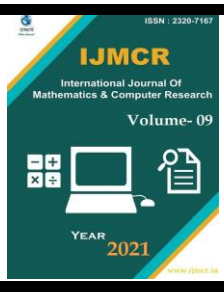
International Journal of Mathematics and Computer Research
ISSN: 2320-7167
Volume 09 Issue 11 November 2021, Page no. – 2455-2468
Index Copernicus ICV: 57.55, Impact Factor: 7.184
DOI: 10.47191/ijmcr/v9i11.02



Manifold Discriminative Learning Inspired Hybrid Beamforming for Millimeter-Wave Massive MIMO Systems

Xiaoping Zhou[1], Yang Liu[2]
[1,2] Shanghai Normal University, shanghai200234, China

ARTICLE INFO

Published Online
02 November 2021

Corresponding Author:
Xiaoping Zhou

ABSTRACT

Millimeter-wave (mmWave) massive MIMO (multiple-input multiple-output) is a promising technology as it provides significant beamforming gains and interference reduction capabilities due to the large number of antennas. However, mmWave massive MIMO is computationally demanding, as the high antenna count results in high-dimensional matrix operations when conventional MIMO processing is applied. Hybrid precoding is an effective solution for the mmWave massive MIMO systems to significantly decrease the number of radio frequency (RF) chains without an apparent sum-rate loss. In this paper, we propose user clustering hybrid precoding to enable efficient and low-complexity operation in high-dimensional mmWave massive MIMO, where a large number of antennas are used in low-dimensional manifolds**.** 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. Through proper user clustering, the hybrid 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.

KEYWORDS: 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 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) [35-39]. 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 II introduces system model and channel models. We focus on dimensionality reduction based on multiuser high-dimension channel in Sections III, and Sections IV describes hybrid precoding algorithm based on channel dimensionality reduction.

Some simulation results are provided in Section V. Finally, we conclude this paper in Section VI.

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. $diag(\square)$ denotes diagonal matrix. $|\mathrm{G}|$ is the cardinality of the set G. $\otimes$ indicates the Kronecker product. $\mathrm{CN}(0,\sigma^2)$ 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 MODELS

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 $\mathrm{G}_1,\cdots,\mathrm{G}_L$ with $g_i = |\mathrm{G}_i|$, $\sum_{i=1}^{L} g_i = K$ and $\mathrm{G}_i \cap \mathrm{G}_{i'} = \varnothing, \forall i \neq i'$. G_i is the ith cluster, where $i = 1,\ldots,L$. The sets $\{\mathrm{G}_1,\cdots,\mathrm{G}_L\}$ are all user clusters.

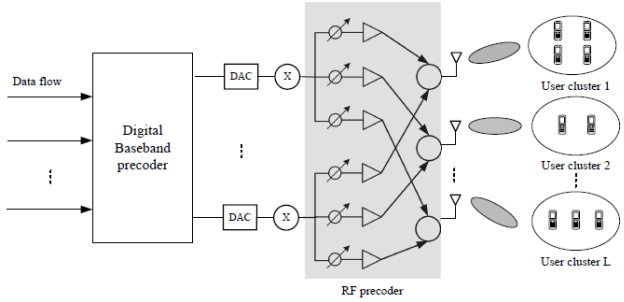


FIGURE 1. Hybrid mmWave massive MIMO system model

Let $u_{b,i,k}, k = 1,\ldots,g_i$ denote the kth user of G_i in the bth cell ($b = 1,2,\cdots,B$). 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

$$y_{b,i,k} = \underbrace{h_{b,i,k}^H F_{b,i,k} W_{b,i,k} x_{b,i,k}}_{\text{Desired signal}} + \underbrace{\sum_{k'=1,k'\neq k}^{|\mathrm{G}_i|} h_{b,i,k}^H F_{b,i,k'} W_{b,i,k'} x_{b,i,k'}}_{\text{Intra-cluster interference}} + \underbrace{\sum_{i'=1,i'\neq i}^{L} h_{b,i'}^H F_{b,i'} W_{b,i'} x_{b,i'}}_{\text{Inter-cluster interference}} + \underbrace{\sum_{b'=1,b'\neq b}^{B} h_{b',i'}^H F_{b',i'} W_{b',i'} x_{b',i'}}_{\text{Inter-cell interference}} + n_{b,i,k} \quad (1)$$

Where $h_{b,i,k} \in \square^{N_t}$ is the channel vector between the BS and user $u_{b,i,k}$. $x_{b,i,k} \in \square^{N_t}$ represents the transmit signal of user $u_{b,i,k}$. $n_{b,i,k} \square \mathrm{CN}(0,\sigma^2)$ is the spatially white additive Gaussian noise. $F_{b,i,k} \in \square^{N_t \times n_{RF,i}}$ is the analog precoding matrix that adaptively steers an $n_{RF,i}$ dimensional RF beamspace for the coverage of G_i with $n_{RF,i} \geq g_i$. $W_{b,i,k} \in \square^{n_{RF,i}}$ is the digital precoding matrix. $\square$ is the set of complex numbers. $\sum_{k'=1,k'\neq k}^{|\mathrm{G}_i|} h_{b,i,k}^H F_{b,\mathrm{G}_{i,k'}} W_{b,i,k'} x_{b,i,k'}$ are intra-cluster interference. $\sum_{i'=1,i'\neq i}^{L} h_{b,i'}^H F_{b,\mathrm{G}_{i'}} W_{b,i'} x_{b,i'}$ are inter-cluster interference. $\sum_{b'=1,b'\neq b}^{B} h_{b',i'}^H F_{b',\mathrm{G}_{i'}} W_{b',i'} x_{b',i'}$ 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[29][40].

B. CHANNEL MODEL

To capture the limited scattering features of multipath fading wideband mmWave channels, the high-dimensional channel vector $h_{i,k}$ between the BS and the user $u_{i,k}$ can be expressed as [41]

$$h_{i,k} = U_{i,k} \Sigma_{i,k}^{1/2} h'_{i,k} \quad (2)$$

Where $h'_{i,k} \sim \mathrm{CN}(0, I_{N_t})$ is represents the small-scale Rayleigh fading channel. I_{N_t} is the $N_t \times N_t$ identity matrix. $U_{i,k} \in \mathrm{C}^{N_t \times r_i}$ is a matrix of eigenvectors corresponding to r_i ($r_i \square N_t$) non-zero eigenvalues of $R_{i,k}$. $\Sigma_{i,k}$ is the transmit correlation matrix composed of r non-zero eigenvalues of $R_{i,k} \in \square^{N_t \times N_t}$, satisfying $\Sigma_k \in \square^{r \times r}$. $R_{i,k}$ is channel covariance matrix for users $u_{i,k}$. According to the one-ring channel model, $R_{i,k} = E\left[h_{i,k} h_{i,k}^{\mathrm{H}}\right]$ is the covariance matrix of the kth user in the ith cluster. Users in the same cluster have the similar transmit covariance matrix, hence, $R_{i,k}$, i.e.,

$$R_{i,k} = U_{i,k} \Sigma_{i,k} U_{i,k}^H \quad (3)$$

Since users in the same user cluster have similar spatial correlations, they have similar local scattering, $R_i = R_{i,k}, \forall k \in G_i$. The correlation coefficient between BS antennas (a'_{RF}, a''_{RF}) is given by:

$$[R_i]_{a'_{RF},a''_{RF}} = \frac{1}{2\Delta_i}\int_{\bar{\theta}_i-\Delta_i}^{\bar{\theta}_i+\Delta_i} e^{\frac{2\pi}{\lambda}d(a'_{RF}-a''_{RF})\sin(\bar{\theta}_i-\theta)}d\theta \qquad (4)$$

where λ indicates the signal wavelength, d indicates the distance between the antennas, θ represents the Angle of Arrival (AoA), $\bar{\theta}_i$ represents the average value of AoA in the cluster, Δ_i represents angle spread (AS) .

