Hybrid Beamforming with Sub-Connected Structure for MmWave Massive Multi-User MIMO Relay Systems

Meng Han, *Student Member, IEEE*, Jianhe Du, *Member, IEEE*, Yang Zhang, Yuanzhi Chen, Xingwang Li, *Senior Member, IEEE*, Khaled M. Rabie, *Senior Member, IEEE*, and Ferdi Kara, *Senior Member, IEEE*.

***Abstract*—In this paper, we investigate the hybrid beamforming design with sub-connected structure for millimeter-wave (mmWave) massive multi-user multiple-input multiple-output (MU-MIMO) amplify-and-forward (AF) relay systems. Considering the constant-modulus and block-diagonal constraints, an iterative algorithm is proposed to sequentially design the analog beamforming of the base station, relay, and users. Next, the relay baseband combiner is designed by transforming the highly complicated non-convex mutual information maximization problem into an easily tractable weighted minimum mean squared error one. To mitigate the inter-user interference, a successive serial interference cancellation based and a piecewise successive approximation based methods are developed for the uniform sparse distribution and dense distribution scenarios, respectively. Simulation results demonstrate that the two proposed hybrid beamforming schemes can achieve good performance in terms of sum-rate and energy efficiency, and outperform benchmarks significantly. In addition, the two proposed schemes are shown to be robust to imperfect channel state information (CSI), even though they are studied based on perfect CSI.**

***Index Terms*—Hybrid beamforming, sub-connected structure, mmwave, massive MU-MIMO, AF relay.**

I. Introduction

THE millimeter-wave (mmWave) frequency band has a vast range of unlicensed spectrum resources, which can solve the problem of lack of spectrum resources in communication systems effectively, and thus it has attracted much attention in both academia and industry [1]. The short wavelength of mmWave promotes its combination with massive multiple-input multiple-output (MIMO) technology. It has been revealed that their integration will significantly improve the spectral efficiency and transmission reliability, and has been considered as a promising key technology for the next generation wireless communication systems [2].

To overcome the high path loss (PL) of mmWave signals, mmWave massive MIMO can provide effective beamforming and multiplexing gain to guarantee sufficient received signal power. However, mmWave communications are fatally hindered by non-line-of-sight (NLoS) transmission because the resulting blockage will cause enormous propagation loss [3]. Fortunately, relay nodes can help to reduce the transmission power and expand the signal coverage, which improves the system performance, particularly at the edge of the cell [4], [5]. It is well known that the full-digital (FD) beamforming can achieve the optimal antenna gain, but needs more radio-frequency (RF) chains (each antenna is connected to one dedicated RF chain) [6]. Moreover, it is shown that RF components may consume up to 70% of the total transceiver energy consumption [7]. Therefore, since the high hardware cost and power consumption are caused by a large number of RF chains in mmWave massive MIMO systems, the FD beamforming becomes unaffordable in practice [8]. To this end, the hybrid beamforming implemented by analog phase shifters (PSs) and baseband (BB) processor has become a potential alternative, since it can achieve the similar performance to that of the FD one while using fewer RF chains [9]. Therefore, it is of great significance to study hybrid beamforming in relay scenarios.

The relay systems can be divided into two categories, i.e., amplify-and-forward (AF) and decode-and-forward (DF) relay systems [10]. For the DF relay, in order to mitigate the self-interference (SI), a zero space SI cancelation algorithm and an orthogonal matching pursuit (OMP) based SI cancelation precoding algorithm are proposed in [11] and [12], respectively. Further, in [13], the authors propose a mixed-connected structure relay system to compromise power consumption and sum-rate performance. However, the DF relay normally leads to higher power consumption and latency than the AF relay due to its complex signal processing, and may introduce performance bottleneck at source-relay link [14]. Therefore, we devote our attention to the AF relay strategy instead of the DF one.

For the mmWave AF relay scenario, the hybrid beamforming can be realized by two typical structures, i.e., the fully-connected structures (where each RF chain is connected to all antennas) [15], and the sub-connected structures (where each RF chain is connected to a subset of antennas) [16]. For

This work was supported in part by the Beijing Natural Science Foundation under Grant L191003, in part by the National Natural Science Foundation of China under Grants U2141233, 62071435, and 61601414, and in part by the Fundamental Research Funds for the Central Universities under Grant CUC220D001. (*Corresponding author: Jianhe Du.*)

M. Han, J. Du, Y. Zhang and Y. Chen are with the State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China (email: hm593@cuc.edu.cn, dujianhe1@gmail.com, zhangyangcuc2021@163.com and chenyuanzhi@cuc.edu.cn).

X. Li is with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China (email: lixingwang@hpu.edu.cn).

Khaled M. Rabie is with the Department of Engineering, Manchester Metropolitan University, M1 5GD Manchester, U.K. (e-mail: K.Rabie@mmu.ac.uk).

Ferdi Kara is with the Department of Computer Engineering, Zonguldak Bulent Ecevit University, Zonguldak 67100, Turkey (e-mail: f.kara@beun.edu.tr).

the AF relay-assisted mmWave networks with fully-connected structure, the authors in [17] jointly designed the hybrid beamforming matrices by using the OMP algorithm for the single-user MIMO system, which was extended by [18] to the multi-user scenario. A joint source and relay beamforming design for one data stream scenario was proposed based on semidefinite programming (SDP) in [19]. To reduce the complexity, the authors in [20] designed the RF and BB beamforming separately, where a minimum mean squared error (MMSE) based design method was utilized for BB filters. An effective algorithm based on alternating direction method of multipliers (ADMM) to improve the system sum-rate performance was studied in [21] and [22], respectively. The work [23] utilized three different methods to design the analog and digital beamforming for different nodes. An artificial noise (AN) aided two-stage secure hybrid beamforming algorithm was proposed in [24], and the scenario of cooperative non-orthogonal multiple access (NOMA) networks was studied in [25]. For the AF relay-assisted mmWave networks with sub-connected structure, hybrid beamforming for single-user scenarios was studied in [26], where the authors reformulated the original problem as three subproblems and proposed an iterative successive approximation (ISA) algorithm. The work [27] designed hybrid beamforming for the downlink single-user scenario by employing the ADMM algorithm. However, the hybrid beamforming designs of the relay station in [26] and [27] are based on the FD beamforming for the base station (BS) and the user. The block-diagonal (BD) based geometric mean decomposition Tomlinson Harashima precoding (GMD-THP) scheme for multi-user scenarios was proposed in [28], but this scheme is non-linear and difficult to be put into practical applications.

To the best of our knowledge, most of researches focus on the AF relay with the fully-connected structure [17]–[25] or in single-user cases [17], [26], [27]. Compared with the sub-connected structure, the fully-connected structure uses more PSs and consumes more energy [29]. Leveraging the hardware complexity and system performance, the sub-connected structure is more energy efficient and easier to be implemented for massive MIMO systems [30], [31]. Besides, the FD beamforming causes high power consumption due to more RF chains. To this end, considering practical constraints (e.g., limited power for relay), the hybrid beamforming for the sub-connected AF relay network is investigated in this paper to serve multi-user in the domain of mmWave massive MIMO systems. Specifically, we jointly optimize the hybrid beamforming designs of BS, relay and users. The design goal is to maximize the system sum-rate performance. To reduce the complexity, the analog and digital beamforming are designed separately. Assuming perfect channel state information (CSI), the analog beamforming of the BS, relay, and users are first designed by continuous serial operations. Then, by utilizing the equivalence between the mutual information and mean squared error (MSE) function, a weighted MMSE (WMMSE) based [32] method is developed to design the relay BB combining matrix. Further, two different BS and users digital beamforming/combining design methods are respectively proposed for two different scenarios to eliminate the inter-user interference (IUI).

The main contributions of our paper can be summarized as follows:

- To satisfy constant-modulus and BD constraints, the idea of successive interference cancellation (SIC) is employed to have a sequential design for the analog beamforming of the BS, relay and users. The proposed algorithm not only shows fast convergence speed, but also achieves the near-optimal performance.
- Based on the equivalence between the mutual information and MSE function, the relay BB combiner is designed by transforming the highly complicated non-convex mutual information maximization problem into an easily tractable WMMSE one.
- Two different user distribution scenarios, i.e., the uniform sparse distribution and dense distribution, are considered in this paper. For the first scenario, a serial SIC (SSIC) based method is proposed to mitigate the interference between adjacent users. For the second scenario, the piecewise successive approximation (PSA) based method is developed to eliminate the interference between different users.
- Simulation results reveal that the two proposed beamforming design schemes have great sum-rate and better energy efficiency performance. Meanwhile, they can support more users, and show great robustness in the case of imperfect CSI.

The rest of this paper is organized as follows: Section II briefly describes the system and channel models of the considered mmWave massive multi-user MIMO (MU-MIMO) relay system. The analog beamforming/combining and relay BB combining designs are shown in Section III. The digital beamforming/combining of BS and users for two different scenarios are discussed in Section IV. Section V shows the numerical simulation results of the proposed schemes, and Section VI draws conclusions.

Notation: Bold upper-case and lower-case letters represent matrices and column vectors, respectively; $(\cdot)^{-1}$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$ denote inversion, transpose and conjugate transpose operations, respectively. $\mathbb{E}[\cdot]$ stands for the expectation. $\mathrm{Tr}(\cdot)$ indicates the trace operator. $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ are the l_2-norm of a vector and the Frobenius norm of a matrix, respectively. $\mathbf{I}_N$ is an identity matrix with size $N \times N$. $\mathbf{0}_{M\times N}$ is an all-zero matrix with size $M \times N$. $\mathbf{a}(j)$ denotes the jth element of the column vector $\mathbf{a}$. $\mathbf{A}(i,j)$, $\mathbf{A}(:,j)$, and $\mathbf{A}(i,:)$ respectively denote the (i,j)th element, jth column vector, and ith row vector of the matrix $\mathbf{A}$. $|\mathbf{A}(i,j)|$ is the amplitude. $\mathbf{A}(i:j,:)$ and $\mathbf{A}(:,i:j)$ represent the matrix consists of vectors from rows i to j and columns i to j of $\mathbf{A}$, respectively. $|\cdot|$ is the determinant operator. $\mathrm{blk}[\mathbf{b}_1,\ldots,\mathbf{b}_K]$ denotes a $KM \times K$ block diagonal matrix, where $\mathbf{b}_k \in \mathbb{C}^M$ for all $k = 1,\ldots,K$. $\angle\mathbf{A}$ denotes the operation of getting the angle of each entry in $\mathbf{A}$.

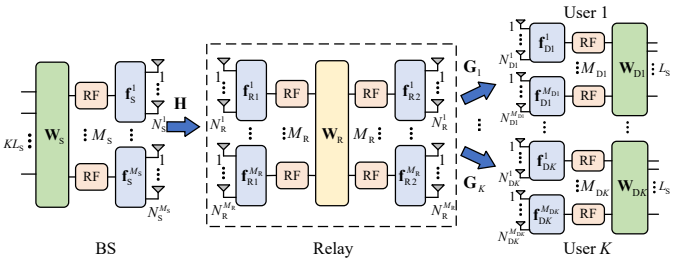


Fig. 1. Block diagram of mmWave massive MU-MIMO relay system with hybrid structure.

II. System Description

A. System Model

The considered mmWave massive MU-MIMO AF relay system using hybrid beamforming is illustrated in Fig. 1[1]. The BS, relay and all users adopt the sub-connected structure, and the relay works in the half-duplex mode [23], [33]. The BS is equipped with M_{S} RF chains and each RF chain is connected to a disjoint subset of $N_{\mathrm{S}}^{m_{\mathrm{S}}}$ antennas, where the total number of antennas is $N_{\mathrm{S}}=\sum_{m_{\mathrm{S}}=1}^{M_{\mathrm{S}}} N_{\mathrm{S}}^{m_{\mathrm{S}}}$. Both the relay receiving and transmitting ends are equipped with M_{R} RF chains and each RF chain is connected to a disjoint subset of $N_{\mathrm{R}}^{m_{\mathrm{R}}}$ antennas[2], where the total number of antennas is $N_{\mathrm{R}}=\sum_{m_{\mathrm{R}}=1}^{M_{\mathrm{R}}} N_{\mathrm{R}}^{m_{\mathrm{R}}}$. Each user k is equipped with $M_{\mathrm{D}k}$ RF chains and each RF chain is connected to a disjoint subset of $N_{\mathrm{D}k}^{m}$ antennas to support L_{S} data streams. The number of antennas for each user is $N_{\mathrm{D}k}=\sum_{m=1}^{M_{\mathrm{D}k}} N_{\mathrm{D}k}^{m}$, and that for all users is $N_{\mathrm{D}}=\sum_{k=1}^{K} N_{\mathrm{D}k}$. The number of RF chains for all users is $M_{\mathrm{D}}=\sum_{k=1}^{K} M_{\mathrm{D}k}$. To enable multi-stream communication, the number of RF chains for all nodes are constrained by $KL_{\mathrm{S}} \leqslant M_{\mathrm{S}} \leqslant N_{\mathrm{S}}$, $KL_{\mathrm{S}} \leqslant M_{\mathrm{R}} \leqslant N_{\mathrm{R}}$ and $L_{\mathrm{S}} \leqslant M_{\mathrm{D}k} \leqslant N_{\mathrm{D}k}$. It is assumed that the overall transmitted vector of K users is $\mathbf{s}=\left[\mathbf{s}_1^{\mathrm{T}},\ldots,\mathbf{s}_K^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{KL_{\mathrm{S}}}$ with normalized power $\mathbb{E}\left[\mathbf{s}\mathbf{s}^{\mathrm{H}}\right]=\frac{1}{KL_{\mathrm{S}}}\mathrm{I}_{KL_{\mathrm{S}}}$, where $\mathbf{s}_k$ represents the L_{S} symbols transmitted to the kth user.

At the BS, the transmitted symbols of K users are first processed by a digital beamformer $\mathbf{W}_{\mathrm{S}} \in \mathbb{C}^{M_{\mathrm{S}} \times KL_{\mathrm{S}}}$ followed by an analog beamformer $\mathbf{F}_{\mathrm{S}} \in \mathbb{C}^{N_{\mathrm{S}} \times M_{\mathrm{S}}}$. Then, the transmitted signal from the BS can be expressed as

$$\mathbf{x}_{\mathrm{S}}=\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}\mathbf{s}. \tag{1}$$

Assume the transmission power constraint at the BS is P_{S}. Then the hybrid beamforming at the BS should meet the following power constraint:

$$\mathbb{E}\left[\mathbf{x}_{\mathrm{S}}\mathbf{x}_{\mathrm{S}}^{\mathrm{H}}\right]=\mathrm{Tr}\left(\mathbf{W}_{\mathrm{S}}^{\mathrm{H}}\mathbf{F}_{\mathrm{S}}^{\mathrm{H}}\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}\right) \leqslant P_{\mathrm{S}}. \tag{2}$$

In the first time slot, the BS transmits data streams of all users to the relay through the channel $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{S}}}$. Therefore, the received symbols at the relay can be given by

$$\mathbf{y}_{\mathrm{R}}=\mathbf{H}\mathbf{x}_{\mathrm{S}}+\mathbf{n}_{\mathrm{R}}, \tag{3}$$

where $\mathbf{n}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}}}$ is the complex additive white Gaussian noise (AWGN) vector at the relay, with each entry following the independent and identically distributed (i.i.d.) $\mathcal{CN}\left(0,\sigma_{\mathrm{R}}^2\right)$. Then the received symbols at the relay are processed by an analog combiner $\mathbf{F}_{\mathrm{R1}} \in \mathbb{C}^{N_{\mathrm{R}} \times M_{\mathrm{R}}}$ followed by a BB combiner $\mathbf{W}_{\mathrm{R}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{R}}}$. During the second time slot, the relay forwards the symbols processed by the analog beamformer $\mathbf{F}_{\mathrm{R2}} \in \mathbb{C}^{N_{\mathrm{R}} \times M_{\mathrm{R}}}$ to users, while the BS remains silent. The forwarded symbols after the analog beamformer at the relay can be expressed as [23], [33]

$$\begin{aligned}\mathbf{x}_{\mathrm{R}}&=\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{y}_{\mathrm{R}}\\&=\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}\mathbf{s}+\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{n}_{\mathrm{R}}.\end{aligned} \tag{4}$$

Assume the transmission power constraint at the relay is P_{R}. Then the power of the transmitted symbols at the relay is constrained as

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}^{\mathrm{H}}\right]=&\mathrm{Tr}\left(\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\left(\mathbf{H}\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}^{\mathrm{H}}\mathbf{F}_{\mathrm{S}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\right.\right.\\&\left.\left.+\ \sigma_{\mathrm{R}}^2\mathbf{I}_{N_{\mathrm{R}}}\right)\left(\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\right)^{\mathrm{H}}\right) \leqslant P_{\mathrm{R}}.\end{aligned} \tag{5}$$

The symbols from the relay are transmitted to kth user through the channel $\mathbf{G}_k \in \mathbb{C}^{N_{\mathrm{D}k} \times N_{\mathrm{R}}}$. The received symbols can hence be written as

$$\begin{aligned}\mathbf{y}_{\mathrm{D}k}=&\mathbf{G}_k\mathbf{x}_{\mathrm{R}}+\mathbf{n}_{\mathrm{D}k}\\=&\mathbf{G}_k\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}\mathbf{s}\\&+\mathbf{G}_k\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{n}_{\mathrm{R}}+\mathbf{n}_{\mathrm{D}k},\end{aligned} \tag{6}$$

TABLE I
Symbolic Representation

Symbol	Definition
$M_{\mathrm{S}}/M_{\mathrm{R}}/M_{\mathrm{D}k}/M_{\mathrm{D}}$	Number of RF chains at the BS/relay/kth user/all users
$N_{\mathrm{S}}^{m_{\mathrm{S}}}/N_{\mathrm{R}}^{m_{\mathrm{R}}}/N_{\mathrm{D}k}^{m}$	Number of antennas connected to each RF chain at the BS/relay/kth user
$N_{\mathrm{S}}/N_{\mathrm{R}}/N_{\mathrm{D}k}/N_{\mathrm{D}}$	Number of antennas at the BS/relay/kth user/all users
L_{S}	Number of data streams transmitted to the kth user
$P_{\mathrm{S}}/P_{\mathrm{R}}$	Transmission power constraint at the BS/relay
$\mathbf{W}_{\mathrm{S}}/\mathbf{F}_{\mathrm{S}}$	Digital/Analog beamformer at the BS
$\mathbf{F}_{\mathrm{R1}}/\mathbf{F}_{\mathrm{R2}}$	Analog combiner/beamformer at the relay
$\mathbf{W}_{\mathrm{R}}$	BB combiner at the relay
$\mathbf{W}_{\mathrm{D}k}/\mathbf{F}_{\mathrm{D}k}$	Digital/Analog combiner at the kth user
$\mathbf{H}$	Channel matrix between the BS and relay
$\mathbf{G}_k$	Channel matrix between the relay and the kth user
$\mathbf{x}_{\mathrm{S}}/\mathbf{x}_{\mathrm{R}}$	Transmitted signal from the BS/relay after hybrid beamformers
$\mathbf{y}_{\mathrm{R}}/\mathbf{y}_{\mathrm{D}k}$	Received signal at the relay/kth user without hybrid combiners
$\mathbf{x}_{\mathrm{D}k}$	Received signal at the kth user after hybrid combiners
$\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$	The uth column of $\mathbf{F}_{\mathrm{R2}}$
$\mathbf{f}_{\mathrm{R2}}^{u}$	The non-zero block of $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$
$\mathbf{F}_{\mathrm{R2}}^{u-1}$	The first $u-1$ columns of $\mathbf{F}_{\mathrm{R2}}$
$\mathbf{W}_{\mathrm{S}}^{k}$	The kth digital beamforming matrix in $\mathbf{W}_{\mathrm{S}}=\left[\mathbf{W}_{\mathrm{S}}^{1},\ldots,\mathbf{W}_{\mathrm{S}}^{K}\right]$
$\bar{\mathbf{H}}=\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{S}}$	The equivalent BB channel in the first time slot
$\bar{\mathbf{G}}_k=\mathbf{F}_{\mathrm{D}k}^{\mathrm{H}}\mathbf{G}_k\mathbf{F}_{\mathrm{R2}}$	The equivalent BB channel in the second time slot

[1]A summary of symbols is listed in Table I.

[2]Note that for the relay, the number of antennas connected to the kth RF chain on the receiving end does not need to be the same as that connected to the kth RF chain on the transmitting end. This setting is just for the convenience of description.

where $\mathbf{n}_{\mathrm{D}k} \in \mathbb{C}^{N_{\mathrm{D}k}}$ is the complex AWGN vector at user k, whose each entry follows i.i.d. $\mathcal{CN}\left(0, \sigma_{\mathrm{D}k}^2\right)$. At the kth user, the received signal after hybrid combiners is given as

$$\begin{aligned}\mathbf{x}_{\mathrm{D}k} =& \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{F}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{y}_{\mathrm{D}k} \\ =& \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \bar{\mathbf{G}}_k \mathbf{W}_{\mathrm{R}} \bar{\mathbf{H}} \mathbf{W}_{\mathrm{S}} \mathbf{s} + \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \bar{\mathbf{G}}_k \mathbf{W}_{\mathrm{R}} \mathbf{F}_{\mathrm{R}1}^{\mathrm{H}} \mathbf{n}_{\mathrm{R}} \\ &+ \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{F}_{\mathrm{D}\,k}^{\mathrm{H}} \mathbf{n}_{\mathrm{D}k}, \end{aligned} \tag{7}$$

where $\mathbf{F}_{\mathrm{D}k} \in \mathbb{C}^{N_{\mathrm{D}k} \times M_{\mathrm{D}k}}$ and $\mathbf{W}_{\mathrm{D}k} \in \mathbb{C}^{M_{\mathrm{D}k} \times L_{\mathrm{S}}}$ represent the analog and digital combiners of the kth user, respectively, $\bar{\mathbf{H}} = \mathbf{F}_{\mathrm{R}1}^{\mathrm{H}} \mathbf{H} \mathbf{F}_{\mathrm{S}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{S}}}$ and $\bar{\mathbf{G}}_k = \mathbf{F}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{G}_k \mathbf{F}_{\mathrm{R}2} \in \mathbb{C}^{M_{\mathrm{D}k} \times M_{\mathrm{R}}}$ represent the equivalent BB channels in the first and second time slots, respectively.

For the sub-connected structure in Fig. 1, each RF chain is connected to a disjoint subset of antennas and each antenna subset has its exclusive phase shifter. Therefore, not only the analog matrices of all nodes are constrained to be BD, but also the non-zero elements in these matrices satisfy the constant-modulus constraints as follows [26]

$$\mathbf{F}_{\mathrm{S}} = \mathrm{blk}\left[\mathbf{f}_{\mathrm{S}}^1, \ldots, \mathbf{f}_{\mathrm{S}}^{M_{\mathrm{S}}}\right], |\mathbf{f}_{\mathrm{S}}^{m_{\mathrm{S}}}(j)| = \frac{1}{\sqrt{N_{\mathrm{S}}^{m_{\mathrm{S}}}}}, \tag{8}$$

$$\mathbf{F}_{\mathrm{R}1} = \mathrm{blk}\left[\mathbf{f}_{\mathrm{R}1}^1, \ldots, \mathbf{f}_{\mathrm{R}1}^{M_{\mathrm{R}}}\right], |\mathbf{f}_{\mathrm{R}1}^{m_{\mathrm{R}}}(j)| = \frac{1}{\sqrt{N_{\mathrm{R}}^{m_{\mathrm{R}}}}}, \tag{9}$$

$$\mathbf{F}_{\mathrm{R}2} = \mathrm{blk}\left[\mathbf{f}_{\mathrm{R}2}^1, \ldots, \mathbf{f}_{\mathrm{R}2}^{M_{\mathrm{R}}}\right], |\mathbf{f}_{\mathrm{R}2}^{m_{\mathrm{R}}}(j)| = \frac{1}{\sqrt{N_{\mathrm{R}}^{m_{\mathrm{R}}}}}, \tag{10}$$

$$\mathbf{F}_{\mathrm{D}k} = \mathrm{blk}\left[\mathbf{f}_{\mathrm{D}k}^1, \ldots, \mathbf{f}_{\mathrm{D}k}^{M_{\mathrm{D}k}}\right], |\mathbf{f}_{\mathrm{D}k}^{m}(j)| = \frac{1}{\sqrt{N_{\mathrm{D}k}^{m}}}, \tag{11}$$

where $\mathbf{f}_{\mathrm{S}}^{m_{\mathrm{S}}} \in \mathbb{C}^{N_{\mathrm{S}}^{m_{\mathrm{S}}}}$ $(m_{\mathrm{S}} = 1, \ldots, M_{\mathrm{S}})$, $\mathbf{f}_{\mathrm{R}1}^{m_{\mathrm{R}}} \in \mathbb{C}^{N_{\mathrm{R}}^{m_{\mathrm{R}}}}$, $\mathbf{f}_{\mathrm{R}2}^{m_{\mathrm{R}}} \in \mathbb{C}^{N_{\mathrm{R}}^{m_{\mathrm{R}}}}$ $(m_{\mathrm{R}} = 1, \ldots, M_{\mathrm{R}})$, and $\mathbf{f}_{\mathrm{D}k}^{m} \in \mathbb{C}^{N_{\mathrm{D}k}^{m}}$ $(m = 1, \ldots, M_{\mathrm{D}k})$. Based on the above hybrid beamforming system model, the achieved sum-rate of the communication system can be derived as [20], [23]

$$\begin{aligned}R = \frac{1}{2} \sum_{k=1}^{K} \log_2 (\big| \mathbf{I}_{L_{\mathrm{S}}} + \mathbf{R}_k^{-1} \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \bar{\mathbf{G}}_k \mathbf{W}_{\mathrm{R}} \bar{\mathbf{H}} \mathbf{W}_{\mathrm{S}} \\ \times \mathbf{W}_{\mathrm{S}}^{\mathrm{H}} \bar{\mathbf{H}}^{\mathrm{H}} \mathbf{W}_{\mathrm{R}}^{\mathrm{H}} \bar{\mathbf{G}}_k^{\mathrm{H}} \mathbf{W}_{\mathrm{D}k} \big|), \end{aligned} \tag{12}$$

where

$$\begin{aligned}\mathbf{R}_k =& \sigma_{\mathrm{R}}^2 \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \bar{\mathbf{G}}_k \mathbf{W}_{\mathrm{R}} \mathbf{F}_{\mathrm{R}1}^{\mathrm{H}} \mathbf{F}_{\mathrm{R}1} \mathbf{W}_{\mathrm{R}}^{\mathrm{H}} \bar{\mathbf{G}}_k^{\mathrm{H}} \mathbf{W}_{\mathrm{D}k} \\ &+ \sigma_{\mathrm{D}k}^2 \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{F}_{\mathrm{D}k}^{\mathrm{H}} \mathbf{F}_{\mathrm{D}k} \mathbf{W}_{\mathrm{D}k} \\ &+ \sum_{i=1, i \neq k}^{K} \mathbf{W}_{\mathrm{D}i}^{\mathrm{H}} \bar{\mathbf{G}}_i \mathbf{W}_{\mathrm{R}} \bar{\mathbf{H}} \mathbf{W}_{\mathrm{S}} \mathbf{W}_{\mathrm{S}}^{\mathrm{H}} \bar{\mathbf{H}}^{\mathrm{H}} \mathbf{W}_{\mathrm{R}}^{\mathrm{H}} \bar{\mathbf{G}}_i^{\mathrm{H}} \mathbf{W}_{\mathrm{D}i} \end{aligned} \tag{13}$$

is the covariance matrix of the colored Gaussian noise at the output of the BB combiner of the kth user.

Generally speaking, since the analog phase shifters are constrained by the BD and constant-modulus constraints, finding the global optima for maximizing the sum-rate in (12) is non-convex and intractable. In this paper, separated analog and digital processing designs are investigated to obtain satisfactory performance according to [34] and [35].

