IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 September 2022, accepted 19 October 2022, date of publication 25 October 2022,
date of current version 28 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3216897

RESEARCH ARTICLE

Manifold Learning Inspired Dynamic Hybrid Precoding With Antenna Partitioning Algorithm for Dual-Hop Hybrid FSO-RF Systems

XIAOPING ZHOU, XUDONG TIAN, LE TONG, AND YANG WANG
College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

Corresponding authors: Xiaoping Zhou (e-mail: zxpshnu@163.com) and Le Tong (e-mail: ton-gle@shnu.edu.cn)

This work was supported in part by the Shanghai Capacity Building Projects in Local Institutions under Grant 19070502900 and in part by the Shanghai Sailing Program under Grant 19YF1437100.

ABSTRACT The decode-and-forward (DF) based free-space optical-radio frequency (FSO-RF) hybrid system combines the advantages of both FSO links and RF links, which makes the system easy to be deployed and enables the extended transmission coverage. In the case of multiple users, the unreasonable dynamic antenna selection algorithm will lead to multi-user interference (MUI) and reduce the spectral efficiency of the system. Therefore, we propose hybrid precoding based on manifold learning with the antenna partitioning algorithm for dual-hop hybrid FSO-RF systems. In the hybrid precoding of dynamic subarray structure, the high-dimensional channels are embedded into the low-dimensional manifolds by the optimized multidimensional scaling (MDS). The potential spatial correlations of the high-dimensional channel are preserved by the scaling by majorizing a complicated function (SMACOF) algorithm. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi-conjugate gradient methods. Meanwhile, an antenna subarray partitioning algorithm is proposed, so that each antenna unit can be assigned to an RF chain based on the increment of the user's maximum signal to interference noise ratio (SINR). By calculating the simulated equivalent channel SINR for the selected users, the antenna division can greatly reduce the computational complexity and the size of the search space, and ensure fairness among users. Simulation results show that this method can obtain almost the best summation rate and higher spectral efficiency compared with the conventional method.

INDEX TERMS Antenna partitioning algorithm, hybrid precoding, hybrid RF/FSO systems, manifold learning.

I. INTRODUCTION

With the increasing demand for faster data rate along with the integration of a wide range of devices into the network, there is a necessity to explore new paradigms of wireless communication. Homogeneous physical layer communication, such as radio frequency (RF), laser, and optical fiber, are morphing into a hybrid combination amongst each other to improve the service provided to the large variety of devices [1]. Dual-hop hybrid free-space optical-radio frequency (FSO-RF) relay systems combine the advantages of free-space optical (FSO) and RF technologies to provide superior performance [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Yanxiang Jiang.

In [4], Safari and Uysal first introduced relaying into a dual-hop hybrid FSO-RF system to improve transmission performance through relaying and subsequently published work studying many dual-hop FSO-RF systems [5], [6], [7]. However, previous studies on dual-hop FSO-RF systems have mainly focused on FSO link or relay selection problems. RF links were not considered. We consider a hybrid FSO/RF system where the RF link utilizes a millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) system and is applied to target multi-user communications.

MmWave systems have been identified as a promising solution to cope with the explosive growth of mobile traffic [8], [9]. For mmWave massive MIMO systems, hybrid precoding is efficient transceivers that can achieve performance

close to that of a fully digital precoding with a limited number of RF chains [10], [11], [12]. In Multi-user mmWave massive MIMO systems, hybrid precoding is mainly used for signaling in a fully connected architecture [12], [13], [14], [15]. In such architecture, each RF chain is connected to all antennas via phase shifters and RF adders, and the number of required analog phase shifters is as large as the number of antenna elements. During the increasing number of transmitting antennas, although the channel capacity will also increase, the excessive number of analog phase shifters and RF adders will lead to high hardware cost and power consumption [16], and the coding matrix will become more complex.

Moreover, it cannot be ignored that RF adders with multiple inputs are computationally complex and the use of a large number of RF adders is not feasible in practical applications. To address these challenges and to reduce the number of phase shifters and RF adders, mapping methods between RF chains and antennas have attracted great attention [17], [18]. Partially connected architectures can significantly reduce the number of phase shifters and eliminate the need for RF adders compared to fully connected architectures. This deployment can effectively lead to a reduction in both hardware cost and power consumption [17], [18], [19], [20]. Partial-connected architectures can be divided into two categories: fixed subarrays and dynamic subarrays. In the fixed subarray architecture, each RF chain is connected with a fixed antenna subset. Meanwhile, each antenna is connected to a single RF chain [21]. In a dynamic subarray scenario, antenna elements are adaptively partitioned into several subsets based on the long-term channel information [22], [23]. For mmWave massive MIMO systems, the dynamic subarray strikes a balance between sum rate and hardware complexity [22]. However, the problem of multi-user hybrid precoding in dynamic subarray architecture is difficult to solve, and the antenna division will potentially lead to unfairness among users and multi-user interference (MUI). There have been some studies on multi-user hybrid precoding in dynamic subarray architectures.

Park, et al. [22] proposed a Minkowski l_1-norm based approach for reducing the complexity of calculating the maximum eigenvalue in the process of assigning antennas based on the exponential channel model. The singular value decomposition (SVD) of the preceding equivalent analog channel in the digital baseband part is used to design the digital precoder. In [24], it is assumed that each antenna is assigned two phase shifters, which gives the analog precoder the ability to control the magnitude, and the problem is modeled as a fitting problem for the best digital precoder. Ultimately, the antenna partitioning problem is transformed into a clustering problem with better system performance relative to the greedy approach. [25] proposed a low complexity greedy (LCG) growth approach to implement dynamic subarray configurations and degrades the computational complexity using the Lanczos method. In [26], an extensive search was used to maximize the analog effective channel gain with virtual path selection (VPS) to accomplish antenna partitioning and analog precoding. The digital precoding was used to suppress the MUI by using the zero-forcing (ZF) criteria based on the analog precoding channel with low dimensions. In addition, [27] presented a multi-user analog precoding technique. In the multi-user MIMO (MU-MIMO) system, based on the channel gain, the antenna is assigned to the RF link with the highest channel gain each time. The analog precoder's phase is then calculated as the quantized phase of the associated column vector. The two sub-steps above were repeated until all MU-MIMO users had been finished. The goal of all current research on dynamic subarrays is system and rate improvement, however this can lead to uneven resource allocation among users, and users at the edge of the base station or with poor channel conditions may become very low.

Recently, manifold learning has been proposed to integrate with mmWave massive MIMO systems. In [28], a manifold optimization (MO) based hybrid precoding algorithm, as well as some low-complexity algorithms, was proposed. A Riemann conjugate gradient manifold algorithm is proposed to solve the generalized eigenvalue problem by maximizing the objective function of each user alternatively [29]. By designing the massive MIMO precoding matrix as an optimization problem at the intersection of oblique and Stiefel manifolds, the steepest descent method is used at the intersection of these two manifolds to obtain the optimal solution [30]. Compared with the methods of gradient descent and constant envelope optimization, the complexity of solving nonlinear least squares problem is much lower with this optimization method. A Riemannian trust-region Newton manifold (RTRNM) is proposed for the optimization beamforming in multi-cluster scenarios [31]. A manifold learning two-tier fully-digital beamforming scheme optimizes resource management in massive MIMO networks [32]. The multi-user high-dimensional channels are reduced using the manifold learning algorithm. It minimizes computing complexity while also preventing inter-cell interference in fully-digital beamforming. It concentrates on the local linear spatial structure between user channels while overlooking the global spatial qualities.

In this paper, we consider the Manifold learning inspired dynamic hybrid precoding for dual-hop FSO-RF systems with antenna partitioning algorithm. For scenarios where users are dense, the users order in the procedure of the antenna selection leads to severe unfairness since the first user is able to choose the whole antenna elements and the other users can only choose the remaining elements, and in which there is also MUI. In the hybrid precoding of dynamic subarray structure, the high-dimensional channels are embedded into the low-dimensional manifolds by the optimized multidimensional scaling (MDS) based on the scaling by majorizing a complicated function (SMACOF) algorithm, while preserving the potential spatial correlation of the high-dimensional channel. In the proposed optimized MDS, we assign different weights according to the channel correlation coefficient and the SMACOF algorithm is used to improve the results of

MDS based on the assigned weights. Then the hybrid precoding was investigated for total rate maximization by appropriate user clustering using the manifold quasi-conjugate gradient method [33]. And we also propose an antenna subarray partitioning algorithm in this process that assigns each antenna unit to an RF chain based on the increment of the user's maximum signal to the interference noise ratio (SINR). By calculating the simulated equivalent channel SINR for the selected users, the computational complexity of antenna assignment and the size of the search space can be greatly reduced. Performance evaluations show that the proposed scheme can obtain a near-optimal sum-rate and considerably higher spectral efficiency than the conventional schemes.

The remainder of this paper is organized as follows. Section II introduces the system model and hybrid precoder design problem. Section III presents dimensionality reduction and clustering of channel matrix for manifold learning. Section IV introduces the hybrid precoding algorithm based on channel dimensionality reduction and subarray antenna partitioning algorithm. In Section V, we present the algorithms computational complexity analysis. Some simulation results are provided in Section VI. Finally, we conclude this paper in Section VII.

Notations: Bold lower-case and upper-case letters are used for vectors and matrices, respectively, while regular letters denote scalars. $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, $tr(\cdot)$, and $\|\cdot\|_F$ are the Hermitian transpose, inverse, transpose, complex conjugate, trace, and Frobenius norm of a matrix, respectively. $E(\cdot)$ is the channel covariance matrix. $|\mathcal{G}|$ is the cardinality of the set $\mathcal{G}$. $\otimes$ indicates the Kronecker product. $\circ$ denotes the Hadamard matrix products. $\mathcal{CN}\left(0, \sigma^2\right)$ represents the zero-mean complex Gaussian distribution with zero mean and the variance σ^2. $\mathbb{C}^{m\times n}$ denotes m × n complex matrices; $\bigcup_{i=1}^{N} X_i$ denotes the union operation of the N sets X_i $(i = 1, \ldots, N)$, and $\cap$ is the intersection of the two sets. The empty set is denoted by $\emptyset$. *span* (Y) denotes the subspace spanned by the column vectors of Y. $\Delta(\cdot)$ indicates gradient. $\nabla(\cdot)$ indicates increment. Ξ denotes the rank. $G_{p,q}^{m,n}\left(x \middle| \begin{matrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{matrix}\right)$ defined the Meijer's G-function.

II. SYSTEM MODEL AND CHANNEL MODEL

The block diagram of a mixed FSO-RF system is presented in Fig. 1 in which the source node (S) communicates with the destination node (D) through a decode-and-forward (DF) relay node (R). S-R link is equipped with a single antenna, and the R-D link is a mmWave massive MIMO system model. The node R has both optical and RF signal processing capabilities. We employ non-coherent intensity modulation with direct detection (IM/DD) receiver at R. After converting the incoming optical signal to an electrical signal, node R utilizes a power splitter to separate out the alternating current (AC) and direct current (DC) components. The unsolicited DC component (which is normally filtered out at the receiver) is applied to the energy harvesting unit which supplies the harvested power to the RF transmitter. The information-bearing AC component is given to the decoder circuit which decodes the information and remodulates using an RF modulation scheme before forwarding it through the RF transmitter. Fig. 2 illustrates the workflow of DF.

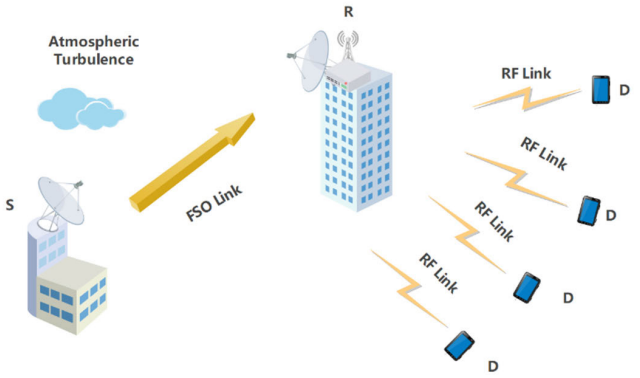


FIGURE 1. Block diagram of mixed FSO-RF system.

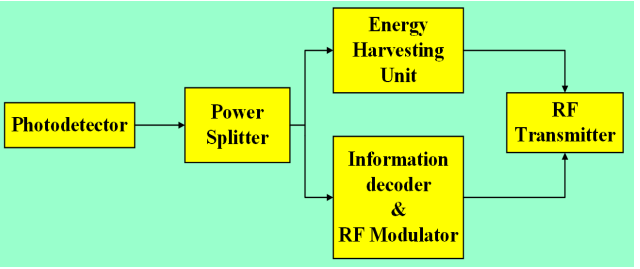


FIGURE 2. DF workflow diagram.

A. FSO SYSTEM MODEL

The FSO channel between S and R is modelled as Gamma-Gamma distribution with pointing error. The statistical behavior of the received optical irradiance is characterized by means of Gamma-Gamma turbulence model which has been widely used to model the FSO channel in the recent literature depending on its doubly stochastic scintillation model [34]. The probability distribution function (PDF) of channel coefficient h_{FSO} is given as [35]

$$f_{h_{FSO}}(h_{FSO}) = \frac{\varsigma^2}{h_{FSO}\Gamma(\alpha)\Gamma(\beta)} G_{1,3}^{3,0}\left(\alpha\beta \frac{h_{FSO}}{I_{FSO}M_o} \middle| \begin{matrix} \varsigma^2+1 \\ \varsigma^2,\alpha,\beta \end{matrix}\right), \tag{1}$$

where $\Gamma(.)$ is the well-known Gamma function [36], $G_{p,q}^{m,n}\left(x \middle| \begin{matrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{matrix}\right)$ is the Meijer's G-function defined in [37]. The atmospheric turbulence induced fading channel gains of FSO links, denoted as I_{FSO} is modeled with pointing errors and path loss. It is assumed that the elements of I is modeled as independent and identically distributed (i.i.d) random variable. $\varsigma = \varpi_e/2\sigma_s$ is the ratio of equivalent beam radius and zero boresight pointing error displacement standard deviation at the photo-detector (PD), the constant term M_o is the power fraction that the detector receives when there is no pointing

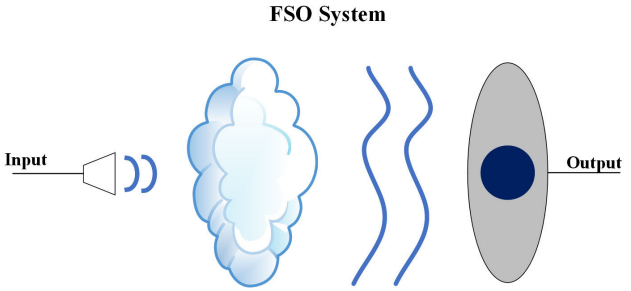


FIGURE 3. FSO system model.

error [38], and $1/\alpha$ and $1/\beta$ are the variances of the large and small turbulence eddies, respectively. The FSO system modeled as show in Fig. 3.

