Performance Optimization for Intelligent Reflecting Surface Assisted Multicast MIMO Networks

Songling Zhang*, Zhaohui Yang†, Mingzhe Chen‡, Danpu Liu*, Kai-Kit Wong†, and H. Vincent Poor‡
*Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing, China
†Department of Electronic and Electrical Engineering, University College London, London, WC1E 6BT, UK
‡Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, 08544, USA
Emails: slzhang@bupt.edu.cn,zhaohui.yang@ucl.ac.uk, mingzhec@princeton.edu, dpliu@bupt.edu.cn,
kai-kit.wong@ucl.ac.uk, poor@princeton.edu

***Abstract*—In this paper, the problem of maximizing the sum rate of all the multicasting groups in an intelligent reflecting surface (IRS)-assisted millimeter wave multicast multiple-input multiple-output communication system is studied. In the considered model, one IRS is deployed to assist the communication from a multi-antenna base station (BS) to the multi-antenna users that are clustered into several groups. Our goal is to maximize the sum rate of all the multicasting groups by jointly optimizing the transmit beamforming matrices of the BS, the receive beamforming matrices of the users, and the phase shifts of the IRS. To solve this non-convex problem, we first use a block diagonalization method to represent the beamforming matrices of the BS and the users by the phase shifts of the IRS. Then, substituting the expressions of the beamforming matrices of the BS and the users, the original sum-rate maximization problem can be transformed into a problem that only needs to optimize the phase shifts of the IRS. To solve the transformed problem, a manifold method is used. Simulation results show that the proposed scheme can achieve up to 14.93% gain in terms of the sum rate of all the multicasting groups compared to the algorithm that optimizes the hybrid beamforming matrices of the BS and the users using our proposed scheme and randomly determines the phase shifts of the IRS.**

I. Introduction

Millimeter wave (mmWave) communications utilize the 30-300 GHz frequency band to achieve multi-gigabit transmissions. However, mmWave suffers from severe path loss and is easily blocked by obstacles due to the short wavelength. To address these problems, massive multiple-input multiple-output (MIMO) and intelligent reflecting surface (IRS) have been proposed [1], [2]. However, deploying IRS and massive MIMO over mmWave communication systems faces several challenges such as IRS location optimization, joint active and passive beamforming design.

A number of existing works [3]–[8] have studied the problems of optimizing the phase-shift matrix of the IRS and beamforming matrix of the transceiver. In [3], the authors maximized the received signal power by jointly optimizing a transmit precoding vector of the base station (BS) and the phase shift coefficients of an IRS. The authors in [4] maximized the spectral efficiency by jointly optimizing the reflection coefficients of the IRS and the hybrid precoder and combiner. The work in [5] studied the hybrid precoding design for an IRS aided multi-user mmWave communication system. A geometric mean decomposition-based beamforming scheme was proposed for IRS-assisted mmWave hybrid MIMO systems in [6]. In [7], the authors optimized channel estimator in a closed form while considering the signal reflection matrix of an IRS and an analog combiner of a receiver. The authors in [8] jointly optimized the coordinated transmit beamforming vectors of the BSs and the reflective beamforming vector of the IRS, so as to maximize the minimum weighted signal-to-interference-plus-noise ratio (SINR) of users. However, most of these existing works [3]–[8] only consider the deployment of IRS over unicast communication networks in which each BS transmits independent data streams to each user.

Multicast enables the BS to transmit a content to multiple users using an identical radio resource, thus improving the spectrum and energy efficiency [9]–[11]. However, deploying IRS over multicast communication systems faces several new challenges. First, the users in a group that have different channel conditions need to be served by a coordinated beamforming matrix, thus complicating the design of the beamforming matrix of the transmitter. Besides, in a multicast system, the data rate of a group is limited by the user with the worst-channel gain [12]. Therefore, in a multicast system, one must maximize the data rate of the user with the worst-channel gain in each group.

The main contribution of this paper is a novel IRS assisted multigroup multicast system. In particular, we consider an IRS-assisted mmWave multicast MIMO communication system. In the considered model, one IRS is used to assist the communication from a multi-antenna BS to multi-antenna users that are clustered into several groups. To maximize the sum rate of all the multicasting groups, we jointly optimize the transmit beamforming matrices of the BS, the receive beamforming matrices of the users, and the phase shifts of the IRS. We formulate an optimization problem with the objective of maximizing the sum rate of all the multicasting groups under amplitude constraint of radio frequency (RF) beamforming matrices, maximum transmit power constraint, and unit-modulus constraint of the IRS phase shifts. To solve this problem, we first use a block diagonalization (BD) method to represent the beamforming matrices of the BS and the users by the phase shifts of the IRS. Then, we substitute the expressions of the beamforming matrices of the BS and the users into the original problem so as to transform it to a problem that only needs to optimize the phase shifts of the IRS. The transformed problem is solved by a manifold method. Simulation results show that the proposed scheme

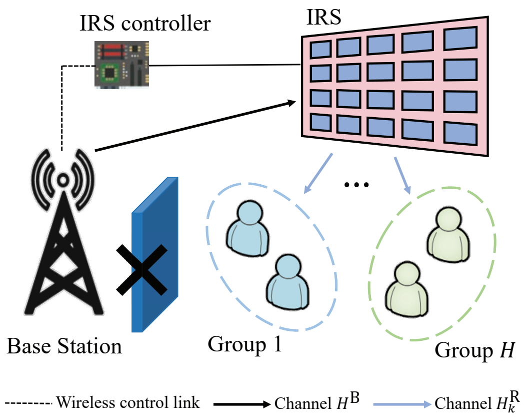

Fig. 1. An IRS-aided mmWave multigroup multicast MIMO communication system.

can achieve up to 14.93% gain in terms of the sum rate of all the multicasting groups compared to the algorithm that optimizes the hybrid beamforming matrices of the BS and the users using our proposed algorithm and randomly determines the phase shifts of the IRS.

