






Domain Selective Precoding in 3-D Massive MIMO Systems

Yunchao Song , *Member, IEEE*, Chen Liu , *Member, IEEE*, Wei Wang , Nan Cheng , *Member, IEEE*, Miao Wang, *Member, IEEE*, Weihua Zhuang , *Fellow, IEEE*, and Xuemin (Sherman) Shen, *Fellow, IEEE*

Abstract—Three-dimensional multiple input multiple output (3-D MIMO) with a large number of active antennas equipped in a uniformly rectangular antenna array has a significant potential in improving the system capacity. In this paper, we present an efficient two-dimensional (2-D) downlink precoding scheme for single-cell 3-D massive MIMO systems. Considering instantaneous channel state information at the base station, we show that either the elevation or azimuth domain can be used for interference cancellation. We divide the interference into two components, one of which is canceled in the elevation domain, while the other is canceled in the azimuth domain. Based on this domain selective (DS) strategy, two DS precoding algorithms are proposed for the single-path scenario based on the zero forcing and signal-to-leakage-plus-noise ratio criteria. Next, we extend our DS precoding scheme to a multi-path scenario. In the proposed algorithms, the precoding vectors of different users are determined in parallel, so as to reduce the computational complexity. Simulation results are provided to demonstrate that the proposed algorithms can achieve better spectral efficiency performance with a low computational complexity.

Index Terms—Three dimensional multiple-input multiple-output (3D MIMO), inter-user interference, 2D precoding, domain selective precoding scheme.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) technology has been widely used to improve the capacity and reliability of wireless communication systems [1]–[4]. By employing tens or hundreds of antennas and simultaneously serves tens of users at the base station (BS), the massive MIMO systems have many advantages such as reduced uplink transmissions power [5], [6] and improved spectral and energy efficiencies, which

make it a promising technology for future wireless communication systems [7]–[10].

Despite the benefits of massive MIMO systems, there are challenges in their practical deployment. One main challenge is that the number of antennas is often limited by the physical space [11]. For example, a uniform linear array (ULA) antenna in a carrier frequency of 2.5 GHz requires a space of about 1.9 m to fit 32 antenna elements with half wavelength spacing between adjacent elements. Three-dimensional MIMO (3D MIMO) systems have recently been proposed to solve this problem [11]–[14]. For 3D massive MIMO systems, a large number of active antenna elements can be placed in a uniformly rectangular antenna array (URA) at the BS, in order to reduce the physical space requirements. Moreover, by exploiting the vertical dimension, the users in the same horizontal direction but at a different height can be served simultaneously in 3D MIMO. However, the interference in 3D MIMO systems exists in both elevation and azimuth domains, which complicates the precoding design.

The precoding design in 3D massive MIMO systems depends on the availability of instantaneous channel state information (CSI) at the BS. In a system where instantaneous CSI is difficult to obtain, such as in a frequency division duplexing (FDD) system, some precoding techniques utilizing the statistical CSI are proposed. Li *et al.* in [15] propose an asymptotical precoding scheme in the 3D massive MIMO system and uses a scheduling algorithm to efficiently reduce the inter-user interference. A 3D beamforming plus joint spatial-division multiplexing (JSDM) transmission algorithm is proposed in [16], [17]. The JSDM scheme divides the users into groups, reduces the inter-group interference by exploiting the statistical CSI, and suppresses the intra-group interference by exploiting the instantaneous CSI [16], [17]. Similarly, Alkhateeb *et al.* in [18] cancels the inter-cell interference by exploiting the statistical CSI, while cancels the intra-cell interference by using the instantaneous CSI.

Considering the BS with instantaneous CSI, such as in a time division duplexing (TDD) system, two kinds of precoding schemes have been proposed to reduce the interference. The first kind is the conventional full precoding algorithm [5], where the structure of 3D MIMO channel is not exploited. This full precoding scheme exhibits excellent spectral efficiency, at the cost of high complexity which is related to the total number of antennas. The second approach is the two-dimensional (2D) precoding, which is often used in the Kronecker product channel model where the precoding in the elevation domain and that

Manuscript received February 14, 2019; revised June 10, 2019; accepted July 5, 2019. Date of publication July 31, 2019; date of current version September 20, 2019. This work was supported in part by the Natural Science Foundation of China under Grants 61771257 and 61871238, in part by the Natural Science Foundation of Jiangsu Higher Education Institutions under Grant 18KJB510027, and in part by the Natural Sciences and Engineering Research Council (NSERC), Canada. The Guest Editor coordinating the review of this manuscript and approving it for publication was Prof. F. Gao. (*Corresponding author: Chen Liu.*)

Y. Song and C. 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).

W. Wang, N. Cheng, W. Zhuang, and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: wei.wang1@uwaterloo.ca; n5cheng@uwaterloo.ca; wzhuang@uwaterloo.ca; sshen@uwaterloo.ca).

M. Wang is with the Department of Electrical and Computer Engineering, Miami University, Oxford, OH 45056 USA (e-mail: Wangm64@MiamiOH.edu).
Digital Object Identifier 10.1109/JSTSP.2019.2930889

in the azimuth domain are designed separately [19]–[23]. In [19], Song *et al.* propose an eigenbeamforming scheme for point to point systems, by using the eigenvector corresponding to the largest eigenvalue of the elevation channel correlation matrix. The 2D matched filter (MF) and 2D zero-forcing (ZF) precoding algorithms [20], [21] are proposed to reduce the inter-user interference in multi-user single-cell systems, which performs well when the elevation beam of each user covers a limited angle, and 2D MF outperforms 2D ZF. Seifi *et al.* [22] proposed a 3D coordinated beamforming scheme to reduce the inter-cell interference in both horizontal and vertical planes of the wireless channel for multi-cell systems. The 2D precoding scheme, although has a significantly reduced computational complexity, achieves a lower spectral efficiency compared with the full precoding algorithm. In [23], a domain selective (DS) 2D precoding scheme and a DS interference cancellation (DSIC) algorithm with group based domain selection scheme are proposed to enhance the spectral efficiency in a single-cell 3D massive MIMO system. However, the heuristic DSIC algorithm still has a spectral efficiency gap with the full precoding scheme.

In this paper, to further enhance the spectral efficiency of 2D precoding in the single-cell 3D massive MIMO system, two domain selective precoding algorithms termed as DSZF algorithm and DS signal-to-leakage-plus-noise ratio (DSSLNR) algorithm are proposed in the single-path channel model. In the DSZF algorithm, all the inter-user interference are divided into two components, one of which is canceled in the elevation domain, and the other is canceled in the azimuth domain. While the DSSLNR uses either elevation or azimuth domain to reduce the impact of interference and noise, such that the users' SLNRs are maximized. Compared with the user group method in [23], the DSZF and DSSLNR algorithms give a more efficient domain selection method to improve the spectral efficiency in the 3D massive MIMO systems. We theoretically prove that our proposed DS precoding algorithms can achieve the same spectral efficiency as the full precoding algorithm when the number of antennas becomes infinity. In designing the DSZF and DSSLNR algorithms, we use a parallel computational strategy to reduce the computational complexity, i.e., the precoding vectors of different users are calculated in parallel. Moreover, we extend the DSSLNR algorithm to the one-ring channel models. Simulations validate that, our proposed DS precoding algorithm results can exhibit excellent spectral efficiency at a low computational complexity, which makes DS precoding scheme efficient in 3D massive MIMO systems.

The remainder of this paper is organized as follows. Section II describes the channel model along with the signal model. In Section III, we derive the DSZF and DSSLNR precoding algorithms in the single-path scenario. We also show that our algorithms can obtain the same performance as the full precoding when the number of antennas approaches to infinity. The DSSLNR algorithm for the one-ring channel model is presented in Section VI. Simulation results are presented in Section V, and we finally conclude this work in Section VI.

Notations: Columns and vectors are denoted in lowercase boldfaced characters, and matrices are represented in uppercase boldfaced characters; \mathbf{a}_i denotes the i th column of the matrix

\mathbf{A} ; $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^\dagger$ indicate the matrix transpose, conjugate, conjugate-transpose, and pseudo inverse operations, respectively; $\|\cdot\|$ denotes the 2-norm operation. The complex number field is represented by \mathbb{C} . We denote $\vec{\mathbf{a}}$ as normalized \mathbf{a} given by $\vec{\mathbf{a}} = \mathbf{a}/\|\mathbf{a}\|$, and \mathcal{K} as set $\{1, 2, \dots, K\}$.

II. SYSTEM AND CHANNEL MODEL

A. System Model

Consider a single-cell system with one BS serving K single-antenna users. The BS is equipped with a URA of $N = M_x M_y$ elements, where M_x and M_y are the numbers of the antennas at the elevation and azimuth dimensions, respectively. In the downlink, the BS applies $N \times K$ precoder \mathbf{W} to transmit information to K users, where \mathbf{w}_k is the k th column of \mathbf{W} . We assume that the instantaneous CSI is available at the BS, which is reasonable in TDD systems [24], [25]. Then, the signal received by user k can be written as

$$y_k = \sum_{j=1}^K \mathbf{h}_k^T \mathbf{w}_j s_j + n_k \quad (1)$$

where \mathbf{h}_k^T denotes the channel vector from the BS to user k , $\mathbf{s} = [s_1, s_2, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$ is the transmitted signal, such that $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = E_s \mathbf{I}$, with \mathbf{I} being the identity matrix, E_s representing the average power of each element in \mathbf{s} , and $\mathbf{n} = [n_1, n_2, \dots, n_K]^T \in \mathbb{C}^{K \times 1}$ is the additive white Gaussian noise with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. The ergodic rate of user k is

$$R_k = \log(1 + \mu_k) \quad (2)$$

where μ_k denotes the signal-to-interference-plus-noise ratio (SINR) of user k , which is given by

$$\mu_k = \frac{|\mathbf{h}_k^T \mathbf{w}_k|^2 E_s}{\sum_{j \neq k} |\mathbf{h}_k^T \mathbf{w}_j|^2 E_s + \sigma^2}. \quad (3)$$

B. Channel Model

We consider a 3D massive MIMO system where a URA is deployed at the BS. Let $D_x = D_y = \lambda/2$ for the sake of simplicity, where D_x and D_y denote the antenna spacings in the elevation and azimuth directions, respectively, and λ is the wave length. We consider a non-dispersive narrow-band channel, with \mathbf{h}_k being expressed as [20], [21], [26]

$$\mathbf{h}_k = \sum_{l=1}^L \alpha_{k,l} \mathbf{l} \mathbf{a}(\theta_{k,l})^e \otimes \mathbf{a}(\theta_{k,l}, \phi_{k,l})^a \quad (4)$$

in which $\alpha_{k,l}$ is the channel path gain, \otimes is the Kronecker product, L is the number of paths, $\theta_{k,l} \in [-\pi, \pi]$ and $\phi_{k,l} \in [0, \pi]$ are the angles of arrival (AoAs) in the elevation and azimuth domains, respectively, and the elevation and azimuth array response vectors are given by

$$\begin{aligned} \mathbf{a}(\theta_{k,l})^e &= \frac{1}{\sqrt{M_x}} \left[1, e^{j u_{k,l}}, \dots, e^{j(M_x-1)u_{k,l}} \right]^T \\ \mathbf{a}(\theta_{k,l}, \phi_{k,l})^a &= \frac{1}{\sqrt{M_y}} \left[1, e^{j v_{k,l}}, \dots, e^{j(M_y-1)v_{k,l}} \right]^T \end{aligned}$$

with $e^{(\cdot)}$ being the natural exponential function, $u_{k,l} = \frac{2\pi D_x}{\lambda} \sin \theta_{k,l}$, $v_{k,l} = \frac{2\pi D_y}{\lambda} \cos \theta_{k,l} \cos \phi_{k,l}$.

Denote $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]$. For the full precoding scheme, ZF is used in designing the precoding matrix \mathbf{W}_{FP} such that $\mathbf{H}^T \mathbf{W}_{FP}$ is a diagonal matrix, which can mitigate the inter-user interference [5], [20], [21], and the precoding vector of user k is given by $\vec{\mathbf{w}}_{k,FP}$, where $\mathbf{w}_{k,FP}$ is the k th column of

$$\mathbf{W}_{FP} = \mathbf{H}^* (\mathbf{H}^T \mathbf{H}^*)^{-1}. \quad (5)$$

This scheme can completely cancel out the inter-user interference, and can achieve high spectral efficiency. However, when the numbers of antennas and users become large, the complexity of this algorithm will be huge, which is prohibitive in massive MIMO systems.

C. 2D Precoding Scheme

We describe the 2D precoding scheme in this subsection. Since the channel matrix has the Kronecker structure, an intuitive way is to formulate the precoding vector in a Kronecker product structure, i.e., $\mathbf{w}_k = \mathbf{w}_k^e \otimes \mathbf{w}_k^a$, and \mathbf{w}_k^e and \mathbf{w}_k^a are the precoding vectors in the elevation and azimuth domains, respectively. Then, the SINR of user k can be written as

$$\mu_k = \frac{|\mathbf{h}_k^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a)|^2 E_s}{\sum_{j \neq k} |\mathbf{h}_k^T (\mathbf{w}_j^e \otimes \mathbf{w}_j^a)|^2 E_s + \sigma^2}. \quad (6)$$

Note that the precoding vector with Kronecker structure is determined by two small size vectors (i.e., \mathbf{w}_k^e and \mathbf{w}_k^a), the complexity of 2D precoding can be significantly reduced as compared with the full precoding scheme, which is attractive in massive MIMO systems. Some algorithms [19]–[23] have been designed to improve the spectral efficiency. However, these algorithms have a performance gap with the full precoding algorithm.

III. DOMAIN DELECTIVE PRECODING IN SINGLE-PATH SCENARIOS

Single-path channel models describe the scenarios where the channel is dominated by one line-of-sight (LOS) or non-LOS (NLOS) path. This is particularly relevant to systems with sparse channels, such as mmWave systems [27]–[31]. Consider a user with a single-path channel defined by its azimuth and elevation angles. The channel vector of user k is expressed as

$$\mathbf{h}_k = \alpha_k \mathbf{a}(\theta_k)^e \otimes \mathbf{a}(\theta_k, \phi_k)^a = \mathbf{h}_k^e \otimes \mathbf{h}_k^a \quad (7)$$

where $\mathbf{h}_k^e = \alpha_k \mathbf{a}(\theta_k)^e$, $\mathbf{h}_k^a = \mathbf{a}(\theta_k, \phi_k)^a$, θ_k and ϕ_k are the elevation and azimuth AOA. We aim to design a precoding vector $\mathbf{w}_k = \mathbf{w}_k^e \otimes \mathbf{w}_k^a$ such that the spectral efficiency is maximized. In the following, two domain selective algorithms are proposed.

A. DSZF Precoding Algorithm

1) *Problem Formulation in DSZF*: Since the inter-user interference is a limiting factor for the spectral efficiency, our DSZF scheme aims to completely cancel out the inter-user interference, which is similar to the conventional ZF precoding. To this end,

we want to design \mathbf{W} to satisfy

$$\mathbf{H}^T \mathbf{W} = \mathbf{D}$$

where \mathbf{D} is a diagonal matrix. With $\mathbf{w}_k = \mathbf{w}_k^e \otimes \mathbf{w}_k^a$, for each k , we have

$$\mathbf{H}^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a) = \mathbf{d}_k \quad (8)$$

with \mathbf{d}_k being the k th column of \mathbf{D} . (8) is a system of bilinear equations [32], which has no exact solution. However, each column of the channel matrix \mathbf{H} in single-path scenarios has the Kronecker product structure, which can be utilized to analyze (8).

From (8), the interference leakages (ILs) from user k are zero, where the IL on user j ($j \neq k$) from user k is defined as $|\mathbf{h}_j^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a)|^2 E_s$. Once all the ILs are mitigated, the inter-user interference is canceled completely, and the SINR of user k reduces to the SNR given by

$$\rho_k = |\mathbf{h}_k^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a)|^2 \cdot E_s / \sigma^2. \quad (9)$$

Our DSZF scheme maximizes the SNR while canceling the ILs. For user k , the problem is formulated as

$$\begin{aligned} \mathcal{P}_1 : \quad & \max_{\|\mathbf{w}_k^e\|=1, \|\mathbf{w}_k^a\|=1} |\mathbf{h}_k^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a)|^2 \\ & \text{s.t. } \mathbf{h}_j^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a) = 0, \forall j \neq k. \end{aligned} \quad (10)$$

To better understand the 2D precoding, we introduce a lemma in [23], which can be used to divide problem \mathcal{P}_1 into the elevation and azimuth parts.

