User Subgrouping in Scalable Cell-Free Massive MIMO Multicasting Systems

Alejandro de la Fuente*, Guillem Femenias†, Felip Riera-Palou†, Giovanni Interdonato‡

*Dept. of Signal Theory and Communications, University Rey Juan Carlos, 28942 Fuenlabrada (Madrid), Spain
†Mobile Communications Group, University of the Balearic Islands, 07122 Mallorca (Illes Balears), Spain
‡Dept. of Electrical and Information Engineering (DIEI), University of Cassino and Southern Lazio, Cassino, Italy
Email: alejandro.fuente@urjc.es

Abstract—Cell-free massive multiple-input multiple-output (CF-mMIMO) is a breakthrough technology for beyond-5G systems, designed to significantly boost the energy and spectral efficiencies of future mobile networks while ensuring a consistent quality of service for all users. Additionally, multicasting has gained considerable attention recently because physical-layer multicasting offers an efficient method for simultaneously serving multiple users with identical service demands by sharing radio resources. Typically, multicast services are delivered either via unicast transmissions or a single multicast transmission. This work, however, introduces a novel subgroup-centric multicast CF-mMIMO framework that divides users into several multicast subgroups based on the similarities in their spatial channel characteristics. This approach allows for efficient sharing of the pilot sequences used for channel estimation and the precoding filters used for data transmission. The proposed framework employs two scalable precoding strategies: centralized improved partial MMSE (IP-MMSE) and distributed conjugate beamforming (CB). Numerical results show that for scenarios where users are uniformly distributed across the service area, unicast transmissions using centralized IP-MMSE precoding are optimal. However, in cases where users are spatially clustered, multicast subgrouping significantly improves the sum spectral efficiency (SE) of the multicast service compared to both unicast and single multicast transmission. Notably, in clustered scenarios, distributed CB precoding outperforms IP-MMSE in terms of per-user SE, making it the best solution for delivering multicast content.

Index Terms—Cell-free massive MIMO, multicasting, user subgrouping, scalability.

I. INTRODUCTION

Cell-free massive multiple-input-multiple-output (CF-mMIMO) is an emerging technique for beyond-5G systems owing to its outstanding enhancements in energy efficiency (EE), spectral efficiency (SE), service quality, and reliability [1]. In a CF-mMIMO system, a large number of access points (APs) are distributed across the network and they are connected to a central processing unit (CPU) via fronthaul links to exchange the channel state information (CSI) and the user-specific data. CF-mMIMO exploits the benefits of massive MIMO (mMIMO) and network MIMO, being able to provide mobile stations (MSs) with nearly uniform

service across the coverage area [2]. The vast majority of cell-free literature explores the improvements attained in the SE, EE, and coverage performance in CF-mMIMO unicast transmissions. Critically, within the vast volume of data traffic that has been predicted by 2029 [3], a significant portion will comprise content that can potentially be shared among groups of users in the network, and therefore, can be leveraged through broadcast/multicast techniques [4], [5]. Multicast in mMIMO systems, using both uncorrelated and correlated Rayleigh fading channels, has been proposed and assessed to provide an efficient use of resources [6], [7]. Authors in [8] first assessed the performance of multicasting in a CF-mMIMO context while proposing a novel downlink (DL) pilot training scheme. In [9], authors propose a weighted max-min power optimization algorithm that improves the performance when increasing the number of AP antennas.

Despite the benefits revealed by these initial studies on CF-mMIMO multicasting, critical issues remain. While precoding and power allocation strategies have been extensively studied in unicast CF-mMIMO scenarios, there is no direct and clear-cut translation between the most efficient configurations for unicast transmissions and those for multicast setups. Furthermore, delivering a common data service to a group of MSs has traditionally been accomplished either through a single multicast stream or by sending a unicast stream to each MS in the multicast group. Subgroup-centric multicast, as proposed in this research work, can be seen as a mechanism to tailor the transmission of shared content to groups of users experiencing similar intragroup and widely different intergroup propagation conditions [7], [10], [11]. Questions regarding the proper management of intragroup and intergroup pilot contamination, the design of common channel estimation processes for all users within the same multicast subgroup, and the impact these common channel estimates might have on the design of the common precoder used to convey the DL multicast payload data to users remain largely unsolved.

Contributions: a novel multicast CF-mMIMO framework is developed considering spatial channel correlation and exploring the advantages and disadvantages of more sophisticated precoding techniques depending on the users' distribution within the CF-mMIMO-multicast framework. Users belonging to a multicast group, intended to receive a common service, must share the same uplink (UL) pilot and DL precoder. Our study demonstrates the pilot contamination significance when multicast users experience widely different propagation conditions. As a countermeasure, our work introduces a novel

subgroup-centric framework where multicast users are divided into subgroups according to a metric of large-scale propagation similarity and proposes the use of a subgroup-specific pilot and precoder. A comprehensive evaluation of two precoding schemes and power control strategies tailored to the proposed framework is presented, namely, centralized improved partial MMSE (IP-MMSE) [12] and distributed conjugate beamforming (CB), with fractional power control strategies inspired by Demir et al. [13]. Simulation results demonstrate the benefits of the proposed user subgrouping and power allocation strategies across various system setups, benchmarking the subgroup-centric multicasting approach against conventional multicasting and unicast transmission strategies, considering both uniform and clustered user distributions.