The parameters α and β are the distance-dependent fading variables which corresponds to atmospheric turbulence conditions as [38]

$$\alpha = \left\{ \exp\left[\frac{0.49\sigma_R^2}{\left(1 + 1.11\sigma_R^{\frac{12}{5}}\right)^{\frac{7}{6}}} \right] - 1 \right\}^{-1}, \tag{2}$$

and

$$\beta = \left\{ \exp\left[\frac{0.51\sigma_R^2}{\left(1 + 0.69\sigma_R^{\frac{12}{5}}\right)^{\frac{5}{6}}} \right] - 1 \right\}^{-1}, \tag{3}$$

where $\sigma_R^2 = 1.23 C_n^2 v^{7/6} d_{SR}^{11/6}$ denoting the Rytov variance, C_n^2, v, and d_{SR} represent the refractive index structure constant, wave number, and the link length, respectively. C_n^2 usually takes values in the range $10^{-17} - 10^{-13}$ for weak up to strong turbulence conditions.

B. DF TRANSMISSION PROTOCOL

The communication between S and D is accomplished in two time slots T_1 and T_2 for first and second hop, respectively. The relay harvests energy during T_1 only as simultaneous harvesting and discharging of power increases the complexity at the node. The S employs sub-carrier intensity modulation (SIM) to convert RF signal vector r with electrical power ρ to an optical signal. A DC bias $A \in [A_{\min}, A_{max}]$ (where $A_{\min}$ and A_{max} are minimum and maximum values of DC bias, respectively) is added to the RF signal to ensure a non-negative optical signal. Let P_S represents the electrical power of S for transmitting the optical signal vector s, then we can write

$$s = \sqrt{P_S}\left[\delta r + A\right], \tag{4}$$

where δ is electrical to optical conversion coefficient. In order to prevent clipping due to non-linearity of laser diode such that it operates in linear region, δ should satisfy the following constraint [39]

$$\delta \leq \min\left(\frac{A - A_{\min}}{\rho}, \frac{A_{\max} - A}{\rho}\right). \tag{5}$$

The electrical signal at the output of the PD, can be expressed as

$$y_{FSO} = h_{FSO} s + n_{FSO}, \tag{6}$$

where $h_{FSO} = \left(P_S^2 \eta^2 / \sigma_{FSO}^2\right) I_{FSO}$ is the channel coefficient of S-D link, where η is optical to electrical conversion coefficient. The noise n_{FSO} is due to circuit noise as well as high-intensity background illumination and is conventionally modelled as being zero mean additive white Gaussian noise. σ_{FSO}^2 is the variance of additive white Gaussian noise with zero mean.

C. HYBRID MMWAVE MASSIVE MIMO SYSTEM MODEL

The channel between R and D is a hybrid mmWave massive MIMO system model consisting of B cells. We assume that a base station (BS) equipped with N_t antenna and N_{tRF} RF chains ($N_t \geq N_{tRF} \geq K$) serves K users. The transmitter employs a uniform planar array (UPA) with $N_t = N_{tx} \times N_{ty}$ antennas and N_{tRF} RF chains to simultaneously transmit N_S data streams to the receiver which is equipped with $N_r = N_{rx} \times N_{ry}$ antennas and N_{rRF} RF chains, i.e. $N_t \gg N_{tRF} \geq N_s$. The hybrid analog and digital beamforming structures are employed by both the transmitter and the receiver, as shown in Fig. 4. To manage the interference and improve the data rate for users, the users are partitioned into G clusters $\mathcal{G}_1, \cdots, \mathcal{G}_G$, $g_i = |\mathcal{G}_i|$, $\sum_{i=1}^{G} g_i = K$ and $\mathcal{G}_i \cap \mathcal{G}_{i'} = \emptyset, \forall i \neq i'$. $\mathcal{G}_i$ is the i-th cluster, where $i = 1, \ldots, G$. The sets $\{\mathcal{G}_1, \cdots, \mathcal{G}_G\}$ are all user clusters.

Let $u_{b,i,k}, k = 1, \ldots, |\mathcal{G}_i|$ denote the k-th user of $\mathcal{G}_i$ in the b-th cell ($b = 1, 2, \cdots, B$). The vector x is firstly precoded by a digital precoding matrix $\mathbf{F}_{BB} \in \mathbb{C}^{N_{tRF} \times N_S}$, in the frequency domain. After conversion to the time domain by the inverse discrete fourier transform (IDFT) and adding the cyclic prefix, the signals are up-converted to the RF domain via N_{tRF} RF chains and further precoded in the analog domain. Let $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_{tRF}}$ be the analog precoding matrix. In order to reduce the hardware complexity and power consumption, N_t low-resolution phase shifters (PSs) are employed to implement the analog beamformers. In this paper, we adopt a dynamic mapping between RF chains and transmit/receive antennas. Specifically, each RF chain can be dynamically connected to a disjoint set of antennas via a switch (SW) network and corresponding low-resolution PSs. While switches designed for mmWave systems can switch at speeds on the order of nanoseconds or even sub-nanoseconds. Therefore, the switching rates and settling times for currently available RF switches are sufficient for the proposed wideband mmWave systems.

In light of these, we propose a dynamic array-of-subarrays hybrid precoding architecture to intelligently adjust the

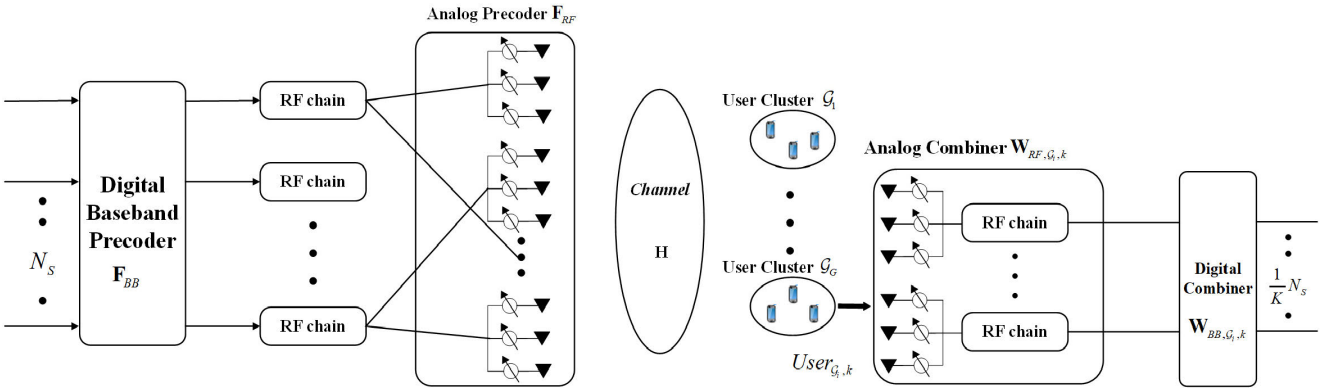


FIGURE 4. Hybrid precoding architecture in MU-MIMO system with dynamic subarrays.

connections between RF chains and subarrays through a network of switches. Specifically, the transmitting antennas are divided into N_q subarrays, where N_q is the number of RF chains. Let $n_t = N_t/N_q$ denote the number of antennas of each subarray. Each RF chain connects to each subarray through a switch. $\mathbf{Q} \in \mathbb{C}^{N_t \times N_{tRF}}$ represents the switch network matrix.

After propagating through the wideband mmWave MIMO channel, the signal is corrupted by additive white Gaussian noise (AGWN). At the receiver the received signal is first processed by N_r low-resolution PSs and SWs which form the analog combiner $\mathbf{W}_{RF} \in \mathbb{C}^{N_{rRF} \times N_r}$. Then the received signal is down-converted to baseband through N_{rRF} RF chains. After removing the cyclic prefix and applying the discrete Fourier transform (DFT), the received frequency-domain signal on each subcarrier is further processed by an individual digital combiner $\mathbf{W}_{BB} \in \mathbb{C}^{N_{rRF} \times N_s}$.

Let S_j denote the collection of subarrays connected to the j-th RF chain. We partition N_q subarrays into N_{tRF} subsets as

$$\bigcup_{j=1}^{N_{tRF}} S_j = \{1, 2, \cdots, N_q\}. \tag{7}$$

Each subarray can be allocated to only one RF chain, i.e.,

$$S_c \cap S_j = \emptyset, \quad c \neq j. \tag{8}$$

Each RF chain will select at least one subarray, i.e.,

$$S_j \neq \emptyset. \tag{9}$$

The overload of RF chains is ignored in this work [27]. Moreover, since phase shifters can only change the phases of the signals, the amplitude of the entries of the analog precoding matrix is constant. Specifically, the hybrid precoder $\mathbf{F}_{RF}$ can be rewritten as $\mathbf{F}_{RF} = \left(\mathbf{f}_{cj}\right)_{N_q \times N_{tRF}}$, where $\mathbf{f}_{cj} \in \mathbb{C}^{n_t \times 1}$ is the analog precoder for the q-th subarray allocated to the j-th RF chain. Then, the unit modulus constraint on $\mathbf{F}_{RF}$ can be expressed by

$$\left|\mathbf{f}_{cj}\right| = \mathbf{1}_{S_j}(c), \tag{10}$$

where the indicator function $\mathbf{1}_{S_j}(c)$ is defined as

$$\mathbf{1}_{S_j}(c) = \begin{cases} 1, & \text{if } c \in S_j, \\ 0, & otherwise, \end{cases} \tag{11}$$

where $\mathbf{1}$ is an $n_t \times 1$ vector having ones in all entries, and $\mathbf{0}$ is an $n_t \times 1$ vector having zeros in all entries. n_t is the number of antennas in each subarray at the transmitter.

For hybrid precoding, the extension from the narrowband to the wideband is not straightforward. This is because in wideband mmWave MIMO systems, the digital precoding matrices $\{\mathbf{F}_{\mathrm{BB}}\}$ are performed on each subcarrier in frequency domain, while the analog precoding matrix $\mathbf{F}_{RF}$ is performed on the entire bandwidth in time domain. Due to the total transmit power constraint, the coupling among the analog precoder, the digital precoder and the subarray design matrix should satisfy

$$\sum_{k=1}^{K} \|\mathbf{F}_{RF}\mathbf{F}_{\mathrm{BB}}\|_F^2 \leq P_R, \tag{12}$$

where P_R denotes the hybrid mmwave massive MIMO chain transmit power.

We adopt a geometric channel model with L scattering clusters and R_l scatterers within the l-th cluster ($1 \leq l \leq L$). The modified delay-τ channel matrix of the k-th user can be written as [40]

$$\mathbf{H}_k(\tau) = \sum_{l=1}^{L} \sum_{r_l=1}^{R_l} \mu_{r_l} p\left(\tau T_s - \tau_l - \tau_{r_l}\right) a_R\left(\phi_{R,l} - \theta_{R,r_l}\right) \times a_T^H\left(\phi_{T,l} - \theta_{T,r_l}\right), \tag{13}$$

where τ_l, $\phi_{R,l}$ and $\phi_{T,l}$ denote the time delay, angles of arrive and departure (AOA/AOD), respectively. The parameters μ_{r_l}, τ_{r_l}, θ_{R,r_l}, and θ_{T,r_l} are the complex path gain, the relative time delay, and relative AOA/AOD shift for r_l scatterers in the l-th cluster, respectively; $p\left(dT_s - \tau_l - \tau_{r_l}\right)$ is a pulse-shaping function for T_s spaced signaling evaluated at $dT_s - \tau_l - \tau_{r_l}$ seconds. Furthermore, $a_R(\cdot)$ and $a_T(\cdot)$ are the receive and transmit array response vectors respectively. For an $N_{tx} \times N_{ty}$

UPA with N_t total antennas, the steering vector of the l-th cluster is given by [18]

$$a(\phi_l, \theta_l) = a_x(\phi_l, \theta_l) \otimes a_y(\theta_l), \tag{14}$$

where $\otimes$ is the Kronecker-product operation. $a_y(\theta_l)$ denotes the elevation steering vector of the l-th cluster, which has a form of

$$a_y(\theta_l) = \left[1, e^{j\frac{2\pi}{\lambda}d_y\cos\theta_l}, \cdots, e^{j(N_{ty}-1)\frac{2\pi}{\lambda}d_y\cos\theta_l}\right]^T, \tag{15}$$

where λ is the wavelength of the signal, d_y is the elevation array aperture. $a_x(\phi_l, \theta_l)$ denotes the azimuth steering vector of the l-th cluster, which is given by

$$a_x(\phi_l, \theta_l) = \left[1, e^{j\frac{2\pi}{\lambda}d_x\sin\theta_l\sin\phi_l}, \cdots, e^{j(N_{tx}-1)\frac{2\pi}{\lambda}d_x\sin\theta_l\sin\phi_l}\right]^T, \tag{16}$$

where d_x is the azimuth array aperture.