Lemma 1: Let $\mathbf{H}^e = [\mathbf{h}_1^e, \mathbf{h}_2^e, \dots, \mathbf{h}_K^e]$, $\mathbf{H}^a = [\mathbf{h}_1^a, \mathbf{h}_2^a, \dots, \mathbf{h}_K^a]$, $\mathbf{W}^e = [\mathbf{w}_1^e, \mathbf{w}_2^e, \dots, \mathbf{w}_K^e]$ and $\mathbf{W}^a = [\mathbf{w}_1^a, \mathbf{w}_2^a, \dots, \mathbf{w}_K^a]$. Define $\mathbf{H} = \mathbf{H}^e \odot \mathbf{H}^a$ and $\mathbf{W} = \mathbf{W}^e \odot \mathbf{W}^a$, where \odot is Khatri-Rao product. We obtain

$$\mathbf{H}^T \mathbf{W} = (\mathbf{H}^e)^T \mathbf{W}^e * (\mathbf{H}^a)^T \mathbf{W}^a \quad (11)$$

where $*$ is the Hardward product.

From Lemma 1, we have

$$\mathbf{h}_j^T (\mathbf{w}_k^e \otimes \mathbf{w}_k^a) = (\mathbf{h}_j^e)^T \mathbf{w}_k^e \cdot (\mathbf{h}_j^a)^T \mathbf{w}_k^a \quad (12)$$

then $\mathbf{h}_j^T \mathbf{w}_k = 0$ ($j \neq k$) can be rewritten as

$$(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0 \text{ or } (\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, \forall j \neq k. \quad (13)$$

Eq. (13) shows that, in the single-path scenario, user k 's IL can be canceled in either elevation domain or azimuth domain. In fact, one may choose two domains to cancel each IL to mitigate the interference; however, it is unnecessary and will reduce the signal energy power. Here, we choose only one domain to cancel the IL.

To select the better domain to cancel the IL, we need to find two sets Φ and $\bar{\Phi}$ such that

$$(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0, j \in \Phi; (\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \in \bar{\Phi} \quad (14)$$

where $\Phi \cap \bar{\Phi} = \emptyset$, and $\Phi \cup \bar{\Phi} = \mathcal{K} \setminus \{k\}$. The first equation means that the IL on user j ($j \in \Phi$) is mitigated in the elevation domain, while the second means that the IL on user j ($j \in \bar{\Phi}$) is mitigated in the azimuth domain.

To completely cancel out all ILs, the elevation channel vectors of users in Φ should not be collinear with that of user k , while the azimuth channel vectors of users in $\bar{\Phi}$ should not be collinear with that of user k . Moreover, the numbers of non-collinear vectors in \mathbf{h}_j^e ($j \in \Phi$) and \mathbf{h}_j^a ($j \in \bar{\Phi}$) should be smaller than M_x and M_y , respectively, so that we can find non-zero vectors \mathbf{w}_k^e and \mathbf{w}_k^a satisfying (14). Since the elevation and azimuth AOA of each user are continuous random variables, the probability that the users have the identical elevation (azimuth) AOAs is zero. We have the following proposition.

Proposition 1: To cancel out the inter-user interference, the DSZF can serve at most $M_x + M_y - 1$ users with probability one.

Proof: See Appendix A. \blacksquare

From Lemma 1, the SNR in (9) can be re-written as

$$\rho_k = \left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \cdot \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2 \cdot E_s / \sigma^2. \quad (15)$$

Problem \mathcal{P}_1 can be equivalently transformed as

$$\begin{aligned} \mathcal{P}_2 : \quad & \max_{\Phi, \bar{\Phi}, \|\mathbf{w}_k^e\|=1, \|\mathbf{w}_k^a\|=1} \left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \cdot \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2 \\ \text{s.t.} \quad & (\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0, j \in \Phi; (\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \in \bar{\Phi}. \end{aligned} \quad (16)$$

It is complexity expensive to get the optimal solution of problem \mathcal{P}_2 , since \mathcal{P}_2 involves two index sets Φ and $\bar{\Phi}$. Below we give an efficient algorithm to solve \mathcal{P}_2 .

2) *Description of DSZF Algorithm:* When the sets Φ and $\bar{\Phi}$ are given, \mathcal{P}_2 can be factored into two problems. The first one involves the elevation precoding vector \mathbf{w}_k^e , i.e.,

$$\begin{aligned} \mathcal{P}_{2.1} : \quad & \max_{\|\mathbf{w}_k^e\|=1} \left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \cdot g_1(\mathbf{w}_k^a) \\ \text{s.t.} \quad & (\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0, j \in \Phi \end{aligned} \quad (17)$$

where $g_1(\mathbf{w}_k^a) = |(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2$ is uncorrelated with \mathbf{w}_k^e . The second problem relates to the azimuth vector \mathbf{w}_k^a , given by

$$\begin{aligned} \mathcal{P}_{2.2} : \quad & \max_{\|\mathbf{w}_k^a\|=1} \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2 \cdot g_2(\mathbf{w}_k^e) \\ \text{s.t.} \quad & (\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \in \bar{\Phi} \end{aligned} \quad (18)$$

where $g_2(\mathbf{w}_k^e) = |(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2$ is uncorrelated with \mathbf{w}_k^a .

Proposition 2: Let $\mathbf{A} \in \mathbb{C}^{N \times K}$ and $\mathbf{b} \in \mathbb{C}^{N \times 1}$ be an arbitrary matrix and an arbitrary vector, respectively. Let \mathbf{f} be the first row of $[\mathbf{b} \ \mathbf{A}]^\dagger$, then $\bar{\mathbf{f}}^T$ is the optimal solution of the following two problems,

a) when \mathbf{A} is not full column rank,

$$\max_{\|\mathbf{x}\|=1} |\mathbf{b}^T \mathbf{x}|^2, \text{ s.t. } \mathbf{A}^T \mathbf{x} = 0; \quad (19)$$

b) when \mathbf{A} is full column rank,

$$\max_{\|\mathbf{x}\|=1} \frac{|\mathbf{b}^T \mathbf{x}|^2}{\|\mathbf{A}^T \mathbf{x}\|^2}. \quad (20)$$

Proof: See Appendix B. \blacksquare

Proposition 2 shows that the ZF precoding solution in each domain is the optimal solution of $\mathcal{P}_{2.1}$ or $\mathcal{P}_{2.2}$, namely, the solutions of $\mathcal{P}_{2.1}$ and $\mathcal{P}_{2.2}$ are $\bar{\mathbf{w}}_\Phi^{e,1}$ and $\bar{\mathbf{w}}_\Phi^{a,1}$ respectively, where $\mathbf{w}_\Phi^{e,1}$ and $\mathbf{w}_\Phi^{a,1}$ are the first row of \mathbf{W}_Φ^e and that of \mathbf{W}_Φ^a , with

$$\mathbf{W}_\Phi^e = [\mathbf{h}_k^e, \mathbf{H}_\Phi^e]^\dagger, \quad \mathbf{W}_\Phi^a = [\mathbf{h}_k^a, \mathbf{H}_\Phi^a]^\dagger. \quad (21)$$

Moreover, it shows that our DSZF scheme has the largest elevation or azimuth signal-to-interference-leakage ratio when the IL cannot be mitigated. Then \mathcal{P}_2 can be transformed as

$$\begin{aligned} \mathcal{P}_3 : \quad & \max_{\Phi, \bar{\Phi}} \left| (\mathbf{h}_k^e)^T \mathbf{w}_{k,op}^e \right|^2 \cdot \left| (\mathbf{h}_k^a)^T \mathbf{w}_{k,op}^a \right|^2 \\ \text{s.t.} \quad & \Phi \cap \bar{\Phi} = \emptyset, \Phi \cup \bar{\Phi} = \mathcal{K} / \{k\} \end{aligned} \quad (22)$$

where $\mathbf{w}_{k,op}^e$ and $\mathbf{w}_{k,op}^a$ are the solutions of $\mathcal{P}_{2.1}$ and $\mathcal{P}_{2.2}$, respectively. To obtain the optimal Φ and $\bar{\Phi}$ in \mathcal{P}_3 , a global search algorithm is usually adopted, which suffers from exponential complexity and is prohibitive in massive MIMO systems. Next, we introduce a low complexity algorithm to obtain the solution.

Considering user k , there are $K - 1$ ILs. We search all these ILs, and decide which IL is canceled in the azimuth domain or the elevation domain. At the initial stage, let $\Phi_0 = \emptyset$, and $\bar{\Phi}_0 = \emptyset$, and the optimal elevation and azimuth precoding vectors $\mathbf{w}_{k,0}^e = (\bar{\mathbf{h}}_k^e)^*$ and $\mathbf{w}_{k,0}^a = (\bar{\mathbf{h}}_k^a)^*$ respectively. In the $(i - 1)$ th iteration, suppose we have Φ_{i-1} and $\bar{\Phi}_{i-1}$, where $\Phi_{i-1} \cap \bar{\Phi}_{i-1} = \emptyset$ and $\Phi_{i-1} \cup \bar{\Phi}_{i-1} = \Omega_{i-1}$, Ω_{i-1} is the index set of the ILs that have been considered. Let $\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e$ denote the matrix whose columns are $\mathbf{h}_j^e, j \in \Phi_{i-1}$, and $\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^a$ the matrix whose columns are $\mathbf{h}_j^a, j \in \bar{\Phi}_{i-1}$. In the i th iteration, we want to cancel the IL on user l (this means the i th element in $\mathcal{K} / \{k\}$ is l). If we cancel it in the elevation domain, then $\Phi_i = \Phi_{i-1} \cup \{l\}$, $\bar{\Phi}_i = \bar{\Phi}_{i-1}$, and $\bar{\mathbf{H}}_{\bar{\Phi}_i}^e$ becomes $[\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e, \mathbf{h}_l^e]$. The azimuth precoding vector remains unchanged, and the elevation precoding vector $\mathbf{w}_{k,i}^e$ should satisfy

$$\begin{aligned} & \max_{\|\mathbf{w}_k^e\|=1} \left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \\ \text{s.t.} \quad & (\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0, j \in \Phi_i. \end{aligned} \quad (23)$$

From Proposition 2, $\mathbf{w}_{k,i}^e$ should be $(\bar{\mathbf{w}}_{\Phi_i}^{e,1})^T$, where $\mathbf{w}_{\Phi_i}^{e,1}$ is the first row of matrix $\mathbf{W}_{\Phi_i}^e = [\mathbf{h}_k^e, \bar{\mathbf{H}}_{\Phi_i}^e]^\dagger = (\bar{\mathbf{H}}_{\Phi_i}^e)^\dagger$. Substituting $\mathbf{w}_{k,i}^e$ and $\mathbf{w}_{k,i}^a$ into (15), we get $p^{e,i}$. Since the inverse operation is expensive, here we propose a low complexity update scheme to obtain $\mathbf{W}_{\Phi_i}^e$ using the lemma in [33]. Let $\bar{\mathbf{H}}_{\Phi_i}^e = [\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e, \mathbf{h}_l^e]$.

Lemma 2: For $\bar{\mathbf{H}}_{\Phi_i}^e = [\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e, \mathbf{h}_l^e]$, we have

$$(\bar{\mathbf{H}}_{\Phi_i}^e)^\dagger = \begin{pmatrix} (\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e)^\dagger \\ \mathbf{b}_m \end{pmatrix} - \mathbf{d}_m \mathbf{b}_m^H \quad (24)$$

where $\mathbf{d}_m = (\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e)^\dagger \mathbf{h}_l^e$,

$$\mathbf{b}_m^H = \begin{cases} (1 + \mathbf{d}_m^H \mathbf{d}_m)^{-1} \mathbf{d}_m^H (\bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e)^\dagger, & \mathbf{h}_l^e - \bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e \mathbf{d}_m = 0, \\ \frac{(\mathbf{h}_l^e - \bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e \mathbf{d}_m)^H}{\|\mathbf{h}_l^e - \bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e \mathbf{d}_m\|^2}, & \mathbf{h}_l^e - \bar{\mathbf{H}}_{\bar{\Phi}_{i-1}}^e \mathbf{d}_m \neq 0. \end{cases}$$

Algorithm 1: The DSZF Algorithm.

Input: $\mathbf{H}^e, \mathbf{H}^a$; **Output:** $\mathbf{W}^e, \mathbf{W}^a$;
 For user $k, k \in \mathcal{K}$, initialize $\mathbf{w}_{k,0}^e = (\bar{\mathbf{h}}_k^e)^*$, $\mathbf{w}_{k,0}^a = (\bar{\mathbf{h}}_k^a)^*$,
 $\Phi_0 = \emptyset$, and $\bar{\Phi}_0 = \emptyset$;
 For $i = 1 : K - 1$
 Assume the i th element in $\mathcal{K}/\{k\}$ (i.e., element l)
 belongs Φ ;
 Denote $\mathbf{w}_{te,e}^a = \mathbf{w}_{k,i-1}^a$, and obtain $\mathbf{W}_{\Phi_i}^e$ using
 Lemma 2;
 Let $\mathbf{w}_{te,e}^e = (\bar{\mathbf{w}}_{\Phi_i}^{e,1})^T$, where $\mathbf{w}_{\Phi_i}^{e,1}$ is the first row of $\mathbf{W}_{\Phi_i}^e$;
 Using $\mathbf{w}_{te,e}^a$ and $\mathbf{w}_{te,e}^e$ to get $p^{e,i}$;
 Assume element l belongs $\bar{\Phi}$;
 Denote $\mathbf{w}_{te,a}^e = \mathbf{w}_{k,i-1}^e$, and obtain $\mathbf{W}_{\bar{\Phi}_i}^a$ using
 Lemma 2;
 Let $\mathbf{w}_{te,a}^a = (\bar{\mathbf{w}}_{\bar{\Phi}_i}^{a,1})^T$, where $\mathbf{w}_{\bar{\Phi}_i}^{a,1}$ is the first row
 of $\mathbf{W}_{\bar{\Phi}_i}^a$;
 Using $\mathbf{w}_{te,a}^e$ and $\mathbf{w}_{te,a}^a$ to get $p^{a,i}$;
 If $p^{e,i} > p^{a,i}$
 $\Phi_i = \Phi_{i-1} \cup l$, $\bar{\Phi}_i = \bar{\Phi}_{i-1}$, $\mathbf{w}_{k,i}^e = \mathbf{w}_{te,e}^e$,
 $\mathbf{w}_{k,i}^a = \mathbf{w}_{te,e}^a$;
 else
 $\Phi_i = \Phi_{i-1}$, $\bar{\Phi}_i = \bar{\Phi}_{i-1} \cup l$, $\mathbf{w}_{k,i}^e = \mathbf{w}_{te,a}^e$,
 $\mathbf{w}_{k,i}^a = \mathbf{w}_{te,a}^a$;
 End
 End

Using Lemma 2, we can obtain $\mathbf{W}_{\Phi_i}^e$ in a recursive way.

Similarly, if the IL on user l is canceled in the azimuth domain, $\Phi_i = \Phi_{i-1}$, $\bar{\Phi}_i = \bar{\Phi}_{i-1} \cup \{l\}$. Then \mathbf{w}_k^e remains unchanged, and \mathbf{w}_k^a should satisfy

$$\begin{aligned} & \max_{\|\mathbf{w}_k^a\|=1} \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2 \\ & \text{s.t. } (\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \in \bar{\Phi}_i. \end{aligned} \quad (25)$$

The solution of (25) is $(\bar{\mathbf{w}}_{\bar{\Phi}_i}^{a,1})^T$, where $\mathbf{w}_{\bar{\Phi}_i}^{a,1}$ is the first row of matrix $\mathbf{W}_{\bar{\Phi}_i}^a = [\mathbf{h}_k^a, \mathbf{H}_{\bar{\Phi}_i}^a]^{\dagger} = (\bar{\mathbf{H}}_{\bar{\Phi}_i}^a)^{\dagger}$. Substituting \mathbf{w}_k^e and \mathbf{w}_k^a into (15), we get $p^{a,i}$. Similarly, we can obtain $\mathbf{W}_{\Phi_i}^a$ from $\mathbf{W}_{\bar{\Phi}_{i-1}}^a$ recursively as in Lemma 2. Then, we only need to choose the larger one between $p^{e,i}$ and $p^{a,i}$. If $p^{e,i}$ is larger than $p^{a,i}$, canceling the IL on user l in the elevation domain will achieve a larger SNR, otherwise canceling the IL in the azimuth domain will achieve a larger SNR.

