Enabling Cell-Free Massive MIMO Systems With Wireless Millimeter Wave Fronthaul

Umut Demirhan[ID] and Ahmed Alkhateeb[ID]

Abstract—Cell-free massive MIMO systems have promising data rate and coverage gains. These systems, however, typically rely on fiber based fronthaul for the communication between the central processing unit and the distributed access points (APs), which increases the infrastructure cost and installation complexity. To address these challenges, this paper proposes two architectures for cell-free massive MIMO systems based on wireless fronthaul that is operating at a higher-band compared to the access links. These dual-band architectures ensure high data rate fronthaul while reducing the infrastructure cost and enhancing the deployment flexibility and adaptability. To investigate the achievable data rates with the proposed architectures, we formulate the end-to-end data rate optimization problem accounting for the various practical aspects of the fronthaul and access links. Then, we develop a low-complexity yet efficient joint beamforming and resource allocation solution for the proposed architectures based on user-centric AP grouping. With this solution, we show that the proposed architectures can achieve comparable data rates to those obtained with optical fiber-based fronthaul under realistic assumptions on the fronthaul bandwidth, hardware constraints, and deployment scenarios. This highlights a promising path for realizing the cell-free massive MIMO gains in practice while reducing the infrastructure and deployment overhead.

Index Terms—Cell free massive MIMO, millimeter wave fronthaul, beamforming.

I. INTRODUCTION

CAN **we realize the cell-free massive MIMO gains with wireless fronthaul?** This paper attempts to answer this important question. Cell-free massive MIMO systems are promising uniform coverage and high data rate gains, even for scenarios with very dense users [1], [2]. A critical challenge, however, with the current cell-free massive MIMO systems is its reliance on optical fiber-based fronthaul for distributing data and synchronization signals from the central processing unit (CPU) to the access points (APs). This highly increases its infrastructure cost and installation time, and limits the flexibility and scalability of the cell-free massive MIMO deployment. To address these challenges, the paper proposes an alternative cell-free massive MIMO architecture

with higher-band fronthaul, e.g., a millimeter wave (mmWave) or terahertz (THz) fronthaul for a sub-6GHz cell-free massive MIMO system. The use of higher frequency band fronthaul has two key advantages: (i) The large bandwidth available at the higher frequency bands provides a high data rate fronthaul and (ii) the higher-band fronthaul signals (lower wavelength) have high capability to synchronize the sub-6GHz AP transceivers. For this proposed architecture, the paper investigates whether it can achieve comparable achieve rates to those obtained by classical fiber-fronthaul based cell-free massive MIMO systems.

A. Related Works

Motivated by its potential to increase the data rate and manage the multi-user interference, the idea of having distributed antennas simultaneously serving the same users have been previously investigated in the network multiple-input multiple-output (MIMO) [3], distributed MIMO [4], and coordinated multipoint with joint transmission (CoMP-JT) [5], [6]. These earlier approaches did not scale well with the increasing number of users and large number of access points (APs), due to the channel estimation and feedback overhead. To overcome this limitation, in [1], [7]–[10], the authors investigated cell-free massive MIMO, where reciprocity of the time-division duplexing (TDD) is leveraged to estimate the downlink channels directly from the joint uplink pilot transmissions. In the proposed cell-free optimization framework, the imperfection of channel state information (CSI) was taken into account, and only the long-term channel coefficients were utilized for the power allocation. Despite the interesting data rate and coverage gains, the realization of cell-free massive MIMO networks suffers from the limitations and high infrastructure cost of the CPU-AP fiber links. To reduce cost, cheaper wire-link alternatives, with limited capacity, were considered. For example, [11] investigated cell-free performance with quantized fronthaul transmissions. However, the approaches in [11]–[14] did not eliminate the wired connection requirement, which is still associated with high infrastructure cost, high installation time, and limited deployment flexibility. To clarify, the joint consideration of the wireless fronthaul and cell-free massive MIMO networks has not been studied in the literature.

Another relevant line of work is presented in the different network architectures that adopted wireless fronthaul/backhaul such as small-cells [15], [16], ultra dense networks [17], heterogeneous networks [18] and the cloud radio access networks (C-RAN) [19]–[21]. In [15]–[18], wireless fronthaul was adopted to show the potential advantages of the

mmWave backhaul. This was however, limited to scenarios with no cooperation between basestations. In C-RAN architectures, multiple (normally a few) base stations coordinate to serve the users. For example, in [19], the time-frequency resources of an OFDMA based C-RAN system consisted of a few base stations were optimized to maximize the weighted sum-rate of large number of users. The solution in [19] assumed the fronthaul transmission to a single base station at a time. To overcome this limitation, the work in [20] proposed a multicast beamforming for the downlink ultra dense C-RAN, and designed user-centric clusters of the base stations for the transmissions. The work in [19], [20], however, was limited to architectures with a few number of base stations. This is partially due to the large channel estimation/feedback overhead associated with the solutions in [19], [20], which limit them from scaling to large number of antennas. This motivates the research for new approaches to enable the potential gains of cell-free massive MIMO systems with the adoption of the wireless fronthaul.

B. Contributions

In this work, we propose a new architecture for cell-free massive MIMO systems where the fronthaul is implemented using high-frequency (e.g., mmWave) wireless links to serve access links at lower frequency (e.g., at a sub-6GHz band). The proposed architecture has two key motivations: (i) The larger available bandwidth at the high frequency band ensures high data rates for the fronthaul and (ii) the small wavelength of the high-frequency signals provides high synchronization accuracy for the access points that are operating at a lower frequency band. Further, the wireless fronthaul enables a modular, flexible and scalable architecture with low infrastructure cost and low installation time. With all these potential gains, an important question is whether this architecture is capable of achieving comparable data rate gains to those achieved with optical fiber-based cell-free massive MIMO systems? To answer this question, we develop an efficient communication scheme, analyze its performance, and draw important insights about the proposed architecture and data rate optimization approaches. The main contributions of the paper can be summarized as follows:

- **Proposing an efficient architecture for cell-free massive MIMO systems based on higher-frequency wireless fronthaul**. The proposed architecture has the potential of ensuring higher data rate fronthaul and high synchronization accuracy for the distributed APs. Further, it requires low infrastructure cost and low installation time and provides interesting flexibility and adaptability gains for cell-free massive MIMO systems.

- **Developing an efficient communication model for the wireless fronthaul based cell free massive MIMO system**. This model assigns a group of APs to each user and optimizes the multicast beamforming at the central processing unit to simultaneously serve the AP group of each user. The adopted system model accounts for the practical constraints on the higher-frequency band (mmWave) beamforming architectures [22].

Specifically, the CPU applies analog beamforming. Given the spatial multiplexing constraints of the analog beamforming, we use time-division multiple-access (TDMA) to multicast the message of each user-centric group.

- **Formulating the end-to-end data rate maximization problem for the proposed architecture and communication model**. The optimization problem takes the wireless fronthaul, access channels, and AP grouping into account. In particular, it aims to determine the user-centric AP group selection, fronthaul beamforming vectors, TDMA schedule, and AP power coefficients to maximize the end-to-end data rates.

- **Developing near-optimal end-to-end data rate maximization solution for the proposed architecture**. The solution adopts an iterative group selection algorithm, which is coupled with the fronthaul and access channel data rate maximization sub-problems. Specifically, in each iteration, the group size and AP selection are determined based on the channel estimates and then the fronthaul/access rates are optimized for the given groups.

- **Extending the architecture and developed solutions to the mixed wireless/wired fronthaul case.** A mixed-fronthaul architecture with wire-connected AP clusters (for example through a radio stripe [2]) is proposed. In this architecture, only the leader AP in each wire-connected cluster has a wireless fronthaul link with the CPU. The proposed data rate maximization solutions are generalized to account for the mixed-fronthaul.

The proposed solutions are extensively evaluated using numerical simulations which draw important insights into the performance of the proposed cell-free architectures. Based on the results, the high-frequency wireless fronthaul can provide sufficient data-rates for the cell-free massive MIMO by taking advantage of the larger bandwidth availability. With the fully wireless architecture, it is possible to achieve data rates with only $10 - 20\%$ degradation compared to the fiber fronthaul based cell-free massive MIMO architectures. Further, the mixed-fronthaul architecture significantly reduces the bandwidth requirements and improves the data-rates. In our simulations, the mixed-fronthaul architecture enabled data rates very similar to the fiber fronthaul based solutions with reasonable fronthaul bandwidth requirements.

Notation: We use the following notation throughout this paper: $\mathbf{a}$ is a vector, a is a scalar, $\mathcal{A}$ is a set of scalars, and $\boldsymbol{\mathcal{A}}$ is a set of vectors or sets. $|\mathbf{a}|$ is the absolute value of $\mathbf{a}$, whereas $\mathbf{a}^T$, $\mathbf{a}^H$ are its transpose and Hermitian. $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathcal{U}[a, b]$ is a uniform random variable in $[a, b]$. $\mathbb{E}[\cdot]$ is used to denote expectation.

II. WIRELESS-FRONTHAULED DUAL-BAND CELL-FREE MASSIVE MIMO: THE KEY IDEA

Cell-free massive MIMO is a promising enabler for the high data rate and coverage requirements in future wireless communications systems. Current cell-free massive MIMO
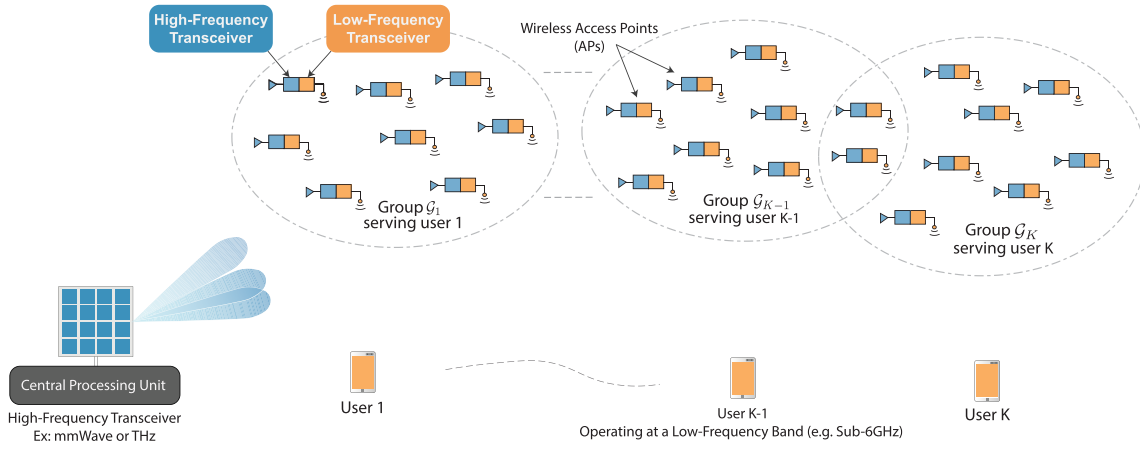
Fig. 1.   An illustration of the proposed architecture where a mmWave CPU provides fronthaul to the different sets of user-centric AP groups via different beamforming vectors. The wireless APs jointly serve their users through a cell-free massive MIMO access channel at sub-6GHz.

