Joint User Association and Hybrid Beamforming Designs for Cell-Free mmWave MIMO Communications

Zihuan Wang, *Graduate Student Member, IEEE*, Ming Li, *Senior Member, IEEE*, Rang Liu, *Graduate Student Member, IEEE*, and Qian Liu, *Member, IEEE*

***Abstract*— Cell-free millimeter-wave (mmWave) multiple-input multiple-output (MIMO) communications have been proposed as promising enablers for the next generation wireless networks. In this paper, we study the user association and hybrid beamforming in cell-free mmWave systems without full channel state information (CSI) acquisition. We consider a cloud radio access network (C-RAN), where multiple remote radio heads (RRHs) are distributed to communicate with users via analog beamforming, and connected to a centralized baseband unit (BBU) through fronthaul links which executes digital beamforming. We aim to jointly design user association, hybrid beamforming, and fronthaul compression with the aid of uplink training. A train-and-design framework is developed to achieve this goal. In particular, we first propose a two-stage uplink training approach to assist RRH-level design, during which the analog beamforming and user association are obtained. After that, digital beamforming and fronthaul compression are optimized at BBU based on the training results. Two performance metrics are considered in this paper, i.e. weighted sum-rate maximization and max-min fairness. Simulation results demonstrate the effectiveness of the proposed train-and-design framework for both sum-rate maximization and max-min fairness performance metrics. It is shown that the proposed algorithms can achieve comparable performance to the full-digital beamformer.**

***Index Terms*— Millimeter-wave (mmWave) MIMO communications, cell-free network, hybrid beamforming, cloud radio access network (C-RAN).**

Manuscript received 27 October 2021; revised 29 March 2022 and 19 July 2022; accepted 22 September 2022. Date of publication 4 October 2022; date of current version 18 November 2022. This work is supported in part by the National Natural Science Foundation of China (Grant No. 61971088, 62071083, U1808206, and U1908214), the Natural Science Foundation of Liaoning Province (Grant No. 2020-MS-108), in part by the Fundamental Research Funds for the Central Universities (Grant No. DUT21GJ208), and in part by Dalian Science and Technology Innovation Project (Grant No. 2020JJ25CY001). An earlier version of this paper was presented in part at the IEEE Vehicular Technology Conference (VTC), Victoria, BC, Canada, October 2020 [DOI: 10.1109/VTC2020-Fall49728.2020.9348506]. The associate editor coordinating the review of this article and approving it for publication was A. Maaref. *(Corresponding authors: Ming Li; Qian Liu.)*

Zihuan Wang is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: zihuanwang@ece.ubc.ca).

Ming Li and Rang Liu are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China (e-mail: mli@dlut.edu.cn; liurang@mail.dlut.edu.cn).

Qian Liu is with the School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China (e-mail: qianliu@dlut.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3211966.

Digital Object Identifier 10.1109/TCOMM.2022.3211966

I. Introduction

THE growing demand on high data-rate wireless communications driven by innovative applications and devices has always been the key promoter for wireless network evolution. Current sub-6 GHz frequency bands are highly congested and the insufficient available bandwidth becomes the primary impediment to meet the requirement of wireless capacity. Therefore, next generation wireless network will inevitably rely on high-frequency millimeter-wave (mmWave) bands to provide abundant spectrum resources [2], [3]. However, for wireless communications operating at mmWave bands, the severe path-loss and low penetration are the major challenges, which hinder the evolution of mmWave communications from theoretical concept to real-world deployment.

To deal with the severe path-loss problem from infrastructure-level, multiple-input multiple-output (MIMO) is employed at base station (BS), which enables energy-focusing beamforming to combat the large attenuation of mmWave channels [4], [5], [6]. For practical implementation of mmWave MIMO communications, energy-efficient analog/digital hybrid beamforming architecture, which significantly reduces the number of radio frequency (RF) chains, is used to provide sufficient channel gain while maintaining affordable hardware cost and energy consumption [7], [8], [9], [10]. Another enabling approach to facilitate mmWave communications is network densification, which leads to a smaller cell size, thus reducing the access distance of users and enhancing the coverage [11], [12]. However, the smaller mmWave cell size may cause higher level inter-cell interference, thus limiting the overall systems performance. In addition, users at cell-edge may still suffer from poor communication qualities and interruptions as mmWave beams are easy to get blocked.

Cell-free network, which breaks the restrictions of traditional cell boundaries, has been proposed as a promising solution to further improve the system throughput [13], [14], [15], [16], [17], [18]. Through multiple BSs serving the same users, the cell-free network not only increases spatial degrees of freedom (DoF), but also enables coordinated transmission by coherent transmission and reception, thus providing high-quality mmWave communications. In order to facilitate the realization of cell-free mmWave MIMO communications, cloud radio access network (C-RAN) is a platform to manage the cooperation among BSs. In the C-RAN system, the base-

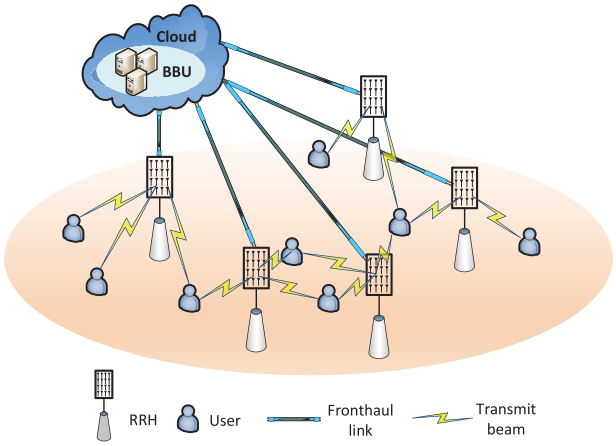


Fig. 1. Cell-free mmWave MIMO systems based on C-RAN.

band signal processing of BSs are migrated to a centralized baseband unit (BBU) in the cloud, while RF processing is carried out at remote radio heads (RRHs) [19], [20]. Therefore, the digital beamforming is implemented at the centralized BBU, and the analog beamforming is distributedly realized at each RRH. The considered C-RAN based cell-free mmWave MIMO system is illustrated in Fig. 1.

A. Related Work

MmWave MIMO communications and hybrid beamforming architecture for point-to-point or cellular systems have been extensively investigated [9], [10], [21], [22], [23], [24], [25]. Given the sparse mmWave channels, codebook-based hybrid beamformer designs are commonly utilized [9], [10], [21], where the columns of the analog beamformer are selected from specific candidate vectors, such as array response vectors of the channel. Then, the digital beamforming is obtained to further enhance the system performance. Note that the design of hybrid beamforming is based on the knowledge of channel state information (CSI), channel estimation via uplink training is usually implemented before the hybrid beamforming design. In [10], the authors proposed an adaptive compressed sensing based algorithm to estimate the parameters of mmWave channel. The extensions of hybrid beamforming design to multi-user scenarios have been studied in [22], [23], and [24]. In [22], the analog and digital beamformer are jointly designed to approximate the optimal minimum mean square error (MMSE) beamformer via orthogonal matching pursuit (OMP) method. In [23], the analog beamformer is first selected from codebook to maximize the channel gains of users, followed by zero-forcing digital beamforming. An iterative design of the hybrid beamformer was proposed in [24] by exploiting the duality of the uplink and downlink multi-user MIMO channels. Finally, the authors in [25] studied the coordinated beamforming for mmWave multi-cell systems, in which generalized eigenvalue decomposition is employed for hybrid beamforming design and analytical approximations of performance are provided.

While hybrid beamforming for point-to-point or cellular systems have been extensively studied, the researches on cell-free network are limited. In [26], [27], [28], and [29], the power control and beamforming designs for cell-free MIMO systems using sub-6 GHz bands have been investigated under the assumption of infinite fronthaul link capacity. However, considering the practical mmWave communication systems, the fronthaul link is usually capacity-constrained [30]. Beamformer and fronthaul compression designs in C-RAN systems have been considered in [31], where full-digital beamformer and fronthaul compression are iteratively calculated based on successive convex approximation (SCA) and weighted minimum-mean-square-error (WMMSE) methods. The extension of this work to hybrid beamforming architecture was investigated in [32] and [33], where the analog and digital beamformers are alternatively optimized based on WMMSE. In [34], the authors proposed to first determine the analog beamformer from a codebook, followed by digital beamforming and compression design based on semidefinite relaxation (SDR) and convex approximation. The max-min fairness problem has been studied in [35], [36], [37], and [38] for cell-free MIMO systems. In [35] and [36], the max-min fairness power control and beamforming designs have been investigated. The authors analytically showed that the problem is a quasi-concave problem and can be solved via second-order cone programming. In [37], the joint beamforming and power control designs for max-min fairness of uplink C-RAN has been investigated. The authors proposed to decompose the design of beamformer and power allocation, and iteratively update the two variables based on uplink-downlink duality. An SDR based approach has been proposed in [38] to jointly select RRH and obtain beamforming in C-RAN systems. These works, however, considered conventional sub-6 GHz bands and ignored the design of fronthaul compression.

Moreover, the aforementioned works assume that either perfect knowledge of CSI or its second-order statistics is available at the cloud. Due to the employment of hybrid beamforming structure, the full CSI estimation for cell-free mmWave systems is challenging. In particular, the hybrid beamforming architecture with limited number of RF chains cannot process and differentiate the signals of all users simultaneously. Moreover, based on the C-RAN architecture, the user association and limited fronthaul capacity also need to be taken into consideration in the hybrid beamforming design. Efficient user association and hybrid beamforming designs for C-RAN enabled cell-free mmWave MIMO systems are expected.

B. Contributions

Motivated by the above discussion, in this paper, we investigate the joint user association and hybrid beamforming designs for cell-free mmWave MIMO systems without full CSI acquisition. Our main contributions can be summarized in the following aspects.

- We develop a train-and-design framework, in which a two-stage uplink training approach is proposed to assist the user association and hybrid beamforming designs. The proposed framework avoids full CSI acquisition and leads to lower computational complexity and training overhead.

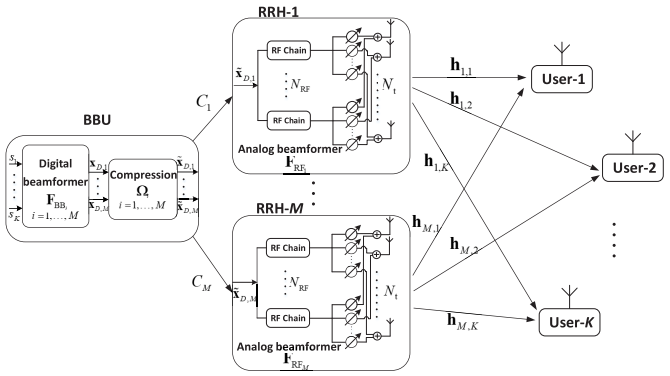


Fig. 2. Hybrid beamforming in cell-free mmWave MIMO systems based on C-RAN architecture.

- We propose a low-complexity hybrid beamforming design. A joint design of RRH analog beamforming and RRH-user association is first developed with the aid of two-stage uplink training. Then, given the training results, the digital beamforming and fronthaul compression are designed based on the obtained effective baseband channel.
- Two performance metrics are considered for the algorithm designs in this paper, i.e. weighted sum-rate maximization and max-min fairness. Simulation results demonstrate the effectiveness of the proposed algorithms, which can achieve comparable performance to the full-digital beamforming scheme.

C. Organization and Notation

The rest of the paper is organized as follows. Sec. II introduces the cell-free mmWave MIMO system model. Sec. III presents the train-and-design framework for hybrid beamforming design to maximize the weighted sum-rate. Then, we further consider the max-min fairness problem for such system in Sec. IV. In Sec. V, we illustrate simulation results of the proposed solution for weighted sum-rate maximization and max-min fairness problems, respectively. Finally, conclusions are drawn in Sec. VI.

Notation: The following notations are used throughout this paper. Boldface lower-case and upper-case letters indicate column vectors and matrices, respectively. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and transpose-conjugate operations, respectively. $\mathbb{E}\{\cdot\}$ represents statistical expectation. $\mathfrak{Re}\{\cdot\}$ extracts the real part of a complex number. $\mathbf{I}_L$ indicates an $L \times L$ identity matrix. $\mathbb{C}$ denotes the set of complex numbers. $|\mathbf{A}|$ denotes the determinant of matrix $\mathbf{A}$. $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. $|a|$ and $\|\mathbf{a}\|$ are the magnitude of a scalar a and norm of a vector $\mathbf{a}$, respectively. Finally, we adopt a MATLAB-like matrix indexing notation: $\mathbf{A}(i,j)$ denotes the element of the i-th row and the j-th column of matrix $\mathbf{A}$; $\mathbf{a}(i)$ denotes the i-th element of vector $\mathbf{a}$.