The DSZF algorithm is summarized in Algorithm 1.

B. DSSLNR Precoding Algorithm

1) *Problem Formulation in DSSLNR:* The DSZF algorithm cancel the inter-user interference in the absence of noise. In this subsection, we take the noise into consideration, and propose a new precoding algorithm, called DSSLNR, to maximize the SLNR of each user. For user k , the SLNR is given by

$$\beta_k = \frac{|\mathbf{h}_k^T \mathbf{w}_k|^2 E_s}{\sum_{j \neq k} |\mathbf{h}_j^T \mathbf{w}_k|^2 E_s + \sigma^2} \quad (26)$$

where $\mathbf{h}_k = \mathbf{h}_k^e \otimes \mathbf{h}_k^a$, $\mathbf{w}_k = \mathbf{w}_k^e \otimes \mathbf{w}_k^a$. Substituting (12) into (26), the SLNR is expressed as

$$\beta_k = \frac{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2 E_s}{\sum_{j \neq k} \left| (\mathbf{h}_j^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_j^a)^T \mathbf{w}_k^a \right|^2 E_s + \sigma^2}.$$

For the ease of notation, denoting $\gamma_k = \beta_k^{-1}$ and $c = \frac{\sigma^2}{E_s}$, we have

$$\begin{aligned} \gamma_k &= \sum_{j \neq k} \frac{\left| (\mathbf{h}_j^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_j^a)^T \mathbf{w}_k^a \right|^2}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2} \\ &+ \frac{c}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2}. \end{aligned}$$

The SLNR maximization problem is formulated as

$$\mathcal{P}_4 : \min \gamma_k, \text{ s.t. } \|\mathbf{w}_k^e\| = 1, \|\mathbf{w}_k^a\| = 1. \quad (27)$$

The term $\frac{|\mathbf{h}_j^e)^T \mathbf{w}_k^e|^2 |\mathbf{h}_j^a)^T \mathbf{w}_k^a|^2}{|\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2 |\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2}$ in γ_k is called user j 's ($j \neq k$) leakage term (LT). Note that each LT is affected by two multipliers, i.e., the azimuth part and the elevation part. In our domain selective scheme, we minimize each LT in γ_k by minimizing one of the multipliers, either in the elevation part or in the azimuth part. In particular, if one multiplier is reduced to a small enough value, while the other multiplier is appropriately upper bounded, the LT becomes small.

2) *Description of DSSLNR Algorithm:* Denote Φ and $\bar{\Phi}$ as the index set of users whose LTs are canceled in the elevation and azimuth domain, respectively. We consider the LTs one by one, and obtain Φ and $\bar{\Phi}$ iteratively, in a manner similar to that of the DSZF algorithm. At the initial stage, we set $\Phi_0 = \emptyset$, and $\bar{\Phi}_0 = \emptyset$, and we have $\mathbf{w}_{k,0}^e = (\bar{\mathbf{h}}_k^e)^*$ and $\mathbf{w}_{k,0}^a = (\bar{\mathbf{h}}_k^a)^*$. In the $(i-1)$ th iteration, assume we have Φ_{i-1} and $\bar{\Phi}_{i-1}$, where $\Phi_{i-1} \cap \bar{\Phi}_{i-1} = \emptyset$ and $\Phi_{i-1} \cup \bar{\Phi}_{i-1} = \Omega_{i-1}$, Ω_{i-1} is the index set of users whose LTs are reduced in the previous $i-1$ iterations. In the i th iteration, we consider the i th LT (user l 's LT), and choose one precoding vector (elevation or azimuth) to reduce the i th LT, which results in smaller $\gamma_{k,i}$, where

$$\begin{aligned} \gamma_{k,i} &= \sum_{j \in \Omega_i} \frac{\left| (\mathbf{h}_j^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_j^a)^T \mathbf{w}_k^a \right|^2}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2} \\ &+ \frac{c}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2 \left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2}. \end{aligned} \quad (28)$$

From above, each LT is reduced in either elevation or azimuth domain. Then, we rewrite (28) as

$$\begin{aligned} \gamma_{k,i} &= \sum_{j \in \Phi_i} a_j^2 \cdot \frac{\left| (\mathbf{h}_j^e)^T \mathbf{w}_k^e \right|^2}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2} + \sum_{j \in \bar{\Phi}_i} b_j^2 \cdot \frac{\left| (\mathbf{h}_j^a)^T \mathbf{w}_k^a \right|^2}{\left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2} \\ &+ c \frac{1}{\left| (\mathbf{h}_k^e)^T \mathbf{w}_k^e \right|^2} \cdot \frac{1}{\left| (\mathbf{h}_k^a)^T \mathbf{w}_k^a \right|^2} \end{aligned} \quad (29)$$

where $a_j = \frac{|(\mathbf{h}_j^e)^T \mathbf{w}_k^e|}{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|}$, $b_j = \frac{|(\mathbf{h}_j^a)^T \mathbf{w}_k^a|}{|(\mathbf{h}_k^a)^T \mathbf{w}_k^a|}$, $\Phi_i \cap \bar{\Phi}_i = \emptyset$ and $\Phi_i \cup \bar{\Phi}_i = \Omega_i$. In (29), we aim to minimize the j th ($j \in \Phi_i$) LT using the elevation precoding vector and minimize the j th ($j \in \bar{\Phi}_i$) LT using the azimuth precoding vector.

In the i th iteration, if we consider to use the elevation precoding vector to reduce the i th LT (denoted as user l 's LT), $\Phi_i = \Phi_{i-1} \cup l$, and $\bar{\Phi}_i = \bar{\Phi}_{i-1}$. Since user j 's ($j \in \bar{\Phi}_i$) LT is minimized using the azimuth precoding vector, $\mathbf{w}_{k,i}^e$ is used to minimize the remaining terms to minimize $\gamma_{k,i}$. The problem is formulated as

$$\mathcal{P}_{5.1} : \min_{\|\mathbf{w}_k^e\|=1} \gamma_{k,\Phi_i}^e = \sum_{j \in \Phi_i} \frac{a_j^2 \cdot |(\mathbf{h}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2} + \frac{c_1}{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2} \quad (30)$$

where $c_1 = c \cdot |(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^{-2}$. We will show that $\mathcal{P}_{5.1}$ is approximately equivalent to minimizing $\gamma_{k,i}$ later. Similarly, using the azimuth precoding vector to reduce this LT, the problem is formulated as

$$\mathcal{P}_{5.2} : \min_{\|\mathbf{w}_k^a\|=1} \gamma_{k,\bar{\Phi}_i}^a = \sum_{j \in \bar{\Phi}_i} \frac{b_j^2 \cdot |(\mathbf{h}_j^a)^T \mathbf{w}_k^a|^2}{|(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2} + \frac{c_2}{|(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2} \quad (31)$$

where $c_2 = c \cdot |(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^{-2}$. Actually, $\mathcal{P}_{5.1}$ and $\mathcal{P}_{5.2}$ can be seen as the elevation and azimuth SLNR maximization problems, respectively, which can be reformulated as

$$\mathcal{P}_{6.1} : \max_{\|\mathbf{w}_k^e\|=1} \frac{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2}{\sum_{j \in \Phi_i} |(a_j \mathbf{h}_j^e)^T \mathbf{w}_k^e|^2 + c_1} \quad (32)$$

$$\mathcal{P}_{6.2} : \max_{\|\mathbf{w}_k^a\|=1} \frac{|(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2}{\sum_{j \in \bar{\Phi}_i} |(b_j \mathbf{h}_j^a)^T \mathbf{w}_k^a|^2 + c_2} \quad (33)$$

Theorem 3: Let $\mathbf{H}_{\Phi_i}^e$ and $\mathbf{H}_{\bar{\Phi}_i}^a$ denote the matrices whose columns are $a_j \mathbf{h}_j^e$, $j \in \Phi_i$ and $b_j \mathbf{h}_j^a$, $j \in \bar{\Phi}_i$, respectively. Let $\tilde{\mathbf{H}}_{\Phi_i}^e = [a_k \mathbf{h}_k^e, \mathbf{H}_{\Phi_i}^e]$ and $\tilde{\mathbf{H}}_{\bar{\Phi}_i}^a = [b_k \mathbf{h}_k^a, \mathbf{H}_{\bar{\Phi}_i}^a]$. The solution of maximizing the elevation or azimuth SLNR is equivalent to the MMSE criterion, which means that the optimal solutions of $\mathcal{P}_{6.1}$ and $\mathcal{P}_{6.2}$ are $\tilde{\mathbf{w}}_{1,\Phi_i}^e$ and $\tilde{\mathbf{w}}_{1,\bar{\Phi}_i}^a$, where \mathbf{w}_{1,Φ_i}^e and $\mathbf{w}_{1,\bar{\Phi}_i}^a$ are the first columns of $\mathbf{W}_{\Phi_i}^e$ and $\mathbf{W}_{\bar{\Phi}_i}^a$, respectively, with

$$\mathbf{W}_{\Phi_i}^e = \left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T + c_1 \mathbf{I} \right)^{-1} \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \quad (34)$$

$$\mathbf{W}_{\bar{\Phi}_i}^a = \left(\left(\tilde{\mathbf{H}}_{\bar{\Phi}_i}^a \right)^* \left(\tilde{\mathbf{H}}_{\bar{\Phi}_i}^a \right)^T + c_2 \mathbf{I} \right)^{-1} \left(\tilde{\mathbf{H}}_{\bar{\Phi}_i}^a \right)^* \quad (35)$$

Proof: See Appendix C. \blacksquare

Remark 1: In our DSSLNR precoding scheme, the elevation channel vectors of users in Φ are not collinear with that of user k . It is because, if the elevation channel vector of user j ($j \neq k$) is collinear with that of user k , the azimuth channel vector of user

j will be not collinear with that of user k , and using the elevation domain to reduce user j 's LT will result in a smaller SLNR than reducing this LT in the azimuth domain, then this LT is reduced in the azimuth domain. Similarly, the azimuth channel vectors of users in $\bar{\Phi}$ are not collinear with that of user k . Hence, we have the following proposition.

Proposition 4: When the numbers of antennas M_x and M_y approach infinity, in the i th iteration, we have

$$\mathbf{w}_{k,i}^e \rightarrow \left(\tilde{\mathbf{h}}_k^e \right)^*, \mathbf{w}_{k,i}^a \rightarrow \left(\tilde{\mathbf{h}}_k^a \right)^*.$$

Proof: See Appendix D. \blacksquare

Remark 2: When using the elevation vector to minimize γ_{k,Φ_i}^e in $\mathcal{P}_{5.1}$, the value of b_j in the second term in (29) is changed, which will enlarge the second term. A similar problem exists in calculating the azimuth precoding vector in $\mathcal{P}_{5.2}$. Actually, from Proposition 4, with an increase of M_x and M_y , we have $\mathbf{w}_{k,i}^e \rightarrow \left(\tilde{\mathbf{h}}_k^e \right)^*$ and $\mathbf{w}_{k,i}^a \rightarrow \left(\tilde{\mathbf{h}}_k^a \right)^*$, and hence a_j (b_j) has approximately the same value in each iteration. In addition, with an increase of the numbers of antennas, \mathbf{h}_j^e ($j \in \Phi_i$) and \mathbf{h}_j^a ($j \in \bar{\Phi}_i$) are different from and orthogonal to \mathbf{h}_k^e and \mathbf{h}_k^a , respectively. Hence, from Proposition 4, $\frac{|(\mathbf{h}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2}$ ($j \in \Phi_i$) and $\frac{|(\mathbf{h}_j^a)^T \mathbf{w}_k^a|^2}{|(\mathbf{h}_k^a)^T \mathbf{w}_k^a|^2}$ ($j \in \bar{\Phi}_i$) will approximately be zero in each iteration. This means that, when minimizing γ_{k,Φ_i}^e in $\mathcal{P}_{5.1}$, the term $b_j^2 \cdot \frac{|(\mathbf{h}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{h}_k^e)^T \mathbf{w}_k^e|^2}$ will approximately be 0, and $\mathcal{P}_{5.1}$ is approximately equivalent to minimizing $\gamma_{k,i}$ in (29). Similarly, minimizing $\gamma_{k,\bar{\Phi}_i}^a$ in $\mathcal{P}_{5.2}$ is approximately equivalent to minimizing $\gamma_{k,i}$ by the azimuth precoding vector.

Next, we can determine in which domain the i th LT in γ_k will be reduced. In the i th iteration, if user l 's LT is reduced in the elevation domain, $\Phi_i = \Phi_{i-1} \cup \{l\}$ and $\bar{\Phi}_i = \bar{\Phi}_{i-1}$. Then, $\mathbf{w}_{k,i}^a = \mathbf{w}_{k,i-1}^a$, and $\mathbf{w}_{k,i}^e$ can be updated by $\mathcal{P}_{6.1}$, which is $\tilde{\mathbf{w}}_{1,\Phi_i}^e$ from Theorem 3, and \mathbf{w}_{1,Φ_i}^e is the first column of

$$\mathbf{W}_{\Phi_i}^e = \left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T + c_1 \mathbf{I} \right)^{-1} \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \quad (36)$$

Note that the columns in $\tilde{\mathbf{H}}_{\Phi_i}^e$ are $a_j \mathbf{h}_j^e$. From Proposition 4, a_j ($j \in \Phi_i$) is approximately constant in each iteration; as a result, we can approximate $\tilde{\mathbf{H}}_{\Phi_i}^e$ by $[\tilde{\mathbf{H}}_{\Phi_{i-1}}^e, a_l \mathbf{h}_l^e]$ to reduce the complexity. Let $\mathbf{h}_l^e = a_l \mathbf{h}_l^e$, then $\tilde{\mathbf{H}}_{\Phi_i}^e = [\tilde{\mathbf{H}}_{\Phi_{i-1}}^e, \mathbf{h}_l^e]$, and we can obtain $\mathbf{W}_{\Phi_i}^e$ from $\mathbf{W}_{\Phi_{i-1}}^e$ in the $i-1$ th iteration as follows.

Lemma 3: Denote $r = \left(\mathbf{h}_l^e \right)^T \left(\mathbf{h}_l^e \right)^* - \mathbf{v}^H \mathbf{G}^{-1} \mathbf{v} + c_1$, where $\mathbf{v} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\mathbf{h}_l^e \right)^*$, $\mathbf{G} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I}$ and $c_1 > 0$. We have $r > 0$.

Proof: See Appendix E. \blacksquare

Using Lemma 3, we have the following theorem.

Theorem 5: For each i , define

$$\mathbf{W}_{\Phi_i}^e = \left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T + c_1 \mathbf{I} \right)^{-1} \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \quad (37)$$

Then, we have

$$\mathbf{W}_{\Phi_i}^e = \left[\mathbf{W}_{\Phi_{i-1}}^e - \mathbf{r}_m \mathbf{p}_m^H, \mathbf{r}_m \right] \quad (38)$$

Algorithm 2: The DSSLNR Algorithm.