Fig. 1. A CF-mMIMO network with multicast user subgrouping (dashed circles) and AP cooperation clustering (blurred colored shapes).

II. System Model and Multicast Subgrouping Framework

We consider a CF-mMIMO system operating in time division duplexing (TDD) [2] that consists of a CPU connected via ideal fronthaul links to L APs, each one equipped with N antennas. The APs are deployed over the coverage area to simultaneously provide, either through multicast or unicast, a common data service to K single-antenna MSs on the same time-frequency resources. The set of MSs is denoted by $\mathcal{K}$ and indexed by $k \in \mathcal{K} = \{1, \ldots, K\}$. The set of APs is denoted by $\mathcal{L}$ and indexed by $l \in \mathcal{L} = \{1, \ldots, L\}$. Each TDD frame is divided into uplink (UL) training phase and DL payload data transmission phase, whose lengths, measured in samples, are denoted as τ_p and τ_d, respectively. The frame length is denoted by $\tau_c = \tau_p + \tau_d$ and assumed to fit the coherence block.

A. Channel Model

A conventional block fading channel model is considered wherein the channel is time-invariant and frequency flat within a time-frequency coherence block, and varies independently over different coherence blocks (block fading). The channel response vector $\boldsymbol{h}_{lk} \in \mathbb{C}^N$ between the AP l and the multicast MS k, in an arbitrary coherence block[1], is distributed as $\boldsymbol{h}_{lk} \sim \mathcal{CN}(\boldsymbol{0}_N, \boldsymbol{R}_{lk})$, where $\boldsymbol{R}_{lk} \in \mathbb{C}^{N \times N}$ is the corresponding positive semi-definite spatial covariance matrix, with average channel gain given by $\beta_{lk} = \text{tr}(\boldsymbol{R}_{lk})/N$. Assuming that the APs and the MSs are well separated, the channel vectors of different AP-MS pairs can be considered to be independently distributed, i.e., $\mathbb{E}\{\boldsymbol{h}_{l'k'}\boldsymbol{h}_{lk}^{\mathsf{H}}\} = \boldsymbol{0}_{N \times N}, \ \forall l'k' \neq lk$. Thus, the channel from MS k to the complete set of APs $l \in \mathcal{L}$, $\boldsymbol{h}_k = [\boldsymbol{h}_{1k}^T \ldots \boldsymbol{h}_{Lk}^T]^T$, is distributed as $\boldsymbol{h}_k \sim \mathcal{CN}(\boldsymbol{0}_{LN}, \boldsymbol{R}_k)$, where $\boldsymbol{R}_k = \text{blkdiag}(\boldsymbol{R}_{1k}, \ldots, \boldsymbol{R}_{Lk}) \in \mathbb{C}^{LN \times LN}$ is the block-diagonal spatial covariance matrix related to MS k.

Channel covariance matrices $\boldsymbol{R}_{lk}, \forall \ k \in \mathcal{K}, \ \forall \ l \in \mathcal{L}$, can be estimated at each AP over a large-scale fading time scale (i.e., over multiple coherence blocks) and thus can be safely assumed to be perfectly known at both the APs and the CPU [14].

B. Subgroup-Centric CF-mMIMO Multicasting

This work proposes delivering a multicast service by subgrouping the K multicast MSs into G disjoint subgroups. In
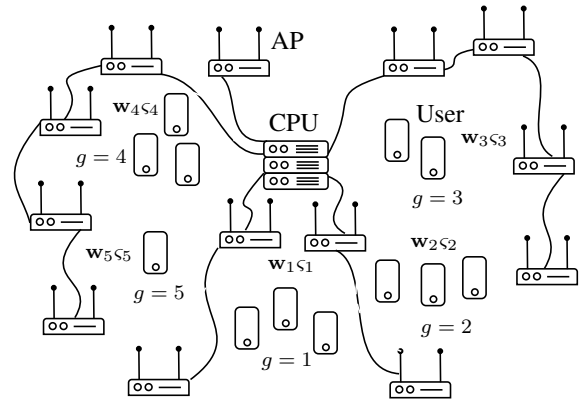
[1]For the sake of clarity, we omit the index identifying the coherence block.

this case, the SE of each multicast subgroup can be adapted to the propagation conditions experienced by the MSs in that particular subgroup, and the multiuser interference reduces to the inter-multicast subgroup interference. We denote the set of multicast subgroups by $\mathcal{G}$ and the subgroups are indexed by $g \in \mathcal{G} = \{1, \ldots, G\}$. The set of MSs in subgroup g is denoted by $\mathcal{K}_g$. Letting $K_g = |\mathcal{K}_g|$ be the number of MSs in subgroup g, it holds that $K = \sum_{g=1}^{G} K_g$. It is interesting to note that in scenarios where the MSs are distributed in spatial clusters, subgrouping the MSs experiencing propagation channels with similar large-scale channel characteristics can bring significant benefits. In this case, MSs in the same multicast subgroup can share the same pilot sequence, thereby controlling pilot contamination.

In [15], the authors benefit from the special attributes of CF-mMIMO channel (distributed transmissions) to propose a K-means-based partitioning algorithm that employs a metric based on the propagation gain vectors $\boldsymbol{\beta}_k = [\beta_{1k} \ldots \beta_{Lk}]^{\mathsf{T}}$ to group MSs to improve the pilot allocation when pilot contamination cannot be avoided. In the present work, the K-means-based protocol described in [15] is suitably adapted to classify MSs experiencing similar large-scale channel characteristics into multicast subgroups that will benefit from sharing the UL pilot sequence to improve both the channel estimation process and the design of the precoding vectors used for DL data transmission, improving the SE results.

