$ by (76), from which Lemma 3 follows. Due to $\mathbf{R} = \sum_{k=1}^{K} \mathbf{R}_i$, it is easy to verify that $\tilde{\mathcal{S}}_k = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap \bigcup_{j\in\Omega_k} span\left(\mathbf{R}_j\right)$. Next, we establish that, for each $j \in \Omega_k$, $span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) \subseteq \bigcup_k \mathcal{S}_k$, from which $\tilde{\mathcal{S}}_k \subseteq \bigcup_k \mathcal{S}_k$, leading to $\bigcup_k \mathcal{S}_k \supseteq \bigcup_k \tilde{\mathcal{S}}_k$. To prove $span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) \subseteq \bigcup_k \mathcal{S}_k$ for each $j \in \Omega_k$, three cases are considered, i.e., $\theta_j < \theta_k, \theta_j = \theta_k$ and $\theta_j > \theta_k$.

a) Case $\theta_j < \theta_k$: Note that the index sets of non-neighbors of user j and k are

$$\bar{\Omega}_k = \{1, 2, \cdots, k_l - 1\} \cup \{k_u + 1, k_u + 2, \cdots, K\} \tag{77a}$$

$$\bar{\Omega}_j = \{1, 2, \cdots, j_l - 1\} \cup \{j_u + 1, k_u + 2, \cdots, K\} \tag{77b}$$

respectively. By $\theta_j < \theta_k$, we can get $j_l < k_l, k < j_u < k_u$. Considering

$$\hat{\mathcal{S}} = span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap \left(span\left(\mathbf{R}_j\right) - span\left(\mathbf{R}_k\right)\right) \tag{78}$$

we have $\hat{\mathcal{S}} \subseteq span\left(\mathbf{R}_j\right) - span\left(\mathbf{R}_k\right)$. It is easy to verify that, when $\theta_j < \theta_k$, $span\left(\mathbf{R}_j\right) - span\left(\mathbf{R}_k\right)$ is orthogonal to both $span\left(\mathbf{R}_k\right)$ and $span\left(\mathbf{R}_l\right)(l > k)$ under OA. Thus, $\hat{\mathcal{S}}$ is orthogonal to $span\left(\mathbf{R}_l\right)(l \geq k)$, and consequently,

$$\begin{aligned}\hat{\mathcal{S}} &= span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) \cap span^{\perp}\left(\sum_{l\geq k} \mathbf{R}_l\right) \\ &= span\left(\mathbf{R}_j\right) \cap span^{\perp}\left(\sum_{l\in\bar{\Omega}_k, l\geq k} \mathbf{R}_l\right).\end{aligned} \tag{79}$$

Note that

$$\begin{aligned}\mathcal{S}_j &= span^{\perp}\left(\bar{\mathbf{R}}_j\right) \cap span\left(\mathbf{R}_j\right) \\ &= span\left(\mathbf{R}_j\right) \cap span^{\perp}\left(\sum_{l\in\bar{\Omega}_j} \mathbf{R}_l\right).\end{aligned} \tag{80}$$

Since

$$\begin{aligned}\left\{l | l \in \bar{\Omega}_k, l \geq k\right\} &= \{1, 2, \cdots, k_l - 1\} \cup \{k, k+1, \cdots, K\} \\ &\supseteq \bar{\Omega}_j\end{aligned} \tag{81}$$

we can get $\hat{\mathcal{S}} \subseteq \mathcal{S}_j$ by comparing (79) with (80), and from (78), we have, for each $\theta_j < \theta_k$,

$$span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) \subseteq \hat{\mathcal{S}} + \mathcal{S}_k \subseteq \mathcal{S}_j \cup \mathcal{S}_k \subseteq \bigcup_k \mathcal{S}_k. \tag{82}$$

b) Case $\theta_j = \theta_k$: When $\theta_j = \theta_k$, user j and user k are the same, and it is easy to verify that $span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) = span^{\perp}\left(\bar{\mathbf{R}}_j\right) \cap span\left(\mathbf{R}_j\right) \subseteq \tilde{\mathcal{S}}_j$.

c) Case $\theta_j > \theta_k$: Since the proof in this case is similar to that in $\theta_j < \theta_k$, we omit the proof for this case, and we have the conclusion that $span^{\perp}\left(\bar{\mathbf{R}}_k\right) \cap span\left(\mathbf{R}_j\right) \subseteq \bigcup_k \mathcal{S}_k$ for each $\theta_j > \theta_k$.