architectures, however, assume that all the antennas are connected to the central processing unit with wired fronthaul, that is typically optical fiber [23]. This complicates the deployment process of these systems and increases its cost and installation time. To overcome these challenges, we propose an alternative architecture for cell-free massive MIMO systems that relies on a wireless fronthaul operating at a much higher frequency band compared to the access band, as depicted in Fig. 1. For example, a mmWave fronthaul channel with sub-6GHz access channels or a terahertz fronthaul supporting mmWave access channel. Thanks to the wireless fronthaul, **the proposed architecture is modular, flexible, scalable, with low cost and installation time**. In particular, replacing the wired-fronthaul with a wireless connection provides modularity and flexibility in the deployment. This way, the costs and installation time can be reduced as the APs can be directly deployed, potentially without any (or with minimal) engineering assistance. Consequently, this allows easier deployment of the APs and potential deployment scalability of the system.[1] While the wireless fronthaul provides these advantages, it may have challenges for the data rates and synchronization, which motivates the use of a higher-frequency fronthaul. The use of a higher frequency band in the fronthaul compared to the access band has the following advantages (for the rest of the paper, and for ease of exposition, we will assume a mmWave fronthaul and a sub-6GHz access channel):

- **High Data-Rate Fronthaul:** The availability of the large bandwidth at the mmWave frequency band enables the central processing unit (which is equipped with a mmWave transceiver) to support high data-rate fronthaul. This way, the crucial large bandwidth availability can be utilized for maintaining the high data rate gains of

cell-free massive MIMO systems, as will be shown in Section VII.

- **High Synchronization Accuracy:** Using a mmWave fronthaul (where signals have a small wavelength) to synchronize the sub-6GHz APs (where signals have a much higher wavelength) has the potential of ensuring precise clock synchronization among the APs, which is essential for the operation of the cell-free massive MIMO systems. Specifically, the synchronization of the APs can be achieved over the mmWave fronthaul with the aid of the CPU via master-slave type algorithms (e.g., [25]) or via network centric solutions (e.g., [26]). The APs use two clock signals, one for the mmWave fronthaul and the other for the sub-6GHz access channel. The APs can synchronize their mmWave clocks and time by referring to the CPU. These signals can then be utilized for more accurate sub-6GHz synchronization by converting the mmWave (smaller wavelength) signal to a large wavelength clock.[2]

Further, with regards to the mmWave fronthaul, we would like to highlight two points: (i) The cost of mmWave transceivers may currently be high. However, it is expected to decrease over time with the mass adoption and production. Further, in Section Section VI, we propose an extension of the architecture that requires much less number of mmWave transceivers. (ii) The potential coverage limitations of the mmWave fronthaul can be reduced with the employment of a large antenna array at the CPU, as will be considered in the following sections. In general, we envision that the proposed architecture will be promising for the scenarios where the CPU is serving APs distributed within a few hundred meters.

Now, with the interesting gains of having wireless fronthaul based cell-free massive MIMO systems, and while mmWave

---

[1]Note that a scalability notion for cell-free massive MIMO was formally defined in [24], which aims to keep the complexity of the channel estimation, signal processing for beamforming, fronthaul signaling and power control optimization limited for asymptotically increasing number of users. It is important to note that this is different than our use of scalability, that is for the deployment of the APs.

[2]For instance, if there are two mmWave clocks that are synchronized with a Δf frequency offset between them, the sub-6GHz clocks obtained from the these (or from the same source) can translate into a smaller frequency difference, possibly at the ratio of the carrier frequencies (e.g., for 30 GHz and 3 GHz carrier frequencies, it could be possible to obtain the clocks with frequency difference $\frac{\Delta f}{10}$). A similar gain is expected if a THz-based fronthaul is leveraged to synchronize a mmWave access channel.

fronthaul may have relatively high data rates compared to other wireless solutions, it is important to answer the question: **Can this wireless fronthaul based architecture achieve the same data rates of the fiber fronthaul based cell-free massive MIMO architectures?** In this paper, we try to answer this question. Towards this objective, we first describe the system model of the proposed mmWave fronthaul based cell-free massive MIMO architecture in Section III and formulate the end-to-end (from the CPU to the users) achievable rate optimization problem in Section IV. Then, we develop an efficient transmission strategy for the proposed architecture in Section V and discuss some important extensions in Section VI. Finally, the achievable rate of the proposed architecture and comparisons with the fiber-based solutions are provided in Section VII.

III. SYSTEM MODEL

We consider the distributed massive MIMO system in Fig. 1, where a CPU communicates with M wireless-APs over a high-frequency (e.g., mmWave) wireless fronthaul and the wireless-APs serve K user equipment (UEs) over a low-frequency (e.g., sub-6GHz) channel. We will use $\mathcal{M} = \{1, \ldots, M\}$ and $\mathcal{K} = \{1, \ldots, K\}$ to denote the sets of M APs and K UEs. For the rest of the paper, we will refer to the downlink channel from the CPU to the APs as the *fronthaul channel* and the downlink channel from APs to the UEs as the *access channel*. Further, for the ease of exposition and without loss of generality, we will assume that the fronthaul channel is operating over a mmWave band, with a bandwidth B^{fh}, and the access channel is adopting a sub-6GHz band, with a bandwidth B^{ac}. The algorithms and results of the paper, however, can be applied to other dual-band architectures, such as terahertz fronthaul with a mmWave access channel.

To beamform the signal to the APs, the CPU employs an antenna array of N elements, while the APs and UEs are, for simplicity, assumed to have single antennas. We consider a user-centric grouping approach where the message of UE k is jointly transmitted by a subset of APs, $\mathcal{G}_k \subseteq \mathcal{M}$. We denote the set of the user-centric groups as $\mathcal{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_K\}$. Note that different groups may include the same APs since multiple APs are utilized in the transmission to every UE. For that, we will also define $\mathcal{U}_m \subseteq \mathcal{M}$ for $m \in \mathcal{M}$ as the set of users that are being served by the AP m. For the distances between the CPU and m-th AP and between the m-th AP and k-th user, we will use d_m^{fh} and d_{mk}^{ac} to denote them. It is important to mention here that we do not assume any knowledge about the positions or distances between the CPU, APs, or the users.

At the Fronthaul Link: The mmWave CPU adopts TDMA to serve the K user-centric groups. In each TDMA slot, the CPU beamforms the signal towards the APs that serve one user. The duration of the TDMA slot allocated for serving APs of UE k is denoted by t_k. Let $\mathbf{h}_m \in \mathbb{C}^N$ denote the channel between the CPU and the mth AP. If the message intended for user k is represented by $q_k^{\mathrm{fh}} \in \mathbb{C}$, with $\mathbb{E}[|q_k^{\mathrm{fh}}|^2] = 1$, then the received signal at the AP m can be written as

$$y_m^{\mathrm{fh}} = \sqrt{\rho^{\mathrm{fh}}} \mathbf{h}_m^H \mathbf{f}_k q_k^{\mathrm{fh}} + w_m^{\mathrm{fh}}, \tag{1}$$

where ρ^{fh} is the normalized fronthaul transmission power and $w_m^{\mathrm{fh}} \sim \mathcal{CN}(0, 1)$ is the receive noise at the mth AP. The vector $\mathbf{f}_k \in \mathbb{C}^N$ is the CPU beamforming vector intended to focus the signal to the APs that serve user k. To satisfy the practical mmWave hardware constraints, we assume that the CPU adopts analog-only beamforming implemented by a network of quantized phase shifters [22]. This means that the beamforming vector $\mathbf{f}_k$ can only be selected from a certain set of vectors, that we define by the codebook $\mathcal{F}$. If each phase shifter has q bits, i.e., 2^q possible phase shift values defined by the set $\mathcal{Q} \in \{0, \frac{\pi}{2^q}, \frac{2\pi}{2^q}, \ldots, \frac{(2^q-1)\pi}{2^q}\}$, then we can write

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{N}}[e^{j\phi_1}, \ldots, e^{j\phi_N}]^T : \phi_n \in \mathcal{Q}, \forall n \in \{1, \ldots, N\} \right\}, \tag{2}$$

and $\mathbf{f}_k \in \mathcal{F}$. We assume that the channels between the CPU and APs are available at the CPU. This is motivated by the stationarity of the CPU and APs and the channel reciprocity assumption.

At the Access Link: Each AP decodes the CPU signal received by its mmWave receiver and prepares it for the transmission over the sub-6GHz access channel. Similar to [1], the APs transmit the weighted sum of the users' messages, where each message is multiplied by a different beamforming and power control coefficient. Since each AP m is assumed to contribute in serving a set of users $\mathcal{U}_m$, the transmitted signal from the mth AP, x_m, can be written as

$$x_m = \sqrt{\rho^{\mathrm{ac}}} \sum_{k \in \mathcal{U}_m} \sqrt{p_{mk}} f_{mk}^{\mathrm{ac}} q_k^{\mathrm{ac}}, \tag{3}$$

where ρ^{ac} is the shared APs coefficient for the transmit power coefficient of the APs, f_{mk}^{ac} and p_{mk} denote the beamforming and power control coefficients of the m-th AP for the k-th UE, and q_k^{ac} represents the intended message for the k-th UE which satisfies $\mathbb{E}\left[|q_k^{\mathrm{ac}}|^2\right] = 1$. After all the APs prepare their messages, they transmit them to their users. Note that all the APs are assumed to maintain sufficient clock synchronization at the sub-6GHz access channel, which is further facilitated by the adoption of mmWave-based synchronization, as briefly highlighted in Section II. Now, if g_{mk} denotes the access channel coefficient between the m-th AP and k-th UE, the received signal at the UE k can be expressed as

$$y_k^{\mathrm{ac}} = \sum_{m=1}^{M} x_m g_{mk} + w_k, \tag{4}$$

where $w_k \sim \mathcal{CN}(0, 1)$ is the receive noise at user k.

Channel Models: For the fronthaul channel, we adopt a geometric channel model [10], [27]. In this model, there are V propagation paths between the CPU and each AP. We denote the large- and small-scale fading and azimuth angle of departure of the path v by $\beta_{m,v}^{\mathrm{fh}}$, $\alpha_{m,v}^{\mathrm{fh}}$ and $\theta_{m,v}$, respectively. With this notation, we can write the channel as

$$\mathbf{h}_m = \sqrt{N} \sum_{v=1}^{V} \sqrt{\beta_{m,v}^{\mathrm{fh}}} \alpha_{m,v}^{\mathrm{fh}} \mathbf{a}(\theta_{m,v}), \tag{5}$$

where $\mathbf{a}(\cdot) \in \mathbb{C}^N$ is the array response vector function. For the access channel, we define the channel coefficient between

AP m and UE k as $g_{mk} = \sqrt{\beta_{mk}^{\mathrm{ac}}} \alpha_{mk}^{\mathrm{ac}}$, where β_{mk}^{ac} and $\alpha_{mk}^{\mathrm{ac}}$ represent the large- and small-scale fading coefficients.