II. System Model

We consider a C-RAN enabled cell-free mmWave MIMO system as shown in Fig. 2. The centralized BBU transmits signals to K single-antenna users through M cooperative RRHs, using the same time-frequency resource. The i-th RRH is connected to the BBU via a fronthaul link of capacity C_i b/s/Hz [30], [31], [32], $i = 1, \ldots, M$. Each RRH is equipped with $N_{\rm t}$ antennas and $N_{\rm RF}$ RF chains, where $N_{\rm RF} \leq K \leq MN_{\rm RF}$. Then, the received signal at the k-th user can be expressed as

$$y_k = \sum_{i=1}^{M} \mathbf{h}_{i,k}^H \mathbf{x}_i + n_k, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{C}^{N_{\rm t} \times 1}$ denotes the transmitted signal of the i-th RRH subject to the transmit power constraint $\mathbb{E}\{\|\mathbf{x}_i\|^2\} \leq P_i$, P_i indicates the maximum power of the i-th RRH, $\mathbf{h}_{i,k}$ is the channel vector from the i-th RRH to the k-th user, $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ indicates the Gaussian additive noise at the k-th user and σ_k^2 is the noise variance. With hybrid beamforming in the C-RAN architecture, the transmitted signal $\mathbf{x}_i$ is a combination result of digital beamforming which is carried out in the centralized BBU, and analog beamforming which is performed at the i-th RRH.

A. Digital Beamforming and Fronthaul Compression

We let s_k, $k = 1, \ldots, K$, denote the transmitted baseband signal to the k-th user. In the BBU, s_k, $\forall k$, is precoded via digital beamformer before being delivered to each RRH. The precoded signals $\mathbf{x}_{{\rm D},i} \in \mathbb{C}^{N_{\rm RF} \times 1}$ for the i-th RRH is

$$\mathbf{x}_{{\rm D},i} = \sum_{k \in \mathcal{K}_i} \mathbf{f}_{{\rm BB}_{i,k}} s_k, \tag{2}$$

where $\mathbf{f}_{{\rm BB}_{i,k}} \in \mathbb{C}^{N_{\rm RF} \times 1}$ is the digital beamformer utilized for the transmission between the i-th RRH and the k-th user. $\mathcal{K}_i$ indicates the set of users served by the i-th RRH. Let $K_i \triangleq |\mathcal{K}_i|$ denote the number of users served by the i-th RRH, and and satisfies $K_i \leq N_{\rm RF}$. We further assume that the number of users served by each RRH is equal to the number of RF chains,[1] i.e., $K_i = N_{\rm RF}$ [23].

The BBU delivers the baseband signals to each RRH through a limited-capacity fronthaul link. From an information theoretic perspective, the effect of compression and quantization can be modeled as a Gaussian test channel [30], with the uncompressed signals as the input and compressed signals as the output. Then, $\mathbf{x}_{{\rm D},i}$ is quantized and compressed as [30]

$$\widetilde{\mathbf{x}}_{{\rm D},i} = \mathbf{x}_{{\rm D},i} + \mathbf{q}_i, \tag{3}$$

where $\mathbf{q}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_i)$ is the quantization noise vector at the i-th RRH and $\mathbf{\Omega}_i \in \mathbb{C}^{N_{\rm RF} \times N_{\rm RF}}$ is the covariance of quantization noise [30], [31], [32]. The data transmission rate after compression, denoted as $c_i(\mathbf{F}_{{\rm BB}_i}, \mathbf{\Omega}_i)$, should not exceed the fronthaul capacity C_i, $\forall i$, to guarantee a reliable information reception at each RRH:

$$c_i(\mathbf{F}_{{\rm BB}_i}, \mathbf{\Omega}_i) = \log_2 \frac{\left|\sum_{k \in \mathcal{K}_i} \mathbf{f}_{{\rm BB}_{i,k}} \mathbf{f}_{{\rm BB}_{i,k}}^H + \mathbf{\Omega}_i\right|}{|\mathbf{\Omega}_i|} \leq C_i, \quad \forall i. \tag{4}$$

[1]We assume the number of users served by each RRH is equal to the number of RF chains to ensure maximum spatial multiplexing efficiency.

B. Analog Beamforming

The baseband signal $\widetilde{\mathbf{x}}_{\mathrm{D},i}$ received at the i-th RRH is processed by analog beamformer $\mathbf{F}_{\mathrm{RF}_i} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{RF}}}$. The transmitted signal $\mathbf{x}_i \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ from RRH-i is given by

$$\mathbf{x}_i = \mathbf{F}_{\mathrm{RF}_i} \widetilde{\mathbf{x}}_{\mathrm{D},i} = \mathbf{F}_{\mathrm{RF}_i} \mathbf{x}_{\mathrm{D},i} + \mathbf{F}_{\mathrm{RF}_i} \mathbf{q}_i. \tag{5}$$

Note that the analog beamforming $\mathbf{F}_{\mathrm{RF}_i}$ is realized by a phase-shifter network, each element $\mathbf{F}_{\mathrm{RF}_i}(m,n)$ can be described as a constant-amplitude complex number, i.e. $|\mathbf{F}_{\mathrm{RF}_i}(m,n)| = \frac{1}{\sqrt{N_t}}$. By substituting (5) into (1), the received signal at the k-th user can be rewritten as

$$\begin{aligned} y_k &= \sum_{i=1}^{M} \mathbf{h}_{i,k}^H \mathbf{x}_i + n_k \\ &= \underbrace{\sum_{i=1}^{M} u_{i,k} \mathbf{h}_{i,k}^H \mathbf{F}_{\mathrm{RF}_i} \mathbf{f}_{\mathrm{BB}_{i,k}} s_k}_{\text{desired signal}} + \underbrace{\sum_{i=1}^{M} \mathbf{h}_{i,k}^H \mathbf{F}_{\mathrm{RF}_i} \mathbf{q}_i}_{\text{quantization noise}} \\ &\quad + \underbrace{\sum_{l=1, l\neq k}^{K} \sum_{i=1}^{M} u_{i,l} \mathbf{h}_{i,k}^H \mathbf{F}_{\mathrm{RF}_i} \mathbf{f}_{\mathrm{BB}_{i,l}} s_l}_{\text{inter-user interference}} + n_k, \end{aligned} \tag{6}$$

where $\mathbf{F}_{\mathrm{BB}_i} \triangleq [\mathbf{f}_{\mathrm{BB}_{i,\mathcal{K}_i(1)}}, \ldots, \mathbf{f}_{\mathrm{BB}_{i,\mathcal{K}_i(N_{\mathrm{RF}})}}]$ is the $N_{\mathrm{RF}} \times N_{\mathrm{RF}}$ digital beamformer of the i-th RRH. In (6), $u_{i,k}$ denotes the RRH-user pairing indicator such that $u_{i,k}$ is equal to 1, if $k \in \mathcal{K}_i$, otherwise $u_{i,k} = 0$. The total power consumed by each RRH can be expressed as

$$\begin{aligned} p_i(\mathbf{F}_{\mathrm{RF}_i}, \mathbf{F}_{\mathrm{BB}_i}, \mathbf{\Omega}_i) &= \mathbb{E}\{\|\mathbf{x}_i\|^2\} \\ &= \mathrm{trace}\left(\mathbf{F}_{\mathrm{RF}_i} \mathbf{F}_{\mathrm{BB}_i} \mathbf{F}_{\mathrm{BB}_i}^H \mathbf{F}_{\mathrm{RF}_i}^H\right) \\ &\quad + \mathrm{trace}\left(\mathbf{F}_{\mathrm{RF}_i} \mathbf{\Omega}_i \mathbf{F}_{\mathrm{RF}_i}^H\right). \end{aligned} \tag{7}$$

C. Channel Model

MmWave channels are known to have limited number of scattering paths. In this paper, we adopt the multipath channel model [9], [23], which allows us to accurately capture the mathematical structure present in mmWave channels. In particular, the downlink mmWave channel vector $\mathbf{h}_{i,k}$ consisted of L paths from the i-th RRH to the k-th user is modeled as

$$\mathbf{h}_{i,k} = \sqrt{\frac{N_{\mathrm{t}}}{L}} \beta_{i,k} \sum_{l=1}^{L} \alpha_{i,k}^l \mathbf{a}(\theta_{i,k}^l), \tag{8}$$

where $\beta_{i,k} \in (0,1]$ is the large-scale fading coefficient which includes the path loss and shadow fading [13], $\alpha_{i,k}^l \sim \mathcal{CN}(0,1)$ is the independent and identically distributed (i.i.d.) complex gain of the l-th propagation path. $\theta_{i,k}^l \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is the angle of departure (AoD). The array response vectors $\mathbf{a}(\theta)$ depends on the antenna array geometry. We assume that the commonly used uniform linear array (ULA) is employed. This transmit antenna array response vector $\mathbf{a}(\theta)$ can be written as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N_{\mathrm{t}}}} [1, e^{j\frac{2\pi}{\lambda} d \sin(\theta)}, \ldots, e^{j(N_{\mathrm{t}}-1)\frac{2\pi}{\lambda} d \sin(\theta)}]^T, \tag{9}$$

where λ is the signal wavelength, and d is the distance between antenna elements.

D. Problem Statement

In this paper, we study the joint design of user association, hybrid beamforming, and fronthaul compression without full CSI acquisition for C-RAN based cell-free mmWave MIMO systems. Conventionally, channel estimation is first executed at the BBU via uplink training. After that, hybrid beamforming and fronthaul compression can be designed based on the estimated CSI. However, this conventional approach is time-consuming and not practical for the considered system due to the employment of hybrid beamforming. To be specific, the hybrid beamforming architecture with limited number of RF chains cannot process and differentiate the signals of all users simultaneously. Full CSI acquisition of each RRH-user pair for such cell-free systems is challenging. In addition, based on the C-RAN architecture, the limited fronthaul capacity and RRH-user association also need to be taken into consideration in the hybrid beamforming design, which makes the problem more difficult to solve.

Therefore, in the following, we propose a train-and-design framework for the C-RAN enabled cell-free mmWave MIMO systems. Specifically, a two-stage uplink training approach is developed to assist the design of user association, hybrid beamforming, and fronthaul compression, during which the RRH analog beamforming and RRH-user association are determined. Then, given the training results and the obtained analog beamforming, the digital beamforming and fronthaul compression are calculated based on the effective baseband channel. While the proposed train-and-design framework decomposes the hybrid beamforming design into analog and digital domains which makes the solution sub-optimal, it can always guarantee comparable performance to the full-digital beamforming benchmark, as will be illustrated in the simulation results. Moreover, the proposed train-and-design framework not only avoids the explicit channel estimation, but also accomplishes the RRH-level design during the training process, which can reduce training overhead and enhance the efficiency of the signal processing design. In the next two sections, we consider two representative performance metrics, i.e. weighted sum-rate maximization and max-min fairness.

III. Proposed Train-and-Design Framework for Weighted Sum-Rate Maximization

A. Weighted Sum-Rate Maximization Problem Formulation

In this section, we investigate the designs of user association, hybrid beamforming, and fronthaul compression with aid of uplink training, for cell-free mmWave MIMO systems to maximize the weighted sum-rate. Given the received signal (6) at each user, the signal to interference-plus-noise ratio (SINR) of the k-the user is (10), as shown at the bottom of the next page. The achievable rate of the k-th user is calculated as

$$R_k = \log_2(1 + \gamma_k). \tag{11}$$

Our goal is to jointly design user association indicator $\{u_{i,k}\}_{i,k}$, hybrid beamforming $\{\mathbf{F}_{\mathrm{RF}_i}, \mathbf{F}_{\mathrm{BB}_i}\}_{i=1}^M$, and quantization noise covariance matrix $\{\mathbf{\Omega}_i\}_{i=1}^M$ to maximize the weighted sum-rate while satisfying the hardware and power

constraints. The objective function can be formulated as

$$\text{P1}: \max_{\substack{\{u_{i,k}\}_{i,k},\\ \{\mathbf{F}_{\text{RF}_i},\mathbf{F}_{\text{BB}_i},\mathbf{\Omega}_i\}_{i=1}^M}} \sum_{k=1}^K \omega_k R_k \tag{12a}$$

$$\text{s.t. } u_{i,k} \in \{0,1\},\ \forall i,k, \tag{12b}$$

$$\sum_{k=1}^K u_{i,k} = N_{\text{RF}}, \quad \forall i, \tag{12c}$$

$$c_i(\mathbf{F}_{\text{BB}_i}, \mathbf{\Omega}_i) \le C_i, \quad \forall i, \tag{12d}$$

$$p_i(\mathbf{F}_{\text{RF}_i}, \mathbf{F}_{\text{BB}_i}, \mathbf{\Omega}_i) \le P_i, \quad \forall i, \tag{12e}$$

$$|\mathbf{F}_{\text{RF}_i}(m,n)| = \frac{1}{\sqrt{N_{\text{t}}}}, \quad \forall m,n,i, \tag{12f}$$

where ω_k is the weight coefficient for the k-th user. Constraints (12b) and (12c) indicate the user-RRH pairing, and (12c) is imposed by the limited number of RF chains. The fronthaul capacity and power constraints are given by (12d) and (12e), respectively. Constraint (12f) is imposed by phase shifter networks. As mentioned before, P1 should be solved based on the availability of CSI, while full CSI acquisition is a challenging task for the considered system. Moreover, even with known CSI, the direct optimal solution to the combinatorial problem P1 is still intractable due to the non-convexity of the objective function and the hardware constraints (12d) and (12f).

