IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 8, 2019, accepted October 22, 2019, date of publication October 28, 2019, date of current version November 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949786

Optimal Hybrid Beamforming Design for Millimeter-Wave Massive Multi-User MIMO Relay Systems

YANG ZHANG[1], JIANHE DU[1], (Member, IEEE), YUANZHI CHEN[1], MENG HAN[1], AND XINGWANG LI[2], (Member, IEEE)

[1]School of Information and Communication Engineering, Communication University of China, Beijing 100024, China
[2]School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

Corresponding author: Jianhe Du (dujianhe1@gmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0502001, in part by the National Natural Science Foundation of China under Grant 61601414, Grant 61701448, and Grant 61702466, in part by the Fundamental Research Funds for the Central Universities under Grant 2018CUCTJ082, Grant CUC19ZD001, and Grant CUC2019D012, in part by the Henan Scientific and Technological Research Project under Grant 182102210307, in part by the Doctoral Scientific Funds of Henan Polytechnic University under Grant B2017-55, and in part by the Key Scientific Research Projects of Higher Education Institutions in Henan Province under Grant 20A510007.

ABSTRACT As a promising technology in the next generation mobile network, millimeter-wave (mmWave) communication can mitigate the spectrum crunch of improving the network capacity by exploiting the large underutilized spectrum bands of the mmWave frequencies. The hybrid (analog/digital) beamforming of multi-data streams are widely used to further the spectrum efficiency of mmWave relay system when faced with the complex environment or long distance communication. This paper investigates the hybrid beamforming scheme for the decode-and-forward (DF) mmWave massive multiple-input multiple-output (MIMO) relay system with mixed structure and full-connected structure. We optimize hybrid beamforming of relay system by maximizing the sum rate of the overall system as an objective function. To reduce the computational complexity, we reformulate the original problem as two single-hop mmWave MIMO sum-rate maximization subproblems. Then, the piecewise successive approximation method is proposed based on the criterion which jointly designs the analog and digital beamforming stages by trying to avoid the loss of information at each stage. The hybrid beamforming of the two subproblems can be solved by the proposed scheme united with the idea of successive interference cancelation (SIC), the baseband block diagonalization (BD) scheme, and waterfilling power allocation method. Finally, simulation results confirm that the proposed optimal method can achieve good performance in hybrid beamforming design of relay system with both mixed and full-connected structures.

INDEX TERMS Millimeter wave, massive MIMO, relay, hybrid beamforming, sum rate.

I. INTRODUCTION

In order to meet the great requirements of wireless communication, mmWave communication has recently gained considerable interest in both academia and industry as a promising candidate in the future cellular network [1], [2]. Due to mmWave has the large band of spectrum resource with 30GHz to 300GHz, it can provide higher data rate in wireless communication and be widely applied in the communication fields such as radar, backhaul, satellite communications, wireless local area network (WLAN), and wireless personal area network (WPAN) [3], [4]. However, the inherent shortcoming of the short wavelength of mmWave makes the signal easily blocked, absorbed, and scattered by obstacles in the transmission, and faces severe propagation path loss in high frequency communication [5], [6]. So, for obtaining the high antenna gain to compensate the propagation path loss using massive multiple-input multiple-output (MIMO), the characteristics of short wavelength of mm-wave can be utilized by mmWave communications, which are applied in line-of-sight (LoS) scenarios [7]–[9].

Faced with the complex environment or long distance communication, the relay assisted communications can be exploited between sources and destinations in mmWave

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

massive MIMO systems to alleviate the high propagation path loss and the serious intermittent blockage effect [10]. With the help of relay, the coverage region of transmission signals in mmWave system will be expanded, and the channel between each two communication nodes can be guaranteed to be LoS at the same time. Contrast to the amplify-and-forward (AF) relaying, the decode-and-forward (DF) relaying can overcome the noise accumulation by regenerating the data at relay [11]. Therefore, the DF mmWave massive MIMO relay system is more suitable for future wireless communication systems, and will be considered in this paper. Similar to the traditional mmWave massive MIMO system design, beamforming technology can also be applied in the mmWave massive MIMO relay system, so as to further improve the performance of the system [12]. However, due to the complex signal processing, the optimal beamforming design of the DF mmWave massive MIMO relay systems is an intractable problem.

It is known to all that the full-digital beamforming can obtain the optimal antenna gain, but it requires the same number of radio frequency (RF) chains as that of the antennas. Due to the high power consumption and production cost of mmWave technology, it is difficult to realize the full-digital beamforming for the mmWave massive MIMO architecture under the existing technical level [13]. To this end, the hybrid (analog/digital) beamforming which demands much fewer RF chains with the same performance to the full-digital beamforming can be adopted in the mmWave massive MIMO system [14]. The hybrid beamforming is a combination of antenna and digital signal processing technologies with the jointly design in the analog and digital domains, where the digital beamforming is achieved in the digital domain by baseband signal microprocessor, while the analog beamforming is realized in the analog domain by phase shifter connected between the RF chains and antennas [15].

The hybrid beamforming design can be implemented by two classical structures: sub-connected structure [16], [17], and full-connected structure [18], [19]. Where the sub-connected structure means that each RF chain is only connected to an independent subset of antennas, and the full-connected structure means that each RF chain is connected to all antennas. The differences between the two structures are that the sub-connected structures use less phase shifters than the full-connected structures with the same number of antennas and RF chains, but in performance, the sub-connected structures are slightly lower than the full-connected ones. The hybrid beamforming of relay system with full-connected structure have been designed in [20] and [21], but they are not based on mmWave large-scale antenna arrays. For the AF mmWave massive MIMO relay system with the full-connected structure, the joint optimal hybrid beamforming designs of downlink single-destination and multi-destination using matching pursuit (MP) algorithms have been studied in [22] and [23], respectively, and a separate design scheme of analog and digital beamforming of downlink multi-destination has been studied in [24]. For the AF mmWave massive MIMO relay system with the sub-connected structure, the hybrid beamforming design of downlink single-destination using iterative successive approximation (ISA) algorithm and alternating direction method of multipliers (ADMM) have been studied in [25] and [26], respectively. However, to the best of our knowledge, there is no existing work on the topic of hybrid beamforming design for DF mmWave massive multi-user (destination) MIMO (MU-MIMO) relay systems with the mixed structure. The mixed structure of the relay system is a compromise between the full-connected and the sub-connected relay system that the source and relay decoding are full-connected, while the relay forwarding and multi-destination are sub-connected. In the case of large antenna array, the mixed structure has the advantages of low computational complexity, low power consumption and simple wiring compared with the full-connected structure. Meanwhile, in terms of performance, the relay system with mixed structure is very close to the full-connected one. However, since the constraints imposed by the structure, designing the optimal hybrid beamforming with the mixed structure relay system is still a challenging problem.

In this paper, we propose an effective hybrid beamforming design scheme to fill the theoretical gap for the mixed structure in mmWave massive MU-MIMO relay system. The proposed optimal design scheme can also be expanded to the full-connected structure. In terms of scheme design, we aim to maximize the sum rate of overall system. Specifically, the main contributions are made by this paper as follows:

- For the mixed structure, the major problems in maximizing the sum rate of the overall system includes ten optimal elements involving power allocation to be solved, and the block-diagonal (BD) and constant-modulus constraints in the analog beamforming design caused by the characteristics of phase shifter and sub-connected structure. To reduce the computational complexity, we reformulate the original optimization problem as two single-hop mmWave sum-rate maximization subproblems according to the communication theory and structure characteristics of DF relay.
- To solve the sum-rate maximization subproblem from the source to the relay decoding, we propose the piecewise successive approximation method based on the criterion which jointly designs the analog and digital beamforming stages by trying to avoid the loss of information at each stage. For the sum-rate maximization subproblem from the relay forwarding to the destinations, the analog beamforming is designed by combining the idea of successive interference cancelation (SIC) with the piecewise successive approximation method, and the baseband BD scheme and waterfilling power allocation method are utilized to solve the digital beamforming and power allocation matrix, respectively.
- For the full-connected structure, the sum-rate maximization original problem of the full-connected structure can also be reformulated as two independent sum-rate

maximization subproblems, and the proposed piecewise successive approximation method is still applied to the hybrid beamforming design of the full-connected structure. Since the proposed method is based on the singular value decomposition, its performance is better than that of the MP method theoretically.
- The performance of the proposed hybrid beamforming design of the full-connected structure is very close to that of the full-digital beamforming. As for the mixed structure, slightly worse in performance but much lower in power consumption than the full-connected structure.

The remainder of this paper is organized as follows. Section II briefly introduces the channel model and system model of mmWave massive MU-MIMO relay system. The original problem of the system design is formulated and discussed in Section III. Section IV presents the proposed hybrid beamforming schemes for the mixed structure and the full-connected structure. The simulation results of the proposed two structures are shown in Section V. Finally, conclusions are drawn in Section VI.

Notations: Bold upper-case and lower-case letters represent matrices and column vectors, respectively; $(\cdot)^{-1}$, $(\cdot)^T$ and $(\cdot)^H$ denote inversion, transpose and conjugate transpose, respectively; The Frobenius norm of the matrix $\mathbf{A}$ is expressed as $\|\mathbf{A}\|_F$, $\mathbf{A}(i,j)$, $\mathbf{A}(:,j)$ and $\mathbf{A}(i,:)$ denote the (i,j)th complex element, jth column vector and ith row vector of the matrix $\mathbf{A}$, and $|\mathbf{A}(i,j)|$ is the amplitude; $A(i:j,:)$ and $A(:,i:j)$ represent the matrix of rows i through j in and the matrix of columns i through j in matrix $\mathbf{A}$, respectively; $\mathbb{E}[\cdot]$ indicates the expectation operator; The determinant and block diagonalization operation of a matrix are respectively expressed as $|\cdot|$ and $\text{blk}(\cdot)$; $\mathbf{I}_N$ is the identity matrix of size $N \times N$; $\mathcal{CN}(0,\sigma^2)$ represents the complex Gaussian distribution with mean 0 and the variance σ^2. $\angle\mathbf{A}$ denotes the operation of getting the angle of each entry in matrix $\mathbf{A}$; $\mathbb{D}^{l\times l}$ and $\mathbb{C}^{m\times n}$ describe a real diagonal matrix of dimension $l \times l$ and a complex matrix of dimension $m \times n$, respectively. $\text{tr}\{\cdot\}$ and $\text{Re}(\cdot)$ denote trace and real part of the matrix, respectively.