Consequently, a multicast subgroup-centric CF-mMIMO is implemented where the MSs in multicast subgroup g are served by a subset of APs somehow readapting the concept of user-centric transmission. We denote the subset of APs serving subgroup g by $\mathcal{L}_g \subseteq \{1, \ldots, L\}$, where $|\mathcal{L}_g| = L_g \leq L$. For later convenience, given a subgroup g, we define the set $\mathcal{S}_g$ as the collection of multicast subgroups served by some (or all) of the APs serving subgroup g, that is, $\mathcal{S}_g = \{c : \mathcal{L}_g \cap \mathcal{L}_c \neq \emptyset\}$. The set of multicast subgroups served by AP l is denoted as $\mathcal{D}_l$. Figure 1 illustrates the system model of CF-mMIMO multicasting with user subgrouping and dynamic cooperation clustering (DCC) [13].

C. Multicast AP Cooperation Clustering and Pilot Allocation

All MSs in the same subgroup are assigned the same pilot sequence chosen among τ_p available mutually orthogonal pilots (pilot sequence size is equal to τ_p samples). Note that, as co-pilot MSs have linearly dependent channel estimates [16], [17], the APs cannot separate the MSs of the same subgroup in the spatial domain. As a different pilot sequence is assigned to each subgroup g, finding the optimal pilot allocation is a combinatorial optimization problem with τ_p^G possible assignments in a setup with G subgroups and τ_p pilots; thus, the complexity of evaluating all of them grows exponentially with the number of subgroups. To address this issue, we apply the DCC and pilot allocation algorithm proposed by Demir *et al.* in [13] but suitably adapting it to use multicast subgroup information instead of individual MS information. The algorithm will iteratively assign pilots to the subgroups by always selecting the one that leads to the least pilot contamination. To that end, pilots are first assigned to subgroups, and then each AP is allowed to serve exactly τ_p subgroups. For every pilot, the AP serves the subgroup with the strongest common average channel gain (i.e., $\frac{1}{K_g} \sum_{k \in \mathcal{K}_g} \beta_{lk}$) among the set of subgroups that have been assigned that pilot.

As shown in Algorithm 1, the multicast subgroup pilot assignment and cooperation clustering creation process consists of two steps. In the first step, the τ_p subgroups with indices from 1 to τ_p are assigned mutually orthogonal pilots, that is, every MS $k \in \mathcal{K}_g$ uses pilot g for $g \in \{1, \ldots, \tau_p\}$. The remaining subgroups, with indices ranging from $\tau_p + 1$ to G, are then assigned pilots one after the other as follows. Assuming that AP l presents an excellent average large-scale propagation gain for MSs in subgroup g, it is expected to contribute strongly to the service of subgroup g and consequently, it is preferable to assign subgroup g to the pilot for which AP l experiences the least pilot contamination. Hence, for each pilot t, AP l computes the sum of the channel gains β_{lk} of the MSs that have already been allocated this pilot, and then identifies the index of the pilot minimizing the pilot interference as

$$\tau \leftarrow \underset{\psi \in \{1, \ldots, \tau_p\}}{\arg\min} \sum_{\substack{c \in \mathcal{G}, \, c \neq g \\ \psi_g = \psi}} \sum_{k \in \mathcal{K}_c} \beta_{lk}. \tag{1}$$

Pilot τ is then assigned to subgroup g and the algorithm continues with the next subgroup.

In the second step of the algorithm, the clusters of APs are created as soon as all the subgroups have been assigned to pilots. Each AP goes through each pilot and identifies which is the subgroup experiencing the largest common average channel gain among those using that pilot. This subgroup will be served by this particular AP. In order to prevent that a subgroup is left unserved, a multicast subgroup g is always served by at least their own Master AP l, which is selected by the CPU according to the rule $l = \underset{l \in \mathcal{L}}{\arg\max} \frac{1}{K_g} \sum_{k \in \mathcal{K}_g} \beta_{lk}$.

D. Uplink Channel Estimation

Let $\psi_g \in \mathbb{C}^{\tau_p \times 1}$ be the pilot sequence assigned to subgroup g, with $\|\psi_g\|_2^2 = 1$. Ideally, pilot sequences should be mutually orthogonal, nonetheless, in practical scenarios it holds that $G > \tau_p$, and a given pilot sequence may be assigned to more