IV. PROBLEM FORMULATION: END-TO-END RATE OPTIMIZATION

In this section, we formulate the optimization problem of the end-to-end achievable rate of the proposed wireless fronthaul based cell-free massive MIMO system. We first present the achievable rates of the access and fronthaul channels. Then, we present a transmission schedule, and formulate the end-to-end rate optimization problem with this schedule. In the following, we omit the superscripts denoting fronthaul and access channel variables as they are easily distinguished from the context.

A. Achievable Rate of the Fronthaul

Considering the system model in Section III, the achievable rate of AP m, that is part of the kth user group $\mathcal{G}_k$, can be written as

$$R_m^{\mathrm{fh}}(\mathbf{f}_k) = \log\left(1 + \rho^{\mathrm{fh}} \left|\mathbf{h}_m^H \mathbf{f}_k\right|^2\right), \qquad (6)$$

where $\mathbf{f}_k$ is the beamforming vector used by the CPU to serve the set of APs in the group $\mathcal{G}_k$. Now, note that the message of the kth user is simultaneously transmitted to all the APs in $\mathcal{G}_k$, and that all the APs in the group should finish receiving the message before they start transmitting it to the user. To incorporate that, we define the effective rate of group k as the minimum rate of the APs in the $\mathcal{G}_k$'s group. Mathematically, we write

$$R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k) = \min_{m \in \mathcal{G}_k} \left\{R_m^{\mathrm{fh}}(\mathbf{f}_k)\right\}. \qquad (7)$$

Moreover, due to the TDMA schedule of the AP groups, the group rates will be further scaled by the TDMA time fractions. Let t_k denote the fraction of TDMA time allocated to group $\mathcal{G}_k$. Then, the time-scaled fronthaul rate of group k can be written as $t_k R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k)$. We note that these time fractions satisfy the constraints $0 < t_k < 1$ and $\sum_{k=1}^K t_k = 1$.

B. Achievable Rate of the Cell-Free Massive MIMO Access Channel

The access channel adopts the main assumptions of the cell-free massive MIMO architectures in [1]. Specifically, we assume the following: (i) The APs synchronously serve the UEs without any cell boundaries, i.e., each AP can serve any UE. (ii) Time-division duplexing (TDD) is adopted for the transmissions, which facilitates the estimation of the downlink access channel coefficients through the uplink pilots. (iii) Only the large-scale fading coefficients are available at the CPU for joint power allocation. Further, we assume that the APs adopt conjugate beamforming for the downlink transmission to the users. More specifically, the uplink pilots are used to estimate the uplink channels (which are also used to construct the downlink channels leveraging channel reciprocity). Then, the information about the large-scale fading coefficients are frequently transmitted to the CPU. The CPU uses this large-scale fading information to determine the access channel

power coefficients. The APs adopt this power allocation while jointly serving their users. Next, we first describe the adopted channel estimation procedure in detail before formulating the achievable rates of the access link.

Channel Estimation: In the adopted protocol, the UEs transmit orthogonal pilot sequences of length L_p, $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K \in \mathbb{C}^{L_p}$, simultaneously to be received by all the APs. If ρ_t denote the power level selected for the pilot transmissions, then the received signal at AP m can be written as

$$\mathbf{y}_m^t = \sqrt{\rho_t L_p} \sum_{k=1}^K g_{mk} \boldsymbol{\psi}_k + \mathbf{n}_m, \qquad (8)$$

where $\mathbf{n}_m \sim \mathcal{CN}(0, \mathbf{I}_{L_p})$ is the receive noise. With this received signal, the MMSE estimator for g_{mk} can be given by

$$\hat{g}_{mk} = \frac{\sqrt{\rho_t L_p} \beta_{mk}}{1 + \rho_t L_p \beta_{mk}} \boldsymbol{\psi}_k^H \mathbf{y}_m^t. \qquad (9)$$

We define the channel estimation error $\tilde{g}_{mk} = g_{mk} - \hat{g}_{mk}$, and note that the estimation $\hat{g}_{mk}$ and the error $\tilde{g}_{mk}$ are uncorrelated thanks to the estimator. Therefore, the distributions of the estimated channel coefficients and the error can be written as $\hat{g}_{mk} \sim \mathcal{CN}(0, \hat{\beta}_{mk})$, $\tilde{g}_{mk} \sim \mathcal{CN}(0, \beta_{mk} - \hat{\beta}_{mk})$ with the variance of the estimator is defined as

$$\hat{\beta}_{mk} = \frac{\rho_t L_p \beta_{mk}^2}{1 + \rho_t L_p \beta_{mk}}. \qquad (10)$$

Achievable Rate: With the adopted conjugate beamforming, the coefficients f_{mn}^{ac} in the received signal equations (3)-(4) can be replaced by $\hat{g}_{mn}^*$. Further, with the described changes on the signal model, the capacity lower bound for the UEs given in [8] becomes valid, and we can express the achievable rate of user k as

$$R_k^{\mathrm{ac}} = \log_2\left(1 + \mathrm{SINR}_k\right), \qquad (11)$$

$$\mathrm{SINR}_k = \frac{\rho^{\mathrm{ac}} \left(\sum_{m=1}^M \sqrt{p_{mk}} \hat{\beta}_{mk}\right)^2}{\rho^{\mathrm{ac}} \sum_{m=1}^M \beta_{mk} \sum_{k'=1}^K p_{mk'} \hat{\beta}_{mk'} + 1}, \qquad (12)$$

where we note here that the power coefficient p_{mk} for the AP m and UE k is set to 0 if the AP is not in the group of that UE, i.e., if $m \notin \mathcal{G}_k$. These power coefficients for each AP m also satisfy

$$\sum_{k=1}^K p_{mk} \hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M}, \qquad (13)$$

which captures the total power constraint of AP m.

C. Transmission Schedule

To achieve the high data rates with proposed transmission scheme, i.e., cell-free massive MIMO access channel and TDMA based fronthaul transmission scheme, a fronthaul transmission schedule needs to be designed. On one hand, the access channel requires the message of each UE to be available at all of the APs in its groups. On the other hand, the fronthaul transmissions are proposed in a TDMA manner, and a set of full transmissions needs to be completed before
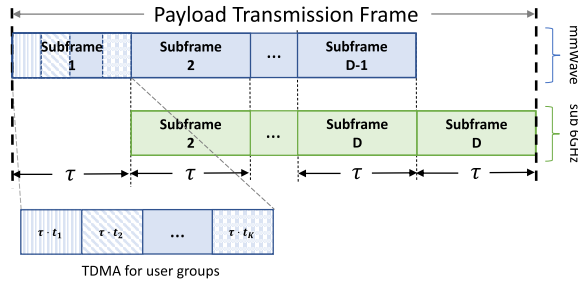
Fig. 2. The timing of the fronthaul and access channel transmissions with the frame structure.

the access channel transmissions. To that end, we propose a frame/subframe structure that allows almost concurrent data rates. Specifically, we consider a frame of downlink data transmission period within the coherence time, that is split into D subframes. Let the duration of this subframe be τ. Then, in the first subframe of a frame, only the fronthaul transmission is carried out. Then, in the following subframes, the APs transmit the data received in the previous subframe, while receiving the data for the access channel transmissions in the next subframe. With large number of subframes and smaller subframe durations, almost concurrent transmission data rates can be achieved. We illustrate the adopted transmission schedule in Fig. 2. Further, with this model, the additional transmission delay of the system can be determined as τ, as the access channel transmissions follow the fronthaul transmissions from the previous subframe. Although this may introduce some delay, the small values can be achieved with a careful fronthaul communication design. Next, we formulate the end-to-end data rate problem with of the proposed system.

D. End-to-End Achievable Rate Optimization

Based on the presented the transmission schedule and achievable rates for the fronthaul and access channels, we now derive the end-to-end achievable rate of the system. With the adopted system model in Section III and the schedule given in Section IV-C, the access channel rate of UE k during the payload transmission frame can be written as $\frac{D-1}{D}B^{\mathrm{ac}}R_k^{\mathrm{ac}}$, and the fronthaul rate of UE k can be written as $\frac{D-1}{D}t_k B^{\mathrm{fh}}R_k^{\mathrm{fh}}$. As the AP operation is essentially a relaying operation between the fronthaul and access channels, the minimum of these rates provides the end-to-end achievable rate.[3] By dropping $\frac{D-1}{D}$ as $D \to \infty$, we express end-to-end rate of UE k by

$$R_k = \min\left\{B^{\mathrm{ac}}R_k^{\mathrm{ac}},\ t_k B^{\mathrm{fh}}R_k^{\mathrm{fh}}(\mathbf{f}_k)\right\}, \tag{14}$$

Note that the end-to-end channel rate accounts for the fronthaul and access bandwidths, which could be significant since the fronthaul bandwidth in our mmWave-based fronthaul is expected to be much larger than the bandwidth of the sub-6GHz access channel. Now, to optimize the end-to-end rate, we adopt the following formulation of the joint max-min fair

[3]This objective is commonly adopted in the relay data rate optimization (e.g., [28]).

rate optimization problem

$$\max_{\mathcal{G},\{\mathbf{f}_k\},\{t_k\},\{p_{mk}\}}\ \min_k\ R_k \tag{15a}$$

$$\text{s.t.}\ \sum_{k\in\mathcal{K}}p_{mk}\hat{\beta}_{mk} \leq 1,\quad \forall m\in\mathcal{M} \tag{15b}$$

$$p_{mk}=0,\quad \forall m\notin\mathcal{G}_k,\ \forall k\in\mathcal{K} \tag{15c}$$

$$0 < t_k < 1,\quad \forall k\in\mathcal{K} \tag{15d}$$

$$\sum_{k\in\mathcal{K}}t_k = 1 \tag{15e}$$

$$\mathbf{f}_k \in \mathcal{F},\quad \forall k\in\mathcal{K} \tag{15f}$$

which aims to jointly optimize the AP groups, the fronthaul beamforming vectors, the fronthaul time allocation for different groups, and the access channel power coefficients. The problem is non-convex and challenging, especially due to the AP grouping and the fronthaul analog-only beamforming. It is worth mentioning here that the AP grouping and analog beamforming can be optimally designed via an exhaustive over all the possible groups and candidate beam codewords, but this will require prohibitive complexity. To reduce this complexity, we propose an iterative sub-optimal solution in the following section.