Instead of directly solving P1, in the following, we propose a sub-optimal solution which decomposes the design into analog and digital domains. First, a two-stage uplink training approach is developed, during which the RRH analog beamforming $\{\mathbf{F}_{\text{RF}_i}\}_{i=1}^M$ and user-RRH association $\{u_{i,k}\}_{i,k}$ are designed to maximize the beamforming gain. Then, the digital beamformer $\{\mathbf{F}_{\text{BB}_i}\}_{i=1}^M$ and quantization noise covariance matrix $\{\mathbf{\Omega}_i\}_{i=1}^M$ are calculated at the BBU based on the training results.

B. Analog Beamforming and User Association via Two-Stage Uplink Training

In this subsection, we demonstrate the detailed designs of analog beamforming and user association for RRHs with the aid of uplink training, which also serves as the foundation for digital beamformer and fronthaul compression design.

Benefiting from the strong directionality of mmWave channel, the achievable rate of each user is highly dependent on the beamforming gain. This is due to the fact that the use of beamforming can mitigate the interference in mmWave communications. Thus, we aim to jointly design analog beamforming and user association at each RRH to maximize the weighted sum-beamforming-gain. Given the sparsity and directionality of mmWave channel shown in (8), the key to achieve high beamforming gain is to align the beamformer with specific AoD. Due to the constraints on the RF hardware, i.e., the constant amplitude and discrete angles for the RF phase shifters, the analog beamforming vectors can be selected from a pre-defined codebook [9]. The codebook is consisted of beamforming vectors which have the same form of the array response vector and can be parameterized by a spatial angle. Based on this fact, we propose a codebook-based approach for analog beamformer selection. The codebook consisting of feasible analog beamformers is defined as [9]

$$\mathcal{F} = \left\{ \mathbf{f}_n = \mathbf{a}\left(\frac{2\pi n}{N_{\text{res}}}\right),\ n = 1,\ldots,N_{\text{res}} \right\}, \tag{13}$$

where N_{res} is the resolution to quantize the spatial angle. Then, the weighted sum-beamforming-gain maximization for the designs of analog beamforming and user association can be formulated as

$$\text{P2}: \max_{\left\{u_{i,k},\mathbf{f}_{\text{RF}_{i,k}}\right\}_{i,k}} \sum_{k=1}^K \omega_k \sum_{i=1}^M u_{i,k} \left|\mathbf{h}_{i,k}^H \mathbf{f}_{\text{RF}_{i,k}}\right|^2 \tag{14a}$$

$$\text{s.t. Constraints (12b) and (12c)},$$

$$\mathbf{f}_{\text{RF}_{i,k}} \in \mathcal{F}, \quad \text{if } u_{i,k} = 1,\ \forall i,k, \tag{14b}$$

where $\mathbf{f}_{\text{RF}_{i,k}}$ is the beamformer vector of the i-th RRH to the k-th user. Furthermore, by exchanging the order of two summations, the objective function of P2 can be rewritten as

$$\sum_{k=1}^K \omega_k \sum_{i=1}^M u_{i,k} \left|\mathbf{h}_{i,k}^H \mathbf{f}_{\text{RF}_{i,k}}\right|^2 = \sum_{i=1}^M \sum_{k=1}^K \omega_k u_{i,k} \left|\mathbf{h}_{i,k}^H \mathbf{f}_{\text{RF}_{i,k}}\right|^2. \tag{15}$$

By this operation, the maximization of the weighted sum-beamforming-gain of all users can be equivalently decomposed into M separate sub-problems, each of which refers to the maximization of the achievable beamforming gain at each RRH. The i-th sub-problem of P2 can be expressed as:

$$\text{P2-}i: \max_{\left\{u_{i,k},\mathbf{f}_{\text{RF}_{i,k}}\right\}_{i,k}} \sum_{k=1}^K u_{i,k}\omega_k \left|\mathbf{h}_{i,k}^H \mathbf{f}_{\text{RF}_{i,k}}\right|^2 \tag{16a}$$

$$\text{s.t. } u_{i,k} \in \{0,1\}, \quad \forall k, \tag{16b}$$

$$\sum_{k=1}^K u_{i,k} = N_{\text{RF}}, \tag{16c}$$

$$\mathbf{f}_{\text{RF}_{i,k}} \in \mathcal{F}, \quad \text{if } u_{i,k} = 1, \quad \forall k. \tag{16d}$$

The optimal solution of P2-i can be obtained by first finding the optimal beamformer $\tilde{\mathbf{f}}_{\text{RF}_{i,k}}$ between the i-th RRH and the k-th user as

$$\tilde{\mathbf{f}}_{\text{RF}_{i,k}} = \underset{\mathbf{f}_{\text{RF}_{i,k}} \in \mathcal{F}}{\text{argmax}} = \left|\mathbf{h}_{i,k}^H \mathbf{f}_{\text{RF}_{i,k}}\right|^2, \quad \forall k, \tag{17}$$

$$\gamma_k = \frac{\left|\sum_{i=1}^M u_{i,k} \mathbf{h}_{i,k}^H \mathbf{F}_{\text{RF}_i} \mathbf{f}_{\text{BB}_{i,k}}\right|^2}{\sum_{l\neq k}^H \left|\sum_{i=1}^M u_{i,l} \mathbf{h}_{i,k}^H \mathbf{F}_{\text{RF}_i} \mathbf{f}_{\text{BB}_{i,l}}\right|^2 + \sum_{i=1}^M \mathbf{h}_{i,k}^H \mathbf{F}_{\text{RF}_i} \mathbf{\Omega}_i \mathbf{F}_{\text{RF}_i}^H \mathbf{h}_{i,k} + \sigma_k^2}. \tag{10}$$

then selecting the N_{RF} analog beamformer which can achieve the largest weighted beamforming gains and setting $u_{i,k}=1$.

Since P2-i should be solved based on the knowledge of channel vector $\mathbf{h}_{i,k}$ which is challenging to estimate, we propose a two-stage uplink training mechanism to assist the above design where the analog beamforming and user association can be obtained during the training procedure without explicit channel estimation. In the first stage, beam training is implemented at the i-th RRH to determine the optimal beamformers for each user and select N_{RF} of them which lead to maximum received signal power. Then, in the second training stage, the i-th RRH identifies the set of users and setting the corresponding indicator $u_{i,k}=1$.

Particularly, during the first stage, all K users simultaneously transmit pilot symbol $\psi_{k,q}^{(1)}$, $\forall k$, to the RRHs at the q-th time slot. To simplify the analysis and expression, we assume that all the users send the same symbol $\psi_{k,q}^{(1)}=\psi^{(1)}$, $\forall k,q$. We let $\rho_k=\omega_k\rho$ be the transmit power of the k-th user in all first-stage transmissions, where ω_k is the weight coefficient and ρ indicates a base power. Based on the hybrid beamforming architecture, N_{RF} measurements per instant can be performed at each RRH. At the q-th time slot, we assume the i-th RRH applies a training beamforming matrix $\mathbf{F}_{i,q}$ whose columns are selected from codebook $\mathcal{F}$ to combine the received signals. Then, the resulting signal can be written as

$$\mathbf{y}_{i,q}^{(1)}=\mathbf{F}_{i,q}^{H}\sum_{k=1}^{K}\sqrt{\rho_k}\mathbf{h}_{i,k}\psi^{(1)}+\mathbf{F}_{i,q}^{H}\mathbf{z}_{i,q}, \tag{18}$$

where $\mathbf{h}_{i,k}$ is the dual uplink channel vector. $\mathbf{z}_{i,q}\sim\mathcal{CN}(\mathbf{0},\sigma_{\mathrm{u}}^2\mathbf{I}_{N_{\mathrm{t}}})$ denotes the noise vector at RRH-i during uplink training. After $\tau_1=\frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}}$ time slots, the i-th RRH can exhaustively search over the codebook and obtain

$$\mathbf{y}_{i}^{(1)}=\sqrt{\tau_1\rho_k}\begin{bmatrix}\mathbf{F}_{i,1}^{H}\\ \vdots\\ \mathbf{F}_{i,\tau_1}^{H}\end{bmatrix}\sum_{k=1}^{K}\mathbf{h}_{i,k}\psi^{(1)}+\begin{bmatrix}\mathbf{F}_{i,1}^{H}\\ \vdots\\ \mathbf{F}_{i,\tau_1}^{H}\end{bmatrix}\mathbf{z}_i \tag{19}$$

$$=\sqrt{\tau_1\rho_k}\begin{bmatrix}\mathbf{f}_{1}^{H}\\ \vdots\\ \mathbf{f}_{N_{\mathrm{res}}}^{H}\end{bmatrix}\sum_{k=1}^{K}\mathbf{h}_{i,k}\psi^{(1)}+\begin{bmatrix}\mathbf{f}_{1}^{H}\\ \vdots\\ \mathbf{f}_{N_{\mathrm{res}}}^{H}\end{bmatrix}\mathbf{z}_i. \tag{20}$$

Note that $\mathbf{y}_i^{(1)}\in\mathbb{C}^{N_{\mathrm{res}}\times 1}$ contains the beam training results, and the square of magnitude of the n-th element $|\mathbf{y}_i^{(1)}(n)|^2$, $n=1,\ldots,N_{\mathrm{res}}$, indicates the received signal power when the n-th candidate beam is utilized to process the pilot at the i-th RRH. We can successively select N_{RF} indices of the beamformer candidates which lead to N_{RF} largest received signal power. The selection of the l-th beam index is given by

$$n_i^{(l)\star}=\underset{\substack{n_i^{(l)}=1,\ldots,N_{\mathrm{res}},n_i^{(l)}\neq n_i^{(j)\star}\\ j=1,\ldots,l-1}}{\operatorname{argmax}}\left|\mathbf{y}_i^{(1)}(n_i^{(l)})\right|^2,\quad\forall i. \tag{21}$$

Therefore, the optimal analog beamformer at the i-th RRH is determined as $\mathbf{F}_{\mathrm{RF}_i}^{\star}=[\mathcal{F}(n_i^{(1)\star}),\ldots,\mathcal{F}(n_i^{(N_{\mathrm{RF}})\star})]$.

After selecting the optimal N_{RF} beamformers for each RRH, in the second training stage, we propose to identify the set of users served by each RRH and set the corresponding user association indicator $u_{i,k}$ as 1. Specifically, each user transmits the orthogonal sequence $\boldsymbol{\psi}_k^{(2)}\in\mathbb{C}^{\tau_2\times 1}$ in successive τ_2 time instants, where $\tau_2\geq K$ should be satisfied to ensure the orthogonality of training sequences. We further assume $\|\boldsymbol{\psi}_k^{(2)}\|^2=1$. RRHs use the optimal analog beamformer $\mathbf{F}_{\mathrm{RF}_i}^{\star}$ to combine the signals. Then, the received signal at the i-th RRH is

$$\mathbf{Y}_i^{(2)}=\sqrt{\tau_2\rho_k}\mathbf{F}_{\mathrm{RF}_i}^{\star H}\sum_{k=1}^{K}\mathbf{h}_{i,k}(\boldsymbol{\psi}_k^{(2)})^H+\mathbf{F}_{\mathrm{RF}_i}^{\star H}\mathbf{Z}_i, \tag{22}$$

where $\mathbf{Z}_i\in\mathbb{C}^{N_{\mathrm{t}}\times\tau_2}$ is the noise matrix given by concatenating τ_2 noise vectors. By multiplying $\frac{\boldsymbol{\psi}_k^{(2)}}{\sqrt{\tau_2\rho_k}}$ to the right of $\mathbf{Y}_i^{(2)}$, the i-th RRH can estimate the effective baseband channel to the k-th user, which is expressed as [18]

$$\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}=\frac{1}{\sqrt{\tau_2\rho_k}}\mathbf{Y}_i^{(2)}\boldsymbol{\psi}_k^{(2)}=\mathbf{F}_{\mathrm{RF}_i}^{\star H}\mathbf{h}_{i,k}+\frac{1}{\sqrt{\tau_2\rho_k}}\mathbf{F}_{\mathrm{RF}_i}^{\star H}\mathbf{Z}_i\boldsymbol{\psi}_k^{(2)}. \tag{23}$$

Because of the directional transmission in mmWave systems, the transmit/receive power mainly concentrates on the analog beamforming directions. Therefore, we propose an energy-based approach to determine the user indices by checking the norm of the effective baseband channel $\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}$. To be specific, at the i-th RRH, the user indices with the N_{RF} largest baseband channel energy will be successively selected. For the m-th user index selection at the i-th RRH, we have

$$k_i^{(m)\star}=\underset{\substack{k_i^{(m)}=1,\ldots,K,k_i^{(m)}\neq k_i^{(j)\star}\\ j=1,\ldots,m-1}}{\operatorname{argmax}}\left|(\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}})^H\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}\right|^2,\quad\forall i. \tag{24}$$

Then, the set of users served by the i-th RRH can be formed as $\mathcal{K}_i=\{k_i^{(m)\star}:m=1,\ldots,N_{\mathrm{RF}}\}$, and the corresponding user association indicator $u_{i,k}$ is set as 1 if $k\in\mathcal{K}_i$. Finally, the user association indicators $u_{i,k}$ and the effective baseband channel $\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}$, $\forall i,k$, are fed back to the BBU, which provide the foundation for the design of digital beamforming and fronthaul compression.