---

Algorithm 1 Multicast subgroup pilot assignment and cooperation clustering

Input: $\tau_p, \beta_{lk}, G, \mathcal{K}_g$
for $g = 1, \ldots, \tau_p$ **do**
    $\psi_g \leftarrow g$
end for
for $g = \tau_p + 1, \ldots, G$ **do**
    $l \leftarrow \arg\max_{l \in \mathcal{L}} \frac{1}{K_g} \sum_{k \in \mathcal{K}_g} \beta_{lk}$
    $\tau \leftarrow \underset{\substack{\psi \in \{1, \ldots, \tau_p\} \\ \psi_g = \psi}}{\arg\min} \sum_{\substack{c \in \mathcal{G}, \, c \neq g}} \sum_{k \in \mathcal{K}_c} \beta_{lk}$
    $\psi_g \leftarrow \tau$
end for
for $l = 1, \ldots, L$ **do**
    for $\psi = 1, \ldots, \tau_p$ **do**
        $c \leftarrow \underset{g \in \{1, \ldots, G\} : \psi_g = \psi}{\arg\max} \frac{1}{K_g} \sum_{k \in \mathcal{K}_g} \beta_{lk}$
        $\mathcal{L}_c \leftarrow \mathcal{L}_c \cup \{l\}$
    end for
end for
for $g = 1, \ldots, G$ **do**
    if $\mathcal{L}_g = \{\emptyset\}$ **then**
        $l \leftarrow \underset{l \in \mathcal{L}}{\arg\max} \frac{1}{K_g} \sum_{k \in \mathcal{K}_g} \beta_{lk}$
        $\mathcal{L}_g \leftarrow \mathcal{L}_g \cup \{l\}$
    end if
end for
Output: Pilot assignment $\psi_1, \ldots, \psi_G$ and DCCs $\mathcal{L}_1, \ldots, \mathcal{L}_G$

---

than one subgroup, thus resulting in the pilot contamination phenomenon. The $N \times \tau_p$ UL received pilot signal matrix at AP l is

$$\boldsymbol{Y}_l = \sqrt{\tau_p P_p} \sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \boldsymbol{h}_{lk} \psi_g^\mathsf{T} + \boldsymbol{N}_l, \tag{2}$$

where P_p is the per pilot-symbol transmit power of every MS and $\boldsymbol{N}_l \in \mathbb{C}^{N \times \tau_p}$ is the additive white Gaussian noise (AWGN) matrix at AP l with i.i.d. elements distributed as $\mathcal{CN}(0, \sigma_u^2)$. To estimate the channel of MSs in subgroup g, the AP projects the received UL training signal on the corresponding complex conjugate of the pilot sequence ψ_g^* to obtain

$$\boldsymbol{y}_l^g = \sqrt{\tau_p P_p} \sum_{k \in \mathcal{K}_g} \boldsymbol{h}_{lk} + \sqrt{\tau_p P_p} \sum_{\substack{c \neq g \\ \psi_c = \psi_g}} \sum_{i \in \mathcal{K}_c} \boldsymbol{h}_{li} + \boldsymbol{n}_{lg}, \tag{3}$$

where $\boldsymbol{n}_{lg} = \boldsymbol{N}_l \psi_g^* \sim \mathcal{CN}(0, \sigma_u^2 \mathbf{I}_N)$.

Since the APs cannot separate co-pilot MSs in the spatial domain as their channel estimates are correlated, we define the composite channel of the $k \in \mathcal{K}_g$ MSs as

$$\boldsymbol{h}_l^g = \frac{\sqrt{\tau_p P_p}}{K_g} \sum_{k \in \mathcal{K}_g} \boldsymbol{h}_{lk}, \tag{4}$$

which is distributed as $\boldsymbol{h}_l^g \sim \mathcal{CN}(\mathbf{0}_N, \boldsymbol{R}_l^g)$, where

$$\boldsymbol{R}_l^g = \frac{\tau_p P_p}{K_g^2} \sum_{k \in \mathcal{K}_g} \boldsymbol{R}_{lk}. \tag{5}$$

The minimum mean square error (MMSE) channel estimate of the composite channel $\boldsymbol{h}_l^g$ can be obtained either at the lth

AP or at the central CPU (in cases where the APs are not equipped with local baseband processors) [17, Sec. 3.2]

$$\hat{\boldsymbol{h}}_l^g = K_g \ \boldsymbol{R}_l^g \ \boldsymbol{\Gamma}_g^{-1} \ \boldsymbol{y}_l^g \,, \qquad (6)$$

where

$$\boldsymbol{\Gamma}_g = \tau_p P_p \sum_{\substack{\forall c \in \mathcal{G} \\ \boldsymbol{\psi}_c = \boldsymbol{\psi}_g}} \sum_{i \in \mathcal{K}_c} \boldsymbol{R}_{li} + \sigma_u^2 \boldsymbol{I}_N. \qquad (7)$$

Note that the composite channel estimate is distributed as $\hat{\boldsymbol{h}}_l^g \sim \mathcal{CN}(\boldsymbol{0}_N, K_g^2 \ \boldsymbol{R}_l^g \ \boldsymbol{\Gamma}_g^{-1} \ \boldsymbol{R}_l^g)$ and the composite channel estimation error as $\tilde{\boldsymbol{h}}_l^g \sim \mathcal{CN}(\boldsymbol{0}_N, \boldsymbol{R}_l^g - K_g^2 \ \boldsymbol{R}_l^g \ \boldsymbol{\Gamma}_g^{-1} \ \boldsymbol{R}_l^g)$.