II. SYSTEM DESCRIPTION

This section briefly introduces the model of DF mmWave massive MU-MIMO relay system, which includes the relay system model with full-connected structure from the source to the destinations, and the relay system model with mixed structure which is the combination of full-connected structure from the source to the relay decoding and sub-connected structure from the relay forwarding to destinations. The channel model used in the system will be introduced firstly.

A. CHANNEL MODEL

It is well known that, unlike microwave wireless communication channels, mmWave channels have limited scattering features [27], and no longer obey the traditional

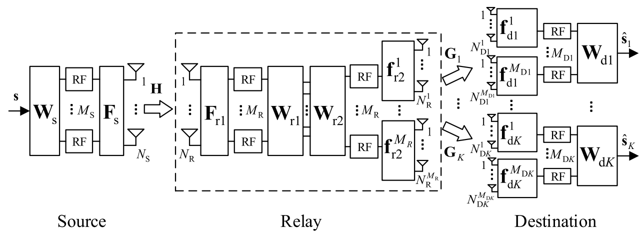

FIGURE 1. Block diagram of the multi-destination relay system with the mixed structure.

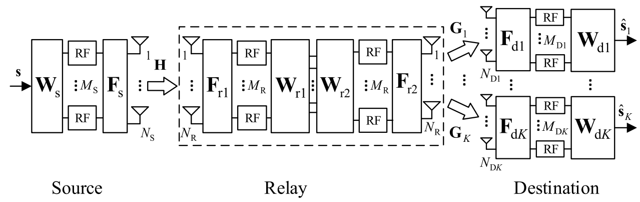

FIGURE 2. Block diagram of the multi-destination relay system with the full-connected structure.

Rayleigh fading distribution [28]. So, mmWave channels often have sparse structures which can be characterized by low-rank matrices [29]. In this paper, we adopt the geometric Saleh-Valenzuela channel model which is more suitable for mmWave communication. As shown in Figs. 1 and 2, we consider the DF mmWave massive MU-MIMO relay system without direct communication link between the source and the destinations, and all the nodes work in half-duplex mode. As can be seen, the system includes the channels $\mathbf{H}$ from the source to the relay and $\mathbf{G}_k$ from the relay to the kth destination. The uniform linear antenna arrays (ULAs) is assumed to be adopted in the narrowband mmWave channel model, where the LoS channels $\mathbf{H}$ and $\mathbf{G}_k$ can be expressed as [30]

$$\mathbf{H} = \sqrt{\frac{N_S N_R}{L_h}} \sum_{l=1}^{L_h} \alpha_l \mathbf{a}_l^R(\theta_l^R)\mathbf{a}_l^S(\phi_l^S)^H, \tag{1}$$

$$\mathbf{G}_k = \sqrt{\frac{N_R N_{Dk}}{L_g}} \sum_{l=1}^{L_g} \gamma_{l,k} \mathbf{a}_l^D(\theta_{l,k}^D)\mathbf{a}_l^R(\phi_{l,k}^R)^H, \tag{2}$$

where L_h and L_g are the number of propagation paths corresponding to finite scatters in the channel $\mathbf{H}$ and $\mathbf{G}_k$, α_l and $\gamma_{l,k}$ are the complex gain of the lth path in $\mathbf{H}$ and $\mathbf{G}_k$ that follow the $\mathcal{CN}(0,1)$ distribution. θ_l and ϕ_l denote the AoA and AoD of the lth path, respectively. The vectors $\mathbf{a}_l(\theta_l)$ and $\mathbf{a}_l(\phi_l)$ are the array response vectors at the receiver and the transmitter, repectively. Especially, when the ULAs is adopted, the array response vectors can be expressed as [31]

$$\mathbf{a}_l^S(\theta) = \frac{1}{\sqrt{N_S}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\theta)}, \ldots, e^{j(N_S-1)\frac{2\pi}{\lambda}d\sin(\theta)}\right]^T,$$
$$\mathbf{a}_l^R(\theta) = \frac{1}{\sqrt{N_R}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\theta)}, \ldots, e^{j(N_R-1)\frac{2\pi}{\lambda}d\sin(\theta)}\right]^T,$$
$$\mathbf{a}_l^D(\theta) = \frac{1}{\sqrt{N_D}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\theta)}, \ldots, e^{j(N_D-1)\frac{2\pi}{\lambda}d\sin(\theta)}\right]^T, \tag{3}$$

where $j = \sqrt{-1}$, d is the distance between adjacent antenna elements, and λ is the transmission wave length. To realize the hybrid beamforming, the channel state information (CSI) is assumed to be known at the source, relay, and the destinations.

B. SYSTEM MODEL

Figs. 1 and 2 show two different structures of DF mmWave MU-MIMO relay system. The relay with mixed and full-connected structures are shown in Figs. 1 and 2, respectively. The differences between the two structures are as follows. 1) For the relay forwarding part with sub-connected structure as shown in Fig.1 is equipped with M_R RF chains, and each RF chain is connected to a subset of N_R^n antennas, while all RF chains are assumed to serve K destinations with sub-connected structure simultaneously. The kth destination is equipped with M_{Dk} RF chains to support L_s data streams, and each RF chain is connected to a subset of N_{Dk}^m antennas. So the total number of antennas at the relay and each destination is $N_R = \sum_{n=1}^{M_R} N_R^n$ and $N_{Dk} = \sum_{m=1}^{M_{Dk}} N_{Dk}^m$, respectively. 2) For the relay forwarding part with full-connected structure as shown in Fig. 2, we assume it to communicate K destinations with full-connected structure simultaneously. At the source and relay decoding part, the mixed structure is the same as the full-connected structure, where the number of antennas and RF chains at the source are denoted as N_S and M_S, respectively.

It is worth nothing that, to achieve high antenna gain and reduce the hardware complexity of the massive MIMO relay system, the number of RF chains at source, relay and destinations are constrained by $M_S \ll N_S$, $M_R \ll N_R$, and $M_{Dk} \leq N_{Dk}$, respectively. The data streams of K destinations are transmitted simultaneously at the source, where the transmitted symbol vector for the kth destination is represented as $\mathbf{s}_k \in \mathbb{C}^{L_s \times 1}$. $\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \ldots, \mathbf{s}_K^T\right]^T \in \mathbb{C}^{KL_s \times 1}$ is the vector for the total transmitted symbols of K destinations, satisfying $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{1}{KL_s} I_{KL_s}$.

In the following, we will focus on the hybrid beamforming of relay system with mixed structure as shown in Fig. 1, the extension to the full-connected structure will be discussed in Section IV-C.

At the source, the transmitted symbol $\mathbf{s}$ with the total transmit power constraint P_s first passes through a diagonal power allocation matrix $\mathbf{P}_s \in \mathbb{D}^{KL_s \times KL_s}$ which distributes power to the $\mathbf{s}_k$ and satisfies $\|\mathbf{P}_s\|_F^2 = P_s$. Then the symbol $\mathbf{s}$ after power allocation is processed by digital beamforming $\mathbf{W}_s \in \mathbb{C}^{M_S \times KL_s}$ and followed by analog beamforming $\mathbf{F}_s \in \mathbb{C}^{N_S \times M_S}$, and finally sent into the channel $\mathbf{H} \in \mathbb{C}^{N_R \times N_S}$ between the source and the relay by the analog antennas. Since the analog beamforming matrix $\mathbf{F}_s$ is constrained by the constant amplitude, its entries have equal magnitudes, i.e., $|\mathbf{F}_s(i,j)| = \frac{1}{\sqrt{N_S}}$. In addition, the digital beamforming matrix $\mathbf{W}_s$ is normalized to satisfy $\|\mathbf{F}_s\mathbf{W}_s\|_F^2 = KL_s$, because the hybrid beamforming at the source should meet the total transmitted power constraint.

At the relay decoding, the transmitted signal from source received by analog antennas of relay can be expressed as

$$\mathbf{y}_r = \mathbf{H}\mathbf{F}_s\mathbf{W}_s\mathbf{P}_s\mathbf{s} + \mathbf{n}_r \in \mathbb{C}^{N_R \times 1}. \tag{4}$$

Subsequently, the symbol $\mathbf{y}_r$ is processed by analog combining $\mathbf{F}_{r1} \in \mathbb{C}^{N_R \times M_R}$ and followed by digital combining $\mathbf{W}_{s1} \in \mathbb{C}^{M_R \times KL_s}$. Similarly, due to the characteristics of the phase shifters, the entries of analog combining matrix $\mathbf{F}_{r1}$ satisfy $|\mathbf{F}_{r1}(i,j)| = \frac{1}{\sqrt{N_R}}$. Finally, the decoded signal at the relay is given by

$$\tilde{\mathbf{s}} = \mathbf{W}_{s1}^H\mathbf{F}_{r1}^H\mathbf{y}_r = \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{H}\mathbf{F}_s\mathbf{W}_s\mathbf{P}_s\mathbf{s} + \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{n}_r, \tag{5}$$

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 σ_R^2 (i.e., $\mathbf{n}_r \sim \mathcal{CN}(0, \sigma_R^2)$).

At the relay forwarding, the decoded symbol $\tilde{\mathbf{s}}$ also firstly goes through the diagonal power allocation matrix $\mathbf{P}_r \in \mathbb{D}^{KL_s \times KL_s}$ satisfying $\|\mathbf{P}_r\|_F^2 = P_r$. Then the symbol $\tilde{\mathbf{s}}$ after power allocation is processed by digital beamforming $\mathbf{W}_{r2} \in \mathbb{C}^{M_R \times KL_s}$ and analog beamforming $\mathbf{F}_{r2} \in \mathbb{C}^{N_R \times M_R}$ in turn, and sent into the kth sub-channel $\mathbf{G}_k \in \mathbb{C}^{N_R \times N_{Dk}}$ between the relay and the kth destination. The entire channel is denoted as $\mathbf{G} = [\mathbf{G}_1^H, \mathbf{G}_2^H, \ldots, \mathbf{G}_K^H]^H$. The transmitted total power constraint at the relay is enforced by normalizing $\mathbf{W}_{r2}$ such that $\|\mathbf{F}_{r2}\mathbf{W}_{r2}\|_F^2 = KL_s$.