From the discussion above, we can show that $\bigcup_k \mathcal{S}_k \supseteq \bigcup_k \tilde{\mathcal{S}}_k$. Combining $\bigcup_k \mathcal{S}_k \supseteq \bigcup_k \tilde{\mathcal{S}}_k$ and (76), we have $\bigcup_k \mathcal{S}_k = \bigcup_k \tilde{\mathcal{S}}_k$, from which Lemma 3 follows.

ACKNOWLEDGMENT

The authors would like to thank Jie Gao for his helpful suggestions.

REFERENCES

[1] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[2] Y. Xu, H. Yang, F. Ren, C. Lin, and X. S. Shen, "Frequency domain packet scheduling with MIMO for 3GPP LTE downlink," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1752–1761, Apr. 2013.

[3] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[4] N. Cheng *et al.*, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 178–185, Feb. 2020.

[5] K. Samdanis and T. Taleb, "The road beyond 5G: A vision and insight of the key technologies," *IEEE Netw.*, vol. 34, no. 2, pp. 135–141, Mar. 2020.

[6] X. Cheng *et al.*, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.

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

[8] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327–1338, Jun. 2017.

[9] W. Wang, Z. Yin, J. Ni, X. Lin, and X. S. Shen, "Against pilot spoofing attack with double channel training in massive MIMO NOMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[10] Y. Chen, C. Liu, Y. Fu, H. Wang, and W.-P. Zhu, "Asymptotic capacity analysis of large-scale MIMO AF relay networks in a composite fading environment," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 8895–8909, Nov. 2016.

[11] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmWave beam alignment via correlated bandit learning," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5894–5908, Dec. 2019.

[12] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.

[13] A. Gokceoglu, E. Bjornson, E. G. Larsson, and M. Valkama, "Spatio-temporal waveform design for multiuser massive MIMO downlink with 1-bit receivers," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 2, pp. 347–362, Mar. 2017.

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

[15] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[16] F. Rottenberg, T. Choi, P. Luo, C. J. Zhang, and A. F. Molisch, "Performance analysis of channel extrapolation in FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2728–2741, Apr. 2020.

[17] 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.

[18] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.

[19] Z. Jiang, S. Chen, A. F. Molisch, R. Vannithamby, S. Zhou, and Z. Niu, "Exploiting wireless channel state information structures beyond linear correlations: A deep learning approach," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 28–34, Mar. 2019.

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

[21] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178–2181, Dec. 2015.

[22] 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.

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

[24] X. Li, S. Jin, H. A. Suraweera, J. Hou, and X. Gao, "Statistical 3-D beamforming for large-scale MIMO downlink systems over Rician fading channels," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1529–1543, Apr. 2016.

[25] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, Jun. 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 Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3029–3041, May 2017.

[27] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sep. 2014.

[28] X. Sun, X. Gao, G. Y. Li, and W. Han, "Agglomerative user clustering and downlink group scheduling for FDD massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[29] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, Jun. 2014.

[30] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, Jun. 2014.

[31] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Multi-layer precoding: A potential solution for full-dimensional massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5810–5824, Sep. 2017.

[32] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2181–2194, Apr. 2017.

[33] Y. Song, C. Liu, and Y. Zou, "The precoding scheme based on domain selective interference cancellation in 3-D massive MIMO," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1228–1231, Jun. 2018.

[34] Y. Song *et al.*, "Domain selective precoding in 3-D massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1103–1118, Sep. 2019.

[35] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[36] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[37] X. Li, S. Jin, X. Gao, and R. W. Heath, Jr., "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.