Input: \mathbf{H}^e , \mathbf{H}^a , P_s/σ^2 ; **Output:** \mathbf{W}^e , \mathbf{W}^a ;
For user $k, k \in \mathcal{K}$, initialize $\mathbf{w}_{k,0}^e = (\bar{\mathbf{h}}_k^e)^*$, $\mathbf{w}_{k,0}^a = (\bar{\mathbf{h}}_k^a)^*$,
 $\Phi_0 = \emptyset$, and $\bar{\Phi}_0 = \emptyset$;
For $i = 1 : K - 1$
 Assume the i th LT (user l 's LT) is reduced in elevation;
 Denote $\mathbf{w}_{te,e}^a = \mathbf{w}_{k,i-1}^a$; Obtain $\mathbf{W}_{\Phi_i}^e$ using Theorem 5;
 Let $\mathbf{w}_{te,e}^e = \bar{\mathbf{w}}_{1,\Phi_i}^e$, where \mathbf{w}_{1,Φ_i}^e is the first column of
 $\mathbf{W}_{\Phi_i}^e$;
 Using $\mathbf{w}_{te,e}^a$ and $\mathbf{w}_{te,e}^e$ to get γ_{k,Φ_i}^e ;
 Assume the i th LT (user l 's LT) is reduced in azimuth;
 Denote $\mathbf{w}_{te,a}^e = \mathbf{w}_{k,i-1}^e$; Obtain $\mathbf{W}_{\bar{\Phi}_i}^a$ using Theorem 5;
 Let $\mathbf{w}_{te,a}^a = \bar{\mathbf{w}}_{1,\bar{\Phi}_i}^a$, where $\mathbf{w}_{1,\bar{\Phi}_i}^a$ is the first column of
 $\mathbf{W}_{\bar{\Phi}_i}^a$;
 Using $\mathbf{w}_{te,a}^e$ and $\mathbf{w}_{te,a}^a$ get $\gamma_{k,\bar{\Phi}_i}^a$;
 If $\gamma_{k,\Phi_{i-1}}^e - \gamma_{k,\Phi_i}^e < \gamma_{k,\bar{\Phi}_{i-1}}^a - \gamma_{k,\bar{\Phi}_i}^a$
 $\bar{\Phi}_i = \bar{\Phi}_{i-1}$, $\Phi_i = \Phi_{i-1} \cup l$, $\mathbf{w}_{k,i}^e = \mathbf{w}_{te,e}^e$ and
 $\mathbf{w}_{k,i}^a = \mathbf{w}_{te,e}^a$;
 else
 $\Phi_i = \Phi_{i-1}$, $\bar{\Phi}_i = \bar{\Phi}_{i-1} \cup l$, $\mathbf{w}_{k,i}^e = \mathbf{w}_{te,a}^e$ and
 $\mathbf{w}_{k,i}^a = \mathbf{w}_{te,a}^a$;
End
End

if $\tilde{\mathbf{H}}_{\Phi_i}^e = [\tilde{\mathbf{H}}_{\Phi_{i-1}}^e, \mathbf{h}_l^e]$, where $\mathbf{p}_m = (\mathbf{W}_{\Phi_{i-1}}^e)^H (\mathbf{h}_l^e)^*$, $\mathbf{r}_m = \frac{\mathbf{q}_m}{(\mathbf{h}_l^e)^T \mathbf{q}_m + c_1}$, $\mathbf{q}_m = (\mathbf{h}_l^e)^* - (\tilde{\mathbf{H}}_{\Phi_{i-1}}^e)^* \mathbf{p}_m$.

Proof: See Appendix F. ■

We can use Theorem 5 to obtain $\mathbf{W}_{\Phi_i}^e$ using $\mathbf{W}_{\Phi_{i-1}}^e$, and substitute $\mathbf{w}_{k,i}^e$ and $\mathbf{w}_{k,i}^a$ into $\mathcal{P}_{5.1}$ to get γ_{k,Φ_i}^e . If we cancel user l 's LT in the azimuth domain, $\bar{\Phi}_i = \bar{\Phi}_{i-1} \cup \{l\}$, $\Phi_i = \Phi_{i-1}$, and $\mathbf{w}_{k,i}^e = \mathbf{w}_{k,i-1}^e$. We update $\mathbf{w}_{k,i}^a$ by $\mathcal{P}_{6.2}$, which is $\bar{\mathbf{w}}_{1,\bar{\Phi}_i}^a$ from Theorem 3, and $\mathbf{w}_{1,\bar{\Phi}_i}^a$ is the first column of $\mathbf{W}_{\bar{\Phi}_i}^a$. By substituting $\mathbf{w}_{k,i}^e$ and $\mathbf{w}_{k,i}^a$ into $\mathcal{P}_{5.2}$, we get $\gamma_{k,\bar{\Phi}_i}^a$. Similarly, we can use $\mathbf{W}_{\bar{\Phi}_{i-1}}^a$ to obtain $\mathbf{W}_{\bar{\Phi}_i}^a$. If $\gamma_{k,\Phi_{i-1}}^e - \gamma_{k,\Phi_i}^e < \gamma_{k,\bar{\Phi}_{i-1}}^a - \gamma_{k,\bar{\Phi}_i}^a$, it means that canceling the LT in the i th iteration in the elevation domain will result in smaller SLNR loss. In this case, we reduce the LT in the elevation domain. Otherwise, we reduce the LT in the azimuth domain.

The DSSLNR algorithm is summarized in Algorithm 2.

C. DS Precoding in Large-Scale Systems

In this subsection, we show our DSZF and DSSLNR algorithms can achieve the same performance as the full precoding when M_x and M_y approach infinity. In the full precoding scheme, the precoding vector of user k is $\bar{\mathbf{w}}_{k,FP}$. We assume all users' channel vectors are not collinear, such that all users can be distinguished. When M_x and M_y approach infinity, \mathbf{h}_i and \mathbf{h}_j are orthogonal for $i \neq j$ [15], [25]. As a result,

$$\mathbf{H}^T \mathbf{H}^* \rightarrow \mathbf{\Lambda}$$

where $\mathbf{\Lambda}$ is a diagonal matrix. Thus, user k 's precoding vector is $\mathbf{w}_k = \bar{\mathbf{h}}_k^*$. Moreover, when M_x and M_y approach infinity, \mathbf{h}_i^e

and \mathbf{h}_j^e are orthogonal if \mathbf{h}_i^e and \mathbf{h}_j^e are not collinear, \mathbf{h}_i^a and \mathbf{h}_j^a are orthogonal if \mathbf{h}_i^a and \mathbf{h}_j^a are not collinear.

In our DSZF precoding, it is easy to verify that, if the elevation channel vector of any user is collinear with that of user k , the IL on this user is canceled in the azimuth domain. Because this user's azimuth channel vector is not collinear with and will be orthogonal to user k 's azimuth channel vector, and using the azimuth domain to cancel the IL will give a larger SNR. Similarly, if the azimuth channel vector of a user is collinear with that of user k , the IL on this user is canceled in the elevation domain. Thus, the azimuth channel vectors of users in set $\bar{\Phi}$ are not collinear with and are orthogonal to the azimuth channel vector of user k , and the elevation channel vectors of users in set Φ are not collinear with and are orthogonal to the elevation channel vector of user k .

From Subsection III-A, the elevation precoding vector \mathbf{w}_k^e of user k in the DSZF precoding is $(\bar{\mathbf{w}}_{\Phi}^{e,1})^T$, where $\mathbf{w}_{\Phi}^{e,1}$ is the first row of the matrix $(\bar{\mathbf{H}}_{\Phi}^e)^\dagger = ((\bar{\mathbf{H}}_{\Phi}^e)^H \bar{\mathbf{H}}_{\Phi}^e)^{-1} (\bar{\mathbf{H}}_{\Phi}^e)^H$, and the azimuth precoding vector \mathbf{w}_k^a is $(\bar{\mathbf{w}}_{\bar{\Phi}}^{a,1})^T$, where $\mathbf{w}_{\bar{\Phi}}^{a,1}$ is the first row of the matrix $(\bar{\mathbf{H}}_{\bar{\Phi}}^a)^\dagger = ((\bar{\mathbf{H}}_{\bar{\Phi}}^a)^H \bar{\mathbf{H}}_{\bar{\Phi}}^a)^{-1} (\bar{\mathbf{H}}_{\bar{\Phi}}^a)^H$. Note that the first columns of $\bar{\mathbf{H}}_{\Phi}^e$ and $\bar{\mathbf{H}}_{\bar{\Phi}}^a$ are \mathbf{h}_k^e and \mathbf{h}_k^a , respectively, and they are orthogonal to the remaining columns of $\bar{\mathbf{H}}_{\Phi}^e$ and $\bar{\mathbf{H}}_{\bar{\Phi}}^a$ respectively. Matrices $(\bar{\mathbf{H}}_{\Phi}^e)^H \bar{\mathbf{H}}_{\Phi}^e$ and $(\bar{\mathbf{H}}_{\bar{\Phi}}^a)^H \bar{\mathbf{H}}_{\bar{\Phi}}^a$ have the following form,

$$(\bar{\mathbf{H}}_{\Phi}^e)^H \bar{\mathbf{H}}_{\Phi}^e \rightarrow \begin{pmatrix} a_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \end{pmatrix}, (\bar{\mathbf{H}}_{\bar{\Phi}}^a)^H \bar{\mathbf{H}}_{\bar{\Phi}}^a \rightarrow \begin{pmatrix} a_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}$$

where a_1 and a_2 are the constants. It is easy to verify that

$$\begin{aligned} ((\bar{\mathbf{H}}_{\Phi}^e)^H \bar{\mathbf{H}}_{\Phi}^e)^{-1} &\rightarrow \begin{pmatrix} a_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1^{-1} \end{pmatrix} \\ ((\bar{\mathbf{H}}_{\bar{\Phi}}^a)^H \bar{\mathbf{H}}_{\bar{\Phi}}^a)^{-1} &\rightarrow \begin{pmatrix} a_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^{-1} \end{pmatrix} \end{aligned}$$

based on which one can obtain

$$\mathbf{w}_k^e \rightarrow (\bar{\mathbf{h}}_k^e)^*, \mathbf{w}_k^a \rightarrow (\bar{\mathbf{h}}_k^a)^*. \quad (39)$$

Considering the DSSLNR algorithm, and using Proposition 4, when M_x and M_y tend to infinity, we have

$$\mathbf{w}_{k,i}^e \rightarrow (\bar{\mathbf{h}}_k^e)^*, \mathbf{w}_{k,i}^a \rightarrow (\bar{\mathbf{h}}_k^a)^*$$

in each iteration. At the end of DSSLNR algorithm, we have

$$\mathbf{w}_k^e \rightarrow (\bar{\mathbf{h}}_k^e)^*, \mathbf{w}_k^a \rightarrow (\bar{\mathbf{h}}_k^a)^*$$

which shows that our DSSLNR precoding algorithm asymptotically approaches the full precoding scheme.

IV. DS PRECODING IN ONE-RING CHANNEL MODEL

We consider the one-ring channel model [16]–[18], [34]–[37] in this section, due to its meaningful geometrical interpretation. This model describes the case when a BS is elevated away from scatterers and communicates with a mobile user that is surrounded by a ring of scatterers. Assuming all paths that reach one user are equal in power [16]–[18], [34]–[37], the channel

response of the l th path (scatterer) from the BS to user k is represented by

$$\mathbf{h}_k^l = \frac{\eta_k e^{j\tau_{k,l}}}{\sqrt{L}} \mathbf{a}(\theta_{k,l})^e \otimes \mathbf{a}(\theta_{k,l}, \phi_{k,l})^a$$

where η_k represents the amplitude of each path, each $\tau_{k,l}$ is uniformly distributed in $[-\pi, \pi]$, $\theta_{k,l} \in [\theta_k - \Delta_E/2, \theta_k + \Delta_E/2]$, and $\phi_{k,l} \in [\phi_k - \Delta_A/2, \phi_k + \Delta_A/2]$, with Δ_E and Δ_A being the angular spreads (ASs) in the elevation and azimuth domains, respectively. θ_k and ϕ_k are determined by the location of the ring. Δ_E and Δ_A are determined by the ring's radius and the distance between the BS and the ring. Then user k 's channel vector is given by

$$\mathbf{h}_k = \sum_{l=1}^L \mathbf{h}_k^l = \frac{\eta_k}{\sqrt{L}} \sum_{l=1}^L e^{j\tau_{k,l}} \mathbf{a}(\theta_{k,l})^e \otimes \mathbf{a}(\theta_{k,l}, \phi_{k,l})^a$$

and it is easy to verify that $\mathbb{E}(\|\mathbf{h}_k\|^2) = \eta_k^2$, where the amplitude η_k is normalized as 1. In fact, the elevation (azimuth) AOA has different distributions in different scenarios. Similar to the one-ring model in [16]–[18], the elevation and azimuth AOAs are assumed to be independent and they are uniformly distributed in $[\theta_k - \Delta_E/2, \theta_k + \Delta_E/2]$ and $[\phi_k - \Delta_A/2, \phi_k + \Delta_A/2]$, respectively (It is shown that similar performances and asymptotic behaviors are achieved by other AoA distributions [16]). In reference channel model [34], [35], the number of paths are assumed to be infinity, while it is not realizable in reality. Here, we use the simulation channel model [36]–[38], where the number of paths is finite in our paper.

In one-ring channel model, we focus on the DSSLNR algorithm, since the noise effect is incorporated and better performance can be achieved. From (26), the SLNR of user k is β_k . Denote the 2D representation of the channel vector \mathbf{h}_k as \mathbf{H}_k , which is expressed as

$$\mathbf{H}_k = \frac{\eta_k}{\sqrt{L}} \sum_{l=1}^L e^{j\tau_{k,l}} \mathbf{a}(\theta_{k,l}, \phi_{k,l})^a (\mathbf{a}(\theta_{k,l})^e)^T. \quad (40)$$

It is easy to verify $\text{vec}(\mathbf{H}_k) = \mathbf{h}_k$, where $\text{vec}(\cdot)$ denotes the vectorization operation. If $\mathbf{w}_k = \mathbf{w}_k^e \otimes \mathbf{w}_k^a$, we have

$$\mathbf{h}_j^T \mathbf{w}_k = (\mathbf{w}_k^a)^T \mathbf{H}_j \mathbf{w}_k^e. \quad (41)$$

Substituting (41) into β_k and denoting $\gamma_k = \beta_k^{-1}$, the SLNR maximization problem can be equivalently written as

$$\min_{\|\mathbf{w}_k^e\|=1, \|\mathbf{w}_k^a\|=1} \gamma_k = \sum_{j \neq k} \frac{|(\mathbf{w}_k^a)^T \mathbf{H}_j \mathbf{w}_k^e|^2}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2} + \frac{c}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2}$$

where $c = \frac{\sigma^2}{E_s}$. Motivated by the DS scheme in the single-path channel, the LTs are divided into two parts, and are removed in either the elevation domain or the azimuth domain, respectively. We need to find two sets Φ and $\bar{\Phi}$, where the j th ($j \in \Phi$) LT and the l th ($l \in \bar{\Phi}$) LT are reduced by the elevation precoding vector

\mathbf{w}_k^e and the azimuth vector \mathbf{w}_k^a , respectively. We re-write γ_k as

$$\gamma_k = \sum_{j \in \Phi} \frac{|(\mathbf{g}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{g}_k^e)^T \mathbf{w}_k^e|^2} + \sum_{j \in \bar{\Phi}} \frac{|(\mathbf{w}_k^a)^T \mathbf{g}_j^a|^2}{|(\mathbf{w}_k^a)^T \mathbf{g}_k^a|^2} + \frac{c}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2} \quad (42)$$

where $(\mathbf{g}_j^e)^T = (\mathbf{w}_k^a)^T \mathbf{H}_j$ and $\mathbf{g}_j^a = \mathbf{H}_j \mathbf{w}_k^e$.

In the one-ring channel model, as the angular spread is usually not large, we have $\mathbf{a}(\theta_{k,l}, \phi_{k,l})^a \approx \mathbf{a}(\theta_k, \phi_k)^a$, and $\mathbf{a}(\theta_{k,l})^e \approx \mathbf{a}(\theta_k)^e$. Hence, $\mathbf{H}_k \approx (\frac{\eta_k}{\sqrt{L}} \sum_{l=1}^L e^{j\tau_{k,l}}) \mathbf{a}(\theta_k, \phi_k)^a (\mathbf{a}(\theta_k)^e)^T$, and $|(\mathbf{w}_k^a)^T \mathbf{H}_j \mathbf{w}_k^e|^2 \approx |(\frac{\eta_k}{\sqrt{L}} \sum_{l=1}^L e^{j\tau_{k,l}}) (\mathbf{w}_k^a)^T \mathbf{a}(\theta_j, \phi_j)^a (\mathbf{a}(\theta_j)^e)^T \mathbf{w}_k^e|^2$. Hence,

$$\frac{|(\mathbf{w}_k^a)^T \mathbf{H}_j \mathbf{w}_k^e|^2}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2} \approx p \frac{|(\mathbf{w}_k^a)^T \mathbf{a}(\theta_j, \phi_j)^a|^2 |(\mathbf{a}(\theta_j)^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{w}_k^a)^T \mathbf{a}(\theta_k, \phi_k)^a|^2 |(\mathbf{a}(\theta_k)^e)^T \mathbf{w}_k^e|^2} \quad (43)$$

where $p = (\sum_{l=1}^L e^{j\tau_{j,l}} / \sum_{l=1}^L e^{j\tau_{k,l}})^2$. Note that the LT is also affected by two multipliers. We can use either elevation or azimuth precoding vector to reduce each multiplier, to reduce the LT. The analysis is similar to that in the single-path scenario. We briefly summarize the process in the following.