B. Channel Model

As shown in Fig. 1, the system includes mmWave channels $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{S}}}$ from the BS to the relay during the first time slot, and $\mathbf{G}_k \in \mathbb{C}^{N_{\mathrm{D}k} \times N_{\mathrm{R}}}$ from the relay to the kth user during the second time slot. To characterize the mathematical structure of mmWave channels, the geometric Saleh-Valenzuela channel model [36] is adopted in this paper. Due to the large PL and shadow fading, there is no direct link between the BS and users. Propagation paths of channels $\mathbf{H}$ and $\mathbf{G}_k$ are respective scattered across a sum of $N_{\mathrm{c}}^{\mathrm{R}}$ and $N_{\mathrm{c}}^{\mathrm{D}k}$ scattering clusters, each of which are respectively assumed to contribute $N_{\mathrm{p}}^{\mathrm{R}}$ and $N_{\mathrm{p}}^{\mathrm{D}k}$ propagation paths, where $k = 1, \ldots, K$. Thus, the normalized discrete-time narrowband channel matrices $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \gamma_{\mathrm{R}} \sum_{i=1}^{N_{\mathrm{c}}^{\mathrm{R}}} \sum_{l=1}^{N_{\mathrm{p}}^{\mathrm{R}}} \alpha_{i,l}^{\mathrm{R}} \Lambda_{\mathrm{r}}^{\mathrm{R}}\left(\theta_{i,l}^{\mathrm{R}}\right) \Lambda_{\mathrm{t}}^{\mathrm{R}}\left(\phi_{i,l}^{\mathrm{R}}\right) \mathbf{a}_{\mathrm{r}}\left(\theta_{i,l}^{\mathrm{R}}\right) \mathbf{a}_{\mathrm{t}}\left(\phi_{i,l}^{\mathrm{R}}\right)^{\mathrm{H}}, \tag{14}$$

and $\mathbf{G}_k$ is defined in the similar way with Dk substituting for the superscript or subscript R, where $\gamma_{\mathrm{R}} = \sqrt{\frac{N_{\mathrm{S}} N_{\mathrm{R}}}{N_{\mathrm{c}}^{\mathrm{R}} N_{\mathrm{p}}^{\mathrm{R}}}}$ and $\gamma_{\mathrm{D}k} = \sqrt{\frac{N_{\mathrm{R}} N_{\mathrm{D}k}}{N_{\mathrm{c}}^{\mathrm{D}k} N_{\mathrm{p}}^{\mathrm{D}k}}}$ represent the normalization factors. $\alpha_{i,l}$ denotes the complex gain of the lth ray in the ith scattering cluster, which follows the independent Gaussian distribution $\mathcal{CN}(0, 1)$. $\theta_{i,l}$ and $\phi_{i,l}$ are the azimuth angles of arrival (AoAs) and departure (AoDs) of the lth ray in the ith scattering cluster, which obey the truncated Laplacian distribution [34]. The functions $\Lambda_{\mathrm{t}}(\phi_{i,l})$ and $\Lambda_{\mathrm{r}}(\theta_{i,l})$ denote the transmit and receive antenna array gain at the corresponding AoDs and AoAs. $\mathbf{a}_{\mathrm{r}}(\theta_{i,l})$ and $\mathbf{a}_{\mathrm{t}}(\phi_{i,l})$ are the corresponding normalized antenna array response vectors, which are 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, the uniform linear arrays (ULAs) are examined in our study. The array response vector of an N-element can be represented as

$$\mathbf{a}^{\mathrm{ULA}}(\varphi) = \frac{1}{\sqrt{N}} \left[1, \mathrm{e}^{j\beta d \sin(\varphi)}, \ldots, \mathrm{e}^{j(N-1)\beta d \sin(\varphi)}\right]^{\mathrm{T}}, \tag{15}$$

where $j = \sqrt{-1}$, $\beta = \frac{2\pi}{\lambda}$, λ is the transmission wave length of the signal, and d is the distance between two adjacent antenna elements, e.g., $d = \frac{\lambda}{2}$.

III. Analog Beamforming/Combining and Relay BB Combining Designs

In this paper, we focus on the hybrid beamforming design of the mmWave massive MU-MIMO relay system with the sub-connected structure. The design goal is to maximize the system sum-rate in (12) by properly designing structures of precoders and combiners.

A. Analog Beamforming/Combining Designs

The objective problem can be formulated as

$$\begin{aligned}(\mathbf{F}_{\mathrm{S}}, \mathbf{F}_{\mathrm{R}1}, \mathbf{F}_{\mathrm{R}2}, \mathbf{F}_{\mathrm{D}}) = & \underset{(\mathbf{F}_{\mathrm{S}}, \mathbf{F}_{\mathrm{R}1}, \mathbf{F}_{\mathrm{R}2}, \mathbf{F}_{\mathrm{D}})}{\arg\max} R \\ & \text{s.t.} \, (8), (9), (10), (11), \end{aligned} \tag{16}$$

where $\mathbf{F}_{\mathrm{D}} = \mathrm{blk}\left[\mathbf{F}_{\mathrm{D}1}, \mathbf{F}_{\mathrm{D}2}, \ldots, \mathbf{F}_{\mathrm{D}K}\right] \in \mathbb{C}^{N_{\mathrm{D}} \times M_{\mathrm{D}}}$. The optimization problem in (16) is non-convex, which makes it intractable to find global optima directly. Fortunately, based

on the information theory [37], the original problem can be simplified by employing the idea of piecewise optimization. Firstly, we analyze the analog beamforming/combining design in the second time slot. Further, the signal transmission in the first time slot can be considered as a single-user data transmission process, and the analog beamforming/combining can be designed by utilizing the same design method.

In the second time slot, the communication process of all users can be decomposed into the inter-connected digital and analog beamforming stages. The analog beamforming stage can be expressed as

$$\hat{\mathbf{s}}_{\mathrm{D}} = \mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{G}\mathbf{F}_{\mathrm{R2}}\mathbf{s}_{\mathrm{R}} + \mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\hat{\mathbf{n}}_{\mathrm{D}}, \tag{17}$$

where $\mathbf{s}_{\mathrm{R}} = \mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\mathbf{s}$, $\hat{\mathbf{n}}_{\mathrm{D}} = \mathbf{G}\mathbf{F}_{\mathrm{R2}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{n}_{\mathrm{R}} + \mathbf{n}_{\mathrm{D}}$, $\mathbf{G} = \left[\mathbf{G}_1^{\mathrm{H}}, \ldots, \mathbf{G}_K^{\mathrm{H}}\right]^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{D}} \times N_{\mathrm{R}}}$ and $\mathbf{n}_{\mathrm{D}} = \left[\mathbf{n}_{\mathrm{D1}}^{\mathrm{T}}, \mathbf{n}_{\mathrm{D2}}^{\mathrm{T}}, \ldots, \mathbf{n}_{\mathrm{D}K}^{\mathrm{T}}\right]^{\mathrm{T}}$. It is assumed that each user's RF chain is connected to a disjoint subset of N_{D}^M antennas, i.e., $N_{\mathrm{D}}^M = N_{\mathrm{D}k}^m$ for all $m = 1, \ldots, M_{\mathrm{D}k}$ and $k = 1, \ldots, K$, and the digital beamforming matrices of all nodes and the analog beamforming matrices in the first time slot are known. The system sum-rate (12) can be represented as the mutual information between $\hat{\mathbf{s}}_{\mathrm{D}}$ and $\mathbf{s}_{\mathrm{R}}$, i.e., $R = \frac{1}{2} I\left(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}\right)$. By assuming $\mathbb{E}\left[\left\|\mathbf{s}_{\mathrm{R}}\mathbf{s}_{\mathrm{R}}^{\mathrm{H}}\right\|_{\mathrm{F}}^2\right] = \beta_{\mathrm{R}}$, which is a constant coefficient, R can be written as

$$R = \frac{1}{2}\log_2\left(\left|\mathbf{I}_{M_{\mathrm{D}}} + \frac{\beta_{\mathrm{R}}}{\hat{\sigma}_{\mathrm{D}}^2}\left(\mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{F}_{\mathrm{D}}\right)^{-1}\mathbf{G}_{\mathrm{D}}\mathbf{F}_{\mathrm{R2}}\mathbf{F}_{\mathrm{R2}}^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\right|\right), \tag{18}$$

where $\hat{\sigma}_{\mathrm{D}}^2$ is the corresponding noise variance at the analog beamforming stage and $\mathbf{G}_{\mathrm{D}} = \mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{G}$. It is observed from (18) that the system sum-rate can reach the maximum value when $\mathbf{F}_{\mathrm{R2}}$ and $\mathbf{F}_{\mathrm{D}}$ are optimal. However, since $\mathbf{F}_{\mathrm{R2}}$ and $\mathbf{F}_{\mathrm{D}}$ are non-convex, it is intractable to find their global optima directly. For the frequency division duplexing (FDD) mode, due to the large number of BS antennas, the required time to send forward-link pilots is so extensive that the forward channel estimation may be infeasible. Therefore, the time division duplexing (TDD) mode is considered in this paper, and the forward channel can be estimated by using the reciprocity of TDD channels. Then according to [20] and [38], $\mathbf{F}_{\mathrm{R2}}$ and $\mathbf{F}_{\mathrm{D}}$ can be iteratively optimized. To better describe the proposed analog beamforming scheme, we first analyze the design of $\mathbf{F}_{\mathrm{R2}}$.

Assuming $\mathbf{F}_{\mathrm{D}}$ is known, the problem of solving $\mathbf{F}_{\mathrm{R2}}$ can be formulated as

$$(\mathcal{P}_1) \quad \mathbf{F}_{\mathrm{R2}} = \underset{\mathbf{F}_{\mathrm{R2}}}{\arg\max} \;\; I\left(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}\right) \qquad \text{s.t.}\,(10)\,. \tag{19}$$

By employing the idea of SIC, we decompose $\mathcal{P}_1$ into a series of subproblems to eliminate the BD constraint (10). Each subproblem corresponds to the optimization for a disjoint subset of antennas connected to one RF chain.

Define $\mathbf{F}_{\mathrm{R2}} = \left[\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1}, \hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right]$, where $\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}$ stands for the M_{R}th column of $\mathbf{F}_{\mathrm{R2}}$, and $\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1} \in \mathbb{C}^{N_{\mathrm{R}} \times (M_{\mathrm{R}}-1)}$ contains the first $M_{\mathrm{R}}-1$ columns of $\mathbf{F}_{\mathrm{R2}}$. Due to the constant-modulus constraints of PSs, we can obtain $\mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{F}_{\mathrm{D}} = \frac{1}{N_{\mathrm{D}}^M}\mathbf{I}_{M_{\mathrm{D}}}$ and $I\left(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}\right)$ can be rewritten as

$$\begin{aligned} I(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}) =& \log_2\left(\left|\mathbf{I}_{M_{\mathrm{D}}} + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\mathbf{G}_{\mathrm{D}}\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1}\left(\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\right.\right. \\ &\left.\left. + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\mathbf{G}_{\mathrm{D}}\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\right|\right) \\ =& \log_2\left(\left|\mathbf{I}_{M_{\mathrm{D}}} + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\mathbf{T}_{M_{\mathrm{R}}-1}^{-1}\mathbf{G}_{\mathrm{D}}\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\right|\right) \\ &+ \log_2\left(\left|\mathbf{T}_{M_{\mathrm{R}}-1}\right|\right) \\ \overset{(a)}{=}& \log_2\left(1 + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\mathbf{T}_{M_{\mathrm{R}}-1}^{-1}\mathbf{G}_{\mathrm{D}}\hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right) \\ &+ \log_2\left(\left|\mathbf{T}_{M_{\mathrm{R}}-1}\right|\right), \end{aligned} \tag{20}$$

where $\mathbf{T}_{M_{\mathrm{R}}-1} = \mathbf{I}_{M_{\mathrm{D}}} + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\mathbf{G}_{\mathrm{D}}\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1}\left(\mathbf{F}_{\mathrm{R2}}^{M_{\mathrm{R}}-1}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}$, and (a) follows from the property $|\mathbf{I} + \mathbf{X}\mathbf{Y}| = |\mathbf{I} + \mathbf{Y}\mathbf{X}|$. The first term on the right side of (a) in (20) is the achievable sum-rate corresponding to the M_{R}th antenna subset. We execute the similar procedures until all M_{R} antenna subsets are considered. Then (20) can be simplified as

$$I(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}) = \sum_{u=1}^{M_{\mathrm{R}}}\log_2\left(1 + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{u}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\mathbf{T}_{u-1}^{-1}\mathbf{G}_{\mathrm{D}}\hat{\mathbf{f}}_{\mathrm{R2}}^{u}\right), \tag{21}$$

where $\mathbf{T}_0 = \mathbf{I}_{M_{\mathrm{D}}}$ and

$$\mathbf{T}_{u-1} = \mathbf{I}_{M_{\mathrm{D}}} + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\mathbf{G}_{\mathrm{D}}\mathbf{F}_{\mathrm{R2}}^{u-1}\left(\mathbf{F}_{\mathrm{R2}}^{u-1}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{D}}^{\mathrm{H}}. \tag{22}$$

It can be seen from (21) that maximizing each of the summation terms in turn, the analog channel capacity can achieve the maximum. Due to the BD constraint (10), only $N_{\mathrm{R}}^{m_{\mathrm{R}}}$ non-zero elements in $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$, i.e., $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}(j) \neq 0$ for $j = \sum_{m_{\mathrm{R}}=1}^{u-1} N_{\mathrm{R}}^{m_{\mathrm{R}}} + 1, \ldots, \sum_{m_{\mathrm{R}}=1}^{u} N_{\mathrm{R}}^{m_{\mathrm{R}}}$ and $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}(j) = 0$ for otherwise, contribute to matrix multiplying operation. Let $\mathbf{Z}_{u-1} = \mathbf{G}_{\mathrm{D}}^{\mathrm{H}}\mathbf{T}_{u-1}^{-1}\mathbf{G}_{\mathrm{D}}$, and $\mathbf{S}_{u-1} \in \mathbb{C}^{N_{\mathrm{R}}^{m_{\mathrm{R}}} \times N_{\mathrm{R}}^{m_{\mathrm{R}}}}$ keeps the non-zero elements of $\mathbf{Z}_{u-1}$, i.e., $\mathbf{S}_{u-1} = \mathbf{\Theta}_{u-1}\mathbf{Z}_{u-1}\mathbf{\Theta}_{u-1}^{\mathrm{H}}$, where $\mathbf{\Theta}_{u-1} = \left[\mathbf{0}_{N_{\mathrm{R}}^{m_{\mathrm{R}}} \times \sum_{m_{\mathrm{R}}=1}^{u-1} N_{\mathrm{R}}^{m_{\mathrm{R}}}}, \mathbf{I}_{N_{\mathrm{R}}^{m_{\mathrm{R}}}}, \mathbf{0}_{N_{\mathrm{R}}^{m_{\mathrm{R}}} \times \left(N_{\mathrm{R}} - \sum_{m_{\mathrm{R}}=1}^{u} N_{\mathrm{R}}^{m_{\mathrm{R}}}\right)}\right]$ is the selection matrix. Hence (21) can be rewritten as

$$I\left(\hat{\mathbf{s}}_{\mathrm{D}}; \mathbf{s}_{\mathrm{R}}\right) = \sum_{u=1}^{M_{\mathrm{R}}}\log_2\left(\left|1 + \frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^M}{\hat{\sigma}_{\mathrm{D}}^2}\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{u}\right)^{\mathrm{H}}\mathbf{S}_{u-1}\hat{\mathbf{f}}_{\mathrm{R2}}^{u}\right|\right). \tag{23}$$