V. PROPOSED SOLUTION

In this section, we develop a suboptimal yet efficient solution for the end-to-end data rate maximization problem of the proposed cell-free massive MIMO architecture. We will then show in Section VII that this developed solution (for the wireless mmWave-fronthauled cell-free massive MIMO architecture) achieves close performance to the upper bound which is given by the fiber-fronthaul based cell-free massive MIMO architecture. To start, we note that the optimization problem in (15) can be written in the equivalent form

$$\max_{\mathcal{G},\{\mathbf{f}_k\},\{p_{mk}\},\{t_k\}}\ \min\ \left\{B^{\mathrm{ac}}R_k^{\mathrm{ac}},\ t_k B^{\mathrm{fh}}R_k^{\mathrm{fh}}(\mathbf{f}_k)\right\}_{k=1}^K$$

$$\text{s.t.}\quad (15\mathrm{b}) - (15\mathrm{f}) \tag{16}$$

where the objective is to maximize the minimum of all the fronthaul and access channel rates. Before making an attempt for the solution, we highlight the following remark.

Remark 1: For the given system model and the max-min fairness problem defined in (16), the separate optimization of the access channel and fronthaul variables depend on the grouping variable $\mathcal{G}$. Nevertheless, for a fixed grouping $\mathcal{G}$, only the objective function (but not the constraints) retains the variables of both the fronthaul ($\mathbf{f}_k$ and t_k) and access channel (p_{mk}), i.e., each constraint affects either the fronthaul channel or the access channel. Therefore, for a given grouping, a two step approach can be developed with the access channel and fronthaul optimization steps to obtain an optimal solution.

Motivated by Remark 1, instead of the joint optimization of the fronthaul and access rates, we first consider a given grouping structure, and design two separate problems that maximize the fronthaul and access rates: (i) The access channel power coefficients are optimized without any consideration of the fronthaul, and (ii) the beamforming vectors for each group and TDMA time fractions are determined to

maximize the fronthaul rate in a fair manner. As this process is conditioned on the grouping structure, an optimization of this grouping is required in order to maximize the end-to-end data rates. For that, we propose an iterative heuristic algorithm for the AP grouping where each iteration involves solving two sub-problems for the fronthaul and access rates. The next three subsections present the details of this proposed approach.

A. Access Channel Rate Optimization

As briefly described, for a given AP grouping selection, we first optimize the power coefficients for the access channel, without any fronthaul limitations. This approach allows us to allocate the power of the APs over the access channel, in a similar way to the standard approaches in the cell-free massive MIMO literature. To formulate the access channel optimization problem, we only keep the access channel related terms and constraints of the original problem defined in (16), and write

$$\max_{\{p_{mk}\}} \min_{k} \quad B^{\mathrm{ac}} R_k^{\mathrm{ac}} \tag{17a}$$

$$\mathrm{s.t.} \quad \sum_{k \in \mathcal{K}} p_{mk} \hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M} \tag{17b}$$

$$p_{mk} = 0, \quad \forall m \notin \mathcal{G}_k, \ \forall k \in \mathcal{K}, \tag{17c}$$

where two further simplifications can be applied to the objective. First, since B^{ac} is a positive constant multiplied with a function of the variables, it can be removed. Second, the objective can be re-written in terms of the SINR values, instead of the rates of the form $\log(1 + \mathrm{SINR})$. This re-formulation does not change the optimal power coefficients, since $\log(1 + \mathrm{SINR})$ is a non-decreasing function of the SINR values. Thus, we simplify (17) and write the grouping sensitive max-min SINR optimization for the access channel as

$$\max_{\{p_{mk}\}} \min_{k} \quad \mathrm{SINR}_k \tag{18a}$$

$$\mathrm{s.t.} \quad \sum_{k=1}^{K} p_{mk} \hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M} \tag{18b}$$

$$p_{mk} = 0, \quad \forall m \notin \mathcal{G}_k, \ \forall k \in \mathcal{K}, \tag{18c}$$

which is in the same form with the power allocation problem of the standard cell-free massive MIMO [1] with the addition of the grouping constraint in (18c). In fact, when all the APs transmit to all UEs, i.e., $\mathcal{G}_k = \mathcal{M} \ \forall k \in \mathcal{K}$, (18c) does not provide any constraints, and the presented optimization of the access channel becomes directly equivalent to the power allocation of cell-free massive MIMO. For the solution of (18), we first elaborate on the additional grouping constraint (18c), which allows an AP to transmit only to the UEs whose groups include that AP, by restricting the power allocated for the other UEs to 0. It is a linear equality constraint which does not affect the convexity of the problem. Hence, for the problem (18), the same solution to the optimal power-allocation of the standard cell-free massive MIMO given in [1] can be applied. Further, different sub-optimal solutions with lower complexity and feedback (e.g., [2], [8], [24]) can be applied instead of the optimal solution of the problem.

B. Fronthaul Channel Rate Optimization

In this subsection, we define an optimization problem to maximize the fronthaul rates of the distributed AP groups in a fair manner. First, by starting with the original formulation given in (16), we write a max-min fair fronthaul optimization problem by eliminating the access channel related constraints and objectives, to get

$$\max_{\{t_k\}\{\mathbf{f}_k\}} \min_{k} \quad t_k R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k) \tag{19a}$$

$$\mathrm{s.t.} \quad \sum_{k \in \mathcal{K}} t_k = 1 \tag{19b}$$

$$t_k \geq 0, \quad \forall k \in \mathcal{K} \tag{19c}$$

$$\mathbf{f}_k \in \mathcal{F}, \quad \forall k \in \mathcal{K}. \tag{19d}$$

For this problem in (19), and given the group selections $\{\mathcal{G}_k\}$, we note that the optimal solution can be obtained by first optimizing the rates of all the groups, $R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k), \forall k$, and then optimization the TDMA time allocation. This is because any set of TDMA time fractions, $\{t_k\}$, does not affect the optimization of the beamforming vectors. Therefore, we will first optimize the beamforming vectors to optimize the group rates individually. Then, we can determine the TDMA time fractions for each group to optimize the time-scaled group rates, as described in the remaining part of this subsection.

Beamforming Optimization: The beamforming vector of each group needs to be optimized to maximize the group rate. Since this rate is determined by the minimum rate of the APs in the group, we can formulate the beamforming optimization problem of any group $\mathcal{G}_k$ as follows

$$\mathbf{f}_k^{\star} = \arg\max_{\mathbf{f}_k} \ \left(R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k) = \min_{m \in \mathcal{G}_k} \left\{ R_m^{\mathrm{fh}}(\mathbf{f}_k) \right\} \right)$$

$$\mathrm{s.t.} \quad \mathbf{f}_k \in \mathcal{F}, \tag{20}$$

which coincides with the well-studied multicast beamforming problem with analog phase-shifters. The optimal solution to the problem can be obtained by exhaustive search which has the complexity increasing exponentially with number of antennas, i.e., $\mathcal{O}(|\mathcal{Q}|^N) = \mathcal{O}(2^{qN})$. To reduce this complexity, several methods have been developed in the literature [29], [30]. In the simulation results in Section VII, we adopt the suboptimal iterative solution proposed in [29] due to its low complexity and good performance. With the optimized beamforming vectors, the optimal time fractions of the TDMA can be determined. Next, we will present two approaches to solve this TDMA time fraction optimization problem.

TDMA Optimization. Approach 1: Given the solution of the problem (20), $\{\mathbf{f}_k^*\}$, the fronthaul rate of each group k becomes a constant denoted by $R^{\mathrm{fh}}(\mathcal{G}_k) = R^{\mathrm{fh}}(\mathcal{G}_k, \mathbf{f}_k^*)$. With these rate constants, we can simplify the fronthaul rate maximization problem given in (19) to be

$$\max_{\{t_k\}} \min_{k} \ t_k R^{\mathrm{fh}}(\mathcal{G}_k) \ \mathrm{s.t.} \sum_{k \in \mathcal{K}} t_k = 1, \ t_k \geq 0 \ \forall k \in \mathcal{K}. \tag{21}$$

As the problem is linear, the optimal solution is obtained at the equality of the time scaled fronthaul rates, i.e., $t_k R_k^{\mathrm{fh}} = t'_k R_{k'}^{\mathrm{fh}}$ $\forall k, k' \in \mathcal{K}$. Therefore, the optimal sum of the time scaled

group rates can be given by

$$R^{\text{fh}}(\boldsymbol{\mathcal{G}}) = \text{HM}\{R^{\text{fh}}(\mathcal{G}_k)\}, \tag{22}$$

where the details of the proof is provided in Appendix A.

Finally, with the obtained access channel and fronthaul rates, we can take the minimum of them to provide an efficient solution to (16) for a given grouping structure.

The provided solution so far only aims to maximize the minimum of the rates, which is also the objective of (16). However, the optimal solution is not necessarily unique, which motivates the search within this set of solutions for one that could further optimize other objectives. In particular, without decreasing the minimum of the end-to-end rates of the given groups, we may still be able to increase the data rates of the other groups by further optimizing the fronthaul TDMA time allocations of these groups. To detail, the solution given in (22) attempts to maximize the minimum of the fronthaul group rates without accounting for the optimized access channel rates. The user rates, however, depend on both the access and fronthaul rates. Therefore, we may be able to further increase the data rates of the users by taking into account the access rates as a restricting constraint. Based on that, we define the following problem which attempts to optimize the fronthaul rate of each group to meet the access rate of the group in a fair manner.

TDMA Optimization. Approach 2: For a fair allocation, the time-scaled rate of the groups should be equal, unless any of the constraints are met. Therefore, the time allocated to each group should be inversely proportional to their rates. To achieve this, we utilize the weighted logarithm function for the objective, i.e., $\sum_k w_k \log(t_k)$, which allows the resources to be allocated fairly, proportional to the weights, w_k. We select the weights as $w_k = \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}$, to allocate the time fractions inversely proportional to the fronthaul rates of the groups, resulting in equal fronthaul rates. In addition, we upper-bound the fronthaul rate of a user by the access channel rate of that user. With this upper-bound, the total time fractions do not necessarily meet the summation equality, $\sum_{k \in \mathcal{K}} t_k = 1$. Hence, we relax the condition by $\sum_{k \in \mathcal{K}} t_k \leq 1$, and write the problem for the time allocation of the TDMA, that maximize the end-to-end rate in a fair manner, as follows

$$\max_{\{t_k\}} \quad \sum_{k \in \mathcal{K}} \frac{1}{R^{\text{fh}}(\mathcal{G}_k)} \log t_k \tag{23a}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} t_k \leq 1 \tag{23b}$$

$$t_k \leq \frac{B^{\text{ac}} R_k^{\text{ac}}}{B^{\text{fh}} R^{\text{fh}}(\mathcal{G}_k)}, \quad \forall k \in \mathcal{K} \tag{23c}$$

$$t_k \geq 0, \quad \forall k \in \mathcal{K}. \tag{23d}$$

We provide the proof of this problem in Appendix B. The resulting solution takes a similar form with the water-filling solution with upper-bounds. Specifically, the solution can be obtained by finding the maximum individual fronthaul rate (water level), $\eta > 0$, that satisfy the following

inequality

$$\sum_{k \in \mathcal{K}} t_k^*(\eta) \leq 1, \text{ with } t_k^*(\eta) = \min\left\{\frac{B^{\text{ac}} R_k^{\text{ac}}}{B^{\text{fh}} R^{\text{fh}}(\mathcal{G}_k)}, \frac{\eta}{R^{\text{fh}}(\mathcal{G}_k)}\right\}. \tag{24}$$

For the optimal η value, η^*, we can obtain the optimal time fractions $t_k^* = t_k^*(\eta^*)$. The given solution allocates the fronthaul rates equally among UEs, until the satisfaction of individual access channel rates or the use of the total time. In the special case of all the fronthaul rates being smaller than the access channel rates, it allocates the rates equally, coinciding with (22).