We determine Φ , $\bar{\Phi}$, \mathbf{w}_k^e and \mathbf{w}_k^a in an iterative manner, i.e., we considered the LTs in γ_k iteratively. In each iteration, if minimizing the LT in the elevation domain has larger SLNR, this LT belongs to Φ , and we use the elevation precoding vector \mathbf{w}_k^e to minimize it. Otherwise, we minimize the LT in the azimuth domain, and the vector \mathbf{w}_k^a is updated via iteration. The process is as follows.

At beginning, denote the SVD of \mathbf{H}_k as $\mathbf{H}_k = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H$. Initialize $\Phi_0 = \emptyset$, $\bar{\Phi}_0 = \emptyset$. To maximize $|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2$, let

$$\mathbf{w}_{k,0}^e = \bar{\mathbf{v}}_1, \mathbf{w}_{k,0}^a = \bar{\mathbf{u}}_1^* \quad (44)$$

where λ_1 is the largest singular value, \mathbf{u}_1 and \mathbf{v}_1 are the columns in \mathbf{U} and \mathbf{V} corresponding to λ_1 respectively. Suppose we have Φ_{i-1} and $\bar{\Phi}_{i-1}$ in the $i-1$ iteration, where $\Phi_{i-1} \cap \bar{\Phi}_{i-1} = \emptyset$ and $\Phi_{i-1} \cup \bar{\Phi}_{i-1} = \Omega_{i-1}$. In the i th iteration, we determine in which domain user l 's LT $\frac{|(\mathbf{w}_k^e)^T \mathbf{H}_l \mathbf{w}_k^e|^2}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2}$ will be reduced such that $\gamma_{k,i}$ is minimized, where

$$\gamma_{k,i} = \sum_{j \in \Phi_i} \frac{|(\mathbf{g}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{g}_k^e)^T \mathbf{w}_k^e|^2} + \sum_{j \in \bar{\Phi}_i} \frac{|(\mathbf{w}_k^a)^T \mathbf{g}_j^a|^2}{|(\mathbf{w}_k^a)^T \mathbf{g}_k^a|^2} + \frac{c}{|(\mathbf{w}_k^a)^T \mathbf{H}_k \mathbf{w}_k^e|^2}.$$

If the LT on user l is reduced in the elevation domain, $\Phi_i = \Phi_{i-1} \cup l$ and $\bar{\Phi}_i = \bar{\Phi}_{i-1}$. Then we use \mathbf{w}_k^e to reduce this LT, and \mathbf{w}_k^a remains unchanged. The problem to determine \mathbf{w}_k^e is

$$\mathcal{P}_{7.1} : \min_{\|\mathbf{w}_k^e\|=1} \gamma_{k,\Phi_i} = \sum_{j \in \Phi_i} \frac{|(\mathbf{g}_j^e)^T \mathbf{w}_k^e|^2}{|(\mathbf{g}_k^e)^T \mathbf{w}_k^e|^2} + \frac{c}{|(\mathbf{g}_k^e)^T \mathbf{w}_k^e|^2}. \quad (45)$$

Note that $\mathcal{P}_{7.1}$ can be rewritten as

$$\max_{\|\mathbf{w}_k^e\|=1} \frac{(\mathbf{w}_k^e)^H \left((\mathbf{g}_k^e)^* (\mathbf{g}_k^e)^T \right) \mathbf{w}_k^e}{(\mathbf{w}_k^e)^H \left(\sum_{j \in \Phi_i} (\mathbf{g}_j^e)^* (\mathbf{g}_j^e)^T \right) \mathbf{w}_k^e + c}. \quad (46)$$

The objective function in (46) is a Rayleigh quotient, and the optimal solution of (46) is $\bar{\mathbf{w}}_{k,\Phi_i}^e$, where

$$\mathbf{w}_{k,\Phi_i}^e = \left((\mathbf{G}_{\Phi_i}^e)^* (\mathbf{G}_{\Phi_i}^e)^T + c\mathbf{I} \right)^{-1} (\mathbf{g}_k^e)^* \quad (47)$$

and $\mathbf{G}_{\Phi_i}^e$ denotes the matrix whose columns are $\mathbf{g}_j^e, j \in \Phi_i$. Substituting $\mathbf{w}_{k,\bar{\Phi}_i}^a$ and \mathbf{w}_{k,Φ_i}^e into γ_{k,Φ_i}^e in $\mathcal{P}_{9.1}$, we have γ_{k,Φ_i}^e . Similar to the single-path scenario, we approximate $\mathbf{G}_{\Phi_i}^e = [\mathbf{G}_{\Phi_{i-1}}^e \mathbf{g}_l^e]$ to reduce complexity. Denoting $\mathbf{A}_{\Phi_i} = ((\mathbf{G}_{\Phi_i}^e)^* (\mathbf{G}_{\Phi_i}^e)^T + c_1\mathbf{I})^{-1}$, \mathbf{A}_{Φ_i} can be obtained by $\mathbf{A}_{\Phi_{i-1}}$ as

$$\begin{aligned} \mathbf{A}_{\Phi_i} &= \left(\mathbf{A}_{\Phi_{i-1}} + (\mathbf{g}_l^e)^* (\mathbf{g}_l^e)^T \right)^{-1} \\ &= \mathbf{A}_{\Phi_{i-1}}^{-1} - \frac{\mathbf{A}_{\Phi_{i-1}}^{-1} (\mathbf{g}_l^e)^* (\mathbf{g}_l^e)^T \mathbf{A}_{\Phi_{i-1}}^{-1}}{1 + (\mathbf{g}_l^e)^T \mathbf{A}_{\Phi_{i-1}}^{-1} (\mathbf{g}_l^e)^*}. \end{aligned} \quad (48)$$

Alternatively, if the LT on user l is minimized in the azimuth domain, $\bar{\Phi}_i = \bar{\Phi}_{i-1} \cup l$, $\Phi_i = \Phi_{i-1}$, and \mathbf{w}_k^e remains unchanged. Hence, we use \mathbf{w}_k^a to minimize this LT, leading to

$$\mathcal{P}_{7.2}: \min_{\|\mathbf{w}_k^a\|=1} \gamma_{k,\bar{\Phi}_i}^a = \sum_{j \in \bar{\Phi}_i} \frac{|(\mathbf{w}_k^a)^T \mathbf{g}_j^a|^2}{|(\mathbf{w}_k^a)^T \mathbf{g}_k^a|^2} + \frac{c}{|(\mathbf{w}_k^a)^T \mathbf{g}_k^a|^2}. \quad (49)$$

Similar to the elevation domain, its optimal solution is $(\bar{\mathbf{w}}_{k,\bar{\Phi}_i}^a)^*$, where

$$\mathbf{w}_{k,\bar{\Phi}_i}^a = \left(\mathbf{G}_{\bar{\Phi}_i}^a (\mathbf{G}_{\bar{\Phi}_i}^a)^H + c\mathbf{I} \right)^{-1} \mathbf{g}_k^a \quad (50)$$

with $\mathbf{G}_{\bar{\Phi}_i}^a$ being the matrix whose columns are $\mathbf{g}_j^a, j \in \bar{\Phi}_i$. Substituting $\mathbf{w}_{k,\bar{\Phi}_i}^a$ and \mathbf{w}_{k,Φ_i}^e in $\mathcal{P}_{7.2}$, we have $\gamma_{k,\bar{\Phi}_i}^a$. Also, we approximate $\mathbf{G}_{\bar{\Phi}_i}^a = [\mathbf{G}_{\bar{\Phi}_{i-1}}^a \mathbf{g}_l^a]$ to reduce complexity. Then, $(\bar{\mathbf{G}}_{\bar{\Phi}_i}^a (\bar{\mathbf{G}}_{\bar{\Phi}_i}^a)^H + c_1\mathbf{I})^{-1}$ can be obtained by $(\bar{\mathbf{G}}_{\bar{\Phi}_{i-1}}^a (\bar{\mathbf{G}}_{\bar{\Phi}_{i-1}}^a)^H + c_1\mathbf{I})^{-1}$, which is similar to (48). If $\gamma_{k,\bar{\Phi}_{i-1}}^a - \gamma_{k,\Phi_i}^e < \gamma_{k,\bar{\Phi}_{i-1}}^a - \gamma_{k,\bar{\Phi}_i}^a$, canceling the LT in the i th iteration in the elevation domain will result in a larger SLNR. Otherwise, we cancel the LT in the azimuth domain. When all the terms in γ_k are considered, stop the iteration. The DSSLNR algorithm in one-ring channel is given in Algorithm 3.

V. SIMULATION RESULTS AND COMPLEXITY ANALYSIS

In this section, we evaluate the performance of our proposed DS precoding algorithms based on simulations. Consider a single-cell system equipped with a rectangular array at a height $H_{BS} = 35$ m and serving K single-antenna users with cell radius r_c . The users are randomly and uniformly distributed in the cell. Each user antenna is located at a height of 1.5 m.

Algorithm 3: The DSSLNR in One-Ring Channel.

Input: $\mathbf{H}_k (k \in \mathcal{K}), P_s/\sigma^2$; **Output:** $\mathbf{W}^e, \mathbf{W}^a$;
For user $k, k \in \mathcal{K}$, initialize $\mathbf{w}_{k,0}^e$ and $\mathbf{w}_{k,0}^a$ by (44),
 $\Phi_0 = \emptyset, \bar{\Phi}_0 = \emptyset$;

For $i = 1 : K - 1$

Assume the i th LT (user l 's LT) is reduced in elevation;
Denote $\mathbf{w}_{te,e}^a = \mathbf{w}_{k,i-1}^a$; Obtain $(\mathbf{g}_l^e)^T = (\mathbf{w}_{te,e}^a)^T \mathbf{H}_l$;
Let $(\mathbf{g}_k^e)^T = (\mathbf{w}_{te,e}^a)^T \mathbf{H}_k$; Use (47), (48) to get \mathbf{w}_{1,Φ_i}^e ;
Let $\mathbf{w}_{te,e}^e = \bar{\mathbf{w}}_{1,\Phi_i}^e$, and use $\mathbf{w}_{te,e}^a$ and $\mathbf{w}_{te,e}^e$ to get γ_{k,Φ_i}^e ;
Assume the i th LT (user l 's LT) is reduced in azimuth;

Denote $\mathbf{w}_{te,a}^e = \mathbf{w}_{k,i-1}^e$; Obtain $\mathbf{g}_l^a = \mathbf{H}_l \mathbf{w}_{te,a}^e$;

Let $\mathbf{g}_l^a = \mathbf{H}_l \mathbf{w}_{te,a}^e$; Use (50) to get $\mathbf{w}_{k,\bar{\Phi}_i}^a$;

Let $\mathbf{w}_{te,a}^a = \bar{\mathbf{w}}_{1,\bar{\Phi}_i}^a$, and use $\mathbf{w}_{te,a}^e$ and $\mathbf{w}_{te,a}^a$ get $\gamma_{k,\bar{\Phi}_i}^a$;

If $\gamma_{k,\bar{\Phi}_{i-1}}^e - \gamma_{k,\Phi_i}^e < \gamma_{k,\bar{\Phi}_{i-1}}^a - \gamma_{k,\bar{\Phi}_i}^a$

$\bar{\Phi}_i = \bar{\Phi}_{i-1}, \Phi_i = \Phi_{i-1} \cup l, \mathbf{w}_{k,i}^e = \mathbf{w}_{te,e}^e$ and

$\mathbf{w}_{k,i}^a = \mathbf{w}_{te,e}^a$;

else

$\bar{\Phi}_i = \bar{\Phi}_{i-1}, \bar{\Phi}_i = \bar{\Phi}_{i-1} \cup l, \mathbf{w}_{k,i}^e = \mathbf{w}_{te,a}^e$ and

$\mathbf{w}_{k,i}^a = \mathbf{w}_{te,a}^a$;

End

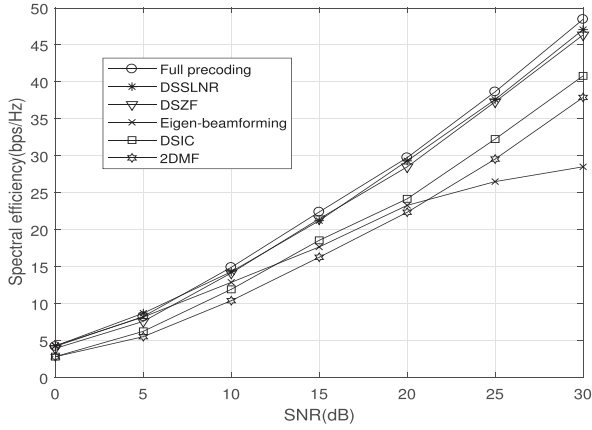
End

A. Results With Single-Path Channels

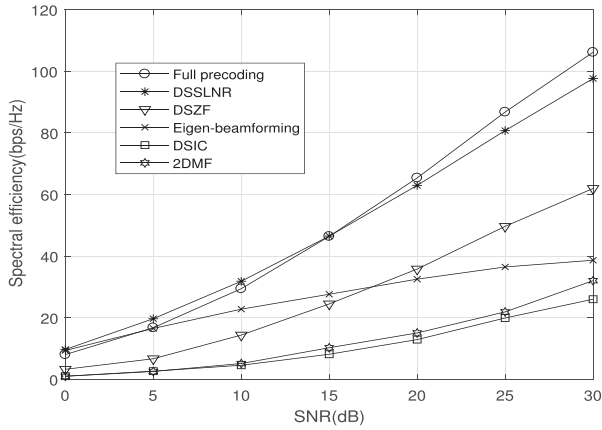
The azimuth and elevation angles are determined based on user locations relative to the BSs. The complex path gain $\alpha_{k,l}$ is normalized as a zero mean Gaussian variable with unit variance. Our DSZF and DSSLNR schemes are compared to 2D MF [21], DSIC [23], eigen-beamforming [19], [39] and the full precoding [5] schemes.

1) *Comparison With Perfect CSI:* In single-path TDD massive MIMO systems, angular-based channel estimation techniques [39], [40] are usually utilized to obtain the complex path gain α_k and the elevation and azimuth AOAs, which can be used to obtain \mathbf{h}_k^e and \mathbf{h}_k^a (Input in Algorithm 1 and 2). The channel estimation error is related with the length of the training sequence and the transmission power, and nearly perfect instantaneous CSI can be obtained when the training sequence is long and the transmission power is large. Here we give the comparison of different algorithms with perfect CSI.

Figs. 1–2 show the achievable rates of our proposed DS precoding algorithms under different SNRs with $r_c = 200$ m. The BS is equipped with a 16×16 antenna array in Fig. 1 and a 30×16 antenna array in Fig. 2. Both Figs. 1 and 2 show that the achievable rate of all the precoding scheme increases as the SNR increases. More importantly, our DSZF and DSSLNR schemes can achieve larger achievable rates than the existing 2D precoding schemes and the eigen-beamforming scheme. Also, DSSLNR can achieve a higher achievable rate than the DSZF schemes due to the fact that the DSSLNR scheme takes the noise into consideration. Moreover, the DSSLNR scheme can approach the full precoding algorithm. The eigen-beamforming deteriorates at high SNR. This is because the eigen-beamforming can not cancel the inter-user interference well, which dominates in high SNR regions.



(a)



(b)

Fig. 1. Comparison of the achievable rates for different SNRs. The BS is assumed to employ 16×16 URAs. (a) $K = 8$. (b) $K = 20$.

Fig. 3 shows the achievable rates of our proposed DS precoding schemes under different numbers of users. The achievable rate of the DSSLNR scheme increases as the number of the users increases, which is similar to that of eigen-beamforming and full precoding schemes. The achievable rate of the DSZF scheme first increases then decreases as the number of the users increases, due to the large inter-user interference. This phenomenon is similar to the ZF precoding scheme in massive MIMO systems, since our DSZF scheme divides the precoding problem into the ZF precoding in the elevation and azimuth domains. The 2D MF performs the ZF precoding on all users' elevation channel vectors, and when the number of users (K) is smaller than M_x , the spectral efficiency performs similar to the ZF precoding, i.e., the spectral efficiency first increases, and then decreases with increasing number of users. When K is larger than M_x , the spectral efficiency increases as the number of users increases. The spectral efficiency of DSIC performs similarly to that of the 2D MF.