Let $x = \sqrt{P_R}y_{FSO}$ denotes the signal forwarded by the node R, then the received signal at D can be expressed as

$$\begin{aligned} y_{b,i,k} &= \mathbf{W}_{BB,b,i,k}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i,k}^H\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b,i,k}\right)\mathbf{F}_{BB,b,i,k}x_{b,i,k} \\ &+ \sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\mathbf{W}_{BB,b,i,k}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i,k}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b,i,k'}\right) \\ &\times\mathbf{F}_{BB,b,i,k'}x_{b,i,k'} \\ &+ \sum_{i'=1,i'\neq i}^{L}\mathbf{W}_{BB,b,i'}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i'}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b,i'}\right)\mathbf{F}_{BB,b,i'}x_{b,i'} \\ &+ \sum_{b'=1,b'\neq b}^{B}\mathbf{W}_{BB,b,i'}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i'}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b',i'}\right)\mathbf{F}_{BB,b',i'}x_{b',i'} \\ &+ \mathbf{W}_{BB,b,i,k}^H\mathbf{W}_{RF}^H n_{b,i,k}, \quad \forall i, \end{aligned} \tag{17}$$

where $\circ$ denotes the Hadamard matrix products. $\mathbf{H}_{b,i,k} \in \mathbb{C}^{N_t\times N_r}$ is the channel vector between the BS and user $u_{b,i,k}$. $n_{b,i,k} \sim \mathcal{CN}(0,\sigma^2)$ is the spatially additive white Gaussian noise with variance $\sigma_{b,i,k}^2$. $\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\mathbf{W}_{BB,b,i,k}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i,k}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b,i,k'}\right)\mathbf{F}_{BB,b,i,k'}x_{b,i,k'}$ are intra-cluster interference, $\sum_{i'=1,i'\neq i}^{L}\mathbf{W}_{BB,b,i'}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i'}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b,i'}\right)\mathbf{F}_{BB,b,i'}x_{b,i'}$ are inter-cluster interference. $\sum_{b'=1,b'\neq b}^{B}\mathbf{W}_{BB,b,i'}^H\mathbf{W}_{RF}^H\mathbf{H}_{b,i'}\left(\mathbf{F}_{RF}\circ\mathbf{Q}_{b',i'}\right)\mathbf{F}_{BB,b',i'}x_{b',i'}$ are inter-cell interference. Although the hybrid method is more accurate than the statistical approach, while generating faster and more generalized results than the deterministic approach, nevertheless it does not provide sufficient intra-cluster angular modeling accuracy necessary for beamforming and inter-cluster interference optimizations [41].

D. PROBLEM FORMULATION

With the mmWave Massive MIMO channel $\mathbf{H}_k$ in (13), the dynamic hybrid beamformer design problem can be formulated as

$$\begin{aligned} &\max_{\{S_j\}}\ \mathcal{R}\left(\{S_j\}\right), \\ &s.t.\quad (7), (8), (9) \end{aligned} \tag{18}$$

where $\mathcal{R}\left(\{S_j\}\right)$ is the optimal spectral efficiency with given partition of subsets, i.e., the optimal objective function of the hybrid precoder design problem. In [42], an alternative minimization-based algorithm will be proposed to solve problem (18) and obtain the designed subarray structure. For the subarray structure obtained in (18), we will further design the hybrid precoders in mmWave massive MIMO systems using the instantaneous channel state information. Mathematically, the hybrid precoder design problem can be formulated in (19), as shown at the bottom of the next page, where $\mathbf{H}_{FSO} = \sqrt{P_R}h_{FSO}$. Unfortunately, the hybrid beamformer design problem (19) is extremely difficult to solve due to not only the non-convex constraints of the analog beamformers but also the coupling of the digital/analog beamformers. To efficiently solve the problem (19), in the section IV, we first transform the original problem into an alternative solvable form and then iteratively design the hybrid precoder.

III. USER CLUSTERS

We assume that there are $|\mathcal{G}_i|$ users in the b-th cell. Since the real relationship between users is not known, and the channel matrix dimension at this time is high, which is inconvenient for user clustering. Let $\mathbf{H}_b$ denote the channel matrix between the BS and the $|\mathcal{G}_i|$ user in b-th cell. Let's denote the matrix corresponding to the b-th cell as $\mathbf{H}_b = \left[\mathbf{H}_{b,1}, \ldots, \mathbf{H}_{b,|\mathcal{G}_i|}\right] \in \mathbb{C}^{N_{r,b}\times N_{t,b}\times|\mathcal{G}_i|}$.

Manifold learning can reduce the dimensionality of high-dimensional data [43]. The basic steps of MDS algorithm are illustrated in Fig. 5.

The MDS uses the similarity between pairs of samples to construct a suitable low-dimensional space. Each sample in the high-dimensional space represents an object, thus the sample-to-sample distance and the similarity between objects are highly correlated. It can be understood that two similar objects are represented in higher dimensional space by two points that are close to each other, and two dissimilar objects are represented in higher dimensional space by two points that are farther away. The distance of samples in the lower dimensional space and the similarity between samples in the higher dimensional space are kept as consistent as possible.

In the matrix $\mathbf{H}_b$, the relative distance between each point can be represented by the Euclidean distance matrix $\mathbf{D}_b = \left[d_{b,k,k'}\right] \in \mathbb{R}^{|\mathcal{G}_i|\times|\mathcal{G}_i|}, k, k' = 1, 2, \cdots, |\mathcal{G}_i|$, where $d_{b,k,k'}$ is the Euclidean distance between the k-th user and the k'-th user, i.e.,

$$d_{b,k,k'} \triangleq \left\|h_{b,k} - h_{b,k'}\right\|. \tag{20}$$

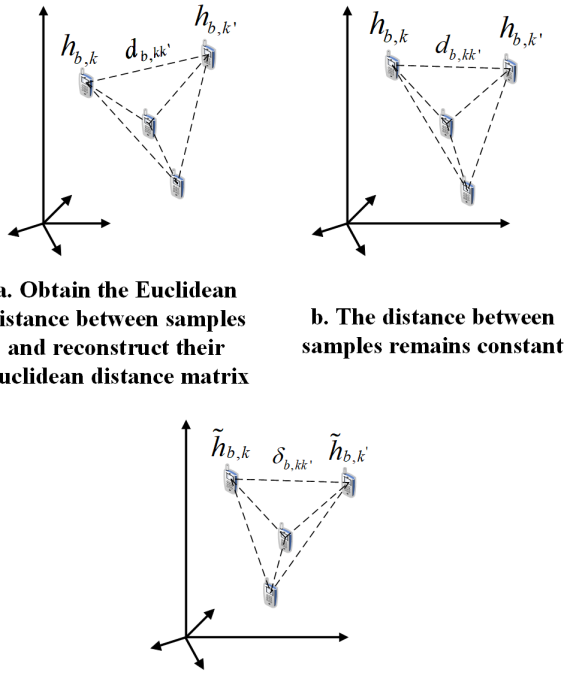


FIGURE 5. The basic steps of MDS algorithm.

We reduce the dimensionality of the data to a low-dimensional space matrix $\widetilde{\mathbf{H}}_b$. The relative position relationship of any two samples in the low-dimensional dimensional space $\widetilde{\mathbf{H}}_b$ is required to be the same as the relative position relationship in the original high-dimensional space. Thus, we have the following expression:

$$\sum_{\min} \left(\left\|\widetilde{h}_{b,k} - \widetilde{h}_{b,k'}\right\| - d_{b,k,k'}\right)^2. \tag{21}$$

Equation (21) can be transformed into

$$\begin{aligned} d^2_{b,k,k'} &= \left\|\widetilde{h}_{b,k} - \widetilde{h}_{b,k'}\right\|^2 \\ &= \left\|\widetilde{h}_{b,k}\right\|^2 + \left\|\widetilde{h}_{b,k'}\right\|^2 - 2\widetilde{h}^T_{b,k}\widetilde{h}_{b,k'}. \end{aligned} \tag{22}$$

Since the points can be translated and rotated in the matrix $\widetilde{\mathbf{H}}_b$ of the low-dimensional space, there are multiple distributions in the low-dimensional space to satisfy the requirements. Without loss of generality, we assume that the instance points in the low-dimensional space are centralized, i.e.

$$\sum_{k=1}^{|\mathcal{G}_i|} \widetilde{h}_{b,k} = 0. \tag{23}$$

Sum the left and right sides of equation (22):

$$\sum_{k=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} = \sum_{k=1}^{|\mathcal{G}_i|} \left\|\widetilde{h}_{b,k}\right\|^2 + |\mathcal{G}_i| \left\|\widetilde{h}_{b,k'}\right\|^2. \tag{24}$$

$$\sum_{k'=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} = \sum_{k'=1}^{|\mathcal{G}_i|} \left\|\widetilde{h}_{b,k'}\right\|^2 + |\mathcal{G}_i| \left\|\widetilde{h}_{b,k}\right\|^2. \tag{25}$$

The summation is performed again for both sides of equation (25):

$$\begin{aligned} \sum_{k=1}^{|\mathcal{G}_i|}\sum_{k'=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} &= \sum_{k=1}^{|\mathcal{G}_i|}\sum_{k'=1}^{|\mathcal{G}_i|} \left\|\widetilde{h}_{b,k'}\right\|^2 + |\mathcal{G}_i| \sum_{k=1}^{|\mathcal{G}_i|} \left\|\widetilde{h}_{b,k}\right\|^2 \\ &= 2\,|\mathcal{G}_i| \sum_{k=1}^{|\mathcal{G}_i|} \left\|\widetilde{h}_{b,k}\right\|^2 \end{aligned} \tag{26}$$

Define the inner product matrix $\mathbf{B} = \mathbf{H}_b^T\mathbf{H}$. Substituting equations (24), (25) and (26) into equation (22), we get:

$$\begin{aligned} \tilde{b}_{b,k,k'} = -\frac{1}{2}\Bigg(&\frac{1}{|\mathcal{G}_i|^2}\sum_{k=1}^{|\mathcal{G}_i|}\sum_{k'=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} - \frac{1}{|\mathcal{G}_i|}\sum_{k=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} \\ &- \frac{1}{|\mathcal{G}_i|}\sum_{k'=1}^{|\mathcal{G}_i|} d^2_{b,k,k'} + d^2_{b,k,k'}\Bigg). \end{aligned} \tag{27}$$

Due to the fact that matrix $\mathbf{B}$ is a symmetric matrix, the eigen-decomposition of matrix $\mathbf{B}$ can be obtained:

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \tag{28}$$

where $\mathbf{\Lambda}$ is the eigenvalue matrix and $\mathbf{V}$ is the eigenvector matrix. Since we reduce the dimensionality of the data into the low-dimensional space, the dimensionality of the original sample space is much larger than that of the low-dimensional space, so we choose the largest eigenvalues as well as eigenvectors for the dimensionality of the matrix $\mathbf{H}_b$ of the low-dimensional space. The data points after dimensionality reduction are represented as

$$\widetilde{\mathbf{H}}_b = \mathbf{V}_* \Lambda_*^{1/2}. \tag{29}$$

where $\mathbf{V}_*$, Λ_* are the largest eigenvalue matrix as well as eigenvector matrix, respectively.

$$\max_{\mathbf{F}_{RF,k},\{\mathbf{F}_{BB,k}\}} \sum_{k=1}^{K} \log\det\left(\mathbf{I} + \frac{\left\|\mathbf{W}^H_{BB,k}\mathbf{W}^H_{RF}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}\mathbf{H}_{FSO}\right\|_F^2 P_k}{\sum\limits_{k'=1,k'\neq k}^{K}\left\|\mathbf{W}^H_{BB,k}\mathbf{W}^H_{RF}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k'}\mathbf{H}_{FSO}\right\|_F^2 P_{k'} + \sigma_k^2}\right),$$
$$s.t. \qquad (\mathbf{10}), (\mathbf{12}) \tag{19}$$

After obtaining the low-dimensional channels $\mathbf{H}_b$, the SMACOF algorithm is carried out to optimize low-dimensional channels $\mathbf{H}_b$. The optimization problem can be made more precise by the channel correlation coefficient by [44]

$$\varphi\left(\mathbf{H}_b\right)=\sum_{k<k'}\omega_{b,k,k'}\left(d_{b,kk'}-\widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right)\right)^2, \tag{30}$$

where $\widetilde{\delta}_{b,k,k'}\triangleq\left\|\widetilde{h}_{b,k}-\widetilde{h}_{b,k'}\right\|$, it is worth noting that we use $\widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right)$ instead of $\widetilde{\delta}_{k,k'}$ in (30) to highlight that $\widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right)$ is a function of $\mathbf{H}_b$. $\omega_{b,k,k'}$ is the channel correlation coefficient between the k-th user and its neighbor k'. In order to calculate the reconstruction weight coefficient $\left\{\omega_{b,k,k'}\right\}_{k\neq k',k'=1}^{|\mathcal{G}_i|}$ between the k-th user and its adjacent user, the linear combination effect is optimal [45]. The stress $\varphi\left(\mathbf{H}_b\right)$ can be decomposed as

$$\begin{aligned}\varphi\left(\mathbf{H}_b\right)=&\sum_{k<k'}\omega_{b,k,k'}d_{b,k,k'}^2+\sum_{k<k'}\omega_{b,k,k'}\widetilde{\delta}_{b,k,k'}^2\left(\mathbf{H}_b\right)\\&-2\sum_{k<k'}\omega_{b,k,k'}d_{b,k,k'}\widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right).\end{aligned} \tag{31}$$

To facilitate the derivation, we denote

$$\sum_{k<k'}\omega_{b,k,k'}d_{b,k,k'}^2=\xi_b^2. \tag{32}$$

$$\sum_{k<k'}\omega_{b,k,k'}\widetilde{\delta}_{b,k,k'}^2\left(\widetilde{\mathbf{H}}_b\right)=\xi^2\left(\widetilde{\mathbf{H}}_b\right). \tag{33}$$

$$\sum_{k<k'}\omega_{b,k,k'}d_{b,k,k'}\widetilde{\delta}_{b,k,k'}\left(\widetilde{\mathbf{H}}_b\right)=\vartheta\left(\widetilde{\mathbf{H}}_b\right). \tag{34}$$