We note that RRHs have the ability to support the uplink training process. As stated in [43], functional splits are exploited in the C-RAN, where some baseband functionalities at the physical layer can be implemented at the RRH. One of promising functional split approaches is to implement resource demapping for the uplink and resource mapping for the downlink at the RRH. This implies that the RRH can distinguish among the different physical channels multiplexed in the resource blocks and implement appropriate analog beamforming. Therefore, the above uplink training procedure can be realized at the RRHs.

C. Digital Beamforming and Fronthaul Compression

Based on the user association indicators $u_{i,k}$, $\forall i,k$, and effective baseband channel $\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}$, $\forall i,k$, obtained through uplink training, the baseband signal processing (including digital beamforming and fronthaul compression for all RRHs) is

designed in the centralized BBU. We first rewrite the received signal at the k-th user as

$$y_k = \sum_{l=1}^{K}\sum_{i=1}^{M} u_{i,l}((\tilde{\mathbf{h}}_{i,k}^{\text{eff}})^H + \mathbf{e}_{i,k}^H)\mathbf{f}_{\text{BB}_{i,l}} s_l + \sum_{i=1}^{M}((\tilde{\mathbf{h}}_{i,k}^{\text{eff}})^H + \mathbf{e}_{i,k}^H)\mathbf{q}_i + n_k, \quad (25)$$

where $\mathbf{e}_{i,k}$ denotes the estimation error vector from the i-th RRH to the k-th user.

Estimation Error Analysis: The perfect effective baseband channel is given by $\mathbf{h}_{i,k}^{\text{eff}} \triangleq \mathbf{F}_{\text{RF}_i}^{\star\,H}\mathbf{h}_{i,k}$, where $\mathbf{F}_{\text{RF}_i}^{\star}$ is the analog beamforming matrix obtained in the first training stage. The estimation error vector can be expressed as $\mathbf{e}_{i,k} \triangleq \mathbf{h}_{i,k}^{\text{eff}} - \tilde{\mathbf{h}}_{i,k}^{\text{eff}}$. Based on the results in (23), we have

$$\mathbf{e}_{i,k} = \mathbf{h}_{i,k}^{\text{eff}} - \tilde{\mathbf{h}}_{i,k}^{\text{eff}} = \frac{1}{\sqrt{\tau_2\rho_k}}\mathbf{F}_{\text{RF}_i}^{\star\,H}\mathbf{Z}_i\boldsymbol{\psi}_k^{(2)}. \quad (26)$$

It can be observed that the estimation error is distributed as $\mathbf{e}_{i,k} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Xi}_{i,k})$, in which the error covariance matrix $\boldsymbol{\Xi}_{i,k}$ is given by

$$\boldsymbol{\Xi}_{i,k} = \frac{\sigma_{\text{u}}^2}{\tau_2\rho_k}\mathbf{F}_{\text{RF}_i}^{\star\,H}\mathbf{F}_{\text{RF}_i}^{\star}, \quad i = 1,\ldots,M,\ k = 1,\ldots,K. \quad (27)$$

From (27), we can see that the error covariance is proportional to the noise power, and inversely proportional to the training sequence length and transmit power during the second training stage.

For the purpose of expression simplicity, we merge the digital beamformer of all RRHs into a single vector as $\mathbf{f}_{\text{BB}_k} \triangleq [\mathbf{f}_{\text{BB}_{1,k}}^T,\ldots,\mathbf{f}_{\text{BB}_{M,k}}^T]^T,\ \forall k$, and define $\mathbf{F}_{\text{BB}} \triangleq [\mathbf{f}_{\text{BB}_1},\ldots,\mathbf{f}_{\text{BB}_K}]$. Similarly, we construct $\tilde{\mathbf{h}}_k \triangleq [(\tilde{\mathbf{h}}_{1,k}^{\text{eff}})^T,\ldots,(\tilde{\mathbf{h}}_{M,k}^{\text{eff}})^T]^T$ and $\mathbf{e}_k \triangleq [\mathbf{e}_{1,k}^T,\ldots,\mathbf{e}_{M,k}^T]^T$. We define the aggregated effective channel vector associated with user-l and user-k as $\bar{\mathbf{h}}_{l,k} \triangleq [u_{1,l}(\tilde{\mathbf{h}}_{1,k}^{\text{eff}})^T,\ldots,u_{M,l}(\tilde{\mathbf{h}}_{M,k}^{\text{eff}})^T]^T$, as well as its estimation error vector $\bar{\mathbf{e}}_{l,k} \triangleq [u_{1,l}\tilde{\mathbf{e}}_{1,k}^T,\ldots,u_{M,l}\tilde{\mathbf{e}}_{M,k}^T]^T$, and the aggregated quantization noise covariance matrix $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1,\ldots,\boldsymbol{\Omega}_M)$, respectively. Then, the digital beamforming and compression design problem can be rewritten as

P3 :

$$\max_{\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}}\ f_o\left(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}\right) \quad (28a)$$

$$\text{s.t. } \log_2\frac{\left|\sum_{k=1}^{K} u_{i,k}\mathbf{E}_i^H\mathbf{f}_{\text{BB}_k}\mathbf{f}_{\text{BB}_k}^H\mathbf{E}_i + \boldsymbol{\Omega}_i\right|}{|\boldsymbol{\Omega}_i|} \le C_i, \quad \forall i, \quad (28b)$$

$$\sum_{k=1}^{K}\text{trace}\left(u_{i,k}\mathbf{F}_{\text{RF}_i}^{\star}\mathbf{E}_i^H\mathbf{f}_{\text{BB}_k}\mathbf{f}_{\text{BB}_k}^H\mathbf{E}_i\mathbf{F}_{\text{RF}_i}^{\star\,H}\right) + \text{trace}\left(\mathbf{F}_{\text{RF}_i}\boldsymbol{\Omega}_i\mathbf{F}_{\text{RF}_i}^H\right) \le P_i, \quad \forall i, \quad (28c)$$

with the objective function f_o defined by

$$f_o\left(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}\right) = \sum_{k=1}^{K}\omega_k\log_2\left(1 + \frac{\left|\bar{\mathbf{h}}_{k,k}^H\mathbf{f}_{\text{BB}_k}\right|^2}{\zeta_k(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega})}\right). \quad (29)$$

The function $\zeta_k(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega})$ is expressed as:

$$\zeta_k(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}) = \sum_{l\neq k}\left|\bar{\mathbf{h}}_{l,k}^H\mathbf{f}_{\text{BB}_l}\right|^2 + \tilde{\mathbf{h}}_k^H\boldsymbol{\Omega}\tilde{\mathbf{h}}_k + \sum_{l=1}^{K}\mathbf{f}_{\text{BB}_l}^H\boldsymbol{\Upsilon}_{l,k}\mathbf{f}_{\text{BB}_l} + \text{trace}(\boldsymbol{\Xi}_k\boldsymbol{\Omega}) + \sigma_k^2, \quad (30)$$

in which $\boldsymbol{\Upsilon}_{l,k} \triangleq \text{diag}(u_{1,l}\boldsymbol{\Xi}_{1,k},\ldots,u_{M,l}\boldsymbol{\Xi}_{M,k})$ and $\boldsymbol{\Xi}_k \triangleq \text{diag}(\boldsymbol{\Xi}_{1,k},\ldots,\boldsymbol{\Xi}_{M,k})$, respectively. In (28b) and (28c), $\mathbf{E}_i \triangleq [\mathbf{0}_{N_{\text{RF}}\times N_{\text{RF}}(i-1)}^H\ \mathbf{I}_{N_{\text{RF}}}\ \mathbf{0}_{N_{\text{RF}}\times N_{\text{RF}}(M-i)}^H]^H$ is defined as a shaping matrix. Finding the global optimum solution of P3 is difficult due to the non-convexity of the objective function and fronthaul constraint. Next, we propose a low-complexity method to find a stationary solution of P3.

Inspired by the fractional programming algorithm introduced in [39] and [40], we apply quadratic transform to the SINR term in (29), and the original objective function in (29) can be reformulated as:

$$f_q(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega},\mathbf{m}) = \sum_{k=1}^{K}\omega_k\log_2\Big(1 + 2\mathfrak{Re}\left\{m_k^*\bar{\mathbf{h}}_{k,k}^H\mathbf{f}_{\text{BB}_k}\right\} - |m_k|^2\zeta_k(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega})\Big), \quad (31)$$

where variable m_k is introduced with respect to each user, and $\mathbf{m}$ is the vector containing of auxiliary variable $\{m_k\}$. Then, the original objective function P3 is equivalent to

$$\text{P4 : } \{\mathbf{F}_{\text{BB}}^{\star},\boldsymbol{\Omega}^{\star},\mathbf{m}^{\star}\} = \arg\max\ f_q\left(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega},\mathbf{m}\right) \quad \text{s.t. Constraints (28b) and (28c).} \quad (32a)$$

The variables in P4 can be optimized through an iterative fashion. When the variables $\mathbf{F}_{\text{BB}}$ and $\boldsymbol{\Omega}$ are fixed, the optimal value of $\mathbf{m}$ can be solved by setting $\frac{\partial f_q}{\partial m_k} = 0$, which is expressed as

$$m_k^{\star} = \frac{\bar{\mathbf{h}}_{k,k}^H\mathbf{f}_{\text{BB}_k}}{\zeta_k(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega})}, \quad k = 1,\ldots,K. \quad (33)$$

We then consider the optimization of $\{\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}\}$ with fixed value of $\mathbf{m}$. The objective $f_q\left(\mathbf{F}_{\text{BB}},\boldsymbol{\Omega},\mathbf{m}\right)$ in P4 is a convex function with respect to $\{\mathbf{F}_{\text{BB}},\boldsymbol{\Omega}\}$, but the fronthaul constraint is still non-convex. We then propose to further transform the constraint (28b) into a solvable term. In particular, for positive definite Hermitian matrices $\mathbf{A},\ \mathbf{B} \in \mathbb{C}^{N\times N}$, we have [31]

$$\log_2|\mathbf{A}| \le \log_2|\mathbf{B}| + \text{trace}\{\mathbf{B}^{-1}\mathbf{A}\} - N, \quad (34)$$

with equality achieved when $\mathbf{A} = \mathbf{B}$. Therefore, based on the above method, we can approximate the fronthaul constraint of the i-th RRH with the following convex expression:

$$\log_2\left|\sum_{k=1}^{K}u_{i,k}\mathbf{E}_i^H\mathbf{f}_{\text{BB}_k}\mathbf{f}_{\text{BB}_k}^H\mathbf{E}_i + \boldsymbol{\Omega}_i\right| - \log_2|\boldsymbol{\Omega}_i| \le g_i(\mathbf{F}_{\text{BB}_i},\boldsymbol{\Omega}_i,\boldsymbol{\Sigma}_i), \quad (35)$$

where $g_i(\mathbf{F}_{\text{BB}_i},\boldsymbol{\Omega}_i,\boldsymbol{\Sigma}_i)$ is an upper bound of fronthaul capacity and is defined as

$$g_i(\mathbf{F}_{\text{BB}_i},\boldsymbol{\Omega}_i,\boldsymbol{\Sigma}_i) = \log_2|\boldsymbol{\Sigma}_i| - \log_2|\boldsymbol{\Omega}_i| - N_{\text{RF}}$$

$$+\operatorname{trace}\left\{\boldsymbol{\Sigma}_i^{-1}\left(\sum_{k=1}^{K} u_{i,k}\mathbf{E}_i^H \mathbf{f}_{\mathrm{BB}_k}\mathbf{f}_{\mathrm{BB}_k}^H \mathbf{E}_i + \boldsymbol{\Omega}_i\right)\right\}. \quad (36)$$

In (35), (36), $\boldsymbol{\Sigma}_i$ is introduced as an auxiliary matrix and the equality in (35) holds only when

$$\boldsymbol{\Sigma}_i^{\star} = \sum_{k=1}^{K} u_{i,k}\mathbf{E}_i^H \mathbf{f}_{\mathrm{BB}_k}\mathbf{f}_{\mathrm{BB}_k}^H \mathbf{E}_i + \boldsymbol{\Omega}_i. \quad (37)$$

Therefore, we replace the original fronthaul link constraint with the upper bound (36), and reformulate problem P4 into P5:

$$\mathrm{P5}: \ \{\mathbf{F}_{\mathrm{BB}}^{\star}, \boldsymbol{\Omega}^{\star}, \boldsymbol{\Sigma}^{\star}\} = \operatorname{argmax} \ f_q\left(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \mathbf{m}^{\star}\right) \quad (38a)$$

$$\text{s.t. Constraint (28c)},$$

$$g_i(\mathbf{F}_{\mathrm{BB}_i}, \boldsymbol{\Omega}_i, \boldsymbol{\Sigma}_i) \le C_i, \quad \forall i. \quad (38b)$$

P5 is convex with respect to $\{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}\}$, and each variable in P5 can be solved in a block coordinate descent manner. To be specific, when $\{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}\}$ are fixed, the optimal value of $\{\boldsymbol{\Sigma}_i\}_{i=1}^M$ is given by (37). On the other hand, when $\{\boldsymbol{\Sigma}_i^{\star}\}_{i=1}^M$ is determined, the optimal values of $\{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}\}$ can be solved using convex optimization solver such as CVX with polynomial complexity. To conclude, the variables $\{\mathbf{m}, \mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}\}$ can be iteratively optimized until the convergence is achieved. The complete train-and-design solution for weighted sum-rate maximization is summarized in Algorithm 1.

Algorithm 1 Proposed Train-and-Design Solution for Weighted Sum-Rate Maximization

1. Joint user association and analog beamforming through two-stage uplink training.

Stage 1: Beam selection.

1) Users continuously send pilot signal $\psi^{(1)}$, $\forall k$, during $\tau_1 = \frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}}$ time slots, while each RRH implements beam sweeping over the codebook. The obtained training signals can be written as in (20).
2) Each RRH determines the optimal analog beamformer $\mathbf{F}_{\mathrm{RF}_i}^{\star}$ based on (21), which results in the largest beamforming gain.

Stage 2: User pairing and effective baseband channel acquisition.

1) Users transmit unique orthogonal sequences $\boldsymbol{\psi}_k^{(2)} \in \mathbb{C}^{\tau_2 \times 1}$ during τ_2 time instants, while each RRH combines the signals with $\mathbf{F}_{\mathrm{RF}_i}^{\star}$.
2) Each RRH obtains the effective baseband channel $\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}$ and determines $\{u_{i,k}\}_{i,k}$ as in (23) and (24), respectively.
3) Each RRH feeds the user association indicator and effective baseband channel back to the BBU through fronthaul link.

2. Design of digital beamforming and fronthaul compression at the BBU.

Initialize $\mathbf{F}_{\mathrm{BB}}$ and $\boldsymbol{\Omega}$ that satisfy the fronthaul link and power constraints.

repeat

Calculate $\boldsymbol{\Sigma}_i^{\star}$ of each RRH as in (37).

Solve P5 assuming fixed $\boldsymbol{\Sigma}^{\star}$ and obtain $\{\mathbf{F}_{\mathrm{BB}}^{\star}, \boldsymbol{\Omega}^{\star}\}$.

Update $m_k^{\star}$ as in (33).

until convergence

D. Convergence Analysis

In this section, we analyze the convergence of the digital beamforming design proposed in Sec. III-C by first showing that the reformulated P4 is equivalent to P3. Then, we will show that the proposed algorithm can obtain a stationary solution of P4 which is also a stationary point of P3.

Given the objective function $f_q(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \mathbf{m})$ in P4, we can calculate the optimal value of m_k by setting $\frac{\partial f_q}{\partial m_k} = 0$, which is given by (33). Then, the optimal value of P4 is exactly equal to that of P3. The equivalence is therefore established. In addition, the solution of P4 lies in a block coordinate descent manner. We introduce a superscript t to each variable as the iteration index. In the t-th iteration, we can obtain the objective value as $f_q(\mathbf{F}_{\mathrm{BB}}^{(t)}, \boldsymbol{\Omega}^{(t)}, \mathbf{m}^{(t)})|_{\boldsymbol{\Sigma}^{(t)}}$, where $\{\mathbf{F}_{\mathrm{BB}}^{(t)}, \boldsymbol{\Omega}^{(t)}\}$ is calculated based on $\boldsymbol{\Sigma}^{(t)}$, and $\mathbf{m}^{(t)}$ is determined by (33) using $\{\mathbf{F}_{\mathrm{BB}}^{(t)}, \boldsymbol{\Omega}^{(t)}\}$. Then, we have

$$\begin{aligned} & f_q(\mathbf{F}_{\mathrm{BB}}^{(t)}, \boldsymbol{\Omega}^{(t)}, \mathbf{m}^{(t)})|_{\boldsymbol{\Sigma}^{(t)}} \\ & \overset{(a)}{\le} f_q(\mathbf{F}_{\mathrm{BB}}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)}, \mathbf{m}^{(t)})|_{\boldsymbol{\Sigma}^{(t+1)}} \\ & \overset{(b)}{\le} f_q(\mathbf{F}_{\mathrm{BB}}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)}, \mathbf{m}^{(t+1)})|_{\boldsymbol{\Sigma}^{(t+1)}}. \end{aligned}$$

(a) holds because at the end of the t-th iteration, we update $\boldsymbol{\Sigma}^{(t+1)}$ as

$$\boldsymbol{\Sigma}_i^{(t+1)} = \sum_{k=1}^{K} u_{i,k}\mathbf{E}_i^H \mathbf{f}_{\mathrm{BB}_k}^{(t)}(\mathbf{f}_{\mathrm{BB}_k}^{(t)})^H \mathbf{E}_i + \boldsymbol{\Omega}_i^{(t)}, \quad \forall i, \quad (39)$$

and (35) holds with equality given $\{\mathbf{F}_{\mathrm{BB}}^{(t)}, \boldsymbol{\Omega}^{(t)}, \boldsymbol{\Sigma}^{(t+1)}\}$. Then, $\{\mathbf{F}_{\mathrm{BB}}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)}\}$ are obtained based on P5 which maximizes f_q with $\boldsymbol{\Sigma}^{(t+1)}$, indicating that the newly obtained objective value $f_q(\mathbf{F}_{\mathrm{BB}}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)}, \mathbf{m}^{(t)})|_{\boldsymbol{\Sigma}^{(t+1)}}$ must increase or at least not decrease. (b) follows since the update of $\mathbf{m}$ maximizes f_q when the other variables are fixed. Thus, the sum-rate objective $f_q(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \mathbf{m})$ in P4 is monotonically non-decreasing after each iteration, and finally converge to a stationary point which is also a stationary solution of P3.

E. Complexity Evaluation

The total required signaling overhead of the proposed approach is $\frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}} + \tau_2$, where τ_2 is the overhead in the second training stage. The computational complexity of the proposed algorithm comes from two parts, i.e., the designs in analog and digital domains. The complexity of jointly determining the user association and analog beamforming is $O(N_{\mathrm{res}} K M N_{\mathrm{RF}} \tau_2^2)$ which is done at the RRHs. For digital beamforming and fronthaul compression design in the BBU, the computational complexity of the proposed design is dominated by the joint optimization of $\{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}\}$, i.e., to solve P5 with fixed $\boldsymbol{\Sigma}$. The solution is obtained by solving a convex optimization problem, which can be efficiently implemented by primal-dual interior point method with approximate

complexity of $O((K+MN_{\mathrm{RF}})^{3.5})$ [41]. Suppose it takes T total number of iterations to converge, the computational complexity of digital signal processing is $O((K+MN_{\mathrm{RF}})^{3.5}T)$, and the overall computational complexity of Algorithm 1 is therefore $O((K+MN_{\mathrm{RF}})^{3.5}T+N_{\mathrm{res}}KMN_{\mathrm{RF}}\tau_2^2)$.

IV. Proposed Train-and-Design Framework for Max-Min Fairness

A. Max-Min Fairness Problem Formulation

In addition to weighted sum-rate maximization, fairness is another important performance metric for wireless communication systems. In this section, we study the max-min fairness problem for the considered cell-free mmWave MIMO system. We aim to jointly design user association and hybrid beamforming along with quantization noise covariance matrix to maximize the minimum SINR among users. The objective function is formulated as

$$\begin{aligned}\mathrm{F1}: \max_{\{u_{i,k}\}_{i,k},\left\{\mathbf{F}_{\mathrm{RF}_i},\mathbf{F}_{\mathrm{BB}_i},\mathbf{\Omega}_i\right\}_{i=1}^{M}} \min_k \frac{1}{\eta_k}\gamma_k \\ \text{s.t. Constraints (12b)-(12f),}\end{aligned} \tag{40a}$$

in which η_k denotes the weight coefficient of the k-th user. F1 is also a combinatorial problem which makes finding the global optimal solution challenging. Similarly, we propose to separately design the analog and digital signal processing based on a train-and-design framework, where a two-stage uplink training scheme is first proposed to assist user association and analog beamforming design. Then, the digital beamforming and fronthaul compression are optimized based on the training results.

B. User Association and Analog Beamforming via Two-Stage Uplink Training

In this subsection, we present the analog beamformer and users association designs with the aid of uplink training. Since the achievable SINR of each user is highly dependent on the beamforming gain, we propose to jointly determine RRH-user pairing $\{u_{i,k}\}_{i,k}$ and RRH analog beamforming $\{\mathbf{F}_{\mathrm{RF}_i}\}_{i=1}^{M}$ to maximize the minimum beamforming gain among users, which can be formulated as

$$\begin{aligned}\mathrm{F2}: \max_{\left\{u_{i,k},\mathbf{f}_{\mathrm{RF}_{i,k}}\right\}_{i,k}} \min_k \frac{1}{\eta_k}\sum_{i=1}^{M} u_{i,k}\left|\mathbf{h}_{i,k}^H\mathbf{f}_{\mathrm{RF}_{i,k}}\right|^2 \\ \text{s.t. Constraints (12b), (12c), and (16d).}\end{aligned} \tag{41a}$$

where $\mathcal{F}$ is a predefined codebook as in (13) for analog beamformer selection. F2 can be solved via exhaustive search over the combinations of every possible $\{u_{i,k}\}$ and analog beamformer candidate, which will cause very high complexity. To alleviate the computational complexity, we propose a low-complexity scheme where optimal beamformers for each RRH-user pair are first selected, followed by a successive elimination procedure to determine the user association indicators.

In particular, we first determine the optimal analog beamformers for all RRH-user pairs while ignoring the user association design. By setting $u_{i,k}=1$, $\forall i,k$, maximizing the minimum beamforming gain is equivalently to finding the optimal analog beamformers for all RRH-user pairs, which can be written as

$$\tilde{\mathbf{f}}_{i,k}=\max_{\mathbf{f}_{i,k}\in\mathcal{F}}\left|\mathbf{h}_{i,k}^H\mathbf{f}_{i,k}\right|^2,\quad\forall i,k. \tag{42}$$

The largest accumulative beamforming gain of each user ξ_k is obtained as

$$\xi_k=\frac{1}{\eta_k}\sum_{i=1}^{M}\left|\mathbf{h}_{i,k}^H\tilde{\mathbf{f}}_{i,k}\right|^2, \tag{43}$$

and the minimum beamforming gain among users is given by

$$\xi_{\min}=\min_k \xi_k. \tag{44}$$

Then, we reconsider the user association design and propose to maximize the minimum accumulative beamforming gain through a successive elimination procedure. Given the optimal analog beamformers in (42), we find the RRH which contributes to the least beamforming gain to the k-th user:

$$\hat{i}_k=\operatorname*{argmin}_{i=1,\ldots,M}\left|\mathbf{h}_{i,k}^H\tilde{\mathbf{f}}_{i,k}\right|^2,\quad\forall k, \tag{45}$$

and calculate the corresponding remaining beamforming gain of each user when eliminating the effect of beamformer $\tilde{\mathbf{f}}_{\hat{i}_k,k}$:

$$\tilde{\xi}_k=\frac{1}{\eta_k}\sum_{i\neq\hat{i}_k}^{M}\left|\mathbf{h}_{i,k}^H\tilde{\mathbf{f}}_{i,k}\right|^2,\quad\forall k. \tag{46}$$

Next, we find the $\bar{k}$-th user whose remaining beamforming gain $\tilde{\xi}_{\bar{k}}$ is the largest, and eliminate $\tilde{\mathbf{f}}_{\hat{i}_{\bar{k}},\bar{k}}$ from the $\hat{i}_{\bar{k}}$-th RRH to the $\bar{k}$-th user. The corresponding user association indicator $u_{\hat{i}_{\bar{k}},\bar{k}}$ is set to 0. Note that if the constraint $\sum_{k=1}^{K}u_{\hat{i}_k,k}=N_{\mathrm{RF}}$ is already satisfied at the $\hat{i}_k$-th RRH, we then remove the effect of the beamformer which results in the second least beamforming gain whose RF chain constraint (12c) is not satisfied at the corresponding RRH. Following this procedure, the beams are successively eliminated until the RF chain constraints (12c) of all the RRH are satisfied. Finally, the analog beamformer at the i-th RRH can be given by $\mathbf{F}_{\mathrm{RF}_i}^{\star}=[\tilde{\mathbf{f}}_{i,\mathcal{K}_i(1)},\ldots,\tilde{\mathbf{f}}_{i,\mathcal{K}_i(N_{\mathrm{RF}})}]$, in which $\mathcal{K}_i$ collects the user indices of non-zero $u_{i,k}$.