E. Downlink Data Transmission and Spectral Efficiency

Multicast subgrouping performs a multicast transmission by employing unique pre-processing schemes and precoding vectors per subgroup. That is, care is taken to pre-process the common information signal into G uncorrelated signals, each directed to the corresponding multicast subgroup. The received DL signal at MS k is

$$y_k = \sum_{l=1}^{L} \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{x}_{lg} + \sum_{l=1}^{L} \sum_{\substack{c=1 \\ c \neq g}}^{G} \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{x}_{lc} + n_k, \qquad (8)$$

where $n_k \sim \mathcal{CN}(0, \sigma_d^2)$ is the AWGN at MS k and $\boldsymbol{x}_{lg} = \boldsymbol{D}_{lg} \mathbf{w}_{lg} \varsigma_g$, with $\mathbf{w}_{lg} \in \mathbb{C}^N$ representing the precoding vector used by AP l towards multicast subgroup g, ς_g denotes the data symbol intended for MSs in subgroup g, with $\mathbb{E}\{|\varsigma_g|^2\} = 1$, and $\mathbb{E}\{\varsigma_g \varsigma_c^*\} = 0$, $\forall \ g \neq c$, and $n_k \sim \mathcal{CN}(0, \sigma_d^2)$ is the AWGN at MS k. The set of diagonal matrices $\boldsymbol{D}_{lg} \in \mathbb{C}^{N \times N}$ are used to describe which APs communicate with which multicast subgroups [18], and are given by $\boldsymbol{D}_{lg} = \boldsymbol{I}_N$ if $l \in \mathcal{L}_g$, or $\boldsymbol{D}_{lg} = \boldsymbol{0}_{N \times N}$, otherwise. Note that, assuming that MS k belongs to multicast subgroup g, the first term in (8) denotes the desired signal, whereas the second term is the inter-subgroup interference.

As we do not transmit DL pilots, the achievable SE of MS $k \in \mathcal{K}_g$ must be obtained assuming that it only knows the average value of its effective DL channel, that is, $\sum_{l=1}^{L} \mathbb{E}\left\{ \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{D}_{lg} \mathbf{w}_{lg} \right\}$ of MS $k \in \mathcal{K}_g$. This is a deterministic number that, thanks to *channel hardening*, can be easily obtained in practice [13, Theorem 6.1], and the corresponding achievable SE can then be obtained as

$$\mathrm{SE}_k = (1 - \tau_p/\tau_c) \log_2 (1 + \gamma_k), \qquad (9)$$

where γ_k is the effective signal-to-interference-plus-noise ratio (SINR) given by

$$\gamma_k = \frac{\left| \sum_{l=1}^{L} \mathbb{E}\left\{ \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{D}_{lg} \mathbf{w}_{lg} \right\} \right|^2}{\sum_{c=1}^{G} \mathbb{E}\left\{ \left| \sum_{l=1}^{L} \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{D}_{lc} \mathbf{w}_{lc} \right|^2 \right\} - \left| \sum_{l=1}^{L} \mathbb{E}\left\{ \boldsymbol{h}_{lk}^{\mathsf{H}} \boldsymbol{D}_{lg} \mathbf{w}_{lg} \right\} \right|^2 + \sigma_d^2}, \qquad (10)$$

and the expectations are with respect to the channel realizations and the channel estimates upon which the precoding vectors are designed.

This is an achievable SE for MS k in subgroup g and holds for any precoding strategy and any multicast DCC approach. Nevertheless, we stress that a subgroup is served by a single multicast transmission, which determines a shared DL SE for all the MSs in the subgroup. For the data to be able to be reliably decoded by all the MSs in the subgroup, the achievable SE of the multicast subgroup (and thus all MSs in this subgroup) will be that of the MS experiencing the worst channel conditions in the subgroup. That is, the effective DL SE achievable by all the MSs in subgroup g is

$$\mathrm{SE}_g = \min_{k \in \mathcal{K}_g} \mathrm{SE}_k. \qquad (11)$$

F. Downlink Precoding and Power Allocation

We consider two scalable precoding strategies for CF-mMIMO: a centralized IP-MMSE precoding strategy [12] and a distributed CB precoder [2]. Both schemes consider that only the APs in $\mathcal{L}_g$ are computing estimates of the channels for MSs in subgroup g and/or send their pilot signals to the CPU.

1) Centralized IP-MMSE precoding: We define a composite channel estimate vector per subgroup as $\check{\boldsymbol{h}}^g = \boldsymbol{D}_g \hat{\boldsymbol{h}}^g$, where $\hat{\boldsymbol{h}}^g = \left[\hat{\boldsymbol{h}}_1^{g,\mathsf{T}} \dots \hat{\boldsymbol{h}}_L^{g,\mathsf{T}} \right]^{\mathsf{T}}$ and $\boldsymbol{D}_g = \mathrm{blkdiag}(\boldsymbol{D}_{1g}, \dots, \boldsymbol{D}_{Lg}) \in \mathbb{C}^{LN \times LN}$. The CPU uses the composite channel estimates to compute the composite precoding vectors. Since all the MSs belonging to subgroup g employ the same pilot, the CPU can easily exploit $\check{\boldsymbol{h}}^g$ to obtain a scaled vector that points out the direction of the precoding vector. Based on a *virtual* multicast UL-DL duality[2], the IP-MMSE combiner [12] is given by

$$\bar{\mathbf{w}}_g = \sqrt{\frac{p_g}{\tau_p P_p}} K_g \left(\sum_{c \in \mathcal{S}_g} \frac{p_c K_c^2}{\tau_p P_p} \check{\boldsymbol{h}}^c (\check{\boldsymbol{h}}^c)^{\mathsf{H}} + \boldsymbol{Z}_{\mathcal{S}_g} + \sigma_u^2 \boldsymbol{I}_{L_g N} \right)^{-1} \check{\boldsymbol{h}}^g,$$

where p_g denotes the total amount of power that would be allocated to MSs in subgroup g in a *virtual* UL payload transmission phase and