II. System Model And Problem Formulation

A. System Model

We consider an IRS-aided mmWave multigroup multicast MIMO communication system in which a BS is equipped with N^{B} antennas serving K users via an IRS, as shown in Fig. 1. The users are divided into H groups. We assume that the users in a group will request the same data streams and the data streams requested by the users in different groups are different. The set of user groups is denoted by $\mathcal{H}=\{1,2,\ldots,H\}$. Meanwhile, the set of users in a group h is denoted as $\mathcal{H}_h$. We also assume that each user can only belong to one group, i.e., $\mathcal{H}_i \cap \mathcal{H}_j=\emptyset$, $\forall i,j \in \mathcal{H}, i\neq j$. In our model, the direct communication link between the BS and a user is blocked due to unfavorable propagation conditions. Each user is equipped with N^{U} antennas and M^{U} RF chains to receive ζ data streams from the BS. The BS simultaneously transmits $H\zeta$ independent data streams to the users by M^{B} RF chains[1].

At the BS, the transmitted data streams of H user groups are precoded by a baseband transmit beamforming matrix $\boldsymbol{F}^{\mathrm{B}}=\left[\boldsymbol{F}_1^{\mathrm{B}},\boldsymbol{F}_2^{\mathrm{B}},\ldots,\boldsymbol{F}_H^{\mathrm{B}}\right]\in\mathbb{C}^{M^{\mathrm{B}}\times H\zeta}$, with $\boldsymbol{F}_h^{\mathrm{B}}$ being the transmit beamforming matrix of group h. After that, each transmitted data stream of H user groups is precoded by an RF transmit beamforming matrix $\boldsymbol{F}^{\mathrm{R}}\in\mathbb{C}^{N^{\mathrm{B}}\times M^{\mathrm{B}}}$.

The received data streams of user k in group h are first processed by an RF receive beamforming matrix $\boldsymbol{W}_k^{\mathrm{R}}\in\mathbb{C}^{N^{\mathrm{U}}\times M^{\mathrm{U}}}$. Then, user k uses a baseband receive beamfoming matrix $\boldsymbol{W}_k^{\mathrm{B}}\in\mathbb{C}^{M^{\mathrm{U}}\times\zeta}$ to recover ζ data streams.

In our model, an IRS is used to enhance the received signal strength of users by reflecting signals from the BS to the users. We assume that the signal power of the multi-reflections (i.e., reflections more than once) on the IRS is ignored due to severe path loss. The phase-shift matrix of the IRS is $\boldsymbol{\Phi}=\operatorname{diag}\left(e^{j\phi_1},\ldots,e^{j\phi_M}\right)\in\mathbb{C}^{M\times M}$, where $\operatorname{diag}\left(e^{j\phi_1},\ldots,e^{j\phi_M}\right)$ is a diagonal matrix of $\left[e^{j\phi_1},\ldots,e^{j\phi_M}\right]$, M is the number of reflecting elements at the IRS, and $\phi_m\in[0,2\pi]$ is the phase shift introduced by element m of the IRS.

[1]The numbers of RF chains are subject to the constraints $H\zeta\leq M^{\mathrm{B}}\leq N^{\mathrm{B}}$ and $\zeta\leq M^{\mathrm{U}}\leq N^{\mathrm{U}}$.

1) Channel Model: The BS and the users employ uniform linear arrays (ULAs), and the IRS uses a uniform planar array (UPA). The normalized array response vector for an ULA is

$$\boldsymbol{a}(r)=\frac{1}{\sqrt{N}}\left[1,\ldots,e^{j\frac{2\pi d}{\lambda}(N-1)\sin(r)}\right]^{\mathrm{T}}, \tag{1}$$

where N is the number of antennas in ULA, d is an interval between two antennas, and λ is the signal wavelength. The normalized array response vector of UPA is

$$\boldsymbol{a}(\theta,\eta)=\frac{1}{\sqrt{F_y\times F_z}}[1,\ldots,e^{j\frac{2\pi d}{\lambda}((F_y-1)\cos\eta\sin\theta+(F_z-1)\sin\eta)}]^{\mathrm{T}}, \tag{2}$$

where $F_y\times F_z$ is the number of elements in UPA, F_y and F_z are respectively the number of elements in the horizontal and vertical directions. The BS-IRS channel $\boldsymbol{H}^{\mathrm{B}}\in\mathbb{C}^{M\times N^{\mathrm{B}}}$ and the channel $\boldsymbol{H}_k^{\mathrm{R}}\in\mathbb{C}^{N^{\mathrm{U}}\times M}$ from the IRS to user k in group h can be respectively given as

$$\boldsymbol{H}^{\mathrm{B}}=\sqrt{\frac{N^{\mathrm{B}}M}{Y}}\sum_{i=1}^{Y}\alpha_i\boldsymbol{a}\left(\theta_i^{\mathrm{A}},\eta_i^{\mathrm{A}}\right)\left(\boldsymbol{a}\left(r_i^{\mathrm{D}}\right)\right)^{\mathrm{H}}, \tag{3}$$

$$\boldsymbol{H}_k^{\mathrm{R}}=\sqrt{\frac{MN^{\mathrm{U}}}{L}}\sum_{i=1}^{L}\beta_i\boldsymbol{a}\left(r_{i,k}^{\mathrm{A}}\right)\left(\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)\right)^{\mathrm{H}}, \tag{4}$$

where Y is the total number of paths (line-of-sight (LOS) and non-line-of-sight (NLOS)) between the BS and the IRS, L is the total number of paths (LOS and NLOS) between the IRS and user k, θ_i^{A} denotes the azimuth angle of arrival of the IRS, θ_i^{D} denotes the azimuth angle of departure of the IRS, η_i^{A} denotes the elevation angle of arrival of the IRS, η_i^{D} denotes the elevation angle of departure of the IRS, $r_{i,k}^{\mathrm{A}}$ represents the arrival angle of user k, r_i^{D} represents the departure angle of the BS, α_i and β_i are complex channel gains. $\boldsymbol{a}\left(r_i^{\mathrm{D}}\right)$ and $\boldsymbol{a}\left(r_{i,k}^{\mathrm{A}}\right)$ denote the normalized array response vectors of the BS and user k, respectively. $\left(\boldsymbol{a}\left(r_i^{\mathrm{D}}\right)\right)^{\mathrm{H}}$ is the Hermitian transpose of $\boldsymbol{a}\left(r_i^{\mathrm{D}}\right)$. $\boldsymbol{a}\left(\theta_i^{\mathrm{A}},\eta_i^{\mathrm{A}}\right)$ represents the normalized array response vector of the IRS over the effective channel from the BS to the IRS. $\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)$ represents the normalized array response vector of the IRS over the effective channel from the IRS to user k. The effective channel from the BS to user k in group h is $\boldsymbol{H}_k=G_{\mathrm{t}}G_{\mathrm{r}}\boldsymbol{H}_k^{\mathrm{R}}\boldsymbol{\Phi}\boldsymbol{H}^{\mathrm{B}}$, where G_{t} and G_{r} are the antenna gains of the BS and each user, respectively.