The optimal solution of $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$ is related to the right singular vector $\mathbf{v}_1$ corresponding to the maximum singular value of $\mathbf{S}_{u-1}$, i.e.,

$$\left(\hat{\mathbf{f}}_{\mathrm{R2}}^{u}\right)_{\mathrm{opt}} = \left[\mathbf{0}_{1 \times \sum_{m_{\mathrm{R}}=1}^{u-1} N_{\mathrm{R}}^{m_{\mathrm{R}}}}, \mathbf{v}_1^{\mathrm{T}}, \mathbf{0}_{1 \times \sum_{m_{\mathrm{R}}=u}^{M_{\mathrm{R}}} N_{\mathrm{R}}^{m_{\mathrm{R}}}}\right]^{\mathrm{T}}. \tag{24}$$

However, the elements in $\mathbf{v}_1$ do not follow the constant-modulus constraint, which is not appropriate for the design of $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$. Therefore, the optimization problem (19) can be reformulated to obtain the MMSE approximation between

$\left(\hat{\mathbf{f}}_{\mathrm{R}2}^{u}\right)_{\text{opt}}$ and $\hat{\mathbf{f}}_{\mathrm{R}2}^{u}$ on the constant-modulus constraint, i.e.,

$$\arg\min_{\hat{\mathbf{f}}_{\mathrm{R}2}^{u}} \mathbb{E}\left[\left\|\left(\hat{\mathbf{f}}_{\mathrm{R}2}^{u}\right)_{\text{opt}}-\hat{\mathbf{f}}_{\mathrm{R}2}^{u}\right\|_2^2\right]. \tag{25}$$

According to [37], $\hat{\mathbf{f}}_{\mathrm{R}2}^{u}$ can be expressed as

$$\hat{\mathbf{f}}_{\mathrm{R}2}^{u}=\frac{1}{\sqrt{N_{\mathrm{R}}^{m_{\mathrm{R}}}}}\mathrm{e}^{j\angle\left(\hat{\mathbf{f}}_{\mathrm{R}2}^{u}\right)_{\text{opt}}}. \tag{26}$$

The analog beamforming $\mathbf{F}_{\mathrm{R}2}$ can be obtained by repeating above processes until its all M_{R} columns are optimized.

By defining $\mathbf{G}_{\mathrm{R}}=\left(\mathbf{G}\mathbf{F}_{\mathrm{R}2}\right)^{\mathrm{H}}$, $I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}_{\mathrm{R}}\right)$ can also be written as

$$I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}_{\mathrm{R}}\right)=\log_2\left(\left|\mathbf{I}_{M_{\mathrm{R}}}+\frac{\beta_{\mathrm{R}}N_{\mathrm{D}}^{M}}{\sigma_{\mathrm{D}}^2}\mathbf{G}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}\mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{G}_{\mathrm{R}}^{\mathrm{H}}\right|\right). \tag{27}$$

Similar to (19), the subproblem of solving $\mathbf{F}_{\mathrm{D}}$ can be formulated as

$$(\mathcal{P}_2)\quad \mathbf{F}_{\mathrm{D}}=\arg\max_{\mathbf{F}_{\mathrm{D}}}\ I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}_{\mathrm{R}}\right)\qquad \text{s.t.}\,(11)\,. \tag{28}$$

The analog combiner $\mathbf{F}_{\mathrm{D}}$ is then obtained by performing similar procedures of solving the problem $(\mathcal{P}_1)$ in (19).

Similarly, assuming the channel $\mathbf{H}$ is known, the analog beamformer and combiner $(\mathbf{F}_{\mathrm{S}},\mathbf{F}_{\mathrm{R}1})$ in the first time slot are obtained by employing the same design method above.

B. Relay BB Combining Design

In this subsection, we focus on designing the relay BB beamforming matrix $\mathbf{W}_{\mathrm{R}}$. The system sum-rate (12) can be expressed by the mutual information between $\mathbf{x}_{\mathrm{D}}=\left[\mathbf{x}_{\mathrm{D}1}^{\mathrm{T}},\ldots,\mathbf{x}_{\mathrm{D}K}^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\mathbf{s}$, i.e., $R=\frac{1}{2}I\left(\mathbf{x}_{\mathrm{D}};\mathbf{s}\right)$. Exploiting the information theory, $I\left(\mathbf{x}_{\mathrm{D}};\mathbf{s}\right)\leqslant I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)$ holds. Define $\bar{\mathbf{G}}=\left[\bar{\mathbf{G}}_1^{\mathrm{H}},\ldots,\bar{\mathbf{G}}_K^{\mathrm{H}}\right]^{\mathrm{H}}$. If $\mathbf{W}_{\mathrm{D}}$ is obtained by the singular value decomposition (SVD) of $\bar{\mathbf{G}}$, the digital combining process can be regarded as information lossless [39]. In this way, maximizing the system sum-rate can be approximately equal to maximize the mutual information between $\hat{\mathbf{s}}_{\mathrm{D}}$ and $\mathbf{s}$, i.e., $I\left(\mathbf{x}_{\mathrm{D}};\mathbf{s}\right)\approx I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)$, and

$$I(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s})=\log_2\left(\left|\mathbf{I}_{KL_{\mathrm{S}}}+\left(\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right|\right), \tag{29}$$

where $\bar{\mathbf{R}}$ denotes the effective noise covariance matrix as

$$\bar{\mathbf{R}}=\sigma_{\mathrm{R}}^2\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R}1}^{\mathrm{H}}\mathbf{F}_{\mathrm{R}1}\mathbf{W}_{\mathrm{R}}^{\mathrm{H}}\bar{\mathbf{G}}^{\mathrm{H}}+\sigma_{\mathrm{D}}^2\mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{F}_{\mathrm{D}}, \tag{30}$$

and $\sigma_{\mathrm{D}}^2=\sum_{k=1}^{K}\sigma_{\mathrm{D}k}^2$. Therefore, the problem of solving $\mathbf{W}_{\mathrm{R}}$ can be formulated as

$$\mathbf{W}_{\mathrm{R}}=\arg\max_{\mathbf{W}_{\mathrm{R}}} I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)\qquad \text{s.t.}\,(5)\,. \tag{31}$$

The MMSE combiner of all K users $\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}=\mathrm{blk}\left[\mathbf{W}_{\mathrm{D}1}^{\mathrm{MMSE}},\ldots,\mathbf{W}_{\mathrm{D}K}^{\mathrm{MMSE}}\right]\in\mathbb{C}^{M_{\mathrm{D}}\times KL_{\mathrm{S}}}$ can be obtained by [32]

$$\begin{aligned}\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}&=\arg\min_{\mathbf{W}_{\mathrm{D}}}\mathbb{E}\left[\left\|\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\hat{\mathbf{s}}_{\mathrm{D}}-\mathbf{s}\right\|_2^2\right]\\&=\left(\left(\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}}\left(\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\left(\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}}+\bar{\mathbf{R}}\right)^{-1}\right)^{\mathrm{H}}.\end{aligned} \tag{32}$$

Hence the MSE function can be formulated as [40]

$$\begin{aligned}\mathbf{E}&=\mathbb{E}\left[\left(\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\hat{\mathbf{s}}_{\mathrm{D}}-\mathbf{s}\right)\left(\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\hat{\mathbf{s}}_{\mathrm{D}}-\mathbf{s}\right)^{\mathrm{H}}\right]\\&=\left(\mathbf{I}_{KL_{\mathrm{S}}}-\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)\\&\quad\left(\mathbf{I}_{KL_{\mathrm{S}}}-\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}}+\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\bar{\mathbf{R}}\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\\&=\left(\mathbf{I}_{KL_{\mathrm{S}}}+\left(\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{-1}.\end{aligned} \tag{33}$$

Clearly, the relationship between the mutual information $I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)$ and (33) can be expressed as [32]

$$I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)=\log_2\left(\left|\mathbf{E}^{-1}\right|\right). \tag{34}$$

To solve the relay BB combining $\mathbf{W}_{\mathrm{R}}$, we need to calculate the partial derivative of $I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)$ with respect to (w.r.t.) $\mathbf{W}_{\mathrm{R}}$, i.e.,

$$\begin{aligned}\frac{\partial I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)}{\partial\mathbf{W}_{\mathrm{R}}}&=\frac{\partial\log_2\left(\left|\mathbf{E}^{-1}\right|\right)}{\partial\mathbf{W}_{\mathrm{R}}}\\&\overset{\text{(a)}}{=}-\frac{\mathrm{Tr}\left(\mathbf{E}^{-1}\partial\mathbf{E}\right)}{\left(\log 2\right)\partial\mathbf{W}_{\mathrm{R}}}\\&\overset{\text{(b)}}{=}\frac{\mathrm{Tr}\left(\mathbf{E}^{-1}\mathbf{E}\partial\mathbf{E}^{-1}\mathbf{E}\right)}{\left(\log 2\right)\partial\mathbf{W}_{\mathrm{R}}},\end{aligned} \tag{35}$$

where (a) and (b) follow from the properties $\partial\log\left(|\mathbf{X}|\right)=\mathrm{Tr}\left(\mathbf{X}^{-1}\partial\mathbf{X}\right)$ and $\partial\left(\mathbf{X}^{-1}\right)=-\mathbf{X}^{-1}\partial\left(\mathbf{X}\right)\mathbf{X}^{-1}$, respectively. By treating $\mathbf{A}=\frac{\mathbf{E}^{-1}}{\log 2}$ as an equivalent constant weight matrix, then (35) can be further written as

$$\begin{aligned}\frac{\partial I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)}{\partial\mathbf{W}_{\mathrm{R}}}&=\frac{\mathrm{Tr}\left(\mathbf{A}\mathbf{E}\partial\mathbf{E}^{-1}\mathbf{E}\right)}{\partial\mathbf{W}_{\mathrm{R}}}\\&=\frac{\mathrm{Tr}\left(\mathbf{A}\mathbf{E}\partial\mathbf{E}^{-1}\mathbf{E}-\left(\partial\mathbf{A}\right)\mathbf{E}\right)}{\partial\mathbf{W}_{\mathrm{R}}}\\&=-\frac{\mathrm{Tr}\left(\left(\partial\mathbf{A}\right)\mathbf{E}+\mathbf{A}\left(\partial\mathbf{E}\right)\right)}{\partial\mathbf{W}_{\mathrm{R}}}.\end{aligned} \tag{36}$$

According to the partial derivative formula of the binary function, i.e., $\partial\left(\mathbf{X}\mathbf{Y}\right)=\left(\partial\mathbf{X}\right)\mathbf{Y}+\mathbf{X}\left(\partial\mathbf{Y}\right)$, (36) can be rewritten as

$$\frac{\partial I\left(\hat{\mathbf{s}}_{\mathrm{D}};\mathbf{s}\right)}{\partial\mathbf{W}_{\mathrm{R}}}=-\frac{\partial\mathrm{Tr}\left(\mathbf{A}\mathbf{E}\right)}{\partial\mathbf{W}_{\mathrm{R}}}. \tag{37}$$

Substituting (37) into (31), the problem (31) can be converted to

$$(\mathcal{P}_3)\quad \mathbf{W}_{\mathrm{R}}=\arg\max_{\mathbf{W}_{\mathrm{R}}}-\mathrm{Tr}\left(\mathbf{A}\mathbf{E}\right)\qquad \text{s.t.}\,(5)\,. \tag{38}$$

The problem $(\mathcal{P}_3)$ is a convex optimization for $\mathbf{W}_{\mathrm{R}}$. Considering the power constraint of relay in (5), the Lagrange function of problem (38) is denoted as

$$\begin{aligned}L\left(\mathbf{W}_{\mathrm{R}},\lambda\right)=&\mathrm{Tr}\left(\mathbf{A}\mathbf{E}\right)+\lambda\left[\mathrm{Tr}\left(\mathbf{F}_{\mathrm{R}2}\mathbf{W}_{\mathrm{R}}\left(\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}^{\mathrm{H}}\bar{\mathbf{H}}^{\mathrm{H}}\right.\right.\right.\\&\left.\left.\left.+\sigma_{\mathrm{R}}^2\mathbf{F}_{\mathrm{R}1}\mathbf{F}_{\mathrm{R}1}^{\mathrm{H}}\right)\left(\mathbf{F}_{\mathrm{R}2}\mathbf{W}_{\mathrm{R}}\right)^{\mathrm{H}}\right)-P_{\mathrm{R}}\right],\end{aligned} \tag{39}$$

where $\lambda\geq 0$ is the Lagrange multiplier, which can be obtained by the bisection method. By calculating the partial derivative

of (39) w.r.t. $\mathbf{W}_{\mathrm{R}}$, the relay BB combiner can be designed as

$$\begin{aligned}\hat{\mathbf{W}}_{\mathrm{R}} =& \left(\bar{\mathbf{G}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\mathbf{A}\left(\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\right)^{\mathrm{H}}\bar{\mathbf{G}} + \lambda\mathbf{F}_{\mathrm{R2}}^{\mathrm{H}}\mathbf{F}_{\mathrm{R2}}\right)^{-1} \\ &\times \bar{\mathbf{G}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}\mathbf{A}\left(\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}} \\ &\times \left(\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\left(\bar{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\right)^{\mathrm{H}} + \sigma_{\mathrm{R}}^2\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{F}_{\mathrm{R1}}\right)^{-1}. \end{aligned} \tag{40}$$

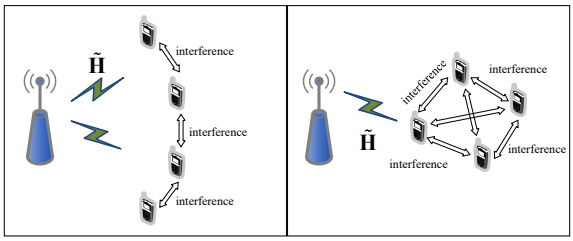
In conclusion, the proposed analog beamforming/combining and relay BB combining algorithms for the massive MU-MIMO relay system are summarized as **Algorithm 1**.

Algorithm 1 : The proposed analog beamforming/combining and relay BB combining algorithms

Input: $\mathbf{H}$, $\mathbf{G}$, $N1_{\mathrm{iter}}$, $N2_{\mathrm{iter}}$, P_{R} and λ;
Randomly generate $\mathbf{F}_{\mathrm{D}}$;
for $i = 1$ to $N1_{\mathrm{iter}}$ **do**
  Compute $\mathbf{G}_{\mathrm{D}} = \mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{G}$;
  for $u = 1$ to M_{R} **do**
    Compute $\mathbf{S}_{u-1} = \mathbf{\Theta}_{u-1}\mathbf{Z}_{u-1}\mathbf{\Theta}_{u-1}^{\mathrm{H}}$;
    Compute the right singular vector $\mathbf{v}_1$ of $\mathbf{S}_{u-1}$;
    Compute $\hat{\mathbf{f}}_{\mathrm{R2}}^{u}$ according to (26);
    Compute $\mathbf{T}_u$ according to (22);
  end for
  Obtain $\mathbf{F}_{\mathrm{R2}} = \left[\hat{\mathbf{f}}_{\mathrm{R2}}^{1}, \ldots, \hat{\mathbf{f}}_{\mathrm{R2}}^{M_{\mathrm{R}}}\right]$;
  Compute $\mathbf{G}_{\mathrm{R}} = \left(\mathbf{G}\mathbf{F}_{\mathrm{R2}}\right)^{\mathrm{H}}$;
  Compute $\mathbf{F}_{\mathrm{D}k} = \left[\hat{\mathbf{f}}_{\mathrm{D}k}^{1}, \ldots, \hat{\mathbf{f}}_{\mathrm{D}k}^{M_{\mathrm{D}_k}}\right]$ and $\mathbf{F}_{\mathrm{D}} = \mathrm{blk}\left[\mathbf{F}_{\mathrm{D}1}, \ldots, \mathbf{F}_{\mathrm{D}K}\right]$ by using steps from 5 to 10, replacing $\mathbf{G}_{\mathrm{D}}$ by $\mathbf{G}_{\mathrm{R}}$ and setting the number of cycles to M_{D};
  Similarly, obtain $\mathbf{F}_{\mathrm{S}} = \left[\hat{\mathbf{f}}_{\mathrm{S}}^{1}, \ldots, \hat{\mathbf{f}}_{\mathrm{S}}^{M_{\mathrm{S}}}\right]$ and $\mathbf{F}_{\mathrm{R1}} = \left[\hat{\mathbf{f}}_{\mathrm{R1}}^{1}, \ldots, \hat{\mathbf{f}}_{\mathrm{R1}}^{M_{\mathrm{R}}}\right]$ by using steps from 4 to 13;
end for
Obtain the equivalent BB channel $\bar{\mathbf{H}} = \mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{S}}$, $\bar{\mathbf{G}}_k = \mathbf{F}_{\mathrm{D}k}^{\mathrm{H}}\mathbf{G}_k\mathbf{F}_{\mathrm{R2}}$, and $\bar{\mathbf{G}} = \left[\bar{\mathbf{G}}_1^{\mathrm{H}}, \ldots, \bar{\mathbf{G}}_K^{\mathrm{H}}\right]^{\mathrm{H}}$;
Randomly generate $\mathbf{W}_{\mathrm{R}}$ and $\mathbf{W}_{\mathrm{S}}$;
for $i = 1$ to $N2_{\mathrm{iter}}$ **do**
  Compute $\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}$ and $\mathbf{E}$ according to (32) and (33), respectively;
  Update $\mathbf{A}$ via $\mathbf{A} = \frac{\mathbf{E}^{-1}}{\log 2}$;
  Obtain $\mathbf{W}_{\mathrm{R}}$ according to (40);
end for
Output: $\mathbf{F}_{\mathrm{S}}$, $\mathbf{F}_{\mathrm{R1}}$, $\mathbf{F}_{\mathrm{R2}}$, $\mathbf{F}_{\mathrm{D}}$, and $\mathbf{W}_{\mathrm{R}}$.

IV. Digital Beamforming/Combining Designs for BS and Users

In this section, we propose two different BS and users digital beamforming/combining design methods for the two following scenarios. The first scenario: the distance between non-adjacent users is large and users are evenly distributed around the relay, i.e., there is only interference between adjacent users. The second scenario: the distance between users is relatively small and users are densely distributed around the relay, i.e., there is mutual interference between different users. Schematic diagram of the two distribution scenarios is shown in Fig. 2.



(a) The first scenario (b) The second scenario

Fig. 2. Schematic diagram of the two distribution scenarios with interference.

Define the equivalent BB channel between the BS and all users as [41]

$$\tilde{\mathbf{H}} = \bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\bar{\mathbf{H}} = \left[\tilde{\mathbf{H}}_1^{\mathrm{H}}, \ldots, \tilde{\mathbf{H}}_K^{\mathrm{H}}\right]^{\mathrm{H}}, \tag{41}$$

where $\tilde{\mathbf{H}}_k$ is the equivalent BB channel from BS to the kth user and $k = 1, \ldots, K$. The considered mmWave relay system can be regarded as a point-to-point single-hop system from BS to users. Therefore, the problem of solving $\mathbf{W}_{\mathrm{D}k}$ and $\mathbf{W}_{\mathrm{S}}$ can be formulated as

$$\begin{aligned}(\mathcal{P}_4) \quad (\mathbf{W}_{\mathrm{D}}, \mathbf{W}_{\mathrm{S}}) &= \arg\max_{\mathbf{W}_{\mathrm{D}}, \mathbf{W}_{\mathrm{S}}} I\left(\mathbf{x}_{\mathrm{D}}; \mathbf{s}\right) \\ &\text{s.t.} \mathbf{W}_{\mathrm{D}} = \mathrm{blk}\left[\mathbf{W}_{\mathrm{D}1}, \ldots, \mathbf{W}_{\mathrm{D}K}\right]. \end{aligned} \tag{42}$$

Assume the digital beamforming $\mathbf{W}_{\mathrm{S}}$ is an unitary matrix[3], i.e., $\mathbf{W}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}^{\mathrm{H}} = \mathbf{I}_{M_{\mathrm{S}}}$, $I\left(\mathbf{x}_{\mathrm{D}}; \mathbf{s}\right)$ can be rewritten as

$$\begin{aligned} I\left(\mathbf{x}_{\mathrm{D}}; \mathbf{s}\right) &= \log_2\left(\left|\mathbf{I}_{KL_{\mathrm{S}}} + \tilde{\mathbf{R}}^{-1}\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\tilde{\mathbf{H}}\mathbf{W}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}^{\mathrm{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}\right|\right) \\ &= \log_2\left(\left|\mathbf{I}_{KL_{\mathrm{S}}} + \tilde{\mathbf{R}}^{-1}\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}\right|\right), \end{aligned} \tag{43}$$

where

$$\tilde{\mathbf{R}} = \sigma_{\mathrm{R}}^2\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\bar{\mathbf{G}}\mathbf{W}_{\mathrm{R}}\mathbf{F}_{\mathrm{R1}}^{\mathrm{H}}\mathbf{F}_{\mathrm{R1}}\mathbf{W}_{\mathrm{R}}^{\mathrm{H}}\bar{\mathbf{G}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}} + \sigma_{\mathrm{D}}^2\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\mathbf{F}_{\mathrm{D}}^{\mathrm{H}}\mathbf{F}_{\mathrm{D}}. \tag{44}$$

If the signal received by the kth user can be guaranteed to be in the null space of other user channels, the IUI can be eliminated [42], i.e.,

$$\mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k\mathbf{W}_{\mathrm{S}}^{i} = \mathbf{0}, \quad \forall k \neq i, \tag{45}$$

where $\mathbf{W}_{\mathrm{S}}^{k} \in \mathbb{C}^{M_{\mathrm{S}} \times L_{\mathrm{S}}}$ is the kth digital beamforming matrix of $\mathbf{W}_{\mathrm{S}} = \left[\mathbf{W}_{\mathrm{S}}^{1}, \ldots, \mathbf{W}_{\mathrm{S}}^{K}\right]$ and $k, i = 1, \ldots, K$. Define the SVD of $\tilde{\mathbf{H}}_k$ as

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k\tilde{\mathbf{\Sigma}}_k\left[\tilde{\mathbf{V}}_k^{(1)}, \tilde{\mathbf{V}}_k^{(0)}\right]^{\mathrm{H}}, \tag{46}$$

where $\tilde{\mathbf{V}}_k^{(1)} = \tilde{\mathbf{V}}_k\left(:, 1 : L_{\mathrm{S}}\right)$ and $\tilde{\mathbf{V}}_k^{(0)} = \tilde{\mathbf{V}}_k\left(:, L_{\mathrm{S}} + 1 : \mathrm{end}\right)$ stand for the subspace orthogonal bases and null space orthogonal bases of $\tilde{\mathbf{H}}_k$, respectively, with $\tilde{\mathbf{H}}_k\tilde{\mathbf{V}}_k^{(0)} = \mathbf{0}$. The digital combining matrix for the kth user is designed as

$$\mathbf{W}_{\mathrm{D}k} = \tilde{\mathbf{U}}_k\left(:, 1 : L_{\mathrm{S}}\right). \tag{47}$$

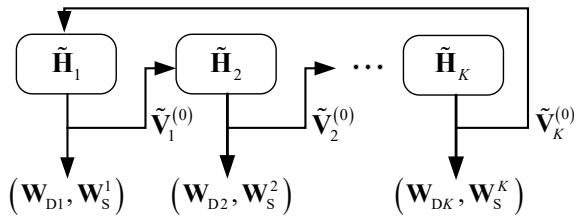


Fig. 3. Diagram of the proposed SSIC-based digital beamforming.

A. The First Scenario

The BB-BD technology is widely used to eliminate IUI. For the first scenario, the BB-BD based SSIC method is proposed to design the digital combining of the BS, and hence the interference between two adjacent users is eliminated successively.

It can be seen from (43) that maximizing $I\left(\mathbf{x}_{\mathrm{D}}; \mathbf{s}\right)$ is equivalent to optimizing $\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}$. Further, $\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}}$ can be written as (48) shown at the top of the next page, where (a) follows that the interference between non-adjacent users is zero in the first scenario. By utilizing SVD, $\mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k\tilde{\mathbf{V}}_{k-1}^{(0)} = \bar{\mathbf{U}}_k\bar{\mathbf{\Sigma}}_k\bar{\mathbf{V}}_k^{\mathrm{H}}$, where $\tilde{\mathbf{V}}_{k-1}^{(0)}$ is the null space orthogonal bases of $\tilde{\mathbf{H}}_{k-1}$ and $\tilde{\mathbf{V}}_0^{(0)} = \tilde{\mathbf{V}}_K^{(0)}$. Therefore, the kth digital beamforming matrix of BS can be designed as

$$\mathbf{W}_{\mathrm{S}}^k = \tilde{\mathbf{V}}_{k-1}^{(0)}\bar{\mathbf{V}}_k\left(:, 1:L_{\mathrm{S}}\right), \tag{49}$$

which satisfies (45).

As mentioned above, the procedures of the proposed SSIC-based digital beamforming/combining method for the first scenario are visually illustrated in Fig. 3 and formally described in **Algorithm 2**.

Algorithm 2 : The proposed SSIC-based digital beamforming/combining method

1: **Input**: $\bar{\mathbf{H}}$, $\bar{\mathbf{G}}$, and $\mathbf{W}_{\mathrm{R}}$;
2: Obtain the equivalent BB channel $\tilde{\mathbf{H}}_k$ according to (41);
3: **for** $k = 1$ to K **do**
4: Obtain $\tilde{\mathbf{V}}_k^{(0)}$ and $\mathbf{W}_{\mathrm{D}k}$ according to (46) and (47), respectively;
5: Perform SVD to $\mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k\tilde{\mathbf{V}}_{k-1}^{(0)}$;
6: Obtain $\mathbf{W}_{\mathrm{S}}^k$ according to (49);
7: **end for**
8: Compute $\mathbf{W}_{\mathrm{S}} = \left[\mathbf{W}_{\mathrm{S}}^1, \ldots, \mathbf{W}_{\mathrm{S}}^K\right]$ and $\mathbf{W}_{\mathrm{D}} = \mathrm{blk}\left[\mathbf{W}_{\mathrm{D}1}, \ldots, \mathbf{W}_{\mathrm{D}K}\right]$;
9: Adjust each column of $\mathbf{W}_{\mathrm{S}}$ to be $\mathbf{W}_{\mathrm{S}}\left(:, i\right) = \frac{\mathbf{W}_{\mathrm{S}}(:,i)}{\|\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}(:,i)\|_{\mathrm{F}}}, i \in \{1, \ldots, KL_{\mathrm{S}}\}$;
10: **Output**: $\mathbf{W}_{\mathrm{S}}$ and $\mathbf{W}_{\mathrm{D}}$.

B. The Second Scenario

If the number of users becomes larger, the BB-BD technology can result in poor performance and high computational complexity since the row subspaces of user channels will overlap significantly [43]. Based on the criterion of avoiding information loss, the PSA-based method is proposed to design the digital beamforming of the BS for the second scenario as illustrated in Fig. 2 and eliminate the interference between different users.

By applying SVD to the composite channel matrix $\tilde{\mathbf{H}}_{\mathrm{comp}}^k = \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k$, we obtain

$$\begin{aligned}\tilde{\mathbf{H}}_{\mathrm{comp}}^k &= \tilde{\mathbf{U}}_{\mathrm{comp}}^k\tilde{\mathbf{\Sigma}}_{\mathrm{comp}}^k\left(\tilde{\mathbf{V}}_{\mathrm{comp}}^k\right)^{\mathrm{H}} \\ &\approx \tilde{\mathbf{U}}_{\mathrm{comp}}^k(:,1:L_{\mathrm{S}})\tilde{\mathbf{\Sigma}}_{\mathrm{comp}}^k(1:L_{\mathrm{S}},1:L_{\mathrm{S}})\left(\tilde{\mathbf{V}}_{\mathrm{comp}}^k(:,1:L_{\mathrm{S}})\right)^{\mathrm{H}}. \end{aligned}\tag{50}$$

Substituting (47) into $\tilde{\mathbf{H}}_{\mathrm{comp}}^k$, we have

$$\tilde{\mathbf{H}}_{\mathrm{comp}}^k \approx \mathbf{I}_{L_{\mathrm{S}}}\tilde{\mathbf{\Sigma}}_k\left(1:L_{\mathrm{S}}, 1:L_{\mathrm{S}}\right)\left(\tilde{\mathbf{V}}_k\left(:, 1:L_{\mathrm{S}}\right)\right)^{\mathrm{H}}. \tag{51}$$

According to (50) and (51), $\tilde{\mathbf{V}}_{\mathrm{comp}}^k\left(:, 1:L_{\mathrm{S}}\right) = \tilde{\mathbf{V}}_k\left(:, 1:L_{\mathrm{S}}\right)$ is obtained. For massive MIMO systems, the first L_{S} right singular vectors belonging to two different user channels are asymptotically mutual orthogonal, which has been proven in [39], i.e.,

$$\lim_{N_{\mathrm{S}}\to\infty}\tilde{\mathbf{V}}_p^{\mathrm{H}}\left(1:L_{\mathrm{S}}, :\right)\tilde{\mathbf{V}}_q\left(:, 1:L_{\mathrm{S}}\right) = \mathbf{0}, \quad p \neq q, \tag{52}$$

where $p, q \in \{1, \ldots, K\}$. The kth digital beamforming matrix of BS can be designed as

$$\mathbf{W}_{\mathrm{S}}^k = \tilde{\mathbf{V}}_{\mathrm{comp}}^k\left(:, 1:L_{\mathrm{S}}\right), \tag{53}$$

which satisfies (45), and $\mathbf{W}_{\mathrm{S}} = \tilde{\mathbf{V}}_{\mathrm{comp}} = \left[\tilde{\mathbf{V}}_{\mathrm{comp}}^1\left(:, 1:L_{\mathrm{S}}\right), \ldots, \tilde{\mathbf{V}}_{\mathrm{comp}}^K\left(:, 1:L_{\mathrm{S}}\right)\right] \in \mathbb{C}^{M_{\mathrm{S}}\times KL_{\mathrm{S}}}$ is an unitary matrix.

To sum up, the proposed PSA-based digital beamforming/combining design algorithm is summarized in **Algorithm 3**.

Algorithm 3 : The proposed PSA-based digital beamforming/combining method

1: **Input**: $\bar{\mathbf{H}}$, $\bar{\mathbf{G}}$, and $\mathbf{W}_{\mathrm{R}}$;
2: Obtain the equivalent BB channel $\tilde{\mathbf{H}}_k$ according to (41);
3: **for** $k = 1$ to K **do**
4: Obtain $\mathbf{W}_{\mathrm{D}k}$ according to (47);
5: Obtain the composite channel matrix $\tilde{\mathbf{H}}_{\mathrm{comp}}^k = \mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k$;
6: Perform SVD to $\tilde{\mathbf{H}}_{\mathrm{comp}}^k$;
7: Obtain $\mathbf{W}_{\mathrm{S}}^k$ according to (53);
8: **end for**
9: Compute $\mathbf{W}_{\mathrm{S}} = \left[\mathbf{W}_{\mathrm{S}}^1, \ldots, \mathbf{W}_{\mathrm{S}}^K\right]$ and $\mathbf{W}_{\mathrm{D}} = \mathrm{blk}\left[\mathbf{W}_{\mathrm{D}1}, \ldots, \mathbf{W}_{\mathrm{D}K}\right]$;
10: Adjust each column of $\mathbf{W}_{\mathrm{S}}$ to be $\mathbf{W}_{\mathrm{S}}\left(:, i\right) = \frac{\mathbf{W}_{\mathrm{S}}(:,i)}{\|\mathbf{F}_{\mathrm{S}}\mathbf{W}_{\mathrm{S}}(:,i)\|_{\mathrm{F}}}, i \in \{1, \ldots, KL_{\mathrm{S}}\}$;
11: **Output**: $\mathbf{W}_{\mathrm{S}}$ and $\mathbf{W}_{\mathrm{D}}$.

C. Complexity

In this subsection, we evaluate the complexity of the proposed methods in terms of the required number of complex multiplications. In **Algorithm 1**, $\mathbf{F}_{\mathrm{S}}$, $\mathbf{F}_{\mathrm{R}1}$, $\mathbf{F}_{\mathrm{R}2}$, $\mathbf{F}_{\mathrm{D}}$,

[3]The digital beamformer $\mathbf{W}_{\mathrm{S}}$ is derived based on SVD, which will be introduced in the following part. Therefore, it is reasonable to assume $\mathbf{W}_{\mathrm{S}}$ is an unitary matrix.

$$\mathbf{W}_{\mathrm{D}}^{\mathrm{H}}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{\mathrm{D}} \overset{\text{(a)}}{=} \begin{bmatrix} \mathbf{W}_{\mathrm{D}1}^{\mathrm{H}}\tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_1^{\mathrm{H}}\mathbf{W}_{\mathrm{D}1} & \mathbf{W}_{\mathrm{D}1}^{\mathrm{H}}\tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_2^{\mathrm{H}}\mathbf{W}_{\mathrm{D}2} & \mathbf{0} & \cdots & \mathbf{W}_{\mathrm{D}1}^{\mathrm{H}}\tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_K^{\mathrm{H}}\mathbf{W}_{\mathrm{D}K} \\ \mathbf{W}_{\mathrm{D}2}^{\mathrm{H}}\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_1^{\mathrm{H}}\mathbf{W}_{\mathrm{D}1} & \mathbf{W}_{\mathrm{D}2}^{\mathrm{H}}\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^{\mathrm{H}}\mathbf{W}_{\mathrm{D}2} & \mathbf{W}_{\mathrm{D}2}^{\mathrm{H}}\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_3^{\mathrm{H}}\mathbf{W}_{\mathrm{D}3} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\mathrm{D}3}^{\mathrm{H}}\tilde{\mathbf{H}}_3\tilde{\mathbf{H}}_2^{\mathrm{H}}\mathbf{W}_{\mathrm{D}2} & \mathbf{W}_{\mathrm{D}3}^{\mathrm{H}}\tilde{\mathbf{H}}_3\tilde{\mathbf{H}}_3^{\mathrm{H}}\mathbf{W}_{\mathrm{D}3} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{\mathrm{D}K}^{\mathrm{H}}\tilde{\mathbf{H}}_K\tilde{\mathbf{H}}_1^{\mathrm{H}}\mathbf{W}_{\mathrm{D}1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{\mathrm{D}K}^{\mathrm{H}}\tilde{\mathbf{H}}_K\tilde{\mathbf{H}}_K^{\mathrm{H}}\mathbf{W}_{\mathrm{D}K} \end{bmatrix} \tag{48}$$

TABLE II
COMPUTATIONAL COMPLEXITY OF THE PROPOSED METHODS

Operation			Complexity
Algorithm 1			
	Step 16	$\bar{\mathbf{H}}$	$M_{\mathrm{R}}N_{\mathrm{S}}N_{\mathrm{R}} + M_{\mathrm{R}}N_{\mathrm{S}}M_{\mathrm{S}}$
		$\bar{\mathbf{G}}$	$\sum_{k=1}^{K}\left(M_{\mathrm{D}k}N_{\mathrm{R}}N_{\mathrm{D}k} + M_{\mathrm{D}k}N_{\mathrm{R}}M_{\mathrm{R}}\right)$
$N2_{\mathrm{iter}}$ iterations	Step 19	$\bar{\mathbf{R}}$	$M_{\mathrm{D}}M_{\mathrm{R}}^2 + M_{\mathrm{D}}M_{\mathrm{R}}N_{\mathrm{R}} + M_{\mathrm{D}}^2N_{\mathrm{R}} + M_{\mathrm{D}}^2N_{\mathrm{D}}$
		$\mathbf{W}_{\mathrm{D}}^{\mathrm{MMSE}}$	$M_{\mathrm{R}}M_{\mathrm{S}}KL_{\mathrm{S}} + M_{\mathrm{D}}M_{\mathrm{R}}KL_{\mathrm{S}} + 2M_{\mathrm{D}}^2KL_{\mathrm{S}} + M_{\mathrm{D}}^3$
	Step 20	$\mathbf{A}$	$M_{\mathrm{D}}^2KL_{\mathrm{S}} + M_{\mathrm{D}}(KL_{\mathrm{S}})^2$
	Step 21	$\mathbf{W}_{\mathrm{R}}$	$M_{\mathrm{D}}M_{\mathrm{R}}KL_{\mathrm{S}} + M_{\mathrm{R}}(KL_{\mathrm{S}})^2 + 2M_{\mathrm{R}}^2N_{\mathrm{R}} + 2M_{\mathrm{R}}^3 + 5M_{\mathrm{R}}^2KL_{\mathrm{S}}$
Algorithm 2			
	Step 2	$\tilde{\mathbf{H}}_k$	$M_{\mathrm{D}k}M_{\mathrm{R}}^2 + M_{\mathrm{D}k}M_{\mathrm{R}}M_{\mathrm{S}}$
K iterations	Step 4	SVD $\left(\tilde{\mathbf{H}}_k\right)$	$\mathcal{O}\left(\min\left(M_{\mathrm{D}k}, M_{\mathrm{S}}\right)M_{\mathrm{D}k}M_{\mathrm{S}}\right)$
	Step 5	SVD $\left(\mathbf{W}_{\mathrm{D}k}^{\mathrm{H}}\tilde{\mathbf{H}}_k\tilde{\mathbf{V}}_{k-1}^{(0)}\right)$	$M_{\mathrm{S}}L_{\mathrm{S}}\left(M_{\mathrm{D}k} + M_{\mathrm{S}} - L_{\mathrm{S}}\right) + \mathcal{O}\left(\min\left(L_{\mathrm{S}}, \left(M_{\mathrm{S}} - L_{\mathrm{S}}\right)\right)L_{\mathrm{S}}\left(M_{\mathrm{S}} - L_{\mathrm{S}}\right)\right)$
	Step 6	$\mathbf{W}_{\mathrm{S}}^{k}$	$M_{\mathrm{S}}L_{\mathrm{S}}\left(M_{\mathrm{S}} - L_{\mathrm{S}}\right)$
Algorithm 3			
	Step 2	$\tilde{\mathbf{H}}_k$	$M_{\mathrm{D}k}M_{\mathrm{R}}^2 + M_{\mathrm{D}k}M_{\mathrm{R}}M_{\mathrm{S}}$
K iterations	Step 4	SVD $\left(\tilde{\mathbf{H}}_k\right)$	$\mathcal{O}\left(\min\left(M_{\mathrm{D}k}, M_{\mathrm{S}}\right)M_{\mathrm{D}k}M_{\mathrm{S}}\right)$
	Step 5	$\tilde{\mathbf{H}}_{\mathrm{comp}}^{k}$	$M_{\mathrm{D}k}M_{\mathrm{S}}L_{\mathrm{S}}$
	Step 6	SVD $\left(\tilde{\mathbf{H}}_{\mathrm{comp}}^{k}\right)$	$\mathcal{O}\left(\min\left(L_{\mathrm{S}}, M_{\mathrm{S}}\right)L_{\mathrm{S}}M_{\mathrm{S}}\right)$

and $\mathbf{W}_{\mathrm{R}}$ are optimized in each iteration. For the complexity to obtain $\mathbf{F}_{\mathrm{R}2}$, the first part originates from the computation of $\mathbf{S}_{u-1}$, which requires $M_{\mathrm{D}}(N_{\mathrm{R}}^{m_{\mathrm{R}}})^2$ times of complex multiplications [29]. The second part comes from the SVD of $\mathbf{S}_{u-1}$, and the complexity is $\mathcal{O}\left((N_{\mathrm{R}}^{m_{\mathrm{R}}})^3\right)$. The third part is to get $\hat{\mathbf{f}}_{\mathrm{R}2}^{u}$. Since only the phase of $\left(\hat{\mathbf{f}}_{\mathrm{R}2}^{u}\right)_{\mathrm{opt}}$ is computed, there is no extra computation. The fourth part comes from the computation of $\mathbf{T}_{u-1}$, which has a complexity of $2m_{\mathrm{R}}N_{\mathrm{R}}M_{\mathrm{D}} + m_{\mathrm{R}}M_{\mathrm{D}}^2$. The overall complexity to obtain $\mathbf{F}_{\mathrm{R}2}$ for M_{R} cycles in each iteration is $\sum_{m_{\mathrm{R}}=1}^{M_{\mathrm{R}}}\left\{M_{\mathrm{D}}(N_{\mathrm{R}}^{m_{\mathrm{R}}})^2 + \mathcal{O}\left((N_{\mathrm{R}}^{m_{\mathrm{R}}})^3\right) + 2\left(m_{\mathrm{R}}-1\right)N_{\mathrm{R}}M_{\mathrm{D}} + \left(m_{\mathrm{R}}-1\right)M_{\mathrm{D}}^2\right\}$. Similarly, the overall complexity to obtain $\mathbf{F}_{\mathrm{D}}$ for M_{D} cycles, $\mathbf{F}_{\mathrm{R}1}$ for M_{S} cycles, and $\mathbf{F}_{\mathrm{S}}$ for M_{R} cycles, in each iteration are $\sum_{m_{\mathrm{D}}=1}^{M_{\mathrm{D}}}\left\{2\left(m_{\mathrm{D}}-1\right)N_{\mathrm{D}}M_{\mathrm{R}} + \left(m_{\mathrm{D}}-1\right)M_{\mathrm{R}}^2\right\} + M_{\mathrm{D}}\left\{M_{\mathrm{R}}\left(N_{\mathrm{D}}^{M}\right)^2 + \mathcal{O}\left(\left(N_{\mathrm{D}}^{M}\right)^3\right)\right\}$, $\sum_{m_{\mathrm{S}}=1}^{M_{\mathrm{S}}}\left\{M_{\mathrm{R}}(N_{\mathrm{S}}^{m_{\mathrm{S}}})^2 + \mathcal{O}\left((N_{\mathrm{S}}^{m_{\mathrm{S}}})^3\right) + 2\left(m_{\mathrm{S}}-1\right)N_{\mathrm{S}}M_{\mathrm{R}} + \left(m_{\mathrm{S}}-1\right)M_{\mathrm{R}}^2\right\}$ and $\sum_{m_{\mathrm{S}}=1}^{M_{\mathrm{R}}}\left\{M_{\mathrm{S}}(N_{\mathrm{R}}^{m_{\mathrm{R}}})^2 + \mathcal{O}\left((N_{\mathrm{R}}^{m_{\mathrm{R}}})^3\right) + 2(m_{\mathrm{R}}-1)N_{\mathrm{R}}M_{\mathrm{S}} + (m_{\mathrm{R}}-1)M_{\mathrm{S}}^2\right\}$, respectively. The computational complexity to obtain $\mathbf{W}_{\mathrm{R}}$ mainly comes from four parts, i.e., steps 16, 19, 20, and 21, and is demonstrated in Table II in detail. Note that the overall computational complexity of **Algorithm 1** also depends on the number of iterations, which will be studied in Section V (see Fig. 11). In addition, the computational complexity of **Algorithm 2** and **Algorithm 3** is also summarized in Table II.

V. NUMERICAL SIMULATIONS

In this section, we evaluate the performance of the proposed hybrid beamforming design methods for mmWave massive MU-MIMO AF relay systems. It is assumed that $N_{\mathrm{S}}^{M} = N_{\mathrm{S}}/M_{\mathrm{S}}$, $N_{\mathrm{R}}^{M} = N_{\mathrm{R}}/M_{\mathrm{R}}$, and $N_{\mathrm{D}}^{M} = N_{\mathrm{D}k}^{m} = N_{\mathrm{D}k}/M_{\mathrm{D}k}$ for all $m = 1, \ldots, M_{\mathrm{D}k}$ and $k = 1, \ldots, K$. The number of antennas at the BS, relay and users are set as $N_{\mathrm{S}} = 128$, $N_{\mathrm{R}} = 48$, and $N_{\mathrm{D}k} = 8$, respectively. It is considered that $K = 8$ users are served simultaneously, and $L_{\mathrm{S}} = 2$ data streams are supported by each user. To reduce the system power consumption, the least number of RF chains is equipped for all nodes [39], i.e., $M_{\mathrm{S}} = M_{\mathrm{R}} = KL_{\mathrm{S}}$ and $M_{\mathrm{D}k} = L_{\mathrm{S}}$. In the simulations, the mmWave channel between the BS and relay contains $N_{\mathrm{c}}^{\mathrm{R}} = 8$ clusters with $N_{\mathrm{p}}^{\mathrm{R}} = 10$ rays per cluster, and that between the relay and user k contains $N_{\mathrm{c}}^{\mathrm{D}k} = 7$ clusters with $N_{\mathrm{p}}^{\mathrm{D}k} = 5$ rays per cluster. The AoAs/AoDs of all channels are assumed to follow the uniform distribution within $[0, 2\pi]$. All simulation results are obtained by averaging over 1,000 random channel realizations based on the MATLAB platform. (Note: Unless otherwise specified, the above parameters are default settings.)

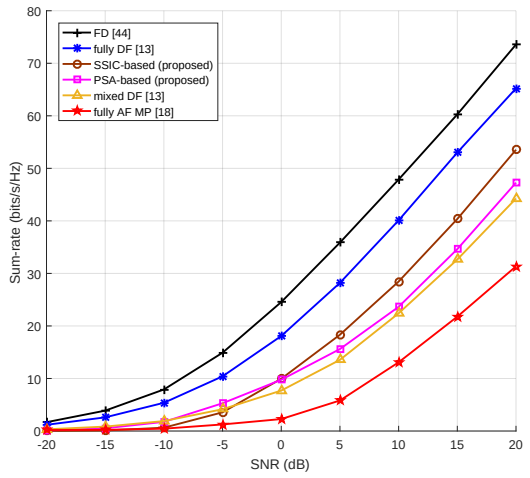


Fig. 4. Sum-rate comparison for different beamforming schemes versus SNR.

For better evaluation, the proposed SSIC-based and PSA-based hybrid beamforming/combining algorithms are compared with four existing beamforming benchmarks: 1) the FD beamforming scheme [44], which provides the performance upper bound; 2) the fully-connected hybrid beamforming based on DF (fully DF) [13]; 3) the fully-connected hybrid beamforming based on AF by using matching pursuit (fully AF MP) [18]; 4) the hybrid beamforming with mixed structure based on DF (mixed DF) [13].

A. Comparison for Different Beamforming Schemes

Fig. 4 compares the sum-rate performance of different beamforming schemes versus signal-to-noise ratio (SNR), where $N_{\mathrm{R}} = 64$ and $N_{\mathrm{D}k} = 16$. It can be seen from Fig. 4 that the sum-rate performance of the two proposed methods outperforms the mixed DF scheme. Since the MP method does not guarantee the optimality [17], [18], its sum-rate is the lowest. In addition, since the evenly distributed scenario can further reduce IUI, the sum-rate performance of the proposed SSIC-based method is superior to that of the proposed PSA-based method. At low SNR, the interference between different users can not be neglected compared to the signal energy, even for the evenly distributed scenario. Therefore, the performance of the proposed SSIC-based method is slightly lower than that of the mixed DF method in the low SNR regime.

B. Performance for Number of Antennas in the BS, Relay and User Ends

Fig. 5a compares the sum-rate performance of different beamforming schemes versus the number of BS antennas, where SNR = 5 dB. As shown in Fig. 5a, when the number of BS antennas increases, the sum-rate performance of different design schemes improves correspondingly. Since the DF-based relay MU-MIMO system can be considered as a series of two (or even more) single-hop MIMO systems, increasing the number of BS antennas only impacts its sum-rate in the first time slot. Hence, the performance improvement of the DF relay system gradually becomes saturated. However, the sum-rate of the overall AF relay system can be affected by the number of BS antennas. Therefore, the sum-rate performance of the proposed SSIC-based and PSA-based schemes improves slowly and always outperforms the mixed DF and fully-connected AF MP methods.

Fig. 5b compares the sum-rate performance of different beamforming schemes versus the number of relay antennas ranging from 32 to 96, where $K = 4$ and SNR = 15 dB. As can be seen from Fig. 5b, since the limitation of sub-connected structures, the performance of the two proposed schemes is improved slowly with the increasing number of relay antennas. However, the two proposed schemes always outperform the mixed DF and fully-connected AF MP methods.

Fig. 5c compares the sum-rate performance of different beamforming schemes versus the number of antennas per user ranging from 4 to 32, where SNR = 5 dB. As shown in Fig. 5c, when the number of antennas per user increases, the sum-rate performance of different design schemes improves correspondingly. For the mixed DF scheme, users adopt fully-connected structures. Therefore, with the increasing number of antennas per user, the performance of the fully-connected and mixed schemes improves faster than that of the two proposed methods.

C. Performance for Number of Relay RF Chains and Users

Fig. 6 compares the sum-rate performance of different beamforming schemes versus the number of relay RF chains ranging from 4 to 16, where $K = 2$ and SNR = 15 dB. It can be observed from Fig. 6 that when the number of RF chains is small, both of the two proposed schemes achieve superior performance compared with the mixed DF method. In addition, due to the large dimensional digital beamforming, the achievable sum-rate of the two proposed methods improves a lot as the number of relay RF chains increases from 4 to 6, and saturates gradually.

Fig. 7 compares the sum-rate performance of different beamforming schemes versus the number of users ranging from 2 to 16, where $N_{\mathrm{D}k} = 4$ and SNR = 10 dB. With the increasing number of users, the row subspace of channel matrices overlaps significantly, and the BB-BD technology can result in poor performance. Therefore, the performance of fully-connected and mixed DF schemes begins to decline when $K = 10$. The performance of the two proposed methods improves gradually and outperforms the fully-connected and mixed DF schemes when the number of users is large enough. In addition, since the interference between different users is rare when K is small, e.g., K is less than or equal to 5, the performance of the SSIC-based method is slightly lower than that of the PSA-based method.

D. Performance for Energy Efficiency

For hybrid beamforming of mmWave relay systems with the sub-connected structure, the energy consumption is a key issue and deserves our consideration. According to [45], the power simulation parameters for all components are set as follows: $P_{\mathrm{BB}} = 10$ W for the digital BB processor , $P_{\mathrm{RF}} = 100$ mW for each RF chain, $P_{\mathrm{PS}} = 10$ mW for the PS, $P_{\mathrm{LNA}} = 100$ mW for the low noise amplifier, $P_{\mathrm{PA}} = 100$ mW for the power

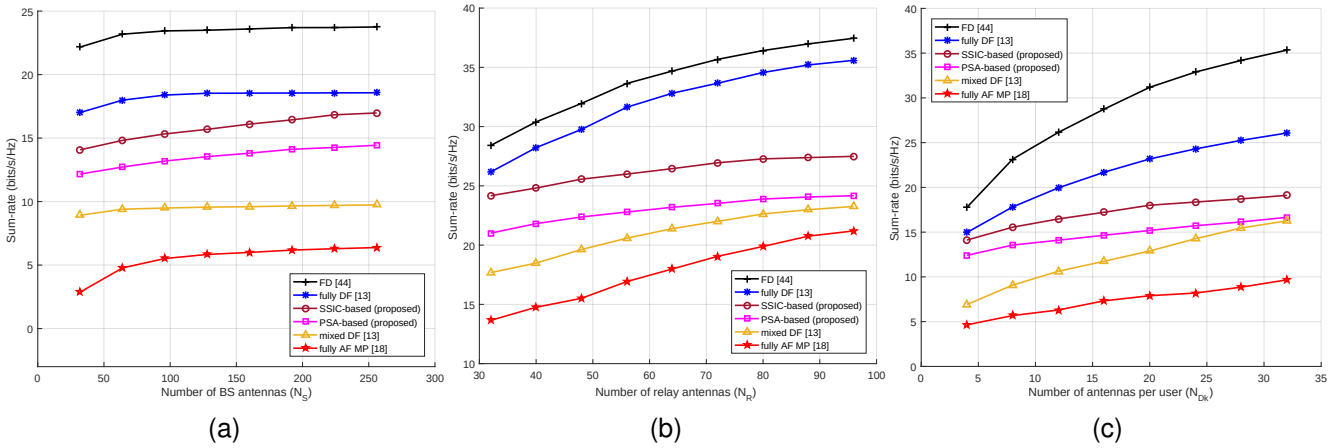


Fig. 5. Sum-rate comparison versus the number of antennas in the BS, relay and user ends.

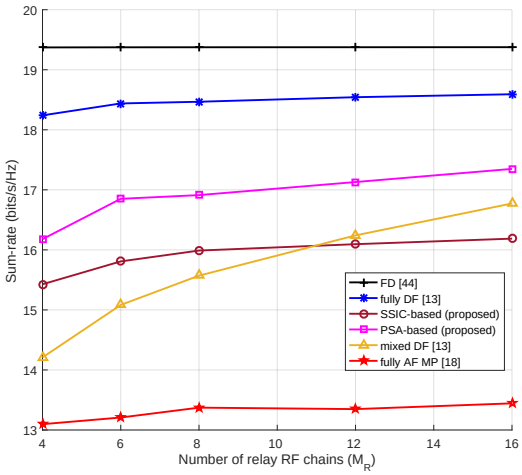


Fig. 6. Sum-rate comparison versus the number of relay RF chains.

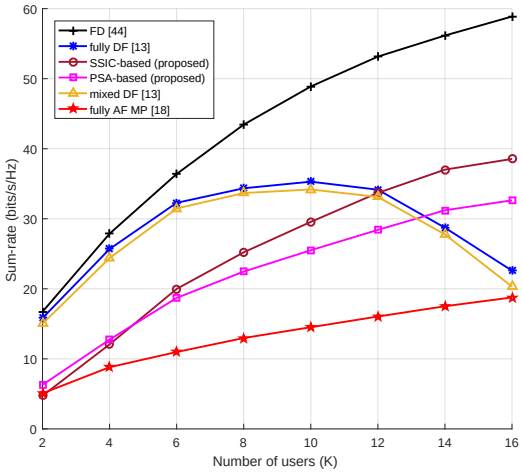


Fig. 7. Sum-rate comparison versus the number of users.

amplifier, $P_{\mathrm{DAC}} = 110$ mW for the digital-to-analog converter, $P_{\mathrm{ADC}} = 200$ mW for the analog-to-digital converter, and $P_{\mathrm{LO}} = 10$ mW for the local oscillator. Systems with different structures have different energy consumption. The energy

TABLE III
ENERGY CONSUMPTION FOR DIFFERENT STRUCTURES

Structure	Energy consumption
P_{FD}	$(K+3)\,P_{\mathrm{BB}} + (2N_{\mathrm{R}} + N_{\mathrm{S}} + N_{\mathrm{D}})\,(P_{\mathrm{RF}} + P_{\mathrm{LO}})$ $+ (N_{\mathrm{R}} + N_{\mathrm{D}})\,(P_{\mathrm{LNA}} + P_{\mathrm{ADC}}) + (N_{\mathrm{R}} + N_{\mathrm{S}})\,(P_{\mathrm{PA}} + P_{\mathrm{DAC}})$
P_{DFF}	$(K+3)\,P_{\mathrm{BB}} + (N_{\mathrm{R}} + N_{\mathrm{D}})\,P_{\mathrm{LNA}} + (N_{\mathrm{R}} + N_{\mathrm{S}})\,P_{\mathrm{PA}} +$ $2M_{\mathrm{R}}\,(P_{\mathrm{ADC}} + P_{\mathrm{DAC}} + 2\,(P_{\mathrm{RF}} + P_{\mathrm{LO}})) +$ $M_{\mathrm{R}}\,(2N_{\mathrm{R}} + N_{\mathrm{S}} + N_{\mathrm{D}k})\,P_{\mathrm{PS}}$
P_{DFM}	$(K+3)\,P_{\mathrm{BB}} + (N_{\mathrm{R}} + N_{\mathrm{D}})\,P_{\mathrm{LNA}} + (N_{\mathrm{R}} + N_{\mathrm{S}})\,P_{\mathrm{PA}} +$ $M_{\mathrm{R}}\,(2P_{\mathrm{ADC}} + P_{\mathrm{DAC}} + 4P_{\mathrm{RF}} + P_{\mathrm{LO}}) +$ $(M_R\,(N_{\mathrm{R}} + N_{\mathrm{S}}) + N_{\mathrm{R}} + N_{\mathrm{D}})\,P_{\mathrm{PS}}$
P_{AFF}	$(K+2)\,P_{\mathrm{BB}} + (N_{\mathrm{R}} + N_{\mathrm{D}})\,P_{\mathrm{LNA}} + (N_{\mathrm{R}} + N_{\mathrm{S}})\,P_{\mathrm{PA}} +$ $2M_{\mathrm{R}}\,(P_{\mathrm{ADC}} + P_{\mathrm{DAC}}) + 4M_{\mathrm{R}}\,(P_{\mathrm{RF}} + P_{\mathrm{LO}}) +$ $M_{\mathrm{R}}\,(2N_{\mathrm{R}} + N_{\mathrm{S}} + N_{\mathrm{D}k})\,P_{\mathrm{PS}}$
P_{AFS}	$(K+2)\,P_{\mathrm{BB}} + (N_{\mathrm{R}} + N_{\mathrm{D}})\,P_{\mathrm{LNA}} + (N_{\mathrm{R}} + N_{\mathrm{S}})\,P_{\mathrm{PA}} +$ $2M_{\mathrm{R}}\,(P_{\mathrm{ADC}} + P_{\mathrm{DAC}}) + 3M_{\mathrm{R}}\,(P_{\mathrm{RF}} + P_{\mathrm{LO}}) +$ $(2N_{\mathrm{R}} + N_{\mathrm{S}} + N_{\mathrm{D}})\,P_{\mathrm{PS}}$

consumption during an unit time interval of the FD (P_{FD}), DF-based with fully-connected (P_{DFF}) and mixed (P_{DFM}), and AF-based fully-connected (P_{AFF}) and sub-connected (P_{AFS}) structures is expressed in Table III. The energy efficiency η is defined as

$$\eta = \frac{R}{P_*}(\mathrm{bps/Hz/J}), \tag{54}$$

where P_* denotes the energy consumption of different systems.