At the kth destination, the received signal from the relay can be expressed as

$$\tilde{\mathbf{y}}_k = \mathbf{G}_k\mathbf{F}_{r2}\mathbf{W}_{r2}\mathbf{P}_r\tilde{\mathbf{s}} + \mathbf{n}_{dk} \in \mathbb{C}^{N_{Dk} \times 1}. \tag{6}$$

Then, after combining by analog matrix $\mathbf{F}_{dk} \in \mathbb{C}^{N_{Dk} \times M_{Dk}}$ and digital matrix $\mathbf{W}_{dk} \in \mathbb{C}^{M_{Dk} \times L_s}$, the detected symbol of the kth destination is given as

$$\begin{aligned}\hat{\mathbf{s}}_k &= \mathbf{W}_{dk}^H\mathbf{F}_{dk}^H\tilde{\mathbf{y}}_k \\ &= \mathbf{W}_{dk}^H\mathbf{F}_{dk}^H\mathbf{G}_k\mathbf{F}_{r2}\mathbf{W}_{r2}\mathbf{P}_r\tilde{\mathbf{s}} + \mathbf{W}_{dk}^H\mathbf{F}_{dk}^H\mathbf{n}_{dk},\end{aligned} \tag{7}$$

where $\mathbf{n}_{dk} \in \mathbb{C}^{N_{Dk} \times 1}$ is the noise vector at the kth destination, with elements following the i.i.d complex Gaussian distribution (i.e., $\mathbf{n}_{dk} \sim \mathcal{CN}(0, \sigma_D^2)$). Due to the particularity of sub-connected structure, the analog beamforming $\mathbf{F}_{r2}$ at the relay forwarding and the analog combining $\mathbf{F}_d$ at the destination are constrained to be block-diagonal as follows:

$$\mathbf{F}_{r2} = \mathrm{blk}(\mathbf{f}_{r2}^1, \mathbf{f}_{r2}^2, \ldots, \mathbf{f}_{r2}^{M_R}), \tag{8}$$

$$\mathbf{F}_d = \mathrm{blk}(\mathbf{F}_{d1}, \mathbf{F}_{d2}, \ldots, \mathbf{F}_{dK}), \tag{9}$$

where $\mathbf{f}_{r2}^n \in \mathbb{C}^{N_R^n \times 1}(n = 1, 2, \ldots, M_R)$, $\mathbf{F}_{dk} \in \mathbb{C}^{N_{Dk} \times M_{Dk}}(k = 1, 2, \ldots, K)$, and $\mathbf{F}_{dk} = \mathrm{blk}(\mathbf{f}_{dk}^1, \ldots, \mathbf{f}_{dk}^{M_{Dk}})$. The corresponding non-zero elements in the $\mathbf{F}_{r2}$ and $\mathbf{F}_d$ satisfy the constant-modulus constraints as follows:

$$\left|\mathbf{f}_{r2}^n(i)\right| = \frac{1}{\sqrt{N_R^n}}, \quad i = 1, 2, \ldots, N_R^n, \tag{10}$$

$$\left|\mathbf{f}_{dk}^m(j)\right| = \frac{1}{\sqrt{N_{Dk}^m}}, \quad j = 1, 2, \ldots, N_{Dk}^m, \tag{11}$$

where $m = 1, 2, \ldots, M_{Dk}$.

The baseband channel for the kth destination can be defined as $\bar{\mathbf{G}}_k = \mathbf{F}_{dk}^H \mathbf{G}_k \mathbf{F}_{r2}$. The SNR $= \frac{P_r}{\sigma_D^2}$ indicates the transmission signal-to-noise ratio (SNR) from the relay to the kth destination.

From the above, the transmitted ith data stream of the kth destination can be further expressed as

$$\begin{aligned}\hat{s}_{k_i} &= \mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,k_i)\sqrt{P_{rk_i}}\tilde{s}_{k_i}\\ &\quad + \sum_{j=1,j\neq i}^{L_s}\mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,k_j)\sqrt{P_{rk_j}}\tilde{s}_{k_j}\\ &\quad + \sum_{m=1,m\neq k}^{K}\sum_{l=1}^{L_s}\mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,m_l)\sqrt{P_{rm_l}}\tilde{s}_{m_l}\\ &\quad + \mathbf{W}_{dk}^H(i,:)\mathbf{F}_{dk}^H\mathbf{n}_{dk}, \qquad (12)\end{aligned}$$

where $k_i = (k-1)L_s + i$, $\tilde{s}_{k_i}$ represents the k_ith entry of the decoded signal $\tilde{\mathbf{s}}$ at the relay in (5), and $\sqrt{P_{rk_i}}$ is the power allocated to the ith data stream of the kth destination. The first term on the right side of the equation in (12) indicates the desired signal, and the other three terms, respectively, represent inner-destination interference, inter-destination interference, and noise. When the Gaussian symbols are assumed to be transmitted, the sum-rate of the transmitted channel between the relay and the destination can be formulated as

$$R = \sum_{k=1}^{K}\sum_{i=1}^{L_s}\log_2(1+\mathrm{SINR}_{k_i}), \qquad (13)$$

where SINR_{k_i} is the signal to interference to noise ratio (SINR) of the signal $\hat{s}_{k_i}$, which can be calculated by the ratio of the desired signal energy of the first term in (12) to the interference plus noise energy of the remaining terms. The SINR_{k_i} is formulated as follows:

$$\begin{aligned}\mathrm{SINR}_{k_i} &= \frac{\tilde{S}_{k_i}}{\tilde{I}_{k_i}+\tilde{N}_{k_i}}\\ \tilde{S}_{k_i} &= \left|\mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,k_i)\sqrt{P_{rk_i}}\right|^2\\ \tilde{I}_{k_i} &= \sum_{j=1,j\neq i}^{L_s}\left|\mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,k_j)\sqrt{P_{rk_j}}\right|^2\\ &\quad + \sum_{m=1,m\neq k}^{K}\sum_{l=1}^{L_s}\left|\mathbf{W}_{dk}^H(i,:)\bar{\mathbf{G}}_k\mathbf{W}_{r2}(:,m_l)\sqrt{P_{rm_l}}\right|^2\\ \tilde{N}_{k_i} &= \sigma_D^2\left\|\mathbf{W}_{dk}^H(i,:)\mathbf{F}_{dk}^H\mathbf{n}_{dk}\right\|_{\mathrm{F}}^2, \qquad (14)\end{aligned}$$

where $k \in \{1,2,\ldots,K\}$, $i \in \{1,2,\ldots,L_s\}$.

For the hybrid beamforming design of relay system with full-connected structure, the analog beamforming is also constrained by the constant-modulus and the transmitted total power constraints, and the constraint expression is the same as that of the above mixed structure. However, the analog beamforming in the full-connected structure is not constrained to be block-diagonal constraints in (10) and (11), so the hybrid beamforming design for the full-connected structure is easier than that for the mixed structure.

III. PROBLEM FORMATION

This section discusses the hybrid beamforming design for the DF mmWave massive MU-MIMO relay system with the mixed structure from the source to the destinations. The achievable sum rate and the mean squared error (MSE) are two optimization objectives for the beamforming design DF problems [25]. Since the former is an important performance evaluation standard in mmWave system, the design goal aims to maximize the sum rate for hybrid beamforming. As the analysis in the previous section, the sum rate problem is formulated as follows:.

$$\begin{aligned}&(\mathcal{P})\ (\mathbf{W}_s,\mathbf{F}_s,(\mathbf{F}_{r1},\mathbf{W}_{r1},\mathbf{W}_{r2},\mathbf{F}_{r2}),\mathbf{F}_d,\mathbf{W}_d,\mathbf{P}_s,\mathbf{P}_r)\\ &\quad = \underset{(\mathbf{W}_s,\mathbf{F}_s,(\mathbf{F}_{r1},\mathbf{W}_{r1},\mathbf{W}_{r2},\mathbf{F}_{r2}),\mathbf{F}_d,\mathbf{W}_d,\mathbf{P}_s,\mathbf{P}_r)}{\arg\max} R\\ &s.t.\ |\mathbf{F}_s(i,j)| = \frac{1}{\sqrt{N_S}},\ |\mathbf{F}_{r1}(i,j)| = \frac{1}{\sqrt{N_R}},\ (10),(11),(*)\\ &\quad (8),(9),\ \mathbf{W}_{r2} = \begin{bmatrix}\mathbf{W}_{r2}^1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \mathbf{W}_{r2}^K\end{bmatrix},\\ &\quad \mathbf{W}_d = \begin{bmatrix}\mathbf{W}_{d1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \mathbf{W}_{dK}\end{bmatrix},\\ &\quad \|\mathbf{P}_s\|_{\mathrm{F}}^2 = P_s,\ \|\mathbf{P}_r\|_{\mathrm{F}}^2 = P_r, \qquad (15)\end{aligned}$$

where $(*)$ represents the constant-modulus constraints. Due to the original problem (15) is non-convex, it is intractable to find the global optimal solution directly [32].

According to the communication mode of the DF relay, the entire system can be decomposed as two relatively independent MIMO cascade sub-systems. Therefore, the overall sum rate of the two sub-systems is maximizing the minimum value between the transmission rate of the sub-system from the source to the relay (R_1) and the sub-system from the relay to the destinations (R_2) without considering the direct link from the source to the destinations [33].

With the help of above analysis, the sum rate of whole system can be expressed as:

$$R = 0.5\min(R_1, R_2). \qquad (16)$$

Since the communication of the relay system is completed in two time slots, the transmitted information rate is half of the overall sum rate compared with a relay-free scenario [34].

According to (12)–(14) in the previous section, we can get the expression of R_1 and R_2 as follows

$$R_1 = \log_2(1+\mathrm{SINR}_{SR}), \qquad (17)$$

$$R_2 = \sum_{k=1}^{K}\sum_{i=1}^{L_s}\log_2(1+\mathrm{SINR}_{RD}^{k_i}), \qquad (18)$$

where SINR_{SR} and SINR_{RD} denote the SINR from source to relay decoding and relay forwarding to destinations, respectively.

To sum up, we reformulate the original optimal problem as two sum-rate maximum subproblems. Then, the detailed expressions of these two subproblems can be given by

$$(\mathcal{P}_1)\ (\mathbf{W}_s, \mathbf{F}_s, \mathbf{F}_{r1}, \mathbf{W}_{r1}, \mathbf{P}_s) = \underset{(\mathbf{W}_s,\mathbf{F}_s,\mathbf{F}_{r1},\mathbf{W}_{r1},\mathbf{P}_s)}{\arg\max} R_1$$
$$s.t.\ |\mathbf{F}_s(i,j)| = \frac{1}{\sqrt{N_S}}, |\mathbf{F}_{r1}(i,j)| = \frac{1}{\sqrt{N_R}}, (*)$$
$$\|\mathbf{P}_s\|_F^2 = P_s, \tag{19}$$

$$(\mathcal{P}_2)\ (\mathbf{W}_{r2}, \mathbf{F}_{r2}, \mathbf{F}_d, \mathbf{W}_d, \mathbf{P}_r) = \underset{(\mathbf{W}_{r2},\mathbf{F}_{r2},\mathbf{F}_d,\mathbf{W}_d,\mathbf{P}_r)}{\arg\max} R_2$$
$$s.t.\ (8), (9), (10), (11),$$
$$\mathbf{W}_d = \begin{bmatrix} \mathbf{W}_{d1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_{dK} \end{bmatrix},$$
$$\mathbf{W}_{r2} = \begin{bmatrix} \mathbf{W}_{r2}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_{r2}^K \end{bmatrix},$$
$$\|\mathbf{P}_r\|_F^2 = P_r, \tag{20}$$

It is observed from (19) and (20) that the two reformulated subproblems $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ are still non-convex. For $(\mathcal{P}_1)$, there is no interference between the destinations due to the mode from source to relay decoding can be considered as a point-to-point single-destination MIMO system. So R_1 can be expressed by the mutual information between $\mathbf{s}$ and $\tilde{\mathbf{s}}$ as follows

$$\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) = \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{1}{\sigma_R^2}\mathbf{R}_{n_r}^{-1}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right|\right), \tag{21}$$

where $\mathbf{R}_{n_r} = \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{F}_{r1}\mathbf{W}_{r1}$, $\tilde{\mathbf{H}} = \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{H}\mathbf{F}_s\mathbf{W}_s\mathbf{P}_s$. To remove the constraint $(*)$ in (19), we assume the optimal solution derived from the unconstrained case in analog beamforming is $\mathbf{F}_s^{opt}$. Then by calculating the minimum mean squared error (MMSE) between $\mathbf{F}_s^{opt}$ and the optimal solution $\mathbf{F}_s$ in the constrained case, the conclusion that the optimal $\mathbf{F}_s$ shares the phase of the corresponding element of $\mathbf{F}_s^{opt}$ can be obtained [35]. Thus, the non-convexity of the constraint $(*)$ in (19) is eliminated, and the optimal solution $\mathbf{F}_s$ in the constrained case is obtained as

$$\mathbf{F}_s = \frac{1}{\sqrt{N_S}}e^{j\angle\mathbf{F}_s^{opt}}. \tag{22}$$

For the design of power allocation matrix, the classic waterfilling power allocation method is used to get it. In this way, solving the non-convex subproblem $(\mathcal{P}_1)$ is transformed into the convex one.

Similarly, the non-convex subproblem $(\mathcal{P}_2)$ can be converted to the convex in the same way. To eliminate the interference between multiple destinations, we use the baseband BD scheme described in [36] to design the digital beamforming. Therefore, the sum rate R_2 expressed by the mutual information of $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ is given by

$$\mathcal{I}(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \sum_{k=1}^{K}\log_2\left(\left|\mathbf{I}_{L_s} + \frac{1}{\sigma_D^2}\mathbf{R}_{ndk}^{-1}\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^H\right|\right)$$
$$= \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{1}{\sigma_D^2}\mathbf{R}_{nd}^{-1}\tilde{\mathbf{G}}\tilde{\mathbf{G}}^H\right|\right), \tag{23}$$

where $\mathbf{R}_{nd} = \mathbf{W}_d^H\mathbf{F}_d^H\mathbf{F}_d\mathbf{W}_d$, $\tilde{\mathbf{G}}_k = \mathbf{W}_{dk}^H\mathbf{F}_{dk}^H\mathbf{G}_k\mathbf{F}_{r2}\mathbf{W}_{r2}\mathbf{P}_r$, $\tilde{\mathbf{G}} = \mathbf{W}_d^H\mathbf{F}_d^H\mathbf{G}\mathbf{F}_{r2}\mathbf{W}_{r2}\mathbf{P}_r$.

Up to now, one can understand that the original hybrid beamforming problem $(\mathcal{P})$ has been decomposed into two subproblems $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$, and each subproblem has been transformed into a convex one. In the next section, we will present the specific design algorithm to solve these two subproblems.

IV. HYBRID BEAMFORMING DESIGN OF RELAY SYSYTEM WITH PERFECT CSI

In this section, we will propose the algorithm for solving the two subproblems to maximize the sum rate of the overall system. As the theoretical solution, we assume that the perfect CSI is known at source, relay, and destinations, and ignore the constant-modulus constraints. Then the optimal hybrid beamforming design of DF mmWave massive MU-MIMO relay system with mixed structure is presented.

A. HYBRID BEAMFORMING DESIGN FROM SOURCE TO RELAY DECODING

For the subproblem $(\mathcal{P}_1)$, it can be considered as a massive single-destination MIMO system with full-connected structure. Inspired by the joint hybrid beamforming design criterion presented in [35], which try to avoid the loss of information at the analog and digital stages, the piecewise successive approximation method is proposed to solve $(\mathcal{P}_1)$. When the number of antennas at the source approaches infinity, the sum rate R_1 can also reach the upper bound of the capacity of the channel $\mathbf{H}$ with equal power allocation.

First, we optimize the analog beamforming matrix $\mathbf{F}_s$ and analog combining matrix $\mathbf{F}_{r1}$ while assuming the digital beamforming matrix $\mathbf{W}_s$ and the digital combining matrix $\mathbf{W}_{r1}$ are fixed. The singular value decomposition (SVD) of $\mathbf{H}$ is defined as $\frac{1}{\sqrt{N_S}}\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$. According to (22), the analog combining matrix $\mathbf{F}_{r1}$ is obtained by $\mathbf{F}_{r1} = \frac{1}{\sqrt{N_R}}e^{j\angle\mathbf{U}(:,1:M_R)}$. When the N_S approaches infinity, the power allocation obtained by using waterfilling will be approximately performed by equal power allocation, i.e., $\mathbf{P}_s \approx \sqrt{\frac{P_s}{KL_s}}\mathbf{I}_{KL_s}$. Therefore, under the assumption that the analog combining matrix $\mathbf{F}_{r1}$ is known at the relay decoding, (21) is further deduced as

$$\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) \approx \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_s}{\sigma_R^2 KL_s}\mathbf{F}_s^H\mathbf{H}_{comp}^H\mathbf{R}_{n_r}^{-1}\mathbf{H}_{comp}\mathbf{F}_s\right|\right), \tag{24}$$

where $\mathbf{H}_{comp} = \mathbf{F}_{r1}^H\mathbf{H}$. In addition, $\mathbf{W}_s$ and $\mathbf{W}_{r1}$ are the unitary matrices,[1] we can obtain

$$\mathbf{R}_{n_r} = \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{F}_{r1}\mathbf{W}_{r1} = \mathbf{I}_{KL_s}, \tag{25}$$

[1]The matrices derived from SVD are unitary matrices, the solving methods of $\mathbf{W}_s$ and $\mathbf{W}_{r1}$ will be introduced in the following subsection.

then (24) is rewritten as

$$\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) \approx \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_s}{\sigma_R^2 KL_s}\mathbf{F}_s^H\mathbf{H}_{comp}^H\mathbf{H}_{comp}\mathbf{F}_s\right|\right). \quad (26)$$

Define the SVD of $\mathbf{H}_{comp}$ as $\frac{1}{\sqrt{N_S}}\mathbf{H}_{comp} = \mathbf{U}_{comp}\Sigma_{comp}\mathbf{V}_{comp}^H$. Then, (26) can be further written as

$$\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) \approx \log_2\left(\left|\mathbf{I}_{KL_S} + \frac{P_S N_S}{\sigma_R^2 KL_s}\mathbf{F}_s^H\mathbf{V}_{comp}\Sigma_{comp}^H\Sigma_{comp}\mathbf{V}_{comp}^H\mathbf{F}_s\right|\right). \quad (27)$$

It can be seen from the expression as (27) that the objective function $\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}})$ is maximised when $\mathbf{F}_s = \mathbf{V}_{comp}(:,1:M_S)$, but $\mathbf{V}_{comp}(:,1:M_S)$ is not satisfy constant-modulus constraints. As in prior work, we use the Frobenius norm to compute the distance between $\mathbf{F}_s$ and $\mathbf{V}_{comp}(:,1:M_S)$. Then, the objective function is converted as

$$\min_{\mathbf{F}_s}\left\|\mathbf{V}_{comp}(:,1:M_S) - \mathbf{F}_s\right\|_F^2. \quad (28)$$

Let $\hat{\mathbf{V}}_{comp} = \mathbf{V}_{comp}(:,1:M_S)$, the objective function (28) can be rewritten as

$$\begin{aligned}\left\|\hat{\mathbf{V}}_{comp} - \mathbf{F}_s\right\|_F^2 &= \mathrm{tr}\left\{\left(\hat{\mathbf{V}}_{comp} - \mathbf{F}_s\right)^H\left(\hat{\mathbf{V}}_{comp} - \mathbf{F}_s\right)\right\}\\ &= 2M_S - \mathrm{tr}\left\{2\mathrm{Re}\left(\mathbf{F}_s^H\hat{\mathbf{V}}_{comp}\right)\right\}\\ &= 2M_S\\ &\quad - 2\sum_{n=1}^{M_S}\sum_{m=1}^{N_S}\mathrm{Re}\left\{\left|\mathbf{F}_s(m,n)\right|\left|\hat{\mathbf{V}}_{comp}(m,n)\right|e^{j\varphi(m,n)}\right\}.\end{aligned} \quad (29)$$

where $\varphi(m,n) = \angle\mathbf{F}_s(m,n) - \angle\hat{\mathbf{V}}_{comp}(m,n)$. It is clear that when $\varphi(m,n) = 0$, i.e., $\angle\mathbf{F}_s(m,n) = \angle\hat{\mathbf{V}}_{comp}(m,n)$, the objective function is minimized. This is consistent with the conclusion of (22), therefore, the optimal analog precoding matrix can be obtained as $\mathbf{F}_s = \frac{1}{\sqrt{N_S}}e^{j\angle\mathbf{V}_{comp}(:,1:M_S)}$.

Similarly, when the analog beamforming matrix $\mathbf{F}_s$ and combining matrix $\mathbf{F}_{r1}$ is known at the source and the relay decoding, respectively, the SVD of the baseband channel matrix $\bar{\mathbf{H}} = \mathbf{F}_{r1}^H\mathbf{H}\mathbf{F}_s$ is defined as $\frac{1}{\sqrt{N_S}}\bar{\mathbf{H}} = \bar{\mathbf{U}}\bar{\Sigma}\bar{\mathbf{V}}^H$, and let $\mathbf{W}_{r1} = \bar{\mathbf{U}}(:,1:KL_s)$. For N_S is large, (21) is rewritten as

$$\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) \approx \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_s N_S}{\sigma_R^2 KL_s}\mathbf{W}_s^H\bar{\mathbf{H}}_{comp}^H\bar{\mathbf{H}}_{comp}\mathbf{W}_s\right|\right). \quad (30)$$

It is clear that (30) and (26) are similar in nature, so the optimal $\mathbf{W}_s$ of maximizing $\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}})$ is equivalent to the first KL_s columns of the right singular vectors (sorted with the corresponding singular value by descending order) of $\frac{1}{\sqrt{N_S}}\bar{\mathbf{H}}_{comp}$, i.e., $\mathbf{W}_s = \bar{\mathbf{V}}_{comp}(:,1:KL_s)$. Since the relay receiver can be regarded as a single user, there is no interference between users, and the loss of information is reduced compared with the traditional SVD method. Therefore, this method can achieve optimal digital beamforming design.

Algorithm 1 Hybrid Beamforming Design From Source to Relay Decoding

Initialize: $\mathbf{H}$;
Analog stage:
Compute $\mathbf{F}_{r1} = \frac{1}{\sqrt{N_R}}e^{j\angle\mathbf{U}(:,1:M_R)}$, where $\mathbf{U}$ is derived by $\frac{1}{\sqrt{N_S}}\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$;
Obtain the composite channel $\mathbf{H}_{comp} = \mathbf{F}_{r1}^H\mathbf{H}$;
Compute $\mathbf{F}_s = \frac{1}{\sqrt{N_S}}e^{j\angle\mathbf{V}_{comp}(:,1:M_S)}$, where $\mathbf{V}_{comp}$ is derived by $\frac{1}{\sqrt{N_S}}\mathbf{H}_{comp} = \mathbf{U}_{comp}\Sigma_{comp}\mathbf{V}_{comp}^H$;
End stage
Obtain the baseband channel $\bar{\mathbf{H}} = \mathbf{F}_{r1}^H\mathbf{H}\mathbf{F}_s$;
Digital stage:
Get $\mathbf{W}_{r1} = \bar{\mathbf{U}}(:,1:KL_s)$, where $\bar{\mathbf{U}}$ is derived by $\frac{1}{\sqrt{N_S}}\bar{\mathbf{H}} = \bar{\mathbf{U}}\bar{\Sigma}\bar{\mathbf{V}}^H$;
Obtain the composite channel $\bar{\mathbf{H}}_{comp} = \mathbf{W}_{r1}^H\bar{\mathbf{H}}$;
Get $\mathbf{W}_s = \bar{\mathbf{V}}_{comp}(:,1:KL_s)$, where $\bar{\mathbf{V}}_{comp}$ is derived by $\frac{1}{\sqrt{N_S}}\bar{\mathbf{H}}_{comp} = \bar{\mathbf{U}}_{comp}\bar{\Sigma}_{comp}\bar{\mathbf{V}}_{comp}^H$;
End stage
Obtain the total equivalent channel $\mathbf{H}_{total} = \mathbf{W}_{r1}^H\mathbf{F}_{r1}^H\mathbf{H}\mathbf{F}_s\mathbf{W}_s$;
Compute $\mathbf{P}_s$ by using waterfilling power allocation of the total equivalent channel $\mathbf{H}_{total}$;
Output: $\mathbf{W}_{r1}, \mathbf{F}_{r1}, \mathbf{F}_s, \mathbf{W}_s, \mathbf{P}_s$.

To sum up, all the obtained beamforming solutions are substituted into (21), which is expressed as

$$\begin{aligned}\mathcal{I}(\mathbf{s};\tilde{\mathbf{s}}) &\approx \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_s\Sigma\Sigma^H}{N_S\sigma_R^2 KL_s}\right|\right)\\ &= \sum_{i=1}^{M_S}\log_2\left(1 + \frac{P_S\gamma_i^2}{N_S\sigma_R^2 KL_s}\right),\end{aligned} \quad (31)$$

where γ_i denotes the ith singular value of $\mathbf{H}$. The expression of (31) is consistent with the upper bound of the capacity of the single-user channel with equal power allocation proposed in [35]. Therefore, the solutions of the subproblem ($\mathcal{P}_1$) is valid, and the overall procedure can be summarized as the following **Algorithm 1**.

B. HYBRID BEAMFORMING DESIGN FROM RELAY FORWARDING TO DESTINATIONS

For the subproblem ($\mathcal{P}_2$), it can be considered as a massive MU-MIMO system with sub-connected structure, and has been converted to a convex problem in the discussion of Section III. To tackle the block-diagonal constraint (8) and (9) in ($\mathcal{P}_2$), the method of SIC proposed in [37] will be utilized for reference. Since the digital beamforming is designed by using the baseband BD scheme, the digital beamforming matrix $\mathbf{W}_{r2}$ and digital combining matrix $\mathbf{W}_d$ are both unitary matrix, i.e., $\mathbf{W}_{r2}^H\mathbf{W}_{r2} = \mathbf{I}_{KL_s}$, $\mathbf{W}_d^H\mathbf{W}_d = \mathbf{I}_{KL_s}$.

Then $\mathbb{E}\left\{\left\|\mathbf{W}_{r2}\mathbf{P}_r\tilde{\mathbf{s}}\right\|_F^2\right\} = \frac{P_r}{KL_s}$, and (23) can be simplified as

$$\mathcal{I}(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\mathbf{F}_d^H\mathbf{G}\mathbf{F}_{r2}\mathbf{F}_{r2}^H\mathbf{G}^H\mathbf{F}_d\right|\right). \quad (32)$$

Obviously, maximizing the mutual information is only related to $\mathbf{F}_d$ and $\mathbf{F}_{r2}$ in (32). Assuming $\mathbf{F}_\mathrm{d}$ is fixed, (32) is a function of the variable $\mathbf{F}_{r2}$ and expressed as

$$\mathcal{I}_R(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\mathbf{G}_d\mathbf{F}_{r2}\mathbf{F}_{r2}^H\mathbf{G}_d^H\right|\right), \quad (33)$$

where $\mathbf{G}_d = \mathbf{F}_d^H\mathbf{G}$. Then for fixed $\mathbf{F}_{r2}$, (32) becomes a function of variable $\mathbf{F}_d$ and can be expressed as

$$\mathcal{I}_D(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\mathbf{F}_d^H\mathbf{G}_r\mathbf{G}_r^H\mathbf{F}_d\right|\right), \quad (34)$$

where $\mathbf{G}_\mathrm{r} = \mathbf{G}\mathbf{F}_{r2}$. According to the properties of determinant: $\det(\mathbf{I}+\mathbf{XY}) = \det(\mathbf{I}+\mathbf{YX})$, when the product of the two matrices $\mathbf{X}$ and $\mathbf{Y}$ is a square matrix, (34) can be converted to

$$\mathcal{I}_D(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \log_2\left(\left|\mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\bar{\mathbf{G}}_r\mathbf{F}_d\mathbf{F}_d^H\bar{\mathbf{G}}_r^H\right|\right), \quad (35)$$

where $\bar{\mathbf{G}}_r = \mathbf{G}_r^H$. Compared (33) with (35), they have the similar form. Therefore, the $\mathbf{F}_d$ and $\mathbf{F}_{r2}$ can be solved by adopting the same method. For simplicity, we only analyze the solution of $\mathbf{F}_d$.

Utilizing the method of SIC proposed in [37] and assuming the same number of RF chains for each destination (i.e., $M_{Dk} = \frac{M_D}{K}, k = 1, \ldots, K$), (35) can be rewritten as

$$\mathcal{I}_D(\tilde{\mathbf{s}};\hat{\mathbf{s}}) = \sum_{k=1}^{K}\sum_{n=1}^{M_{DK}}\log_2\left(\left|1 + \frac{P_r}{\sigma_D^2 KL_s}\mathbf{f}_{dk}^{nH}\bar{\mathbf{G}}_r^H\mathbf{T}_{n-1}^{-1}\bar{\mathbf{G}}_r\mathbf{f}_{dk}^n\right|\right), \quad (36)$$

where $\mathbf{f}_{dk}^n$ is the nth column of the analog combining $\mathbf{F}_{dk}$ for the kth destination, i.e., $\mathbf{F}_{dk} = \mathrm{blk}\left[\mathbf{f}_{dk}^1, \mathbf{f}_{dk}^2, \ldots, \mathbf{f}_{dk}^{M_{DK}}\right]$, $k = 1, \ldots, K$, and $\mathbf{T}_n = \mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\bar{\mathbf{G}}_r\mathbf{f}_{dk}^n\mathbf{f}_{\mathrm{d}k}^{nH}\bar{\mathbf{G}}_r^H$, $\mathbf{T}_0 = \mathbf{I}_{KL_s}$. We define the similar term in the iteration as $\mathbf{G}_{n-1} = \bar{\mathbf{G}}_r^H\mathbf{T}_{n-1}^{-1}\bar{\mathbf{G}}_r$, and assume that each destination has the same number of RF chains and antennas, i.e., $N_{Dk}^m = \frac{N_D}{KM_{DK}} = N_{DK}, m = 1, \ldots, M_{DK}$. Therefore, only the first N_{DK} elements in $\mathbf{f}_{dk}^n$ are non-zero, and the sub-matrix $\mathbf{S}$ of $\mathbf{G}_{n-1}$ in each iteration is $\mathbf{G}_{n-1}(N_{DK}\times(n-1)+1 : N_{DK}\times n, N_{DK}\times(n-1)+1 : N_{DK}\times n)$. It is known from (22) that the solution $\mathbf{f}_{dk}^n$ shares the phase of the corresponding element of the maximum right singular vector $\mathbf{v}_1$ of $\mathbf{S}$, so each analog solution vector can be expressed as

$$\mathbf{f}_n = \frac{1}{\sqrt{N_{DK}}}e^{j\angle\mathbf{v}_1}. \quad (37)$$

In summary, the complete procedure of hybrid beamforming design for the subproblem $(\mathcal{P}_2)$ is provided as **Algorithm 2**.

Algorithm 2 Hybrid Beamforming Design From Relay Forwarding to Destinations

1: **Initialize**: $\mathbf{G}, P_r, \sigma_D^2, N_{LOOP}$;
2: **Analog stage**:
3: Randomly generate the initial matrix $\mathbf{F}_{r2}$ (satisfied with the constant-modulus constraint and diagonalization structure);
4: **for** $m = 1$ to N_{LOOP} **do**
5: Compute $\bar{\mathbf{G}}_r = \mathbf{F}_{r2}^H\mathbf{G}^H$;
6: **for** $n = 1$ to KM_{DK} **do**
7: Compute $\mathbf{G}_{n-1} = \bar{\mathbf{G}}_r^H(\mathbf{I}_{KL_s} + \frac{P_r}{\sigma_D^2 KL_s}\bar{\mathbf{G}}_r\mathbf{P}_n\mathbf{P}_n^H\bar{\mathbf{G}}_r^H)^{-1}\bar{\mathbf{G}}_r$;
8: Acquire the effective sub-matrix of $\mathbf{G}_{n-1}$: $\mathbf{S} = \mathbf{G}_{n-1}(N_{DK}\times(n-1)+1 : N_{DK}\times n, N_{DK}\times(n-1)+1 : N_{DK}\times n)$;
9: Obtain the maximum right singular vector $\mathbf{v}_1$ of $\mathbf{S}$;
10: Compute $\mathbf{f}_n = \frac{1}{\sqrt{N_{DK}}}e^{j\angle\mathbf{v}_1}$;
11: Compute $\mathbf{p}_n = [\mathbf{0}_{1\times N_{DK}(n-1)}, \mathbf{f}_n^H, \mathbf{0}_{1\times N_{DK}(KM_{DK}-1)}]^H$;
12: **end for**
13: Obtain $\mathbf{F}_{r2} = \mathbf{P}_n$;
14: Compute $\mathbf{G}_d = \mathbf{F}_d^H\mathbf{G}$;
15: Compute $\mathbf{F}_d$ by utilizing the steps from 6 to 12, and substitute $\mathbf{G}_d$ for $\bar{\mathbf{G}}_r$, M_R cycles;
16: **end for**
17: **End stage**
18: **Digital stage**:
19: Obtain the baseband channel $\bar{\mathbf{G}} = \mathbf{F}_d^H\mathbf{G}\mathbf{F}_{r2}$;
20: Compute $\mathbf{W}_{r2} = \mathrm{blk}\left[\mathbf{W}_{r2}^1, \ldots, \mathbf{W}_{r2}^K\right]$ and $\mathbf{W}_d = \mathrm{blk}\left[\mathbf{W}_{d1}, \ldots, \mathbf{W}_{dK}\right]$ by using the baseband BD scheme;
21: **End stage**
22: Obtain the total equivalent channel $\mathbf{H}_{total} = \mathbf{W}_d^H\mathbf{F}_d^H\mathbf{G}\mathbf{F}_{r2}\mathbf{W}_{r2}$;
23: Compute $\mathbf{P}_r$ by using waterfilling power allocation of the total equivalent channel $\mathbf{H}_{total}$;
24: **Output**: $\mathbf{W}_d, \mathbf{F}_d, \mathbf{F}_{r2}, \mathbf{W}_{r2}, \mathbf{P}_r$.

C. HYBRID BEAMFORMING DESIGN FOR RELAY SYSTEM WITH THE FULL-CONNECTED STRUCTURE

As all mentioned in the Section II, for the full-connected structure, the hybrid beamforming design is simpler than the sub-connected structure due to the fact that the block-diagonal constraints are not required. However, the hybrid beamforming design problem of the full-connected structure is still challenging due to the constant-modulus constraints and the interferences of inter-destination and inner-destination. Luckily, the proposed optimal design scheme for sub-connected structure can be extended to solve the hybrid beamforming design problem of the full-connected structure, and meets the design requirements.

When the multi-destination relay system with full-connected structure is shown in Figs. 2, the hybrid beamforming design problem of relay system based on the maximum sum rate can be expressed as follows:

$$\begin{aligned}(\mathcal{Q})\ &(\mathbf{W}_s, \mathbf{F}_s, (\mathbf{F}_{r1}, \mathbf{W}_{r1}, \mathbf{W}_{r2}, \mathbf{F}_{r2}), \mathbf{F}_d, \mathbf{W}_d, \mathbf{P}_s, \mathbf{P}_r)\\ &= \underset{(\mathbf{W}_s,\mathbf{F}_s,(\mathbf{F}_{r1},\mathbf{W}_{r1},\mathbf{W}_{r2},\mathbf{F}_{r2}),\mathbf{F}_d,\mathbf{W}_d,\mathbf{P}_s,\mathbf{P}_r)}{\arg\max} R\\ &s.t.\ |\mathbf{F}_s(i,j)| = \frac{1}{\sqrt{N_S}}, |\mathbf{F}_{r1}(i,j)| = \frac{1}{\sqrt{N_R}}, (10), (11), (*)\\ &\quad \mathbf{W}_{r2} = \mathrm{blk}\left(\mathbf{W}_{r2}^1, \mathbf{W}_{r2}^2, \ldots, \mathbf{W}_{r2}^K\right),\\ &\quad \mathbf{W}_d = \mathrm{blk}\left(\mathbf{W}_d^1, \mathbf{W}_d^2, \ldots, \mathbf{W}_d^K\right),\\ &\quad \|\mathbf{P}_s\|_F^2 = P_s, \|\mathbf{P}_r\|_F^2 = P_r, \end{aligned} \tag{38}$$

Similar to reformulation of the original optimal problem $(\mathcal{P})$ given in Section III, the problem $\mathcal{Q}$ can also be reformulated as two subproblems and expressed as

$$\begin{aligned}(\mathcal{Q}_1)\ &(\mathbf{W}_s, \mathbf{F}_s, \mathbf{F}_{r1}, \mathbf{W}_{r1}, \mathbf{P}_s) = \underset{(\mathbf{W}_s,\mathbf{F}_s,\mathbf{F}_{r1},\mathbf{W}_{r1},\mathbf{P}_s)}{\arg\max} R_1\\ &s.t.\ |\mathbf{F}_s(i,j)| = \frac{1}{\sqrt{N_S}}, |\mathbf{F}_{r1}(i,j)| = \frac{1}{\sqrt{N_R}}, (*)\\ &\quad \|\mathbf{P}_s\|_F^2 = P_s, \end{aligned} \tag{39}$$

$$\begin{aligned}(\mathcal{Q}_2)\ &(\mathbf{W}_{r2}, \mathbf{F}_{r2}, \mathbf{F}_d, \mathbf{W}_d, \mathbf{P}_r) = \underset{(\mathbf{W}_{r2},\mathbf{F}_{r2},\mathbf{F}_d,\mathbf{W}_d,\mathbf{P}_r)}{\arg\max} R_2\\ &s.t.\ (10), (11),\\ &\quad \mathbf{W}_{r2} = \mathrm{blk}\left(\mathbf{W}_{r2}^1, \mathbf{W}_{r2}^2, \ldots, \mathbf{W}_{r2}^K\right),\\ &\quad \mathbf{W}_d = \mathrm{blk}\left(\mathbf{W}_d^1, \mathbf{W}_d^2, \ldots, \mathbf{W}_d^K\right),\\ &\quad \|\mathbf{P}_r\|_F^2 = P_r. \end{aligned} \tag{40}$$

For the DF relay system, the full-connected and mixed structures have the same structure from the source to the relay decoding, and the same beamforming design method proposed in section IV-A can be adopted for $(\mathcal{Q}_1)$. As for the subproblem $(\mathcal{Q}_2)$, the method of SVD of each destination's channel $\mathbf{G}_k$ can be used in the analog beamforming design and expressed as

$$\frac{1}{\sqrt{N_R}}\mathbf{G}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^H, \tag{41}$$

$$\frac{1}{\sqrt{N_R}}\mathbf{G}_{comp} = \mathbf{U}_{comp} \Sigma_{comp} \mathbf{V}_{comp}^H, \tag{42}$$

where

$$\mathbf{G}_{comp} = \mathbf{F}_d^H \mathbf{G} = \begin{bmatrix} \mathbf{F}_{d1}^H & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{F}_{dK}^H \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_K \end{bmatrix}. \tag{43}$$

Similarly, the optimal solution of the analog beamforming design can be obtained by defining $\mathbf{F}_{dk} = \mathbf{U}_k(:, 1 : M_{Dk})$ and $\mathbf{F}_{r2} = \mathbf{V}_{comp}(:, 1 : M_R)$. The solutions $\mathbf{W}_{dk}$ and $\mathbf{W}_{r2}$ designed in the digital beamforming can be obtained by exploiting the baseband BD scheme, and the power allocation matrix $\mathbf{P}_r$ can be designed by using the waterfilling power allocation method. Thus, the two subproblems of full-connected structure are effectively solved by the proposed method.

V. NUMERICAL SIMULATION

To evaluate the effectiveness of the proposed scheme, we present numerical results of the hybrid beamforming design for DF mmWave massive MU-MIMO relay system. All simulation results were obtained by averaging over 2,000 random channel realizations based on MATLAB platform. For simplicity, the number of antennas connected to each RF chain at the relay forwarding are assumed to be the same, i.e., $N_R^n = \frac{N_R}{M_R}$, $n = 1, \ldots, M_R$, meanwhile each destination has the same number of RF chains and antennas connected to each RF chain, i.e., $N_{Dk}^n = \frac{N_{Dk}}{M_{Dk}} = N_{DK}$, $n = 1, \ldots, M_{Dk}$, $M_{Dk} = M_{DK}$. The number of antennas at the source, relay, and each destination are set as $N_S = 128$, $N_R = 32$, and $N_{Dk} = 32$, respectively. The corresponding number of RF chains are $M_S = 16$, $M_R = 16$, and $M_{Dk} = 2$. During each simulation, the 4-destination MU-MIMO relay system is assumed to simultaneously transmit $L_s = 1$ stream per destination. The transmission power at the source and relay is $P_s = P_r$, and the noise variable is $\sigma_R^2 = \sigma_D^2$. The ULAs at the source, relay, and each destination have an antenna spacing $d = \frac{\lambda}{2}$. The number of first-hop channel scattering paths between the source and relay is $L_h = 16$, and for the second-hop channel between the relay and each destination, there are $L_{gk} = 4$ scattering paths. All the AoAs/AoDs are assumed to be uniformly distributed in $[0, 2\pi]$.

It is worth noting that we focus on the hybrid beamforming design of DF relay system with mixed structure in the paper. Since the optimal full-digital beamforming scheme implemented with the full-digital BD technology and water-filling power allocation algorithm can achieve the best performance of the mmWave system, the optimal full-digital beamforming scheme is considered as the upper bound of the hybrid beamforming design. Meanwhile, compared with the mixed structure, the full-connected structure mentioned in the Section IV-C is only different in terms of the analog beamforming and combining from the relay forwarding to destinations. However, the performance of full-connected structure outperforms that of the mixed one, so it can be used as a performance comparison. In addition, for comparing with the AF relay system, the mmWave massive MU-MIMO relay system with full-connected structure proposed in [23] is also included in the simulation.

A. PERFORMANCE FOR SUM RATE

Since our objective function is the sum rate between the source and the destinations, we compare the sum rate performance of different beamforming schemes. Fig. 3 compares the sum rate performance of different beamforming schemes versus SNR for $K = 4$. It can be seen clearly that the sum rate performance of the proposed full-connected hybrid beamforming design is close to that of the full-digital beamforming. Therefore, the proposed scheme for the full-connected structure is an ideal solution in the DF mmWave relay system. However, due to the high hardware complexity, hybrid beamforming systems with sub-connected structure are often used in practice, which will further reduce the performance

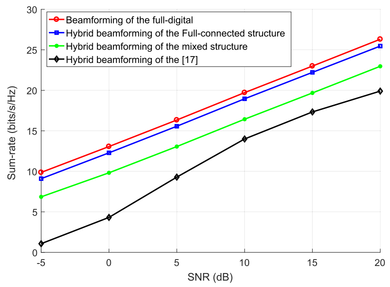

FIGURE 3. Sum rate comparison for different beamforming schemes versus SNR, where $K = 4$.

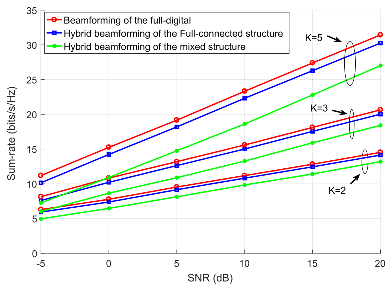

FIGURE 4. Sum rate comparison for the proposed mixed and full-connected hybrid beamforming schemes and the full-digital beamforming versus SNR, where $K = 2, 3, 5$.

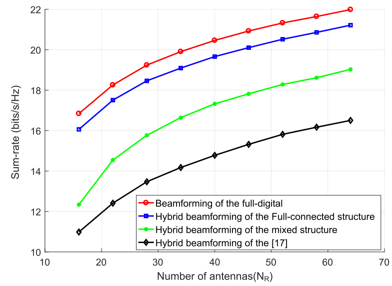

FIGURE 5. Sum rate comparison for different beamforming schemes versus the antennas at the relay, where SNR = 10dB, $N_R = 2M_R$.

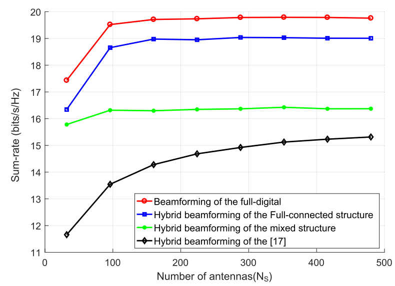

FIGURE 6. Sum rate comparison for different beamforming schemes versus the antennas at source, where SNR = 10dB.

of the system. The mixed structure is a compromise between the full-connected and sub-connected in structure, and the performance gap between the mixed and the full-connected is reduced within 3dB. In addition, the proposed scheme for mixed structure outperforms the MP scheme in [23], so the DF mmWave massive MU-MIMO relay system with mixed structure can be considered in practical application scenarios.

B. PERFORMANCE FOR NUMBER OF DESTINATIONS

Fig. 4 compares the sum rates achieved by the proposed mixed and full-connected hybrid beamforming schemes with the full-digital beamforming when the number of destinations K is different. It can be seen that the sum rate performance gap between the proposed mixed and full-connected beamforming schemes is narrow when $K = 2$ and $K = 3$, and both of the proposed schemes obtain close performance to the full-digital one. However, when $K = 5$, the gap between the proposed schemes becomes large and exceeds 3dB. Therefore, the hybrid beamforming design schemes proposed in this paper meet the best performance when the number of destinations is less than or equal to $K = 4$.

C. PERFORMANCE FOR NUMBER OF ANTENNAS AT RELAY

Fig. 5 compares the sum rate performance of different beamforming schemes versus the antennas at the relay, where SNR = 10dB, and the number of antennas is twice the number of RF chains at the relay. It can be clearly found that the performance of different design schemes increases correspondingly as the increasing number of RF chains. In addition, the proposed schemes for both the mixed and full-connected structures outperform the MP scheme in [23]. As the number of antennas at relay increases, the antenna gain increases correspondingly, which is the expected result.

D. PERFORMANCE FOR NUMBER OF ANTENNAS AT SOURCE

Fig. 6 compares the sum rate performance of different beamforming schemes versus the antennas at the source, where SNR = 10dB. As can be seen from this figure, when the number of antennas at source is small, the sum rate of different beamforming schemes increases radically with the augment of the number of antennas at source. When the number of antennas at source is large, the sum rate of different schemes will be stabilized. Due to the DF relay system can be decomposed as two (or more) different MIMO sub-systems in series, increasing the number of antennas at source will improve the sum rate performance from the source to the relay decoding, and the sum rate of overall system can also be improved within a certain range. However, the performance from the relay forwarding to the destinations will not be increased. Therefore, the sum rate of different schemes will be saturated for the number of antennas at source is large enough.

E. PERFORMANCE FOR POWER EFFICIENCY OF RELAY

As mentioned in Section I, for the hybrid beamforming of relay system with different connection structures, the power consumption is a key issue which should be considered. In this subsection, we aim at comparing the proposed mixed and full-connected structures in terms of power efficiency of relay [38].

Considering the relay in the hybrid beamforming design, the power consumption mainly includes the following parts: a)the low noise amplifier (LNA) and the power amplifier (PA) connected to each antenna on the receiver node and the transmitter node, respectively; b) the phase shifters (PS) and the RF chains (RF) on both receiver and transmitter nodes; c) the digital baseband (BB) processor; d) digital-to-analog converters (DAC) on the receiver node and analog-to-digital converters (ADC) on the transmitter node.

According to the power consumption of the above five parts, the total power consumption P_{total} of the DF relay can be written as

$$P_{total} = 2P_{BB} + 2M_R P_{RF} + N_R(P_{LNA} + P_{PA}) + 2M_{PS}P_{PS} + M_R(P_{DAC} + P_{ADC}), \quad (44)$$

where P_{BB}, P_{RF}, P_{LNA}, P_{PA}, P_{PS}, P_{DAC}, and P_{ADC} indicate the power consumed by the digital baseband processor, the each RF chains, the each low noise amplifier, the each power amplifier, the each phase shifters, the each digital-to-analog converter, and the each analog-to-digital converter, respectively. For the relay with the mixed and the full-connected structures, the number of phase shifters M_{PS} can be expressed as:

$$M_{PS} = \begin{cases} M_R N_S, & \text{full-connected structure,} \\ N_S, & \text{mixed structure.} \end{cases} \quad (45)$$

Since the sub-connected structure is adopted by the relay forwarding of mixed structure, it can be seen from (45) that the relay with sub-connected structure needs less phase shifters than the full-connected one, which indicates that the power consumed by the relay with sub-connected structure is lower than that of the full-connected one.

To better compare the power consumption performance of the above two structures, the power efficiency η is employed as the measurement standard and expressed as [39]

$$\eta = \frac{R}{P_{total}} \text{ (bps/Hz/J).} \quad (46)$$

Fig. 7 compares the power efficiency performance of the proposed relay hybrid beamforming with the mixed and the full-connected structures. The simulation parameters of power consumption are set by [40], as follows: $P_{BB} = 10$W, $P_{RF} = 100$mW, $P_{LNA} = P_{PA} = 100$mW, $P_{PS} = 10$mW, $P_{DAC} = 110$mW, $P_{ADC} = 200$mW. It can be see clearly from the simulation results that for the proposed mixed and full-connected structures, the power efficiency will decrease tremendously as the increasing of the number of RF chains M_R. However, the relay power efficiency of the mixed structure decreases more slowly than that of the full-connected structure, due to the fact that the number of phase shifters of the sub-connected structure at the relay forwarding is equal to the number of antennas, and relatively independent of the number of RF chains. Therefore, it is proved that the relay system with mixed structure is not only lower in power consumption but also higher in power efficiency than the relay system with full-connected structure.

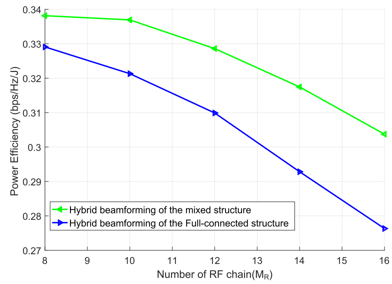

FIGURE 7. Energy efficiency comparison for the proposed relay hybrid beamforming with the mixed and full-connected structures versus the RF chains, where SNR = 10dB and $N_R = 2M_R$.

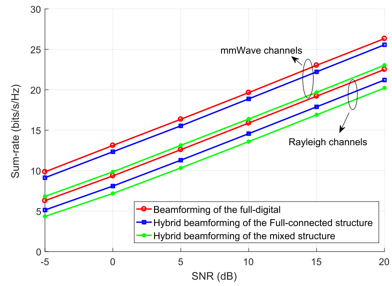

FIGURE 8. Sum rate comparison for the proposed hybrid beamforming schemes of relay system with the mixed and full-connected structures and the full-digital beamforming in the different channel distributions versus SNR.

F. PERFORMANCE FOR DIFFERENT CHANNEL DISTRIBUTIONS

To emulate the high and low frequency communication systems in some practical application scenarios, Fig. 8 shows the sum rate performance of relay system with different structures in the mmWave channel (high frequency) and Rayleigh channel (low frequency). In Rayleigh channel, the sum rate performance of relay system with the mixed structure is very close to that of the proposed full-connected structure. Therefore, the optimal hybrid beamforming scheme of relay with the mixed structure can be deployed for the low frequency communication. On the contrary, in the high frequency communication scenario, the optimal hybrid beamforming design of relay system with the full-connected structure can be utilized.

G. PERFORMANCE FOR SUPPORTED DATA STREAMS

Fig. 9 compares the sum rate performance of relay system with the mixed and the full-connected structures when the number of data streams supported by each destination is different, where SNR = 10dB and the number of destination $K = 1$. It can be seen that the sum rate of the mixed structure is approximately equal to that of the full-connected

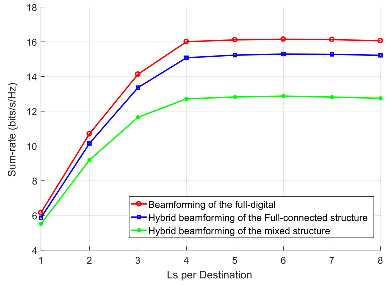

FIGURE 9. Sum rate comparison for the proposed mixed and full-connected beamforming schemes and the full-digital beamforming versus the supported data streams of each destination, where SNR = 10dB and $K = 1$.