In Fig. 4, we evaluate the impact of the number of elevation antennas, using the same system and channel parameters as in Fig. 3. The achievable rates of all the DS-based precoding schemes, the eigen-beamforming scheme and the full precoding scheme increases as the number of elevation antennas increases,

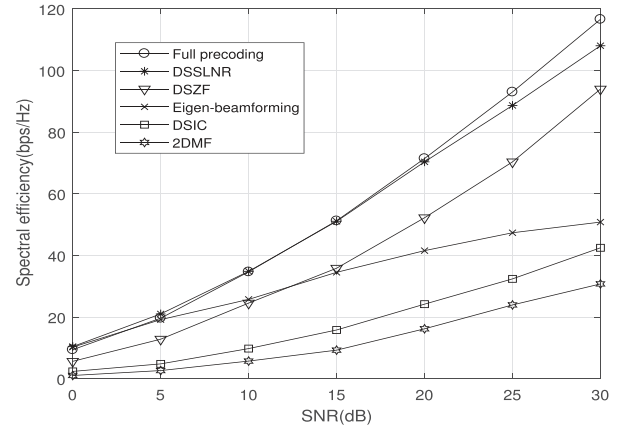


Fig. 2. Comparison of the achievable rates for different SNRs. The BS is assumed to employ 30×16 URAs, and $K = 20$.

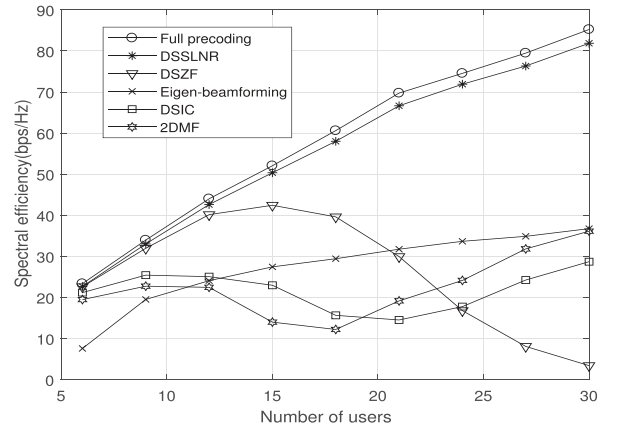


Fig. 3. Comparison of the achievable rates for different number of users. The BS is assumed to employ 16×16 URAs, and SNR = 20 dB.

while that of the 2D MF remains unchanged. Our DSSLNR scheme has the largest achievable rate among the 2D precoding schemes, and the achievable rate of DSSLNR approaches that of the full precoding scheme.

2) *Comparison With Imperfect CSI*: In massive MIMO systems, the training interval length and the transmit power are limited. In this case, there will exist estimation errors. Next, we show the impact of imperfect CSI to our proposed precoding schemes. In single-path scenario, using angular-based channel estimation techniques, we obtain the estimation of the complex gain α_k as $\tilde{\alpha}_k$, and the estimations of elevation and azimuth AOs as $\tilde{\theta}_k$ and $\tilde{\phi}_k$, respectively, then the channel vector of user k obtained by the BS is given by

$$\hat{\mathbf{h}}_k = \tilde{\alpha}_k \mathbf{a}(\tilde{\theta}_k)^e \otimes \mathbf{a}(\tilde{\theta}_k, \tilde{\phi}_k)^a. \quad (51)$$

We denote $\tilde{\alpha}_k = \alpha_k + \varepsilon_k$, $\tilde{\theta}_k = \theta_k + \delta_e$ and $\tilde{\phi}_k = \phi_k + \delta_a$, where ε_k , δ_e , and δ_a are the estimation errors in path gain, elevation AOA, and azimuth AOA, respectively. The error ε_k in path gain is assumed to be zero mean Gaussian variable [41] with variance σ_ε^2 . δ_e and δ_a are assumed to be uniformly distributed in $[-\delta/2, \delta/2]$ where δ is a constant [41]. In sparse channel model,

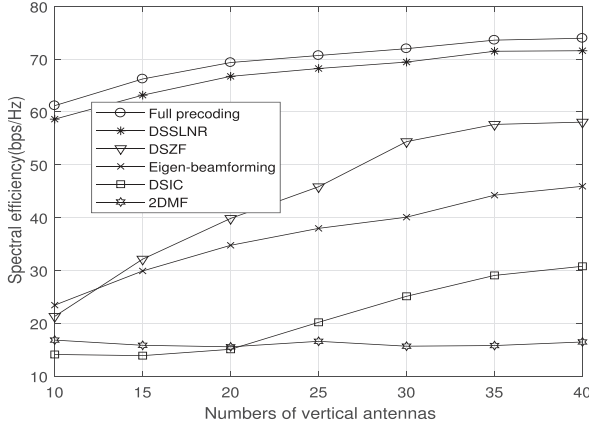


Fig. 4. Comparison of the achievable rates for different number of vertical antennas. $M_y = 16$, $K = 20$, and $\text{SNR} = 20$ dB.

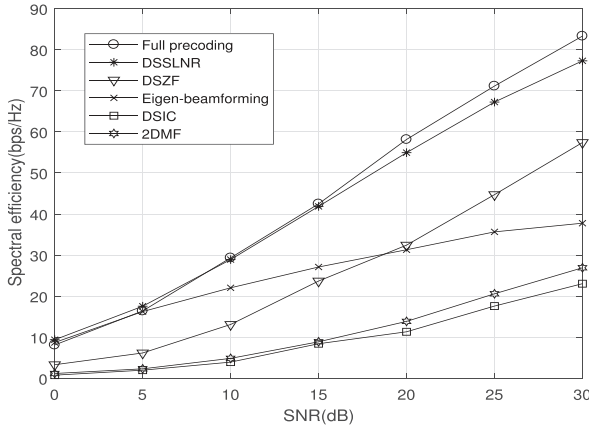


Fig. 5. Comparison of the achievable rates for different SNRs with $\sigma_\varepsilon^2 = 0.01$ and $\delta = 2^\circ$. The BS is assumed to employ 16×16 URAs and $K = 20$.

the channel estimation error can be very small through effective channel estimation techniques [39], [40].

Fig. 5 and 6 compare the achievable rates under imperfect CSI, using the same system and channel parameters as in Fig. 1(b). In Fig. 5, we give the comparison of different precoding algorithms under different SNRs with $\delta = 2^\circ$, $\sigma_\varepsilon^2 = 0.01$. Similar to the comparison with perfect CSI, the achievable rates of all precoding schemes increases as the SNR increases, and our DSZF and DSSLNR schemes can achieve a larger achievable rate than the existing 2D precoding schemes and the eigen-beamforming scheme. The DSSLNR scheme also can approach the full precoding algorithm with imperfect CSI. Fig. 6 compares the spectral efficiency of our DS-based precoding algorithms with perfect CSI and imperfect CSI under different δ , and σ_ε^2 is set to 0.01 in imperfect CSI situation. Our DSSLNR algorithm can achieve larger achievable rate than the DSZF algorithm. With the increase of δ , the estimation error increases, and the interference becomes larger, resulting in lower achievable rate.

B. Results With One-Ring Channels

For the one-ring channel model, the azimuth and elevation angles are geometrically determined based on user locations

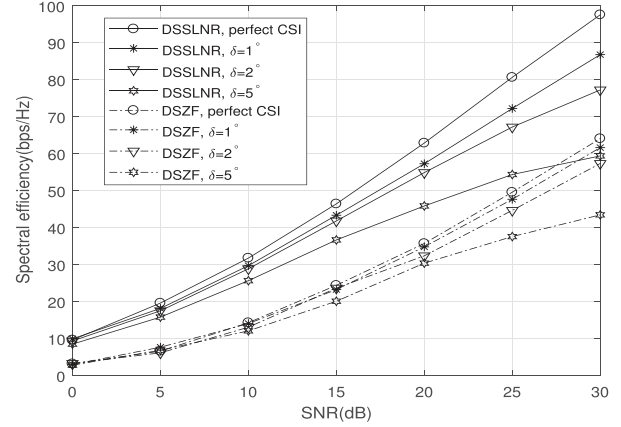


Fig. 6. Comparison of the achievable rates of DS-based algorithms for different SNRs with different δ . The BS is assumed to employ 16×16 URAs, $K = 20$ and $\sigma_\varepsilon^2 = 0.01$.

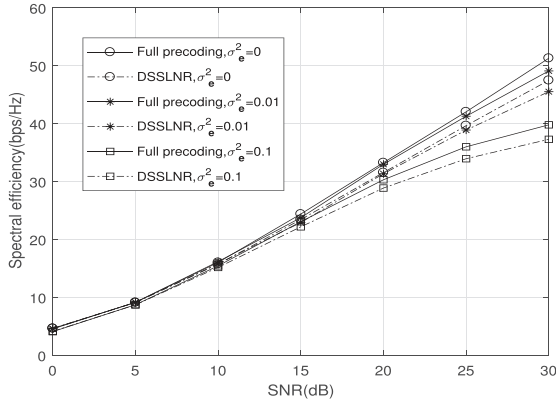
relative to the BS, and the angular spreads in the elevation and azimuth domain are set to Δ_A and Δ_E respectively. The BS randomly serves $K = 20$ users. The number of paths is set to 25 [36]. From Algorithm 3, we need to obtain \mathbf{H}_k (i.e., \mathbf{h}_k) for each user, which can be obtained by linear MMSE based channel estimations [42] or channel covariance matrix based channel estimations [43]. Our DSSLNR is compared with the full precoding scheme [5], and both the perfect and imperfect CSI are considered. In one-ring channel model, the channel vector of user k obtained by the BS is given by

$$\hat{\mathbf{h}}_k = \mathbf{h}_k + \mathbf{e}_k \quad (52)$$

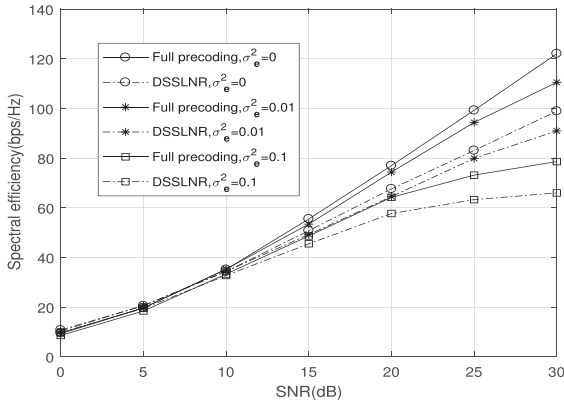
where \mathbf{e}_k is the estimation error. For simplicity, \mathbf{e}_k is assumed to be distributed as a zero mean Gaussian random variable [15]. Note that $\mathbb{E}(\|\mathbf{h}_k\|^2) = 1$. Here the covariance matrix of \mathbf{e}_k is set to $\frac{\sigma_\varepsilon^2}{M_x M_y} \mathbf{I}$, such that $\mathbb{E}(\|\mathbf{e}_k\|^2) = \sigma_\varepsilon^2$, where $\sigma_\varepsilon^2 = 0$ means perfect CSI.

Figure 7 shows the achievable rate of our proposed domain selective precoding scheme under different SNRs with both perfect and imperfect CSI, where the BS is equipped with 16×16 antennas. The ASs are set to 5° in both the elevation and azimuth domains. It is shown that the achievable rate of our DSSLNR precoding algorithm increases as the SNR increases. Moreover, our DSSLNR algorithm can achieve a spectral efficiency similar to that of the full precoding scheme when the number of users is small, while it has a spectral efficiency gap from that of the full precoding scheme when the number of users becomes large. In fact, due to the Kronecker structure of the precoding vector, the 2D precoding scheme performs worse than the full precoding algorithm. Moreover, the achievable rates of both our DSSLNR and full precoding algorithms decrease as the channel estimation error becomes large.

Fig. 8 shows the achievable rate of our proposed DSSLNR algorithm under different ASs, where the BS employs a 16×16 antenna array, $K = 20$, and the SNR is 20 dB. The elevation AS and azimuth AS are the same. The DSSLNR performs worse than the full precoding scheme does. The achievable rate of the DSSLNR algorithm decreases as the AS increases. The larger the AS is, the worse the performance of the 2D precoding



(a)



(b)

Fig. 7. Comparison of the achievable rates for different SNRs. The BS is assumed to employ 16×16 URAs and $\Delta_A = \Delta_E = 5^\circ$. (a) $K = 8$. (b) $K = 20$.

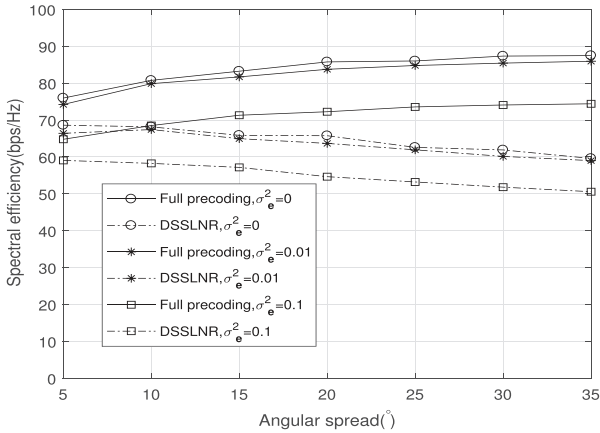


Fig. 8. Comparison of the achievable rates different angular spread. The BS is assumed to employ 16×16 URAs, SNR=20 dB, and $K = 20$.

algorithms will be. This is because, when the AS becomes large, the Kronecker structure of the channel matrix becomes less accurate, and the Kronecker structure based precoding scheme gives a worse performance. Similar to Fig. 7, with the increase of channel estimation error, the achievable rates of both our DSSLNR and full precoding algorithms decrease.

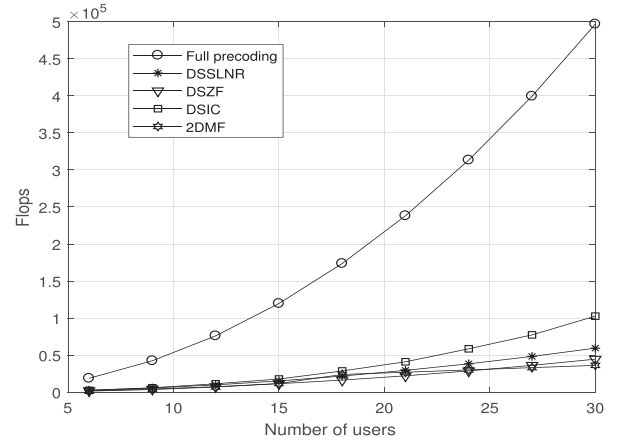


Fig. 9. Complexity comparison for different number of users with $M_x = M_y = 16$.

C. Complexity Analysis

In this subsection, we analyze the complexity of our DS-based precoding algorithms. Since the precoding vectors of different users are calculated in parallel, we denote the computational complexity as the largest one among the complexities of calculating all users' precoding vectors. For each user, the calculation of the precoding vector requires $K - 1$ iterations.

First of all, we consider the DSZF algorithm in a single-path scenario. In the i th iteration, when calculating $(\bar{\mathbf{H}}_{\Phi_i}^e)^\dagger$, we need to calculate \mathbf{d}_m which costs $2M_x|\Phi_{i-1}|$ floating point operations (flops) [44]. If $\mathbf{h}_l^e - \bar{\mathbf{H}}_{\Phi_{i-1}}^e \mathbf{d}_m \neq 0$, calculating $\mathbf{h}_l^e - \bar{\mathbf{H}}_{\Phi_{i-1}}^e \mathbf{d}_m$ will cost $2M_x|\Phi_{i-1}| + M_x$ flops, and obtaining \mathbf{b}_m will cost $2M_x|\Phi_{i-1}| + 4M_x$ flops. Then $(\bar{\mathbf{H}}_{\Phi_i}^e)^\dagger$ costs $6M_x|\Phi_{i-1}| + 4M_x$ flops. If $\mathbf{h}_l^e - \bar{\mathbf{H}}_{\Phi_{i-1}}^e \mathbf{d}_m = 0$, obtaining \mathbf{b}_m will cost $6M_x|\Phi_{i-1}| + 3M_x + 1$ flops, and the complexity of obtaining $(\bar{\mathbf{H}}_{\Phi_i}^e)^\dagger$ requires $8M_x|\Phi_{i-1}| + 3M_x + 1$ flops. Similar complexity will be required for the azimuth precoding. Obtaining $p^{e,i}$ and $p^{a,i}$ will cost $2M_x$ and $2M_y$ flops, respectively. Summing all the components for each iteration, we get the total complexity.

