


**Please cite the Published Version**

Han, Meng, Du, Jianhe, Zhang, Yang, Li, Xingwang, Rabie, Khaled  and Nauryzbayev, Galymzhan (2021) Efficient Hybrid Beamforming Design in mmWave Massive MU-MIMO DF Relay Systems with the Mixed-Structure. IEEE Access, 9. pp. 66141-66153. ISSN 2169-3536

**DOI:** <https://doi.org/10.1109/ACCESS.2021.3073847>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Version:** Published Version

**Downloaded from:** <https://e-space.mmu.ac.uk/635048/>

**Usage rights:**  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

**Additional Information:** This is an open access article which first appeared in IEEE Access

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Received April 3, 2021, accepted April 11, 2021, date of publication April 16, 2021, date of current version May 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073847

# Efficient Hybrid Beamforming Design in mmWave Massive MU-MIMO DF Relay Systems With the Mixed-Structure

MENG HAN<sup>1</sup>, (Student Member, IEEE), JIANHE DU<sup>1</sup>, (Member, IEEE),  
YANG ZHANG<sup>1</sup>, (Student Member, IEEE), XINGWANG LI<sup>2</sup>, (Senior Member, IEEE),  
KHALED M. RABIE<sup>3</sup>, (Senior Member, IEEE),  
AND GALYMZHAN NAURYZBAYEV<sup>4</sup>, (Senior Member, IEEE)

<sup>1</sup>School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

<sup>2</sup>School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

<sup>3</sup>Department of Engineering, Manchester Metropolitan University, Manchester M1 5GD, U.K.

<sup>4</sup>School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan 010000, Kazakhstan

Corresponding authors: Jianhe Du (dujianhe1@gmail.com) and Yang Zhang (zhynkt2017@cuc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601414, Grant 61701448, and Grant 61702466; in part by the National Key Research and Development Program of China under Grant 2016YFB0502001; and in part by the Fundamental Research Fund for the Central Universities under Grant CUC200A011.

**ABSTRACT** In this paper, we consider the decode-and-forward (DF) relay system in millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems, and propose a hybrid beamforming design method for the mixed structure, which contains the fully-connected and sub-connected structures. To satisfy constant-modulus and block-diagonalization (BD) constraints, the analog beamforming is designed by the idea of sorted successive interference cancellation (SSIC). More specifically, the proposed method first sorts the capacities of different analog sub-channels in descending order, and then designs the analog beamforming serially according to the order of the capacities. To efficiently mitigate the inner-user and inter-user interference, we propose a modified baseband BD technology to reduce the information loss in digital beamforming design, thereby improving the system capacity. In addition, the proposed hybrid beamforming algorithm is designed by considering both uniform linear arrays (ULAs) and uniform planar arrays (UPAs). Simulation results demonstrate that the proposed hybrid beamforming design scheme can obtain good performance in terms of the achievable sum-rate and power efficiency in ULAs and UPAs.

**INDEX TERMS** DF relay, mmWave, massive MIMO, hybrid beamforming, mixed structure.

## I. INTRODUCTION

5G seeks to meet communication needs in the scenarios with large connection, high bandwidth and low latency [1]. Compared with 5G, 6G will further improve information transmission rate, signal coverage, delay and intelligence [2]. Millimeter wave (mmWave) covers the spectrum from 30 GHz to 300 GHz, and meets the bandwidth requirements for 5G/6G services [3]. However, mmWave signals are easily blocked, absorbed and scattered by obstacles during transmission, which leads to severe propagation path losses in high frequency communications. Due to the shorter wavelength of mmWave, a large number of antennas are allowed to be

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Khawaja<sup>1</sup>.

equipped at transmitting and receiving ends, i.e., massive multiple-input multiple-output (MIMO) techniques.

In the face of complex environments or long-distance communications, relays can be used in mmWave massive MIMO systems to assist the communication between source and destination nodes. With the help of relays, the number of transmission signals in the coverage area will be expanded [4]. It can also be ensured that channels between any two communication nodes are in line of sight (LoS). MmWave massive MIMO relay systems can reduce propagation path losses and severe intermittent blocking effects. There are two commonly used relays, i.e., amplify-and-forward (AF) and decode-and-forward (DF) relays. AF relays amplify the power of the received signal and forward it. DF relays decode the received signal and then re-encode and forward it, which exhibit the

digital nature. In contrast to AF relays, the signal processing method of DF relays is complicated. However, DF relays can overcome the noise accumulation by regenerating data at relays and increase the possibility of adaptive modulation and coding [5]–[7].

The precoding technique can increase the power gain required for transmission, which helps to compensate for propagation path losses of mmWave signals in wireless channels. It is also conducive to reduce severe signal attenuation caused by atmospheric absorption and rainfall. In general, the full-digital (FD) precoding scheme is optimal in the matter of flexibility and performance [8]–[10]. However, the FD precoding architecture asks a dedicated radio frequency (RF) chain to be assigned for each antenna, and thus is more complex in hardware implementation. It is impractical for massive MIMO antenna arrays from the perspective of cost and power consumption. In order to leverage the hardware complexity and system performance, the hybrid precoding scheme falls into place. It has fewer RF chains, but its performance is close to that of the FD precoding scheme. The hybrid precoding technique has become a popular RF architecture in future mmWave massive MIMO communication systems [11]–[15].

In [11], an asymptotically optimal hybrid precoding scheme with closed-form solution was proposed for the downlink massive multi-user (MU) MIMO system. The scheme has superior performance and low complexity. Its sum-rate is close to channel capacity with massive antennas at the base station. In [12], a two-stage hybrid precoding design based on the signal-to-leakage-plus-noise ratio metric was proposed for frequency-division duplexing massive MU-MIMO systems, and then was extended to multi-cell systems. This hybrid precoding method can significantly reduce the downlink training and uplink feedback overhead. The work in [13] proposed a joint hybrid precoding scheme for large-scale MIMO systems by exploiting the concept of equivalent channel. Its system spectral efficiency is enhanced. In [14], a hybrid beamforming scheme with partial interfering beam feedback was proposed for the codebook based MU-MIMO system, and it outperforms an existing hybrid precoding scheme based on channel reconstruction. The research conducted in [15] proposed a two-stage hybrid beamforming design for the mmWave massive MU-MIMO system with sub-connected structure. The hybrid precoding scheme more accurately approximates to that of the FD system.

The hybrid precoding design can be realized by two classic structures: the sub-connected and fully-connected structure. The sub-connected structure means that each RF chain is only connected to an independent subset of antennas, while the fully-connected structure means that each RF chain is connected to all antennas. For the AF mmWave massive MIMO relay system with fully-connected structure, [16] and [17] studied the joint optimal hybrid precoding design for downlink single-user (SU) and MU scenarios. The works [18] and [19] focused on the AF mmWave massive MIMO relay system with sub-connected structure in the SU scenario.

The research in [20] investigated the hybrid beamforming for multi-hop AF relay systems and channel errors were also taken into account.

In [21], a mixed connected structure, which contains the fully-connected and sub-connected structures, was proposed and a matrix factorization based near-optimal hybrid precoding design was designed for mmWave massive MIMO systems. This mixed connected structure shows a lower hardware complexity in comparison with the fully-connected structure, and its spectral efficiency is better than that of the sub-connected structure. The work [22] proposed a generalized sub-array-connected (GSAC) architecture and a beamsteering codebook for the hybrid precoding aided GSAC architecture in mmWave massive MIMO systems, which improves the energy efficiency.

To the best of the authors' knowledge, there are few research works on the hybrid precoding design of mmWave massive MU-MIMO DF relay system with mixed structure, which motivates our work. For communication systems with large-scale antennas, compared with the fully-connected structure, the mixed structure exhibits advantages of low computational complexity, low power consumption, and simple wiring. At the same time, communication systems with the mixed structure can achieve great performance close to that with the fully-connected structure. In this paper, we study the hybrid beamforming design in a mmWave massive MU-MIMO DF relay system with the mixed structure, where the source node sends signals to users with the aiding of a DF relay. The main contributions of this work can be summarized as follows:

- 1) Considering the uniform linear arrays (ULAs) and uniform planar arrays (UPAs), the hybrid beamforming design of mmWave massive MU-MIMO DF relay system with mixed fully-connected and sub-connected structure is investigated in this paper. By exploiting the idea of joint hierarchical optimization, we recursively combine the channel with optimized precoders/combiners to decompose the total sum-rate optimization problem with non-convex constraints into a series of simple optimization problems.
- 2) To satisfy constant-modulus and block-diagonalization (BD) constraints, the sorted successive interference cancellation (SSIC) method is proposed to design the analog beamforming with sub-connected structure. Then the piecewise success approximation is utilized to obtain the analog beamforming with fully-connected structure. The modified BD technology is utilized to design the digital beamforming.
- 3) In the case of large-scale antenna arrays, simulation results show the proposed hybrid beamforming algorithm for the mixed structure can achieve great sum-rate performance in both ULAs and UPAs, and has the advantages of low computational complexity and low power consumption compared with the fully-connected structure. In addition, the proposed hybrid beamforming design method can support more users than other

design methods when the system sum-rate is steadily increasing.

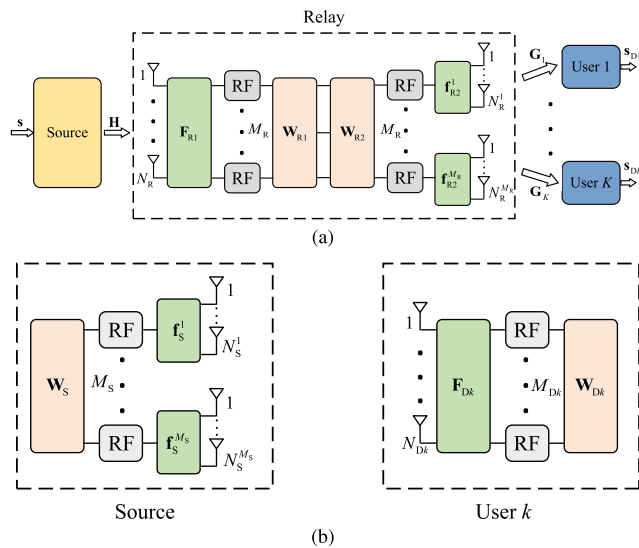
The remainder of this paper is organized as follows. In Section II, we introduce the considered massive MU-MIMO DF relay system with mixed structure and mmWave channel model. The proposed hybrid beamforming design scheme is shown in Section III. Simulation results are demonstrated in Section IV, and conclusions are drawn in Section V.