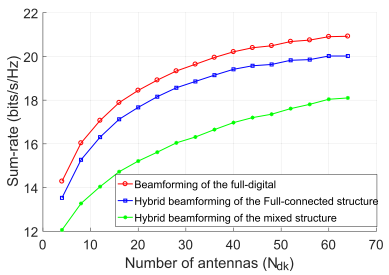

FIGURE 10. Sum rate comparison for the proposed mixed and full-connected beamforming schemes and the full-digital beamforming versus the number of each destinationąŕs antennas, where SNR = 10dB, $M_S = M_R = 16$, and $M_{Dk} = 2$.

structure, and close to that of the full-digital beamforming as the number of data streams supported by each destination is small, i.e., $L_s = 1$ and $L_s = 2$. Therefore, the relay system with the mixed structure is more suitable for the transmission with fewer data streams. It can also be seen from Fig. 9 that when the number of data streams per destination is $L_s = 4$, the sum rate of the two proposed designs will be saturated. Therefore, the maximum number of data streams per user in the proposed optimal hybrid beamforming scheme of relay system is four.

H. PERFORMANCE FOR NUMBER OF ANTENNAS AT DESTINATION

Fig. 10 compares the sum rate performance of relay hybrid beamforming schemes with the mixed and the full-connected structures when the number of each destinationąŕs antennas is different. As can be seen from the Fig. 10, when the number of each destinationąŕs antennas increases, the sum rate of the relay system with mixed and full-connected structures also grow correspondingly. The performance of full-connected structure is very close to that of the full-digital beamforming. Although the sum rate of relay system with mixed structure is lower than that of the full-connected, it also shows very good performance, and the gap between them is about 2dB. At the same time, with the increase of the number of antennas, the performance gap between the relay beamforming system with the mixed structure and the full-digital beamforming is gradually narrowed.

VI. CONCLUSION

In this paper, we have proposed the optimal hybrid beamforming design for the DF mmWave massive MU-MIMO relay systems. The optimal hybrid beamforming design aims to maximize the sum rate from the source to the destinations. To solve the intractable optimal hybrid beamforming design problem for the relay system with the mixed and full-connected structures, the piecewise successive approximation method was proposed based on the criterion which jointly designs the analog and digital beamforming stages by trying to avoid the loss of information at each stage. Meanwhile, the idea of successive interference cancelation, the baseband BD scheme, and waterfilling power allocation method were united for the optimal hybrid beamforming design. The high-approximate optimal solution was obtained by the proposed design scheme respectively in the two kinds of relay systems. Simulation results validate that the proposed optimal design scheme achieves superior performance in terms of the sum rate in both mixed and full-connected structures, so a more suitable relay structure can be chosen according to the requirements of different scenarios in actual application. In the future, the simpler optimal hybrid beamforming design for obtaining the stable higher performance of the mmWave MIMO relay systems can be studied based on the work in this paper.

REFERENCES

[1] G. Zhang, k. Saito, W. Fan, X. Cai, P. Hanpinitsak, J.-I. Takada, and G. F. Pedersen, "Experimental characterization of millimeter-wave indoor propagation channels at 28 GHz," *IEEE Access*, vol. 6, pp. 76516–76526, 2018.

[2] R. He, B. Ai, G. L. Stüber, G. Wang, and Z. Zhong, "Geometrical-based modeling for millimeter-wave MIMO mobile-to-mobile channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 2848–2863, Apr. 2018.

[3] R. Chakraborty, N. Kumari, M. Mousam, and A. Mukherjee, "The Future of 5G and Millimeter Waves," in *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, Coimbatore, India, Mar. 2018, pp. 1679–1683.

[4] S. Ahmed, M. Sadekm, A. Zekry, and H. Elhennawy, "Hybrid analog and digital beamforming for space-constrained and energy-efficient massive MIMO wireless systems," in *Proc. 40th Int. Conf. Telecommun. Signal Process. (TSP)*, Barcelona, Spain, Jul. 2017, pp. 186–189.

[5] M. Alouzi and F. Chan, "Millimeter wave massive MIMO with Alamouti code and imperfect channel state information," in *Proc. IEEE 5G World Forum (5GWF)*, Silicon Valley, CA, USA, Jul. 2018, pp. 507–511.

[6] M. K. Samimi and T. S. Rappaport, "3-D statistical channel model for millimeter-wave outdoor mobile broadband communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2430–2436.

[7] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. Amsterdam, The Netherlands: Elsevier, 2016.

[8] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[9] S. Suyama, T. Okuyama, Y. Inoue, and Y. Kishiyama, "5G multi-antenna technology," *NTT DOCOMO Tech. J.*, vol. 17, no. 4, pp. 29–39, Apr. 2016.

[10] Z. X. Wei, X. Zhu, S. Sun, Y. Huang, A. Al-Tahmeesschi, and Y. Jiang, "Energy-efficiency of millimeter-wave full-duplex relaying systems: Challenges and solutions," *IEEE Access*, vol. 4, pp. 4848–4860, 2016.

[11] M. K. 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.

[12] S. Gong, C. Xing, Z. Fei, and S. Ma, "Millimeter-wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2059–2071, Mar. 2017.

[13] G. Femenias and F. Riera-Palou, ''Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity,'' *IEEE Access*, vol. 7, pp. 44596–44612, 2019.
[14] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, ''Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems,'' *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1199–1215, Feb. 2019.
[15] F. Sohrabi and W. Yu, ''Hybrid digital and analog beamforming design for large-scale antenna arrays,'' *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
[16] Y. Lu, C. Cheng, J. Yang, and G. Gui, ''Improved hybrid precoding scheme for mmWave large-scale MIMO systems,'' *IEEE Access*, vol. 7, pp. 12027–12034, 2019.
[17] M. Hanif, H.-C. Yang, G. Boudreau, E. Sich, and H. Seyedmehdi, ''Low-complexity hybrid precoding for multi-user massive MIMO systems: A hybrid EGT/ZF approach,'' *IET Commun.*, vol. 11, no. 5, pp. 765–771, Mar. 2017.
[18] Y.-Y. Lee, C.-H. Wang, Y.-H. Huang, ''A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems,'' *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 305–317, Jan. 2015.
[19] Y. Lin, S. Jin, and X. H. You, ''Hybrid beamforming based on uniform channel decomposition for millimeter wave massive MIMO systems,'' in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Hangzhou, China, Oct. 2018, pp. 1–6
[20] Z. Luo, C. Zhan, L. Zhang, and R. Zhang, ''Robust hybrid beamforming in millimeter wave relay networks with imperfect CSI,'' *IEEE Access*, vol. 6, pp. 73093–73101, 2018.
[21] M. Fozooni, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, ''Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation,'' *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 107–123, Jan. 2019.
[22] 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)*, Sydney, NSW, Australia, Jun. 2014, pp. 5838–5842.
[23] X. Xue, T. E. Bogale, X. Wang, Y. Wang, and L. B. Le, ''Hybrid analog-digital beamforming for multiuser MIMO millimeter wave relay systems,'' in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, China, Shenzhen, Nov. 2015, pp. 1–7.
[24] J.-S. Sheu, ''Hybrid digital and analogue beamforming design for millimeter wave relaying systems,'' *J. Commun. Netw.*, vol. 19, no. 5, pp. 461–469, Oct. 2017.
[25] X. Xue, Y. Wang, L. Dai, and C. Masouros, ''Relay hybrid precoding design in millimeter-wave massive MIMO systems,'' in *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2011–2016, Apr. 2018.
[26] 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)*, Chicago, IL, USA, Aug. 2018, pp. 1–5.
[27] C. Hu, J. Liu, X. Liao, Y. Liu, and J. Wang, ''A novel equivalent baseband channel of hybrid beamforming in massive multiuser MIMO systems,'' *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 764–767, Apr. 2018.
[28] 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.
[29] T. E. Bogale and L. B. Le, ''Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital,'' in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 4066–4071.
[30] C.-E. Chen, ''An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems,'' *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, Jun. 2015.
[31] C. A. Balanis, *Antenna Theory Analysis Design*. Hoboken, NJ, USA: Wiley, 2012.
[32] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., ''Spatially sparse precoding in millimeter wave MIMO systems,'' *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, ''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.
[38] S. Buzzi and C. D'Andrea, ''Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmWave systems,'' *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 385–396, Jun. 2018.
[39] L.-F. Lin, W.-H. Chung, H.-J. Chen, and T.-S. Lee, ''Energy efficient hybrid precoding for multi-user massive MIMO systems using low-resolution ADCs,'' in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2016, pp. 115–120.
[40] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Upper Saddle River, NJ, USA: Prentice-Hall, 2015.

YANG ZHANG 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.

JIANHE DU 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, China. His research interests mainly include channel estimation, massive MIMO communication, MIMO relay, and tensor-based signal processing applied to wireless communications. He has published several journal articles in IEEE ACCESS, the IEEE COMMUNICATIONS LETTERS, and *IET Communications*. He was a Guest Editor of the ''Special Issue Applications of Tensor Models in Wireless Communications and Mobile Computing'' of the *Wireless Communications and Mobile Computing Journal*.

YUANZHI CHEN received the B.Sc. and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1988 and 1993, respectively. From 1994 to 1996, he was a Postdoctoral Research Fellow with Beihang University. He is currently a Professor with the Faculty of Science and Technology, Communication University of China, Beijing, China. His research interests mainly include wireless communication systems, signal processing, and massive MIMO.

MENG HAN is currently pursuing the master's degree with the School of Information and Communication Engineering, Communication University of China, Beijing, China. Her current research interests include massive MIMO systems, cooperative communication technique, channel estimation, and tensor-based technique.

XINGWANG LI (S'12–M'15) received the B.Sc. degree from Henan Polytechnic University, in 2007, the M.Sc. degree from the University of Electronic Science and Technology of China, in 2010, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2015. From 2010 to 2012, he was an Engineer with Comba Telecom Ltd., Guangzhou, China. From 2017 to 2018, he was a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He was a Visiting Scholar with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, from 2016 to 2018. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His current research interests include MIMO communication, cooperative communication, hardware constrained communication, non-orthogonal multiple access, physical-layer security, unmanned aerial vehicles, and the Internet of Things. He has served as a TPC member for many conferences, such as the IEEE/CIC International Conference on Communications in China (ICCC'2019) and the IEEE Global Communications Conference 2018 (Globecom'18). He is also an Editor on the Editorial Board of IEEE ACCESS, *Computer Communications*, and *KSII Transactions on Internet and Information Systems*.