III. OUR PROPOSED USER CLUSTERING HYBIRD 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 DISCRIMINANTIVE 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 $\left\{\left(h_{i,k}, m_{k,j}\right)\right\}_{i=1}^{L}$ are constructed, as shown in Fig. 2.

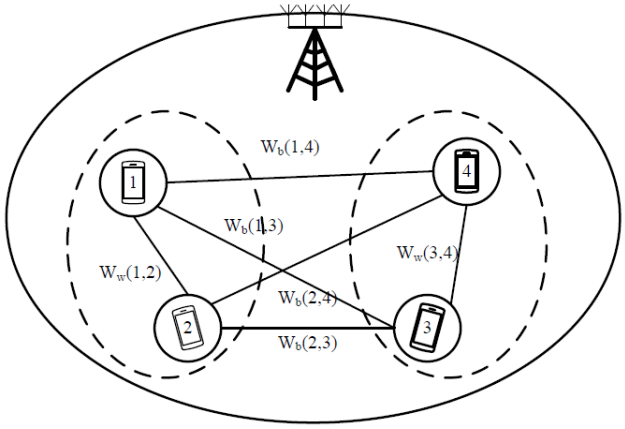


FIGURE 2. User cluster undirected characteristic graph

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

$$\begin{cases} 0 \le m_{\xi,k,j} \le 1, & k, j \text{ in the intra-cluster} \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

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 $m_{\zeta,k,j}$ of the inter-cluster is defined as follows

$$\begin{cases} 0 \le m_{\zeta,k,j} \le 1, & k, j \text{ in the inter-cluster} \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

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.

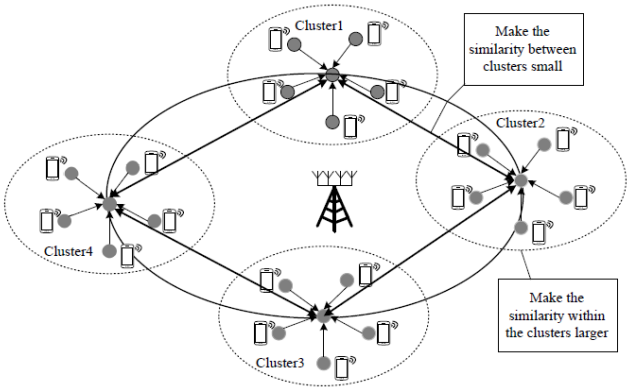


FIGURE 3. Schematic diagram of dimension reduction

Some existing manifold learning algorithms, such as LLE [42], 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:

$$\tilde{h}_k = A^T h_k \qquad (7)$$

Where A is projection matrix, $\tilde{h}_k$ is the kth user low-dimensional mapping of the high-dimensional channels h_k. The original high-dimensional channels h_k can be transformed into a 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., $\forall h_k, h_j \left(k \neq j\right)$ of the intra-cluster, the objective function of the intra-cluster can be obtained as

$$\max_A \sum_{k,j} \left(\tilde{h}_k - \tilde{h}_j\right)^2 m_{\xi,k,j} \qquad (8)$$

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

$$\begin{aligned} &\max_A \sum_{k,j} \left(\tilde{h}_k - \tilde{h}_j\right)^2 m_{\xi,k,j} \\ &= \max_A \frac{1}{2} \sum_{k,j} \left(A^T h_k - A^T h_j\right)^2 m_{\xi,k,j} \\ &= \max_A A^{\mathrm{T}} H D_\xi H^{\mathrm{T}} A - A^{\mathrm{T}} H M_\xi H^{\mathrm{T}} A \\ &= \max_A A^{\mathrm{T}} S_{\xi,\mathrm{local}} A \end{aligned} \qquad (9)$$

where $S_{\xi,\mathrm{local}} = H\left(D_\xi - M_\xi\right)H^{\mathrm{T}}$ is local manifold structure of the intra-cluster, D_ξ is diagonal matrix and $D_\xi = \sum_{k \neq j} M_\xi(k,j)$.

The criteria of measuring the similarity degree between users is the distance function and the similarity coefficient function. Since $span(U) \mapsto UU^T$, $\forall U_{i,k}, U_{i',k'}$, the similarity measurement function between any two users based on the distance of subspace projection matrix can be expressed as

$$\begin{aligned} &d_{pm}\left(U_k U_k^T, V_i V_i^T\right) \\ &= \frac{1}{\sqrt{2}} \left\| U_k U_k^T - V_i V_i^T \right\|_F^2 \\ &= \frac{1}{\sqrt{2}} tr\left(\psi_{k,i} \psi_{k,i}^T\right) \end{aligned} \qquad (10)$$

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 ith 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 $S_{\xi,\mathrm{global}}$ of intra-cluster is measured as

$$S_{\xi,\mathrm{global}} = \sum_{i=1}^{L} \sum_{k \in G_i} \frac{1}{\sqrt{2}} \frac{1}{g_i} tr\left(\psi_{k,i} \psi_{k,i}^T\right) \qquad (11)$$

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]

$$\eta_\xi = \lambda S_{\xi,\mathrm{global}} + (1-\lambda) S_{\xi,\mathrm{local}} \qquad (12)$$

Where λ are constants.

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

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

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., $\forall h_k, h_j \left(k \neq j\right)$ of the inter-cluster, the objective function of the inter-cluster can be obtained as

$$\max_A \sum_{k,j} \left(\tilde{h}_k - \tilde{h}_j\right)^2 m_{\zeta,k,j} \qquad (14)$$

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

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

Where $S_{\zeta,\mathrm{local}} = H\left(D_\zeta - M_\zeta\right)H^{\mathrm{T}}$ is local manifold structure of the inter-cluster, D_ζ is diagonal matrix and $D_\zeta = \sum_{k \neq j} M_\zeta(k,j)$.

The global inter-cluster $S_{\zeta,\mathrm{global}}$ is measured as

$$S_{\zeta,\mathrm{global}} = \sum_{i=1}^{L} \sum_{k \notin G_i} \frac{1}{\sqrt{2}} \frac{1}{K - g_{i'}} tr\left(\psi_{k,i} \psi_{k,i}^T\right) \qquad (16)$$

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

$$\eta_\zeta = \wp S_{\zeta,\mathrm{global}} + (1-\wp) S_{\zeta,\mathrm{local}} \qquad (17)$$

Where $\wp$ are constants.