**Notations:** Bold lower-case letters and boldface capitals stand for vectors and matrices, respectively.  $E[\cdot]$  represents the expectation.  $(\cdot)^T$ ,  $(\cdot)^{-1}$ ,  $(\cdot)^H$ ,  $\text{tr}\{\cdot\}$  and  $\|\cdot\|_F$  denote the transpose, inversion, conjugate transpose, trace and Frobenius norm of a matrix, respectively.  $\mathbf{I}_N$  is an  $N \times N$  identity matrix and  $\mathbf{0}_{M \times N}$  is a  $M \times N$  all-zero matrix.  $\mathbb{C}^{m \times n}$  and  $\mathbb{Z}^{m \times n}$  represent a  $m \times n$ -dimensional complex space and an integer space, respectively.  $\mathbb{D}^{m \times m}$  describes a  $m \times m$  diagonal matrix.  $\text{blk}[\cdot]$  indicates the block-diagonal operator.  $\angle \mathbf{X}$  denotes a matrix forming by  $e^{j\phi_{i,j}}$ , where  $\phi_{i,j}$  is the phase of the  $(i, j)$ -th element of  $\mathbf{X}$ .

## II. SYSTEM DESCRIPTION

### A. SYSTEM MODEL

The block diagram of the considered mmWave massive MU-MIMO DF relay system with mixed structure is shown in Fig. 1. The relay with mixed structure is depicted in Fig. 1(a), where its receiving end adopts the fully-connected structure and transmitting end adopts the sub-connected structure. In Fig. 1(b), we show the transmitter and receiver structures of the source node and  $k$ -th user end. The source node adopts the sub-connected structure. It is equipped with  $M_S$  RF chains and each RF chain is connected to a disjoint subset of  $N_S^{ms}$  antennas, where the total number of antennas is  $N_S = \sum_{m_S=1}^{M_S} N_S^{m_S}$ . The relay receiving end is



**FIGURE 1.** The structure of the mmWave massive MU-MIMO DF relay system: (a) Relay hybrid beamforming with mixed structure; (b) Transmitter and receiver structures of the source node and  $k$ -th user, respectively.

equipped with  $M_R$  RF chains and  $N_R$  receiving antennas. The relay transmitting end is equipped with  $M_R$  RF chains. Each RF chain is connected to a disjoint subset of  $N_R^{m_R}$  antennas and the total number of antennas is  $N_R = \sum_{m_R=1}^{M_R} N_R^{m_R}$ . The

relay is assumed to serve  $K$  users with the fully-connected structure. For the  $k$ -th user,  $N_{Dk}$  antennas and  $M_{Dk}$  RF chains are installed to support  $L_S$  data streams. The total number of antennas is  $N_D = \sum_{k=1}^K N_{Dk}$ . In order to reduce the hardware implementation complexity and make sure effective multi-stream communication, the number of RF chains for all nodes are constrained by  $KL_S \leq M_S \leq N_S$ ,  $KL_S \leq M_R \leq N_R$  and  $L_S \leq M_{Dk} \leq N_{Dk}$ .

The overall symbol vector transmitted to all  $K$  users is  $\mathbf{s} = [\mathbf{s}_1^T, \dots, \mathbf{s}_K^T]$  and  $\mathbf{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{KL_S} \mathbf{I}_{KL_S}$ , in which  $\mathbf{s}_k \in \mathbb{C}^{L_S \times 1}$  and  $k = 1, 2, \dots, K$ . At the source node, the diagonal power allocation matrix  $\mathbf{P}_S = [\mathbf{P}_S^1, \dots, \mathbf{P}_S^K] \in \mathbb{D}^{KL_S \times KL_S}$  first acts on  $\mathbf{s}$  with power constraint  $\|\mathbf{P}_S\|_F^2 = P_S$ . Then, signals after power allocation are processed by a hybrid precoder which is composed of a digital precoder  $\mathbf{W}_S \in \mathbb{C}^{M_S \times KL_S}$  and an analog precoder  $\mathbf{F}_S \in \mathbb{C}^{N_S \times M_S}$ , i.e.,

$$\mathbf{x}_S = \mathbf{F}_S \mathbf{W}_S \mathbf{P}_S \mathbf{s}. \quad (1)$$

The analog precoder  $\mathbf{F}_S$  is implemented by analog phase shifters. Therefore, its non-zero elements follow constant-modulus constraints. The total transmitted power constraints at the source can be enforced by normalizing the digital precoder  $\mathbf{W}_S$  such that  $\|\mathbf{F}_S \mathbf{W}_S\|_F^2 = KL_S$ . Furthermore, since each RF chain is connected only to a subset of antennas in the sub-connected structure, the analog precoder  $\mathbf{F}_S$  at the source end is constrained as the following BD form

$$\mathbf{F}_S = \text{blk}(\mathbf{f}_S^1, \dots, \mathbf{f}_S^{M_S}), \quad (2)$$

where  $\mathbf{f}_S^{m_S} \in \mathbb{C}^{N_S^{m_S} \times 1}$  and  $m_S = 1, \dots, M_S$ . Its non-zero elements are subject to

$$|\mathbf{f}_S^{m_S}(i)| = \frac{1}{\sqrt{N_S^{m_S}}}, \quad i = 1, 2, \dots, N_S^{m_S}. \quad (3)$$

$\mathbf{x}_S$  is then transmitted via the channel  $\mathbf{H} \in \mathbb{C}^{N_R \times N_S}$  from the source to relay. Hence the received signal can be expressed as

$$\mathbf{y}_R = \mathbf{H} \mathbf{x}_S + \mathbf{n}_R, \quad (4)$$

where  $\mathbf{n}_R \in \mathbb{C}^{N_R \times 1}$  is the complex additive white Gaussian noise vector whose elements follow the independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and variance  $\sigma_R^2$ . At the relay receiving end,  $\mathbf{y}_R$  is decoded by an analog combiner  $\mathbf{F}_{R1} \in \mathbb{C}^{N_R \times M_R}$  and a digital combiner  $\mathbf{W}_{R1} \in \mathbb{C}^{M_R \times KL_S}$ . Hence the decoded signal in the DF relay can be given by

$$\begin{aligned} \mathbf{s}_R &= \mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{y}_R \\ &= \mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{H} \mathbf{x}_S + \mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{n}_R. \end{aligned} \quad (5)$$

Similar to the above treatments,  $\mathbf{s}_R$  first passes through a diagonal power allocation matrix  $\mathbf{P}_R \in \mathbb{D}^{KL_S \times KL_S}$  which satisfies  $\|\mathbf{P}_R\|_F^2 = P_R$ , and then is processed by a digital precoder  $\mathbf{W}_{R2} \in \mathbb{C}^{M_R \times KL_S}$  and an analog precoder  $\mathbf{F}_{R2} \in \mathbb{C}^{N_R \times M_R}$ , i.e.,

$$\mathbf{x}_R = \mathbf{F}_{R2} \mathbf{W}_{R2} \mathbf{P}_R \mathbf{s}_R. \quad (6)$$

The total transmitted power constraints at the relay are enforced by normalizing the digital precoders  $\mathbf{W}_{R2}$  such that  $\|\mathbf{F}_{R2} \mathbf{W}_{R2}\|_F^2 = KL_S$ . Since the relay transmitting end adopts the sub-connected structure, the analog precoders  $\mathbf{F}_{R2}$  is also in the BD form

$$\mathbf{F}_{R2} = \text{blk}(\mathbf{f}_{R2}^1, \dots, \mathbf{f}_{R2}^{M_R}), \quad (7)$$

$$|\mathbf{f}_{R2}^{m_R}(j)| = \frac{1}{\sqrt{N_R^{m_R}}}, \quad j = 1, 2, \dots, N_R^{m_R}, \quad (8)$$

where  $\mathbf{f}_{R2}^{m_R} \in \mathbb{C}^{N_R^{m_R} \times 1}$  and  $m_R = 1, \dots, M_R$ . After transmitting via the channel  $\mathbf{G}_k \in \mathbb{C}^{N_{Dk} \times N_{Dk}}$  from relay to the  $k$ -th user, the received signal can be expressed as

$$\mathbf{y}_{Dk} = \mathbf{G}_k \mathbf{x}_R + \mathbf{n}_{Dk}, \quad (9)$$

where  $\mathbf{n}_{Dk} \sim \mathcal{CN}(0, \sigma_D^2)$ . At the  $k$ -th user end,  $\mathbf{y}_{Dk}$  is processed by an analog combiner  $\mathbf{F}_{Dk} \in \mathbb{C}^{N_{Dk} \times M_{Dk}}$  and a digital combiner  $\mathbf{W}_{Dk} \in \mathbb{C}^{M_{Dk} \times L_S}$ . The resulting signal can be shown as

$$\begin{aligned} \mathbf{s}_{Dk} &= \mathbf{W}_{Dk}^H \mathbf{F}_{Dk}^H \mathbf{y}_{Dk} \\ &= \mathbf{W}_{Dk}^H \mathbf{F}_{Dk}^H \mathbf{G}_k \mathbf{x}_R + \mathbf{W}_{Dk}^H \mathbf{F}_{Dk}^H \mathbf{n}_{Dk}. \end{aligned} \quad (10)$$

By defining  $\bar{\mathbf{G}}_k = \mathbf{F}_{Dk}^H \mathbf{G}_k \mathbf{F}_{R2}$ , the received  $i$ -th data stream  $\mathbf{s}_{Dki}$  for the  $k$ -th user can be expressed as

$$\begin{aligned} \mathbf{s}_{Dki} &= \underbrace{\mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, k_i) \sqrt{P_{Rk_i}} \mathbf{s}_{Rk_i}}_{\text{desired signal}} \\ &+ \underbrace{\sum_{j=1, j \neq i}^{L_S} \mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, k_j) \sqrt{P_{Rk_j}} \mathbf{s}_{Rk_j}}_{\text{inner-user interference}} \\ &+ \underbrace{\sum_{m=1, m \neq k}^K \sum_{l=1}^{L_S} \mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, m_l) \sqrt{P_{Rm_l}} \mathbf{s}_{Rm_l}}_{\text{inter-user interference}} \\ &+ \underbrace{\mathbf{W}_{Dk}^H(i, :) \mathbf{F}_{Dk}^H \mathbf{n}_{Dk}}_{\text{noise}}, \end{aligned} \quad (11)$$

where  $k_i = (k-1)L_S + i$ ,  $i = 1, \dots, L_S$ , and  $\sqrt{P_{Rk_i}}$  is the power allocated to the  $i$ -th data stream for the  $k$ -th user. When Gaussian symbols are transmitted in the considered system, the achievable sum-rate from the relay to the  $k$ -th user can be given by

$$R = \sum_{k=1}^K \sum_{i=1}^{L_S} \log_2(1 + \text{SINR}_{k_i}), \quad (12)$$

where  $\text{SINR}_{k_i}$  is the signal-to-interference and noise ratio (SINR) of  $\mathbf{s}_{Dki}$ . This can be computed by the ratio of the

desired signal energy to the interference plus noise energy, and is formulated as

$$\begin{aligned} \text{SINR}_{k_i} &= \frac{S_{Dk_i}}{I_{Dk_i} + N_{Dk_i}}, \\ S_{Dk_i} &= \left| \mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, k_i) \sqrt{P_{Rk_i}} \right|^2, \\ I_{Dk_i} &= \sum_{j=1, j \neq i}^{L_S} \left| \mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, k_j) \sqrt{P_{Rk_j}} \right|^2 \\ &+ \sum_{m=1, m \neq k}^K \sum_{l=1}^{L_S} \left| \mathbf{W}_{Dk}^H(i, :) \bar{\mathbf{G}}_k \mathbf{W}_{R2}(:, m_l) \sqrt{P_{Rm_l}} \right|^2, \\ N_{Dk_i} &= \sigma_D^2 \left\| \mathbf{W}_{Dk}^H(i, :) \mathbf{F}_{Dk}^H \right\|_F^2, \end{aligned} \quad (13)$$

where  $k = 1, \dots, K$  and  $i = 1, \dots, L_S$ .