2) Data Rate Model: We assume that the BS obtains the channel state information (CSI). The BS is responsible for designing the reflection coefficients of the IRS. As a result, the detected data of user k in group h is given by

$$\hat{\boldsymbol{s}}_{k,h}=\left(\boldsymbol{W}_k^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}\boldsymbol{s}+\left(\boldsymbol{W}_k^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{n}_k, \tag{5}$$

where $\boldsymbol{s}=\left[\boldsymbol{s}_1^{\mathrm{T}},\ldots,\boldsymbol{s}_H^{\mathrm{T}}\right]^{\mathrm{T}}\in\mathbb{C}^{H\zeta\times 1}$ represents the data streams to be transmitted to all users, with $\boldsymbol{s}_h=\left[s_{h,1},\ldots,s_{h,\zeta}\right]^{\mathrm{T}}\in\mathbb{C}^{\zeta\times 1}$ being ζ streams that will be transmitted to each user in group h. $\boldsymbol{n}_k\in\mathbb{C}^{N^{\mathrm{U}}\times 1}$ is an additive white Gaussian noise vector of user k. Each element of $\boldsymbol{n}_k$

follows the independent and identically distributed complex Gaussian distribution with zero mean and variance σ^2. In (5), the first term represents the signal received by user k. The second term is the noise received by user k. The estimated data stream i received by user k in group h can be expressed as

$$\begin{aligned}\hat{s}_{ik,h} &= \left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{h_i}^{\mathrm{B}}s_{h,i}\\ &+\sum_{j=1,j\neq i}^{\zeta}\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{h_j}^{\mathrm{B}}s_{h,i}\\ &+\sum_{m=1,m\notin\mathcal{H}_h}^{H}\sum_{l=1}^{\zeta}\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{m_l}^{\mathrm{B}}s_{m,l}\\ &+\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{n}_k,\end{aligned}\tag{6}$$

where $h_i=(h-1)\zeta+i$, $\boldsymbol{w}_{k,i}^{\mathrm{B}}$ denotes row i of matrix $\boldsymbol{W}_k^{\mathrm{B}}$, and $\bar{\boldsymbol{f}}_{h_i}^{\mathrm{B}}$ denotes column h_i of matrix $\boldsymbol{F}^{\mathrm{B}}$. In (6), the first term represents the desired signal. The second term is the interference caused by other streams of user k. The third term is the interference caused by the users from other groups. The fourth term is the noise. The SINR of user k in group h receiving data stream i is given by

$$\begin{aligned}&\xi_{ik,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)=\\ &\qquad\frac{\left|\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{h_i}^{\mathrm{B}}\right|^2}{I_{ik,h}+J_{ik,h}+\sigma^2},\end{aligned}\tag{7}$$

where $I_{ik,h}$ is short for $I_{ik,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)$ and $I_{ik,h}=\sum_{j=1,j\neq i}^{\zeta}\left|\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{h_j}^{\mathrm{B}}\right|^2$ represents the interference from user itself, $J_{ik,h}$ is short for $J_{ik,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)$ and $J_{ik,h}=\sum_{m=1,m\notin\mathcal{H}_h}^{H}\sum_{l=1}^{\zeta}\left|\left(\boldsymbol{w}_{k,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_k^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}^{\mathrm{R}}\bar{\boldsymbol{f}}_{m_l}^{\mathrm{B}}\right|^2$ represents the interference from other groups. The achievable data rate of user k in group h is given by

$$\begin{aligned}&R_{k,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)=\\ &\qquad\sum_{i=1}^{\zeta}\log_2\left(1+\xi_{ik,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)\right),\end{aligned}\tag{8}$$

Due to the nature of the multicast mechanism, the achievable data rate of group h depends on the user with minimum data rate, which is defined as follows:

$$\min_{k\in\mathcal{H}_h}\left\{R_{k,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)\right\}.\tag{9}$$

B. Problem Formulation

Next, we introduce our optimziation problem. Our goal is to maximize the sum rate of all the multicasting groups via jointly optimizing the transmit beamforming matrices $\boldsymbol{F}^{\mathrm{B}}$, $\boldsymbol{F}^{\mathrm{R}}$, the receive beamforming matrices $\boldsymbol{W}^{\mathrm{R}}$, $\boldsymbol{W}^{\mathrm{B}}$, and the phase shift $\boldsymbol{\nu}$ of the IRS. Mathematically, the optimization problem is formulated as

$$\max_{\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}^{\mathrm{B}},\boldsymbol{\nu}}\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\left\{R_{k,h}\left(\boldsymbol{W}_k^{\mathrm{R}},\boldsymbol{W}_k^{\mathrm{B}},\boldsymbol{\nu},\boldsymbol{F}^{\mathrm{R}},\boldsymbol{F}_h^{\mathrm{B}}\right)\right\}\tag{10}$$

$$\text{s.t.}\quad \left\|\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}\right\|_F^2\leq P,\tag{10a}$$

$$\left|\boldsymbol{F}^{\mathrm{R}}(i,j)\right|=\left|\boldsymbol{W}_k^{\mathrm{R}}(i,j)\right|=1,\forall i,j,\tag{10b}$$

$$0\leq\phi_m\leq 2\pi, m=1,\ldots,M,\tag{10c}$$

where P is the maximum transmit power of the BS, $\left\|\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}\right\|_F$ is the Frobenius norm of $\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}$, $\boldsymbol{\nu}=\left[e^{j\phi_1},\ldots,e^{j\phi_M}\right]^{\mathrm{H}}$, $\boldsymbol{F}^{\mathrm{R}}(i,j)$ denotes the element (i,j) of matrix $\boldsymbol{F}^{\mathrm{R}}$, with $\left|\boldsymbol{F}^{\mathrm{R}}(i,j)\right|$ being the amplitude of $\boldsymbol{F}^{\mathrm{R}}(i,j)$. The maximum transmit power constraint of the BS is given in (10a). Constraint (10b) represents the amplitude constraints of the RF beamforming matrices of the BS and each user, while (10c) shows the phase shift limits of the IRS. Due to non-convex objective function (10) and non-convex constraints (10a)-(10c), problem (10) is non-convex, and hence it is hard to solve. Next, we introduce an efficient scheme to solve problem (10).