C. End-to-End Optimization Through Iterative Group Selection

With the proposed fixed-group end-to-end data rate maximization, the efficient selection of the groups becomes crucial in achieving high data rates. However, the group selection problem is a combinatorial problem that has high complexity given the large number of APs and UEs. In particular, there are 2^M possible selection of groups for each UE, leading to a total of 2^{MK} distinct selections for the K users. Moreover, each selection of the groups needs to be utilized with the proposed end-to-end data rate optimization which results in prohibitive optimization complexity for practical systems. Hence, an efficient design for the group selection is required. To this end, we propose a low-complexity yet efficient solution that is motivated by understanding the end-to-end achievable rate optimization problem and the proposed cell-free massive MIMO architecture. Next, we present the proposed approach which has two key points, namely the group selection and the group size optimization.

Group Selection: We adopt the following group selection approach: The group of each user is selected as the G APs with the maximum channel gains. Here, it is important to clarify two points: (i) The group size G is assumed to be fixed for the sake of simplicity and low-complexity solution. Note that the impact of this constraint is expected to be marginal given the high density of the APs. (ii) The selection of the APs with the maximum channel gains will likely lead to a set of APs that are close to each other, which will lead to efficient beamforming design via more focused beams. Next, we mathematically describe the proposed approach. Let $\{\beta_{1k}^{(o)}, \ldots, \beta_{Mk}^{(o)}\}$ define the ordered set of channel coefficients between user k and the M APs. This set adopts a descending order, i.e., the channel coefficients satisfy $\beta_{mk}^{(o)} \geq \beta_{m'k}^{(o)}$ for any $m \leq m'$. In addition, to formalize the mapping between the original channel coefficients sets $\{\beta_{1k}, \ldots, \beta_{Mk}\}$ and the ordered channel coefficients set $\{\beta_{1k}^{(o)}, \ldots, \beta_{Mk}^{(o)}\}$, we define the permutation $\varsigma_k(.)$ such that $\beta_{m'k}^{(o)} = \beta_{mk}$ if $\varsigma_k(m') = m$, i.e., if the mth AP has the m'th highest channel gain. Using these definitions, for a given group size G, the group of UE k can be determined as $\mathcal{G}_k = \bigcup_{m'=1}^{G} \varsigma_k(m')$, which simply selects the G APs with the best channel gains. With this approach, the group selection is reduced to the selection of the parameter G, and it needs to be determined carefully.
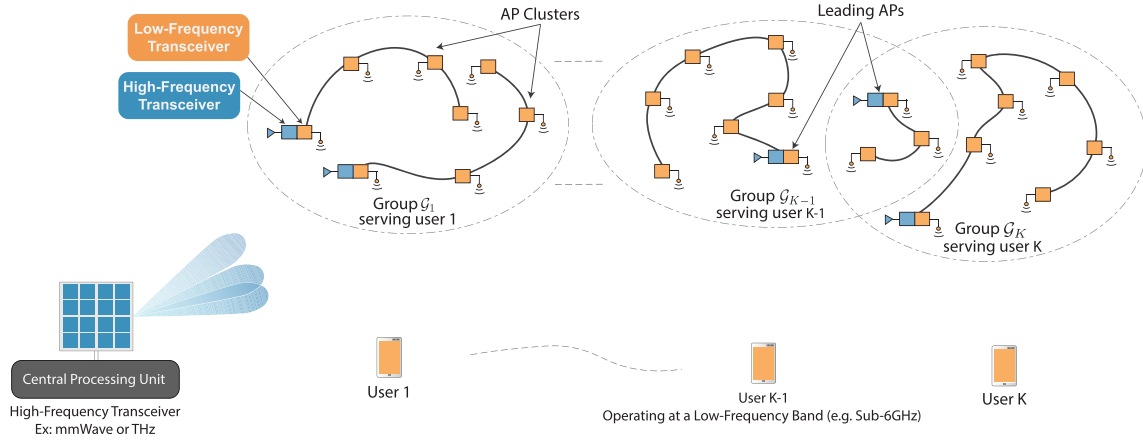
Fig. 3. An illustration of the proposed mixed-fronthaul architecture. In this architecture, each subset of APs are connected together via an optical fiber forming a *cluster*. Each cluster has one leading AP that is responsible for the wireless fronthaul communication with the central processing unit. For the access link, each user is served by a group of AP clusters. This architecture has the potential of reducing the fronthaul cost/complexity while maintaining the data rate/coverage gains as illustrated in Section VII.

Group Size Optimization: The number of APs per group, G, needs to be optimized to maximize the end-to-end rate. Here, it is interesting to highlight the following trade-off: A small group size may lead to more optimized CPU-APs beamforming design and hence high fronthaul rates. At the same time, it may also result in low APs-user beamforming gain and achievable access channel rates. Therefore, to select G, one possible approach is the trial of different group size values from a pre-determined interval (for a given AP structure). Another approach is to start from a certain group size G and then increase/decrease the group size depending on the relation between the access and fronthaul rates. More specifically, if the sum fronthaul rate is larger than the sum access channel rate, i.e., $R^{\mathrm{fh}} \geq R^{\mathrm{acc}}$, the group size is increased. Otherwise, if $R^{\mathrm{fh}} < R^{\mathrm{acc}}$, the group size is decreased. Through the iterations of these updates, the group value can be locally optimized.[4]

VI. MIXED-FRONTHAUL CELL-FREE MASSIVE MIMO ARCHITECTURE

In the previous sections, we proposed and designed a cell-free massive MIMO architecture with wireless higher-band fronthaul. While this architecture is flexible, scalable, and has the potential of reducing the installation cost and time of cell-free massive MIMO systems, it has a few drawbacks: (i) Adding a dual-band relay to each remote wireless AP adds an extra cost to the system and (ii) separately powering the distributed APs may require additional infrastructure cost. To address these points while maintaining the same promising gains, we propose a modified architecture with partially connected APs. In this section, we first describe the proposed architecture in Section VI-A and then briefly present its rate optimization approach in Section VI-B. The performance evaluation of this architecture is then detailed in Section VII-C.

[4]Although the iterative algorithm is applied through the sum data rates, it still achieves some fairness between the users since the group size is fixed.

A. Description of the Proposed Mixed-Fronthul Architecture

To reduce the cost of the distributed APs while maintaining the potential deployment/operational gains of using wireless fronthaul, we propose the alternative cell-free massive MIMO architecture depicted in Fig. 3. In this architecture, every cluster/set of distributed APs is connected with wired connections, e.g., an optical fiber. Further, each set of APs includes one leading AP that has a wireless-fronthaul to the central unit. This leading AP will be responsible for transmitting/receiving the cluster data to/from the central unit. It is worth noting here that the connection between the APs in each cluster could be realized using radio stripes [2]. With this implementation, the proposed mixed-fronthaul cell-free massive MIMO architectures combines the gains of the radio stripes and the installation cost/flexibility gains of the higher-band wireless fronthaul, shaping a promising and practical solution for cell-free massive MIMO systems.

From an operation perspective, the proposed mixed-fronthaul architecture reduces the number of APs that are simultaneously using the wireless fronthaul, which relaxes the fronthaul requirements in terms of the equipment cost, the required bandwidth, and the beamforming design. Therefore, this mixed-fronthaul architecture has the potential of further improving the wireless fronthaul rates since better beams can be utilized by the central unit in serving the clusters leading APs. In terms of the end-to-end communication model, the main difference between the mixed-fronthaul architecture and the original proposed architecture in Section II is that all the APs in one cluster communicate with the central unit through one leading AP. Therefore, and to simplify the end-to-end rate optimization, we assume that all the APs in one cluster will serve the same user. This approach is similar to [31], where the user-centric groups for cell-free massive MIMO are selected from the set of APs connected to the same CPU rather than individual APs. If the number of APs in each cluster is large, however, it could be important to relax this constraint, i.e., to allow any user to be served by only a subset of the cluster APs. In the next subsection,

we elaborate more on the proposed rate-optimization approach.

B. Proposed Rate Optimization Approach

In this subsection, we extend the rate optimization solution presented in Section V to the modified cell-free massive MIMO architecture with mixed-fronthaul. First, we assume that all the APs in one cluster will be serving the same set of UEs. This is motivated by the negligible data transmission cost within the same cluster. Hence, the group of APs serving each user k will be determined by selecting the set of clusters (instead of selecting individual APs in the separate-AP architecture in Section II-Section V). Formally, assume that there are L distinct AP clusters, such that $\mathcal{C}_l \subseteq \mathcal{M}$, $\forall l \in \{1, \dots, L\}$ with $\bigcup_{l=1}^{L} \mathcal{C}_l = \mathcal{M}$ and $\mathcal{C}_l \cap \mathcal{C}_{l'} = \emptyset$, $\forall l \neq l'$. Then, the APs group of user k, which we denote as $\mathcal{G}_k^{\text{cluster}}$ to emphasize that they are selected from the available AP *clusters* will satisfy $\mathcal{G}_k^{\text{cluster}} \subseteq \{1, \dots, L\}$. Further, we let a_l represent the leading AP of the lth cluster. Next, we revisit the fronthaul/access rate optimization and the group selection methodology from the lens of the mixed fronthaul architecture.

Fronthaul Rate Optimization: The initial fronthaul rate optimization was defined in (19), which aims to optimize the time and beams of the group $\mathcal{G}_k$. With the mixed architecture, the CPU only needs to transmit the data to the leading APs of the clusters of the k-th UE. To simplify the description, we define the user-centric fronthaul groups $\mathcal{G}_k^{\text{fh}} = \{a_l : l \in \mathcal{G}_k^{\text{cluster}}\}$ of the leading APs of the clusters. This notation allows a straightforward use of the fronthaul formulation in (19), only by adopting $\mathcal{G}_k^{\text{fh}}$ instead of $\mathcal{G}_k$. With this fronthaul group definition, the rest of the fronthaul optimization methodology can be carried out in the same form.