It is obvious that ξ_b is a constant independent of $\mathbf{H}_b$ and $\xi^2\left(\mathbf{H}_b\right)$ can be derived as

$$\begin{aligned}\xi^2\left(\mathbf{H}_b\right)&=\sum_{k<k'}\omega_{b,k,k'}\left(\widetilde{h}_{b,k}-\widetilde{h}_{b,k'}\right)^T\left(\widetilde{h}_{b,k}-\widetilde{h}_{b,k'}\right)\\&=\sum_{k<k'}\omega_{b,k,k'}\left(\mathbf{e}_{b,k}-\mathbf{e}_{b,k'}\right)^T\mathbf{T}\mathbf{T}^T\left(\mathbf{e}_{b,k}-\mathbf{e}_{b,k'}\right)\\&=tr\left(\mathbf{H}_b^T\omega_{b,k,k'}\mathbf{O}_{b,k,k'}\mathbf{H}_b\right),\end{aligned} \tag{35}$$

where $\mathbf{O}_{b,k,k'}=\left(\mathbf{e}_{b,k}-\mathbf{e}_{b,k'}\right)\left(\mathbf{e}_{b,k}-\mathbf{e}_{b,k'}\right)^T$ whose elements equal 1 at $o_{k,k}$, $o_{k',k'}$, -1 at $o_{k,k'}$, $o_{k',k}$, and 0 elsewhere. Furthermore, we define

$$\mathbf{E}=\sum_{k<k'}\omega_{b,k,k'}\mathbf{O}_{b,k,k'}, \tag{36}$$

as the weighted sum of row and column centered matrices $\mathbf{O}_{k,k'}$, thus we can rewrite

$$\xi^2\left(\widetilde{\mathbf{H}}_b\right)=tr\left(\widetilde{\mathbf{H}}_b^T\mathbf{E}\widetilde{\mathbf{H}}_b\right). \tag{37}$$

Similarly, we can also rewrite $\vartheta\left(\mathbf{H}_b\right)$ as

$$\vartheta\left(\mathbf{H}_b\right)=tr\left(\mathbf{H}_b^T f\left(\mathbf{H}_b\right)\mathbf{H}_b\right), \tag{38}$$

with

$$f\left(\mathbf{H}_b\right)=\sum_{k<k'}\omega_{b,k,k'}\tilde{s}_{b,k,k'}\left(\mathbf{H}_b\right)\mathbf{O}_{b,k,k'}, \tag{39}$$

where

$$\tilde{s}_{b,k,k'}\left(\mathbf{H}_b\right)=\begin{cases}d_{b,k,k'}/\widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right), & \widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right)\neq 0,\\0, & \widetilde{\delta}_{b,k,k'}\left(\mathbf{H}_b\right)=0.\end{cases} \tag{40}$$

Then, substitute (32), (37) and (38) into (40), we thus have

$$\varphi\left(\mathbf{H}_b\right)=\xi_\delta^2+tr\left(\mathbf{H}_b^T\mathbf{E}\mathbf{H}_b\right)-2tr\left(\mathbf{H}_b^T f\left(\mathbf{H}_b\right)\mathbf{H}_b\right). \tag{41}$$

It can be seen from (41) that the second part is quadratic in $\mathbf{H}_b$ while the third part is not as $f\left(\mathbf{H}_b\right)$ varies with $\mathbf{H}_b$. The SMACOF algorithm changes (41) into a quadratic function about $\mathbf{H}_b$ by introducing the supporting point denoted by $\mathbf{C}$. Then, according to the Cauchy-Schwartz inequality, we have

$$\begin{aligned}\varphi\left(\mathbf{H}_b\right)&\leq\xi_\delta^2+tr\left(\mathbf{H}_b^T\mathbf{E}\mathbf{H}_b\right)-2tr\left(\widetilde{\mathbf{H}}_b^T f\left(\mathbf{C}\right)\mathbf{C}\right)\\&=\iota\left(\widetilde{\mathbf{H}}_b,\mathbf{C}\right).\end{aligned} \tag{42}$$

The minimum can be find by

$$\frac{\partial\iota\left(\widetilde{\mathbf{H}}_b,\mathbf{C}\right)}{\partial\widetilde{\mathbf{H}}_b^T}=2\mathbf{E}\widetilde{\mathbf{H}}_b^T-2f\left(\mathbf{C}\right)\mathbf{C}=0. \tag{43}$$

After that, we can get

$$\widetilde{\mathbf{H}}_b^T=\mathbf{E}^+f\left(\mathbf{C}\right)\mathbf{C}, \tag{44}$$

where

$$\mathbf{E}^+=\left(\mathbf{E}+\frac{1}{K}11^T\right)^{-1}-\frac{1}{K}11^T, \tag{45}$$

where 1 is the vector of ones. Then iterate the formulas as follows.

$$\begin{cases}\mathbf{C}=\mathbf{H}_b^{(\widetilde{t}-1)},\\\mathbf{H}_b^{(\widetilde{t})}=\mathbf{E}^+f\left(\mathbf{C}\right)\mathbf{C},\end{cases} \tag{46}$$

where $\widetilde{t}=1,2,\cdots,\widetilde{t}_{\max}$ and $\widetilde{t}_{\max}$ is the maximum number of iterations. The iteration stops until $\varphi\left(\mathbf{H}_b^{(\widetilde{t}-1)}\right)-\varphi\left(\mathbf{H}_b^{(\widetilde{t})}\right)<\lambda$, where λ is the threshold to control the accuracy. When the iteration stops, $\mathbf{H}_b^{(\widetilde{t})}$ is returned as low-dimensional channels $\mathbf{H}_b$ of the b-th cell.

According to the similarity of the optimized low-dimensional channels $\mathbf{H}_b$ by (46), this paper draws on the method in the literature [46] to perform clustering by maximizing the degree of cohesion of the cluster.

We consider the total number of users participating in clustering is $|\mathcal{G}_i|$ in low-dimensional channels matrix $\mathbf{H}_b$. Cluster probability vector p is a $|\mathcal{G}_i|-$ *dimensional* vector, where element $p_{b,i,k}$ is probability that k-th user appears in cluster $p_{b,i}$.

Particularly, when $p_{b,i,k}=1,\forall k'\neq k,p_{b,i,k'}=0$, it means that only k-th user is included in the cluster, and no other users belong to the cluster.

Excellent clustering should ensure that users in the same category have good cohesion. Reflected in the channel

matrix, these channels should have a greater degree of similarity in channel characteristics. (47) can express the degree of cohesion of the cluster corresponding to the vector g [46].

$$z(p) = p^T \cdot \mathbf{H}_b \cdot p. \tag{47}$$

The larger the $z(p)$, the closer the relationship between channels, and the more likely it is to become a cluster. In this way, the problem of user clustering is transformed into finding a suitable vector p, so that the cohesion of the cluster $z(p)$ reaches the maximum, as in (48) shown.

$$\max_p z(p) = p^T \cdot \mathbf{H}_b \cdot p. \tag{48}$$

In order to solve equation (48), the replicator dynamics method in evolutionary game theory is adopted, that is, iterative (49) is used to iteratively solve (48).

$$p_{b,i,k}\left(\tilde{t}+1\right) = p_{b,i,k}\left(\tilde{t}\right) \cdot \frac{\left(\mathbf{H}_b \cdot p\left(\tilde{t}\right)\right)_{b,i,k}}{p\left(\tilde{t}\right)^T \cdot \mathbf{H}_b \cdot p\left(\tilde{t}\right)}, \quad k = 1, \cdots, |\mathcal{G}_i|, \tag{49}$$

where $\tilde{t}$ is the number of iterations, $p_{b,i,k}\left(\tilde{t}\right)$ is the probability that k-th user belongs to the cluster $p_{b,i}$ in the $\tilde{t}$ -th iteration, $\mathbf{H}_b \cdot p\left(\tilde{t}\right)$ denotes the similarity between k-th user and other users in the $\tilde{t}$ -th iteration, $p\left(\tilde{t}\right)^T \cdot \mathbf{H}_b \cdot p\left(\tilde{t}\right)$ denotes the degree of similarity between all users. (48) increases continuously with the iterative process and converges to a stable value [46], p corresponding to this value is the desired cluster probability vector. At this time, the non-zero elements in the vector p are the users belonging to this cluster. Remove users who have completed clustering, establish a new clustering probability vector and channel matrix, and repeat the above process until all users are clustered.

The specific clustering process algorithm as follows:

Algorithm 1 User Clustering Algorithm

1. **Input**: low-dimensional channels matrix $\mathbf{H}_b$, cluster probability vector p;
2. $i = 1$;
3. **while** $|\mathcal{G}_i| \geq 1$ **do**
4. Initialize vector p;
5. The new vector p is obtained by iterating equation (48) with $\mathbf{H}_b$ and p until convergence
6. Users corresponding to non-zero elements in vector g are put into cluster $\mathcal{G}_i$;
7. Delete the information of the clustered channels in vector p and matrix $\mathbf{H}_b$, and obtain a new vector p and matrix $\mathbf{H}_b$;
8. $i = i + 1$;
9. **return** $\mathcal{G}_i$;
10. **Output**: Cluster set $\mathcal{G}$;

$\mathbf{H}_{b,\mathcal{G}}$ is the $\mathcal{G}_i$ in the b-th cell low-dimensional mapping of the high-dimensional channels $\mathbf{H}_{b,\mathcal{G}_i}$.

IV. MANIFOLD DISCRIMINANTIVE LEARNING FOR HYBRID PRECODING

On the basis of manifold discriminative learning for global dimensionality reduction and user clustering, we investigate the sum-rate maximization problem for hybrid precoding.

The first step: Our objective is to design the precoding matrices $\mathbf{W}_{BB,\mathcal{G}_i}^H, \mathbf{W}_{RF}^H, \mathbf{F}_{BB,\mathcal{G}_i}, \mathbf{F}_{RF,\mathcal{G}_i}, \mathbf{Q}_{\mathcal{G}_i}$ such that they manage intra-cluster interference and inter-cluster interference. In order to improve the spectral efficiency of the systems, the design of each cluster analog precoding should strike a balance between optimizing self-transmission and the interference. By modeling each user set as a manifold, the received signal of the i-th cluster can be represented as

$$\begin{aligned}\tilde{y}_{\mathcal{G}_i} = {} & \mathbf{W}_{BB,\mathcal{G}_i}^H \mathbf{W}_{RF}^H \tilde{H}_{\mathcal{G}_i}^H \left(\mathbf{F}_{RF} \circ \mathbf{Q}_{\mathcal{G}_i}\right) \mathbf{F}_{BB,\mathcal{G}_i} x_{\mathcal{G}_i} \\ & + \sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|} \mathbf{W}_{BB,\mathcal{G}_i,k'}^H \mathbf{W}_{RF}^H \tilde{H}_{\mathcal{G}_i,k}^H \left(\mathbf{F}_{RF} \circ \mathbf{Q}_{\mathcal{G}_i,k'}\right) \\ & \times \mathbf{F}_{BB,\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'} \\ & + \sum_{i'=1,i'\neq i}^{G} \mathbf{W}_{BB,\mathcal{G}_{i'}}^H \mathbf{W}_{RF}^H \tilde{H}_{\mathcal{G}_{i'}}^H \left(\mathbf{F}_{RF} \circ \mathbf{Q}_{\mathcal{G}_{i'}}\right) \\ & \times \mathbf{F}_{BB,\mathcal{G}_{i'}} x_{\mathcal{G}_{i'}} + n_{\mathcal{G}_i},\end{aligned} \tag{50}$$

where $\tilde{y}_{\mathcal{G}_i} = \left[\tilde{y}_{\mathcal{G}_i,1}^T, \cdots, \tilde{y}_{\mathcal{G}_i,g_i}^T\right]^T$ represents the received signal, $\tilde{H}_{\mathcal{G}_i} = \left[\tilde{H}_{\mathcal{G}_i,1}, \cdots, \tilde{H}_{\mathcal{G}_i,g_i}\right]$ represents the channel matrix for the ***i*th** cluster, $\mathbf{F}_{BB,\mathcal{G}_i} = \left[\mathbf{F}_{BB,\mathcal{G}_i,1}, \cdots, \mathbf{F}_{BB,\mathcal{G}_i,g_i}\right]$, $\mathbf{F}_{RF,\mathcal{G}_i} = \left[\mathbf{F}_{RF,\mathcal{G}_i,1}, \cdots, \mathbf{F}_{RF,\mathcal{G}_i,g_i}\right]$ $\mathbf{W}_{RF,\mathcal{G}_i} = \left[\mathbf{W}_{RF,\mathcal{G}_i,1}, \ldots, \mathbf{W}_{RF,\mathcal{G}_i,g_i}\right]$, and $\mathbf{W}_{BB,\mathcal{G}_i} = diag\left(\mathbf{W}_{BB,\mathcal{G}_i,1}, \ldots, \mathbf{W}_{BB,\mathcal{G}_i,g_i}\right)$. $\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|} \mathbf{W}_{BB,\mathcal{G}_i,k'}^H \mathbf{W}_{RF}^H \tilde{H}_{\mathcal{G}_i,k}^H \left(\mathbf{F}_{RF} \circ \mathbf{Q}_{\mathcal{G}_i,k'}\right) \mathbf{F}_{BB,\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'}$ are the intra-cluster interference, $\sum_{i'=1,i'\neq i}^{G} \mathbf{W}_{BB,\mathcal{G}_{i'}}^H \mathbf{W}_{RF}^H \tilde{H}_{\mathcal{G}_{i'}}^H \left(\mathbf{F}_{RF} \circ \mathbf{Q}_{\mathcal{G}_{i'}}\right) \mathbf{F}_{BB,\mathcal{G}_{i'}} x_{\mathcal{G}_{i'}}$ are the inter-cluster interference after the low-dimensional mapping. In order to adapt to special scenarios and requirements, the hybrid precoding matrix can be determined by per-cluster processing (PCP). The goal of PCP is to balance the performance and complexity by effectively separating the clusters in the RF beam domain