$$\boldsymbol{Z}_{\mathcal{S}_g} = \sum_{c \in \mathcal{S}_g} \frac{p_c}{\tau_p P_p} K_c^2 \boldsymbol{D}_g \tilde{\boldsymbol{R}}^c \boldsymbol{D}_g + \sum_{d \notin \mathcal{S}_g} \frac{p_d}{\tau_p P_p} K_d^2 \boldsymbol{D}_g \boldsymbol{R}^d \boldsymbol{D}_g,$$

where $\tilde{\boldsymbol{R}}^c = \mathrm{blkdiag}(\tilde{\boldsymbol{R}}_1^c, \dots, \tilde{\boldsymbol{R}}_L^c) \in \mathbb{C}^{LN \times LN}$ denotes the error correlation matrix of the composite channel $\boldsymbol{h}_c$ and $\boldsymbol{R}^d = \mathrm{blkdiag}(\boldsymbol{R}_1^d, \dots, \boldsymbol{R}_L^d) \in \mathbb{C}^{LN \times LN}$ the covariance matrix of the composite channels of the interfering groups.

The centralized precoding vector used to multicast data to MSs in subgroup g can be expressed as

$$\mathbf{w}_g = \sqrt{\rho_g} \frac{\bar{\mathbf{w}}_g}{\sqrt{\mathbb{E}\{\|\bar{\mathbf{w}}_g\|^2\}}}, \qquad (12)$$

where $\rho_g \geq 0$ denotes the DL transmit power allocated to subgroup g. Note that the normalization in (12) guarantees that $\mathbb{E}\{\|\mathbf{w}_g\|^2\} = \rho_g$ [13].

We propose an *inter-subgroup* fractional DL power control to select the power allocation coefficients proportionally to the

---

[2]UL-DL duality on the channels defined by (4) is utilized to design the *virtual* combiners.

trace of $\boldsymbol{R}_l^g$. To satisfy the power constraint at each AP, the power allocated to subgroup g is obtained as

$$\rho_g = P_{\text{dl}} \frac{\left[\sum\limits_{l \in \mathcal{L}_g} \text{tr}\left(\boldsymbol{R}_l^g\right)\right]^\nu \omega_g^{-\kappa}}{\max\limits_{\ell \in \mathcal{L}_g} \sum\limits_{c \in \mathcal{D}_\ell} \left[\sum\limits_{l \in \mathcal{L}_c} \text{tr}\left(\boldsymbol{R}_l^g\right)\right]^\nu \omega_c^{1-\kappa}}, \quad (13)$$

where P_{dl} is the total amount of power an AP can transmit, $\nu \in [-1, 1]$ is the parameter tuning the power allocation in accordance to different policies, i.e., $\nu < 0$ aims at max-min fairness (MMF) characteristics. We also define the largest fraction of ρ_g that can be used at any of the serving APs as

$$\omega_g = \max\limits_{l \in \mathcal{L}_g} \mathbb{E}\{\|\bar{\mathbf{w}}_{lg}\|^2\}, \quad (14)$$

which is used as an additional tuning parameter with an exponent $0 \le \kappa \le 1$ that reshapes the ratio of power allocation between different subgroups, where $\bar{\mathbf{w}}_g = \left[\bar{\mathbf{w}}_{1g} \ldots \bar{\mathbf{w}}_{Lg}\right]$.

2) Distributed CB precoding: Distributed operation offers the benefit of deploying new APs without having to upgrade the computational power of the CPU since each AP contains a local processor that can perform its associated baseband processing tasks and locally designs its transmitted signal [13]. The distributed CB precoding is given by

$$\mathbf{w}_{lg} = \sqrt{\rho_{lg}} \frac{\boldsymbol{D}_{lg} \hat{\boldsymbol{h}}_l^g}{\sqrt{\mathbb{E}\{\|\boldsymbol{D}_{lg} \hat{\boldsymbol{h}}_l^g\|^2\}}}, \quad (15)$$

with $\mathbb{E}\{\|\boldsymbol{D}_{lg} \hat{\boldsymbol{h}}_l^g\|^2\} = K_g^2 \text{tr}(\boldsymbol{D}_{lg} \boldsymbol{R}_l^g \boldsymbol{\Gamma}_g^{-1} \boldsymbol{R}_l^g \boldsymbol{D}_{lg})$. To satisfy the per-AP power constraint, we consider the following adaptive power allocation (APA) policy [19]

$$\rho_{lg} = \begin{cases} P_{dl} \dfrac{\left[\text{tr}\left(\boldsymbol{R}_l^g\right)\right]^\nu}{\sum_{g \in \mathcal{D}_l} \left[\text{tr}\left(\boldsymbol{R}_l^g\right)\right]^\nu}, & \text{if } g \in \mathcal{D}_l \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

III. Numerical Results

We consider a CF-mMIMO network with $L = 100$ APs, each one equipped with $N = 4$ antennas, uniformly distributed at random within a square coverage area of side 1000 m. To approximate a coverage area without boundaries, the nominal area is wrapped-around by 8 identical neighbor replicas. The path-loss in dB is given by $-30.5 - 36.7\log_{10}(d) + F$, where d is the 3D distance between the MS and the AP and F is the shadow fading, whose standard deviation is 4 dB, and the decorrelation distance is 9 m [2]. We consider a coherence block of $\tau_c = 200$ samples and a maximum pilot length of $\tau_p = 20$ samples. The pilot transmit power per MS is $P_p = 100$ mW, while the maximum DL power per AP is $P_{dl} = 200$ mW. The *virtual* UL power per subgroup to build the centralized IP-MMSE precoder [12] is $p_g = 100$ mW. The DL power control parameters are set targeting user fairness: $\nu = -0.5$ and $\kappa = 0.5$ for centralized IP-MMSE precoding, and $\nu = 0.5$ for distributed CB precoding [13]. The angular standard deviation in the local scattering model is $15°$. We consider a noise power spectral density of -174 dBm/Hz, a receiver noise figure of 7 dB, and an operating bandwidth of 20 MHz

[13]. Each simulation result has been obtained as the average of 250 different snapshots of randomly deployed MSs and APs with 500 channel realizations for each snapshot.

Figure 2 illustrates the cumulative distribution function (CDF) of the sum SE achieved by 100 and 500 independently and uniformly distributed random multicast MSs within the coverage area. We evaluate the performance of either unicast, single multicast or multicast subgrouping transmissions. The results show that when the multicast MSs are uniformly distributed, employing unicast transmission, where the channel estimation and the precoding are reasonably accurate despite the pilot contamination (i.e., $\tau_p = 20$ and $K = [100, 500]$ in Figs. [2a, 2b], respectively), performs significantly better than any option employing multicast transmission, under both centralized IP-MMSE and distributed CB precoding. Irrespective of the precoder, increasing the number of multicast subgroups improves the performance, but the highest sum SE is achieved by unicast transmission. Furthermore, note that the sum SE achieved using IP-MMSE is significantly higher than that achieved with CB, thus centralized IP-MMSE with unicast transmissions will deliver the highest sum SE when the multicast MSs are uniformly distributed.

To validate the utilization of multicast transmissions, we deploy scenarios where the multicast MSs are located in square cluster areas of side 10 m, thus resulting in bunches of users located very close to one another. This situation can extremely affect the channel estimation and the precoding because of the pilot contamination. Figure 3 shows the CDF of the sum SE of the multicast service when the MSs are placed in 10 spatial clusters of 50 MSs each. We observe how the unicast transmissions are severely degraded because of the strong effect of pilot contamination among MSs placed in the same spatial cluster (i.e., 50 MSs and $\tau_p = 20$ orthogonal pilot sequences). We also notice that using a single multicast transmission does not lead to the best performance, and creating multicast subgroups outperforms both unicast and single multicast transmissions. Fig. 3a shows that centralized IP-MMSE precoding, when transmitting to a large number of multicast subgroups, results in the highest sum SE. Remarkably, this strategy tends to approach unicast transmission while preventing pilot contamination among users located in the same spatial cluster (i.e., 100 multicast subgroups). IP-MMSE allows the systems to treat the interference from very close subgroups, and the only reason to not use unicast is the strong pilot contamination from nearby users. In contrast, Fig. 3b reveals that when using CB precoding, splitting the MSs into 30 subgroups presents the best trade-off between signal and interference. Indeed, a smaller number of subgroups deteriorates the desired signal due to less accurate channel estimates while a larger number of subgroups increases the interference received from close subgroups.

IV. Conclusion

This work proposes a novel framework to assess the performance of scalable multicast techniques in CF-mMIMO networks when using different precoders and power allocations. A subgrouping technique has been introduced whereby multicast MSs are separated based on their spatial location aiming at improving the performance of the multicast service. Results
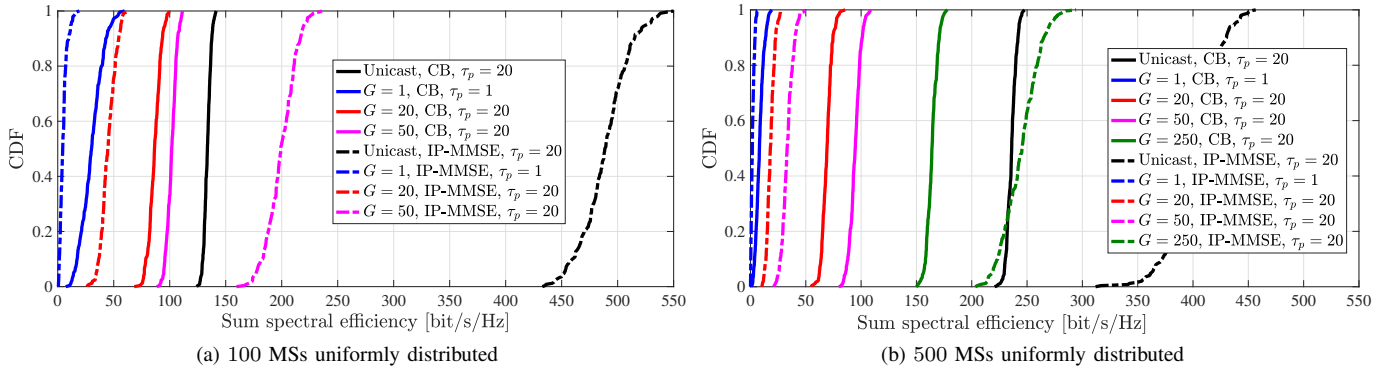
(a) 100 MSs uniformly distributed