Access Rate Optimization: Similar to the fronthaul, we let $\mathcal{G}_k^{\text{ac}}$ denote the set of APs serving each user k over the access channel. This can be described as $\mathcal{G}_k^{\text{ac}} = \{\mathcal{C}_l : l \in \mathcal{G}_k^{\text{cluster}}\}$. With this notation, we can utilize the original access channel formulation in (17) and the adopted rate optimization approach, by replacing the AP groups $\{\mathcal{G}_k\}$ with the access channel groups $\{\mathcal{G}_k^{\text{ac}}\}$.

Group Selection: For the selection of the groups, as the user groups are defined in terms of the clusters, a metric for each cluster is needed. For this purpose, we consider the sum of the channel gains of the APs of the clusters, i.e., $\bar{\beta}_{lk} = \sum_{m \in \mathcal{C}_l} \beta_{mk}$, and select the G clusters with the maximum value as the clusters of a UE, $\mathcal{G}_k^{\text{cluster}}$. After determining the active clusters of UEs, the obtained groups of clusters can be converted to the fronthaul and access channel groups of the APs by earlier definitions. Then, the access channel and fronthaul optimization problems can be utilized over the access channel and fronthaul groups, completing an iteration of end-to-end data rate maximization. The described process will evaluated in the numerical results section to show the performance of connected APs.

VII. Numerical Results

In this section, we evaluate the performance of the proposed cell-free massive MIMO architectures and the developed rate

Parameter	Fronthaul	Access Channel
Frequency ($f_c^{\text{fh}}, f_c^{\text{ac}}$)	28 GHz	3.5 GHz
Bandwidth (B)	240 MHz	20 MHz
Power ($\tilde{\rho}^{\text{fh}}, \tilde{\rho}^{\text{ac}}$)	30 dBm	10 dBm
Noise Figure (σ^2)	9dB	
Antenna Spacing (d_A)	$\lambda/2$	N/A
CPU Antennas (N)	128	

optimization algorithms. Further, these proposed solutions are compared with the classical fiber-fronthaul based cell-free massive MIMO systems. The flow of this section is as follows. First, we present the adopted system setup and parameters in Section VII-A. Then, we evaluate the performance of the two proposed cell-free massive MIMO architectures, namely with separate APs and connected APs, in Sections VII-B and VII-C.

A. Simulation Setup

In the simulation, we considered the proposed cell-free massive MIMO architectures in Section II (for the separate APs case) and in Section VI (for the connected APs case). For both architectures, we adopt mmWave fronthaul channels at $f^{\text{fh}} = 28$ GHz carrier frequency, and access channels at $f^{\text{ac}} = 3.5$ GHz. The bandwidth allocated for the access channels is $B^{\text{ac}} = 20$ MHz, and various values for the fronthaul bandwidth are considered, as will be discussed in the results. The noise figure is set to 9 dB for both the fronthaul and access links. The central unit is deploying a uniform linear array with $N = 128$ elements and half wavelength antenna spacing. For the phase-shift set of analog beamforming at the CPU, we use $q = 3$ bit uniform quantizers. The transmit power of the central unit is assumed to be $\tilde{\rho}^{\text{fh}} = 30$ dBm for the fronthaul links while the APs and users transmit powers are set to $\tilde{\rho}^{\text{ac}} = \tilde{\rho}^t = 10$ dBm for the access links and pilot transmissions. The noise power is determined based on the bandwidth of the fronthaul and access channels, following $\sigma_n^2 = \sigma^2 \cdot B \cdot k \cdot T$, with the Boltzmann constant k, the channel bandwidth B, the temperature $T = 290$ K, and the noise figure σ^2. The normalized power levels of the different channels are then determined by $\tilde{\rho}/\sigma_n^2$ with the corresponding parameters of the specific channel. These baseline system parameters are summarized in Table I. The values that are selected differently from the baseline parameters in different figures are explicitly stated.

At the fronthaul, we consider a LOS multi-path channel with $V = 1 + V_{\text{NLOS}}$ paths, considering the potential availability of the LOS path with the fixed placement of the CPU and APs. The parameters of the NLOS path are determined randomly, where the number of paths $V_{\text{NLOS}} \sim \mathcal{U}[1, 6]$, angle-of-departure $\theta_{m,v} \sim \mathcal{U}[\pi/2, \pi/2]$ and the small-scale fading $\alpha_{m,v}^{\text{fh}} \sim \mathcal{CN}(0, 1)$ for $v > 1$. The LOS path parameters are determined from the geometry of the simulation model with $\alpha_{m,1}^{\text{fh}} = 1$. For the large-scale channel coefficients of LOS and NLOS paths, we adopt the Urban Micro (UMi) street-canyon model given in 3GPP 38.901 [32] as follows:

$$\text{PL}_{\text{LOS}}(d, f_c) = 32.4 + 21 \log_{10} d + 21 \log_{10} f_c + \mathcal{X}$$
$$\text{PL}_{\text{NLOS}}(d, f_c) = 32.4 + 31.9 \log_{10} d + 21 \log_{10} f_c + \mathcal{X},$$
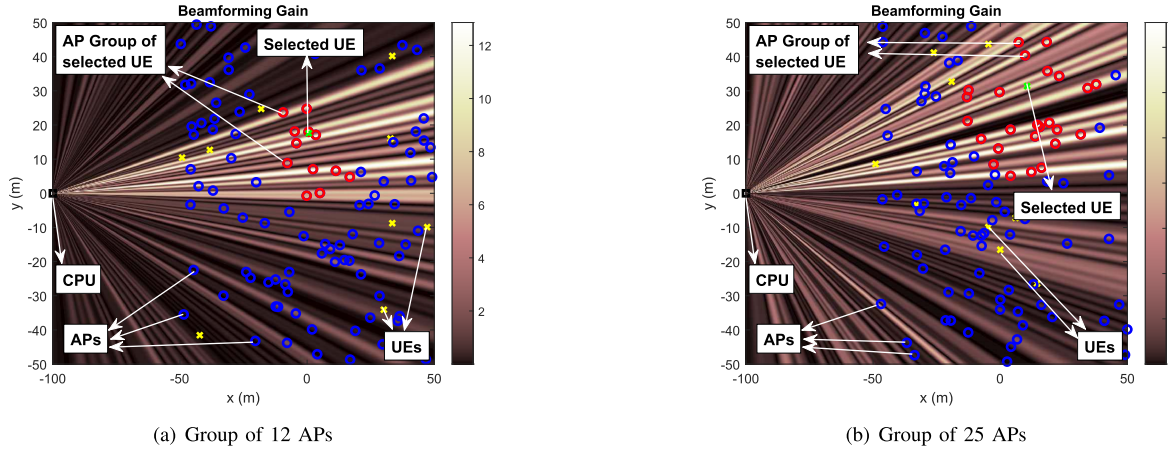
(a) Group of 12 APs   (b) Group of 25 APs

Fig. 4. The beamforming gains of the optimized beamforming vectors for the groups of (a) 12 APs and (b) 25 APs. The given scenario adopts randomly placed $M = 100$ APs and $K = 10$ UEs. The beamforming gain of a group depends on the group size, decreasing with larger groups.

where f_c is in GHz and d is in meters. $\mathcal{X}$ (dB) is the shadow fading effect determined by a Gaussian random variable with zero mean and standard deviation 4 and 8.2 for LOS and NLOS paths of the fronthaul. For the access channel, we adopt $\text{PL}_{\text{NLOS}}(d_{mk}^{\text{ac}}, f_c^{\text{ac}})$, however, apply a spatially correlated shadow fading exactly as described in [1] with $\delta = 0.5$ and decorrelation distance 100m. Next, we evaluate the performance of the proposed solutions.

B. Evaluating the Proposed Architecture With Separate APs (Wireless-Only Fronthaul)

The objective of this subsection is to evaluate the performance of the proposed cell-free massive MIMO architecture with wireless fronthaul (and separate APs). In particular, we want to draw some insights into how this architecture performs if compared with the classical fiber-fronthaul based cell-free massive MIMO architecture and whether it can achieve comparable data rates. To do that, we consider a setup where the UEs and APs are randomly placed over an area of $100m \times 100m$ centered at $(0, 0)$. Unless otherwise stated, the following simulations assume that $K = 10$ UEs and $M = 100$ APs are placed in the considered area. The central unit (CPU) is located at $\mathbf{z}_0 = (x_0, y_0) = (-D, 0)$, i.e., at a distance $D = 100$m away from the center of the square area. The simulations are averaged over 250 realizations. Each realization drops the APs and UEs randomly adopting a uniform probability distribution. With this setup, we study the following important questions.

What Is the Performance of the Fronthaul Beamforming? To evaluate the performance of the adopted fronthaul beamforming approach[5] in our wireless-fronthaul based cell-free massive MIMO architecture, we plot the achievable beamforming gain for the considered deployment in Fig. 4. This figure shows the beamforming gain in x-y coordinates for two different AP group sizes, namely a group of 12 APs
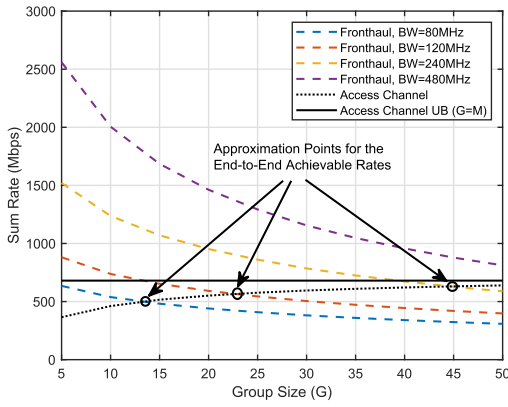
---

[5] As mentioned in Section V-B, we adopt the suboptimal approach in [29] to optimize the beams. The approach starts from a beam vector, and iteratively updates each beamforming coefficient to maximize the rate until the convergence. To ensure a good performance, this solution is applied for 100 starting points, and the best performing beam is adopted in the simulations.

in Fig. 4(a) and a group of 25 APs in Fig. 4(b). These figures provide a visual verification that the central unit focuses it multi-cast beam towards the APs in the group of interest. In terms of beamforming gain, it varies from 12 (i.e., around 11dB) in the case of 12 APs group to 4 (i.e., around 6dB) in the 25 APs/group case. This highlights the trade-off between the fronthaul rate and access rate as the group sizes increases, as this decreases the fronthaul rate while increasing the access rate. It is worth mentioning here that this fronthaul beamforming can be further improved when adopting hybrid precoding approaches [22], [33]. Also, leveraging machine learning could enable the autonomous design of these beamforming vectors/codebooks [34]–[36]. This can further reduce the deployment overhead.

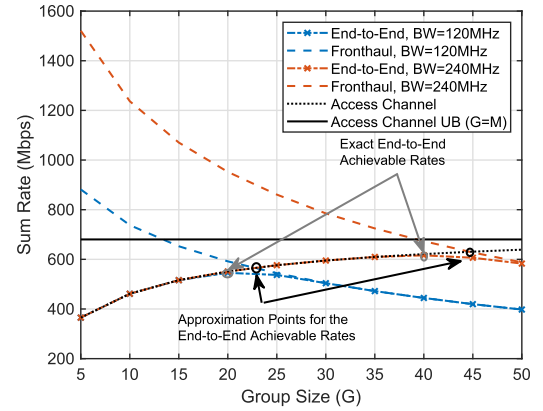**What Are the Achievable Fronthaul and Access Rates?** Now, we want to evaluate the fronthaul and access achievable rates using the developed rate optimization solutions in Section V. Adopting the same setup described earlier in this section with 10 UEs and 100 APs, we plot the achievable fronthaul/access sum-rates in Fig. 5(a) versus the AP group size. These achievable sum-rates (sum of the rates of UEs) are obtained from the solutions of the fronthaul (Approach 1) and cell-free massive MIMO solutions, and present the upper-bounds on the end-to-end rate. They reflect the behavior of the system within a transmission frame and their intersection point can be considered as a good approximation for the achievable end-to-end sum-rate, as will be discussed shortly in Fig. 5(b). The group for each user is determined based on the channel gain criterion described in Section V-C, the access rate is optimized based on the solution presented in Section V-A, and the fronthaul time allocation optimization is implemented based on Approach 1 in V-B. As an upper bound on the cell-free massive MIMO system, we show fully activated APs, i.e., $G = M = 100$, without any fronthaul limitations. This curve is abbreviated as Access Channel UB in the figure. As shown in Fig. 5(a), as the group size increases, the access rate increases because of the higher access channel beamforming gain and the fronthaul rate decreases because of the multicasting to more APs with a lower beamforming gain. With high enough fronthaul bandwidth, e.g., 240MHz,

(a) Sum fronthaul and access channel rates

(b) Sum fronthaul, access channel and end-to-end rates

Fig. 5. The access channel and fronthaul sum data rates of the proposed architecture with different group size and fronthaul bandwidth values. (a) shows the change in the fronthaul rates based on the available bandwidth an group size. (b) includes the sum of the end-to-end data rates reflecting the exact performance.

the proposed rate optimization approach and cell-free massive MIMO architecture intersects at 630Mbps sum-rate. This is very close to the 680Mbps sum-rate achieved with 100 active APs without any fronthaul limitations. These results highlight the promising data rates of the proposed rate optimization solutions.
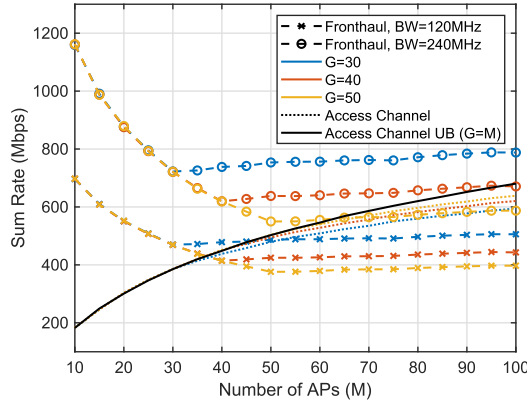
Can the Proposed Architecture With Wireless Fronthaul Approach the Data Rates of the Classical Fiber-Fronthaul Based Cell-Free Massive MIMO Architecture? This is a key question that we are trying to answer in this paper. Essentially, the proposed architecture with wireless higher-band fronthaul has clear gains in terms of the installation cost/time/flexibility, but could it also provide comparable achievable rates to the fiber-based architecture? In Fig. 5(a), we provided some insights into the answer of this question by separately evaluating the achievable fronthaul and access rates. The exact achievable rate, however, is the one with matched fronthaul and access rates. To evaluate this exact rate, we adopt the fronthaul rate optimization (Approach 2) in Section V-B which ensures that the fronthaul rate of each link does not exceed its corresponding access rate. In Fig. 5(b), we plot the achievable rates for the access and fronthaul links using Approach 2. This rate then represents the exact end-to-end achievable rates using the proposed architecture. For reference, we also plot the fronthaul rates using Approach 1. As shown in Fig. 5(b), with sufficient fronthaul bandwidth, the exact achievable end-to-end data rates using the proposed architecture are very comparable to the classical fiber-based cell-free massive MIMO architecture. For example, with 240MHz fronthaul bandwidth, the proposed solution achieves around 620Mbps sum-rate compared to 680Mbps for the fully active APs without any fronthaul limitations. This difference can be further reduced by increasing the fronthaul bandwidth. It is also good to note that the exact rate is very close to the approximation discussed in Fig. 5(a), i.e., 630.

Can the Proposed Architecture Support a Large Number of Distributed APs and UEs? The scalability is a key objective of cell-free massive MIMO systems to be able to support more users/larger areas and to increase the beamforming gains/achievable rates. To draw some insights into

the capability of the proposed architecture in supporting large numbers of APs and UEs, we plot the fronthaul/access rates vs the number of APs in Fig. 6(a), and the number of UEs in Fig. 6(b). As shown in Fig. 6(a), for each group size, the access rates increase with denser APs due to the higher access beamforming gains. At the fronthaul, the rate decreases rapidly for $M \leq G$. In this region, $M \leq G$, the group size is taken as $G = M$. Therefore, every new AP introduced into the system also increases the effective group size by adding APs to the groups. This decreases the fronthaul beamforming gains. For larger values of M, the fronthaul rates slightly increase, as the area containing a single AP group becomes smaller, allowing higher beamforming gains. Overall, however, the end-to-end data rates, which are approximately given by the intersections of the fronthaul and access channel rates, are increasing with more APs. This is also very comparable to the increase in data rates experienced by the fiber-based architecture, which highlights the potential of the proposed solution. In Fig. 6(b), we investigate the effect of increasing number of UEs. In the figure, the sum fronthaul rate with a fixed group size is almost constant for different number of UEs, mainly due to the fronthaul resources and TDMA based communication design. However, with smaller group size values, the achievable maximum fronthaul rate increases, while less number of APs serve each UE and the access channel rate decreases. At the access channel, with the scaling of UEs, the effect of the group size increases. However, even with $K = 50$ APs, the 80% of the data rate with fiber-fronthaul can be obtained while only using 240MHz fronthaul bandwidth.

C. Evaluating the Proposed Architecture With Connected APs (Mixed-Fronthaul)

In this subsection, we evaluate the performance of the proposed modified architecture in Section VI with connected APs. As shown in Fig. 7, we consider a square-grid of APs, where the APs are located uniformly in an 10×10 grid with equal distance between the rows and columns. For simplicity, we assume that the 10 APs in each line in the y-axis form
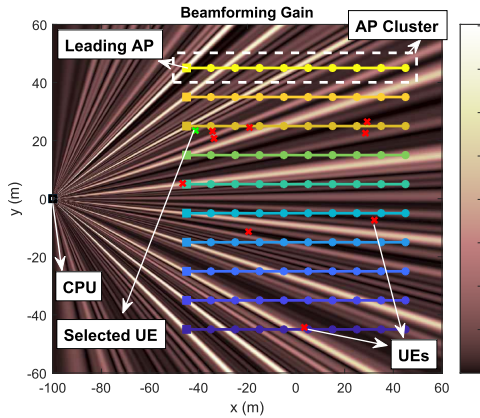
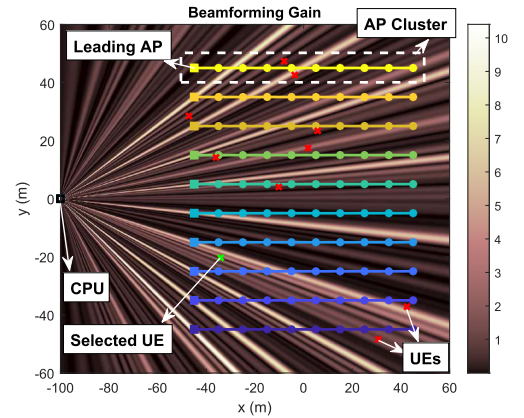(a) Achievable rates with varying number of APs
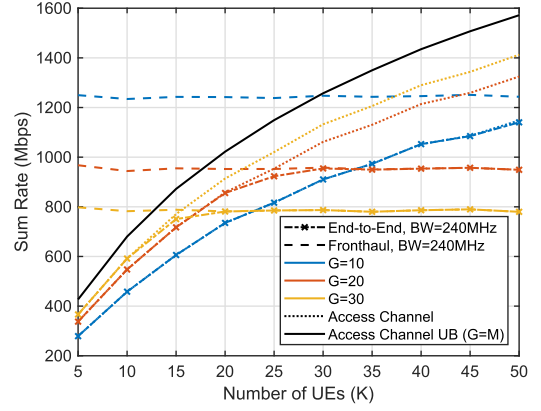
(b) Achievable rates with varying number of UEs

Fig. 6.   The access channel and fronthaul sum data rates of the proposed architecture with different number of (a) APs with $K = 10$ UEs, and (b) UEs with $M = 100$ APs. The figures show the relationship of group sizes, G, to different number of APs and UEs.

(a) Group size of 10 clusters

(b) Group size of 9 clusters

Fig. 7.   The beamforming gains with the optimized beamforming vectors are shown for the mixed-fronthaul architecture. In (a), all of the clusters are activated, hence the beam gain is maximized towards all of the leading APs. In (b), the bottom 9 clusters are activated, and the beam gain is not maximized for the leading AP of the top cluster.

a cluster. The leaders of the clusters (that communicate with the central unit through the mmWave fronthaul) are selected as the closest APs to the central unit. Further, we assume that $K = 10$ UEs are randomly located in the $100m \times 100m$ area of interest. The central unit is located at the origin, as depicted in Fig. 7, and is equipped with a ULA of $N = 128$ antenna elements.

When Does the Achievable Data Rate With the Proposed Architecture Approach the Optimal Solution? To evaluate the achievable data rates with the proposed connected-APs cell-free massive MIMO architecture, we adopt the setup described in the previous paragraph, which is also depicted in Fig. 7. For this setup, we first investigate the achievable beamforming gains for different numbers of AP clusters per group. To illustrate, we consider one UE and plot the beamforming gains for the clusters serving this UE in two scenarios: (i) When a group of 10 AP clusters (i.e., all the clusters) jointly serve this UE as shown in Fig. 7(a) and (ii) when a group of the 9 bottom clusters serve the UE, as shown in Fig. 7(b). As expected, when the number of AP clusters per group increases, the beamforming gain decreases. Next, we calculate the achievable fronthaul and access rates of the
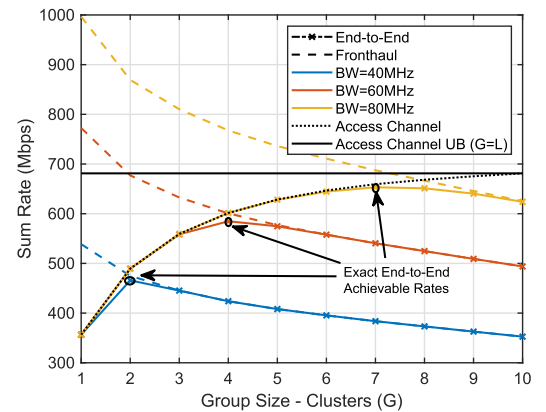

Fig. 8.   For the mixed-fronthaul architecture, the fronthaul, access channel and end-to-end sum rates are shown for different number of group size (of clusters). In the figure, the 10×10 APs rectangular grid scenario illustrated in Fig. 7 is adopted and each cluster consists of 10 APs.

same setup in Fig. 8 for different numbers of AP clusters per group. As shown in this figure, the modified architecture with connected APs has clear data rate gains compared to the

separate APs architecture studied in Fig. 5. With only 80MHz fronthaul bandwidth, the proposed connected-APs architecture achieves nearly the same end-to-end rates obtained by the upper bound, which is defined by the classical fiber-fronthaul based cell-free massive MIMO architecture with fully active APs. This provides a promising solution for flexible, scalable, and efficient cell-free massive MIMO systems.