In order to achieve precoding, perfect channel state information (CSI) is assumed to be known for all nodes. In practical systems, CSI received at the relay and user ends can be obtained via training. Then through the limited feedback, CSI can be shared from the relay to source and the user ends to relay.

### B. CHANNEL MODEL

MmWave channels exhibit the characteristics of high free-space path losses and limited scattering or spatial selectivity [23], [24]. In addition, large tightly-packed antenna arrays are often implemented in mmWave systems, which results in high levels of antenna correlation. The traditional statistical fading distribution, such as Rayleigh fading distribution [25], is no longer suitable for modeling mmWave channels. In this paper, the narrowband clustered channel based on geometric Saleh-Valenzuela model [23], [26] is adopted to accurately reflect the mathematical structure of mmWave communications.

It is assumed that channel matrices  $\mathbf{H}$  and  $\mathbf{G}_k$  are respective a sum of  $N_{cR}$  and  $N_{ck}$  scattering clusters, each of which respectively contribute  $N_{pR}$  and  $N_{pk}$  propagation paths, where  $k = 1, \dots, K$ . Hence the normalized narrowband channels from source to relay and from relay to the  $k$ -th user can be expressed as [23]

$$\mathbf{H} = \sqrt{\frac{N_S N_R}{N_{cR} N_{pR}}} \sum_{i=1}^{N_{cR}} \sum_{l=1}^{N_{pR}} \alpha_{i,l} \mathbf{a}(\theta_{i,l}^{R1}, \varphi_{i,l}^{R1}) \mathbf{a}(\theta_{i,l}^{S}, \varphi_{i,l}^{S})^H, \quad (14)$$

$$\mathbf{G}_k = \sqrt{\frac{N_R N_{Dk}}{N_{ck} N_{pk}}} \sum_{i=1}^{N_{ck}} \sum_{l=1}^{N_{pk}} \gamma_{i,l}^k \mathbf{a}(\theta_{i,l}^{Dk}, \varphi_{i,l}^{Dk}) \mathbf{a}(\theta_{i,l}^{R2}, \varphi_{i,l}^{R2})^H, \quad (15)$$

where  $\alpha_{i,l}$  and  $\gamma_{i,l}^k$  denote complex gains of the  $i$ -th path in the  $l$ -th cluster, and they follow the independent Gaussian distribution  $\mathcal{CN}(0, 1)$ .  $\theta_{i,l}$  and  $\varphi_{i,l}$  with different superscripts are the azimuth and elevation angles of arrival/departure (AoAs/AoDs) of the  $i$ -th path in the  $l$ -th cluster, respectively. They obey the truncated Laplacian distribution [27], [28], which has been found to be a good fit for a variety of propagation scenarios.  $\mathbf{a}(\theta_{i,l}, \varphi_{i,l})$  stands for the normalized array

response vector with the azimuth and elevation angles  $\theta_{i,l}$  and  $\varphi_{i,l}$ . They are independent of antenna element properties and only subject to the antenna array structure. The hybrid precoding scheme derived in this paper can be applied to arbitrary antenna geometries. For the sake of simplicity but without loss of generality, ULAs and UPAs are examined in our study. For an ULA with  $U$  elements, the array response vector can be written as

$$\mathbf{a}_{\text{ULA}}(\theta) = \sqrt{\frac{1}{U}} [1, e^{j\zeta \sin \theta}, \dots, e^{j(U-1)\zeta \sin \theta}]^T, \quad (16)$$

where  $\zeta = 2\pi d/\lambda$ ,  $d$  indicates the spacing between two neighboring antenna elements and  $\lambda$  is the signal wavelength. Note that the elevation dimension is ignored, since the ULA response vector is invariant in the elevation domain. For an UPA with  $W_1$  and  $W_2$  elements on two arbitrary axes and  $W_1 W_2 = U$ , the array response vector can be given by

$$\mathbf{a}_{\text{UPA}}(\theta, \varphi) = \sqrt{\frac{1}{U}} [1, \dots, e^{j\zeta(w_1 \sin \theta \sin \varphi + w_2 \cos \varphi)}, \dots, e^{j\zeta((W_1-1) \sin \theta \sin \varphi + (W_2-1) \cos \varphi)}]^T, \quad (17)$$

where  $0 \leq w_1 \leq (W_1 - 1)$ ,  $0 \leq w_2 \leq (W_2 - 1)$  and  $w_1, w_2 \in \mathbb{Z}$ .

### III. HYBRID PRECODING DESIGN

This section discusses the hybrid beamforming design of the considered mmWave massive MU-MIMO DF relay system with mixed structure. The achievable sum-rate is an important performance evaluation standard for communication systems. The design goal is to maximize the sum-rate shown in (12) by properly designing structures of precoders and combiners. The optimization problem can be mathematically formulated as

$$\begin{aligned} (\mathcal{P}) \quad & (\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}, \mathbf{F}_{R2}, \mathbf{W}_{R2}, (\mathbf{F}_{Dk}, \mathbf{W}_{Dk})_{k=1,\dots,K}, \mathbf{P}_S, \mathbf{P}_R) \\ & \arg \max_{(\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}, \mathbf{F}_{R2}, \mathbf{W}_{R2}, (\mathbf{F}_{Dk}, \mathbf{W}_{Dk})_{k=1,\dots,K}, \mathbf{P}_S, \mathbf{P}_R)} R \\ \text{s.t.} \quad & |\mathbf{F}_{R1}(i, j)| = \frac{1}{\sqrt{N_R}}, \\ & |\mathbf{F}_{Dk}(i, j)| = \frac{1}{\sqrt{N_{Dk}}}, \\ & |\mathbf{f}_S^{ms}(i)| = \frac{1}{\sqrt{N_S^{ms}}}, \\ & i = 1, 2, \dots, N_S^{ms}, \\ & |\mathbf{f}_{R2}^{mr}(j)| = \frac{1}{\sqrt{N_R^{mr}}}, \\ & j = 1, 2, \dots, N_R^{mr}, \\ & \mathbf{F}_S = \text{blk}(\mathbf{f}_S^1, \dots, \mathbf{f}_S^{M_S}), \\ & \mathbf{F}_{R2} = \text{blk}(\mathbf{f}_{R2}^1, \dots, \mathbf{f}_{R2}^{M_R}), \\ & \|\mathbf{F}_S \mathbf{W}_S\|_F^2 = \|\mathbf{F}_{R2} \mathbf{W}_{R2}\|_F^2 = KL_S, \\ & \|\mathbf{P}_S\|_F^2 = P_S, \quad \|\mathbf{P}_R\|_F^2 = P_R. \end{aligned} \quad (18)$$

Both the objective function and constraints in (18), as well as the original optimization problem, are non-convex. It is difficult to jointly optimize variables  $(\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}, \mathbf{F}_{R2}, \mathbf{W}_{R2}, (\mathbf{F}_{Dk}, \mathbf{W}_{Dk})_{k=1,\dots,K}, \mathbf{P}_S, \mathbf{P}_R)$  and search the global optimal solution in problem (18). On the basis of the communication mode of DF relay systems, the entire signal transmission scheme can be decomposed into two independent cascade subsystems. The first subsystem is from source to relay, and the second one is from relay to user ends. The corresponding transmission rates are  $R_1$  and  $R_2$ , respectively. Therefore, the original sum-rate maximization problem can be reconstructed to maximize the minimum value of  $R_1$  and  $R_2$  [29]. Due to the fact that the relay communication is completed in two time slots, the transmission rate is half of the overall sum-rate compared with a relay-free scenario [30]. The sum-rate of the entire system is expressed as

$$R = 0.5 \min(R_1, R_2). \quad (19)$$

Similar to (11)-(13),  $R_1$  and  $R_2$  can be given by

$$R_1 = \log_2(1 + \text{SINR}_{\text{SR}}), \quad (20)$$

$$R_2 = \sum_{k=1}^K \sum_{i=1}^{L_S} \log_2\left(1 + \text{SINR}_{\text{RD}}^{k_i}\right), \quad (21)$$

where  $\text{SINR}_{\text{SR}}$  and  $\text{SINR}_{\text{RD}}^{k_i}$  respectively represent SINR from source to relay decoding and relay forwarding to users. The original sum-rate maximization problem (18) is then reformulated as the following two separate sub-problems

$$\begin{aligned} (\mathcal{P}_1) \quad & (\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}, \mathbf{P}_S) \\ & = \arg \max_{(\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}, \mathbf{P}_S)} R_1 \\ \text{s.t.} \quad & |\mathbf{F}_{R1}(i, j)| = \frac{1}{\sqrt{N_R}}, \\ & |\mathbf{f}_S^{ms}(i)| = \frac{1}{\sqrt{N_S^{ms}}}, \quad i = 1, 2, \dots, N_S^{ms}, \\ & \mathbf{F}_S = \text{blk}(\mathbf{f}_S^1, \dots, \mathbf{f}_S^{M_S}), \\ & \|\mathbf{F}_S \mathbf{W}_S\|_F^2 = KL_S, \\ & \|\mathbf{P}_S\|_F^2 = P_S. \end{aligned} \quad (22)$$

$$\begin{aligned} (\mathcal{P}_2) \quad & (\mathbf{F}_{R2}, \mathbf{W}_{R2}, (\mathbf{F}_{Dk}, \mathbf{W}_{Dk})_{k=1,\dots,K}, \mathbf{P}_R) \\ & = \arg \max_{(\mathbf{F}_{R2}, \mathbf{W}_{R2}, (\mathbf{F}_{Dk}, \mathbf{W}_{Dk})_{k=1,\dots,K}, \mathbf{P}_R)} R_2 \\ \text{s.t.} \quad & |\mathbf{F}_{Dk}(i, j)| = \frac{1}{\sqrt{N_{Dk}}}, \\ & |\mathbf{f}_{R2}^{mr}(j)| = \frac{1}{\sqrt{N_R^{mr}}}, \quad j = 1, 2, \dots, N_R^{mr}, \\ & \mathbf{F}_{R2} = \text{blk}(\mathbf{f}_{R2}^1, \dots, \mathbf{f}_{R2}^{M_R}), \\ & \|\mathbf{F}_{R2} \mathbf{W}_{R2}\|_F^2 = KL_S, \\ & \|\mathbf{P}_R\|_F^2 = P_R. \end{aligned} \quad (23)$$

The two reformulated sub-problems  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are still non-convex, and corresponding specific design algorithms will be presented to solve these two sub-problems.