In the following, we propose a two-stage uplink training approach to facilitate the above user association and analog beamforming design. After that, the training results, and the design of user association and analog beamforming also provide as foundation for digital processing optimization.

In the first stage, the joint design of analog beamformer and user association is obtained at the BBU through beam training which is different from the training process in Sec. III-B. Particularly, all K users simultaneously transmit orthogonal pilot sequences $\boldsymbol{\varphi}_k^{(1)}\in\mathbb{C}^{\kappa_1\times 1}$ with transmit power $\rho_k=\frac{\rho}{\eta_k}$ at the q-th time block, where ρ is a base power and η_k is the weight coefficient of the k-th user. $\|\boldsymbol{\varphi}_k^{(1)}\|^2=1$ and $\kappa_1\geq K$. The i-th RRH employs a training beamforming $\mathbf{F}_{i,q}$ which contains the beam candidates selected from $\mathcal{F}$ to combine

the signals. Then, the received signal at the i-th RRH after combining can be expressed as

$$\mathbf{Y}_{i,q}^{(1)} = \sqrt{\kappa_1}\mathbf{F}_{i,q}^H \sum_{k=1}^{K} \sqrt{\rho_k}\mathbf{h}_{i,k}(\boldsymbol{\varphi}_k^{(1)})^H + \mathbf{F}_{i,q}^H \mathbf{Z}_{i,q}, \quad (47)$$

$\mathbf{Z}_{i,q} \in \mathbb{C}^{N_{\mathrm{t}} \times \kappa_1}$ denotes the noise matrix during the q-th time block. After Q time blocks transmission, the received signals at the i-th RRH, $i = 1, \ldots, M$, are stored as $\mathbf{Y}_i^{(1)} \triangleq [(\mathbf{Y}_{i,1}^{(1)})^H, \ldots, (\mathbf{Y}_{i,Q}^{(1)})^H]^H$ of dimension $N_{\mathrm{res}} \times \kappa_1$:

$$\begin{aligned}\mathbf{Y}_i^{(1)} &= \sqrt{Q\kappa_1\rho_k} \begin{bmatrix} \mathbf{F}_{i,1}^H \\ \vdots \\ \mathbf{F}_{i,Q}^H \end{bmatrix} \sum_{k=1}^{K} \mathbf{h}_{i,k}(\boldsymbol{\varphi}_k^{(1)})^H + \begin{bmatrix} \mathbf{F}_{i,1}^H \\ \vdots \\ \mathbf{F}_{i,Q}^H \end{bmatrix} \mathbf{Z}_i \\ &= \sqrt{Q\kappa_1\rho_k} \begin{bmatrix} \mathbf{f}_1^H \\ \vdots \\ \mathbf{f}_{N_{\mathrm{res}}}^H \end{bmatrix} \sum_{k=1}^{K} \mathbf{h}_{i,k}(\boldsymbol{\varphi}_k^{(1)})^H + \begin{bmatrix} \mathbf{f}_1^H \\ \vdots \\ \mathbf{f}_{N_{\mathrm{res}}}^H \end{bmatrix} \mathbf{Z}_i. \end{aligned} \quad (48)$$

By right multiplying $\frac{\boldsymbol{\varphi}_k^{(1)}}{\sqrt{Q\kappa_1\rho_k\eta_k}}$ to $\mathbf{Y}_i^{(1)}$, the beamforming gain of each beam candidate in $\mathcal{F}$ of the i-th RRH for the k-th user can be obtained as

$$\begin{aligned}\hat{\mathbf{h}}_{i,k} &= \frac{1}{\sqrt{Q\kappa_1\rho_k\eta_k}} \mathbf{Y}_i^{(1)} \boldsymbol{\varphi}_k^{(1)} \\ &= \begin{bmatrix} \mathbf{F}_{i,1}^H \\ \vdots \\ \mathbf{F}_{i,Q}^H \end{bmatrix} \sqrt{\frac{1}{\eta_k}} \mathbf{h}_{i,k} + \frac{1}{\sqrt{Q\kappa_1\rho}} \begin{bmatrix} \mathbf{F}_{i,1}^H \\ \vdots \\ \mathbf{F}_{i,Q}^H \end{bmatrix} \mathbf{Z}_i \boldsymbol{\varphi}_k^{(1)} \quad (49) \\ &= \begin{bmatrix} \mathbf{f}_1^H \\ \vdots \\ \mathbf{f}_{N_{\mathrm{res}}}^H \end{bmatrix} \sqrt{\frac{1}{\eta_k}} \mathbf{h}_{i,k} + \frac{1}{\sqrt{Q\kappa_1\rho}} \begin{bmatrix} \mathbf{f}_1^H \\ \vdots \\ \mathbf{f}_{N_{\mathrm{res}}}^H \end{bmatrix} \mathbf{Z}_i \boldsymbol{\varphi}_k^{(1)}, \quad (50) \end{aligned}$$

where the d-th element of $\hat{\mathbf{h}}_{i,k}$, $d = 1, \ldots, N_{\mathrm{res}}$, is the beamforming gain between the i-th RRH and the k-th user when the d-th analog beamformer in $\mathcal{F}$ is adopted. The obtained $\hat{\mathbf{h}}_{i,k}$, $\forall i, k$, are fed back to the BBU through fronthaul link for joint RRH-user pairing and analog beamforming designs.

Given the training results at the BBU, we first select the optimal analog beamformer based on $\hat{\mathbf{h}}_{i,k}$, $\forall i, k$, without considering user association. The index of the optimal beam which leads to the largest gain between the i-th RRH and the k-th user can be selected as

$$d_{i,k}^{\star} = \underset{d=1,\ldots,N_{\mathrm{res}}}{\operatorname{argmax}} |\hat{\mathbf{h}}_{i,k}(d)|^2, \quad \forall i, k. \quad (51)$$

The corresponding analog beamformer between the i-th RRH and k-th user is given by $\tilde{\mathbf{f}}_{i,k} = \mathcal{F}(d_{i,k}^{\star})$. Then, the RRH-user association can be determined based on previously introduced successive elimination scheme by substituting the term $\mathbf{h}_{i,k}^H \tilde{\mathbf{f}}_{i,k}$ in (43)-(46) with the training value $\hat{\mathbf{h}}_{i,k}(d_{i,k}^{\star})$, $\forall i, k$. After that, the BBU can determine the optimal analog beamforming matrix for each RRH as $\mathbf{F}_{\mathrm{RF}_i}^{\star} = [\tilde{\mathbf{f}}_{i,\mathcal{K}_i(1)}, \ldots, \tilde{\mathbf{f}}_{i,\mathcal{K}_i(N_{\mathrm{RF}})}]$, in which $\mathcal{K}_i = \{k : \text{if } u_{i,k} = 1\}$, and delivers the corresponding beam indices to each RRH through fronthaul link.

In the second stage, we aim to acquire the effective baseband channel between each RRH and each user, which is utilized for the design of digital signal processing. While the effective baseband channel can be estimated from $\hat{\mathbf{h}}_{i,k}$ from the first stage, it requires a large amount of signaling overhead in order to guarantee a low estimation error which is not efficient. We will provide a detailed explanation at the end of this subsection. Instead, we propose to acquire the effective baseband channel in the second training stage. Similar to the training procedure in Sec. III-B, users send orthogonal sequences $\boldsymbol{\varphi}_k^{(2)} \in \mathbb{C}^{\kappa_2 \times 1}$ with transmit power ρ_k, and each RRH adopts $\mathbf{F}_{\mathrm{RF}_i}^{\star}$ to combine the signals. The received signal at the i-th RRH can be written as

$$\mathbf{Y}_i^{(2)} = \sqrt{\kappa_2\rho_k}\mathbf{F}_{\mathrm{RF}_i}^{\star H} \sum_{k=1}^{K} \mathbf{h}_{i,k}(\boldsymbol{\varphi}_k^{(2)})^H + \mathbf{F}_{\mathrm{RF}_i}^{\star H} \mathbf{Z}_i. \quad (52)$$

The effective baseband channel from the i-th RRH to the k-th user is obtained by multiplying $\frac{\boldsymbol{\varphi}_k^{(2)}}{\sqrt{\kappa_2\rho_k}}$ to the right of $\mathbf{Y}_i^{(2)}$, which is expressed as

$$\begin{aligned}\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}} &= \frac{1}{\sqrt{\kappa_2\rho_k}} \mathbf{Y}_i^{(2)} \boldsymbol{\varphi}_k^{(2)} \\ &= \mathbf{F}_{\mathrm{RF}_i}^{\star H} \mathbf{h}_{i,k} + \frac{1}{\sqrt{\kappa_2\rho_k}} \mathbf{F}_{\mathrm{RF}_i}^{\star H} \mathbf{Z}_i \boldsymbol{\varphi}_k^{(2)}. \end{aligned} \quad (53)$$

Each RRH feeds the effective baseband channel $\{\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}\}_{i,k}$ back to the BBU for digital signal processing designs.

C. Digital Beamforming and Fronthaul Compression

Given the effective baseband channel $\{\tilde{\mathbf{h}}_{i,k}^{\mathrm{eff}}\}_{i,k}$ and user association $\{u_{i,k}\}_{i,k}$ obtained through uplink training, the digital beamforming and fronthaul compression design is based on

$$\begin{aligned}\mathrm{F3}: \max_{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}} \min_k \ & \frac{1}{\eta_k} \frac{|\bar{\mathbf{h}}_{k,k}^H \mathbf{f}_{\mathrm{BB}_k}|^2}{\zeta_k(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega})} \\ \text{s.t. Constraints (28b) and (28c)}, & \quad (54a)\end{aligned}$$

where $\zeta_k(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega})$ is defined in (30).

Regarding as the non-convex objective function, we still apply fractional programming to transform the sum-of-ratio term into a convex expression, which is given by

$$\begin{aligned}f_\alpha(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = \frac{2}{\sqrt{\eta_k}} \mathfrak{Re}\Big\{\alpha_k^* \bar{\mathbf{h}}_{k,k} \mathbf{f}_{\mathrm{BB}_k}\Big\} \\ - |\alpha_k|^2 \zeta_k(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}), \quad (55)\end{aligned}$$

in which $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_K]$ is introduced as an auxiliary variable to each user. In addition, the fronthaul constraint can be transformed as in (36). Finally, the max-min fair problem is reformulated as

$$\begin{aligned}\mathrm{F4}: \max_{\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}} \ & t \quad (56a) \\ \text{s.t. Constraints (28c) and (38b)}, & \\ f_\alpha(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) \geq t, \ & \forall k. \quad (56b)\end{aligned}$$

The above problem can be solved through an iterative procedure. Particularly, the optimal value of α_k, $k = 1, \ldots 1, K$, is given by setting $\frac{\partial f_\alpha}{\alpha_k} = 0$:

$$\alpha_k^{\star} = \frac{\bar{\mathbf{h}}_{k,k}^H \mathbf{f}_{\mathrm{BB}_k}}{\sqrt{\eta_k}\zeta_k(\mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Omega})}. \quad (57)$$

In addition, the optimal $\mathbf{\Sigma}_i$, $i = 1, \ldots, M$, is given by (37). Finally, the optimal value of $\{\mathbf{F}_{\mathrm{BB}}, \mathbf{\Omega}\}$ with fixed $\{\mathbf{\Sigma}, \boldsymbol{\alpha}\}$ can be obtained by solving F4 through CVX.

The total required signaling overhead is $\frac{N_{\mathrm{res}}\kappa_1}{N_{\mathrm{RF}}} + \kappa_2$, where κ_1 and κ_2 are the sequence lengths in the first and second training stages. We can also obtain that the overall computational complexity of max-min fairness problem is $O((K + MN_{\mathrm{RF}})^{3.5}T + N_{\mathrm{res}}KMN_{\mathrm{RF}}(\kappa_1^2 + \kappa_2^2))$.

As we mentioned before, from a mathematical perspective, this two-stage training scheme can be integrated into one stage, since we can estimate the effective baseband channel from $\hat{\mathbf{h}}_{i,k}$ obtained in the first stage. However, this may result in a higher signaling overhead. To be specific, if we integrate the total training procedure into one stage, the signaling overhead will be $\frac{N_{\mathrm{res}}\kappa}{N_{\mathrm{RF}}}$, where κ is the length of sequence sent by the RRH. From (53), we notice that the estimation error power for the two cases are in proportion to $\frac{\sigma_{\mathrm{u}}^2}{\kappa_2\rho_k}$ and $\frac{\sigma_{\mathrm{u}}^2}{\kappa\rho_k}$, respectively. Given a fixed transmit power ρ_k during training, when we increase κ_2 or κ by 1 in order to reduce the estimation error, the total overhead for the proposed two-stage training (given by $\frac{N_{\mathrm{res}}\kappa_1}{N_{\mathrm{RF}}} + \kappa_2$) is only increased by 1. While for the one-stage case, the total overhead (given by $\frac{N_{\mathrm{res}}\kappa}{N_{\mathrm{RF}}}$) is increased by a multiplication of $\frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}}$ (usually $\frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}} > 10$). Therefore, to guarantee a low estimation error, we divide the training into two stages.

V. Simulation Results

In this section, we provide numerical results to evaluate the performance of the proposed train-and-design strategies on a system setup of $K = 4$ single-antenna users and $M = 3$ RRHs. Each RRH is equipped with $N_{\mathrm{RF}} = 3$ RF chains and $N_{\mathrm{t}} = 16$ antennas [32]. We consider the half wavelength-spaced ULA model of the RRH antennas, where the AoDs follow uniform distributions in the range $\theta_k \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. We further assume all the RRHs have the same fronthaul link capacity $C_i = C$ and the same maximum transmit power $P_i = P$, $\forall i$. We further assume $\sigma_k = \sigma, \forall k$. In the simulations, we consider the fronthaul capacity of $C = 3$ and 5 b/s/Hz [32], respectively.

A. Weighted Sum-Rate Maximization Problem

For weighted sum-rate maximization problem, we assume that all users have the same weight coefficient, i.e. $\omega_k = 1$, $\forall k$, without loss of generality. The size of codebook is set to be $N_{\mathrm{res}} = 64$ [9], [10]. In addition, we define the uplink SNR during training as $\gamma = \frac{\rho}{\sigma_{\mathrm{u}}^2}$ and consider the cases of $\gamma = 10, 20$ dB [42], respectively.

Fig. 3 shows the achievable sum-rate versus the downlink SNR varying from -5 dB to 50 dB, which is defined as SNR $= \frac{P}{\sigma^2}$ [32]. The fronthaul capacities of $C = 3$ and 5 b/s/Hz are considered and represented as dash lines and solid lines, respectively. For comparison purpose, we include the simulation results of the following approaches:

1) Full-digital beamforming with perfect CSI (referred as "Full-digital"), which is served as a benchmark;
2) The WMMSE based hybrid beamforming scheme in [32] with perfect CSI (referred as "WMMSE");

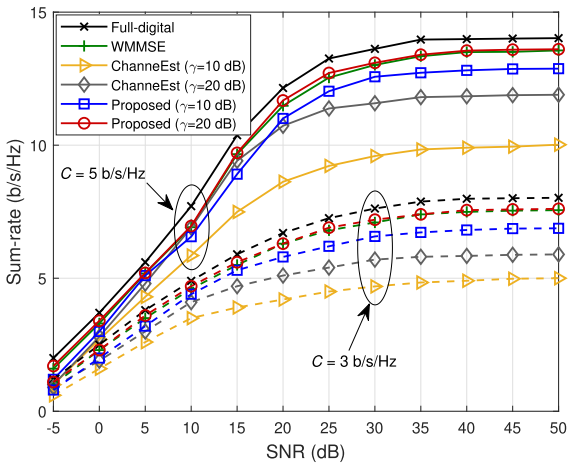


Fig. 3. Sum-rate versus SNR ($M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$).

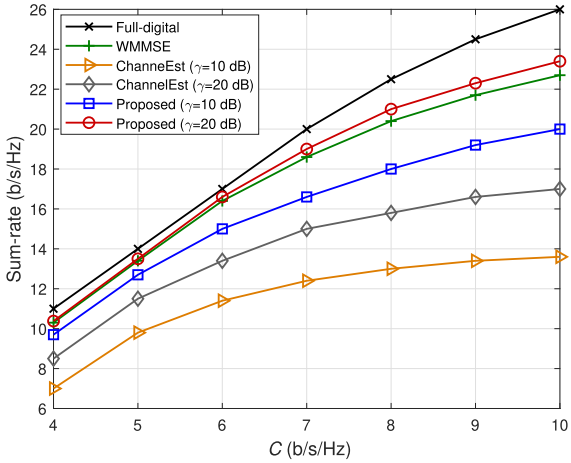


Fig. 4. Sum-rate versus C ($M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$, SNR $=$ 30 dB).

3) The channel estimation approach (referred as "ChannelEst"), which first acquires the CSI of each RRH-user pair using the algorithm introduced in [10], then design the hybrid beamforming and fronthaul compression using the WMMSE approach.

It can be observed from Fig. 3 that the proposed scheme outperforms the channel estimation based approach, which demonstrates the advantages of the proposed train-and-design solution. Although user association and analog beamforming are obtained without explicit CSI, the proposed algorithm also achieves satisfactory performance. Moreover, the propose scheme with $\gamma = 20$ dB even has better performance than the WMMSE based hybrid beamforming design with perfect CSI knowledge. This is due to the fact that benefiting from the strong sparsity and directionality of mmWave channels, the proposed analog beamformer design can significantly enhance the beamforming gain and consequently improve the estimation accuracy of effective baseband channel. Although WMMSE assumes perfect CSI, the relaxation of RF constraint results in a performance degradation of the analog beamforming design. While we propose to decompose the design in analog and digital domains, it can be observed that

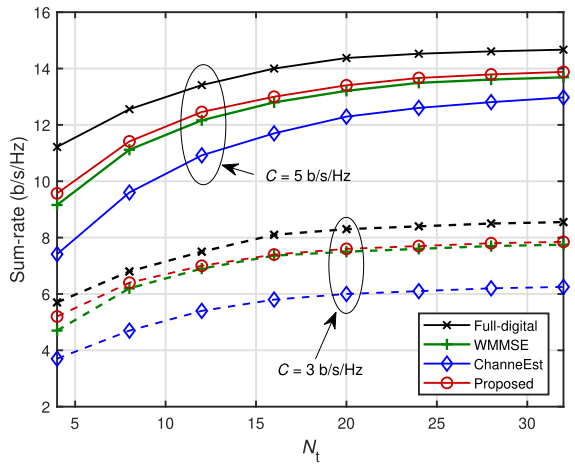


Fig. 5. Sum-rate versus N_{t} ($M = 3$, $N_{\mathrm{RF}} = 3$, $K = 4$, SNR $= 30$ dB, $\gamma = 20$ dB).

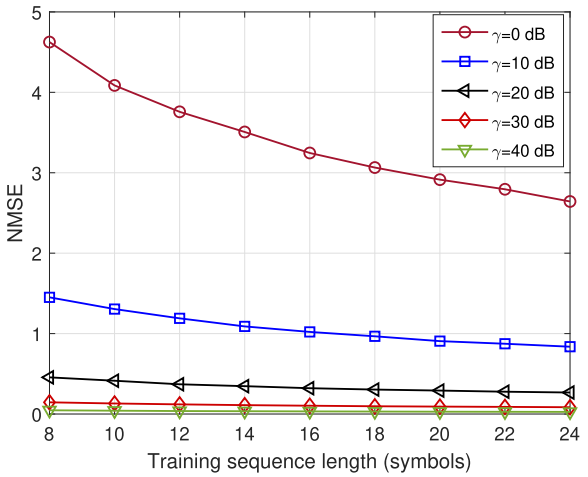
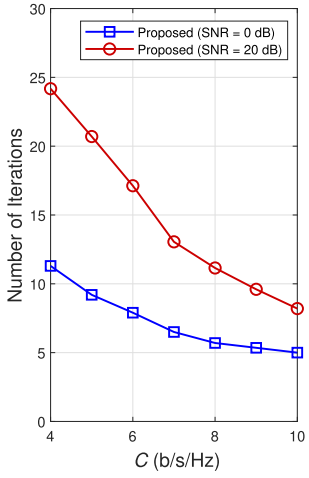
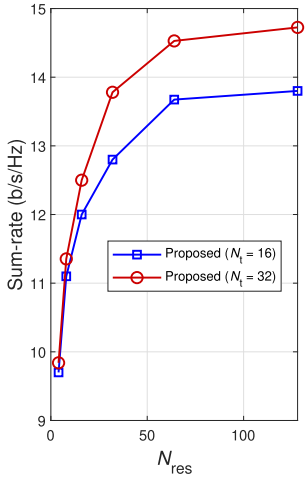


Fig. 7. NMSE of the effective baseband channel estimation versus the training sequence length τ_2.



(a) Number of iterations versus fronthaul capacity C.



(b) Sum-rate versus the codebook resolution N_{res}.

Fig. 6. The system setting is given by $M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$, $\gamma = 20$ dB.

the proposed scheme achieves performance close to the full-digital case, which justify the effectiveness of this sub-optimal design. In addition, it is shown that with increasing SNR, the achievable sum-rate will grow rapidly at low SNR while tend to become saturated at high SNR due to the limited fronthaul capacity.

In Fig. 4, we illustrate the sum-rate as a function of the fronthaul capacity C for the considered cell-free system, where the downlink SNR is fixed at 30 dB. We can see that our proposed design still exhibits better performance than the channel estimation based approach. Besides, it can be seen from the figure that the proposed approach and channel estimation based approaches exhibit relatively slow increase of sum-rate compared to the full-digital beamforming. This performance loss is caused by the effect of imperfect knowledge of CSI.

The achievable sum-rate versus the number of antennas is evaluated in Fig. 5. We can observe that the sum-rate increases as the number of antennas grows. Under high SNR condition, the performance will finally achieve saturation due to the limited capacity of fronthaul link.

Fig. 6a illustrates the number of iterations to achieve convergence versus the fronthaul capacity. The figure shows that the proposed algorithm converges within a few tens of iterations. In addition, as C increases, the fronthaul constraint becomes less strict and it requires less iterations for the proposed algorithm to reach convergence.

The impact of the codebook resolution is evaluated in Fig. 6b, where the resolution N_{res} is set between $8 \sim 128$. This can be realized by phase shifters of resolution $3 \sim 7$ bits. The results show that higher resolution of codebook can result in higher achievable sum-rate. We can also observe that a codebook of size 64 realized by 6-bit phase shifters is sufficient to accomplish a satisfactory performance.

Finally, we evaluate the estimation error of the effective baseband channel. We show the normalized mean square error (NMSE) of the effective baseband channel estimation which is defined as

$$NMSE = \frac{\sum_{i=1}^{M}\sum_{k=1}^{K}\mathbb{E}\left[||\mathbf{e}_{i,k}||^2\right]}{\sum_{i=1}^{M}\sum_{k=1}^{K}\mathbb{E}\left[||\mathbf{h}_{i,k}^{\mathrm{eff}}||^2\right]}. \tag{58}$$

In Fig. 7, we provide the NMSE performance under different training sequence length τ_2 and transmit power during the second training stage. We assume $\rho_k = \rho$, for $k = 1, \ldots, K$. We can observe that the NMSE will decrease as the transmit power or the training sequence length increases. When $\gamma \geq 10$ dB, the NMSE is sufficiently small and the proposed algorithm can achieve comparable performance to the full-digital scheme.

B. Max-Min Fairness Problem

In this subsection, we evaluate the performance of the proposed algorithm based on max-min fairness metric. For performance comparison, we include the full-digital beamforming based on full CSI as a performance benchmark. In addition, we include the max-min fair power allocation

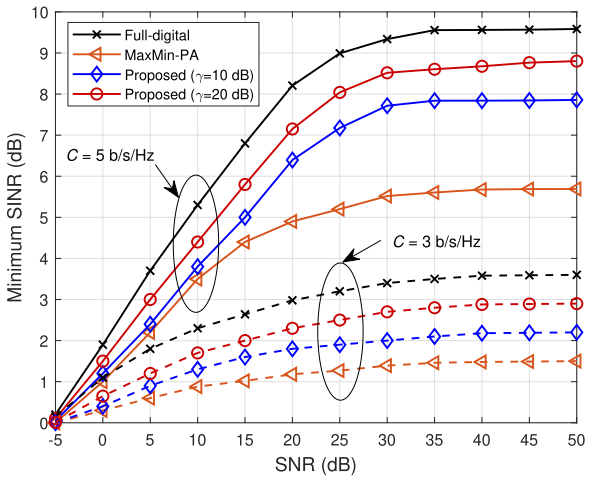


Fig. 8. Minimum SINR versus SNR ($M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$).