Fig. 8 compares the energy efficiency of different beamforming schemes versus the number of relay RF chains ranging from 4 to 16, where $K = 2$ and SNR = 15 dB. As shown in Fig. 8, when the number of RF chains increases, the energy efficiency of the DF-based and AF-based fully-connected structures declines tremendously due to high energy consumption. The energy consumption of FD system is independent of the number of relay RF chains, and hence its energy efficiency always remains stable. For the DF-based mixed and AF-based sub-connected structures, the energy efficiency remains almost stable over different number of RF chains. Specially, the two proposed schemes have better energy efficiency performance among these methods. In addition, it can be seen from Fig. 6 that when the number of relay RF chains is large enough, the

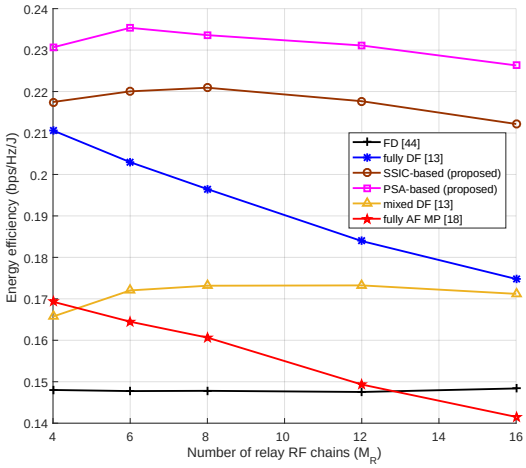


Fig. 8. Energy efficiency comparison versus the number of relay RF chains.

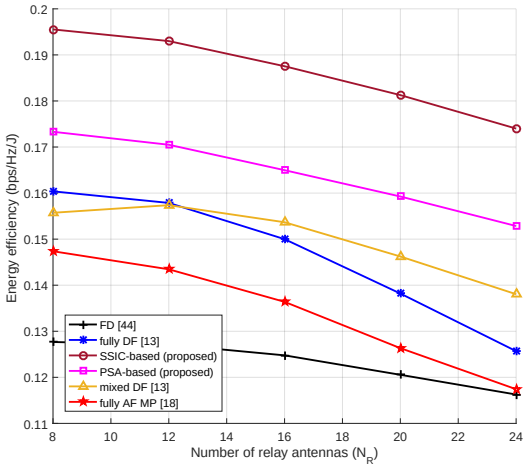


Fig. 9. Energy efficiency comparison versus the number of relay antennas.

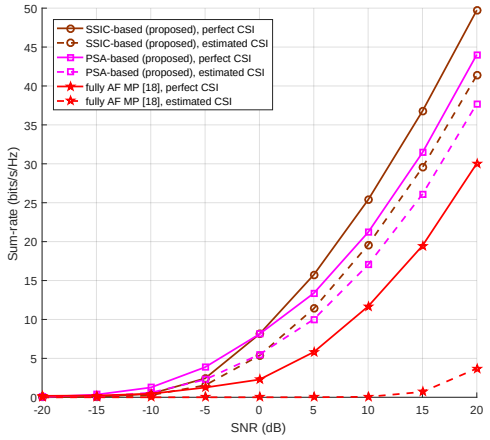


Fig. 10. Sum-rate comparison versus SNR with estimated CSI.

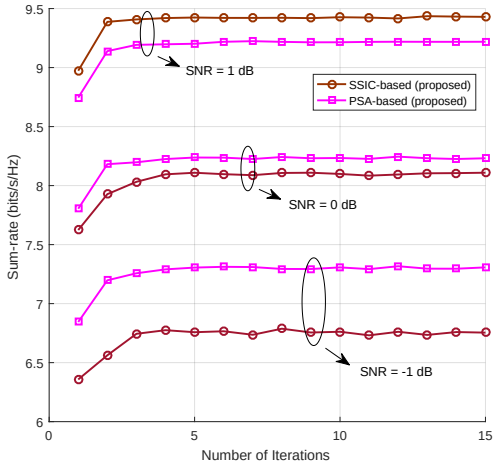


Fig. 11. Convergence speed of the two proposed methods at the analog beamforming stage.

fully-connected AF MP method shows the worst performance due to its high energy consumption.

Fig. 9 further shows the energy efficiency performance versus the number of relay antennas, where $K = 4$, $N_{\mathrm{R}} = 4M_{\mathrm{R}}$, $L_{\mathrm{S}} = 1$ and SNR = 15 dB. As can be seen from Fig. 9, the two proposed methods are superior to other comparative algorithms. According to Table III, P_{DFM} owns less energy consumption than P_{DFF}. Hence the energy efficiency of the mixed DF scheme degrades more slowly than that of the fully-connected DF scheme.

E. Performance for Robustness and Iteration

In practice, CSI is unknown and needs to be estimated. For the two-hop relay systems, the CSI from the BS to relay and from the relay to users can be respectively estimated by adopting point-to-point methods [46]–[48] at the relay and users. Then the obtained CSI can be shared with the BS and relay via feedback from the relay to BS and from the users to relay [26], respectively. In this example, we study the impact of the CSI estimation on the performance of the two proposed hybrid beamforming methods. The CSI is estimated by the pilot-assisted orthogonal matching pursuit [46] method when SNR $= 10$dB. The pilot length is set as 128 to estimate the CSI from the BS to relay, and is set as 48 to estimate the CSI from the relay to users. Fig. 10 demonstrates the sum-rate curves of the two proposed methods and MP algorithm with perfect and estimated CSI. As shown in Fig. 10, the performance of the two proposed schemes is slightly affected by the CSI accuracy. For instance, at the sum-rate of 25bits/s/Hz, the SNR gaps between perfect CSI and estimated CSI are about 2.7 dB and 2.3 dB for the proposed SSIC-based and PSA-based methods, respectively. In contrast, the sum-rate of the MP algorithm degrades dramatically in the case of estimated CSI compared with the perfect CSI scenario. Therefore, both of the two proposed methods are shown to be robust to imperfect CSI.

The convergence speed of the two proposed methods at the analog beamforming stage is depicted in Fig. 11. The sum-rate performance of the proposed SSIC-based and PSA-based methods will reach saturation after 5 and 6 iterations, respectively. Both the two proposed methods show fast convergence speed.

VI. Conclusion

In this paper, we have investigated the hybrid beamforming design with sub-connected structure for mmWave massive

MU-MIMO AF relay systems. Assuming perfect CSI, the analog beamforming of the BS, relay and users has been iteratively designed by employing the idea of SIC. Then, the highly complicated non-convex mutual information maximization problem has been transformed into an easily tractable WMMSE one, and hence the relay BB combining matrix is obtained. Further, we have considered two different user distribution scenarios, and developed a SSIC-based and a PSA-based methods to design digital beamformers at the BS and users. Simulations have demonstrated that the two proposed hybrid beamforming schemes can achieve good performance. Although the two proposed schemes are derived based on perfect CSI, they are also shown to be robust to imperfect CSI. Perspective of our work contains an extension to wideband reconfigurable intelligent surface aided mmWave massive MU-MIMO systems, and the hybrid structure is also considered at the BS.

REFERENCES

[1] X. Yang, M. Matthaiou, J. Yang, C.-K. Wen, F. Gao, and S. Jin, "Hardware-constrained millimeter-wave systems for 5G: Challenges, opportunities, and solutions," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 44–50, 2019.

[2] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 886–901, 2019.

[3] H.-M. Wang, K.-W. Huang, and T. A. Tsiftsis, "Base station cooperation in millimeter wave cellular networks: Performance enhancement of cell-edge users," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5124–5139, 2018.

[4] J.-H. Lee, K.-H. Park, Y.-C. Ko, and M.-S. Alouini, "Throughput maximization of mixed FSO/RF UAV-aided mobile relaying with a buffer," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 683–694, 2021.

[5] J. Du, Y. Zhang, Y. Chen, X. Li, Y. Cheng, and M. Rajesh, "Hybrid beamforming NOMA for mmWave half-duplex UAV relay-assisted B5G/6G IoT networks," *Comput. Commun.*, vol. 180, pp. 232–242, 2021.

[6] C. Guo, L. Tian, Z. H. Jiang, and W. Hong, "A Self-Calibration Method for 5G Full-Digital TDD Beamforming Systems Using an Embedded Transmission Line," *IEEE Trans. Antennas Propag.*, vol. 69, no. 5, pp. 2648–2659, 2021.

[7] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, 2004.

[8] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.

[9] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, "A block sparsity based estimator for mmWave massive MIMO channels with beam squint," *IEEE Trans. Signal Process.*, vol. 68, pp. 49–64, 2020.

[10] Z. Wei, X. Zhu, S. Sun, Y. Huang, L. Dong, and Y. Jiang, "Full-duplex versus half-duplex amplify-and-forward relaying: Which is more energy efficient in 60-GHz dual-hop indoor wireless systems?" *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2936–2947, 2015.

[11] S. Nishikawa and G. Abreu, "Manifold optimization-based hybrid TX/RX precoding for FD relay mmWave systems," in *2018 15th Workshop on Positioning, Navigation and Communications (WPNC)*. IEEE, 2018, pp. 1–6.

[12] Y. Zhang, M. Xiao, S. Han, M. Skoglund, and W. Meng, "On precoding and energy efficiency of full-duplex millimeter-wave relays," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1943–1956, 2019.

[13] 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. 157 212–157 225, 2019.

[14] Z. 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.

[15] Y.-Y. Lee, C.-H. Wang, and 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, 2014.

[16] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, 2018.

[17] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014, pp. 5838–5842.

[18] X. Xue, T. E. Bogale, X. Wang, Y. Wang, and B. Le Long, "Hybrid analog-digital beamforming for multiuser MIMO millimeter wave relay systems," in *2015 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2015, pp. 1–7.

[19] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint source and relay precoding in multiantenna millimeter-wave systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4924–4937, 2016.

[20] J.-S. Sheu, "Hybrid digital and analogue beamforming design for millimeter wave relaying systems," *J. Commun. Networks*, vol. 19, no. 5, pp. 461–469, 2017.

[21] C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Hybrid analog-digital transceiver designs for mmwave amplify-and-forward relaying systems," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2018, pp. 1–6.

[22] Z. Luo and H. Liu, "Robust hybrid transceiver designs for millimeter wave AF cooperative systems," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–5.

[23] L. Jiang and H. Jafarkhani, "MmWave amplify-and-forward MIMO relay networks with hybrid precoding/combining design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1333–1346, 2020.

[24] S. Wang, X. Xu, K. Huang, X. Ji, Y. Chen, and L. Jin, "Artificial noise aided hybrid analog-digital beamforming for secure transmission in MIMO millimeter wave relay systems," *IEEE Access*, vol. 7, pp. 28 597–28 606, 2019.

[25] 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, 2019.

[26] 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, 2018.

[27] W. Xu, Y. Wang, and X. Xue, "ADMM for hybrid precoding of relay in millimeter-wave massive MIMO system," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–5.

[28] N. Song, T. Yang, and H. Sun, "Efficient Hybrid Beamforming for Relay Assisted Millimeter-Wave Multi-User Massive MIMO," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[29] 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, 2016.

[30] J. Du, W. Xu, H. Shen, X. Dong, and C. Zhao, "Hybrid precoding architecture for massive multiuser MIMO with dissipation: Sub-connected or fully connected structures?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5465–5479, 2018.

[31] W. Xu, J. Liu, S. Jin, and X. Dong, "Spectral and energy efficiency of multi-pair massive MIMO relay network with hybrid processing," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3794–3809, 2017.

[32] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[33] 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, 2019.

[34] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, 2015.

[35] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, 2014.

[36] C. Cano, G. H. Sim, A. Asadi, and X. Vilajosana, "A Channel Measurement Campaign for mmWave Communication in Industrial Settings," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 299–315, 2021.
[37] 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, 2018.
[38] 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. 13 482–13 496, 2020.
[39] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
[40] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 141–154, 2005.
[41] E. Sharma, A. S. Chauhan, and R. Budhiraja, "Transceiver design for massive MIMO two-way half-duplex AF hybrid relay with MIMO users," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8759–8774, 2019.
[42] 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, 2019.
[43] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 953–961, 2008.
[44] M. R. Khandaker and Y. Rong, "Joint transceiver optimization for multiuser MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5977–5986, 2012.
[45] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmWave," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 756–770, 2019.
[46] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
[47] J. Du, M. Han, Y. Chen, L. Jin, and F. Gao, "Tensor-based joint channel estimation and symbol detection for time-varying mmWave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 6251–6266, 2021.
[48] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.