**A. HYBRID PRECODING DESIGN FROM THE SOURCE TO RELAY DECODING**

In this subsection, we commence with the optimization of the first sub-problem  $\mathcal{P}_1$ . The communication link from the source to relay can be regarded as a point-to-point massive MIMO system. Since the source node adopts the sub-connected structure, the analog precoder  $\mathbf{F}_S$  is in the BD form with a constant amplitude. Due to these non-convex constraints, it is difficult to maximize the sum-rate in (22).

With the special BD structure of the hybrid precoding matrix  $\mathbf{F}_S$ , it is observed that the precoder for different subset of antennas is independent. Inspired by the design idea based on the successive interference cancellation (SIC) joint hierarchical optimization algorithm [20], [22], [31], the complex optimization problem  $\mathcal{P}_1$  can be decomposed into a series of sub-optimal sum-rate optimization problems. Considering that each RF chain is connected to a subset of disjoint antennas, we can first only optimize the sum-rate corresponding to the selected antenna subset while without considering others. After that, the second antenna subset is selected and its sum-rate is optimized. Repeat the above processes until the sum-rate corresponding to all antenna subsets is optimized.

The traditional SIC method is optimized in the recursive order. However, the CSI for different antenna subsets is significantly different. For our hybrid precoding design scheme, all  $M_S$  antenna subsets are first sorted according to the difference in their channel capacity. Then we perform the above mentioned SSIC optimization processes in the arrangement order till the sum-rate corresponding to the last antenna subset is optimized.

After the analog precoder  $\mathbf{F}_S$  is obtained, the analog combiner  $\mathbf{F}_{R1}$ , digital combiner  $\mathbf{W}_{R1}$  and digital precoder  $\mathbf{W}_S$  are sequentially optimized based on the idea of joint hierarchical optimization.

$C_{m_S}$  is defined as the achievable sum-rate corresponding to the  $m_S$ -th antenna subset, in which  $m_S = 1, \dots, M_S$ . During the optimization processes, the digital precoding matrix is assumed to be fixed. Hence the objective in (22) can be expressed as

$$\begin{aligned} \mathbf{F}_S &= \arg \max_{\mathbf{F}_S} C_{SR} = \sum_{m_S=1}^{M_S} C_{m_S} \\ \text{s.t. } |\mathbf{f}_S^{m_S}(i)| &= \frac{1}{\sqrt{N_S^{m_S}}}, \quad i = 1, 2, \dots, N_S^{m_S}, \\ \mathbf{F}_S &= \text{blk}(\mathbf{f}_S^1, \dots, \mathbf{f}_S^{M_S}). \end{aligned} \quad (24)$$

Define the hybrid precoding matrix  $\mathbf{F}_S = [\tilde{\mathbf{F}}_S^{M_S-1}, \tilde{\mathbf{F}}_S^{M_S}]$ , where  $\tilde{\mathbf{F}}_S^{M_S}$  and  $\tilde{\mathbf{F}}_S^{M_S-1}$  denote the  $M_S$ -th column and an  $N_S \times (M_S - 1)$  matrix containing the first  $M_S - 1$  columns of  $\mathbf{F}_S$ , respectively. Hence the sum-rate in (24) can be rewritten

as

$$\begin{aligned} C_{SR} &= \log_2(|\mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} \mathbf{F}_S \mathbf{F}_S^H \mathbf{H}^H|) = \log_2 \\ &\times (|\mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} [\tilde{\mathbf{F}}_S^{M_S-1}, \tilde{\mathbf{F}}_S^{M_S}] [\tilde{\mathbf{F}}_S^{M_S-1}, \tilde{\mathbf{F}}_S^{M_S}]^H \mathbf{H}^H|) \\ &= \log_2(|\mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S-1} (\tilde{\mathbf{F}}_S^{M_S-1})^H \mathbf{H}^H \\ &+ \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S} (\tilde{\mathbf{F}}_S^{M_S})^H \mathbf{H}^H|). \end{aligned} \quad (25)$$

Define an auxiliary matrix

$$\mathbf{S}_{M_S-1} = \mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S-1} (\tilde{\mathbf{F}}_S^{M_S-1})^H \mathbf{H}^H. \quad (26)$$

Due to  $|\mathbf{I} + \mathbf{X}\mathbf{Y}| = |\mathbf{I} + \mathbf{Y}\mathbf{X}|$ , (25) can be simplified as

$$\begin{aligned} C_{SR} &= \log_2(|\mathbf{S}_{M_S-1}|) \\ &+ \log_2 \left( |\mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{S}_{M_S-1}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S} (\tilde{\mathbf{F}}_S^{M_S})^H \mathbf{H}^H| \right) \\ &\stackrel{(a)}{=} \log_2(|\mathbf{S}_{M_S-1}|) \\ &+ \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 KL_S} (\tilde{\mathbf{F}}_S^{M_S})^H \mathbf{H}^H \mathbf{S}_{M_S-1}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S} \right). \end{aligned} \quad (27)$$

The first term on the right side of (a) in (27) is in the same form as that in (25), and the second term  $1 + \frac{P_S}{\sigma_R^2 KL_S} (\tilde{\mathbf{F}}_S^{M_S})^H \mathbf{H}^H \mathbf{S}_{M_S-1}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S}$  denotes the achievable sum-rate of the  $M_S$ -th antenna subset. Further, we decompose  $\log_2(|\mathbf{S}_{M_S-1}|)$  using the similar way in (27) as

$$\begin{aligned} \log_2(|\mathbf{S}_{M_S-1}|) &= \log_2(|\mathbf{S}_{M_S-2}|) \\ &+ \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 KL_S} (\tilde{\mathbf{F}}_S^{M_S-1})^H \mathbf{H}^H \mathbf{S}_{M_S-2}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{M_S-1} \right). \end{aligned} \quad (28)$$

Such the similar procedure will be executed until all  $M_S$  antenna subsets are considered. Then  $C_{SR}$  can be rewritten as

$$C_{SR} = \sum_{m_S=1}^{M_S} \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 KL_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{H}^H \mathbf{S}_{m_S-1}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{m_S} \right), \quad (29)$$

where  $\mathbf{S}_{m_S} = \mathbf{I}_{KL_S} + \frac{P_S}{\sigma_R^2 KL_S} \mathbf{H} \tilde{\mathbf{F}}_S^{m_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{H}^H$  and  $\mathbf{S}_1 = \mathbf{I}_{M_S}$ .

According to the above analysis, the sum-rate of the first selected antenna subset to be optimized can be expressed as

$$C_{m_S}^{\max} = \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 KL_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{H}^H \mathbf{S}_{m_S-1}^{-1} \mathbf{H} \tilde{\mathbf{F}}_S^{m_S} \right), \quad (30)$$

where  $C_{m_S}^{\max} = \max\{C_1, \dots, C_{M_S}\}$  and  $\mathbf{T}_{m_S-1} = \mathbf{H}^H \mathbf{S}_{m_S-1}^{-1} \mathbf{H}$ . Let  $\mathbf{G}_{m_S-1} \in \mathbb{C}^{N_S^{m_S} \times N_S^{m_S}}$  keep the rows and columns of  $\mathbf{T}_{m_S-1}$  from  $\left( \sum_{i=1}^{m_S-1} N_S^i + 1 \right)$  to  $\sum_{i=1}^{m_S} N_S^i$  and define

$\mathbf{R} = [\mathbf{0}_{\sum_{i=1}^{M_S-1} N_S^i \times N_S^{m_S}}, \mathbf{I}_{N_S^{m_S}}, \mathbf{0}_{\sum_{i=m_S}^{M_S} N_S^i \times N_S^{m_S}}]^T$  as the corresponding selection matrix. Then  $\mathbf{G}_{m_S-1}$  can be expressed as

$$\mathbf{G}_{m_S-1} = \mathbf{R}^H \mathbf{T}_{m_S-1} \mathbf{R} = \mathbf{R}^H \mathbf{H}^H \mathbf{S}_{m_S-1}^{-1} \mathbf{H} \mathbf{R}. \quad (31)$$

Therefore, (30) can be rewritten as

$$C_{m_S}^{\max} = \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 K L_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{G}_{m_S-1} \tilde{\mathbf{F}}_S^{m_S} \right). \quad (32)$$

Applying the singular value decomposition (SVD) to  $\mathbf{G}_{m_S-1}$ , we can obtain  $\mathbf{G}_{m_S-1} = \mathbf{V} \Sigma \mathbf{V}^H$ , in which  $\Sigma \in \mathbb{C}^{N_S^{m_S} \times N_S^{m_S}}$  is a diagonal matrix containing the singular values of  $\mathbf{G}_{m_S-1}$  arranged in descending order, and  $\mathbf{V} \in \mathbb{C}^{N_S^{m_S} \times N_S^{m_S}}$  denotes the right singular matrix. Hence the optimal solution of (24) can be obtained by

$$\left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_1 \\ \mathbf{0} \end{bmatrix}, \quad (33)$$

where  $\mathbf{v}_1$  represents the first column of  $\mathbf{V}$ . The elements of  $\mathbf{v}_1$  do not follow the constant-modulus constraint, which is not appropriate for the design of  $\tilde{\mathbf{F}}_S^{m_S}$ . However, we can explore a suitable substitute of  $\mathbf{v}_1$  to make  $\tilde{\mathbf{F}}_S^{m_S}$  close enough to its optimal solution  $\left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}}$ .

Matrices  $\Sigma$  and  $\mathbf{V}$  can be further divided into  $\Sigma = \text{blk}(\Sigma_1, \Sigma_2)$  and  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$ , respectively. Then (32) can be rewritten as

$$\begin{aligned} C_{m_S}^{\max} &= \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 K L_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{V} \Sigma \mathbf{V}^H \tilde{\mathbf{F}}_S^{m_S} \right) \\ &= \log_2 \left( 1 + \sum_{i=1}^2 \frac{P_S}{\sigma_R^2 K L_S} (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{v}_i \Sigma_i \mathbf{v}_i^H \tilde{\mathbf{F}}_S^{m_S} \right). \end{aligned} \quad (34)$$

In order to find the optimal solution of  $\tilde{\mathbf{F}}_S^{m_S}$ , it is reasonable to assume that  $\tilde{\mathbf{F}}_S^{m_S}$  is orthogonal to  $\mathbf{v}_2$ , i.e.,  $\tilde{\mathbf{F}}_S^{m_S} \mathbf{v}_2 \approx 0$ . According to the effective theory of high signal-to-noise ratio (SNR) approximation, i.e.,

$$\left( 1 + \frac{P_S}{\sigma_R^2 K L_S} \Sigma_1 \right)^{-1} \frac{P_S}{\sigma_R^2 K L_S} \Sigma_1 \approx 1, \quad (35)$$

(34) can be expressed as

$$\begin{aligned} C_{m_S}^{\max} &\approx \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 K L_S} \Sigma_1 (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{v}_1 \mathbf{v}_1^H \tilde{\mathbf{F}}_S^{m_S} \right) \\ &\approx \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 K L_S} \Sigma_1 \right) + \log_2 \left( (\tilde{\mathbf{F}}_S^{m_S})^H \mathbf{v}_1 \mathbf{v}_1^H \tilde{\mathbf{F}}_S^{m_S} \right). \end{aligned} \quad (36)$$

It can be observed from (36) that maximizing  $C_{m_S}^{\max}$  is equivalent to minimizing the mean-square-error (MSE) between  $\tilde{\mathbf{F}}_S^{m_S}$  and  $\left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}}$  on the constraint of constant-modulus. Define  $\mathbf{F}_S^{\text{opt}}$  as the optimal solution of  $\mathbf{F}_S$ . As a result, the optimization problem (24) can be further formulated as

$$\arg \min_{\mathbf{F}_S} E \left\{ \left\| \mathbf{F}_S^{\text{opt}} - \mathbf{F}_S \right\|_F^2 \right\}. \quad (37)$$

The MSE function can be expressed as

$$\begin{aligned} E \left\{ \left\| \mathbf{F}_S^{\text{opt}} - \mathbf{F}_S \right\|_F^2 \right\} &= \text{tr} \left\{ \left( \mathbf{F}_S^{\text{opt}} - \mathbf{F}_S \right)^H \left( \mathbf{F}_S^{\text{opt}} - \mathbf{F}_S \right) \right\} \\ &= 2M_S - \text{tr} \left\{ 2 \text{Re} \left( \mathbf{F}_S^H \mathbf{F}_S^{\text{opt}} \right) \right\} \\ &= 2M_S - 2 \sum_{n_S=1}^{N_S} \sum_{m_S=1}^{M_S} \\ &\quad \times \text{Re} \left( \left| \mathbf{F}_S(n_S, m_S) \right| \left| \mathbf{F}_S^{\text{opt}}(n_S, m_S) \right| \right. \\ &\quad \left. \times e^{j\varphi(n_S, m_S)} \right), \end{aligned} \quad (38)$$

where  $\varphi(n_S, m_S) = \angle \mathbf{F}_S(n_S, m_S) - \angle \mathbf{F}_S^{\text{opt}}(n_S, m_S)$ . It is clear that when  $\varphi(n_S, m_S) = 0$ , i.e., each column of  $\mathbf{F}_S$  shares the same phase with that of  $\mathbf{F}_S^{\text{opt}}$ , the objective function in (37) is minimized. Therefore, the analog precoding vector  $\tilde{\mathbf{F}}_S^{m_S}$  can be chosen as

$$\tilde{\mathbf{F}}_S^{m_S} = \frac{1}{\sqrt{N_S}} e^{j\angle \left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}}}, \quad (39)$$

where  $\angle \left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}}$  represents the phase vector of  $\left( \tilde{\mathbf{F}}_S^{m_S} \right)_{\text{opt}}$  and  $m_S = 1, \dots, M_S$ .

According to the above analysis, the optimization problem (24) can be transformed into a series of sub-problems which can be optimized one by one. The optimization procedure of the proposed analog precoder  $\mathbf{F}_S$  design scheme is summarized and shown in Fig. 2. All  $M_S$  antenna subsets are first sorted on the basis of their corresponding channel capacity. Then we update  $\mathbf{S}_{m_S}$  and optimize  $\tilde{\mathbf{F}}_S^{m_S}$  in the order of arrangement. Repeat above processes until the sum-rate corresponding to all antenna subsets is optimized.

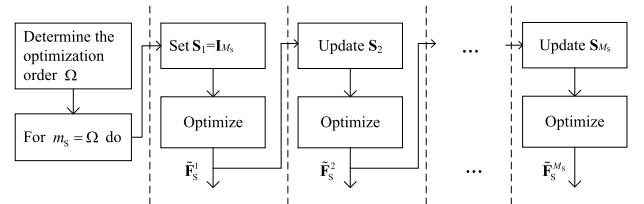


FIGURE 2. Diagram of the proposed analog precoder  $\mathbf{F}_S$  design scheme.

According to the idea of joint hierarchical optimization, we sequentially optimize  $\mathbf{F}_{R1}$ ,  $\mathbf{W}_{R1}$  and  $\mathbf{W}_S$ . Applying SVD to  $\mathbf{H}\mathbf{F}_S$ , we have  $\mathbf{H}\mathbf{F}_S = \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^H$ , in which  $\tilde{\Sigma}$  is a diagonal matrix,  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$  are respective the left and right singular matrices. The analog combiner  $\mathbf{F}_{R1}$  is also confined to the constant-modulus constraint. Therefore,  $\mathbf{F}_{R1}$  is obtained by

$$\mathbf{F}_{R1} = \frac{1}{\sqrt{N_R}} e^{j\angle \tilde{\mathbf{U}}(:, 1:M_R)}, \quad (40)$$

which is similar to the optimization of  $\mathbf{F}_S$ . Then we perform SVD to  $\mathbf{F}_{R1}^H \mathbf{H}\mathbf{F}_S$ , and take the first  $K L_S$  column of its left singular matrix  $\tilde{\mathbf{U}}$  to optimize  $\mathbf{W}_{R1}$ , i.e.,

$$\mathbf{W}_{R1} = \tilde{\mathbf{U}}(:, 1:K L_S). \quad (41)$$



Via calculating the SVD of  $\mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{H} \mathbf{F}_S$ , the optimization of  $\mathbf{W}_S$  can be obtained by

$$\mathbf{W}_S = \mathbf{V}(:, 1 : KL_S), \quad (42)$$

where  $\mathbf{V}$  is the right singular matrix.

By exploiting the idea of joint hierarchical optimization, we recursively combine the channel with optimized precoders/combiners, and perform SVD to them. Thereby, all optimized precoders/combiners can be obtained. Since the optimization order depends on the channel capacity of different antenna subsets, the proposed precoders and combiners design scheme reduces the capacity loss. Finally, the classic water filling power allocation method is used to get the design of power allocation matrix. The proposed hybrid precoding design from the source to relay decoding scheme is summarized and shown in Table 1.

**TABLE 1. The proposed hybrid precoding design from the source to relay decoding.**

<b>Input:</b> $M_S, N_S^{m_S}$ and $\mathbf{H}$ .
1. Determine the optimization order $\Omega$ by sorting the capacities of different analog sub-channels in descending order;
2. <b>For</b> $m_S = \Omega$ <b>do</b>
3. Apply SVD to $\mathbf{G}_{m_S-1}$ ;
4. Compute $(\tilde{\mathbf{F}}_S^{m_S})_{\text{opt}}$ by (33);
5. Compute $\tilde{\mathbf{F}}_S^{m_S}$ by (39);
6. <b>End for</b>
7. $\mathbf{F}_S = \text{blk}(\tilde{\mathbf{F}}_S^1, \dots, \tilde{\mathbf{F}}_S^{M_S})$ ;
8. Apply SVD to $\mathbf{H} \mathbf{F}_S$ and obtain $\mathbf{F}_{R1}$ by (40);
9. Apply SVD to $\mathbf{F}_{R1}^H \mathbf{H} \mathbf{F}_S$ and obtain $\mathbf{W}_{R1}$ by (41);
10. Apply SVD to $\mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{H} \mathbf{F}_S$ and obtain $\mathbf{W}_S$ by (42);
11. Compute the total equivalent baseband channel $\mathbf{H}_{\text{total}} = \mathbf{W}_{R1}^H \mathbf{F}_{R1}^H \mathbf{H} \mathbf{F}_S \mathbf{W}_S$ ;
12. Compute $\mathbf{P}_S$ by using the water filling power allocation method.
<b>Output:</b> $\mathbf{W}_S, \mathbf{F}_S, \mathbf{F}_{R1}, \mathbf{W}_{R1}$ and $\mathbf{P}_S$ .

### B. HYBRID PRECODING DESIGN FROM THE RELAY FORWARDING TO USERS

We now focus on solving the sub-problem  $\mathcal{P}_2$ . It can be regarded as a massive MU-MIMO system with a sub-connected structure. The improved SSIC algorithm is considered to optimize the analog precoder  $\mathbf{F}_{R2}$ . If other variables are fixed in (23), the problem can be transformed into:

$$\begin{aligned} \mathbf{F}_{R2} &= \arg \max_{\mathbf{F}_{R2}} C_{\max} \\ \text{s.t. } |\mathbf{f}_{R2}^{m_R}(j)| &= \frac{1}{\sqrt{N_R^{m_R}}}, \quad j = 1, 2, \dots, N_R^{m_R} \\ \mathbf{F}_{R2} &= \text{blk}(\mathbf{f}_{R2}^1, \dots, \mathbf{f}_{R2}^{M_R}), \end{aligned} \quad (43)$$

where

$$\begin{aligned} C_{\max} &= \sum_{m_R=1}^{M_R} C_{m_R} \\ &= \sum_{m_R=1}^{M_R} \log_2 \left( 1 + \frac{P_R}{\sigma_D^2 K L_S} (\tilde{\mathbf{F}}_{R2}^{m_R})^H \mathbf{G}^H \mathbf{G} \tilde{\mathbf{F}}_{R2}^{m_R} \right). \end{aligned} \quad (44)$$

The whole process is the same as that of solving the problem  $\mathcal{P}_1$ , and hence is omitted here for the sake of brevity. According to the flowchart in Fig. 2,  $\mathbf{F}_{R2}$  is obtained based on the improved SSIC.

After that, the analog precoding matrix  $\mathbf{F}_{R2}$  is combined with the channel matrix  $\mathbf{G}_k$  as  $\mathbf{G}_k \mathbf{F}_{R2} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ . The analog combiner  $\mathbf{F}_{Dk}$  has the constant-modulus constraint. It can be obtained by

$$\mathbf{F}_{Dk} = \frac{1}{\sqrt{N_{Dk}}} e^{j\angle \mathbf{U}(:, 1 : M_{Dk})}. \quad (45)$$

Then we perform SVD to  $\mathbf{F}_{Dk}^H \mathbf{G}_k \mathbf{F}_{R2}$ , and take the first  $L_S$  column of its left singular matrix  $\mathbf{U}_k$  to optimize  $\mathbf{W}_{Dk}$ , i.e.,

$$\mathbf{W}_{Dk} = \mathbf{U}_k(:, 1 : L_S). \quad (46)$$

The analog compound matrix  $\mathbf{F}_D = \text{blk}(\mathbf{F}_{D1}, \dots, \mathbf{F}_{DK})$  and digital compound matrix  $\mathbf{W}_D = \text{blk}(\mathbf{W}_{D1}, \dots, \mathbf{W}_{DK})$  are obtained by optimizing  $\mathbf{F}_{Dk}$  and  $\mathbf{W}_{Dk}$  for  $k = 1, \dots, K$ .