VIII. CONCLUSION

In this work, we proposed two wireless-fronthaul based architectures for cell-free massive MIMO systems where the fronthaul link is operating at a higher band (e.g., mmWave) compared to the access band (e.g., sub-6GHz). We formulated the end-to-end data rate optimization problems for these architectures based on user-centric AP grouping. Then, we developed efficient transmission and resource allocation strategies that can achieve near-optimal performance. Simulation results showed that the proposed architectures can achieve data rates that approach the optimal rates obtained with optical fiber-based fronthaul with reasonable assumptions on the number or CPU antennas, coverage area, number of users, etc. The results also showed that these architecture are capable of supporting large numbers of APs under realistic fronthaul bandwidth requirements. For the future work, it is interesting to develop more advanced fronthaul beamforming solutions using hybrid analog/digital architectures. In addition, more practical studies adopting more realistic cell-free constraints (e.g., pilot contamination), practical issues of the mmWave systems, and energy efficiency of the architecture are interesting directions for future extensions. Further, the design of machine learning based adaptive AP grouping and beamforming optimization approaches for the proposed wireless-fronthaul based cell-free massive MIMO architectures can be investigated.

APPENDIX A

For the solution of the problem in (21), we first introduce a slack variable [37], η, to represent minimum time-scaled data rate of the groups and write the equivalent formulation

$$\max_{\{t_k\},\eta} \eta \quad \text{s.t.} \quad t_k R^{\text{fh}}(\mathcal{G}_k) \geq \eta \ \forall k \in \mathcal{K},$$
$$\sum_{k\in\mathcal{K}} t_k = 1, \quad t_k \geq 0 \ \ \forall k \in \mathcal{K}. \quad (25)$$

This formulation is a linear program and the solution can be attained by finding maximum η. Let us assume that there exists an optimal η, $\eta^\star$, that satisfies the minimum rate constraint of (25), i.e., $t_k R^{\text{fh}}(\mathcal{G}_k) \geq \eta$. Due to the maximization objective, this constraint is satisfied at the equality, and it is possible to obtain the exact solution thanks to the linear structure of the problem. Hence, for the equality, we can write $\eta^\star = t_k^\star R^{\text{fh}}(\mathcal{G}_k)$, where $t_k^\star$ denotes the optimal time fraction for group k. Then, we have $t_k^\star = \frac{\eta^\star}{R^{\text{fh}}(\mathcal{G}_k)}$. By replacing t_k with $t_k^\star$ in the time constraint of (25), $\sum_{k\in\mathcal{K}} t_k = 1$, we obtain

$$\eta^\star = \frac{1}{\sum_{k\in\mathcal{K}} \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}} = \frac{\text{HM}\{R^{\text{fh}}(\mathcal{G})\}}{K} \quad (26)$$

which leads to the solution given in (22).

APPENDIX B

The problem given in (23) is in the form

$$\max_{\{t_k\}} \sum_{k\in\mathcal{K}} w_k \log t_k$$
$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} t_k \leq 1, \ t_k \leq \bar{t}_k \ \forall k \in \mathcal{K}, \ \ t_k \geq 0 \ \forall k \in \mathcal{K} \quad (27)$$

for the constants defined by $w_k = \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}$ and $\bar{t}_k = \frac{B^{\text{ac}} R_k^{\text{ac}}}{B^{\text{fh}} R^{\text{fh}}(\mathcal{G}_k)}$. The problem is concave due to the sum of logarithms in the objective, and it can be solved by the Karush-Kuhn-Tucker (KKT) conditions [37]. For the solution, we first write the Lagrangian function[6] as

$$\mathcal{L}(\mathbf{t}, \boldsymbol{\lambda}, \nu) = \sum_{k\in\mathcal{K}} w_k \log t_k - \nu\left(\sum_{k\in\mathcal{K}} t_k - 1\right) - \sum_{k\in\mathcal{K}} \lambda_k(t_k - \bar{t}_k), \quad (28)$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]$ and ν are the Lagrange multipliers. We note that the partial derivative of the Lagrangian function is $\frac{\partial \mathcal{L}(\mathbf{t},\boldsymbol{\lambda},\nu)}{\partial t_k} = \frac{w_k}{t_k} - \nu - \lambda_k$, and write the KKT conditions as

$$\frac{w_k}{t_k^\star} - \nu - \lambda_k = 0, \quad \text{(Stationarity)}$$

$$\sum_{k\in\mathcal{K}} t_k^\star \leq 1, \quad t_k^\star \leq \bar{t}_k, \quad \lambda_k, \nu \geq 0, \quad \text{(Feasibility)}$$

$$\lambda_k(t_k^\star - \bar{t}_k) = 0, \quad \nu\left(\sum_{k\in\mathcal{K}} t_k^\star - 1\right) = 0. \quad \text{(C. Slackness)}$$

From the stationarity condition, we have $\nu = \frac{w_k}{t_k} - \lambda_k$. Combining this with the complementary slackness condition of λ_k, we obtain

$$t_k^\star = \begin{cases} \dfrac{w_k}{\nu} & \text{if } \lambda = 0, \\ \bar{t}_k & \text{if } \lambda > 0. \end{cases} \quad (29)$$

This can be simplified to $t_k^\star = \min\{\frac{w_k}{\nu}, \bar{t}_k\}$. Finally, we define $\eta = \frac{1}{\nu}$, and place $t_k^\star$ in terms of η in the first (primal) feasibility condition, and obtain the solution given in (24).

REFERENCES

[1] 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.

[2] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 197, Dec. 2019.

[3] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. Conf. Rec. 41st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2007, pp. 83–87.

[4] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2112–2127, Oct. 2013.

[5] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

---

[6]We ignore the constraint $t_k \geq 0$ since the logarithm is only defined in this region, and any solution obtained from the formulation will satisfy the condition.

[6] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[7] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2017.

[8] 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.

[9] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[10] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1465–1470.

[11] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max–min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6796–6815, Oct. 2019.

[12] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–7.

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

[14] G. Femenias and F. Riera-Palou, "Reduced-complexity downlink cell-free mmWave massive MIMO systems with fronthaul constraints," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, A Coruna, Spain, Sep. 2019, pp. 1–5.

[15] Y. Zhu, Y. Niu, J. Li, D. O. Wu, Y. Li, and D. Jin, "QoS-aware scheduling for small cell millimeter wave mesh backhaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[16] W. Hao and S. Yang, "Small cell cluster-based resource allocation for wireless backhaul in two-tier heterogeneous networks with massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 509–523, Jan. 2018.

[17] 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.

[18] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Mar. 2016.

[19] R. G. Stephen and R. Zhang, "Joint millimeter-wave fronthaul and OFDMA resource allocation in ultra-dense CRAN," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1411–1423, Mar. 2017.

[20] B. Hu, C. Hua, J. Zhang, C. Chen, and X. Guan, "Joint fronthaul multicast beamforming and user-centric clustering in downlink C-RANs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5395–5409, Aug. 2017.

[21] B. Hu, C. Hua, C. Chen, and X. Guan, "Joint beamformer design for wireless fronthaul and access links in C-RANs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2869–2881, May 2018.

[22] A. Alkhateeb, G. Leus, and R. W. Heath, "Multi-layer precoding for full-dimensional massive MIMO systems," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 815–819.

[23] T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures," *J. Opt. Commun. Netw.*, vol. 7, no. 11, pp. B38–B45, Nov. 2015.

[24] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[25] D. Shrestha, Z. Pang, and D. Dzung, "Precise clock synchronization in high performance wireless communication for time sensitive networking," *IEEE Access*, vol. 6, pp. 8944–8953, 2018.

[26] R. Rogalin *et al.*, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.

[27] R. W. Heath, 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.

[28] A. Ikhlef, D. S. Michalopoulos, and R. Schober, "Max-max relay selection for relays with buffers," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1124–1135, Mar. 2012.

[29] Z. Wang, Q. Liu, M. Li, and W. Kellerer, "Energy efficient analog beamformer design for mmWave multicast transmission," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 552–564, Jun. 2019.

[30] J.-C. Chen and J.-H. Wen, "Improved analog beamformer design for millimeter-wave multicast transmission with large antenna arrays," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1825–1829, Aug. 2020.

[31] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[32] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, 3GPP document TR 38.901, version 14.1.1 Release 14, 2017.

[33] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 800–805.

[34] M. Alrabeiah, Y. Zhang, and A. Alkhateeb, "Neural networks based beam codebooks: Learning mmWave massive MIMO beams that adapt to deployment and hardware," 2020, *arXiv:2006.14501*.

[35] Y. Zhang and A. Alkhateeb, "Learning reflection beamforming codebooks for arbitrary RIS and non-stationary channels," 2021, *arXiv:2109.14909*.

[36] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Reinforcement learning of beam codebooks in millimeter wave and terahertz MIMO systems," 2021, *arXiv:2102.11392*.

[37] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Umut Demirhan received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with Arizona State University, Tempe, AZ, USA. He has held research internships at CommScope, Nokia Bell Labs, and Samsung Research America. His research interests include the broad areas of wireless communications, signal processing, and machine learning.

Ahmed Alkhateeb received the Ph.D. degree in electrical engineering from The University of Texas at Austin, USA, in August 2016. From September 2016 and December 2017, he was a Wireless Communications Researcher at the Connectivity Laboratory, Facebook, Menlo Park, CA, USA. He joined Arizona State University (ASU) in Spring 2018, where he is currently an Assistant Professor with the School of Electrical, Computer and Energy Engineering. He has held research and development internships at Futurewei Technologies, Chicago, IL, USA, and Samsung Research America (SRA), Dallas, TX, USA. His research interests are in the broad areas of wireless communications, communication theory, signal processing, machine learning, and applied math. He was a recipient of the 2012 MCD Fellowship from The University of Texas at Austin, the 2016 IEEE Signal Processing Society Young Author Best Paper Award for his work on hybrid precoding and channel estimation in millimeter wave communication systems, and the 2021 NSF CAREER Award.