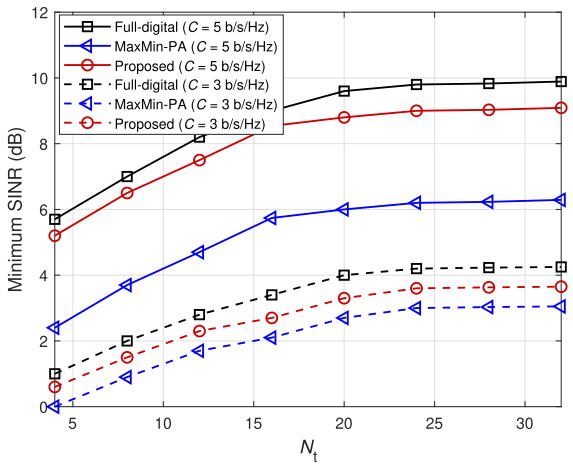


Fig. 10. Minimum SINR versus N_{t} ($M = 3$, $N_{\mathrm{RF}} = 3$, $K = 4$, SNR = 30 dB, $\gamma = 20$ dB).

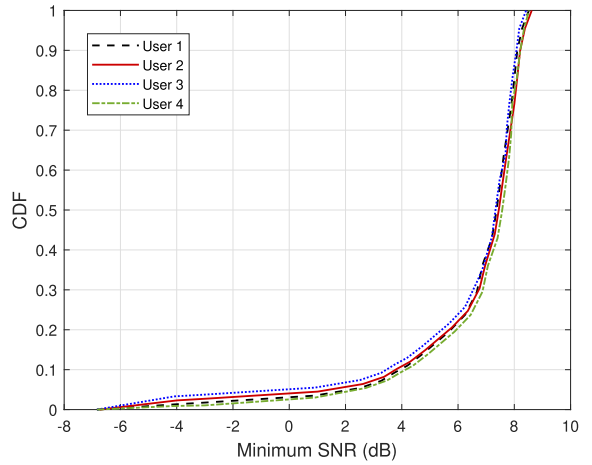


Fig. 9. CDF of SINR ($M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$, SNR = 30 dB, $C = 5$ b/s/Hz, $\gamma = 20$ dB).

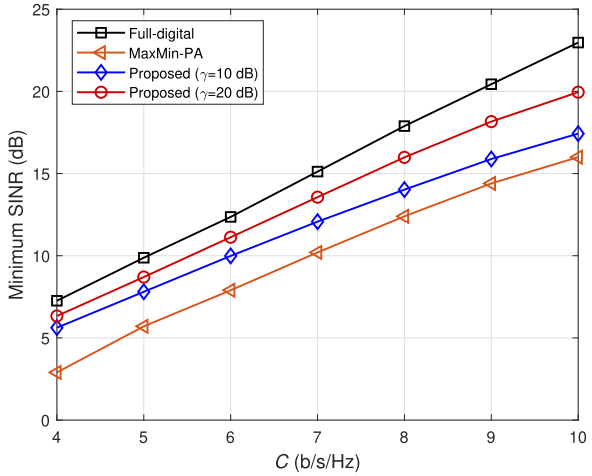


Fig. 11. Minimum SINR versus C ($M = 3$, $N_{\mathrm{RF}} = 3$, $N_{\mathrm{t}} = 16$, $K = 4$, SNR = 30 dB).

(MaxMin-PA) algorithm proposed in [17] with perfect CSI to serve as baseline.

In Fig. 8, we plot the minimum SINR among users versus SNR. The results show that the proposed algorithm outperforms the MaxMin-PA algorithm and can achieve comparable fairness to the full-digital scheme. The minimum SINR among users will reach saturation round SNR of 30 dB due to the fronthaul link capacity constraint, and not increase much with larger SNR.

Fig. 9 depicts cumulative distribution function (CDF) of SINR for different users under $C = 5$ bit/s/Hz. We can observe that different users exhibit similar CDF performance, which demonstrate that the proposed algorithm can well guarantee the fairness among users.

Then, in Fig. 10, we illustrate the minimum SINR as a function of the number of antennas. It can be seen from the figure that the minimum SINR improves with increasing number of antennas, and the proposed algorithm can always achieve comparable performance to the full-digital beamforming scheme and outperform the MaxMin-PA baseline. In addition, under high SNR condition, the finite-capacity of fronthaul link becomes the main obstacle to further improve the system performance.

Finally, the achievable minimum SINR among users with respect to the fronthaul link capacity C is illustrated in Fig 11, where the SNR is fixed at 30 dB. Similar conclusions can be drawn that the proposed approach has satisfactory performance.

C. Training Overhead

The total required signaling of the proposed approach is $\frac{N_{\mathrm{res}}}{N_{\mathrm{RF}}} + \tau_2$ for sum-rate maximization problem, and $\frac{N_{\mathrm{res}}\kappa_1}{N_{\mathrm{RF}}} + \kappa_2$ for max-min fairness problem, respectively. The resolution of the codebook is set to $N_{\mathrm{res}} = 64$. To support $K = 4$ users, κ_1 is usually given by $\kappa_1 = 4$. τ_2 (κ_2) is between $4 \sim 20$. The number of required training symbols is in the range of $30 \sim 50$ for sum-rate maximization problem and $90 \sim 110$ for max-min fairness problem, respectively. Considering the channel coherence time for mmWave systems is around 1 ms and on the order of hundreds of symbols, this training overhead is acceptable [44], [45]. For comparison, if we use

the compressed channel estimation [10] to estimate the channel vector of each user successively, the required training overhead is around 20 per user and 80 for 4 users. We can notice that our proposed approach for sum-rate maximization problem has a much lower training overhead compared with compressed channel estimation. While the training overhead of compressed channel estimation is smaller than the max-min fairness case, the approach in [10] requires feedback from users during the training, which will make the system more complicated and increase the training overhead.

VI. Conclusion

In this paper, we studied the user association, hybrid beamforming, and fronthaul compression designs for C-RAN enabled cell-free mmWave MIMO systems without full CSI acquisition. A train-and-design was introduced to address this problem. In particular, we proposed to jointly design the user association and analog beamforming with the assistance of a two-stage uplink training. Then, digital beamforming and fronthaul compression were optimized based on the training results. Two performance metrics were considered in this paper, i.e., weighted sum-rate maximization and max-min fairness. Simulation results demonstrated the advantages of the proposed solution, which can achieve better performance than the channel estimation based design and has close performance to the full-digital beamforming with perfect CSI.

References

[1] Z. Wang, R. Liu, H. Li, M. Li, and Q. Liu, "Hybrid beamforming design for C-RAN based mmWave cell-free systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Victoria, BC, Canada, Oct. 2020, pp. 1–5.

[2] T. S. Rappaport, G. R. Maccartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[3] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[6] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, 2nd Quart., 2018.

[7] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] Z. Wang, M. Li, Q. Liu, and A. L. Swindlehurst, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

[9] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[11] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016.

[12] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[13] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[14] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, 2021.

[15] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.

[16] E. Bjornson and L. Sanguinetti, "Cell-free versus cellular massive MIMO: What processing is needed for cell-free to win?" in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.

[17] G. Femenias and F. Riera-Palou, "Reduced-complexity downlink cell-free mmWave massive MIMO systems with fronthaul constraints," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, A Coruña, Spain, Sep. 2019, pp. 1–5.

[18] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44596–44612, 2019.

[19] A. Checko et al., "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

[20] X. Wang et al., "Virtualized cloud radio access network for 5G transport," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 202–209, Sep. 2017.

[21] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcicy, and R. W. Heath, Jr., "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1340–1345.

[22] D. H. N. Nguyen, L. B. Le, and T. Le-Ngoc, "Hybrid MMSE precoding for mmWave multiuser MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[23] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[24] Z. Wang, M. Li, X. Tian, and Q. Liu, "Iterative hybrid precoder and combiner design for mmWave multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1581–1584, Jul. 2017.

[25] S. Sun, T. S. Rappaport, M. Shafi, and H. Tataria, "Analytical framework of hybrid beamforming in multi-cell millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7528–7543, Sep. 2018.

[26] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[27] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.

[28] J. Kim, H.-W. Lee, and S. Chong, "Virtual cell beamforming in cooperative networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1126–1138, Jun. 2014.

[29] C.-S. Lee and W.-H. Chung, "Hybrid RF-baseband precoding for cooperative multiuser massive MIMO systems with limited RF chains," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1575–1589, Apr. 2017.

[30] S.-H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Oct. 2014.

[31] Y. Zhou and W. Yu, "Fronthaul compression and transmit beamforming optimization for multi-antenna uplink C-RAN," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4138–4151, Aug. 2016.

[32] J. Kim, S.-H. Park, O. Simeone, I. Lee, and S. S. Shitz, "Joint design of fronthauling and hybrid beamforming for downlink C-RAN systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4423–4434, Jun. 2019.

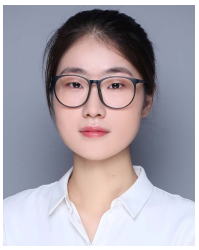
[33] A. Liu, X. Chen, W. Yu, V. K. N. Lau, and M.-J. Zhao, "Two-timescale hybrid compression and forward for massive MIMO aided C-RAN," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2484–2498, May 2019.

[34] Q. Hou, S. He, Y. Huang, H. Wang, and L. Yang, "Joint user scheduling and hybrid precoding design for MIMO C-RAN," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–6.

[35] T. C. Mai, H. Q. Ngo, and T. Q. Duong, "Downlink spectral efficiency of cell-free massive MIMO systems with multi-antenna users," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4803–4815, Aug. 2020.
[36] A. Zhou, J. Wu, E. G. Larsson, and P. Fan, "Max-min optimal beamforming for cell-free massive MIMO," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2344–2348, Oct. 2020.
[37] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max–min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.
[38] Q.-D. Vu, K.-G. Nguyen, and M. Juntti, "Weighted max–min fairness for C-RAN multicasting under limited fronthaul constraints," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1534–1548, Apr. 2018.
[39] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[40] K. Shen and W. Yu, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, May 2018.
[41] F. A. Potra and S. J. Wright, "Interior-point methods," *J. Comput. Appl. Math.*, vol. 124, nos. 1–2, pp. 281–302, Dec. 2000.
[42] A. Manoj and A. P. Kannu, "Channel estimation strategies for multi-user mmWave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5678–5690, Nov. 2018.
[43] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *J. Commun. Netw.*, vol. 18, no. 2, pp. 135–149, Apr. 2016.
[44] M. Kokshoorn, H. Chen, Y. Li, and B. Vucetic, "Beam-on-graph: Simultaneous channel estimation for mmWave MIMO systems with multiple users," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 2931–2946, Jul. 2018.
[45] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 2909–2913.

Zihuan Wang (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electronics information engineering from the Dalian University of Technology, Dalian, China, in 2017 and 2020, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The University of British Columbia (UBC), Vancouver, BC, Canada. Her research interests include machine learning and optimization for wireless networks, and integrated sensing and communications. She has received UBC's Four Year Fellowship (2020–2024) and the Graduate Support Initiative Award from the Faculty of Applied Science at UBC (2021–2022). She received the Best Paper Award at the IEEE ICC 2022.

Ming Li (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo (SUNY-Buffalo), Buffalo, NY, USA, in 2005 and 2010, respectively. From January 2011 to August 2013, he was a Post-Doctoral Research Associate with the Department of Electrical Engineering, SUNY-Buffalo. From August 2013 to June 2014, he joined Qualcomm Technologies Inc., as a Senior Engineer. Since June 2014, he has been with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, where he is currently a Professor. His current research interests include general areas of communication theory and signal processing with applications to integrated sensing and communication, reconfigurable intelligent surfaces, mmWave communications, massive MIMO systems, and secure wireless communications. He has served as a TPC chair/member of various international flagship conferences. He was a recipient of an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS.



Rang Liu (Graduate Student Member, IEEE) received the B.S. degree in electronics information engineering from the Dalian University of Technology, Dalian, China, in 2018, where she is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. Her current research interests include signal processing, massive MIMO systems, reconfigurable intelligent surfaces, and integrated sensing and communication. She was a recipient of the National Scholarship in 2020.

Qian Liu (Member, IEEE) received the B.S. and M.S. degrees from the Dalian University of Technology, Dalian, China, in 2006 and 2009, respectively, and the Ph.D. degree from The State University of New York at Buffalo (SUNY-Buffalo), Buffalo, NY, USA, in 2013. She was a Post-Doctoral Fellow at the Ubiquitous Multimedia Laboratory, SUNY-Buffalo, from 2013 to 2015. She is currently an Associate Professor at the Department of Computer Science and Technology, Dalian University of Technology. Her current research interests include haptic communications and signal processing, wireless multimedia communications, and haptic-oriented human–computer interaction. She was an Alexander von Humboldt Fellow at the Chair of Media Technology and the Chair of Communication Networks, Technical University of Munich, from 2016 to 2017. She provides services to the IEEE Haptic Codec Task Group as a Secretary for standardizing haptic codecs in the tactile internet. She also served as the Technical Program Co-Chair for the 2017 IEEE Haptic Audio Visual Environments and Games (HAVE 2017), HAVE 2018, AsiaHaptics 2020, and AsiaHaptics 2022.