The signal estimate $\hat{x}_{\mathcal{G}_i}$ for $\mathcal{G}_i$ can be expressed in (51), as shown at the bottom of the next page, where $\mathbf{F}_{\mathcal{G}_i} = \left(\mathbf{F}_{RF,\mathcal{G}_i} \circ \mathbf{Q}_{\mathcal{G}_i}\right) \mathbf{F}_{BB,\mathcal{G}_i}$, $\mathbf{W}_{\mathcal{G}_i}^H = \mathbf{W}_{BB,\mathcal{G}_i}^H \mathbf{W}_{RF,\mathcal{G}_i}^H$, $\zeta_{\mathcal{G}_i}$ is a scaling equalization that is jointly optimized with the hybrid precoding. The conditional mean square error (MSE) for $\mathcal{G}_i$ is defined as

$$\begin{aligned}& \varepsilon\left(\tilde{\mathbf{F}}_{\mathcal{G}_i}, \tilde{\mathbf{W}}_{\mathcal{G}_i}, \zeta_{\mathcal{G}_i}\right) \\ & = E\left[\left\|x_{\mathcal{G}_i} - \hat{x}_{\mathcal{G}_i}\right\|^2\right] \\ & = E\left[\left\|x_{\mathcal{G}_i} - \zeta_{\mathcal{G}_i}^{-1}\left(\mathbf{W}_{\mathcal{G}_i}^H \tilde{\mathbf{H}}_{\mathcal{G}_i}^H \mathbf{F}_{\mathcal{G}_i} x_{\mathcal{G}_i}\right)\right\|^2\right]\end{aligned}$$

$$+E\left[\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\left\|\zeta_{\mathcal{G}_i}^{-1}\mathbf{W}_{\mathcal{G}_i,k'}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_i,k}^{H}\mathbf{F}_{\mathcal{G}_{i,k'}}x_{\mathcal{G}_i,k'}\right\|\right]$$
$$+E\left[\sum_{i'=1,i'\neq i,}^{G}\left\|\zeta_{\mathcal{G}_i}^{-1}\mathbf{W}_{\mathcal{G}_{i'}}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_{i'}}^{H}\mathbf{F}_{\mathcal{G}_{i'}}x_{\mathcal{G}_{i'}}\right\|^2+\zeta_{\mathcal{G}_i}^{-2}n_{\mathcal{G}_i}\right]. \tag{52}$$

The conditional MSE in (52) is simplified as

$$\varepsilon\left(\tilde{\mathbf{F}}_{\mathcal{G}_i},\tilde{\mathbf{W}}_{\mathcal{G}_i},\zeta_{\mathcal{G}_i}\right)=\varepsilon_{\mathcal{G}_i}^{(1)}+\varepsilon_{\mathcal{G}_i}^{(2)}, \tag{53}$$

where

$$\varepsilon_{\mathcal{G}_i}^{(1)}=E\left[\left\|x_{\mathcal{G}_i}-\zeta_{\mathcal{G}_i}^{-1}\left(\mathbf{W}_{\mathcal{G}_i}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_i}^{H}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right\|^2\right]. \tag{54}$$

$$\varepsilon_{\mathcal{G}_i}^{(2)}=E\left[\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\left\|\zeta_{\mathcal{G}_i}^{-1}\mathbf{W}_{\mathcal{G}_i,k'}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_i,k}^{H}\mathbf{F}_{\mathcal{G}_i,k'}x_{\mathcal{G}_i,k'}\right\|\right]$$
$$+E\left[\sum_{i'=1,i'\neq i,}^{G}\left\|\zeta_{\mathcal{G}_i}^{-1}\mathbf{W}_{\mathcal{G}_{i'}}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_{i'}}^{H}\mathbf{F}_{\mathcal{G}_{i'}}x_{\mathcal{G}_{i'}}\right\|^2+\zeta_{\mathcal{G}_i}^{-2}n_{\mathcal{G}_i}\right]. \tag{55}$$

The optimal scaling factor $\zeta_{\mathcal{G}_i}$ can be obtained from the base station transmission power with $tr\left(\mathbf{F}_{\mathcal{G}_i}\mathbf{W}_{\mathcal{G}_i}\mathbf{W}_{\mathcal{G}_i}^{H}\mathbf{F}_{\mathcal{G}_i}^{H}\right)\leq P_R$ as

$$\zeta_{\mathcal{G}_i}=\sqrt{\frac{P_R}{\sum_{i=1}^{G}tr\left(\mathbf{F}_{\mathcal{G}_i}\mathbf{W}_{\mathcal{G}_i}\mathbf{W}_{\mathcal{G}_i}^{H}\mathbf{F}_{\mathcal{G}_i}^{H}\right)}}. \tag{56}$$

Accordingly, equation (54) can be expressed in (57), as shown at the bottom of the next page.

After simple mathematical derivation, equation (55) can be expressed in (58), as shown at the bottom of the next page.

The second step: Our objective is to design the precoding matrices manage intra-cell interference and inter-cell interference. In order to adapt to different scenarios and requirements, the hybrid precoding matrix can be determined by joint processing (JCP).

The received signal for the *b***th** cell is given as

$$\tilde{y}_b=\mathbf{W}_b^{H}\tilde{\mathbf{H}}_b^{H}\mathbf{F}_bx_b+\sum_{b'=1,b'\neq b}^{B}\mathbf{W}_{b'}^{H}\tilde{\mathbf{H}}_{b'}^{H}\mathbf{F}_{b'}x_{b'}+n_b, \tag{59}$$

where $\tilde{y}_b=\left[\tilde{y}_{b,\mathcal{G}_1}^{T},\cdots,\tilde{y}_{b,\mathcal{G}_i}^{T}\right]^{T}$ represents the received signal, $\tilde{\mathbf{H}}_b=\left[\tilde{\mathbf{H}}_{b,\mathcal{G}_1},\cdots,\tilde{\mathbf{H}}_{b,\mathcal{G}_i}\right]$ represents the channel matrix for the b-th cell, $\mathbf{F}_b=\left[\mathbf{F}_{b,\mathcal{G}_1},\cdots,\mathbf{F}_{b,\mathcal{G}_i}\right]$ and $\mathbf{W}_b=diag\left(\mathbf{W}_{b,\mathcal{G}_1},\ldots,\mathbf{W}_{b,\mathcal{G}_i}\right)$ represent the precoding matrix and the combiner matrix respectively. n_b represents the spatially additive white Gaussian noise for the b-th cell. Thus, the estimation of the received signal in the b-th cell is given as

$$\hat{x}_b=\zeta_b^{-1}\left(\mathbf{W}_b^{H}\tilde{\mathbf{H}}_b^{H}\mathbf{F}_bx_b+\sum_{b'=1,b'\neq b}^{B}\mathbf{W}_{b'}^{H}\tilde{\mathbf{H}}_{b'}^{H}\mathbf{F}_{b'}x_{b'}+n_b\right), \tag{60}$$

where ζ_b^{-1} is the b-th cell scaling factor that is jointly optimized with the hybrid precoding. The design of precoding is jointly derived across all G user clusters is derived in (61), as shown at the bottom of the next page.

In order to eliminate intra-cluster interference, inter-cluster interference and inter-cell interference, the precoding is conducted as a multiplication of two precoding, i.e., $\tilde{\mathbf{F}}_{BB,b}=\tilde{\mathbf{F}}_{BB,b}^{(1)}\tilde{\mathbf{F}}_{BB,b}^{(2)}$. $\tilde{\mathbf{F}}_{BB,b}^{(1)}$ and $\tilde{\mathbf{F}}_{BB,b}^{(2)}$represent the first and the second precoding matrix of the b-th cell respectively, where $\tilde{\mathbf{F}}_{RF,b}$ represents the digital precoding matrix of the b-th cell. The signal space of $\tilde{\mathbf{F}}_{BB,b}^{(1)}=\left[\tilde{\mathbf{F}}_{BB,b,\mathcal{G}_1}^{(1)},\cdots,\tilde{\mathbf{F}}_{BB,b,\mathcal{G}_{i'}}^{(1)},\cdots,\tilde{\mathbf{F}}_{BB,b,\mathcal{G}_i}^{(1)}\right]$ is mapped to the channel null space of all remaining user groups $\mathcal{G}_{i'}$, namely:

$$\tilde{\mathbf{F}}_{BB,b,\mathcal{G}_i'}^{(1)}\subset Span^{\perp}\{\tilde{U}_{b,\mathcal{G}_i''}^{*}(\mathcal{G}_i',\mathcal{G}_i''\in b,\mathcal{G}_i'\neq\mathcal{G}_i'')\}, \tag{62}$$

where $\tilde{\mathbf{U}}_{b,\mathcal{G}_i''}^{*}$ is a matrix comprising dominant eigenvectors corresponding to the $r_{b,\mathcal{G}_i''}^{*}<r_{b,\mathcal{G}_i''}$ dominant eigenvalues of $\mathbf{R}_{b,\mathcal{G}_i''}$. $\mathbf{R}_{b,\mathcal{G}_i''}=E[\tilde{H}_{b,\mathcal{G}_i''}\tilde{H}_{b,\mathcal{G}_i''}^{H}]$ is the channel covariance matrix of the b-**th** cell. The idea of formula (62) is to design the pre-beamforming matrix to concentrating the inter-cell transmission energy in the specific direction. The inter-cell interference is reduced by leaving slots in the spatial domain.

In order to realize (62) based on the approach of block diagonalization [47], we define a matrix of eigenmodes of equivalent interference channel covariance for the b-th cell as follows:

$$\Xi_{b,\mathcal{G}_i'}=[\tilde{\mathbf{U}}_{b,1}^{*},\tilde{\mathbf{U}}_{b,2}^{*},\ldots,\tilde{\mathbf{U}}_{b,\mathcal{G}_i'-1}^{*},\tilde{\mathbf{U}}_{b,\mathcal{G}_i'+1}^{*},\ldots,\tilde{\mathbf{U}}_{b,\mathcal{G}_i}^{*}], \tag{63}$$

where $\Xi_{b,\mathcal{G}_i'}$ is rank $r_{b,\mathcal{G}_i''}^{*}\times(\mathcal{G}_i-1)$, $r_{b,\mathcal{G}_i''}^{*}$ is the dominant eigenvalues of $\mathbf{R}_{b,\mathcal{G}_i''}$. For the singular value decomposition (SVD) of $\Xi_{b,\mathcal{G}_i'}$, let $\Phi_{b,\mathcal{G}_i'}^{(0)}$ denote the left eigenvectors corresponding to the zero singular values. And $\Phi_{b,\mathcal{G}_i'}^{(0)}$ can be approximated as the orthogonal basis of null space of the channel vectors for other user cells. $\Phi_{b,\mathcal{G}_i'}^{(0)}=null(\Xi_{b,\mathcal{G}_i'})$. Based on the Karhunen-Loeve decomposition, the equivalent channel covariance matrix $\tilde{\mathbf{R}}_{b,\mathcal{G}_i'}$ is given by:

$$\tilde{\mathbf{R}}_{b,\mathcal{G}_i'}=(\Phi_{b,\mathcal{G}_i'}^{(0)})^{H}\tilde{\mathbf{U}}_{b,\mathcal{G}_i'}\Lambda_{b,\mathcal{G}_i'}\tilde{\mathbf{U}}_{b,\mathcal{G}_i'}^{H}\Phi_{b,\mathcal{G}_i'}^{(0)}. \tag{64}$$

$$\hat{x}_{\mathcal{G}_i}=\zeta_{\mathcal{G}_i}^{-1}\left(\mathbf{W}_{\mathcal{G}_i}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_i}^{H}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}+\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\mathbf{W}_{\mathcal{G}_i,k'}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_i,k}^{H}\mathbf{F}_{\mathcal{G}_i,k'}x_{\mathcal{G}_i,k'}+\times\sum_{i'=1,i'\neq i}^{G}\mathbf{W}_{\mathcal{G}_{i'}}^{H}\tilde{\mathbf{H}}_{\mathcal{G}_{i'}}^{H}\mathbf{F}_{\mathcal{G}_{i'}}x_{\mathcal{G}_{i'}}+n_{\mathcal{G}_i}\right), \tag{51}$$

Then, the SVD of (64) is carried out. Let $\tilde{\mathbf{U}}_{b,\mathcal{G}_i'}$, contains the dominant $r^*_{b,\mathcal{G}_i'}$ eigenmodes of $\mathbf{R}_{b,\mathcal{G}_i'}$. The first precoding matrix $\tilde{\mathbf{F}}^{(1)}_{BB,b,\mathcal{G}_i'}$ is given by:

$$\tilde{\mathbf{F}}^{(1)}_{BB,b,\mathcal{G}_i'} = \Phi^{(0)}_{b,\mathcal{G}_i'}\tilde{\mathbf{U}}_{b,\mathcal{G}_i'}, \tag{65}$$