Next, we consider the complexity of DSSLNR algorithm. In the i th iteration, to obtain $\mathbf{W}_{\Phi_i}^e$, we need to calculate \mathbf{p}_m which costs $2M_x|\Phi_{i-1}|$ flops. Calculating \mathbf{q}_m costs $2M_x|\Phi_{i-1}| + M_x$ flops, and obtaining \mathbf{r}_m costs $3M_x + 1$ flops. The complexity of obtaining $\mathbf{W}_{\Phi_i}^e$ entails $6M_x|\Phi_{i-1}| + 4M_x + 1$ flops. Similar complexity is required for the azimuth precoding. Summing all components for each iteration, we get the total complexity. In Fig. 9, we show the comparison of the number of flops of each precoding algorithm. Our precoding algorithm has a similar complexity with other 2D precoding algorithms, while the full precoding has much higher complexity.

In the one-ring channel model, the complexity in the initial stage is $O(M_x^2 M_y)$, and the complexity in each iteration is $O(M_x^2 + M_y^2)$. Since there are $K - 1$ iterations, the complexity of DSSLNR algorithm in the one-ring channel scenario is $O((K - 1)(M_x^2 + M_y^2) + M_x^2 M_y)$. From the above observations, the complexity of our DSSLNR is cubic, which is much smaller than that of the full precoding algorithm.

VI. CONCLUSION

In this paper, we have proposed two DS-based precoding scheme termed as DSZF and DSSLNR for 3D massive MIMO systems. We have mathematically analyzed the two DS precoding algorithms, and derived one efficient method to select a better domain to cancel the interference. Moreover, we have theoretically proved that our DS precoding algorithms asymptotically approach the full precoding scheme when the number of antennas becomes infinity. Simulation results have demonstrated that our proposed algorithms can achieve better spectral efficiency performance than existing 2D precoding with a low computational complexity. In single-path and one-ring channel 3D TDD massive MIMO systems, since the instantaneous CSI is available at the BS, the proposed DS precoding algorithms can be applied to improve the spectral efficiency at a low computational complexity. However, in the FDD massive MIMO systems, obtaining the instantaneous CSI is very expensive. Therefore, using statistical CSI to design precoding should be an efficient way to reduce the cost of channel estimation. For our future work, we will design the DS precoding scheme utilizing the statistical CSI in FDD 3D massive MIMO systems.

APPENDIX

A. Proof of Proposition 1

To cancel out the inter-user interference, $(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0$ or $(\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \neq k$ for each k , should be satisfied. For user k , suppose the ILs on user j ($j \in \Phi$) are canceled in the elevation domain, then $(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0, j \in \Phi$. If the ILs on user j ($j \in \bar{\Phi}$) are canceled in the azimuth part, then $(\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0, j \in \bar{\Phi}$.

If the number of users is larger than $M_x + M_y - 1$, either the elevation will suffer more than $M_x - 1$ ILs or the azimuth will suffer more than $M_y - 1$ ILs. Since all users have different elevation (azimuth) array response vectors with probability one, we cannot find non-zero vectors \mathbf{w}_k^e or \mathbf{w}_k^a such that $(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0 (j \in \Phi)$ or $(\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0 (j \in \bar{\Phi})$. The ILs will not be completely canceled.

If the number of users is no larger than $M_x + M_y - 1$, for user k , the number of ILs is not larger than $M_x + M_y - 2$. We can set Φ and $\bar{\Phi}$ such that $|\Phi| < M_x$ and $|\bar{\Phi}| < M_y$. Since all users have different elevation (azimuth) array response vectors with probability one, we almost can find vectors \mathbf{w}_k^e and \mathbf{w}_k^a such that $(\mathbf{h}_j^e)^T \mathbf{w}_k^e = 0 (j \in \Phi)$ and $(\mathbf{h}_j^a)^T \mathbf{w}_k^a = 0 (j \in \bar{\Phi})$.

B. Proof of Proposition 2

Denote $\tilde{\mathbf{A}} = \mathbf{A}^*$ and $\tilde{\mathbf{b}} = \mathbf{b}^*$. The projection matrix of $\tilde{\mathbf{A}}$ is $\mathbf{P} = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H$ [44]. If the unit norm vector \mathbf{x} satisfies $\mathbf{A}^T \mathbf{x} = 0$ ($\tilde{\mathbf{A}}^H \mathbf{x} = 0$), then

$$\tilde{\mathbf{b}}^H \mathbf{x} = \tilde{\mathbf{b}}^H (\mathbf{I} - \mathbf{P}) \mathbf{x} + \tilde{\mathbf{b}}^H \mathbf{P} \mathbf{x} = \tilde{\mathbf{b}}^H (\mathbf{I} - \mathbf{P}) \mathbf{x}. \quad (53)$$

To maximize $|\mathbf{b}^T \mathbf{x}|^2$ ($|\tilde{\mathbf{b}}^H \mathbf{x}|^2$), $\mathbf{x} = \vec{\mathbf{x}}$ where $\vec{\mathbf{x}} = (\mathbf{I} - \mathbf{P}) \tilde{\mathbf{b}}$. Note $[\mathbf{b} \ \mathbf{A}]^\dagger = ([\mathbf{b} \ \mathbf{A}]^H [\mathbf{b} \ \mathbf{A}])^{-1} [\mathbf{b} \ \mathbf{A}]^H$, and denote

$$[\mathbf{b} \ \mathbf{A}]^H [\mathbf{b} \ \mathbf{A}] = \begin{pmatrix} v & \mathbf{r}^H \\ \mathbf{r} & \mathbf{Z} \end{pmatrix} \quad (54)$$

where $v = \mathbf{b}^H \mathbf{b}$, $\mathbf{r} = \mathbf{A}^H \mathbf{b}$ and $\mathbf{Z} = \mathbf{A}^H \mathbf{A}$. Then, we have

$$\left([\mathbf{b} \ \mathbf{A}]^H [\mathbf{b} \ \mathbf{A}] \right)^{-1} = \begin{pmatrix} k^{-1} & -k^{-1} \mathbf{r}^H \mathbf{Z}^{-1} \\ -k^{-1} \mathbf{Z}^{-1} \mathbf{r} & (\mathbf{Z} - v^{-1} \mathbf{r} \mathbf{r}^H)^{-1} \end{pmatrix} \quad (55)$$

where $k = v - \mathbf{r}^H \mathbf{Z}^{-1} \mathbf{r}$. Consequently, the first row of $[\mathbf{b} \ \mathbf{A}]^\dagger$ is given by

$$\begin{aligned} \mathbf{f} &= [k^{-1} \quad -k^{-1} \mathbf{r}^H \mathbf{Z}^{-1}] [\mathbf{b} \ \mathbf{A}]^H \\ &= k^{-1} \left(\mathbf{b}^H - \mathbf{b}^H \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \right) \end{aligned} \quad (56)$$

or

$$\mathbf{f}^T = k^{-1} \left(\mathbf{b}^* - \mathbf{A}^* (\mathbf{A}^T \mathbf{A}^*)^{-1} \mathbf{A}^T \mathbf{b}^* \right) = k^{-1} (\mathbf{I} - \mathbf{P}) \tilde{\mathbf{b}} \quad (57)$$

which means $\vec{\mathbf{x}} = \vec{\mathbf{f}}^T$.

If $N \leq K$, $[\mathbf{b} \ \mathbf{A}]^\dagger = [\mathbf{b} \ \mathbf{A}]^H (\mathbf{b} \mathbf{b}^H + \mathbf{A} \mathbf{A}^H)^{-1}$, and the first row of $[\mathbf{b} \ \mathbf{A}]^\dagger$ is $\mathbf{f} = \mathbf{b}^H (\mathbf{b} \mathbf{b}^H + \mathbf{A} \mathbf{A}^H)^{-1}$. According to the Sherman-Morrison formula [44], we have

$$\begin{aligned} \mathbf{f} &= \mathbf{b}^H \left((\mathbf{A} \mathbf{A}^H)^{-1} - d \cdot (\mathbf{A} \mathbf{A}^H)^{-1} \mathbf{b} \mathbf{b}^H (\mathbf{A} \mathbf{A}^H)^{-1} \right) \\ &= q \mathbf{b}^H (\mathbf{A} \mathbf{A}^H)^{-1} \end{aligned} \quad (58)$$

where $d = 1 + \mathbf{b}^H (\mathbf{A} \mathbf{A}^H)^{-1} \mathbf{b}$ and $q = 1 - d \cdot \mathbf{b}^H (\mathbf{A} \mathbf{A}^H)^{-1} \mathbf{b}$. Thus, $\mathbf{f}^T = q (\mathbf{A}^* \mathbf{A}^T)^{-1} \mathbf{b}^*$. Note that the optimal solution of $\frac{\mathbf{x}^H (\mathbf{b}^* \mathbf{b}^T) \mathbf{x}}{\mathbf{x}^H \mathbf{A}^* \mathbf{A}^T \mathbf{x}}$ is $\vec{\mathbf{x}}^o$, where

$$\vec{\mathbf{x}}^o = (\mathbf{A}^* \mathbf{A}^T)^{-1} \mathbf{b}^*. \quad (59)$$

We have $\vec{\mathbf{f}}^T = \vec{\mathbf{x}}^o$, which means $\vec{\mathbf{f}}^T$ also maximizes $\frac{|\mathbf{b}^T \mathbf{x}|^2}{\mathbf{x}^H \mathbf{A}^* \mathbf{A}^T \mathbf{x}}$.

C. Proof of Theorem 3

Since the proof in the azimuth domain is similar to that in the elevation domain, we only consider the elevation domain. According to the Rayleigh quotient theorem, we have the solution of $\mathcal{P}_{6.1}$, which can be expressed as $\vec{\mathbf{w}}_{k,1}^e$ and

$$\vec{\mathbf{w}}_k^e = \left(\sum_{j \in \Phi} a_j^2 (\mathbf{h}_j^e)^* (\mathbf{h}_j^e)^T + c_1 \mathbf{I} \right)^{-1} (\mathbf{h}_k^e)^*. \quad (60)$$

For the channel matrix $(\tilde{\mathbf{H}}_\Phi^e)^T$, user k 's precoding vector in the MMSE scheme is $\vec{\mathbf{w}}_{k,1}^e$, where $\mathbf{w}_{k,1}^e$ is the first column of the matrix $((\tilde{\mathbf{H}}_\Phi^e)^* (\tilde{\mathbf{H}}_\Phi^e)^T + c_1 \mathbf{I})^{-1} (\tilde{\mathbf{H}}_\Phi^e)^*$, namely

$$\mathbf{w}_{k,1}^e = \left(\sum_{j \in \Phi} a_j^2 (\mathbf{h}_j^e)^* (\mathbf{h}_j^e)^T + c_1 \mathbf{I} + a_k^2 (\mathbf{h}_k^e)^* (\mathbf{h}_k^e)^T \right)^{-1} (\mathbf{h}_k^e)^*. \quad (61)$$

Denote $\mathbf{B} = \sum_{j \in \Phi} a_j^2 (\mathbf{h}_j^e)^* (\mathbf{h}_j^e)^T + c_1 \mathbf{I}$. According to Sherman-Morrison formula, we can obtain

$$\left(\mathbf{B} + a_k^2 (\mathbf{h}_k^e)^* (\mathbf{h}_k^e)^T \right)^{-1} = \mathbf{B}^{-1} + g \mathbf{B}^{-1} (\mathbf{h}_k^e)^* (\mathbf{h}_k^e)^T \mathbf{B}^{-1}$$

where g is a constant. Then,

$$\mathbf{w}_{k,1}^e = \left(\mathbf{B}^{-1} + g \mathbf{B}^{-1} (\mathbf{h}_k^e)^* (\mathbf{h}_k^e)^T \mathbf{B}^{-1} \right) (\mathbf{h}_k^e)^* = \tilde{g} \mathbf{B}^{-1} (\mathbf{h}_k^e)^*$$

where $\tilde{g} = 1 + g(\mathbf{h}_k^e)^T \mathbf{B}^{-1}(\mathbf{h}_k^e)^*$. Thus, we have $\vec{\mathbf{w}}_k^e = \vec{\mathbf{w}}_{k,1}^e$, which means $\mathbf{w}_k^e = \vec{\mathbf{w}}_{k,1}^e$.

D. Proof of Proposition 4

Since the proof in the azimuth domain is similar to that in the elevation domain, we only consider the elevation domain. When M_x approaches infinity, \mathbf{h}_i^e and \mathbf{h}_j^e are orthogonal if \mathbf{h}_i^e and \mathbf{h}_j^e are not collinear [15].

In the i th iteration of our DSSLNR algorithm, the elevation precoding matrix $\mathbf{w}_{k,i}^e$ is $\vec{\mathbf{w}}_{1,\Phi_i}^e$, where \mathbf{w}_{1,Φ_i}^e is the first column of matrix $\mathbf{W}_{\Phi_i}^e$ in (34). According to the matrix inversion lemma [45], we can rewrite (34) as

$$\mathbf{W}_{\Phi_i}^e = \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} \right)^{-1}. \quad (62)$$

Note that the first column of $\tilde{\mathbf{H}}_{\Phi_i}^e$ is $a_k \mathbf{h}_k^e$, and it is orthogonal to the remaining columns of $\tilde{\mathbf{H}}_{\Phi_i}^e$. As a result, we have

$$\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} = \begin{pmatrix} b_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_1 \end{pmatrix}$$

where b_1 is a constant. It is easy to verify that

$$\left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} \right)^{-1} = \begin{pmatrix} b_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_1^{-1} \end{pmatrix}.$$

Therefore, we have $\mathbf{w}_{k,i}^e = \left(\mathbf{h}_k^e \right)^*$.

E. Proof of Lemma 3

Since $\mathbf{v} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^*$, $\mathbf{G} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I}$, we can obtain

$$r = \left(\underline{\mathbf{h}}_l^e \right)^T \mathbf{T} \left(\underline{\mathbf{h}}_l^e \right)^* + c_1 \quad (63)$$

where $\mathbf{T} = \mathbf{I} - \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I} \right)^{-1} \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T$. Denote the singular value decomposition of $\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H$, we obtain $\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I} = \mathbf{V} \left(\mathbf{\Lambda}^H \mathbf{\Lambda} + c_1 \mathbf{I} \right) \mathbf{V}^H + c_1 \mathbf{I} = \mathbf{V} \left(\mathbf{\Lambda}^H \mathbf{\Lambda} + c_1 \mathbf{I} \right) \mathbf{V}^H$. Then,

$$\begin{aligned} \mathbf{T} &= \mathbf{I} - \mathbf{U} \mathbf{\Lambda}^H \left(\mathbf{\Lambda}^H \mathbf{\Lambda} + c_1 \mathbf{I} \right)^{-1} \mathbf{\Lambda}^H \mathbf{U}^H \\ &= \mathbf{U} \left(\mathbf{I} - \mathbf{\Lambda}^H \left(\mathbf{\Lambda}^H \mathbf{\Lambda} + c_1 \mathbf{I} \right)^{-1} \mathbf{\Lambda}^H \right) \mathbf{U}^H. \end{aligned} \quad (64)$$

It is easy to verify that $\mathbf{\Lambda}^H \left(\mathbf{\Lambda}^H \mathbf{\Lambda} + c_1 \mathbf{I} \right)^{-1} \mathbf{\Lambda}^H$ is a diagonal matrix, with each diagonal entry being within $[0, 1)$, then \mathbf{T} is a positive definite Hermitian matrix, and we have $\left(\underline{\mathbf{h}}_l^e \right)^T \mathbf{T} \left(\underline{\mathbf{h}}_l^e \right)^* > 0$ which means $r > 0$ from (63).