In order to obtain the optimal digital precoder  $\mathbf{W}_{R2}$ , the modified BD technique is adopted. The MU-MIMO channel can be divided into multiple SU-MIMO channels, which is the main idea of applying BD technique. If the signal received by the  $k$ -th user can be guaranteed to be in the null space of other user channels, the inter-user interference can be eliminated [32]. First of all, define  $\mathbf{G}_{\text{int},k} = \mathbf{W}_D^H \mathbf{F}_D^H \mathbf{G}_k \mathbf{F}_{R2}$ ,  $k \in \{1, \dots, K\}$ . The constraint can be expressed as  $\mathbf{G}_{\text{int},j} \mathbf{W}_{R2}^k = 0, \forall j \neq k$ , where  $\mathbf{W}_{R2}^k = \mathbf{W}_{R2}(:, ((k-1)L_S + 1) : kL_S)$  and  $j, k \in \{1, \dots, K\}$ . Define  $\tilde{\mathbf{G}}_k = [\mathbf{G}_{\text{int},1}^T, \dots, \mathbf{G}_{\text{int},k-1}^T, \mathbf{G}_{\text{int},k}^T, \dots, \mathbf{G}_{\text{int},K}^T]^T$  should fall in the null space of  $\tilde{\mathbf{G}}_k$ . Applying SVD to  $\tilde{\mathbf{G}}_k$ , we can get

$$\tilde{\mathbf{G}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k [\tilde{\mathbf{V}}_k^{(1)}, \tilde{\mathbf{V}}_k^{(0)}]^H \quad (47)$$

and

$$\begin{aligned} \tilde{\mathbf{G}}_k \tilde{\mathbf{V}}_k^{(0)} &= \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k [\tilde{\mathbf{V}}_k^{(1)}, \tilde{\mathbf{V}}_k^{(0)}]^H \tilde{\mathbf{V}}_k^{(0)} \\ &= \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k (\tilde{\mathbf{V}}_k^{(1)})^H \tilde{\mathbf{V}}_k^{(0)} \\ &= 0, \end{aligned} \quad (48)$$

where  $\tilde{\mathbf{V}}_k^{(1)} = \tilde{\mathbf{V}}_k(:, 1 : (K-1)L_S)$  and  $\tilde{\mathbf{V}}_k^{(0)} = \tilde{\mathbf{V}}_k(:, (K-1)L_S + 1 : \text{end})$  represent the subspace orthogonal bases and the null space orthogonal bases of  $\tilde{\mathbf{G}}_k$ , respectively. Decomposing  $\mathbf{G}_{\text{int},k} \tilde{\mathbf{V}}_k^{(0)}$  by SVD yields

$$\mathbf{G}_{\text{int},k} \tilde{\mathbf{V}}_k^{(0)} = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k [\tilde{\mathbf{V}}_k^{(1)}, \tilde{\mathbf{V}}_k^{(0)}]^H. \quad (49)$$

To eliminate inter-user interference,  $\tilde{\mathbf{V}}_k^{(1)}$  corresponding to the non-zero singular values is taken as the precoding matrix. The final digital precoder is given by

$$\mathbf{W}_{R2}^k = \tilde{\mathbf{V}}_k^{(0)} \tilde{\mathbf{V}}_k^{(1)}. \quad (50)$$

The water filling power allocation method is then performed to obtain the power allocation matrix. The proposed hybrid precoding design from the relay forwarding to users scheme is summarized and shown in Table 2.

**TABLE 2.** The proposed hybrid precoding design from the relay forwarding to users.

<b>Input:</b> $M_R, N_R^{mR}, K$ and $\mathbf{G}_k$ for all $k = 1, \dots, K$ .
1. Refer to <b>Table 1</b> to compute $\mathbf{F}_{R2}$ ;
2. <b>For</b> $k = 1, \dots, K$ <b>do</b>
3. Apply SVD to $\mathbf{G}_k \mathbf{F}_{R2}$ and obtain $\mathbf{F}_{Dk}$ by (45);
4. Apply SVD to $\mathbf{F}_{Dk}^H \mathbf{G}_k \mathbf{F}_{R2}$ and obtain $\mathbf{W}_{Dk}$ by (46);
5. <b>End for</b>
6. $\mathbf{F}_D = \text{blk}(\mathbf{F}_{D1}, \dots, \mathbf{F}_{DK})$ and $\mathbf{W}_D = \text{blk}(\mathbf{W}_{D1}, \dots, \mathbf{W}_{DK})$ ;
7. Define $\mathbf{G}_{\text{int},k} = \mathbf{W}_D^H \mathbf{F}_D^H \mathbf{G}_k \mathbf{F}_{R2}$ , $k \in \{1, \dots, K\}$ and $\bar{\mathbf{G}}_k = [\mathbf{G}_{\text{int},1}^T, \dots, \mathbf{G}_{\text{int},k-1}^T, \mathbf{G}_{\text{int},k}^T, \dots, \mathbf{G}_{\text{int},K}^T]^T$ ;
8. <b>For</b> $k = 1, \dots, K$ <b>do</b>
9. Apply SVD to $\bar{\mathbf{G}}_k$ and obtain $\bar{\mathbf{V}}_k^{(0)}$ as in (47) and (48);
10. Apply SVD to $\mathbf{G}_{\text{int},k} \bar{\mathbf{V}}_k^{(0)}$ and obtain $\tilde{\mathbf{V}}_k^{(1)}$ as in (49);
11. Compute $\mathbf{W}_{R2}^k$ by (50);
12. <b>End for</b>
13. Compute the total equivalent baseband channel $\mathbf{G}_{\text{total}} = \mathbf{W}_D^H \mathbf{F}_D^H \mathbf{G}_k \mathbf{F}_{R2} \mathbf{W}_{R2}$ ;
14. Compute $\mathbf{P}_R$ by using the water filling power allocation method.
<b>Output:</b> $\mathbf{W}_D, \mathbf{F}_D, \mathbf{F}_{R2}, \mathbf{W}_{R2}$ and $\mathbf{P}_R$ .

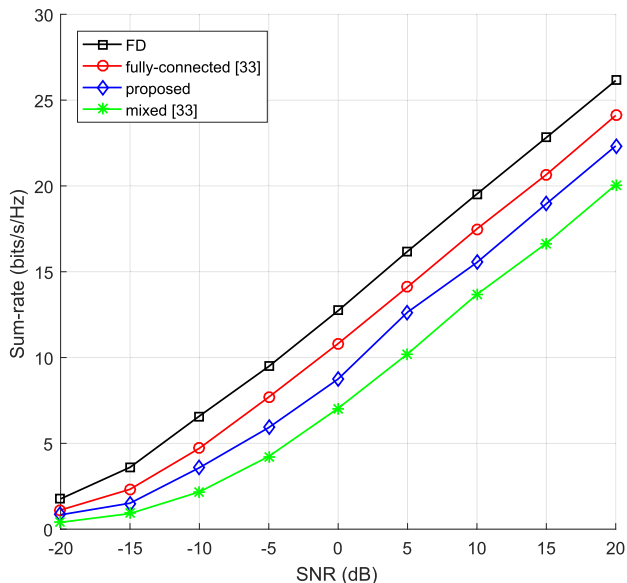
**IV. SIMULATION RESULT**

In this section, we evaluate the performance of the proposed hybrid beamforming scheme in a DF relay system connected with the mixed structure. The corresponding simulation results are described below. All simulation results are based on the MATLAB platform for the average value experiment of 1000 channels. For simplicity, the propagation environment is modeled as  $N_c = 8$  clusters with  $N_p = 10$  rays per cluster, and the inter-element spacing  $d$  is assumed to be half wavelength. AoAs and AoDs of each element are uniformly distributed in  $[0, 2\pi]$ , respectively. The specific setting parameters are as follows:  $N_S = 128, N_R = 64, N_{Dk} = 16, M_S = 8, M_R = 8, M_{Dk} = 2, N_S^{mS} = 16, N_R^{mR} = 8$ . During each simulation, it is assumed that  $K = 4$  users simultaneously send  $L_S = 1$  data stream. The transmission power at the source and relay are equal, i.e.,  $P_S = P_R$ , and the noise variable is  $\sigma_R^2 = \sigma_D^2$ .

In order to clearly verify the effectiveness of the proposed algorithm, we compare the performance of the proposed method with other algorithms. The optimal curve can be known as the use of FD precoding in the source and relay communication, and the dirty paper coding (DPC) algorithm in the relay and destination communication. This curve is defined as FD in simulations as the performance upper limit of the DF relay system with mixed structure. In addition, the hybrid beamforming design schemes of DF relay system with fully-connected and mixed structures proposed in [33] will be compared as a benchmark.

**A. SUM-RATE COMPARISON FOR DIFFERENT SNR**

The objective function studied in Section II is the sum-rate between source and destination. Therefore, we compare the overall rate performance of different hybrid precoding algorithms. Fig. 3 shows the performance comparison of various algorithms versus SNR, where the antenna array is arranged with ULA. Through the comparison of simulation curves,



**FIGURE 3.** Sum-rate comparison for different hybrid precoding schemes with ULA at each node.

it can be clearly seen that the sum-rate performance of the proposed hybrid precoding algorithm is better than the mixed scheme, and very close to the fully-connected scheme. Therefore, the proposed hybrid precoding algorithm in the DF mmWave relay system with mixed structure is a good solution. The obvious gap with the optimal FD can be summarized as the structural difference adopted under the mmWave DF relay system. The difference between sub-connected and fully-connected structures has led to a decline in performance. On the other hand, it can be known that the mixed structure is a compromise between fully-connected and sub-connected structures. Due to the increasing implementation cost of the fully-connected structure, the proposed mixed structure has more practical application prospects.

Fig. 4 shows the sum-rate performance comparison for different hybrid precoding schemes versus SNR, where the antenna arrays is arranged with UPA. It can be seen from Fig. 4 that when the antenna layout is changed from a linear array to a planar array, the performance of the proposed hybrid precoding algorithm and the mixed scheme drop slightly, but that of fully-connected scheme improves correspondingly. In addition, Fig. 4 also verifies that although the performance of the proposed mixed structure in mmWave DF relay system is slightly lower than that in Fig. 3, its trend still shows that the proposed algorithm is objectively effective. The change of antenna deployment mode also greatly improves the space utilization of equipment.

Note that since the optimal FD precoding scheme asks a dedicated RF chain to be assigned for each antenna, it is more complex in hardware implementation. The mixed structure sacrifices a little degree of freedom, which results in a relatively poor sum-rate performance. However, the proposed mixed structure can reduce hardware complexity and power consumption (see Fig. 9 and Fig. 10) without an obvious performance loss, it is more practically attractive.

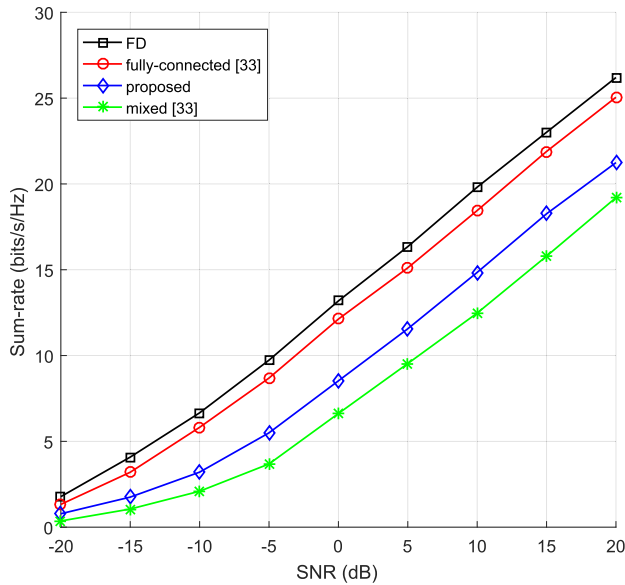


FIGURE 4. Sum-rate comparison for different hybrid precoding schemes with UPA at each node.

**B. SUM-RATE COMPARISON FOR DIFFERENT NUMBER OF SOURCE NODE ANTENNAS**

Fig. 5 compares the sum-rate performance of different beamforming schemes versus the number of source node antennas with ULA and UPA. In order to better compare simulation results under the two arrays, SNR is set as 5dB in ULA and -5dB in UPA. As can be observed from Fig. 5 that when the number of source node antennas increases, the sum-rate performance of different design schemes keep saturate due to the sum-rate from the source node to relay is higher than that from relay to users. However, in UPA, the performance of FD and fully-connected schemes improve obviously when the number of source node antennas is small, and will saturate

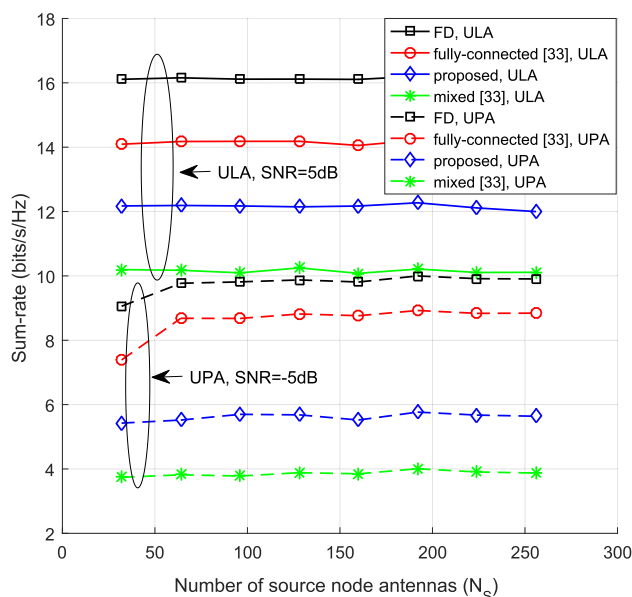


FIGURE 5. Sum-rate comparison for different beamforming schemes versus the number of source node antennas with ULA and UPA.

with increasing number. In addition, the performance of the proposed method always outperforms the mixed scheme both in ULA and UPA.

**C. SUM-RATE COMPARISON FOR DIFFERENT NUMBER OF RELAY ANTENNAS**

Fig. 6 plots the sum-rate performance of the comparative algorithms, when the number of relay antennas ranges from 32 to 256, where SNR = 5dB in ULA and SNR = -5dB in UPA. It can be seen from Fig. 6, when the number of relay antennas increases, the sum-rate performance of all comparative algorithms improves as antenna gain increases, which is expected. Since the DF relay system can be considered as a series of two single-hop MIMO systems, the performance of different beamforming schemes is improved slowly with the increasing the number of relay antennas.

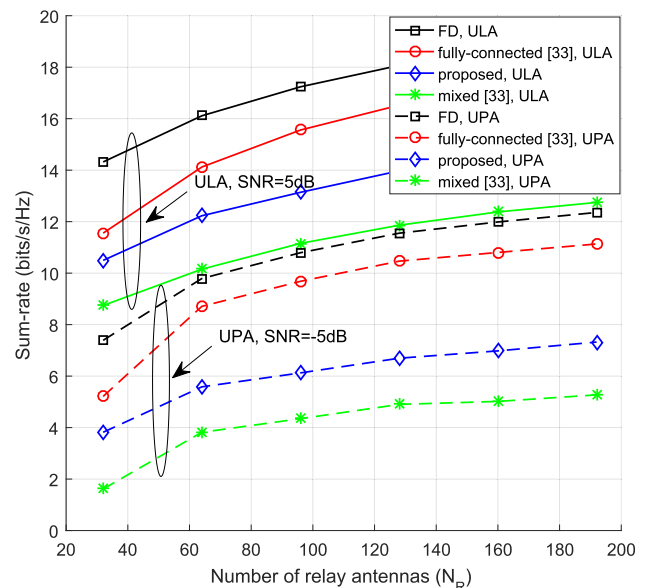


FIGURE 6. Sum-rate comparison for different beamforming schemes versus the number of relay antennas with ULA and UPA.

**D. SUM-RATE COMPARISON FOR DIFFERENT NUMBER OF RELAY RF CHAINS**

Fig. 7 presents the sum-rates achieved by different beamforming schemes when the number of relay RF chains ranges from 8 to 32, where  $N_R = 4M_R$ . Since our proposed method is designed to maximize the sum-rate between source and user node after RF beamforming/combining, the gap between our method and the FD scheme is more-or-less fixed, which is caused by the analog processing. However, the proposed scheme with ULA is closer to the fully-connected scheme compared with UPA.

**E. SUM-RATE COMPARISON FOR DIFFERENT NUMBER OF USERS**

Fig. 8 compares the sum-rate performance of different beamforming schemes versus the number of users with SNR = 5dB in ULA and SNR = -5dB in UPA, where the number of

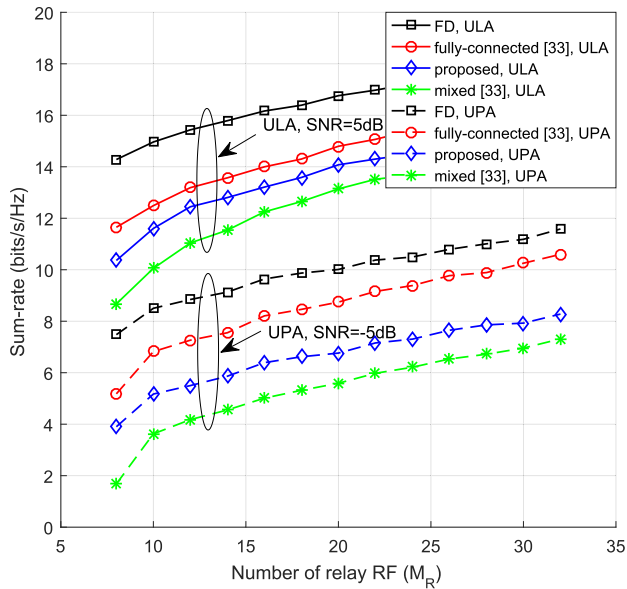


FIGURE 7. Sum-rate comparison for different beamforming schemes versus the number of relay RF chains with ULA and UPA, where  $N_R = 4M_R$ .

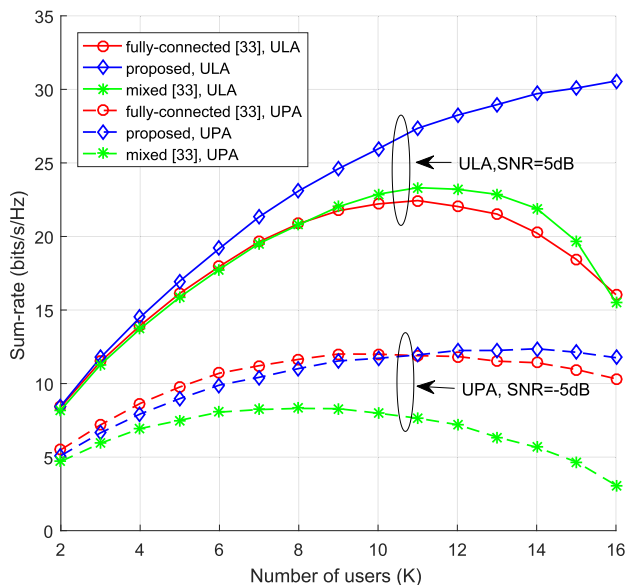


FIGURE 8. Sum-rate comparison for different beamforming schemes versus the number of users with ULA and UPA, where  $M_S = M_R = 32$ .

users changes from 2 to 16, and that of RF chains at source and relay are  $M_S = M_R = 32$ . Since the row subspace of channel matrices overlap significantly if the number of users becomes large, the baseband BD technology can result in a poor performance. However, since the proposed scheme adopts the criterion which tries to avoid the information loss, the performance of the proposed scheme is superior to fully-connected and mixed schemes with increasing number of users as shown in Fig. 8. In addition, the performance of the proposed scheme with UPA is slightly lower than that with ULA, but it outperforms the fully-connected method after  $K = 11$ .

### F. POWER EFFICIENT COMPARISON FOR DIFFERENT NUMBER OF RELAY RF CHAINS

The power consumption is a key issue which should be considered for both sub-connected and fully-connected structures. Fig. 9 presents power efficient comparison for different beamforming schemes versus the number of RF chains at relay with ULA, where  $N_R = 4M_R$  and  $SNR = 5dB$ . As shown in Fig. 9, the power efficiency performance of different beamforming schemes increase tremendously with increasing number of RF chains. However, since the proposed mixed structure has less power consumption than other

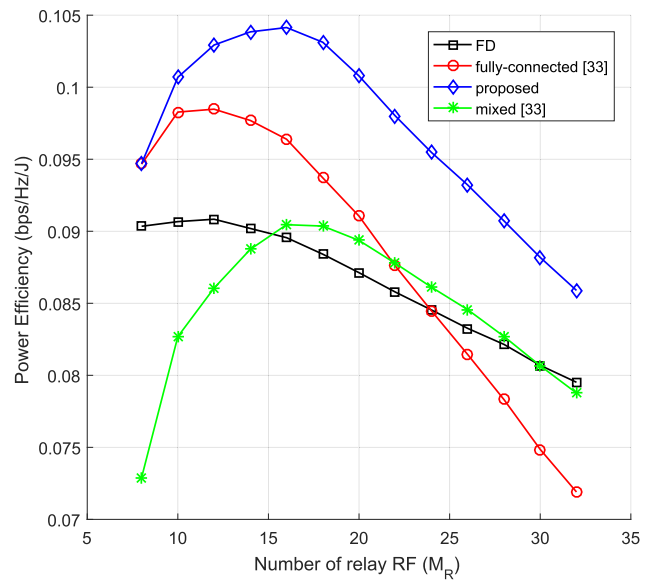


FIGURE 9. Power efficient comparison for different beamforming schemes versus the number of RF chains at relay with ULA, where  $N_R = 4M_R$  and  $SNR = 5dB$ .

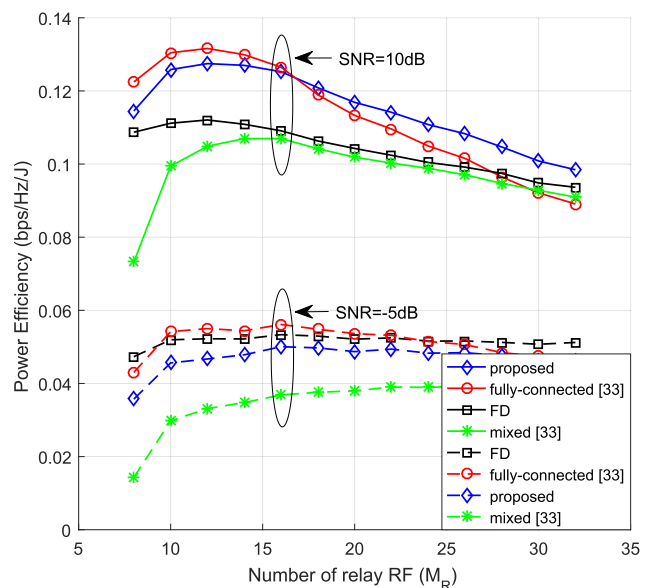


FIGURE 10. Power efficient comparison for different beamforming schemes versus the number of RF chains at relay with UPA, where  $N_R = 4M_R$ ,  $SNR = 10dB$  and  $-5dB$ .

schemes, the power efficiency performance of the proposed beamforming scheme is the highest. In addition, with increasing number of relay RF chains, the power consumption generated by phase shifters of the fully-connected structure is dominant, thus its power efficiency performance is gradually lower than that of FD scheme.

Fig. 10 presents power efficient comparison for different beamforming schemes versus the number of RF chains at relay with UPA, where  $N_R = 4M_R$  and SNR = 10dB and -5dB. Although the proposed mixed structure with UPA has less power consumption, its sum-rate performance is low compared with ULA. Therefore, the performance of the proposed algorithm is poor when SNR = -5dB as shown in Fig. 10. However, as SNR increases, e.g., SNR = 10dB, the sum-rate performance of the proposed scheme increases, thus its power efficiency performance improves and outperforms other schemes.

## V. CONCLUSION

In this paper, we considered the mixed-structure DF relay systems in the domain of mmWave massive MU-MIMO. The hybrid beamforming design is to maximize the sum-rate between the source node and users. To solve this challenging hybrid beamforming design problem, an efficient sorted serial design method is proposed to design the analog beamforming of each node. Further, to mitigate the inner and inter-user interference and increase the number of users carried by the system, a modified baseband BD technology is proposed to design the digital beamforming at each node. In addition, the proposed hybrid beamforming algorithm is designed by considering both ULAs and UPAs. Simulation results demonstrate that the proposed hybrid beamforming scheme can achieve superior performance in terms of achievable sum-rate and power efficiency in both ULAs and UPAs. In the future, a hybrid beamforming design of non-orthogonal multiple access (NOMA) based relay systems [34], [35] will be studied based on this paper. NOMA can be realized in code, power, or other domains. By using NOMA in power distribution, the number of users can be increased. Imperfect CSI and imperfect SIC are also considered in that work.

## REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.
- [3] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.
- [4] X. Li, M. Zhao, M. Zeng, S. Mumtaz, V. G. Menon, Z. Ding, and O. A. Dobre, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," *IEEE Trans. Commun.*, early access, Jan. 11, 2021, doi: 10.1109/TCOMM.2021.3050503.
- [5] M. Arti and M. R. Bhatnagar, "Performance analysis of hop-by-hop beamforming and combining in DF MIMO relay system over Nakagami-m fading channels," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2080–2083, Nov. 2013.
- [6] Z. Yi and I.-M. Kim, "Optimum beamforming in the broadcasting phase of bidirectional cooperative communication with multiple decode-and-forward relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5806–5812, Dec. 2009.
- [7] M. R. Bhatnagar and M. K. Arti, "Selection beamforming and combining in decode-and-forward MIMO relay networks," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1556–1559, Aug. 2013.
- [8] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.
- [9] J. Du, M. Han, L. Jin, Y. Hua, and X. Li, "Semi-blind receivers for multi-user massive MIMO relay systems based on block Tucker2-PARAFAC tensor model," *IEEE Access*, vol. 8, pp. 32170–32186, 2020.
- [10] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3403–3418, Jul. 2018.
- [11] X. Wu, D. Liu, and F. Yin, "Hybrid beamforming for multi-user massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3879–3891, Sep. 2018.
- [12] A. Almradi, M. Matthaiou, P. Xiao, and V. F. Fusco, "Hybrid precoding for massive MIMO with low rank channels: A two-stage user scheduling approach," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4816–4831, Aug. 2020.
- [13] S. Wang, M. He, Y. Zhang, and R. Ruby, "Equivalent channel-based joint hybrid precoding/combining for large-scale MIMO systems," *Phys. Commun.*, vol. 47, Aug. 2021, Art. no. 101287.
- [14] S. S. Nair and S. Bhashyam, "Hybrid beamforming in MU-MIMO using partial interfering beam feedback," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1548–1552, Jul. 2020.
- [15] Y. Zhang, J. Du, Y. Chen, X. Li, K. M. Rabie, and R. Khkrel, "Dual-iterative hybrid beamforming design for millimeter-wave massive multi-user MIMO systems with sub-connected structure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13482–13496, Nov. 2020.
- [16] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5838–5842.
- [17] X. Xue, T. E. Bogale, X. Wang, Y. Wang, and B. L. Long, "Hybrid analog-digital beamforming for multiuser MIMO millimeter wave relay systems," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–7.
- [18] X. Xue, Y. Wang, L. Dai, and C. Masouros, "Relay hybrid precoding design in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2011–2026, Apr. 2018.
- [19] W. Xu, Y. Wang, and X. Xue, "ADMM for hybrid precoding of relay in millimeter-wave massive MIMO system," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.
- [20] C. Xing, X. Zhao, S. Wang, W. Xu, S. X. Ng, and S. Chen, "Hybrid transceiver optimization for multi-hop communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1880–1895, Aug. 2020.
- [21] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmWave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 662–674, Feb. 2018.
- [22] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Millimeter-wave massive MIMO systems relying on generalized sub-array-connected hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8940–8950, Sep. 2019.
- [23] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [24] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
- [25] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [26] M. Majidzadeh, A. Moilanen, N. Tervo, H. Pennanen, A. Tolli, and M. Latva-aho, "Hybrid beamforming for single-user MIMO with partially connected RF architecture," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–6.
- [27] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

- [28] A. Forenza, D. J. Love, and R. W. Heath, Jr., "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.
- [29] J. Gao, S. A. Vorobyov, H. Jiang, J. Zhang, and M. Haardt, "Sum-rate maximization with minimum power consumption for MIMO DF two-way relaying—Part I: Relay optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3563–3577, Jul. 2013.
- [30] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.
- [31] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [32] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8315–8328, Dec. 2019.
- [33] Y. Zhang, J. Du, Y. Chen, M. Han, and X. Li, "Optimal hybrid beamforming design for millimeter-wave massive multi-user MIMO relay systems," *IEEE Access*, vol. 7, pp. 157212–157225, 2019.
- [34] O. Abbasi, A. Ebrahimi, and N. Mokari, "NOMA inspired cooperative relaying system using an AF relay," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 261–264, Feb. 2019.
- [35] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 680–695, Jan. 2020.



**XINGWANG LI** (Senior Member, IEEE) received the B.Sc. degree in communication engineering from Henan Polytechnic University, Jiaozuo, China, in 2007, the M.Sc. degree from the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China (UESTC), in 2010, and the Ph.D. degree in communication and information system from the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications (BUPT), in 2015. From 2017 to 2018, he was a Visiting Scholar with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast (QUB), Belfast, U.K. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University. He worked on several funded research projects on the wireless communications areas. He has several papers published in journal and conferences, and authored several patents. His research interests include MIMO communication, cooperative communication, hardware constrained communication, NOMA, physical layer security, UAV, FSO communications, and performance analysis of fading channels. He is a TPC Member of IEEE/CIC ICCS workshop 19' and IEEE Globecom Workshop 18'. He also serves as an Associate Editor for the IEEE Access and an Editor for the *KSII Transactions on Internet and Information Systems*.



**MENG HAN** (Student Member, IEEE) received the B.Sc. degree from the Communication University of China, Beijing, China, in 2019, where she is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. She has published a journal article in IEEE Access and a conference paper in ICCT. Her current research interests include hybrid beamforming technique, tensor-based technique, massive MIMO systems, cooperative communication technique, and channel estimation.



**JIANHE DU** (Member, IEEE) received the B.Sc. and M.Sc. degrees from Yunnan University, Yunnan, China, in 2007 and 2010, respectively, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015. He is currently an Associate Professor with the School of Information and Engineering, Communication University of China, Beijing. He has published several journal articles in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE ACCESS, IEEE COMMUNICATIONS LETTERS, and *IET Communications*. His research interests mainly include channel estimation, massive MIMO communication, MIMO relay, and tensorbased signal processing applied to wireless communication. He was a Guest Editor for the special issue "Applications of Tensor Models in Wireless Communications and Mobile Computing" of the *Wireless Communications and Mobile Computing* journal.



**YANG ZHANG** (Student Member, IEEE) received the M.Sc. degree from the Guilin University of Electronic Technology, Guilin, China, in 2013. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Communication University of China, Beijing, China. His current main research interests include millimeter wave massive MIMO communication, channel estimation, MIMO relay, and beamforming technology for wireless networks.



**KHALED M. RABIE** (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical and electronic engineering from the University of Tripoli, Tripoli, Libya, in 2008, and the M.Sc. and Ph.D. degrees in communication engineering from The University of Manchester, Manchester, U.K., in 2010 and 2015, respectively. He is currently a Postdoctoral Research Associate with Manchester Metropolitan University (MMU), Manchester. His research interests include signal processing and the analysis of power line and wireless communication networks. He is a Fellow of the U.K. Higher Education Academy. He was a recipient of the Best Student Paper Award at the IEEE ISPLC, TX, USA, 2015, and the MMU Outstanding Knowledge Exchange Project Award, in 2016. He is currently the Program Chair of the IEEE ISPLC 2018, the IEEE CSNDSP 2018 Co-Chair of the Green Communications and Networks track, and the Publicity Chair of the INISCOM 2018. He is also an Associate Editor of IEEE Access and an Editor of the *Physical Communication* journal (Elsevier).



**GALYMZHAN NAURYZBAYEV** (Senior Member, IEEE) received the B.Sc. and M.Sc. (Hons.) degrees in radio engineering, electronics, and telecommunications from the Almaty University of Power Engineering and Telecommunication, Almaty, Kazakhstan, in 2009 and 2011, respectively, and the Ph.D. degree in wireless communications from The University of Manchester, U.K., in 2016. From 2016 to 2018, he held several academic and research positions with Nazarbayev University, Kazakhstan, L.N. Gumilyov Eurasian National University, Kazakhstan, and Hamad Bin Khalifa University, Qatar. In 2019, he joined Nazarbayev University, as an Assistant Professor. His research interest includes the area of wireless communication systems, with particular focus on multi-user MIMO systems, cognitive radio, signal processing, energy harvesting, visible light communications, NOMA, interference mitigation, and so on. He served as a Technical Program Committee Member on numerous IEEE flagship conferences.