$\tilde{\mathbf{F}}^{(2)}_{BB,b}$ can be obtained as

$$\tilde{\mathbf{F}}^{(2)}_{BB,b} = \tilde{\mathbf{H}}^H_{eq}\left(\tilde{\mathbf{H}}_{eq}\tilde{\mathbf{H}}^H_{eq} + \Upsilon_b^{-1}\mathbf{I}_b\right)^{-1}, \tag{66}$$

where Υ_b^{-1} is regularization factor, which depends on noise variance and base station transmit power. The equivalent channel matrix $\tilde{\mathbf{H}}_{eq}$ after analog precoding denote as

$$\tilde{\mathbf{H}}_{eq} = \tilde{\mathbf{W}}^H_{BB,b}\tilde{\mathbf{W}}^H_{RF}\tilde{H}^H_b\left(\tilde{\mathbf{F}}_{RF,b}\circ\tilde{\mathbf{Q}}_b\right), \tag{67}$$

where $\mathbf{I}_b$ can be expressed as

$$\mathbf{I}_b = \sum_{b'=1,b'\neq b}^{B}\mathbf{W}^H_{BB,b'}\mathbf{W}^H_{RF}\tilde{\mathbf{H}}^H_{b'}\left(\mathbf{F}_{RF,b'}\circ\mathbf{Q}_{b'}\right)\mathbf{F}_{BB,b'}x_{b'} + n_b. \tag{68}$$

$$\begin{aligned}
\varepsilon^{(1)}_{\mathcal{G}_i} &= E\left[\left\|x_{\mathcal{G}_i} - \zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right\|^2\right] \\
&= E\left\{tr\left[\left(x_{\mathcal{G}_i} - \zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right)^H\left(x_{\mathcal{G}_i} - \zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right)\right]\right\} \\
&= E\left\{tr\left[\left(x^H_{\mathcal{G}_i} - \zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)^H\right)\left(x_{\mathcal{G}_i} - \zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right)\right]\right\} \\
&= E\left\{tr\left(x^H_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right\} - E\left\{tr\left[\zeta^{-1}_{\mathcal{G}_i}x^H_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right]\right\} \\
&\quad - E\left\{tr\left[\zeta^{-1}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)^H x_{\mathcal{G}_i}\right]\right\} \\
&\quad + E\left\{tr\left[\zeta^{-2}_{\mathcal{G}_i}\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)^H\left(\mathbf{W}^H_{\mathcal{G}_i}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_i}x_{\mathcal{G}_i}\right)\right]\right\}.
\end{aligned} \tag{57}$$

$$\begin{aligned}
\varepsilon^{(2)}_{\mathcal{G}_i} &= E\left[\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\left\|\zeta^{-1}_{\mathcal{G}_i}\mathbf{W}^H_{\mathcal{G}_{i,k'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_{i,k}}\mathbf{F}_{\mathcal{G}_{i,k'}}x_{\mathcal{G}_{i,k'}}\right\|\right] \\
&\quad + \sum_{i'=1,i'\neq i}^{G}E\left[\left\|\zeta^{-1}_{\mathcal{G}_i}\mathbf{W}^H_{\mathcal{G}_{i'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_i}\mathbf{F}_{\mathcal{G}_{i'}}x_{\mathcal{G}_{i'}}\right\|^2_F + n_{\mathcal{G}_i}\right] \\
&= \sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\zeta^{-2}_{\mathcal{G}_i}tr\left[\mathbf{W}^H_{\mathcal{G}_{i,k'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_{i,k}}\mathbf{F}_{\mathcal{G}_{i,k'}}E\left(x_{\mathcal{G}_{i,k'}}x^H_{\mathcal{G}_{i,k'}}\right)\mathbf{F}^H_{\mathcal{G}_{i,k'}}\tilde{\mathbf{H}}_{\mathcal{G}_{i,k}}\mathbf{W}_{\mathcal{G}_{i,k'}}\right] \\
&\quad + \sum_{i'=1,i'\neq i}^{G}\zeta^{-2}_{\mathcal{G}_i}tr\left[\mathbf{W}^H_{\mathcal{G}_{i'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_{i'}}\mathbf{F}_{\mathcal{G}_{i'}}E\left(x_{\mathcal{G}_{i'}}x^H_{\mathcal{G}_{i'}}\right)\mathbf{F}^H_{\mathcal{G}_{i'}}\tilde{\mathbf{H}}_{\mathcal{G}_{i'}}\mathbf{W}_{\mathcal{G}_{i'}}\right] + \zeta^{-2}_{\mathcal{G}_i}g_{\mathcal{G}_i}\sigma^2_{\mathcal{G}_i} \\
&= \sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|}\zeta^{-2}_{\mathcal{G}_i}g_i tr\left(\mathbf{W}^H_{\mathcal{G}_{i,k'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_{i,k}}\mathbf{F}_{\mathcal{G}_{i,k'}}\mathbf{F}^H_{\mathcal{G}_{i,k'}}\tilde{\mathbf{H}}_{\mathcal{G}_{i,k}}\tilde{\mathbf{W}}_{\mathcal{G}_{i,k'}}\right) \\
&\quad + \sum_{i'=1,i'\neq i}^{G}\zeta^{-2}_{\mathcal{G}_i}g_{i'}tr\left(\mathbf{W}^H_{\mathcal{G}_{i'}}\tilde{\mathbf{H}}^H_{\mathcal{G}_{i'}}\mathbf{F}_{\mathcal{G}_{i'}}\mathbf{F}^H_{\mathcal{G}_{i'}}\tilde{\mathbf{H}}_{\mathcal{G}_{i'}}\mathbf{W}_{\mathcal{G}_{i'}}\right) + \zeta^{-2}_{\mathcal{G}_i}g_{\mathcal{G}_i}\sigma^2_{\mathcal{G}_i}.
\end{aligned} \tag{58}$$

$$\begin{aligned}
\mathcal{E}_b\left(\tilde{\mathbf{F}}_b, \tilde{\mathbf{W}}_b, \beta_b\right) &\triangleq E\left[\left\|x_b - \hat{x}_b\right\|^2\right] \\
&= E\left[\left\|x_b - \zeta_b^{-1}\left(\mathbf{W}^H_b\tilde{\mathbf{H}}^H_b\mathbf{F}_b x_b + \sum_{b'=1,b'\neq b}^{B}\mathbf{W}^H_{b'}\tilde{\mathbf{H}}^H_{b'}\mathbf{F}_{b'}x_{b'} + n_b\right)\right\|^2\right].
\end{aligned} \tag{61}$$

Therefore, the optimization problem under the multi-cell scenario can be transformed into

$$\begin{aligned} &\underset{\tilde{\mathbf{F}}_{RF,b}}{\operatorname{argmin}}\ J\left(\tilde{F}_{RF,b}\right) \triangleq \sum_{i=1}^{G} \varepsilon_{\mathcal{G}_i}^{(1)} + \sum_{i=1}^{G} \varepsilon_{\mathcal{G}_i}^{(2)} + \sum_{b\in B} \mathcal{E}_b \\ &s.t. \qquad \tilde{\mathbf{F}}_{RF,b}^{\mathbf{H}} \tilde{\mathbf{F}}_{RF,b} = \mathbf{I}_b, \end{aligned} \tag{69}$$

where $\tilde{\mathbf{F}}_{RF,b} = \left[\tilde{\mathbf{F}}_{RF,\mathcal{G}_1}, \cdots, \tilde{\mathbf{F}}_{RF,\mathcal{G}_i}\right]$. The analog precoding matrix $\tilde{F}_{RF,b}$ is design to avoid the intra-cluster interference and inter-cluster interference based on $J\left(\tilde{\mathbf{F}}_{RF,b}\right)$ optimization problem in single cell scenario. The Euclidean conjugate gradient of $J\left(\tilde{\mathbf{F}}_{RF,b}\right)$ can be expressed as

$$\Delta J\left(\tilde{\mathbf{F}}_{RF,b}\right) = \frac{\partial J\left(\tilde{\mathbf{F}}_{RF,b}\right)}{\partial \tilde{\mathbf{F}}_{RF,b}^{*}}. \tag{70}$$

In the next step, the direction vector is updated by using gradient as

$$\mathbf{Z}_{b,t+1} = -\Delta J\left(\tilde{\mathbf{F}}_{RF,b,t+1}\right) + \Gamma_{b,t}\mathbf{Z}_{b,t}, \tag{71}$$

where

$$\Gamma_{b,t} = \frac{\left\|\Delta J\left(\tilde{\mathbf{F}}_{RF,b,t+1}\right)\right\|_F^2}{\left\|\Delta J\left(\tilde{\mathbf{F}}_{RF,b,t}\right)\right\|_F^2}. \tag{72}$$

The manifold quasi-conjugate gradient algorithm based on implicit vector transmission applied is as follows:

Update the analog precoding matrix until convergence to satisfy the error threshold condition, the algorithm ends.

For the intra-cluster, it has been proved that the channel correlation between the intra-cluster users. And its nearby inter-cluster users are much larger than that of the non-adjacent clusters. The interference intensity is the same. Therefore, the interference caused by remote user clusters to intra-cluster users is negligible. Therefore, the SINR for a user cluster $\mathcal{G}_i$ in the b-th cell is given by:

$$SINR_{\mathcal{G}_i} = \frac{\left|\tilde{\mathbf{W}}_{\mathcal{G}_i}^H \tilde{\mathbf{H}}_{\mathcal{G}_i}^H \tilde{\mathbf{F}}_{\mathcal{G}_i}\right|^2 P_{\mathcal{G}_i}}{\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|} \left|\mathbf{IN}_{\mathcal{G}_i,k'}\right|^2 P_{\mathcal{G}_i,k'} + \sum_{i'=1,i'\neq i}^{G} \left|\mathbf{IN}_{\mathcal{G}_{i'}}\right|^2 P_{\mathcal{G}_{i'}} + \sigma_{\mathcal{G}_i}^2}, \tag{73}$$

where $\mathbf{IN}_{\mathcal{G}_i,k'} = \tilde{\mathbf{W}}_{\mathcal{G}_i,k}^H \tilde{H}_{\mathcal{G}_i,k}^H \tilde{\mathbf{F}}_{\mathcal{G}_{i,k'}}$, $\mathbf{IN}_{\mathcal{G}_{i'}} = \tilde{\mathbf{W}}_{\mathcal{G}_{i'}}^H \tilde{H}_{\mathcal{G}_{i'}}^H \tilde{\mathbf{F}}_{\mathcal{G}_{i'}}$, $P_{\mathcal{G}_i}$ are the transmit power of the $\mathcal{G}_i$-th cluster, $P_{\mathcal{G}_i,k'}$ and $P_{\mathcal{G}_{i'}}$ are the transmit power of the k'-th user in the i'-th cluster and the transmit power of the $\mathcal{G}_{i'}$-th cluster, respectively.

Then, we investigate how to design combiner matrix $\mathbf{W}$. The design of the precoding process and the combining process is usually decomposed as two similar problems and solved separately [48]. Since we focus on the precoding process, we assume that the combining matrices $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ have been determined and are known, which is a common practice in the hybrid precoding studies [17], [48].

Algorithm 2 The Manifold Quasi-Conjugate Gradient Algorithm Based on Implicit Vector Transmission

1. **Input**: the analog precoding matrix $\tilde{\mathbf{F}}_{RF,b,t}$, error threshold ε, the initial gradient $\mathbf{Z}_{b,1}$, the regularization factor matrix $\mathbf{\Upsilon}_{b,t}$, the number of initialization iterations $\tilde{t}$;
2. **Initialize** the analog precoding matrix $\tilde{\mathbf{F}}_{RF,b,1} = \left[\tilde{\mathbf{F}}_{RF,\mathcal{G}_1,1}, \cdots, \tilde{\mathbf{F}}_{RF,\mathcal{G}_i,1}\right]$, error threshold $\varepsilon \in (0,1)$, the initial gradient $\mathbf{Z}_{b,1} = \left[Z_{\mathcal{G}_1,1}, \cdots, Z_{\mathcal{G}_i,1}\right]$, where $Z_{\mathcal{G}_i,1} = -\Delta J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,1}\right)$, the number of initialization iterations $t = 1$;
3. **If** $\left\|\Delta J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}}\right)\right\| \leq \varepsilon, i = 1, \cdots, G$, stop;
4. **else**
   search $\mathbf{\Upsilon}_{b,\tilde{t}} = \left[\mathbf{\Upsilon}_{\mathcal{G}_1,\tilde{t}}, \cdots, \mathbf{\Upsilon}_{\mathcal{G}_i,\tilde{t}}\right]$ satisfying $J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}} + \Upsilon_{\mathcal{G}_i,\tilde{t}} Z_{\mathcal{G}_i,\tilde{t}}\right) = \min_{\mathbf{\Upsilon}_{\mathcal{G}_i,\tilde{t}}\geq 0} J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}} + \mathbf{\Upsilon}_{\mathcal{G}_i,\tilde{t}} Z_{\mathcal{G}_i,\tilde{t}}\right), i = 1, \cdots, G$;
5. Update the analog precoding matrix $\tilde{\mathbf{F}}_{RF,b,\tilde{t}+1}$ using $\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}+1} = \tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}} + \Upsilon_{\mathcal{G}_i,\tilde{t}} Z_{\mathcal{G}_i,\tilde{t}}, i = 1, \cdots, G$;
6. **If** $\tilde{t} < N_{tRF}$,
   perform step 8;
7. **else**
   repeat step 9;
8. Update $Z_{\mathcal{G}_i,\tilde{t}+1} = -\Delta J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}+1}\right) + \Gamma_{\mathcal{G}_i,\tilde{t}} Z_{\mathcal{G}_i,\tilde{t}}, i = 1, \cdots, G$, where $\Gamma_{\mathcal{G}_i} = \frac{\left\|\Delta J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}+1}\right)\right\|_F^2}{\left\|\Delta J\left(\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}}\right)\right\|_F^2}$; Update the number of iterations $\tilde{t} = \tilde{t} + 1$, repeat step 2;
9. Update $\tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}} = \tilde{\mathbf{F}}_{RF,\mathcal{G}_i,\tilde{t}+1}, Z_{\mathcal{G}_i,\tilde{t}} = -\Delta J\left(\tilde{F}_{RF,\mathcal{G}_i,\tilde{t}}\right), \tilde{t} = 1, i = 1, \cdots, G$, repeat step 2.
10. **end if**
11. **Until** convergence, **end if**
12. **Output**: the analog precoding matrix $\tilde{\mathbf{F}}_{RF,b,\tilde{t}}$;

This scheme selects N antenna elements with the maximum amplitude for each user based on the channel of the multi-user MIMO system.

Because each antenna element is divided to subarrays based on the maximum SINR increment of all served users, fairness among users is ensured.

Hybrid precoding in a dynamic subarray architecture enables a compromise between performance and hardware complexity in multi-user millimeter-wave large MIMO systems.

In this paper, the objective of multi-user selection is to select a set of MU-MIMO users with minimum inter-user interference and maximum target channel gain

from K alternative users so that each RF link transmits data for one user and the system performance is optimal. This is because inter-user interference can severely degrade the system performance when the base station transmits signals to multiple users in the same time slot.

To maximize the sum rate of MU-MIMO users, the algorithm divides the dynamic subarrays according to the maximum SINR increment of the selected MU-MIMO users.