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

$$m_{\zeta,k,j} = \exp\left(-s'' / d_{k,j}\right) \qquad (18)$$

where s'' is constants.

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

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

$$J(A) = \max_A \frac{A^{\mathrm{T}} \left(\lambda S_{\xi,\mathrm{whole}} + (1-\lambda) S_{\xi,\mathrm{local}}\right) A}{A^{\mathrm{T}} \left(\wp S_{\zeta,\mathrm{whole}} + (1-\wp) S_{\zeta,\mathrm{local}}\right) A} \qquad (20)$$

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

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 $A=[A_1,\cdots,A_n]$. 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:

$$J\left(\mu_{i,k}\right)=\sum_{k=1}^{K}\sum_{i=1}^{L}\mu_{i,k}^{\Im}d_{i,k} \qquad (21)$$

Where $d_{i,k}=\frac{1}{\sqrt{2}}tr\left(\psi_{k,i}\psi_{k,i}^{T}\right)$ 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 $\Im$ is a constant. The parameter $\Im$ controls the fuzziness of the resulting partition, and $\Im$=2 is used in this study. The cost function $J\left(\mu_{i,k}\right)$ 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:

$$\mu_{i,k}=\frac{1}{\sum_{i'=1}^{L}\left(d_{i,k}/d_{i,i'}\right)^{1/(\Im-1)}} \qquad (22)$$

and

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

Where $d_{i,i'}=\frac{1}{\sqrt{2}}tr\left(\psi_{i,i'}\psi_{i,i'}^{T}\right)$ is the similarity measurement function between the *i*th cluster center and the *i'*th cluster center. One of the important characteristics of the intra-cluster is that neighboring users are highly correlated. In other words, these neighboring users possess similar feature values, and the probability that they belong to the same cluster is great. This spatial relationship is important in clustering. To exploit the spatial information, a spatial membership function $\bar{\mu}_{i,k}$ is defined as

$$\bar{\mu}_{i,k}=\sum_{k'\in G_i}\mu_{i,k'} \qquad (24)$$

The spatial membership function $\bar{\mu}_{i,k}$ represents the probability that user $u_{i,k}$ belongs to the *i*th cluster. The spatial membership function of a pixel user a cluster is large if the majority of its neighborhood belongs to the same clusters. The spatial membership function is incorporated into membership function as follows:

$$\mu'_{i,k}=\frac{\mu_{i,k}^{\partial}\bar{\mu}_{i,k}^{\varepsilon}}{\sum_{i'=1}^{L}\mu_{i',k}^{\partial},\bar{\mu}_{i',k}^{\varepsilon}} \qquad (25)$$

where ∂ and ε are parameters to control the relative importance of both functions.

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 $\left\{\left(h_{i,k},m_{k,j}\right)\right\}_{i=1}^{L}$;

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 $d_{pm}\left(U_kU_k^T,V_iV_i^T\right)=\frac{1}{\sqrt{2}}tr\left(\psi_{k,i}\psi_{k,i}^{T}\right)$, 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 $k(k=1,2,...,K)$, the membership functions $\mu_{i,k}^{(0)}$ and the center point $V_i^{(0)}(i=1,2,...,L)$ of each user cluster in turn.

Step 5: Calculate the spatial membership function $\bar{\mu}_{i,k}^{(0)}$, $\mu'^{(0)}_{i,k}$ and update the center point $V_i^{(0)}(i=1,2,...,L)$ of each user cluster with $V_{i,k}V_{i,k}^{T}=\frac{\sum_{k=1}^{K}\left(\mu'_{i,k}\right)^{\Im}U_{i,k}U_{i,k}^{T}}{\sum_{k=1}^{K}\left(\mu'_{i,k}\right)^{\Im}}$;

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: $\left\|(U_k\Sigma_k^{1/2})^HV_i\right\|_F^2$ 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;

10: Calculate the $S_{\zeta,\text{whole}}$, $S_{\xi,\text{local}}$, $S_{\zeta,\text{local}}$ and $S_{\zeta,\text{whole}}$ 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 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.

1). Single cell scenario

In the special case where only one cell of users is scheduled for transmission, eigen-beamforming satisfies such a stronger condition. Our objective is to design the precoding matrices $\tilde{F}_{\mathrm{G}_i}\tilde{W}_{\mathrm{G}_i}$, 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

$$\tilde{y}_{\mathrm{G}_i} = \underbrace{\tilde{H}_{\mathrm{G}_i}^H \tilde{F}_{\mathrm{G}_i}\tilde{W}_{\mathrm{G}_i} x_{\mathrm{G}_i}}_{\text{Desired signal}} + \underbrace{\sum_{k'=1,k'\neq k}^{|\mathrm{G}_i|} \tilde{H}_{\mathrm{G}_i,k}^H \tilde{F}_{\mathrm{G}_i,k'}\tilde{W}_{\mathrm{G}_i,k'} x_{\mathrm{G}_i,k'}}_{\text{Intra-cluster interference}} + \underbrace{\sum_{i'=1,i'\neq i}^{L} \tilde{H}_{\mathrm{G}_{i'}}^H \tilde{F}_{\mathrm{G}_{i'}}\tilde{W}_{\mathrm{G}_{i'}} x_{\mathrm{G}_{i'}}}_{\text{Inter-cluster int erference}} + n_{\mathrm{G}_i} \tag{26}$$

Where $\tilde{y}_{\mathrm{G}_i} = \left[\tilde{y}_{\mathrm{G}_i,1}^T,\cdots,\tilde{y}_{\mathrm{G}_i,g_i}^T\right]^T$ represents the received signal, $\tilde{H}_{\mathrm{G}_i} = \left[\tilde{H}_{\mathrm{G}_i,1},\cdots,\tilde{H}_{\mathrm{G}_i,g_i}\right]$ represents the channel matrix for the *i*th cluster, $\tilde{F}_{\mathrm{G}_i} = \left[\tilde{F}_{\mathrm{G}_i,1},\cdots,\tilde{F}_{\mathrm{G}_i,g_i}\right]$ and $\tilde{W}_{\mathrm{G}_i} = diag\left(\tilde{W}_{\mathrm{G}_i,1},\ldots,\tilde{W}_{\mathrm{G}_i,g_i}\right)$. $\sum_{k'=1,k'\neq k}^{|\mathrm{G}_i|} \tilde{H}_{\mathrm{G}_i,k}^H \tilde{F}_{\mathrm{G}_i,k'}\tilde{W}_{\mathrm{G}_i,k'} x_{\mathrm{G}_i,k'}$ are the intra-cluster interference, $\sum_{i'=1,i'\neq i}^{L} \tilde{H}_{\mathrm{G}_{i'}}^H \tilde{F}_{\mathrm{G}_{i'}}\tilde{W}_{\mathrm{G}_{i'}} x_{\mathrm{G}_{i'}}$ 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 $\tilde{F}_{\mathrm{G}_i}$ of each cluster is calculated according to manifold quasi-conjugate gradient algorithm, while the digital precoding matrix $\tilde{W}_{\mathrm{G}_i}$ 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

$$\tilde{H}_{eq} = \begin{bmatrix} \tilde{H}_{\mathrm{G}_1}^H\tilde{F}_{\mathrm{G}_1} & \tilde{H}_{\mathrm{G}_1}^H\tilde{F}_{\mathrm{G}_2} & \cdots & \tilde{H}_{\mathrm{G}_1}^H\tilde{F}_{\mathrm{G}_L} \\ \tilde{H}_{\mathrm{G}_2}^H\tilde{F}_{\mathrm{G}_1} & \tilde{H}_{\mathrm{G}_2}^H\tilde{F}_{\mathrm{G}_2} & \cdots & \tilde{H}_{\mathrm{G}_2}^H\tilde{F}_{\mathrm{G}_L} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{H}_{\mathrm{G}_L}^H\tilde{F}_{\mathrm{G}_1} & \tilde{H}_{\mathrm{G}_L}^H\tilde{F}_{\mathrm{G}_2} & \cdots & \tilde{H}_{\mathrm{G}_L}^H\tilde{F}_{\mathrm{G}_L} \end{bmatrix} \tag{27}$$

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

$$\tilde{H}_{eq} \approx \begin{bmatrix} \tilde{H}_{\mathrm{G}_1}^H\tilde{F}_{\mathrm{G}_1} & 0 & \cdots & 0 \\ 0 & \tilde{H}_{\mathrm{G}_2}^H\tilde{F}_{\mathrm{G}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{H}_{\mathrm{G}_L}^H\tilde{F}_{\mathrm{G}_L} \end{bmatrix} \tag{28}$$

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

$$\tilde{W} = \mathrm{diag}\left(\tilde{W}_{\mathrm{G}_1},\ldots,\tilde{W}_{\mathrm{G}_L}\right) \tag{29}$$

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

$$\hat{x}_{\mathrm{G}_i} = \beta_{\mathrm{G}_i}^{-1}\left(\tilde{H}_{\mathrm{G}_i}^H\tilde{F}_{\mathrm{G}_i}\tilde{W}_{\mathrm{G}_i}x_{\mathrm{G}_i} + \sum_{k'=1,k'\neq k}^{|\mathrm{G}_i|}\tilde{H}_{\mathrm{G}_i,k}^H\tilde{F}_{\mathrm{G}_{i,k'}}\tilde{W}_{\mathrm{G}_i,k'}x_{\mathrm{G}_i,k'} + \sum_{i'=1,i'\neq i}^{L}\tilde{H}_{\mathrm{G}_{i'}}^H\tilde{F}_{\mathrm{G}_{i'}}\tilde{W}_{\mathrm{G}_{i'}}x_{\mathrm{G}_{i'}} + n_{\mathrm{G}_i}\right) \tag{30}$$

Where β_{G_i} is a scaling equalization that is jointly optimized with the hybrid precoding. The conditional mean square error (MSE) for G_i is defined as

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

$$E\left[\sum_{i'=1,i'\neq i,}^{L}\left\|\beta_{G}^{-1}\tilde{H}_{G,}^{H}\tilde{F}_{G,}\tilde{W}_{G,}x_{G,}\right\|^{2}+\beta_{G}^{-2}n_{G}\right] \quad (31)$$

The conditional MSE in (26) is simplified as

$$\varepsilon\left(\tilde{F}_{G},\tilde{W}_{G},\beta_{G}\right)=\varepsilon_{G}^{(1)}+\varepsilon_{G}^{(2)} \quad (32)$$

Where

$$\varepsilon_{G}^{(1)}=E\left[\left\|x_{G}-\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right\|^{2}\right] \quad (33)$$

$$\varepsilon_{G}^{(2)}=E\left[\sum_{k'=1,k'\neq k}^{|G|}\left\|\beta_{G}^{-1}\tilde{H}_{G,k}^{H}\tilde{F}_{G,k'}\tilde{W}_{G,k'}x_{G,k'}\right\|\right]+$$

$$E\left[\sum_{i'=1,i'\neq i,}^{L}\left\|\beta_{G}^{-1}\tilde{H}_{G,}^{H}\tilde{F}_{G,}\tilde{W}_{G,}x_{G,}\right\|^{2}+\beta_{G}^{-2}n_{G}\right] \quad (34)$$

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

$$\begin{aligned}\tilde{W}_{G}'&=\left(\tilde{H}_{eq_{G}}^{H}\tilde{H}_{eq_{G}}+\gamma_{G}^{-1}I_{G}\right)^{-1}\tilde{H}_{eq_{G}}^{H}\\&=\left(\tilde{F}_{G}^{H}\tilde{H}_{G}\tilde{H}_{G}^{H}\tilde{F}_{G}+\gamma_{G}^{-1}I_{G}\right)^{-1}\tilde{F}_{G}^{H}\tilde{H}_{G}\end{aligned} \quad (35)$$

Where γ_{G}^{-1} is regularization factor, which depends on noise variance and base station transmit power. I_{G} can be expressed as

$$I_{G}=\sum_{k'=1,k'\neq k}^{|G|}H_{G,k}^{H}F_{G,k'}W_{G,k'}x_{G,k'}+\sum_{i'=1,i'\neq i,}^{L}\tilde{H}_{G,}^{H}\tilde{F}_{G,}\tilde{W}_{G,}x_{G,}+n_{G} \quad (36)$$

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

$$\beta_{G}=\sqrt{\frac{P_{tol}}{\sum_{i=1}^{L}tr\left(\tilde{F}_{G}\tilde{W}_{G}\tilde{W}_{G}^{H}\tilde{F}_{G}^{H}\right)}} \quad (37)$$

Accordingly, equation (33) can be expressed as

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

$$=E\left\{\mathrm{tr}\left[\left(x_{G}-\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right)^{H}\left(x_{G}-\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right)\right]\right\}$$

$$=E\left\{\mathrm{tr}\left[\left(x_{G}^{H}-\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)^{H}\right)\left(x_{G}-\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right)\right]\right\}$$

$$=E\left\{\mathrm{tr}\left(x_{G}^{H}x_{G}\right)\right\}-E\left\{\mathrm{tr}\left[\beta_{G}^{-1}x_{G}^{H}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right]\right\}-$$

$$E\left\{\mathrm{tr}\left[\beta_{G}^{-1}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)^{H}x_{G}\right]\right\}+E\left\{\mathrm{tr}\left[\beta_{G}^{-2}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)^{H}\left(\tilde{H}_{G}^{H}\tilde{F}_{G}\tilde{W}_{G}x_{G}\right)\right]\right\} \quad (38)$$

After simple mathematical derivation, equation (34) can be expressed as

$$\varepsilon_{G}^{(2)}=E\left[\sum_{k'=1,k'\neq k}^{|G|}\left\|\beta_{G}^{-1}\tilde{H}_{G,k}^{H}\tilde{F}_{G,k'}\tilde{W}_{G,k'}x_{G,k'}\right\|\right]+$$

$$\sum_{i'=1,i'\neq i}^{L}E\left[\left\|\beta_{G}^{-1}\tilde{H}_{G}^{H}\tilde{F}_{G,}\tilde{W}_{G,}x_{G,}\right\|_{F}^{2}+n_{G}\right]$$

$$=\sum_{k'=1,k'\neq k}^{|G|}\beta_{G}^{-2}tr\left[\tilde{H}_{G,k}^{H}\tilde{F}_{G,k'}\tilde{W}_{G,k'}E\left(x_{G,k'}x_{G,k'}^{H}\right)\tilde{W}_{G,k'}^{H}\tilde{F}_{G,k'}^{H}\tilde{H}_{G,k}\right]+$$

$$\sum_{i'=1,i'\neq i}^{L}\beta_{G}^{-2}\,\mathrm{tr}\left[\tilde{H}_{G,}^{H}\tilde{F}_{G,}\tilde{W}_{G,}E\left(x_{G,}x_{G,}^{H}\right)\tilde{W}_{G,}^{H}\tilde{F}_{G,}^{H}\tilde{H}_{G,}\right]+\beta_{G}^{-2}g_{G}\sigma^{2}$$

$$=\sum_{k'=1,k'\neq k}^{|G|}\beta_{G}^{-2}g_{i}\,\mathrm{tr}\left(\tilde{H}_{G,k}^{H}\tilde{F}_{G,k'}\tilde{W}_{G,k'}\tilde{W}_{G,k'}^{H}\tilde{F}_{G,k'}^{H}\tilde{H}_{G,k}\right)+$$

$$\sum_{i'=1,i'\neq i}^{L}\beta_{G}^{-2}g_{i'}\,\mathrm{tr}\left(\tilde{H}_{G,}^{H}\tilde{F}_{G,}\tilde{W}_{G,}\tilde{W}_{G,}^{H}\tilde{F}_{G,}^{H}\tilde{H}_{G,}\right)+\beta_{G}^{-2}g_{G}\sigma_{n}^{2} \quad (39)$$

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

$$\arg\min_{\tilde{F}_{G}}J\left(\tilde{F}_{G}\right)=\sum_{i=1}^{L}\varepsilon_{G}^{(1)}+\sum_{i=1}^{L}\varepsilon_{G}^{(2)}\ \text{ s.t. }\tilde{F}_{G}^{H}\tilde{F}_{G}=I_{n_{t}} \quad (40)$$

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 $J\left(\tilde{F}_{G}\right)$ can be expressed as

$$\nabla J\left(\tilde{F}_{G}\right)=\frac{\partial J\left(\tilde{F}_{G}\right)}{\partial\tilde{F}_{G}^{*}} \quad (41)$$

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

$$Z_{G,t+1}=-\nabla J\left(\tilde{F}_{G,t+1}\right)+\Gamma_{G,t}Z_{G,t} \quad (42)$$

where

$$\Gamma_{G,t}=\frac{\left\|\nabla J\left(\tilde{F}_{G,t+1}\right)\right\|_{F}^{2}}{\left\|\nabla J\left(\tilde{F}_{G,t}\right)\right\|_{F}^{2}} \quad (43)$$

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

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

Step 2: If $\left\|\nabla J\left(\tilde{F}_{G,1}\right)\right\|\leq\varepsilon$, stop; Otherwise, search $\Upsilon_{G,t}$ satisfying $J\left(\tilde{F}_{G,t}+\Upsilon_{G,t}Z_{G,t}\right)=\min_{\Upsilon_{G,t}\geq 0}J\left(\tilde{F}_{G,t}+\Upsilon_{G,t}Z_{G,t}\right)$;

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

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

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

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

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

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

2) Multi-cell scenario

In order to adapt to different scenarios and requirements, the hybrid precoding matrix can be determined by joint processing (JCP) and PCP.

The received signal for the bth cell is given as

$$\tilde{y}_b=\underbrace{\tilde{H}_b^H\tilde{F}_b\tilde{W}_b x_b}_{\text{Cell signal}}+\underbrace{\sum_{b'=1,b'\neq b}^{B}\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}x_{b'}}_{\text{Inter-cell interference}}+n_b \quad (44)$$

Where $\tilde{y}_b=\left[\tilde{y}_{b,G_1}^T,\tilde{y}_{b,G_1}^T,\cdots\tilde{y}_{b,G_L}^T\right]^T$ represents the received signal, $\tilde{H}_b=\left[\tilde{H}_{b,G_1},\cdots,\tilde{H}_{b,G_1}\right]$ represents the channel matrix for the bth cell, $\tilde{F}_b=\left[\tilde{F}_{b,G_1},\tilde{F}_{b,G_2},\cdots,\tilde{F}_{b,G_1}\right]$ and $\tilde{W}_b=diag\left(\tilde{W}_{b,G_1},\ldots,\tilde{W}_{b,G_1}\right)$ represent analog precoding matrix and digital precoding matrix respectively. Thus, the estimation of the received signal in the bth cell is given as

$$x_b=\beta_b^{-1}\left(\tilde{H}_b^H\tilde{F}_b\tilde{W}_b x_b+\sum_{b'=1,b'\neq b}^{B}\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}x_{b'}+n_b\right) \quad (45)$$

Where β_b^{-1} is the bth cell scaling factor that is jointly optimized with the hybrid precoding. The design of precoding is jointly derived across all L user clusters.

$$\mathrm{E}_b\left(\tilde{F}_b,\tilde{W}_b,\beta_b\right)\triangleq E\left[\left\|x_b-\hat{x}_b\right\|^2\right]$$
$$=E\left[\left\|x_b-\beta_b^{-1}\left(\tilde{H}_b^H\tilde{F}_b\tilde{W}_b x_b+\sum_{b'=1,b'\neq b}^{B}\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}x_{b'}+n_b\right)\right\|^2\right] \quad (46)$$

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., $\tilde{W}_b=\tilde{W}_b^{(1)}\tilde{W}_b^{(2)}$. $\tilde{W}_b^{(1)}$ and $\tilde{W}_b^{(2)}$ represent the first and the second precoding matrix of the bth cell respectively. Therefore, equation (44) can be rewritten as

$$\tilde{y}_b=\underbrace{\tilde{H}_b^H\tilde{F}_b\tilde{W}_b^{(1)}\tilde{W}_b^{(2)}x_b}_{\text{Cell signal}}+\underbrace{\sum_{b'=1,b'\neq b}^{B}\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}^{(1)}\tilde{W}_{b'}^{(2)}x_{b'}}_{\text{Inter-cell interference}}+n_b \quad (47)$$

Where $\tilde{F}_b$ represents the analog precoding matrix of the bth cell. To obtain (47), the signal space of $\tilde{W}_b^{(1)}$ is mapped to the channel null space of all remaining user groups b', namely:

$$\tilde{W}_b^{(1)}\subset Span^{\perp}\{\tilde{U}_{b'}^{0}(b'\in B)\} \quad (48)$$

Where $\tilde{U}_b^{0}$ is a matrix comprising dominant eigenvectors corresponding to the $r_b^*<r_b$ dominant eigenvalues of R_b. $R_b=E[\tilde{H}_b^{0}\tilde{H}_b^{0H}]$ is the channel covariance matrix of the bth cell. $\tilde{U}_{b'}^{0}$ is the dominant eigenspaces of the corresponding channel covariance matrix $R_{b'}$. $R_{b'}$ is the covariance matrix of the inter-cell channels. The idea of formula (48) 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 (48) based on the approach of block diagonalization [30], we define a matrix of eigenmodes of equivalent interference channel covariance for the bth cell as follows:

$$\Xi_b=[\tilde{U}_1^{0},\tilde{U}_2^{0},\ldots,\tilde{U}_{b-1}^{0},\tilde{U}_{b+1}^{0},\ldots,\tilde{U}_B^{0}] \quad (49)$$

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

$$\tilde{R}_b^{0}=(\Phi_b^{(0)})^H\tilde{U}_b^{0}\Lambda_b\tilde{U}_b^{0H}\Phi_b^{(0)} \quad (50)$$

Then, the SVD of (28) is carried out. Let $\tilde{U}_b^{0}$ contains the dominant r_b^* eigenmodes of R_b. The first precoding matrix $\tilde{W}_b^{(1)}$ is given by:

$$\tilde{W}_b^{(1)}=\Phi_b^{(0)}\tilde{U}_b \quad (51)$$

$\tilde{W}_b^{(2)\prime}$ can be obtained by KKT conditions as

$$\tilde{W}_b^{(2)\prime}=\tilde{H}_{eq}^H\left(\tilde{H}_{eq}\tilde{H}_{eq}^H+\gamma_b^{-1}I_b\right)^{-1}$$
$$=\tilde{F}_b^{\mathrm{H}}\tilde{H}_{b,b}^{\mathrm{H}}\left(\tilde{H}_{b,b}^{\mathrm{H}}\tilde{F}_b\tilde{F}_b^{\mathrm{H}}\tilde{H}_{b,b}+\gamma_b^{-1}I_b\right)^{-1} \quad (52)$$

Where γ_b^{-1} is regularization factor, which depends on noise variance and base station transmit power. I_b can be expressed as

$$I_b=\sum_{b'=1,b'\neq b}^{B}\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}x_{b'}+n_b \quad (53)$$

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

$$J\left(\tilde{F}_b\right)=\min_{\tilde{F}_b^{(2)}}\sum_{b\in B}\mathrm{E}_b\left(\tilde{F}_b\right)\ \text{s.t.}\ \tilde{F}_b^{\mathrm{H}}\tilde{F}_b=I_{n_t} \quad (54)$$

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

$$\nabla J\left(\tilde{F}_b\right)=\frac{\partial J\left(\tilde{F}_b\right)}{\partial \tilde{F}_b^*} \qquad (55)$$

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

$$Z_{b,t+1}=-\nabla J\left(\tilde{F}_{b,t+1}\right)+\Gamma_{b,t}Z_{b,t} \qquad (56)$$

Where

$$\Gamma_{b,t}=\frac{\left\|\nabla J\left(\tilde{F}_{b,t+1}\right)\right\|_F^2}{\left\|\nabla J\left(\tilde{F}_{b,t}\right)\right\|_F^2} \qquad (57)$$

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

Step 1: Initialize the analog precoding matrix $\tilde{F}_{b,1}=\left[\tilde{F}_{\mathrm{G}_1,1},\cdots,\tilde{F}_{\mathrm{G}_L,1}\right]$, error threshold $\varepsilon\in(0,1)$, the initial gradient $Z_{b,1}=\left[Z_{\mathrm{G}_1,1},\cdots,Z_{\mathrm{G}_L,1}\right]$,where $Z_{\mathrm{G}_1,1}=-\nabla J\left(\tilde{F}_{\mathrm{G}_1,1}\right)$, the number of initialization iterations $t=1$;

Step 2: If $\left\|\nabla J\left(\tilde{F}_{\mathrm{G}_1,1}\right)\right\|\le\varepsilon$, $i=1,\cdots,L$, stop; otherwise, search $\Upsilon_{b,t}=\left[\Upsilon_{\mathrm{G}_1,t},\cdots,\Upsilon_{\mathrm{G}_L,t}\right]$ satisfying $J\left(\tilde{F}_{\mathrm{G}_i,t}+\Upsilon_{\mathrm{G}_i,t}Z_{\mathrm{G}_i,t}\right)=\min_{\Upsilon_{\mathrm{G}_i,t}\ge 0}J\left(\tilde{F}_{\mathrm{G}_i,t}+\Upsilon_{\mathrm{G}_i,t}Z_{\mathrm{G}_i,t}\right),i=1,\cdots,L$;

Step 3: Update the analog precoding matrix $\tilde{F}_{b,t+1}$ using (43) $\tilde{F}_{\mathrm{G}_i,t+1}=\tilde{F}_{\mathrm{G}_i,t}+\Upsilon_{\mathrm{G}_i,t}Z_{\mathrm{G}_i,t}$, $i=1,\cdots,L$;

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

Step 5: Update $Z_{\mathrm{G}_i,t+1}=-\nabla J\left(\tilde{F}_{\mathrm{G}_i,t+1}\right)+\Gamma_{\mathrm{G}_i,t}Z_{\mathrm{G}_i,t}$, $i=1,\cdots,L$, where

$\Gamma_{\mathrm{G}_i}=\frac{\left\|\nabla J\left(\tilde{F}_{\mathrm{G}_i,t+1}\right)\right\|_F^2}{\left\|\nabla J\left(\tilde{F}_{\mathrm{G}_i,t+1}\right)\right\|_F^2}$, and $\Gamma_{b,t}=\left[\Gamma_{\mathrm{G}_1,t},\cdots,\Gamma_{\mathrm{G}_L,t}\right]$; Update the number of iterations $t=t+1$, repeat 2;

Step 6: Update $\tilde{F}_{\mathrm{G}_i,1}=\tilde{F}_{\mathrm{G}_i,t+1}$, $Z_{\mathrm{G}_i,1}=-\nabla J\left(\tilde{F}_{\mathrm{G}_i,1}\right)$, $t=1,i=1,\cdots,L$, repeat step 2.

Update the analog precoding matrix until convergence meets 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 G_i in the bth cell is given by:

$$SINR_{\mathrm{G}_i}=\frac{\left|\tilde{H}_{\mathrm{G}_i}^H\tilde{F}_{\mathrm{G}_i}\tilde{W}_{\mathrm{G}_i}\right|^2P_{\mathrm{G}_i}}{\sum_{k'=1,k'\ne k}^{|\mathrm{G}_i|}\left|IN_{\mathrm{G}_i,k'}\right|^2P_{\mathrm{G}_i,k'}+\sum_{i'=1,i'\ne i}^{L}\left|IN_{\mathrm{G}_{i'}}\right|^2P_{\mathrm{G}_{i'}}+\sigma_{\mathrm{G}_i}^2}$$

(58)

Where $IN_{\mathrm{G}_i,k'}=\tilde{H}_{\mathrm{G}_i,k}^H\tilde{F}_{\mathrm{G}_i,k'}\tilde{W}_{\mathrm{G}_i,k'}$, $IN_{\mathrm{G}_{i'}}=\tilde{H}_{\mathrm{G}_{i'}}^H\tilde{F}_{\mathrm{G}_{i'}}\tilde{W}_{\mathrm{G}_{i'}}$, P_{G_i} are the transmit power of the G_ith cluster, $P_{\mathrm{G}_i,k'}$ and $P_{\mathrm{G}_{i'}}$ are the transmit power of the k'th user in the ith cluster and the transmit power of the $\mathrm{G}_{i'}$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. Therefore, the SINR for a bth cell from inter-cell interference is expressed as:

$$SINR_b=\frac{\left|\tilde{H}_b^H\tilde{F}_b\tilde{W}_b^{(1)}\tilde{W}_b^{(2)}\right|^2P_b}{\sum_{b'=1,b'\ne b}^{B}\left|\tilde{H}_{b'}^H\tilde{F}_{b'}\tilde{W}_{b'}^{(1)}\tilde{W}_{b'}^{(2)}\right|^2P_{b'}+\sigma_b^2} \qquad (59)$$

The capacity of mmWave massive MIMO system can be expressed as

$$SUM=\sum_{\mathrm{G}_i=\mathrm{G}_1}^{\mathrm{G}_L}\log_2(1+SINR_{\mathrm{G}_i})+\sum_{b=1}^{B}\log_2(1+SINR_b)$$

$$=\sum_{\mathrm{G}_i=\mathrm{G}_1}^{\mathrm{G}_L}\log_2(1+SINR_{\mathrm{G}_i})+\sum_{b=1}^{B}\log_2(1+SINR_b) \qquad (60)$$

(60) can be written as

$$SUM=\sum_{\mathrm{G}_i=\mathrm{G}_1}^{\mathrm{G}_L}\log_2\left(1+\frac{\left|\hat{H}_{\mathrm{G}_i}^H\hat{F}_{\mathrm{G}_i}\hat{W}_{\mathrm{G}_i}\right|^2P_{\mathrm{G}_i}}{\sum_{k'=1,k'\ne k}^{|\mathrm{G}_i|}\left|IN_{\mathrm{G}_i,k'}\right|^2P_{\mathrm{G}_i,k'}+\sum_{i'=1,i'\ne i}^{L}\left|IN_{\mathrm{G}_{i'}}\right|^2P_{\mathrm{G}_{i'}}+\sigma_{\mathrm{G}_i}^2}\right)$$

$$+\sum_{b=1}^{B}\log_2\left(1+\frac{\left|\hat{H}_b^H\hat{F}_b\hat{W}_b^{(1)}\hat{W}_b^{(2)}\right|^2P_b}{\sum_{b'=1,b'\ne b}^{B}\left|\hat{H}_{b'}^H\hat{F}_{b'}\hat{W}_{b'}^{(1)}\hat{W}_{b'}^{(2)}\right|^2P_{b'}+\sigma_b^2}\right) \qquad (61)$$

V. 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 Δ=8. The complex gain of each path follows the distribution $\mathcal{CN}(0,1)$. The uniform linear array (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 in single-cell scenario. 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 in single-cell scenario. The achievable sum-rate of mmWave massive MIMO systems is improved by user clustering hybrid precoding.

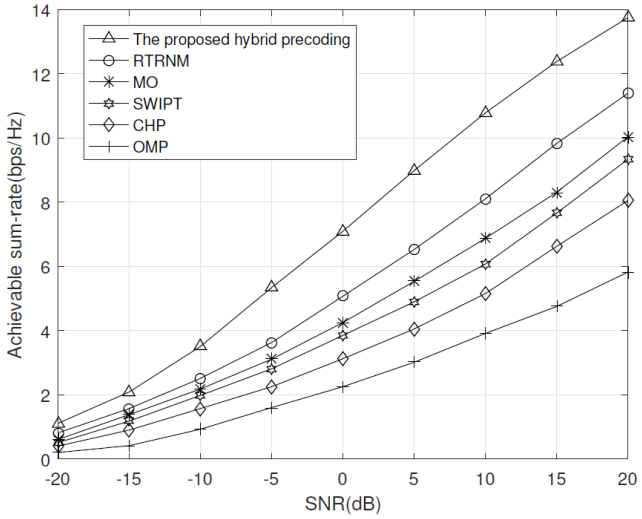


FIGURE 4. Achievable sum-rate comparison of different hybrid precoding in single-cell scenario.

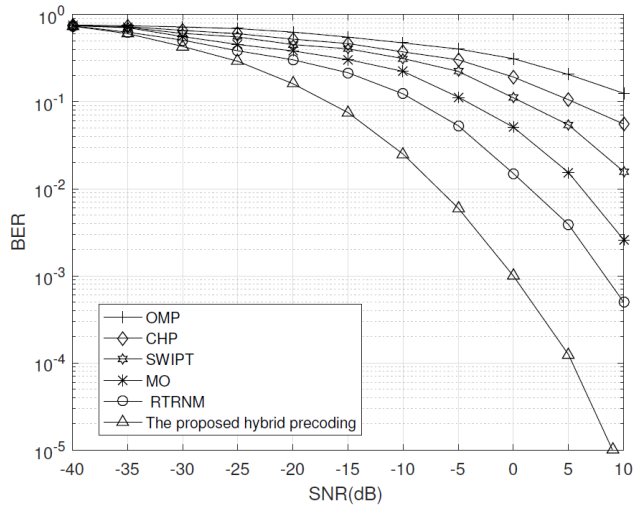


FIGURE 5. BER performance comparison of different hybrid precoding in single-cell scenario

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.

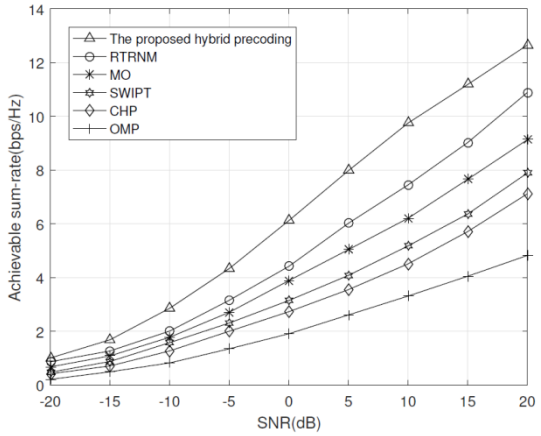


FIGURE 6. Achievable sum-rate comparison of different hybrid precoding in multi-cell scenario.

In Fig. 6, we set $N_t = 256$, $N_{RF} = 64$, and $K = 64$ in multi-cell scenario. Fig. 6, clearly shows that the proposed solution can achieve a considerably higher sum-rate than other existing hybrid precodings. This is because most of the high-dimensional channels are embedded in the low-dimensional manifolds by manifold discriminative learning in multi-cell scenario, 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. The proposed scheme can better eliminate intra-cluster, inter-cluster and inter-cell interferences. We propose user clustering hybrid precoding to enable efficient operation in high-dimensional mmWave massive MIMO, where a large number of antennas are used in low-dimensional manifolds**.**

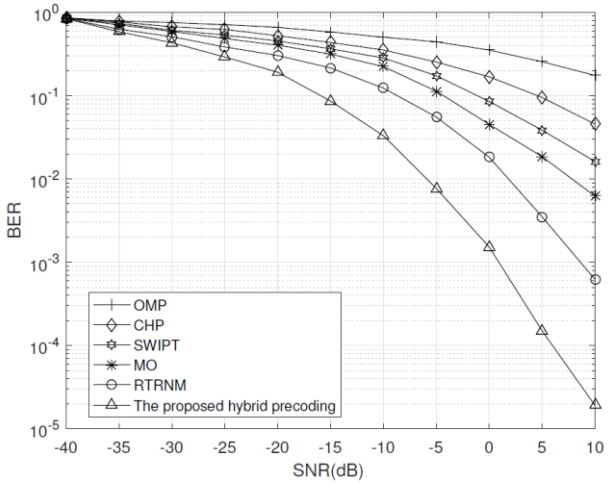


FIGURE.7.BER performance comparison of different hybrid precoding schemes in multi-cell scenario

In Fig. 7, we compare the BER performance of different hybrid precoder schemes in multi-cell scenario, where the same channel parameters as considered in Fig. 6 are used. From Fig. 7, similar conclusions to those observed for Fig.4 can be obtained in multi-cell scenario. In low-dimensional manifolds, the intra-cell channels become more clustered and the separability of embedded features is enhanced. The proposed scheme not only reduces the computational complexity in mmWave massive MIMO system, but also performs well in inter-user interference.

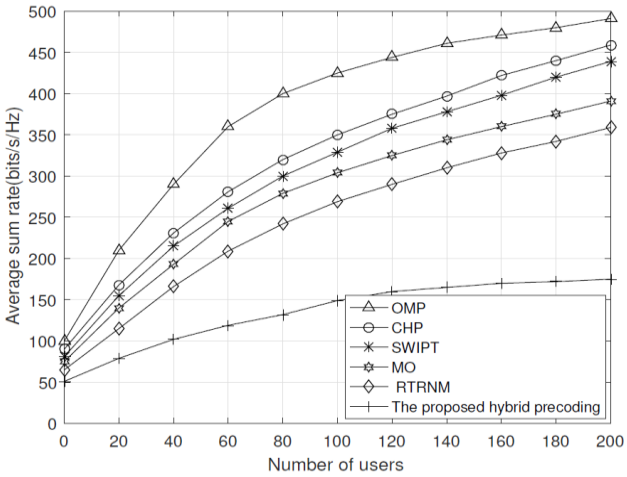


FIGURE 8. Average sum rate of two-tier system with different precoding

As shown in Fig. 8, we compare the average sum rate for the proposed scheme, and other existing precodings with different numbers of users in multi-cell scenario. We set $n_{RF,i} \geq g_i$ in each cluster. It is observed from Fig. 8 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.

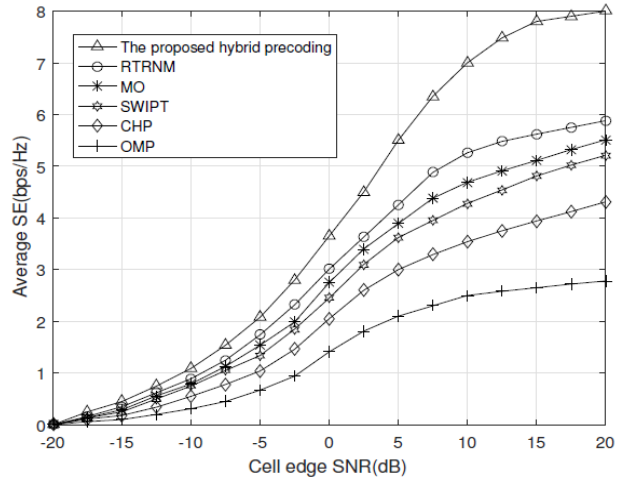


FIGURE .9. Average SE. versus cell edge SNR.

Fig. 9 shows the effect of SNR on the system average SE is given with increasing cell edge SNR. Fig. 10 presents the average SE for different numbers of BS antennas . It can be observed that the proposed scheme provides a significantly higher average SE than other existing schemes in multi-cell scenario. From Fig. 9, 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. From Fig. 10, the proposed scheme enables efficient and where a large number of antennas are used in multiple low-dimensional manifolds.

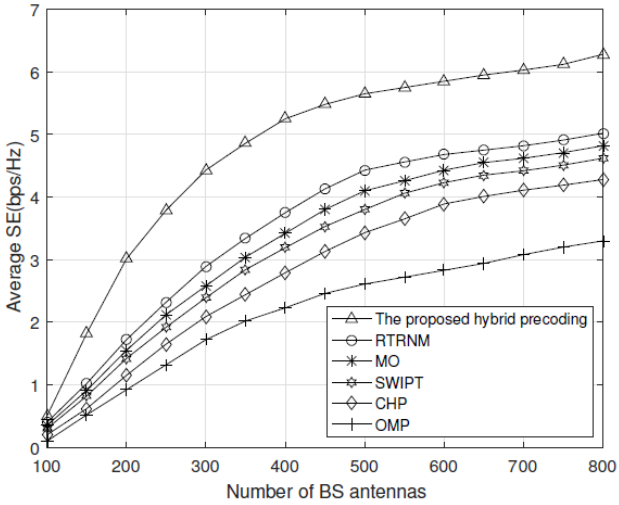


FIGURE 10. Average SE. versus number of BS antennas, high SNR.

VI. CONCLUSION AND FUTURE WORK

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.

In the future, the time correlation of the channel is considered in future work. The proposed scheme studies uniformly distributed users, while the actual user distribution has a

uniformly distributed scenario. These questions will be further studied by us.

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 Com- mun.*, 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)*, pp. 1-6. USA: Washington, Dec. 2016.
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. El 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)*, pp. 1-6, USA: Washington, Dec. 2016.
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, 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，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 Li 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, 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, 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, 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, D. E. Vittorio, "Millimeter-wve popagation: caracterization 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, 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 , B. Shim, "Principal component analysis based broadband hybrid precoding for millimeter-wave Massive MIMO systems," *IEEE Trans. on Wire. Commun.*, pp.1-1, June 2020.
35. Y. Li , G. Cao, 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, 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, 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. H. Cheng, N. Xiong, A. V. Vasilakos, "Nodes organization for channel assignment with topology preservation in multi-radio wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 5, pp. 760-773, July 2012.
39. H. Cheng, N. Xiong, G. Chen, "Channel Assignment with Topology Preservation for Multi-radio Wireless Mesh Networks," *Journal of Commun.* vol. 10, no. 5, pp. 63-70, Jan. 2010.
40. P. Juan, M. José-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.
41. A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
42. Dan Meng, Guitao Cao, Wenming 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.