III. Proposed Scheme

In this section, we first use the phase shift $\boldsymbol{\nu}$ of the IRS to represent the fully digital transmit beamforming matrix of the BS and receive beamforming matrix of the users. Then, we substitute them in (10) to transform problem (10). To solve the transformed problem, the phase shift $\boldsymbol{\nu}$ of the IRS is optimized by a manifold method. Finally, we introduce the entire algorithm used to solve problem (10).

A. Block Diagonalization Method

1) Simplification of Optimization Problem: $\boldsymbol{W}_k^{\mathrm{B}}\boldsymbol{W}_k^{\mathrm{R}}$ and $\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}$ are regarded as a whole, that is, problem (10) is solved as a problem of fully digital beamforming. Once the fully digital transmit beamforming matrix and receive beamforming matrix are obtained, we can use the algorithm in [13] to find the hybrid transmit beamforming matrices and receive beamforming matrices to approximate the fully digital transmit beamforming matrix and receive beamforming matrix, as done in [14]. Although the fully digital architecture has better performance, the cost and energy consumption of hardware implementation are too high, so the hybrid architecture is chosen. Let $\boldsymbol{B}=[\boldsymbol{B}_1,\ldots,\boldsymbol{B}_H]\in\mathbb{C}^{N^{\mathrm{B}}\times H\zeta}$ be a fully digital transmit beamforming matrix which has the same size as the hybrid transmit beamforming matrix $\boldsymbol{F}^{\mathrm{R}}\boldsymbol{F}^{\mathrm{B}}$ and $\boldsymbol{J}_k\in\mathbb{C}^{N^{\mathrm{U}}\times\zeta}$ be a fully digital receive beamforming matrix of user k in group h. The size of $\boldsymbol{J}_k$ is equal to that of the hybrid receive beamforming matrix $\boldsymbol{W}_k^{\mathrm{R}}\boldsymbol{W}_k^{\mathrm{B}}$. Substituting $\boldsymbol{B}$ and $\boldsymbol{J}_k$ in (10), problem (10) can be transformed as

$$\max_{\boldsymbol{B},\boldsymbol{J},\boldsymbol{\nu}}\quad\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\left\{R_{k,h}\left(\boldsymbol{B}_h,\boldsymbol{J}_k,\boldsymbol{\nu}\right)\right\}\tag{11}$$

$$\text{s.t.}\quad (10\text{c}),$$

$$\|\boldsymbol{B}\|_{\mathrm{F}}^2\leq P,\tag{11a}$$

where $\boldsymbol{J}=\mathrm{diag}\left(\boldsymbol{J}_1,\ldots,\boldsymbol{J}_K\right)$.

2) Optimization of $\boldsymbol{B}$ and $\boldsymbol{J}$: Due to the low complexity of a BD method, we use it to find the relationship between $\boldsymbol{\nu}$ and the fully digital transmit beamforming matrix $\boldsymbol{B}$ of the BS as well as the receive beamforming matrix $\boldsymbol{J}$ of the users.

Lemma 1: Given $\boldsymbol{\nu}$ and the power allocation matrix $\boldsymbol{G}_h = \mathrm{diag}\left(p_1, \ldots, p_\zeta\right)$ in group h, where $p_i = \frac{P}{H\zeta}$ is the transmit power of data stream i, $\boldsymbol{B}$ and $\boldsymbol{J}$ can be given by

$$\boldsymbol{B}_h\left(\boldsymbol{\nu}\right) = \tilde{\boldsymbol{V}}_h^{(0)}\left(\frac{\boldsymbol{V}_1^{(1)}+\cdots+\boldsymbol{V}_K^{(1)}}{\sqrt{|\mathcal{H}_h|}}\right)\sqrt{\frac{P}{H\zeta}}, \tag{12}$$

$$\boldsymbol{J}_k\left(\boldsymbol{\nu}\right) = \boldsymbol{U}_k^{(1)}, \tag{13}$$

where $\tilde{\boldsymbol{V}}_h^{(0)} = \mathrm{null}\left(\tilde{\boldsymbol{H}}_h\right)$, $\boldsymbol{U}_k^{(1)}$, and $\boldsymbol{V}_k^{(1)}$ can be obtained by singular value decomposition (SVD) of $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}$ with $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)} = \left[\boldsymbol{U}_k^{(1)}, \boldsymbol{U}_k^{(2)}\right]\begin{bmatrix}\boldsymbol{\Sigma}_k^{(1)} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_k^{(2)}\end{bmatrix}\left[\boldsymbol{V}_k^{(1)}, \boldsymbol{V}_k^{(2)}\right]^{\mathrm{H}}$.

Proof: To prove Lemma 1, we first define $\tilde{\boldsymbol{H}}_h$ as

$$\tilde{\boldsymbol{H}}_h \triangleq \left[\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{h-1}, \boldsymbol{H}_{h+1}, \ldots, \boldsymbol{H}_H\right]^{\mathrm{T}}, \tag{14}$$

where $\boldsymbol{H}_{h-1}$ is the effective channels of all users in group $h-1$. We assume that the rank of $\tilde{\boldsymbol{H}}_h$ is $\tilde{L}_k$. Next, we introduce the use of BD method to represent the transmit beamforming matrices of the BS and the receive beamforming matrices of the users by the phase shifts of the IRS. To eliminate the inter-group interference, we define $\tilde{\boldsymbol{V}}_h^{(0)} \in \mathbb{C}^{N^{\mathrm{B}}\times\left(N^{\mathrm{B}}-\tilde{L}_k\right)}$ as

$$\tilde{\boldsymbol{V}}_h^{(0)} = \mathrm{null}\left(\tilde{\boldsymbol{H}}_h\right), \tag{15}$$

where $\mathrm{null}\left(\tilde{\boldsymbol{H}}_h\right)$ represents that $\tilde{\boldsymbol{V}}_h^{(0)}$ lies in the null space of $\tilde{\boldsymbol{H}}_h$. Hence, we have $\tilde{\boldsymbol{H}}_h\tilde{\boldsymbol{V}}_h^{(0)} = \boldsymbol{0}$. The self interference of each user can be eliminated by the SVD of $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}$, which is

$$\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)} = \left[\boldsymbol{U}_k^{(1)}, \boldsymbol{U}_k^{(2)}\right]\begin{bmatrix}\boldsymbol{\Sigma}_k^{(1)} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_k^{(2)}\end{bmatrix}\left[\boldsymbol{V}_k^{(1)}, \boldsymbol{V}_k^{(2)}\right]^{\mathrm{H}}. \tag{16}$$