[38] D. Yu *et al.*, "Stable local broadcast in multihop wireless networks under SINR," *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 1278–1291, Jun. 2018.

[39] Y. Wang, D. Yu, Q. Liu, and F. C. M. Lau, "Inductive coloring: Implementing basic communication primitives with Rayleigh-fading interference," in *Proc. IEEE 35th Annu. IEEE Int. Conf. Comput. Commun. INFOCOM*, Apr. 2016, pp. 1–9.

[40] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

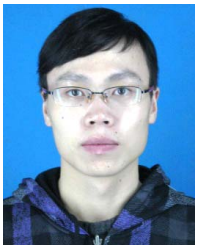
[41] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Danvers, MA, USA: Wiley, 1973.

[42] J. Li, X. Wu, and P. Viswanath, *OFDMA Mobile Broadband Communications: A Systems Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[43] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (Release 6)*, document TR 25.996, V6.1.0., 3GPP, Sep. 2003.

[44] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.

[45] X. Zhang, *Matrix Analysis and Applications*. Beijing, China: Tsinghua Univ. Press, 2013.



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. He is currently a Visiting Scholar with the BBCR Lab, Department of ECE, University of Waterloo, Canada. Since 2017, he has been an Instructor with the College of Electronic and Optical Engineering, NUPT. His research interests include massive MIMO systems and millimeter wave communications.

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 in 2005. In 1988, he joined NUPT, where he has been a Professor since 2002. His current research interests include massive MIMO systems, 3-D MIMO systems, and millimeter wave communications.

Yiliang Liu (Student Member, IEEE) received the B.E. and M.Sc. degrees in computer science and communication engineering from Jiangsu University, Zhenjiang, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Communication Research Centre, Harbin Institute of Technology, China. He was a Visiting Research Student with the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan, from 2014 to 2015, and the Department of Electrical and Computer Engineering, University of Waterloo, Canada, from 2018 to 2019. His research interests include security of wireless communications, physical-layer security, and intelligent connected vehicles.

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, in 2016. He worked as a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, from 2017 to 2019. He is currently a Professor with the State Key Laboratory of ISN, School of Telecommunication Engineering, Xidian University, Shaanxi, China. His current research interests include B5G/6G, space-air-ground integrated networks, big data in vehicular networks, self-driving systems, performance analysis, MAC, opportunistic communications, and applications of AI for vehicular networks.

Yongming Huang (Senior Member, IEEE) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2007. Since March 2007, he has been a Faculty Member with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. From 2008 to 2009, he visited the Signal Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include MIMO wireless communications, cooperative wireless communications, and millimeter wave wireless communications. He has published over 200 peer-reviewed articles and holds over 60 invention patents. He submitted around 20 technical contributions to IEEE standards. He was awarded a Certificate of Appreciation for outstanding contribution to the development of IEEE standard 802.11aj. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE WIRELESS COMMUNICATIONS LETTERS, the *EURASIP Journal on Advances in Signal Processing*, and the *EURASIP Journal on Wireless Communications and Networking* and a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990.

He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include network resource management, wireless network security, the Internet of Things, 5G and beyond, and vehicular ad-hoc and sensor networks.

Dr. Shen is a member of the IEEE Fellow Selection Committee. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and the Communications Society. He received the R. A. Fessenden Award in 2019 from the IEEE, Canada, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and the Education Award in 2017 from the IEEE Communications Society, and the Technical Recognition Award from the Wireless Communications Technical Committee in 2019 and the AHSN Technical Committee in 2013. He also received the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario. He has served as the Technical Program Committee Chair/Co-Chair of the IEEE GLOBECOM'16, the IEEE INFOCOM'14, the IEEE VTC'10 Fall, and the IEEE GLOBECOM'07, the Symposia Chair of the IEEE ICC'10, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He was/is the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, the IEEE NETWORK, *IET Communications*, and *Peer-to-Peer Networking and Applications*. He is elected as the IEEE Communications Society Vice President of Technical and Educational Activities, the Vice President of Publications, a Member-at-Large on the Board of Governors, and the Chair of the Distinguished Lecturer Selection Committee.