(b) 500 MSs uniformly distributed

Fig. 2. CDF of the sum SE. Uniform distribution of multicast MSs. Unicast vs multicast with CB and IP-MMSE precoding. $L=100$ APs, $N=4$
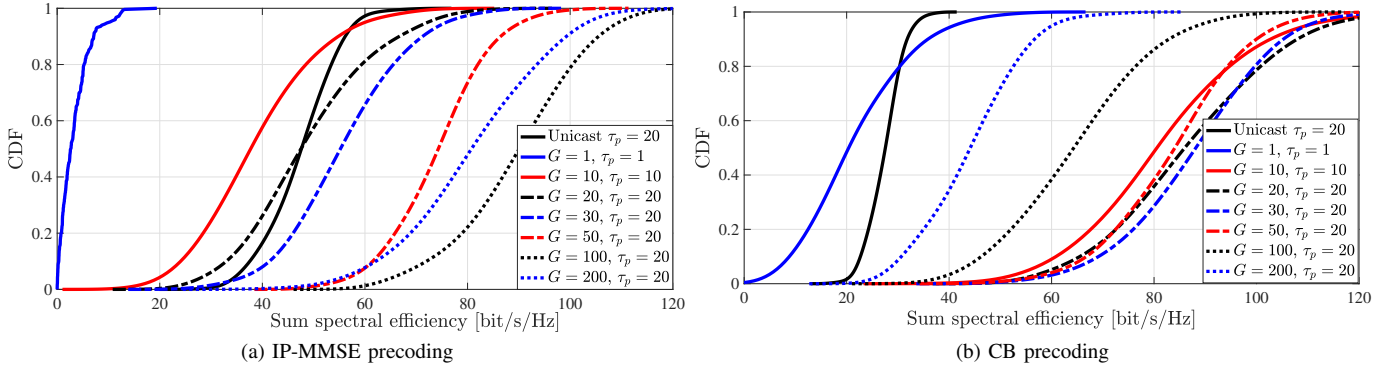

(a) IP-MMSE precoding

(b) CB precoding

Fig. 3. CDF of the sum SE. Clustered distribution of multicast MSs with 10 spatial clusters of 50 MSs. Unicast vs multicast with CB and IP-MMSE precoding. $L=100$ APs, $N=4$.

have shown that unicast transmissions are preferable when the multicast MSs are uniformly distributed. On the contrary, when the multicast MSs tend to form spatial clusters, unicast transmissions are severely degraded by pilot contamination, while multicasting is able to sustain considerable higher rates. Further research will focus on the assessment of heterogeneous deployments where uniform placement of users is combined with the existence of hotspots with larger user density.

REFERENCES

[1] H. Q. Ngo, G. Interdonato, E. G. Larsson, G. Caire, and J. G. Andrews, "Ultradense cell-free massive MIMO for 6G: Technical overview and open questions," *Proceedings of the IEEE*, pp. 1–27, 2024, early access.

[2] 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, 2017.

[3] Ericsson, "Ericsson Mobility Report June 2024," White Paper, Jun 2024, [Online] Available at: https://www.ericsson.com/en/reports-and-papers/mobility-report.

[4] A. de la Fuente, R. P. Leal, and A. G. Armada, "New Technologies and Trends for Next Generation Mobile Broadcasting Services," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 217–223, Nov 2016.

[5] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over Emerging 5G Networks: Challenges and Perspectives," *IEEE Netw.*, vol. 31, no. 2, pp. 80–89, March 2017.

[6] M. Sadeghi, E. Björnson, E. G. Larsson, C. Yuen, and T. L. Marzetta, "Max–Min Fair Transmit Precoding for Multi-Group Multicasting in Massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1358–1373, Feb 2018.

[7] A. de la Fuente, G. Interdonato, and G. Araniti, "User Subgrouping and Power Control for Multicast Massive MIMO Over Spatially Correlated Channels," *IEEE Trans. Broadcast.*, pp. 1–14, 2022.

[8] T. X. Doan, H. Q. Ngo, T. Q. Duong, and K. Tourki, "On the Performance of Multigroup Multicast Cell-Free Massive MIMO," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2642–2645, 2017.

[9] Y. Zhang, H. Cao, and L. Yang, "Max-Min Power Optimization in Multigroup Multicast Cell-Free Massive MIMO," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, 2019, pp. 1–6.

[10] G. Araniti, M. Condoluci, L. Militano, and A. Iera, "Adaptive Resource Allocation to Multicast Services in LTE Systems," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 658–664, Dec 2013.

[11] A. de la Fuente, G. Femenias, F. Riera-Palou, and A. G. Armada, "Subband CQI Feedback-Based Multicast Resource Allocation in MIMO-OFDMA Networks," *IEEE Trans. Broadcast.*, vol. 64, no. 4, pp. 846–864, Dec 2018.

[12] G. Femenias, F. Riera-Palou, and E. Björnson, "Another Twist to the Scalability in Cell-Free Massive MIMO Networks," *IEEE Trans. Commun.*, vol. 71, no. 11, pp. 6793–6804, 2023.

[13] Ö. T. Demir, E. Björnson, L. Sanguinetti *et al.*, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021.

[14] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. Asilomar Conf. Signals, Syst. and Comput.*, Nov. 2016, pp. 974–978.

[15] F. Riera-Palou, G. Femenias, A. G. Armada, and A. Pérez-Neira, "Clustered Cell-Free Massive MIMO," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.

[16] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge University Press, 2016.

[17] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

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

[19] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability Aspects of Cell-Free Massive MIMO," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2019, pp. 1–6.