We assume that the rank of $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}$ is L_k, the column vectors of $\boldsymbol{U}_k^{(1)} \in \mathbb{C}^{N^{\mathrm{U}}\times\zeta}$, $\boldsymbol{U}_k^{(2)} \in \mathbb{C}^{N^{\mathrm{U}}\times(L_k-\zeta)}$, $\boldsymbol{V}_k^{(1)} \in \mathbb{C}^{\left(N^{\mathrm{B}}-\tilde{L}_k\right)\times\zeta}$, and $\boldsymbol{V}_k^{(2)} \in \mathbb{C}^{\left(N^{\mathrm{B}}-\tilde{L}_k\right)\times(L_k-\zeta)}$ can form orthonormal sets, $\boldsymbol{\Sigma}_k^{(1)} \in \mathbb{C}^{\zeta\times\zeta}$ and $\boldsymbol{\Sigma}_k^{(2)} \in \mathbb{C}^{(L_k-\zeta)\times(L_k-\zeta)}$ are diagonal matrices of singular values. $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$ must be designed to cancel the inter-group interference as well as the interference of all users in this group. Thus, $\tilde{\boldsymbol{V}}_h^{(0)}$ and $\sum\limits_{i=1}^{K}\boldsymbol{V}_i^{(1)}$ must be included in $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$, which can be given by

$$\boldsymbol{B}_h\left(\boldsymbol{\nu}\right) = \tilde{\boldsymbol{V}}_h^{(0)}\left(\frac{\boldsymbol{V}_1^{(1)}+\cdots+\boldsymbol{V}_K^{(1)}}{\sqrt{|\mathcal{H}_h|}}\right)\boldsymbol{G}_h^{1/2}, \tag{17}$$

where $|\mathcal{H}_h|$ is the number of users in group h, $\frac{1}{\sqrt{|\mathcal{H}_h|}}$ ensures that the power of $\left(\frac{\boldsymbol{V}_1^{(1)}+\cdots+\boldsymbol{V}_K^{(1)}}{\sqrt{|\mathcal{H}_h|}}\right)$ is unit, and $\boldsymbol{G}_h$ is the power allocation matrix in group h. To eliminate the interference of all users in this group, the fully digital receive beamforming matrix $\boldsymbol{J}_k$ of user k in group h is written as

$$\boldsymbol{J}_k\left(\boldsymbol{\nu}\right) = \boldsymbol{U}_k^{(1)}. \tag{18}$$

Substituting $\boldsymbol{G}_h$ into (17), we have

$$\boldsymbol{B}_h\left(\boldsymbol{\nu}\right) = \tilde{\boldsymbol{V}}_h^{(0)}\left(\frac{\boldsymbol{V}_1^{(1)}+\cdots+\boldsymbol{V}_K^{(1)}}{\sqrt{|\mathcal{H}_h|}}\right)\sqrt{\frac{P}{H\zeta}}. \tag{19}$$

This completes the proof. ■

From Lemma 1, we can see that $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$ mainly depends on the orthogonal bases of the null space of users in other groups, the orthogonal bases of the subspace of users in group h, the maximum transmit power of the BS and number of groups, $\boldsymbol{J}_k\left(\boldsymbol{\nu}\right)$ depends on the effective channel of user k and the orthogonal bases of the null space of users in other groups.

3) Simplification of Problem (11): Based on the Lemma 1, the interference caused by other groups $J_{ik,h}$ and user itself $I_{ik,h}$ can be eliminated by the fully digital transmit beamforming matrix $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$ and receive beamforming matrix $\boldsymbol{J}_k\left(\boldsymbol{\nu}\right)$. Substituting $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$ and $\boldsymbol{J}_k\left(\boldsymbol{\nu}\right)$ into (11), the achievable data rate of user k in group h can be rewritten as follows:

$$\begin{aligned}&R_{k,h}\left(\boldsymbol{B}_h\left(\boldsymbol{\nu}\right), \boldsymbol{J}_k\left(\boldsymbol{\nu}\right), \boldsymbol{\nu}\right) = \\ &\qquad\sum_{i=1}^{\zeta}\log_2\left(1+\left|\left(\boldsymbol{j}_{k,i}\left(\boldsymbol{\nu}\right)\right)^{\mathrm{H}}\boldsymbol{H}_k\bar{\boldsymbol{b}}_{h,i}\left(\boldsymbol{\nu}\right)\right|^2/\sigma^2\right),\end{aligned} \tag{20}$$

Let $\left|\left(\boldsymbol{j}_{k,i}\left(\boldsymbol{\nu}\right)\right)^{\mathrm{H}}\boldsymbol{H}_k\bar{\boldsymbol{b}}_{h,i}\left(\boldsymbol{\nu}\right)\right|^2/\sigma^2 = \lambda_i$, (20) can be rewritten by

$$\begin{aligned}R_{k,h}\left(\boldsymbol{B}_h\left(\boldsymbol{\nu}\right), \boldsymbol{J}_k\left(\boldsymbol{\nu}\right), \boldsymbol{\nu}\right) &= \sum_{i=1}^{\zeta}\log_2\left(1+\lambda_i\right),\\ &= \log_2\left(\left(1+\lambda_1\right)\times\cdots\times\left(1+\lambda_\zeta\right)\right),\\ &= \log_2\begin{vmatrix}1+\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1+\lambda_\zeta\end{vmatrix}.\end{aligned} \tag{21}$$

Therefore, we have

$$\begin{aligned}&R_{k,h}\left(\boldsymbol{B}_h\left(\boldsymbol{\nu}\right), \boldsymbol{J}_k\left(\boldsymbol{\nu}\right), \boldsymbol{\nu}\right)\\ &= \log_2\det\left(\boldsymbol{I}_\zeta+\left|\left(\boldsymbol{J}_k\left(\boldsymbol{\nu}\right)\right)^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)\right|^2/\sigma^2\right),\end{aligned} \tag{22}$$

where $\det(\cdot)$ represents the determinant of a square matrix, and $\boldsymbol{I}_\zeta$ is an $\zeta\times\zeta$ identity matrix. Substituting $\boldsymbol{B}_h\left(\boldsymbol{\nu}\right)$ and $\boldsymbol{J}_k\left(\boldsymbol{\nu}\right)$ into (22), we have

$$\begin{aligned}&R_{k,h}\left(\boldsymbol{B}_h\left(\boldsymbol{\nu}\right), \boldsymbol{J}_k\left(\boldsymbol{\nu}\right), \boldsymbol{\nu}\right)\\ &= \log_2\det\left(\boldsymbol{I}_\zeta+\frac{\left|\left(\boldsymbol{U}_k^{(1)}\right)^{\mathrm{H}}\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}\left(\frac{\sum_{i=1}^{K}\boldsymbol{V}_i^{(1)}}{\sqrt{|\mathcal{H}_h|}}\right)\sqrt{\frac{P}{H\zeta}}\right|^2}{\sigma^2}\right).\end{aligned} \tag{23}$$

Since $\boldsymbol{H}_i\left(\boldsymbol{H}_k\right)^{\mathrm{H}}=\mathbf{0}$ $(i\neq k)$, we have $\boldsymbol{H}_k\left(\boldsymbol{V}_i^{(1)}\right)^{\mathrm{H}}=\mathbf{0}$ $(i\neq k)$. Substituting (16) into (23), we have

$$R_{k,h}(\boldsymbol{B}_h(\boldsymbol{\nu}),\boldsymbol{J}_k(\boldsymbol{\nu}),\boldsymbol{\nu})=\log_2\det\left(\boldsymbol{I}_\zeta+\frac{P}{|\mathcal{H}_h|H\zeta\sigma^2}\left(\boldsymbol{\Sigma}_k^{(1)}\right)^2\right). \tag{24}$$

Then, the optimization problem in (11) can be transformed as

$$\begin{aligned}\max_{\boldsymbol{\nu}}\ &\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\left\{\log_2\det\left(\boldsymbol{I}_\zeta+\frac{P}{|\mathcal{H}_h|H\zeta\sigma^2}\left(\boldsymbol{\Sigma}_k^{(1)}\right)^2\right)\right\}\\ \text{s.t.}\quad &(10\text{c})\,.\end{aligned} \tag{25}$$

B. Phase Optimization with Manifold Method

1) Approximation of $\boldsymbol{\Sigma}_k^{(1)}$: Since $\boldsymbol{\Sigma}_k^{(1)}$ in (25) is unknown, we use the function of phase shift to represent $\boldsymbol{\Sigma}_k^{(1)}$, which is proved in Theorem 1.

Theorem 1: When the numbers of antennas of the BS and the users are more than 128, $\boldsymbol{\Sigma}_k^{(1)}(i,j)\approx\beta_i\alpha_j\boldsymbol{\nu}^{\mathrm{H}}\boldsymbol{c}^{ij}$, where $\boldsymbol{c}^{ij}=\left(\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)\right)^*\circ\boldsymbol{a}\left(\theta_j^{\mathrm{A}},\eta_j^{\mathrm{A}}\right)$ with $\left(\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)\right)^*$ being the conjugate of $\left(\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)\right)$ and $\circ$ being the Hadamard product.

Proof: See Appendix A. ■

From Theorem 1, we can see that $\boldsymbol{\Sigma}_k^{(1)}$ depends on the distance α_j between the BS and the IRS, the distance β_i between the IRS and user k, the angle $\boldsymbol{a}\left(\theta_j^{\mathrm{A}},\eta_j^{\mathrm{A}}\right)$ from the BS to the IRS, the phase shifts of the IRS, and the angle $\boldsymbol{a}\left(\theta_i^{\mathrm{D}},\eta_i^{\mathrm{D}}\right)$ from the IRS to user k.

2) Problem Transformation: Based on Theorem 1, the optimization problem (25) can be rewritten as

$$\max_{\boldsymbol{\nu}}\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\sum_{i=1}^{\zeta}\log_2\left(1+\frac{P}{|\mathcal{H}_h|H\zeta\sigma^2}\left|\boldsymbol{D}_k(i,i)\right|^2\right) \tag{26}$$

$$\text{s.t.}\quad (10\text{c})\,,$$

$$|d_{ij}|=\left|\boldsymbol{\nu}^H\boldsymbol{c}^{ij}\right|<\tau,\quad\forall i\neq j, \tag{26a}$$

where $d_{ii}=\boldsymbol{\nu}^H\boldsymbol{c}^{ii}$, $\boldsymbol{D}_k(i,i)=\alpha_i\beta_i d_{ii}$ $(i\in\{1,\ldots,\zeta\})$, and τ is a small positive value. Constraint (26a) is to make sure that $\boldsymbol{D}_k$ is approximately a non-square diagonal matrix such that $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}=\boldsymbol{A}_k\boldsymbol{D}_k\left(\boldsymbol{A}\right)^{\mathrm{H}}\left[\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1};\ldots;\boldsymbol{z}_{K\zeta}\right]$ can be treated as an approximation of the truncated SVD of $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}$, where $\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}$ denotes row $(K-|\mathcal{H}_h|)\zeta+1$ of matrix $\boldsymbol{Z}$. Constraint (26a) can be removed and this omission does not affect the validity of our proposed solution [4]. Hence, problem (26) can be rewritten as follows:

$$\begin{aligned}\max_{\boldsymbol{\nu}}\ &\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\sum_{i=1}^{\zeta}\log_2\left(1+\frac{P}{|\mathcal{H}_h|H\zeta\sigma^2}\left|\boldsymbol{D}_k(i,i)\right|^2\right)\\ \text{s.t.}\quad &(10\text{c})\,.\end{aligned} \tag{27}$$

Substituting $\boldsymbol{D}_k(i,i)=\alpha_i\beta_i\boldsymbol{\nu}^H\boldsymbol{c}^{ii}$ into (27), the problem (27) can be transformed as follows:

$$\begin{aligned}\max_{\boldsymbol{\nu}}\ &\sum_{h=1}^{H}\min_{k\in\mathcal{H}_h}\left\{\sum_{i=1}^{\zeta}\log_2\left(1+b_i\boldsymbol{\nu}^H\boldsymbol{C}^{ii}\boldsymbol{\nu}\right)\right\}\\ \text{s.t.}\quad &(10\text{c})\,,\end{aligned} \tag{28}$$

Algorithm 1 Proposed Scheme for Solving Problem (10)

1: **Input:** $\boldsymbol{H}^{\mathrm{B}},\boldsymbol{H}_k^{\mathrm{R}},\zeta,P,\sigma^2$.
2: Calculate $\boldsymbol{B}_h(\boldsymbol{\nu})$ and $\boldsymbol{J}_k(\boldsymbol{\nu})$ by Lemma 1.
3: Find the phase shift $\hat{\boldsymbol{\nu}}$ of the IRS by solving problem (28).
4: Obtain $\boldsymbol{B}_h(\hat{\boldsymbol{\nu}})$ and $\boldsymbol{J}_k(\hat{\boldsymbol{\nu}})$ by Lemma 1.
5: Calculate $\hat{\boldsymbol{F}}^{\mathrm{B}},\hat{\boldsymbol{F}}^{\mathrm{R}},\hat{\boldsymbol{W}}^{\mathrm{B}}$, and $\hat{\boldsymbol{W}}^{\mathrm{R}}$ by the algorithm in [13].
6: **Output:** $\hat{\boldsymbol{\nu}},\hat{\boldsymbol{F}}^{\mathrm{B}},\hat{\boldsymbol{F}}^{\mathrm{R}},\hat{\boldsymbol{W}}^{\mathrm{B}},\hat{\boldsymbol{W}}^{\mathrm{R}}$.

TABLE I
SIMULATION PARAMETERS

Parameters	Values	Parameters	Values
M	16×16	N^{B}	64
N^{U}	64	M^{B}	8
M^{U}	4	ζ	4
$Y=L$	7	G_{t}	24.5 dBi
G_{r}	0 dBi	σ^2	-90 dBm

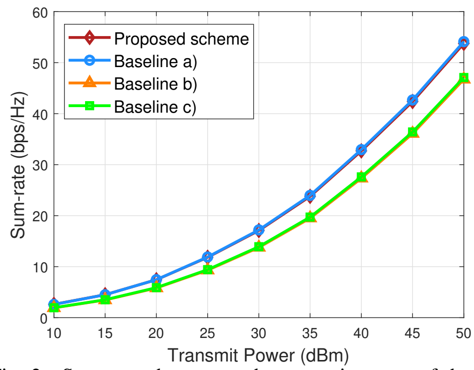

Fig. 2. Sum-rate changes as the transmit power of the BS.

where $\boldsymbol{C}^{ii}\triangleq\boldsymbol{c}^{ii}\left(\boldsymbol{c}^{ii}\right)^{\mathrm{H}}$ and $b_i\triangleq\frac{P}{|\mathcal{H}_h|H\zeta\sigma^2}|\alpha_i\beta_i|^2$. Since constraint (10c) is non-convex unit modulus, we use a manifold method [15] to solve problem (28).

Given $\hat{\boldsymbol{\nu}}$ obtained via solving problem (28), we can use Lemma 1 to calculate $\boldsymbol{B}_h(\hat{\boldsymbol{\nu}})$ and $\boldsymbol{J}_k(\hat{\boldsymbol{\nu}})$. Then, we use the algorithm in [13] to find $\hat{\boldsymbol{F}}^{\mathrm{B}},\hat{\boldsymbol{F}}^{\mathrm{R}},\hat{\boldsymbol{W}}^{\mathrm{B}}$, and $\hat{\boldsymbol{W}}^{\mathrm{R}}$.

IV. Simulation Results

In our simulations, the coordinates of the BS and the IRS are (2m, 0m, 10m) and (0m, 148m, 10m), respectively. All users are randomly distributed in a circle centered at (7m, 148m, 1.8m) with radius 10 m. The values of other parameters are defined in Table I. For comparison purposes, we use three baselines: a) the fully digital beamforming matrices of the BS and the users as well as the phase of each element of the IRS are optimized by the proposed scheme (using proposed scheme to solve problem (11)), b) the hybrid beamforming matrices of the BS and the users are determined by the proposed algorithm and the phase shifts of the IRS are randomly determined, and c) the fully digital beamforming matrices of the BS and the users are determined by the proposed scheme while the phase shifts of the IRS are randomly determined.

Fig. 2 shows how the sum rate of all users changes as the transmit power of the BS varies. From Fig. 2, we see that, the performance achieved by the proposed algorithm is similar to the optimal performance achieved by baseline a). The is because the proposed shceme can find the optimal hybrid beamforming matrices to represent the fully digital

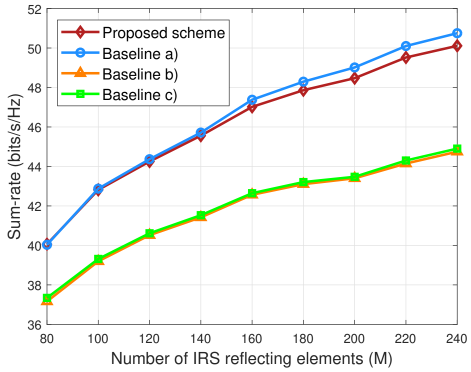

Fig. 3. Sum-rate versus number of reflecting elements at the IRS.

matrices. Fig. 2 also shows that compared to baseline b), the proposed scheme can achieve up to 14.93% gain in terms of the sum rate of all users when P=50 dBm and M=256. This is because the proposed scheme optimizes the phase shifts of the IRS by a manifold method.

In Fig. 3, we show how the sum of all users' data rates changes as the number of reflecting elements at the IRS varies. Fig. 3 shows that the proposed scheme can achieve up to 14% gain in terms of the sum of all users' data rates compared to baseline b) when M=240. This is due to the fact that in the proposed scheme, the phase of each element of the IRS is optimized by a manifold method, which can align the angles of the cascaded channel and well deal with module-one constraint.

V. Conclusions

In this paper, we have developed a novel framework for an IRS-assisted mmWave multigroup multicast MIMO communication system. The transmit beamforming matrices of the BS, the receive beamforming matrices of the users, and the phase shifts of the IRS were jointly optimized to maximize the sum rate of all users. We have used a BD method to represent the beamforming matrices of the BS and the users by the IRS phase shifts. Then, we have transformed the original problem to a problem that only needs to optimize the IRS phase shifts. The transformed problem is solved by a manifold method. Simulation results show that the proposed scheme can achieve significant performance gains compared to baselines.

Appendix A

To prove Theorem 1, we first define the effective channel $\boldsymbol{H}_k$ as done in [4]

$$\boldsymbol{H}_k = G_{\rm t}G_{\rm r}\boldsymbol{H}_k^{\rm R}\boldsymbol{\Phi}\boldsymbol{H}^{\rm B} = \boldsymbol{A}_k\boldsymbol{D}_k\left(\boldsymbol{A}\right)^{\rm H}, \tag{29}$$

where $\boldsymbol{A}_k = \left[\boldsymbol{a}\left(r_{1,k}^{\rm A}\right), \ldots, \boldsymbol{a}\left(r_{L,k}^{\rm A}\right)\right]$ is a array response matrix of user k, $\boldsymbol{A} = \left[\boldsymbol{a}\left(r_1^{\rm D}\right), \ldots, \boldsymbol{a}\left(r_Y^{\rm D}\right)\right]$ is a array response matrix of the BS, and $\boldsymbol{D}_k$ is an $Y \times L$ matrix with element $\boldsymbol{D}_k\left(i,j\right) = \beta_i\alpha_j d_{ij}$, where $d_{ij} = \left(\boldsymbol{a}\left(\theta_i^{\rm D},\eta_i^{\rm D}\right)\right)^{\rm H}\boldsymbol{\Phi}\boldsymbol{a}\left(\theta_j^{\rm A},\eta_j^{\rm A}\right) = \boldsymbol{\nu}^{\rm H}\boldsymbol{c}^{ij}$.

Given the effective channel $\boldsymbol{H}_k$, we define $\tilde{\boldsymbol{H}}_h$ as

$$\begin{aligned}\tilde{\boldsymbol{H}}_h &= \left[\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{h-1}, \boldsymbol{H}_{h+1}, \ldots, \boldsymbol{H}_H\right]^{\rm T} \\ &= \begin{bmatrix}\boldsymbol{A}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{A}_K\end{bmatrix}\boldsymbol{P}\begin{bmatrix}\tilde{\boldsymbol{\Sigma}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{Q}\begin{bmatrix}\left(\boldsymbol{A}\right)^{\rm H} \\ \vdots \\ \left(\boldsymbol{A}\right)^{\rm H}\end{bmatrix}.\end{aligned} \tag{30}$$

Let $\boldsymbol{Z} = \boldsymbol{Q}\begin{bmatrix}\left(\boldsymbol{A}\right)^{\rm H} \\ \vdots \\ \left(\boldsymbol{A}\right)^{\rm H}\end{bmatrix}$, $\tilde{\boldsymbol{V}}_h^{(0)} = \left[\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}; \ldots; \boldsymbol{z}_{K\zeta}\right]$, with $\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}$ being row $\left(K-|\mathcal{H}_h|\right)\zeta+1$ of matrix $\boldsymbol{Z}$. Based on (29), $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)} = \boldsymbol{A}_k\boldsymbol{D}_k\left(\boldsymbol{A}\right)^{\rm H}\left[\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}; \ldots; \boldsymbol{z}_{K\zeta}\right]$. For ULA with N antennas, when N is more than 128, the column vectors of $\boldsymbol{A}_k$ and row vectors of $\left(\boldsymbol{A}\right)^{\rm H}\left[\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}; \ldots; \boldsymbol{z}_{K\zeta}\right]$ can form orthonormal sets [4]. Hence, $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)} = \boldsymbol{A}_k\boldsymbol{D}_k\left(\boldsymbol{A}\right)^{\rm H}\left[\boldsymbol{z}_{(K-|\mathcal{H}_h|)\zeta+1}; \ldots; \boldsymbol{z}_{K\zeta}\right]$ can be considered as an approximation of the truncated SVD of $\boldsymbol{H}_k\tilde{\boldsymbol{V}}_h^{(0)}$, and $\boldsymbol{D}_k$ can represent $\boldsymbol{\Sigma}_k^{(1)}$.

References

[1] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.

[2] H. Du, J. Zhang, J. Cheng, and B. Ai, "Millimeter wave communications with reconfigurable intelligent surfaces: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2752–2768, Apr. 2021.

[3] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14 960–14 973, Dec. 2020.

[4] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, Feb. 2021.

[5] C. Pradhan, A. Li, L. Song, B. Vucetic, and Y. Li, "Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041–1045, Jul. 2020.

[6] K. Ying, Z. Gao, S. Lyu, Y. Wu, H. Wang, and M.-S. Alouini, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19 530–19 539, Jan. 2020.

[7] W. Zhang, J. Xu, W. Xu, D. W. K. Ng, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, Feb. 2021.

[8] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1379–1393, Feb. 2021.

[9] W. Huang, Y. Huang, S. He, and L. Yang, "Cloud and edge multicast beamforming for cache-enabled ultra-dense networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3481–3485, Mar. 2020.

[10] C. Kai, W. Ding, and W. Huang, "Max-min fairness in IRS-aided MISO broadcast channel via joint transmit and reflective beamforming," in *Proc. IEEE Global Commun. Conf.*, Taipei, Taiwan, Dec. 2020, pp. 1–6.

[11] Q. Tao, S. Zhang, C. Zhong, and R. Zhang, "Intelligent reflecting surface aided multicasting with random passive beamforming," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 92–96, Jan. 2021.

[12] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast MISO communication systems," *IEEE Trans. Signal Processing*, vol. 68, pp. 3236–3251, Apr. 2020.

[13] H. Kasai, "Fast optimization algorithm on complex oblique manifold for hybrid precoding in millimeter wave MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Anaheim, USA, Nov. 2018, pp. 1266–1270.

[14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[15] R. M. P.-A. Absil and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.