F. Proof of Theorem 5

According to the matrix inversion lemma [45], we have

$$\mathbf{W}_{\Phi_{i-1}}^e = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I} \right)^{-1}$$

$$\mathbf{W}_{\Phi_i}^e = \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* \left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} \right)^{-1}.$$

Denote

$$\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} = \begin{pmatrix} \mathbf{G} & \mathbf{v} \\ \mathbf{v}^H & u \end{pmatrix} \quad (65)$$

where $u = \left(\underline{\mathbf{h}}_l^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^* + c_1$, $\mathbf{v} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^*$, $\mathbf{G} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* + c_1 \mathbf{I}$. According to the block matrix inversion formula, we have

$$\begin{aligned} &\left(\left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^T \left(\tilde{\mathbf{H}}_{\Phi_i}^e \right)^* + c_1 \mathbf{I} \right)^{-1} \\ &= \begin{pmatrix} \left(\mathbf{G} - u^{-1} \mathbf{v} \mathbf{v}^H \right)^{-1} & -r^{-1} \mathbf{G}^{-1} \mathbf{v} \\ -r^{-1} \mathbf{v}^H \mathbf{G}^{-1} & r^{-1} \end{pmatrix} \end{aligned}$$

where $r = u - \mathbf{v}^H \mathbf{G}^{-1} \mathbf{v}$. As $\tilde{\mathbf{H}}_{\Phi_i}^e = \left[\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \quad \underline{\mathbf{h}}_l^e \right]$, we have

$$\begin{aligned} \mathbf{W}_{\Phi_i}^e &= \left[\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\underline{\mathbf{h}}_l^e \right)^* \right] \\ &\times \begin{pmatrix} \left(\mathbf{G} - u^{-1} \mathbf{v} \mathbf{v}^H \right)^{-1} & -r^{-1} \mathbf{G}^{-1} \mathbf{v} \\ -r^{-1} \mathbf{v}^H \mathbf{G}^{-1} & r^{-1} \end{pmatrix}. \end{aligned}$$

Denote

$$\begin{aligned} \mathbf{T}_1 &= \left[\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\underline{\mathbf{h}}_l^e \right)^* \right] \begin{pmatrix} \left(\mathbf{G} - u^{-1} \mathbf{v} \mathbf{v}^H \right)^{-1} \\ -r^{-1} \mathbf{v}^H \mathbf{G}^{-1} \end{pmatrix} \\ &= \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\mathbf{G} - u^{-1} \mathbf{v} \mathbf{v}^H \right)^{-1} - \left(\underline{\mathbf{h}}_l^e \right)^* r^{-1} \mathbf{v}^H \mathbf{G}^{-1}. \end{aligned} \quad (66)$$

According to the Sherman-Morrison formula [29], we have

$$\left(\mathbf{G} - u^{-1} \mathbf{v} \mathbf{v}^H \right)^{-1} = \mathbf{G}^{-1} + \frac{1}{r} \mathbf{G}^{-1} \mathbf{v} \mathbf{v}^H \mathbf{G}^{-1}. \quad (67)$$

Substituting (67) into (66), we have

$$\begin{aligned} \mathbf{T}_1 &= \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \mathbf{G}^{-1} + \frac{\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \mathbf{G}^{-1} \mathbf{v} \mathbf{v}^H \mathbf{G}^{-1}}{r} \\ &\quad - \frac{\left(\underline{\mathbf{h}}_l^e \right)^* \mathbf{v}^H \mathbf{G}^{-1}}{r} \\ &= \mathbf{W}_{\Phi_{i-1}}^e \\ &\quad - \frac{\left(\underline{\mathbf{h}}_l^e \right)^* - \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \mathbf{G}^{-1} \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^* \mathbf{v}^H \mathbf{G}^{-1}}{\left(\underline{\mathbf{h}}_l^e \right)^T \left(\left(\underline{\mathbf{h}}_l^e \right)^* - \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\mathbf{W}_{\Phi_{i-1}}^e \right)^H \left(\underline{\mathbf{h}}_l^e \right)^* \right) + c_1} \\ &= \mathbf{W}_{\Phi_{i-1}}^e - \frac{\mathbf{q}_m \mathbf{P}_m^H}{\left(\underline{\mathbf{h}}_l^e \right)^T \mathbf{q}_m + c_1}. \end{aligned}$$

Denote

$$\begin{aligned} \mathbf{T}_2 &= \left[\left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \left(\underline{\mathbf{h}}_l^e \right)^* \right] \begin{pmatrix} -r^{-1} \mathbf{G}^{-1} \mathbf{v} \\ r^{-1} \end{pmatrix} \\ &= -r^{-1} \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \mathbf{G}^{-1} \mathbf{v} + \left(\underline{\mathbf{h}}_l^e \right)^* r^{-1}. \end{aligned} \quad (68)$$

Noting that $\mathbf{v} = \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^*$, (68) can be written as

$$\mathbf{T}_2 = \frac{\left(\underline{\mathbf{h}}_l^e \right)^* - \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^* \mathbf{G}^{-1} \left(\tilde{\mathbf{H}}_{\Phi_{i-1}}^e \right)^T \left(\underline{\mathbf{h}}_l^e \right)^*}{u - \mathbf{v}^H \mathbf{G}^{-1} \mathbf{v}}$$

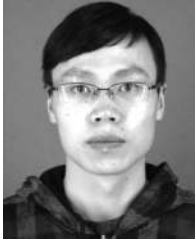
$$\begin{aligned}
&= \frac{(\mathbf{h}_l^e)^* - (\tilde{\mathbf{H}}_{\Phi_{i-1}}^e)^* (\mathbf{W}_{\Phi_{i-1}}^e)^H (\mathbf{h}_l^e)^*}{(\mathbf{h}_l^e)^T (\mathbf{h}_l^e)^* - (\mathbf{h}_l^e)^T (\tilde{\mathbf{H}}_{\Phi_{i-1}}^e)^* (\mathbf{W}_{\Phi_{i-1}}^e)^H (\mathbf{h}_l^e)^* + c_1} \\
&= \frac{\mathbf{q}_m}{(\mathbf{h}_l^e)^T \mathbf{q}_m + c_1}. \tag{69}
\end{aligned}$$

From $\mathbf{W}_{\Phi_i}^e = [\mathbf{T}_1, \mathbf{T}_2]$, we can obtain (38).

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 586–595, Dec. 1999.
- [2] C. Xing, S. Ma, and Y. Zhou, "Matrix-monotonic optimization for MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 334–348, Jan. 2015.
- [3] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution HSPA and LTE for Mobile Broadband*. New York, NY, USA: Academic, 2008.
- [4] L. Xiao, T. Chen, G. Han, W. Zhuang, and L. Sun, "Game theoretic study on channel-based authentication in MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7474–7484, Aug. 2017.
- [5] 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.
- [6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [7] W. Wang, N. Cheng, K. C. Teh, X. Lin, W. Zhuang, and X. Shen, "On countermeasures of pilot spoofing attack in massive MIMO systems: A double channel training based approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6697–6708, Jul. 2019.
- [8] Y. Han, S. Jin, J. Zhang, J. Zhang, and K. K. Wong, "DFT-based hybrid beamforming multiuser systems: Rate analysis and beam selection," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 514–528, Jun. 2018.
- [9] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [10] W. Wang, K. C. Teh, S. Luo, and K. H. Li, "Physical layer security in heterogeneous networks with pilot attack: A stochastic geometry approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6437–6449, Dec. 2018.
- [11] Y. Nam *et al.*, "Full-dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–179, Jun. 2013.
- [12] Y. Kim *et al.*, "Full-dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 92–100, Jun. 2014.
- [13] S. M. Razavizadeh, M. Ahn, and I. Lee, "Three-dimensional beamforming: A new enabling technology for 5G wireless networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 94–101, Nov. 2014.
- [14] Q. U. A. Nadeem, A. Kammoun, M. Debbah, and M. Alouini, "Design of 5G full dimension massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 726–740, Oct. 2018.
- [15] 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.
- [16] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [17] J. Nam, A. Adhikary, J. 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.
- [18] 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.
- [19] Y. Song, S. Nagata, H. Jiang, and L. Chen, "CSI-RS design for 3D MIMO in future LTE-advanced," in *Proc. IEEE Int. Conf. Commun.*, Sydney, NSW, Australia, Jun. 2014, pp. 5101–5106.
- [20] Z. Wang, W. Liu, C. Qian, S. Chen, and L. Hanzo, "Two-Dimensional precoding for 3-D massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5488–5493, Jun. 2017.
- [21] W. Liu, Z. Wang, C. Sun, S. Chen, and L. Hanzo, "Structured non-uniformly spaced rectangular antenna array design for FD-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3252–3266, May 2017.
- [22] N. Seifi, J. Zhang, R. W. Heath, Jr. T. Svensson, and M. Coldrey, "Coordinated 3D beamforming for interference management in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5396–5410, Oct. 2014.
- [23] 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.
- [24] 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 Proc.*, vol. 11, no. 2, pp. 347–362, Mar. 2017.
- [25] 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.
- [26] A. Kammoun, H. Khanfir, Z. Altman, M. Debbah, and M. Kamoun, "Preliminary results on 3D channel modeling: From theory to standardization," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1219–1229, Sep. 2017.
- [27] W. Wu, Q. Shen, M. Wang, and X. Shen, "Performance analysis of IEEE 802.11.ad downlink hybrid beamforming," *Proc. IEEE Int. Conf. Commun.*, Paris, France, Jul. 2017, pp. 1–6.
- [28] S. Hur *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 454–469, Apr. 2016.
- [29] J. Qiao, X. Shen, J. W. Mark, Q. Shen, Y. He, and L. Lei, "Enabling device-to-device communications in millimeter-wave 5G cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 209–215, Jan. 2015.
- [30] T. Bai, A. Alkhateeb, and R. W. Heath, Jr., "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.
- [31] Z. Shi, R. Lu, J. Chen, and X. Shen, "Three-dimensional spatial multiplexing for directional millimeter-wave communications in multi-cubicle office environments," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 4384–4389.
- [32] S. Cohen and C. Tomasi, "Systems of bilinear equations," *Comput. Sci. Dept.*, Stanford Univ., Stanford, CA, USA, Tech. Rep. CS-TR-97-1588, 1997.
- [33] X. Zhang, *Matrix Analysis and Applications*, 2nd ed. Beijing, China: Tsinghua Univ. Press, 2013.
- [34] S. Yoo, J. Lee, and K. Kim, "Geometry-based one-ring models for MIMO systems: Modeling accuracy assessment and improvement," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4583–4597, Jul. 2016.
- [35] R. He, B. Ai, G. L. Stuber, G. Wang, and Z. Zhong, "Geometrical-based modeling for millimeter-wave MIMO mobile-to-mobile channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 2848–2863, Apr. 2018.
- [36] M. Pätzold and B. O. Hogstad, "A space-time channel simulator for MIMO channels based on the geometrical one-ring scattering model," *Wireless Commun. Mobile Comput., Special Issue Multiple-Input Multiple-Output (MIMO) Commun.*, vol. 4, no. 7, pp. 727–737, Nov. 2004.
- [37] B. O. Hogstad, M. Pätzold, N. Youssef, and D. Kim, "A MIMO mobile-to-mobile channel model: Part II-The simulation model," in *Proc. IEEE 16th Int. Symp. Personal, Indoor Mobile Radio Commun.*, Berlin, Germany, Sep. 2005, pp. 562–567.
- [38] M. Pätzold, B. O. Hogstad, and N. Youssef, "Modeling, analysis, and simulation of MIMO mobile-to-mobile fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 510–520, Feb. 2008.
- [39] R. Shafin, L. Liu, J. Zhang, and Y. C. Wu, "DoA estimation and capacity analysis for 3-D millimeter wave massive-MIMO/FD-MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6963–6978, Oct. 2016.
- [40] D. Fan *et al.*, "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.
- [41] J. Choi, G. Lee, and B. L. Evans, "Two-stage analog combining in hybrid beamforming systems with low-resolution ADCs," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2410–2425, May 2019.
- [42] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

- [43] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.
- [44] G. H. Golub, and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.
- [45] G. Xu *et al.*, "Full dimension MIMO (FD-MIMO): Demonstrating commercial feasibility," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1876–1886, Aug. 2017.



Yunchao Song (M'17) received the B.E. degree in electronic science and technology and the Ph.D. degree in circuits and systems from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2010 and 2016, respectively. Since 2017, he has been an Instructor with the College of Electronic and Optical Engineering, NUPT. He is currently a Visiting Scholar with BCCR Lab, Department of ECE, University of Waterloo, Waterloo, ON, Canada. His research interests are 3-D MIMO systems and millimeter wave communications.



Chen Liu (M'18) received the B.E. degree in electrical and information engineering from the Nanjing Institute of Technology (Southeast University), Nanjing, China, in 1985 and the M.S. degree in circuits and systems from Anhui University, Hefei, China, in 1988. He received the Ph.D. degree in signal and information processing from Southeast University, China, in 2005. He joined NUPT in 1988 where he has been a Professor since 2002. His current research interests include massive MIMO systems, 3-D MIMO systems, and millimeter wave communications.

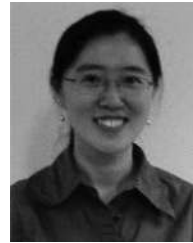


Wei Wang (S'14) received the B.E. degree in information countermeasure technology and the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2018. He is currently a Post-doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include wireless communications, space-air-ground integrated networks,

wireless security, and physical layer security. Dr. Wang was the recipient of the IEEE Student Travel Grants for IEEE International Conference on Communications'17, and the Chinese Government Award for outstanding self-financed students abroad in 2019.

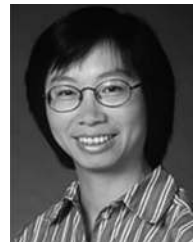


Nan Cheng (M'16) received the B.E. and M.S. degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently a Joint Professor with the School of Telecommunication, Xidian University, Xi'an, China. He is also a Joint Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, and with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include performance analysis, MAC, opportunistic communication for vehicular networks, unmanned aerial vehicles, and application of artificial intelligence (AI) for wireless networks.



traffic control, and routing protocol design for vehicular networks.

Miao Wang (M'15) received the B.Sc. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007, the M.Sc. degree from Beihang University, Beijing, China, in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2015. She is an Assistant Professor with the Department of Electrical and Computer Engineering, Miami University, Oxford, OH, USA. Her current research interests include Internet of Things, electric vehicles charging/discharging in smart grid,



2017 ten N2Women (Stars in Computer Networking and Communications), and a co-recipient of several Best Paper Awards from IEEE conferences. She was an IEEE Communications Society Distinguished Lecturer from 2008 to 2011. She was the Editor-in-Chief for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2013, Technical Program Chair/Co-Chair of IEEE Vehicular Technology Conference Fall 2016 and Fall 2017, and the Technical Program Symposia Chair of the IEEE GLOBECOM 2011. She is an elected member in the Board of Governors and VP Publications of the IEEE Vehicular Technology Society.

Weihua Zhuang (M'93–SM'01–F'08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair in Wireless Communication Networks. Ms. Zhuang is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. She was the recipient of 2017 Technical Recognition Award from IEEE Communications Society Ad Hoc & Sensor Networks Technical Committee, one of



works, smart grid, and vehicular ad hoc and sensor networks. Dr. Shen is a registered Professional Engineer of ON, Canada, an Engineering Institute of Canada Fellow, and a Canadian Academy of Engineering Fellow, and a Royal Society of Canada Fellow.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree in electrical engineering from Dalian Maritime University, Dalian, China, in 1982, and the M.Sc. and Ph.D. degrees in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1987 and 1990, respectively. He is currently a University Professor and an Associate Chair for Graduate Studies with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on resource management, wireless network security, social networks,

He was the recipient of the Outstanding Performance Award five times from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, the Excellent Graduate Supervision Award in 2006, the Joseph LoCicero Award in 2015 and the Education Award in 2017 from the IEEE Communications Society, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, and the R.A. Fessenden Award in 2019 from IEEE, Canada. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society. He is the Editor-in-Chief for the IEEE INTERNET OF THINGS JOURNAL and the Vice President on Publications of the IEEE Communications Society. He served as the Technical Program Committee Chair/Co-Chair for the IEEE Globecom'16, the IEEE Infocom'14, the IEEE Vehicular Technology Conference'10 Fall, the IEEE Globecom'07, the Symposia Chair for the IEEE International Conference on Communications'10, the Tutorial Chair for the IEEE Vehicular Technology Conference'11 Spring, the Chair for the IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking.