

Hybrid Multi-User Precoding with Manifold Discriminative Learning for Millimeter-Wave Massive MIMO Systems

Xiaoping Zhou, Bin Wang, Jing Zhang, Qian Zhang, and Yang Wang
Shanghai Normal University, Shanghai 200234, China

Email: {zxpshnu, 15021222186, wye-mailbox}@163.com, binwang@shnu.edu.cn, qianzhang@shnu.edu.cn

Abstract—In large-array millimeter-wave (mmWave) systems, hybrid multi-user precoding is one of the most attractive research topics. This paper first presents a low-dimensional manifolds architecture for the analog precoder. An objective function is formulated to maximize the Energy Efficiency (EE) in consideration of the insertion loss for hybrid multi-user precoder. The optimal scheme is intractable to achieve, so that we present a user clustering hybrid precoding scheme. By modeling each user set as a manifold, we formulate the problem as clustering-oriented multi-manifolds learning. We discuss the effect of non-ideal factors on the EE performance. Through proper user clustering, the hybrid multi-user precoding is investigated for the sum-rate maximization problem by manifold quasi conjugate gradient methods. The high signal to interference plus noise ratio (SINR) is achieved and the computational complexity is reduced by avoiding the conventional schemes to deal with high-dimensional channel parameters. Performance evaluations show that the proposed scheme can obtain near-optimal sum-rate and considerably higher spectral efficiency than some existing solutions.

Index Terms—mmWave massive MIMO; manifold discriminant analysis; hybrid precoding; user clustering

I. INTRODUCTION

Millimeter-wave (mmWave) massive MIMO (multiple-input multiple-output) communication is a promising technology for next generation wireless communication owing to its abundant frequency spectrum resource [1]-[3]. Due to the high carrier frequency, mmWave signal suffers from high propagation loss so that large-scale antenna arrays are leveraged for path compensation [4]. However, a large number of antennas could lead to the severe hardware cost and power consumption if each antenna requires a Radio Frequency (RF) chain as in conventional fully-digital MIMO systems [5]. To overcome this problem, hybrid MIMO has been emerging to trade off hardware cost with the Spectral Efficiency (SE) and Energy Efficiency (EE) [6]-[8]. Nevertheless, how to design the hybrid precoding over broadband channels is challenging.

How to obtain the optimal precoding matrix is the key issue for hybrid precoding. The large antenna arrays

challenge the low-complexity design of hybrid precoding [9]. In particular, the hybrid precoding may require matrix operations with a scale of antenna size, which is generally large in mmWave communication [8]. To reduce the complexity of hybrid precoding in mmWave massive MIMO system, some advanced schemes based on the beamspace hybrid precoding have been proposed [10]-[12]. The key ideas of [13]-[17] are to efficiently explore the sparsity of beamspace channel by sparse signal processing techniques. The problem of finding the optimal precoder with a hybrid architecture is posed as a sparse reconstruction problem in [13], [14], leading to algorithms and solutions based on basis pursuit methods. Specifically, a compressive sensing-based hybrid precoding has been proposed in [15], [16], where the channel sparsity is ingeniously exploited to design hybrid precoding with the aid of orthogonal matching pursuit (OMP) algorithm. In multi-user scenario, a low-complexity multi-user hybrid precoding for mmWave systems has been investigated in [17]. A Kronecker decomposition for hybrid beamforming (KDHB) for multi-cell multiuser massive MIMO systems over mmWave channels characterized by sparse propagation paths is proposed [18].

However, considering the limited beamspace resolution, the sparsity of beamspace channel may be impaired by power leakage, indicating that the beamspace channel is not ideally sparse and there are many small nonzero entries. Therefore, some works have considered hybrid precoding for practical interference mmWave channels [19], [20]. Handling interference is challenging due to the large channel dimensionality and the high complexity associated with implementing large precoding matrices [21]. To address the high interference problem, a closed-form wideband hybrid precoding solution was proposed in [22]-[25]. An analytical framework of hybrid beamforming (AFHB) in multi-cell millimeter-wave systems was proposed [26]. The general methodology analytically computes the expected per-cell spectral efficiency of an mmWave multi-cell single-stream system using phase-shifter-based analog beamforming and regularized zero-forcing digital beamforming.

Very recently, manifold learning has been proposed to integrate with mmWave massive MIMO systems. In [27], a manifold optimization (MO) based hybrid precoding algorithm, as well as some low-complexity algorithms, was proposed. A Riemannian conjugate gradient

Manuscript received March 6, 2021; revised September 13, 2021.

This work was supported by the Shanghai Capacity Building Projects in Local Institutions under Grant NO.19070502900.

Corresponding author email: zxpshnu@163.com, binwang@shnu.edu.cn.

doi:10.12720/jcm.16.10.411-422

manifold algorithm is proposed by viewing the feasible region of the constant envelope problem as a complex circle manifold [28]. A Riemannian vector perturbation manifold is explored by jointing design of hybrid RF-baseband precoding for multi-user massive MIMO systems [29]. The nonlinear least squares problem is solved with much lower complexity than both gradient descent and constant envelope optimization. A Riemannian Trust-Region Newton Manifold (RTRNM) is proposed for the optimization beamforming in multi-cluster scenarios [30]. The optimization beamforming is utilized to mitigate inter-cell interference by dividing multi-users into multi-clusters with spatial correlation. However, the multi-user high-dimensional channels are not embedded in the low-dimensional subspaces to achieve dimensionality reduction. A manifold learning two-tier fully-digital beamforming scheme optimizes resource management in massive MIMO networks [31]. The manifold learning algorithm is used to reduce the multi-user high-dimensional channels. It reduces the computational complexity while mitigating inter-cell interference-based fully-digital beamforming. It focuses on the local linear spatial structure between user channels, and ignores the global spatial characteristics.

In this paper, we propose user clustering hybrid precoding to enable efficient and low-complexity operation in mmWave massive MIMO, where a large number of antennas are embedded in low-dimensional subspaces. The mmWave channel measurement results show that the mmWave has a diffuse scattering phenomenon on the surface of the rough scatterer, and the scattering range will increase as the wavelength decreases [32]. For scenarios where users are dense, when there is not enough space between users, diffuse scattering may cause adjacent users to receive signals of the same path. Therefore, it causes serious inter-user interference. Our objective is to design the hybrid precoding matrices, such that (i) they manage the intra-cell and inter-cell interferences with low requirements on the channel knowledge, and (ii) they can be implemented using low complexity hybrid analog/digital architectures, i.e., with a small number of RF chains compared to the number of antennas. A discriminative learning method is presented, called Manifold Discriminant Analysis (MDA) [33], to solve the problem of set classification. By modeling each user set as a manifold, we formulate the problem as clustering-oriented multi-manifolds learning. The manifold discriminative learning seek to learn the embedding low-dimensional manifolds, where manifolds with different user cluster labels are better separated, and the local spatial correlation of the high-dimensional channels within each manifold is enhanced. Most of the high-dimensional channels are embedded in the low-dimensional manifolds by manifold discriminative learning, while retaining the potential spatial correlation of the high-dimensional channels. The nonlinearity of high-dimensional channel is transformed into global and local nonlinearity to achieve dimensionality reduction. In

low-dimensional manifolds, the intra-cluster channels become more clustered and the separability of embedded features is enhanced. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi conjugate gradient methods [34]. In order to improve the spectral efficiency of the system, the design of each cluster analog RF precoder should strike a balance between optimizing self-transmission and the interference. The digital precoding matrix is obtained by Karush Kuhn Tucker (KKT) [35]-[37]. The high signal to interference plus noise ratio (SINR) is achieved and the computational complexity is reduced by avoiding the conventional schemes to deal with high-dimensional channel parameters. Performance evaluations show that the proposed scheme can obtain near-optimal sum-rate and considerably higher spectral efficiency than the conventional schemes.

The remainder of this paper is organized as follows. Section 2 introduces system model and channel models. We focus on dimensionality reduction based on multiuser high-dimension channel in Sections 3, and Sections 4 describes hybrid precoding algorithm based on channel dimensionality reduction. Some simulation results are provided in Section 5. Finally, we conclude this paper in Section 6.

Notations: Upper and lower-case boldface letters represents matrices and vectors, respectively. $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, $tr(\cdot)$, and $\|\cdot\|_F$ are the Hermitian transpose, inverse, transpose, complex conjugate, trace, and Forbenius norm of a matrix, respectively. $E(\cdot)$ is the expectation. $diag(\cdot)$ denotes diagonal matrix. $|\mathcal{G}|$ is the cardinality of the set \mathcal{G} . \otimes indicates the Kronecker product. $\mathcal{CN}(0, \sigma^2)$ represents the zero-mean complex Gaussian distribution with zero mean and the variance σ^2 . $span(Y)$ denotes the subspace spanned by the column vectors of Y . $\nabla(\cdot)$ indicates gradient. Finally, I_N denotes the $N \times N$ identity matrix.

II. SYSTEM MODEL AND CHANNEL MODEL

A. System Model

We consider 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_{RF} RF chains ($N_t \geq N_{RF} \geq K$) serves K single-antenna users, as shown in Fig. 1. To manage the interference and improve the data rate for users, the users are partitioned into L clusters $\mathcal{G}_1, \dots, \mathcal{G}_L$ with $g_i = |\mathcal{G}_i|$, $\sum_{i=1}^L g_i = K$ and $\mathcal{G}_i \cap \mathcal{G}_{i'} = \emptyset, \forall i \neq i'$. \mathcal{G}_i is i th cluster, where $i = 1, \dots, L$. The sets $\{\mathcal{G}_1, \dots, \mathcal{G}_L\}$ are all user clusters.

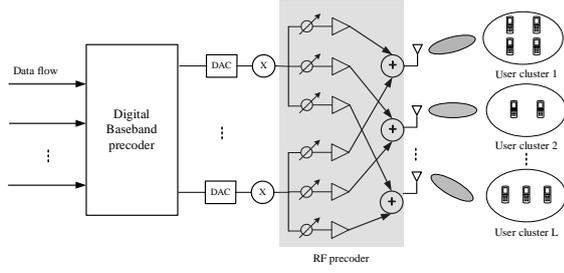


Fig. 1. Hybrid mmWave massive MIMO system model

Let $u_{b,i,k}$, $k=1, \dots, g_i$ denote the k th user of \mathcal{G}_i in the b th cell ($b=1, 2, \dots, B$). The hybrid precoding is performed in two stages: digital precoding in the baseband domain and analog precoding in the RF domain. In a downlink system, the transmit symbols are first applied with digital precoders and the resulting signals are fed to RF chains. The output of the RF chains is processed using analog precoding and subsequently fed to the antenna elements. The transmitted signal vector $x_{b,i,k}$ at the BS is firstly precoded with a digital precoding $W_{b,i,k}$. The resulting signals are fed to analog precoding $F_{b,i,k}$. The received signal $y_{b,i,k}$ of user $u_{b,i,k}$ can be given by

$$y_{b,i,k} = h_{b,i,k}^H F_{b,i,k} W_{b,i,k} x_{b,i,k} + \sum_{k'=1, k' \neq k}^{g_i} h_{b,i,k}^H F_{b,i,k'} W_{b,i,k'} x_{b,i,k'} + \sum_{i'=1, i' \neq i}^L h_{b,i',k}^H F_{b,i'} W_{b,i',k} x_{b,i'} + \sum_{b'=1, b' \neq b}^B h_{b',i,k}^H F_{b',i} W_{b',i,k} x_{b',i} + n_{b,i,k} \quad (1)$$

where $h_{b,i,k} \in \mathbb{C}^{N_i}$ is the channel vector between the BS and user $u_{b,i,k}$. $x_{b,i,k} \in \mathbb{C}^{N_i}$ represents the transmit signal of user $u_{b,i,k}$. $n_{b,i,k} \sim \mathcal{CN}(0, \sigma^2)$ is the spatially white additive Gaussian noise. $F_{b,i,k} \in \mathbb{C}^{N_i \times n_{RF,i}}$ is the analog precoding matrix that adaptively steers an $n_{RF,i}$ dimensional RF beamspace for the coverage of \mathcal{G}_i with $n_{RF,i} \geq g_i$. $W_{b,i,k} \in \mathbb{C}^{n_{RF,i}}$ is the digital precoding matrix. \mathbb{C} is the set of complex numbers.

$\sum_{k'=1, k' \neq k}^{g_i} h_{b,i,k}^H F_{b,i,k'} W_{b,i,k'} x_{b,i,k'}$ are intra-cluster interference.

$\sum_{i'=1, i' \neq i}^L h_{b,i',k}^H F_{b,i'} W_{b,i',k} x_{b,i'}$ are inter-cluster interference.

$\sum_{b'=1, b' \neq b}^B h_{b',i,k}^H F_{b',i} W_{b',i,k} x_{b',i}$ 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 [29], [38].

B. Channel Model

To capture the limited spatial selectivity or scattering characteristics in mmWave massive MIMO channel, we adopt a widely used Saleh-Valenzuela (SV) model in this paper [39], where the channel matrix $h_{i,k}$ of the k th user in i th cluster can be expressed as:

$$h_{i,k} = \sum_{l=1}^{N_l} \alpha_{i,l} \mathbf{a}_r(\theta_{r,i,l}) \mathbf{a}_t(\theta_{t,i,l}) \quad (2)$$

where N_l denotes the number of paths, $\alpha_{i,l}$ is the complex gain of the l th path, $\mathbf{a}_r(\theta_{r,i,l})$ and $\mathbf{a}_t(\theta_{t,i,l})$ are the array response vectors at the user and the BS, respectively, where $\theta_{r,i,l}$ denotes the angle of arrival (AoA) at the user, and $\theta_{t,i,l}$ is the angle of departure (AoD) at the BS. For the simple uniform linear line (ULA) antenna array of N elements, the array response vector is

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j(2\pi/\lambda)d_{ULA} \sin(\theta)}, \dots, e^{j(N-1)(2\pi/\lambda)d_{ULA} \sin(\theta)} \right]^T \quad (3)$$

where λ is the wavelength, and d_{ULA} denotes the antenna spacing. Because of the limited spatial scattering in mmWave propagation, the mmWave massive MIMO channel $h_{i,k}$ is low-rank. As a result, we can leverage a finite number of RF chains to achieve the near-optimal throughput.

III. USER CLUSTERING HYBRID PRECODING SCHEME

Our objective is to design the hybrid precoding matrices, such that (i) they manage the intra-cluster, inter-cluster and inter-cell interference with low requirements on the channel knowledge, and (ii) they can be implemented using low complexity hybrid analog/digital architectures, i.e., with a small number of RF chains compared to the number of antennas. Next, we present the main idea of hybrid precoding based on manifold discriminative learning, a potential solution to achieve these objectives.

A. Manifold Discriminative Learning for User Clusters

As the number of service antennas and users tend to infinity in the mmWave massive MIMO system, the performance is limited by directed inter-cell and intra-cell interferences. The high-dimensional channel matrix requires high complexity hybrid analog/digital architectures. By modeling each user set as a manifold, we formulate the problem as clustering-oriented manifold discriminative learning.

The undirected similarity graph of multi-users is represented by graph embedding method. By modeling each user set as a manifold, the user channel characteristic graphs $\{(h_{i,k}, m_{k,j})\}_{i=1}^L$ are constructed, as shown in Fig. 2.

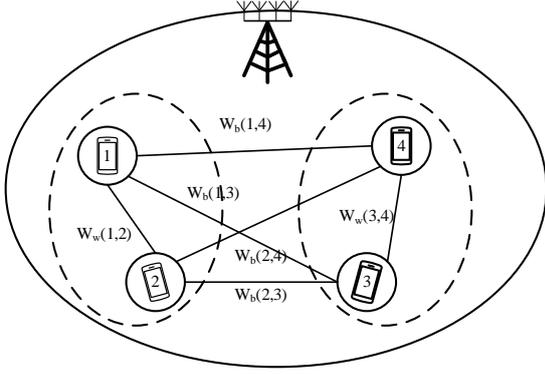


Fig. 2. User cluster undirected characteristic graph

$\mu_{i,k}^{(0)}$ represents the intra-cluster channel weight function between user k and user j . $m_{\xi,k,j}$ represents the inter-cluster channel weight function between user k and j . The sets of the cluster channel weight functions are $M = \{m_{k,j} : k, j \in (1, \dots, K)\}$. The weight function $m_{\xi,k,j}$ of the intra-cluster is defined as follows:

$$\begin{cases} 0 < m_{\xi,k,j} \leq 1, & k, j \text{ in the intra-cluster} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The weight functions of the intra-cluster show that when user k and j are the same cluster, the weight is larger; when user k and j are the different cluster, the weight is 0.

The weight function $m_{\xi,k,j}$ of the inter-cluster is defined as follows:

$$\begin{cases} 0 < m_{\xi,k,j} \leq 1, & k, j \text{ in the inter-cluster} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The weight functions of the inter-cluster show that when user k and j are different cluster, the weight is larger; when user k and j are the same cluster, the weight is 0. The manifold discriminative learning seek to learn the embedding low-dimensional manifolds, where manifolds with different user cluster labels are better separated, and the local spatial correlation of the high-dimensional channels within each manifold is enhanced.

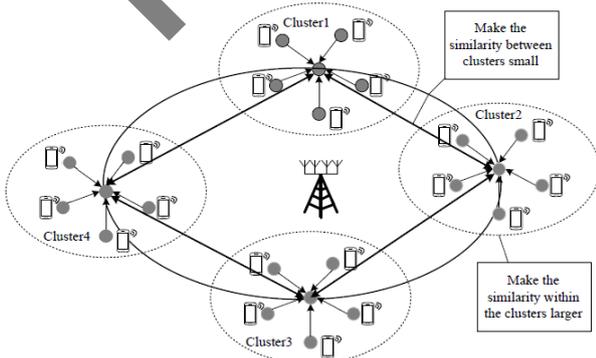


Fig. 3. Schematic diagram of dimension reduction

Some existing manifold learning algorithms, such as LLE [40], can't retain the complete global nonlinear channel structure of user clusters.

We propose to perform the manifold discriminative learning for global dimensionality reduction. The high-dimensional channels are embedded in the low dimensional manifolds, as shown in Fig. 2. In order to reveal the potential non-linear manifold structure of high-dimensional channels, intra-cluster graph and inter-cluster graph are constructed by using the label information of user characteristics. In addition, it can make the low-dimensional channels more clustered, and enhance the separability of embedded low dimensional channels. The RF eigen-beamformers are shown as an optimal solution for user cluster transmission. The channel eigenvector learning corresponding to the maximum eigenvalue is taken as the spatial direction. In theory, the main direction learned is the beamforming. Multi-users of the same cluster have highly correlated transmission paths. We seek to learn a generic mapping A that is defined as:

$$\tilde{h}_k = A^T h_k \quad (6)$$

where A is projection matrix, \tilde{h}_k is the k th user low-dimensional mapping of the high-dimensional channels h_k . The original high-dimensional channels h_k can be transformed into the low-dimensional channels \tilde{h}_k . The relative spatial relationship of neighboring users in high-dimensional channels remains unchanged in low-dimensional manifolds. In order to maintain the manifold structure of the high-dimensional channels, the optimization problem is the projection direction of manifold, i.e., $\forall h_k, h_j (k \neq j)$ of the intra-cluster, the objective function of the intra-cluster can be obtained as

$$\max_A \sum_{k,j} (\tilde{h}_k - \tilde{h}_j)^2 m_{\xi,k,j} \quad (7)$$

Therefore, the projection is posed as a solution maximizing the sum across all uses of the intra-cluster, i.e.,

$$\begin{aligned} & \max_A \sum_{k,j} (\tilde{h}_k - \tilde{h}_j)^2 m_{\xi,k,j} \\ & = \max_A \frac{1}{2} \sum_{k,j} (A^T h_k - A^T h_j)^2 m_{\xi,k,j} \\ & = \max_A A^T H D_{\xi} H^T A - A^T H M_{\xi} H^T A \\ & = \max_A A^T S_{\xi, \text{local}} A \end{aligned} \quad (8)$$

where $S_{\xi, \text{local}} = H(D_{\xi} - M_{\xi})H^T$ is local manifold structure of the intra-cluster, D_{ξ} is diagonal matrix and

$$D_{\xi} = \sum_{k \neq j} M_{\xi}(k, j).$$

According to the SV model, $R_{i,k} = E[h_{i,k} h_{i,k}^H]$ is the covariance matrix of the k th user in the i th cluster. Users

in the same cluster have the similar transmit covariance matrix, hence, $R_{i,k}$, i.e.,

$$R_{i,k} = U_{i,k} \Lambda_{i,k} U_{i,k}^H \quad (9)$$

where $U_{i,k} \in \mathbb{C}^{N_i \times r_i}$ is a matrix of eigenvectors corresponding to r_i ($r_i \ll N_i$) non-zero eigenvalues of $R_{i,k}$. $\Lambda_{i,k}$ is diagonal matrix whose elements are the nonzero eigenvalues of $R_{i,k} \in \mathbb{C}^{N_i \times N_i}$ corresponding to the nonzero eigenvalues, satisfying $\Lambda_k \in \mathbb{C}^{r \times r}$. Since users in the same user cluster have similar spatial correlations, they have similar local scattering, $R_i = R_{i,k}, \forall k \in \mathcal{G}_i$. The criteria of measuring the similarity degree between users is the distance function and the similarity coefficient function. Since $\text{span}(U) \mapsto UU^T$, $\forall U_{i,k}, U_{i',k'}$, the similarity measurement function between any two users based on the distance of subspace projection matrix can be expressed as

$$d_{pm}(U_k U_k^T, V_i V_i^T) = \frac{1}{\sqrt{2}} \|U_k U_k^T - V_i V_i^T\|_F^2 \quad (10)$$

$$= \frac{1}{\sqrt{2}} \text{tr}(\psi_{k,i} \psi_{k,i}^T)$$

where U_k is eigenvectors matrix of R_k in any cluster, i.e., $R_k = U_k \Lambda_k U_k^H$, and V_i is eigenvectors matrix of the i th cluster center R_i . $\psi_{k,i} = U_k U_k^T - V_i V_i^T$ is the symmetric positive semidefinite matrix that needs to be learned. The global manifold structure $S_{\xi, \text{global}}$ of intra-cluster is measured as

$$S_{\xi, \text{global}} = \sum_{i=1}^L \sum_{k \in \mathcal{G}_i} \frac{1}{\sqrt{2}} \frac{1}{g_i} \text{tr}(\psi_{k,i} \psi_{k,i}^T) \quad (11)$$

To effectively utilize the global characteristics and local manifold structure of intra-cluster channels, we can get the intra-cluster dispersion η_{ξ} by combining equations (9) and (11)[33]

$$\eta_{\xi} = \nu S_{\xi, \text{global}} + (1-\nu) S_{\xi, \text{local}} \quad (12)$$

where ν are constants.

The weight functions $m_{\xi, k, j}$ of the intra-cluster can be obtained as

$$m_{\xi, k, j} = \exp(-d_{k, j} / s') \quad (13)$$

where s' is constants, $d_{k, j}$ is the similarity measurement function between user k and user j .

In order to maintain the manifold structure of the inter-cluster user channels, the optimization problem is the projection direction of manifold, i.e., $\forall h_k, h_j (k \neq j)$ of the inter-cluster, the objective function of the inter-cluster can be obtained as

$$\max_A \sum_{k, j} (\tilde{h}_k - \tilde{h}_j)^2 m_{\xi, k, j} \quad (14)$$

Therefore, the projection is posed as a solution maximizing the sum across all uses of the inter-cluster, i.e.,

$$\begin{aligned} & \max_A \sum_{k, j} (\tilde{h}_k - \tilde{h}_j)^2 m_{\xi, k, j} \\ & = \frac{1}{2} \max_A \sum_{k, j} (A^T h_k - A^T h_j)^2 m_{\xi, k, j} \\ & = \max_A A^T H D_{\xi} H^T A - A^T H M_{\xi} H^T A \\ & = \max_A A^T S_{\xi, \text{local}} A \end{aligned} \quad (15)$$

where $S_{\xi, \text{local}} = H(D_{\xi} - M_{\xi})H^T$ is local manifold structure of the inter-cluster, D_{ξ} is diagonal matrix and $D_{\xi} = \sum_{k \neq j} M_{\xi}(k, j)$. The global inter-cluster $S_{\xi, \text{global}}$ is measured as

$$S_{\xi, \text{global}} = \sum_{i=1}^L \sum_{k \in \mathcal{G}_i} \frac{1}{\sqrt{2}} \frac{1}{K - g_i} \text{tr}(\psi_{k,i} \psi_{k,i}^T) \quad (16)$$

To effectively utilize the global characteristics and local manifold structure of inter-cluster channels, we can get the inter-cluster dispersion η_{ξ} by combining equations (15) and (16)

$$\eta_{\xi} = \phi S_{\xi, \text{global}} + (1-\phi) S_{\xi, \text{local}} \quad (17)$$

where ϕ are constants.

The weight functions $m_{\xi, k, j}$ of the inter-cluster can be obtained as

$$m_{\xi, k, j} = \exp(-s'' / d_{k, j}) \quad (18)$$

where s'' is constants.

The discriminative function $J(A)$ is transformed as:

$$J(A) = \max_A \frac{A^T \eta_{\xi} A}{A^T \eta_{\xi} A} \quad (19)$$

$$J(A) = \max_A \frac{A^T (\nu S_{\xi, \text{whole}} + (1-\nu) S_{\xi, \text{local}}) A}{A^T (\phi S_{\xi, \text{whole}} + (1-\phi) S_{\xi, \text{local}}) A} \quad (20)$$

$$\text{s.t. } \tilde{h}_{i, k} = A^T h_{i, k}$$

According to equation (20), the low-dimensional mapping of the k th user channel matrix $\tilde{h}_{i, k}$ is determined by the projection matrix A . By solving the generalized eigenvalues of the discriminative function, we can obtain the projection matrix $A = [A_1, \dots, A_n]$. n is the dimensionality reduction of user channel matrix. After user clustering, the channel correlation of users in the same cluster is enhanced.

Then, according to the intra-cluster graph and inter-cluster graph constructed by using the label information of user characteristics, the user clusters can be divided more accurately with lower complexity. Based on the maximum and minimum distances and the weighted likelihood similarity criterion, an improved spatial fuzzy c-means clustering algorithm is proposed. The algorithm

is an iterative optimization that minimizes the cost function defined as follows:

$$J(\mu_{i,k}) = \sum_{k=1}^K \sum_{i=1}^L \mu_{i,k}^{\mathfrak{S}} d_{i,k} \quad (21)$$

where $d_{i,k} = \frac{1}{\sqrt{2}} \text{tr}(\psi_{k,i} \psi_{k,i}^T)$ is the similarity measurement function between the k th user and the i th cluster center. $\mu_{i,k}$ represents the membership function of user $u_{i,k}$ in the i th cluster, and \mathfrak{S} is a constant. The parameter \mathfrak{S} controls the fuzziness of the resulting partition, and $\mathfrak{S}=2$ is used in this study. The cost function $J(\mu_{i,k})$ is minimized when user $u_{i,k}$ close to the cluster center is assigned high membership values, and low membership values are assigned to user $u_{i,k}$ far from the cluster center. The membership function represents the probability that a user $u_{i,k}$ belongs to a specific cluster. The membership functions and cluster centers are updated by the following:

$$\mu_{i,k} = \frac{1}{\sum_{i'=1}^L (d_{i,k} / d_{i,i'})^{1/(\mathfrak{S}-1)}} \quad (22)$$

and

$$V_{i,k} V_{i,k}^T = \frac{\sum_{k=1}^K \mu_{i,k}^{\mathfrak{S}} U_{i,k} U_{i,k}^T}{\sum_{k=1}^K \mu_{i,k}^{\mathfrak{S}}} \quad (23)$$

where $d_{i,i'} = \frac{1}{\sqrt{2}} \text{tr}(\psi_{i,i'} \psi_{i,i'}^T)$ is the similarity measurement function between the i th cluster center and the i' th cluster center. In summary, by modeling each user set as a manifold, the process of clustering-oriented manifold discriminative learning is as follows:

Step 1: Construct the user channel characteristic graphs $\{(h_{i,k}, m_{k,i})\}_{i=1}^L$;

Step 2: Find out the two most distant U_i and $U_{i'}$, and use them as the central point of the initial user clusters, i.e., $V_1^{(0)} = U_i, V_2^{(0)} = U_{i'}$. The number of the user clusters is $i = 2$;

Step 3: According to Euclidean distance criterion $d_{pm}(U_k U_k^T, V_i V_i^T) = \frac{1}{\sqrt{2}} \text{tr}(\psi_{k,i} \psi_{k,i}^T)$, all users are clustered into i user clusters;

Step 4: In the i user clusters that completed the clustering, the weakest similar point (i.e., the point with the largest distance) is found in each user cluster, and i user clusters are obtained. Then we calculate the sum distance $d_{i,k}$ between the user $k(k=1, 2, \dots, K)$, the membership functions $\mu_{i,k}^{(0)}$ and the center point $V_i^{(0)} (i=1, 2, \dots, L)$ of each user cluster in turn.

Step 5: Calculate the spatial membership function and update the center point $V_i^{(0)} (i=1, 2, \dots, L)$ of each user

$$\text{cluster with } V_{i,k} V_{i,k}^T = \frac{\sum_{k=1}^K (\mu_{i,k}^{\mathfrak{S}}) U_{i,k} U_{i,k}^T}{\sum_{k=1}^K (\mu_{i,k}^{\mathfrak{S}})};$$

Then the maximum value among $d_{i,k}$ is found. $V_{i+1}^{(0)} = \arg \max_k d_{i,k}$. All users into $(i+1)$ are redivided into different user clusters;

Step 6: When the current number of user groups $i = i+1 \geq L$ is true, perform step 5; otherwise repeat step 3;

Step 7: $\|(U_k \Sigma_k^{1/2})^H V_i\|_F^2$ is computed, and each user is assigned to the user clusters with the largest similarity coefficient;

Step 8: Output cluster result, and the number of users in each cluster;

Step 9: Calculate the $m_{\xi,k,j}$ and $m_{\zeta,k,j}$ according to equation (13) and (18); Construct intra-cluster graph and inter-cluster graph by using the label information of user characteristics;

Step 10: Calculate the $S_{\zeta, \text{whole}}$, $S_{\xi, \text{local}}$, $S_{\zeta, \text{local}}$ and $S_{\xi, \text{whole}}$ according to equation (9), (11), (15) and (16);

Step 11: Calculate the η_{ζ} and η_{ξ} according to equation (12) and (17);

Step 12: Optimize the discriminative function $J(A)$ according to equation (20);

Step 13: According to the obtained projection matrix, get the projection in low-dimensional subspace $\tilde{h}_{i,k}$.

B. Manifold Discriminative Learning for User Clusters

On the basis of manifold discriminative learning for global dimensionality reduction and user clustering, we investigate the sum-rate maximization problem for hybrid precoding. Our objective is to design the precoding matrices $\tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i}$, 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

$$\tilde{y}_{\mathcal{G}_i} = \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} + \sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \tilde{H}_{\mathcal{G}_i, k}^H \tilde{F}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'} x_{\mathcal{G}_i, k'} + \sum_{i'=1, i' \neq i}^L \tilde{H}_{\mathcal{G}_i, i'}^H \tilde{F}_{\mathcal{G}_i, i'} \tilde{W}_{\mathcal{G}_i, i'} x_{\mathcal{G}_i, i'} + n_{\mathcal{G}_i} \quad (24)$$

where $\tilde{y}_{\mathcal{G}_i} = [\tilde{y}_{\mathcal{G}_i, 1}^T, \dots, \tilde{y}_{\mathcal{G}_i, g_i}^T]^T$ represents the received signal, $\tilde{H}_{\mathcal{G}_i} = [\tilde{H}_{\mathcal{G}_i, 1}, \dots, \tilde{H}_{\mathcal{G}_i, g_i}]$ represents the channel

matrix for the i th cluster, $\tilde{F}_{\mathcal{G}_i} = [\tilde{F}_{\mathcal{G}_i,1}, \dots, \tilde{F}_{\mathcal{G}_i,g_i}]$ and $\tilde{W}_{\mathcal{G}_i} = \text{diag}(\tilde{W}_{\mathcal{G}_i,1}, \dots, \tilde{W}_{\mathcal{G}_i,g_i})$.

$\sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \tilde{H}_{\mathcal{G}_i,k}^H \tilde{F}_{\mathcal{G}_i,k'} \tilde{W}_{\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'}$ are the intra-cluster

interference, $\sum_{i'=1, i' \neq i}^L \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i}$ 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.

In PCP mode, the analog precoding matrix $\tilde{F}_{\mathcal{G}_i}$ of each cluster is calculated according to manifold quasi-conjugate gradient algorithm, while the digital precoding matrix $\tilde{W}_{\mathcal{G}_i}$ is calculated by each user cluster according to their equivalent channel matrix. Let $\tilde{H}_{eq} = \tilde{H}^H \tilde{F}$ denote the equivalent channel matrix after analog precoding, and it is an approximate block diagonal matrix, which can be expressed as

$$\tilde{H}_{eq} = \begin{bmatrix} \tilde{H}_{\mathcal{G}_1}^H \tilde{F}_{\mathcal{G}_1} & \tilde{H}_{\mathcal{G}_1}^H \tilde{F}_{\mathcal{G}_2} & \dots & \tilde{H}_{\mathcal{G}_1}^H \tilde{F}_{\mathcal{G}_L} \\ \tilde{H}_{\mathcal{G}_2}^H \tilde{F}_{\mathcal{G}_1} & \tilde{H}_{\mathcal{G}_2}^H \tilde{F}_{\mathcal{G}_2} & \dots & \tilde{H}_{\mathcal{G}_2}^H \tilde{F}_{\mathcal{G}_L} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{H}_{\mathcal{G}_L}^H \tilde{F}_{\mathcal{G}_1} & \tilde{H}_{\mathcal{G}_L}^H \tilde{F}_{\mathcal{G}_2} & \dots & \tilde{H}_{\mathcal{G}_L}^H \tilde{F}_{\mathcal{G}_L} \end{bmatrix} \quad (25)$$

where $\tilde{H}_{eq_{\mathcal{G}_i}} = \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i}$ represents the diagonal elements of the matrix in (22), off-diagonal elements of the matrix $\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i}$ ($i \neq i'$) represents the interference channel matrix between user clusters. After analog precoding, the inter-cluster interference is eliminated, that is, $\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_j} \approx 0$. \tilde{H}_{eq} can be expressed as

$$\tilde{H}_{eq} \approx \begin{bmatrix} \tilde{H}_{\mathcal{G}_1}^H \tilde{F}_{\mathcal{G}_1} & 0 & \dots & 0 \\ 0 & \tilde{H}_{\mathcal{G}_2}^H \tilde{F}_{\mathcal{G}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{H}_{\mathcal{G}_L}^H \tilde{F}_{\mathcal{G}_L} \end{bmatrix} \quad (26)$$

The digital precoding matrix \tilde{W} is a block diagonal matrix, which can be expressed as

$$\tilde{W} = \text{diag}(\tilde{W}_{\mathcal{G}_1}, \dots, \tilde{W}_{\mathcal{G}_L}) \quad (27)$$

With scalar equalization $\beta_{\mathcal{G}_i}^{-1}$, the signal estimate $\hat{x}_{\mathcal{G}_i}$ for \mathcal{G}_i can be expressed as

$$\hat{x}_{\mathcal{G}_i} = \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} + \sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \tilde{H}_{\mathcal{G}_i,k}^H \tilde{F}_{\mathcal{G}_i,k'} \tilde{W}_{\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'} + \sum_{i'=1, i' \neq i}^L \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} + n_{\mathcal{G}_i} \right) \quad (28)$$

where $\beta_{\mathcal{G}_i}$ 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

$$\begin{aligned} \varepsilon(\tilde{F}_{\mathcal{G}_i}, \tilde{W}_{\mathcal{G}_i}, \beta_{\mathcal{G}_i}) &= E \left[\|x_{\mathcal{G}_i} - \hat{x}_{\mathcal{G}_i}\|^2 \right] \\ &= E \left[\|x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} (\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i})\|^2 \right] + \\ &E \left[\sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \|\beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i,k}^H \tilde{F}_{\mathcal{G}_i,k'} \tilde{W}_{\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'}\|^2 \right] + \\ &E \left[\sum_{i'=1, i' \neq i}^L \|\beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i}\|^2 + \beta_{\mathcal{G}_i}^{-2} n_{\mathcal{G}_i} \right] \end{aligned} \quad (29)$$

The conditional MSE in (24) is simplified as

$$\varepsilon(\tilde{F}_{\mathcal{G}_i}, \tilde{W}_{\mathcal{G}_i}, \beta_{\mathcal{G}_i}) = \varepsilon_{\mathcal{G}_i}^{(1)} + \varepsilon_{\mathcal{G}_i}^{(2)} \quad (30)$$

where

$$\varepsilon_{\mathcal{G}_i}^{(1)} = E \left[\|x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} (\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i})\|^2 \right] \quad (31)$$

$$\begin{aligned} \varepsilon_{\mathcal{G}_i}^{(2)} &= E \left[\sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \|\beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i,k}^H \tilde{F}_{\mathcal{G}_i,k'} \tilde{W}_{\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'}\|^2 \right] + \\ &E \left[\sum_{i'=1, i' \neq i}^L \|\beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i}\|^2 + \beta_{\mathcal{G}_i}^{-2} n_{\mathcal{G}_i} \right] \end{aligned} \quad (32)$$

Therefore, the hybrid precoding based on interference leakage is jointly optimized with $\tilde{F}_{\mathcal{G}_i}$, $\tilde{W}_{\mathcal{G}_i}$, $\beta_{\mathcal{G}_i}$. According to the literature [20], $\tilde{W}_{\mathcal{G}_i}$ can be decomposition into $\tilde{W}_{\mathcal{G}_i} = \beta_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i}'$, where $\tilde{W}_{\mathcal{G}_i}'$ is an unnormalized digital precoding matrix, which can be obtained by KKT conditions as

$$\begin{aligned} \tilde{W}_{\mathcal{G}_i}' &= (\tilde{H}_{eq_{\mathcal{G}_i}}^H \tilde{H}_{eq_{\mathcal{G}_i}} + \gamma_{\mathcal{G}_i}^{-1} I_{\mathcal{G}_i})^{-1} \tilde{H}_{eq_{\mathcal{G}_i}}^H \\ &= (\tilde{F}_{\mathcal{G}_i}^H \tilde{H}_{\mathcal{G}_i} \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} + \gamma_{\mathcal{G}_i}^{-1} I_{\mathcal{G}_i})^{-1} \tilde{F}_{\mathcal{G}_i}^H \tilde{H}_{\mathcal{G}_i} \end{aligned} \quad (33)$$

where $\gamma_{\mathcal{G}_i}^{-1}$ is regularization factor, which depends on noise variance and base station transmit power. $I_{\mathcal{G}_i}$ can be expressed as

$$I_{\mathcal{G}_i} = \sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} H_{\mathcal{G}_i,k}^H F_{\mathcal{G}_i,k'} W_{\mathcal{G}_i,k'} x_{\mathcal{G}_i,k'} + \sum_{i'=1, i' \neq i}^L \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} + n_{\mathcal{G}_i} \quad (34)$$

The optimal value given in [13] is $\gamma_{\mathcal{G}_i}^{-1} = P_{tot} / K \sigma^2$. P_{tot} is the total power of the transmitted signal. The optimal scaling factor $\beta_{\mathcal{G}_i}$ can be obtained from the base station transmission power with $\text{tr}(\tilde{F} \tilde{W} \tilde{W}^H \tilde{F}^H) \leq P_{tot}$ as

$$\beta_{\mathcal{G}_i} = \sqrt{\frac{P_{tot}}{\sum_{i=1}^L \text{tr}(\tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i}^H)}} \quad (35)$$

Accordingly, equation (31) can be expressed as

$$\begin{aligned}
 \varepsilon_{\mathcal{G}_i}^{(1)} &= E \left[\left\| x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right\|^2 \right] \\
 &= E \left\{ \text{tr} \left[\left(x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right)^H \left(x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right) \right] \right\} \\
 &= E \left\{ \text{tr} \left[\left(x_{\mathcal{G}_i}^H - \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right)^H \right) \left(x_{\mathcal{G}_i} - \beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right) \right] \right\} \\
 &= E \left\{ \text{tr} \left(x_{\mathcal{G}_i}^H x_{\mathcal{G}_i} \right) \right\} - E \left\{ \text{tr} \left[\beta_{\mathcal{G}_i}^{-1} x_{\mathcal{G}_i}^H \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right] \right\} - \\
 &E \left\{ \text{tr} \left[\beta_{\mathcal{G}_i}^{-1} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right)^H x_{\mathcal{G}_i} \right] \right\} + \\
 &E \left\{ \text{tr} \left[\beta_{\mathcal{G}_i}^{-2} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right)^H \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right) \right] \right\} \quad (36)
 \end{aligned}$$

After simple mathematical derivation, equation (32) can be expressed as

$$\begin{aligned}
 \varepsilon_{\mathcal{G}_i}^{(2)} &= E \left[\sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \left\| \beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'} x_{\mathcal{G}_i, k'} \right\|^2 \right] + \\
 &\sum_{i=1, i' \neq i}^L E \left[\left\| \beta_{\mathcal{G}_i}^{-1} \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} x_{\mathcal{G}_i} \right\|_F^2 + n_{\mathcal{G}_i} \right] \\
 &= \sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \beta_{\mathcal{G}_i}^{-2} \text{tr} \left[\tilde{H}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'} E \left(x_{\mathcal{G}_i, k'} x_{\mathcal{G}_i, k'}^H \right) \tilde{W}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'}^H \tilde{H}_{\mathcal{G}_i, k'} \right] + \\
 &\sum_{i'=1, i' \neq i}^L \beta_{\mathcal{G}_i}^{-2} \text{tr} \left[\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} E \left(x_{\mathcal{G}_i} x_{\mathcal{G}_i}^H \right) \tilde{W}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i}^H \tilde{H}_{\mathcal{G}_i} \right] + \beta_{\mathcal{G}_i}^{-2} g_{\mathcal{G}_i} \sigma^2 \\
 &= \sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \beta_{\mathcal{G}_i}^{-2} g_i \text{tr} \left(\tilde{H}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'}^H \tilde{H}_{\mathcal{G}_i, k'} \right) + \\
 &\sum_{i'=1, i' \neq i}^L \beta_{\mathcal{G}_i}^{-2} g_{i'} \text{tr} \left(\tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i}^H \tilde{H}_{\mathcal{G}_i} \right) + \beta_{\mathcal{G}_i}^{-2} g_{\mathcal{G}_i} \sigma_n^2 \quad (37)
 \end{aligned}$$

Let $J(\tilde{F}_{\mathcal{G}_i})$ represent the objective function. The hybrid precoding optimization problem based on interference leakage under orthogonal constraints is

$$\arg \min_{\tilde{F}_{\mathcal{G}_i}} J(\tilde{F}_{\mathcal{G}_i}) \triangleq \sum_{i=1}^L \varepsilon_{\mathcal{G}_i}^{(1)} + \sum_{i=1}^L \varepsilon_{\mathcal{G}_i}^{(2)} \quad \text{s.t.} \quad \tilde{F}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} = I_{n_i} \quad (38)$$

It can be seen that the solution of the objective function is a convex optimization problem. It is essentially to find a radio frequency precoding matrix so that the objective function obtains a minimum value. This problem can be equivalent to an unconstrained optimization problem, which can be solved by using manifold optimization algorithms [35]. The Euclidean conjugate gradient of $J(\tilde{F}_{\mathcal{G}_i})$ can be expressed as

$$\nabla J(\tilde{F}_{\mathcal{G}_i}) = \frac{\partial J(\tilde{F}_{\mathcal{G}_i})}{\partial \tilde{F}_{\mathcal{G}_i}^*} \quad (39)$$

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

$$Z_{\mathcal{G}_i, t+1} = -\nabla J(\tilde{F}_{\mathcal{G}_i, t+1}) + \Gamma_{\mathcal{G}_i, t} Z_{\mathcal{G}_i, t} \quad (40)$$

where

$$\Gamma_{\mathcal{G}_i, t} = \frac{\left\| \nabla J(\tilde{F}_{\mathcal{G}_i, t+1}) \right\|_F^2}{\left\| \nabla J(\tilde{F}_{\mathcal{G}_i, t}) \right\|_F^2} \quad (41)$$

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

Step 1: Initialize the analog precoding matrix $\tilde{F}_{\mathcal{G}_i, 1}$, error threshold $\epsilon \in (0, 1)$, the initial gradient $Z_{\mathcal{G}_i, 1} = -\nabla J(\tilde{F}_{\mathcal{G}_i, 1})$, the number of initialization iterations $t = 1$;

Step 2: If $\left\| \nabla J(\tilde{F}_{\mathcal{G}_i, t}) \right\| \leq \epsilon$, stop; Otherwise, search $\Upsilon_{\mathcal{G}_i, t}$ satisfying

$$J(\tilde{F}_{\mathcal{G}_i, t} + \Upsilon_{\mathcal{G}_i, t} Z_{\mathcal{G}_i, t}) = \min_{\Upsilon_{\mathcal{G}_i, t} \geq 0} J(\tilde{F}_{\mathcal{G}_i, t} + \Upsilon_{\mathcal{G}_i, t} Z_{\mathcal{G}_i, t});$$

Step 3: Update the analog precoding matrix $\tilde{F}_{\mathcal{G}_i, t+1}$ using $\tilde{F}_{\mathcal{G}_i, t+1} = \tilde{F}_{\mathcal{G}_i, t} + \Upsilon_{\mathcal{G}_i, t} Z_{\mathcal{G}_i, t}$;

Step 4: If $t < n$, perform step 5; otherwise repeat step 6;

Step 5: Update $Z_{\mathcal{G}_i, t+1} = -\nabla J(\tilde{F}_{\mathcal{G}_i, t+1}) + \Gamma_{\mathcal{G}_i, t} Z_{\mathcal{G}_i, t}$, where

$$\Gamma_{\mathcal{G}_i, t} = \frac{\left\| \nabla J(\tilde{F}_{\mathcal{G}_i, t+1}) \right\|_F^2}{\left\| \nabla J(\tilde{F}_{\mathcal{G}_i, t}) \right\|_F^2}; \quad \text{Update the number of iterations } t = t + 1, \text{ repeat step 2;}$$

Step 6. Update $\tilde{F}_{\mathcal{G}_i, 1} = \tilde{F}_{\mathcal{G}_i, t+1}$, $Z_{\mathcal{G}_i, 1} = -\nabla J(\tilde{F}_{\mathcal{G}_i, 1})$, $t = 1$, repeat step 2.

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:

$$\text{SINR}_{\mathcal{G}_i} = \frac{\left| \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i} \right|^2 P_{\mathcal{G}_i}}{\sum_{k'=1, k' \neq k}^{|\mathcal{G}_i|} \left| IN_{\mathcal{G}_i, k'} \right|^2 P_{\mathcal{G}_i, k'} + \sum_{i'=1, i' \neq i}^L \left| IN_{\mathcal{G}_i} \right|^2 P_{\mathcal{G}_i} + \sigma_{\mathcal{G}_i}^2} \quad (42)$$

where $IN_{\mathcal{G}_i, k'} = \tilde{H}_{\mathcal{G}_i, k'}^H \tilde{F}_{\mathcal{G}_i, k'} \tilde{W}_{\mathcal{G}_i, k'}$, $IN_{\mathcal{G}_i} = \tilde{H}_{\mathcal{G}_i}^H \tilde{F}_{\mathcal{G}_i} \tilde{W}_{\mathcal{G}_i}$, $P_{\mathcal{G}_i}$ 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. 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.

The capacity of mmWave massive MIMO system can be expressed as

$$\text{SUM} = \sum_{\mathcal{G}_i} \log_2(1 + \text{SINR}_{\mathcal{G}_i}) \quad (43)$$

(43) can be written as

$$SUM = \sum_{G_t=G_r} \log_2 \left(1 + \frac{|\tilde{H}_{G_t}^H \tilde{F}_{G_t} \tilde{W}_{G_t}|^2 P_{G_t}}{\sum_{k'=1, k' \neq k}^{G_t} |IN_{G_t, k'}|^2 P_{G_t, k'} + \sum_{i'=1, i' \neq i}^L |IN_{G_t, i'}|^2 P_{G_t, i'} + \sigma_{G_t}^2} \right) \quad (44)$$

IV. SIMULATION RESULTS

In this section, we will investigate the SE, and BER performance of the proposed hybrid precoder design. We compare our proposed solution with some existing solutions i.e., OMP, KDHB, AFHB, MO and RTRNM. The basic simulation parameters are as follows:

The carrier frequency is 60 GHz. The AoAs and AoDs are uniformly distributed in $[0, 2\pi]$, and a common AS $\Delta=8$. The complex gain of each path follows the distribution $\mathcal{CN}(0,1)$. The ULA is adopted in simulations [27]. In this setting, there is considerable overlap between channel power azimuth spectra, which results in strong inter-cluster interference.

Fig. 4 presents the achievable sum-rate achieved by the proposed hybrid precoding compared with some existing solutions for the mmWave massive MIMO system. We set $N_t = 128$, $N_{RF} = 32$, and $K = 32$. From Fig. 4 we can observe that the proposed hybrid precoding can achieve considerably higher sum-rate than other existing precodings against different signal-to-noise ratio (SNR). This is mainly because the performance of other schemes is limited by the resolution of the multi-user high-dimensional channels nonlinearity. By modeling each user set as a manifold, we formulate the problem as clustering-oriented multi-manifolds learning. A clustered user geometry model is researched for some high-density hotspot scenarios of the cell. The proposed scheme can better eliminate intra-cluster and inter-cluster interferences. The achievable sum-rate of mmWave massive MIMO systems is improved by user clustering hybrid precoding.

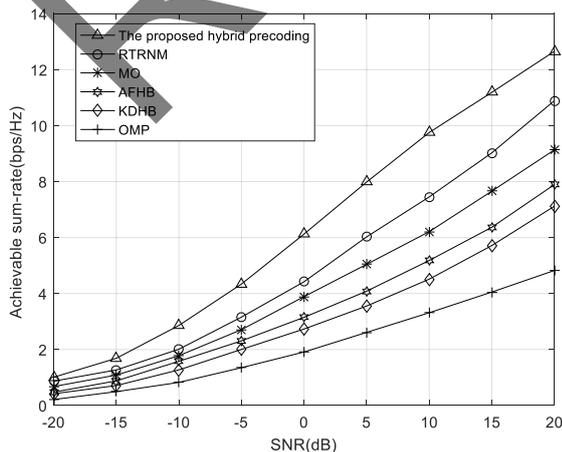


Fig. 4. Achievable sum-rate comparison of different hybrid precoding.

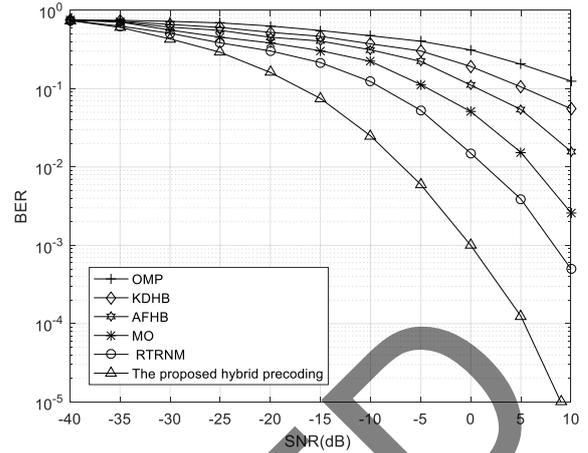


Fig. 5. BER performance comparison of different hybrid precoding

In Fig. 5, we compare the BER performance of different hybrid precoder schemes, where the same channel parameters as considered in Fig. 4 are used for single-cell scenario. From Fig. 5, similar conclusions to those observed for Fig. 4 can be obtained with different SNR. In particular, it can be seen that our proposed-based manifold discriminative learning scheme achieve a better BER performance than other schemes. The proposed scheme improves beamspace resolution and reduces the influence of power leakage on beamspace channel.

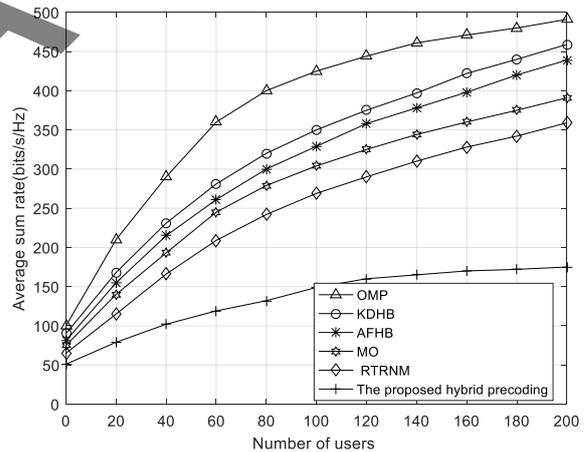


Fig. 6. Average sum rate of two-tier system with different precoding

As shown in Fig. 6, we compare the average sum rate for the proposed scheme, and other existing precoding schemes with different numbers of users. We set $n_{RF, i} \geq g_i$ in each cluster. It is observed from Fig. 6 that the proposed scheme outperforms other schemes. This is mainly because as users increase, the performance of other schemes is limited by the resolution of the multi-user high-dimensional channels nonlinearity. The proposed scheme can better eliminate intra-cluster, inter-cluster and inter-cell interferences. The average sum rate of mmWave massive MIMO systems is improved by user clustering hybrid precoding.

Fig. 7 shows the effect of SNR on the system average SE is given with increasing cell edge SNR. It can be

observed that the proposed scheme provides a significantly higher average SE than other existing schemes. From Fig. 7, we find that each user high-dimensional channels and its neighbor user high-dimensional channels are located in a global and local nonlinear neighborhood by the proposed scheme with manifold discriminative learning. The clustered user geometry model is researched for some high-density hotspot scenarios of the cell. The proposed scheme manages the multi-user and inter-cell interference and improves the data rate for cell-edge users.

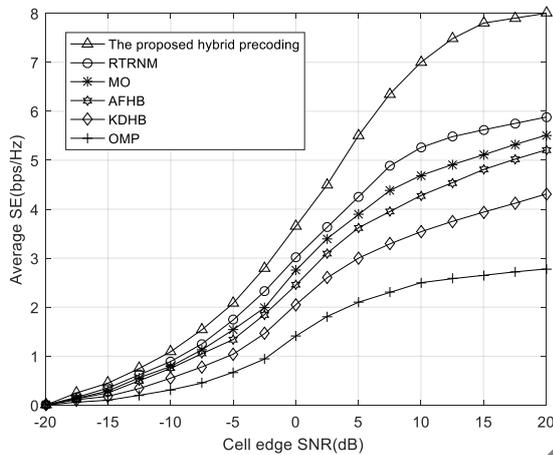


Fig. 7. Average SE. versus cell edge SNR.

V. CONCLUSION

A user clustering hybrid precoding scheme is proposed to enable efficient and low-complexity operation in large scale dimensional mmWave massive MIMO, where a large number of antennas are used in multiple low-dimensional manifolds. For the BS of mmWave massive MIMO, manifold discriminative learning is used to obtain low-dimensional channel matrix. Then user clustering hybrid precoding is studied for the transmitted signal based on the low-dimensional channel matrix. The manifold discriminative learning seek to learn the embedding low-dimensional subspace, where manifolds with different user cluster labels are better separated, and the local spatial correlation of the high-dimensional channels within each manifold is enhanced. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi conjugate gradient methods. The simulation results show that the proposed techniques not only reduce the computational complexity in mmWave massive MIMO system, but also perform well in robustness.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Our project research team jointly conducted the presented study, carried out numerical calculations and analyzed the results obtained both individually and cooperatively. All authors had approved the final version of the paper.

REFERENCES

- [1] X. Wang, M. Jia, Q. Guo, I. W.H. Ho, and J. Wu, "Joint power, original bandwidth, and detected hole bandwidth allocation for multi-homing heterogeneous networks based on cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2777-2790, Mar. 2019.
- [2] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 168-178, Jan. 2015.
- [3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366-385, Mar. 2014.
- [4] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Massive MIMO channel estimation for millimeter wave systems via matrix completion," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1675-1679, Nov. 2018.
- [5] Y. Sun and C. Qi, "Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications," *IEEE Wireless Commun. Lett.*, vol. 21, no. 8, pp. 1883-1886, Oct. 2017.
- [6] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432-1443, Jul. 2017.
- [7] A. Liu, V. K. N. Lau, and M. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 432-444, Jun. 2018.
- [8] S. He, C. Qi, Y. Wu, and Y. Huang, "Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems," *IEEE Access*, vol. 4, pp. 9895-9905, 2016.
- [9] G. C. Alexandropoulos and S. Chouvardas, "Low complexity channel estimation for millimeter wave systems with hybrid A/D antenna processing," in *Proc. IEEE Global Comm. Workshops (GC Wkshps)*, USA: Washington, Dec. 2016, pp. 1-6.
- [10] 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.
- [11] X. Yu, J. Zhang, and K. B. Letaief, "A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 282-297, May 2018.
- [12] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167-19181, Sept. 2017.

- [13] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
- [14] C. Huang, L. Liu, C. Yuen, and S. Sun, "A LSE and sparse message passing-based channel estimation for mmWave MIMO systems," in *Proc. IEEE Global Comm. Workshops (GC Wkshps)*, USA: Washington, Dec. 2016, pp. 1-6.
- [15] Z. Gao, L. Dai, S. Han, C. I. Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 144-153, Jun. 2018.
- [16] C. H. Chen, C. Tsai, Y. Liu, W. Hung, and A. Wu, "Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeterwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1412-1424, Mar. 2017.
- [17] Y. Huang, J. Zhang, and M. Xiao, "Constant envelope hybrid precoding for directional millimeter-wave communications," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 845-859, Apr. 2018.
- [18] G. Zhu and K. Huang, "Hybrid beamforming via the kronecker decomposition for the millimeter-wave massive MIMO systems," *IEEE Journal on Sel. Areas in Commun.*, vol. 35, no. 9, pp. 2097-2114, Sept. 2017.
- [19] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289-5304, Oct. 2017.
- [20] M. Kim and Y. Lee, "MSE-based hybrid RF/Baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714-2720, Jun. 2015.
- [21] T. Mir, M. Z. Siddiqi, and U. Mir, "Machine learning inspired hybrid precoding for wideband millimeter-wave massive MIMO systems," *IEEE Access*, vol. 7, pp. 62852-62864, May 2019.
- [22] J. Zhang, Y. Huang, J. Wang, and L. Yang, "Hybrid precoding for wideband millimeter-Wave systems with finite resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11285-11290, Nov. 2018.
- [23] 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.
- [24] H. Li, M. Li, and Q. Liu, "Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems," *IEEE Trans. on Commun.*, vol. 68, no. 1, pp. 602 – 614, Jan. 2020.
- [25] 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, July 2019.
- [26] S. Sun and T. S. Rappaport, "Analytical framework of hybrid beamforming in multi-cell millimeter-wave systems," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 11, pp. 7528-7543, Nov. 2018.
- [27] 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, Apr. 2016.
- [28] J. C. Chen, "Low-PAPR precoding design for massive multiuser MIMO systems via Riemannian manifold optimization," *IEEE Commun. Letters*, vol. 21, no. 4, pp. 945-948, Jan. 2017.
- [29] R. Mai and T. N. Le, "Two-timescale hybrid RF-baseband precoding with MMSE-VP for multi-user massive MIMO broadcast channels," *IEEE Trans. on Wire. Commun.*, vol. 17, no. 7, pp. 4462-4476, Apr. 2018.
- [30] T. Lin, J. Cong, and Y. Zhu, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3693-3708, Jan. 2018.
- [31] X. Zhou, *et al.*, "A manifold learning two-tier beamforming scheme optimizes resource management in massive MIMO networks," *IEEE Access*, vol. 8, pp. 22976-22987, Jan. 2020.
- [32] S. Sana and D. E. Vittorio, "Millimeter-wave propagation: Characterization and modeling toward fifth-generation systems. [Wireless Corner]," *IEEE Antennas and Propag. Mag.*, vol. 58, no. 6, pp. 115-127, Dec. 2016.
- [33] J. Feng, J. Wang, H. G. Zhang, and Z. Y. Han, "Fault diagnosis method of joint fisher discriminant analysis based on the local and global manifold learning and its kernel version," *IEEE Trans. on Auto. Sci. and Eng.*, vol. 13, no. 1, pp. 122-133, Jan. 2016.
- [34] Y. Sun, Z. Gao, H. Wang, and B. Shim, "Principal component analysis based broadband hybrid precoding for millimeter-wave massive MIMO systems," *IEEE Trans. on Wire. Commun.*, p. 1, June 2020.
- [35] Y. Li, G. Cao, and W. Cao, "LMDAPNet: A novel manifold-based deep learning network," *IEEE Access*, vol. 8, pp. 65938-65946, April 2020.
- [36] S. E. Selvan, U. Amato, and K. A. Gallivan, "Descent algorithms on oblique manifold for source-adaptive ICA contrast," *IEEE Trans. on Neural Net. and Learn. Sys.*, vol. 23, no. 12, pp. 1930-1947, Dec. 2012.
- [37] S. Lu, M. Hong, and Z. Wang, "A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality," *IEEE Trans. on Signal Proc.*, vol. 65, no. 12, pp. 3120-3135, June 2017.
- [38] P. Juan and M. Jos é-María, "On the importance of diffuse scattering model parameterization in indoor wireless channels at mm-Wave frequencies," *IEEE Access*, vol. 4, pp. 688-701, Feb. 2016.
- [39] X. Gao and L. Dai, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 998-1009, 2016.
- [40] D. Meng, G. Cao, and W. Cao, "Supervised feature learning network based on the improved LLE for face recognition," in *Proc. International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, China, Jul. 2016, pp. 306-311.

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.



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 Jiaotong University of Science and Technology, China. He is currently a

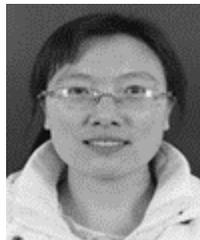
Full Professor with the Shanghai Normal University of Information Science and Technology, Shanghai, China. His current research interests include mobile communication systems, image processing, parameter estimation, electrostatic discharge, and so on.



Bin Wang received her Ph.D. degree from Department of Automation, Shanghai Jiao Tong University, Shanghai, China. She is now the lecturer of Shanghai Normal University, China. Her research interests include computer vision, machine learning, and image processing, multimedia analysis.



Jing Zhang was born in ANHUI, China, in 1997. She is currently pursuing the B.S. degree from the Shanghai Normal University (SHNU), Shanghai, China. Her current research interests include massive MIMO, millimeter-wave communications, and tensor analysis.



Qian Zhang is now the associate professor of Shanghai Normal University, China. She received her P.H.D. from Shanghai University in China. Her research interest fields include video processing.

RETRACTED