The simulated beamforming is used to weight the transmitted signal of the antenna array, which is then filtered in the null domain. By phase-shifting network processing without amplitude modulation, the antenna array can emit a narrow beam with concentrated energy in the desired direction.

$$\tilde{\mathbf{Q}}_{\mathcal{G}_i} = \underset{\mathbf{Q}_{\mathcal{G}_i}}{\operatorname{argmax}} \left(SINR_{\mathcal{G}_i} \right). \tag{74}$$

However, using the maximum SINR criterion can lead to unfairness among users, since the SINR value will be higher for users who are divided into more antennas. On the other hand, the first antenna can generate more SINR increments than the others [10]. Considering the user fairness and the objective function of the maximum sum rate criterion, the dynamic subarray assigns each antenna unit to the RF chain according to the maximum SINR increment. The SINR increment $\nabla SINR_{\mathcal{G}_i}$ of the antenna subarray $\mathbf{Q}_{\mathcal{G}_i}$ can be defined as

$$\nabla SINR_{\mathcal{G}_i} = SINR\left(\mathbf{Q}_{\mathcal{G}_i} \cup n_t'\right) - SINR\left(\mathbf{Q}_{\mathcal{G}_i}\right), \tag{75}$$

where $\left(\mathbf{Q}_{\mathcal{G}_i} \cup n_t'\right)$ means that antenna n_t' is added into subarray $\mathbf{Q}_{\mathcal{G}_i}$. Thus, the optimal subarray $\tilde{\mathbf{Q}}_{\mathcal{G}_i}$ can be rewritten as

$$\tilde{\mathbf{Q}}_{\mathcal{G}_i} = \underset{\mathbf{Q}_{\mathcal{G}_i}}{\operatorname{argmax}} \nabla SINR_i. \tag{76}$$

The algorithm loops as follows:

In the initial stage, the dynamic subarray for each user is the null set and the candidate antenna set contains all antenna elements. When adding antennas to the dynamic subarray, the SINR and SINR increment values are updated separately for each user. Then, the subarray $\tilde{\mathbf{Q}}_i$ with the largest SINR increment value is found and that antenna element is assigned to that subarray. It should be noted that only one antenna is assigned in each antenna assignment phase and the other antennas remain unchanged. The above process is performed iteratively until all antenna assignments are completed.

At the initial stage, the dynamic subarray of each user is an empty set and the candidate antenna set contains all antenna elements. Then, the algorithm updates SINR and SINR increment values of each user respectively when an antenna is added into the dynamic subarray. At last, the algorithm finds the subarray $\tilde{\mathbf{Q}}_{\mathcal{G}_i}$ with the maximal SINR increment value and assigns this antenna element to the optimal subarray $\tilde{\mathbf{Q}}_{\mathcal{G}_i}$. Note that only one antenna is assigned and other antennas remain unchanged at each antenna selection stage. The above process is performed iteratively until all antennas are assigned.

Algorithm 3 Antenna Partitioning Algorithm

1. **Input**: $N_t, G, \mathcal{G}_i, K, \tilde{\mathbf{F}}_{BB}, \tilde{\mathbf{F}}_{RF}, \tilde{\mathbf{H}}_k, \tilde{\mathbf{W}}_{BB}, \tilde{\mathbf{W}}_{RF}$;
2. **Initialize** the analog precoding matrix $N_t, G, \mathcal{G}_i, K, \mathbf{Q}_0 = \{1, \cdots, N_t\}, \mathbf{Q}_{\mathcal{G}_i,1}, \cdots, \mathbf{Q}_{\mathcal{G}_i,g_i} = \phi; i = 1, \ldots, G, n_t = 1;$
3. **Repeat**
4. **for** $n_t = 1 : N_t$ **do**
5. **for** $i = 1 : G$ **do**
6. **for** $k = 1 : K$ **do**
7. **Calculate**

$$SINR\left(\mathbf{Q}_k\right) = \sum_{k=1}^{K} \times \frac{\left\| \tilde{\mathbf{W}}_{BB,k}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_k \left(\tilde{\mathbf{F}}_{RF} \circ \mathbf{Q}_k \right) \tilde{\mathbf{F}}_{BB,k} \right\|_F^2 P_k}{\sum_{k'=1,k' \neq k}^{K} \left\| \tilde{\mathbf{W}}_{BB,k}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_k \left(\tilde{\mathbf{F}}_{RF} \circ \mathbf{Q}_k \right) \tilde{\mathbf{F}}_{BB,k'} \right\|_F^2 P_{k'} + \sigma_k^2}$$

8. Adding antennas to dynamic subarrays

$$SINR\left(\mathbf{Q}_k \cup n_t\right) = \sum_{k=1}^{K} \frac{\left\| \tilde{\mathbf{W}}_{BB,k}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_k \left(\tilde{\mathbf{F}}_{RF} \circ \left(\mathbf{Q}_k \cup n_t' \right) \right) \tilde{\mathbf{F}}_{BB,k} \right\|_F^2 P_k}{\sum_{k'=1,k' \neq k}^{K} \left\| \tilde{\mathbf{W}}_{BB,k}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_k \left(\tilde{\mathbf{F}}_{RF} \circ \left(\mathbf{Q}_k \cup n_t' \right) \right) \tilde{\mathbf{F}}_{BB,k'} \right\|_F^2 P_{k'} + \sigma_k^2}$$

9. Update the SINR and SINR incremental values for each user separately

$$\begin{aligned} SINR\left(\mathbf{Q}_{\mathcal{G}_i}\right) &= \sum_{k=1}^{g_i} SINR\left(\mathbf{Q}_k\right) \\ SINR\left(\mathbf{Q}_{\mathcal{G}_i} \cup n_t'\right) &= \sum_{k=1}^{g_i} SINR\left(\mathbf{Q}_k \cup n_t'\right) \\ \nabla SINR_{\mathcal{G}_i} &= SINR\left(Q_{\mathcal{G}_i} \cup n_t'\right) - SINR\left(Q_{\mathcal{G}_i}\right) \end{aligned}$$

10. Update, until the maximum SINR increment value is found

$$\tilde{\mathbf{Q}}_{\mathcal{G}_i} = \underset{\mathbf{Q}_{\mathcal{G}_i}}{\operatorname{argmax}} \nabla SINR_{\mathcal{G}_i}$$

assign that antenna element to that subarray

$$\tilde{\mathbf{Q}}_{\mathcal{G}_i} \leftarrow \tilde{\mathbf{Q}}_{\mathcal{G}_i} \cup n_t, \mathbf{Q}_0 \leftarrow \mathbf{Q}_0 / n_t$$

11. **end for**
12. **end for**
13. **Until** convergence, **end if**
14. **Output**: $\tilde{\mathbf{Q}} = \left[\tilde{\mathbf{Q}}_{\mathcal{G}_1}, \cdots, \tilde{\mathbf{Q}}_{\mathcal{G}_G} \right]$;

For the inter-cell, its nearby inter-cell users are much larger than that of the non-adjacent cells. The interference intensity is the same. Therefore, the SINR for a b-th cell from inter-cell interference is expressed as:

$$SINR_b = \frac{\left| \tilde{\mathbf{W}}_{BB,b}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_b^H \left(\tilde{\mathbf{F}}_{BB,b}^{(1)} \tilde{\mathbf{F}}_{BB,b}^{(2)} \circ \tilde{\mathbf{Q}}_b \right) \right|^2 P_b}{\sum_{b'=1,b' \neq b}^{B} \left| \tilde{\mathbf{W}}_{BB,b'}^H \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{H}}_{b'}^H \left(\tilde{\mathbf{F}}_{BB,b'}^{(1)} \tilde{\mathbf{F}}_{BB,b'}^{(2)} \circ \tilde{\mathbf{Q}}_{b'} \right) \right|^2 P_{b'} + \sigma_b^2}. \tag{77}$$

The capacity of mmWave massive MIMO system can be expressed as

$$SUM = \sum_{\mathcal{G}_i=\mathcal{G}_1}^{\mathcal{G}_L} \log_2(1 + SINR_{\mathcal{G}_i}) + \sum_{b=1}^{B} \log_2(1 + SINR_b) = \sum_{\mathcal{G}_i=\mathcal{G}_1}^{\mathcal{G}_G} \log_2\left(1 + SINR_{\mathcal{G}_i}\right) + \sum_{b=1}^{B} \log_2(1 + SINR_b), \tag{78}$$

(78) can be written in (79), as shown at the bottom of the next page.

V. COMPUTATIONAL COMPLEXITY ANALYSIS

The quantitative analysis of complexity is shown in this part in order to make the complexity of the proposed algorithm increasingly clear. Complexity is measured in terms of the number of flops. A complex addition and multiplication have 2 and 6 flops, respectively [49]. The flop counts of several matrixes operations in the signal detection algorithms are given as follows [50].

- The flop count for SVD of the matrix is $24N_{tRF}N_t^2 + 48N_{tRF}^2N_t + 54N_{tRF}^3$.
- The Moore-Pseudo of a matrix generally is calculated by the Greville method, and the flop count is $12N_{tRF}N_t^2 - 5N_{tRF}N_t - N_t^2$

Without loss of generality, we assume that the antennas at transmitter and receiver are in the similar magnitude, namely $N_{tRF} = \mathcal{O}(N_t)$.

In [46], for algorithm 1, it only requires to calculate the Moore-Pseudo of $(K-1)N_{tRF}N_t$ channel matrix. Then a total number of Moore-Pseudo for this algorithm is K when there are K users in b-th cell. So, the computational complexity is $\mathcal{O}\left(4KN_{tRF}^2N_t^2\right)$.

For algorithm 2, it just requires one SVD operation for the $N_{tRF}N_t$ matrix $\tilde{\mathbf{F}}_{RF,b}$. Water-filling needs $2K^2N_{tRF}^2+6KN_{tRF}$. So, the computational complexity is $\mathcal{O}\left(K^2N_{tRF}^3N_t^2\right)$.

For algorithm 3, Consider a multi-user downlink system with N_{tRF} RF chains and N_t transmit antennas. The problem in (76) is an optimization problem, which the exhaustive search is required to find the optimal solution from all probable cases. Such that, the computational complexity is given by $\mathcal{O}\left(KN_{tRF}N_t^2\right)$. We compare the computational complexity of the proposed algorithm with other conventional schemes in the next section.

VI. SIMULATION RESULTS

In this section, we present simulation results to examine the estimation performance of the proposed scheme. We compare the method in this paper with several traditional methods, i.e., MO [17], RTRNM [20], LCG [30] and VPS [31].

Without loss of generality, The essential simulation parameters are the same as those in [23] and [51], and are provided in Table 1. In the simulations, the geometric channel model with L scattering clusters is adopted as described in Subsection 2. The complex gain of each path obeys the distribution $CN(0, 1)$. The ULA is adopted in simulations [17].

TABLE 1. Simulation parameters.

Number of antennas at user, N_r / K	2
Number of scattering clusters, L	4
Range of azimuth angle	uniformly distribution in $[-180^o, 180^o]$
Antenna spacing	0.5λ
Carrier frequency	60GHz
Power consumption of RF chain, P_R	250mW
Power consumption of PS, P_{PS}	1mW
Power amplifier efficiency, η_p	0.38

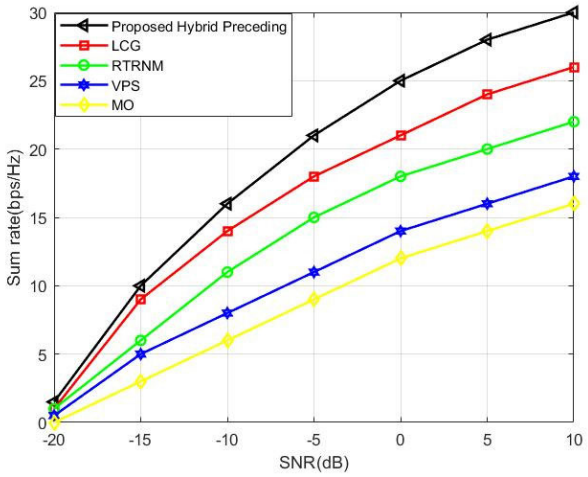


FIGURE 6. Sum-rate comparison of different schemes when BS equips 64 antennas (ULA).

Fig.6 shows the differents of sum rate performance with traditional schemes in the millimeter-wave massive MIMO system of hybrid precoding. Let $N_t = 64$, $N_{tRF} = 16$, and $K = 16$. As observed from Fig.6 the sum rate of the proposed solution significantly outperforms traditional methods. The reason for this situation is that the antenna elements in the proposed solution are adaptively partitioned to RF chains according to the long-term channel information. It can greatly reduce the size of the search space and the calculation complexity. In Fig.7, let $N_t = 128$. From Fig. 7, In the case of different signal-to-noise ratios, a conclusion identical to Fig. 6 can be drawn. More importantly, the performance gaps of three array architectures are more obvious with $N_t = 128$ chains observed from Fig. 7.

Fig.8 compares the sum rate of precoding between this paper and the traditional scheme for different number of users. The proposed antenna partitioning algorithm guarantees the user fairness since each antenna element is allocated to acquire the maximal SINR increment of all selected users. The order of users in the antenna selection process in other schemes leads to serious inequities, because the first user can select the entire antenna unit, while other users can only select the remaining units.

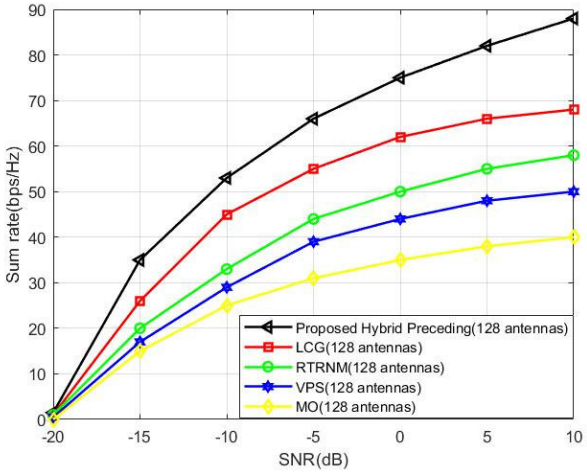


FIGURE 7. Sum-rate comparison of different schemes when BS equips 128 antennas (ULA).

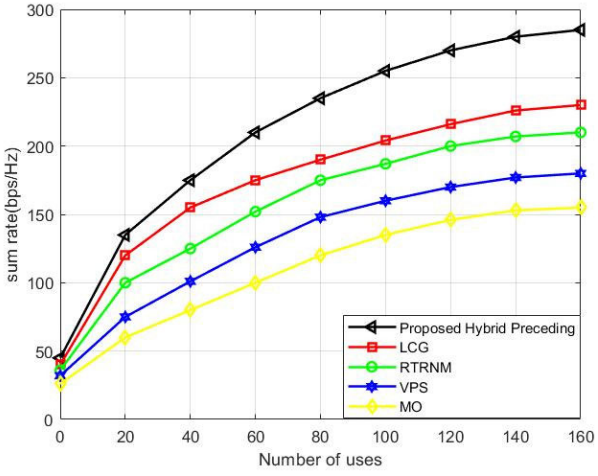


FIGURE 8. Sum rate vs. number of transmit antennas.

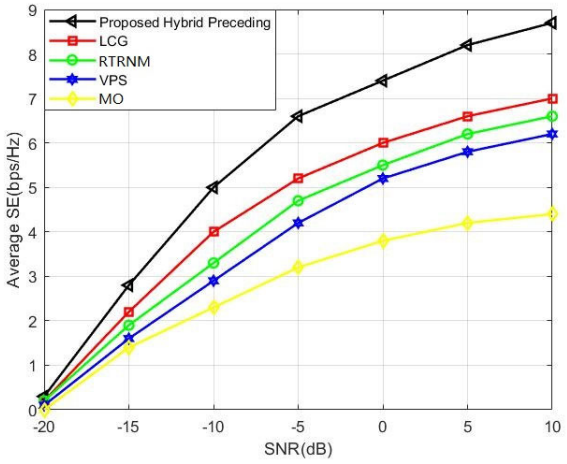


FIGURE 9. Average SE. versus SNR.

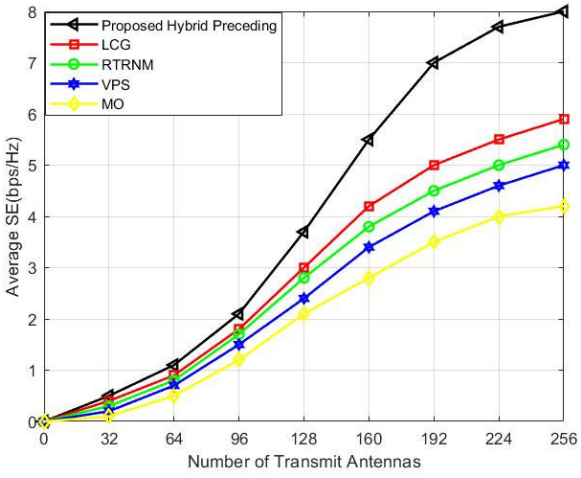


FIGURE 10. Average SE comparison against the numbers of transmit antennas.

Fig.9 is the change trend of the system average SE ratio as the signal-to-noise ratio changes. Fig.10 is the average SE when the BS antenna changes. From the figure we can learn that the average SE achieved by the proposed method is significantly higher than that of other conventional schemes. In Fig.9, the SINR increment is maximum for all MU-MIMO users, thus the adaptive antenna gains more array processing gain and increases the user data rate. Fig.10 show that the improvement of SE is remarkable with the number of antennas increasing. The proposed method can effectively and extensively utilize antennas in multiple low-dimensional manifolds.

$$SUM = \sum_{\mathcal{G}_i=\mathcal{G}_1}^{\mathcal{G}_G} \log_2 \left(1 + \frac{\left| \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} \right|^2 P_{\mathcal{G}_i}}{\sum_{k'=1,k'\neq k}^{|\mathcal{G}_i|} \left| IN_{\mathcal{G}_i,k'} \right|^2 P_{\mathcal{G}_i,k'} + \sum_{i'=1,i'\neq i}^{G} \left| IN_{\mathcal{G}_{i'}} \right|^2 P_{\mathcal{G}_{i'}} + \sigma_{\mathcal{G}_i}^2} \right) + \sum_{b=1}^{B} \log_2 \left(1 + \frac{\left| \tilde{H}_b^H \tilde{F}_b \tilde{W}_b^{(1)} \tilde{W}_b^{(2)} \right|^2 P_b}{\sum_{b'=1,b'\neq b}^{B} \left| \tilde{H}_{b'}^H \tilde{F}_{b'} \tilde{W}_{b'}^{(1)} \tilde{W}_{b'}^{(2)} \right|^2 P_{b'} + \sigma_b^2} \right). \quad (79)$$

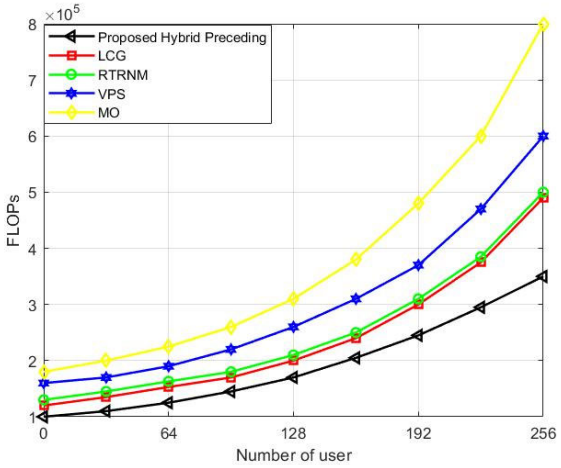


FIGURE 11. The computational complexity.

In Fig.11, it shows a simulation of the computational complexity for the our proposed method and other conventional schemes. According to Fig.11, it can be seen that all the computational complexity increase along with the number of users increasing. Compared with the other conventional schemes, the computational complexity of the proposed method is greatly reduced with the number of users increasing. Thus it is confirmed that the proposed method has the characteristics of lower complexity.

VII. CONCLUSION

In the FSO-RF system for the multi-user scenario, we chose mmWave massive MIMO system in RF link and proposed a hybrid precoding scheme base on manifold learning with antenna subarray partitioning algorithm in this paper. The channel matrix for mmWave massive MIMO is obtained by used manifold learning to obtained the low-dimensional channel matrix. Then user clustering hybrid precoding researched the transmission signal of low-dimensional channel matrix. Manifolds with different user cluster labels were easier to distinguish, and the local spatial correlation of high-dimensional channels in each manifold was enhanced. The antenna subarray partitioning algorithm not only reduces the complexity of antenna assignment, but also ensures fairness among users due to the fact that each antenna element was partitioned to subarrays based on the maximum SINR increment of all served users. Meanwhile, the problem of maximizing the total rate by hybrid precoding was investigated by a manifold quasi-conjugate gradient method through appropriate user clustering. Simulation results show that the hybrid precoding scheme with antenna subarray partitioning algorithm had a significantly better sum-rate and SE than other traditional methods.

REFERENCES

[1] K. Swaminathan, A. Saini, and S. Bitragunta, "A novel hybrid FSO/mmWave/RF system design and analysis for next generation wireless communication," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2019, pp. 1–6.

[2] H. Kaushal and G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 57–96, 1st Quart., 2017.

[3] M. Z. Chowdhury, M. K. Hasan, M. Shahjalal, M. T. Hossan, and Y. M. Jang, "Optical wireless hybrid networks: Trends, opportunities, challenges, and research directions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 930–966, 2nd Quart., 2020.

[4] M. Safari and M. Uysal, "Relay-assisted free-space optical communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5441–5449, Dec. 2008.

[5] I. S. Ansari, F. Yilmaz, and M.-S. Alouini, "On the performance of mixed RF/FSO dual-hop transmission systems," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[6] E. Zedini, H. Soury, and M.-S. Alouini, "On the performance analysis of dual-hop mixed FSO/RF systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3679–3689, May 2016.

[7] B. Bag, A. Das, A. Chandra, and C. Bose, "Capacity analysis for Rayleigh/gamma-gamma mixed RF/FSO link with fixed-gain AF relay," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw.*, 2017, pp. 1747–1757.

[8] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[9] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2012.

[11] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Feb. 2016.

[12] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2013.

[13] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[14] C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Hybrid analog-digital transceiver designs for mmWave amplify-and-forward relaying systems," in *Proc. 41st Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2018, pp. 1–6.

[15] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "Hybrid analog-digital transceiver designs for cognitive radio millimiter wave systems," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove CA, USA, Nov. 2016, pp. 1785–1789.

[16] J. Jiang, Y. Yuan, and L. Zhen, "Multi-user hybrid precoding for dynamic subarrays in mmWave massive MIMO systems," *IEEE Access*, vol. 7, pp. 101718–101728, 2019.

[17] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,' *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[18] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

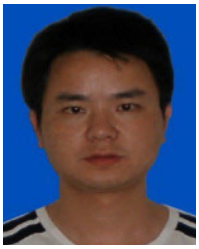
[19] O. El Ayach, R. W. Heath, Jr., S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3476–3480.

[20] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Millimeter-wave massive MIMO systems relying on generalized sub-array-connected hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8940–8950, Sep. 2019.

[21] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog-digital transceivers for single-and multi-carrier large antenna array systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.

[22] S. Park, A. Alkhateeb, and R. W. Heath, Jr., "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[23] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarray architecture for wideband hybrid precoding in millimeter wave massive MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 600–604.
[24] X. Yu, J. Zhang, and K. B. Letaief, "Doubling phase shifters for efficient hybrid precoder design in millimeter-wave communication systems," *J. Commun. Inf. Netw.*, vol. 4, no. 2, pp. 51–67, Jun. 2019.
[25] K. Xu, F.-C. Zheng, P. Cao, H. Xu, and X. Zhu, "A low complexity greedy algorithm for dynamic subarrays in mmWave MIMO systems," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.
[26] A. Li and C. Masouros, "Hybrid analog-digital millimeter-wave MU-MIMO transmission with virtual path selection," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 438–441, Feb. 2017.
[27] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.
[28] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Feb. 2016.
[29] C. Wang, A.-A. Lu, X. Gao, and Z. Ding, "Robust precoding for 3D massive MIMO configuration with matrix manifold optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3423–3437, May 2022.
[30] W. Guo, A.-A. Lu, X. Meng, X. Gao, and N. Ma, "Broad coverage precoding design for massive MIMO with manifold optimization," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2792–2806, Apr. 2019.
[31] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. Ben Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019.
[32] X. Zhou, P. Wang, Z. Yang, L. Tong, Y. Wang, C. Yang, N. Xiong, and H. Gao, "A manifold learning two-tier beamforming scheme optimizes resource management in massive MIMO networks," *IEEE Access*, vol. 8, pp. 22976–22987, 2020.
[33] Y. Sun, Z. Gao, H. Wang, B. Shim, G. Gui, G. Mao, and F. Adachi, "Principal component analysis-based broadband hybrid precoding for millimeter-wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6331–6346, Oct. 2020.
[34] M. Abramowitz and I. A. Stegun, "Handbook of mathematical functions with formulas, graphs, and mathematical tables," Appl. Math. Ser., Nat. Bur. Standards, Gaithersburg, MD, USA, Tech. Rep. 55, 1972.
[35] I. S. Ansari, F. Yilmaz, and M.-S. Alouini, "Performance analysis of FSO links over unified gamma-gamma turbulence channels," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
[36] S. C. Tokgoz, S. Althunibat, S. Yarkan, and K. A. Qaraqe, "Physical layer security of hybrid FSO-mmWave communications in presence of correlated wiretap channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–7.
[37] Wolfram. *The Wolfram Functions Site*. Accessed: Feb. 23, 2022. [Online]. Available: https://functions.wolfram.com
[38] Q. Sun, Z. Zhang, Y. Zhang, M. Lopez-Benitez, and J. Zhang, "Performance analysis of dual-hop wireless systems over mixed FSO/RF fading channel," *IEEE Access*, vol. 9, pp. 85529–85542, 2021.
[39] J. Chen, L. Yang, W. Wang, H.-C. Yang, Y. Liu, M. O. Hasna, and M.-S. Alouini, "A novel energy harvesting scheme for mixed FSO-RF relaying systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8259–8263, Aug. 2019.
[40] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
[41] J. Pascual-García, J. M. Molina-García-Pardo, M. T. Martínez-Inglés, J. V. Rodríguez, and N. Saurín-Serrano, "On the importance of diffuse scattering model parameterization in indoor wireless channels at mm-wave frequencies," *IEEE Access*, vol. 4, pp. 688–701, 2016.
[42] F. Yang, J.-B. Wang, M. Cheng, J.-Y. Wang, M. Lin, and J. Cheng, "A partially dynamic subarrays structure for wideband mmWave MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7578–7592, Dec. 2020.
[43] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
[44] J. D. Leeuw and P. Mair, "Multidimensional scaling using majorization: SMACOF in R," *J. Stat. Softw.*, vol. 31, no. 3, p. 26, Aug. 2009.
[45] X. Zhou, H. Liu, B. Wang, J. Huang, and Y. Wang, "Tensor dictionary manifold learning for channel estimation and interference elimination of multi-user millimeter-wave massive MIMO systems," *IEEE Access*, vol. 10, pp. 5343–5358, 2022.
[46] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, Jan. 2007.
[47] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
[48] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
[49] G. H. Golub and C. F. Vanloan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The John Hopkins Univ. Press, 1996.
[50] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.
[51] D. H. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, Jr., "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, 2017.



XIAOPING ZHOU received the Ph.D. degree in information and communication engineering from the University of Shanghai, Shanghai, China, in 2011. From 2011 to 2013, he was a Postdoctoral Fellow with the Communication Laboratory, Shanghai Jiao Tong University of Science and Technology, China. He is currently a Full Professor with the Shanghai Normal University of Information Science and Technology, Shanghai. His current research interests include mobile communication systems, image processing, parameter estimation, and electrostatic discharge.

XUDONG TIAN was born in Henan, China, in 1996. He is currently pursuing the M.S. degree with Shanghai Normal University (SHNU), Shanghai, China. His current research interests include massive MIMO, millimeter-wave communications, manifold learning, and intelligent reflecting surface.

LE TONG received the B.S. and M.S. degrees in electrical engineering from Donghua University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2017. She is currently a Lecturer at Shanghai Normal University. Her main research interests include optimization of intelligent communication systems, digital image processing, and machine vision.

YANG WANG received the Ph.D. degree in information and communication engineering from the Shanghai Institute of Microsystems and Information Technology, Chinese Academy of Sciences, in 2015. His current research interests include electromagnetic inverse scattering for imaging and computational processing.