Cluster Formation in Scalable Cell-free Massive MIMO Networks

Charmae Franchesca Mendoza†, Stefan Schwarz† and Markus Rupp
†Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion
Institute of Telecommunications, Technische Universität (TU) Wien
Email: {charmae.mendoza,stefan.schwarz,markus.rupp}@tuwien.ac.at

***Abstract*—Inter-cell interference remains to be a bottleneck for conventional cellular networks as cell-edge users continue to suffer from poor performance. Cell-free massive MIMO is a novel network architecture that suppresses inter-cell interference by eliminating cell boundaries. It promises uniform performance throughout the coverage area, enabled by the coherent joint transmission from multiple distributed antennas. To make the network scalable, a user-centric approach is adopted where each user is served by a cluster of nearby access points (APs). In this work, we study the impact of cluster formation on the total fronthaul requirement, which is a limiting factor in practical coordinated distributed systems. We also investigate its effect on the guaranteed quality of service (QoS) of the network. Using our proposed algorithms, we look into the optimal cluster sizes for a given scenario and show that good performance can be achieved even with relatively small user-centric clusters, which then translates to fronthaul savings.**

***Index Terms*—cell-free massive MIMO, user-centric, clustering, fronthaul, max-min SINR, quality of service**

I. Introduction

The continuous growth of data traffic has been a driving force behind the technological advancements in wireless communications. To cope with the traffic demand, the capacity of a mobile network is increased by allocating more bandwidth, utilizing the spectrum more efficiently and deploying smaller cells (network densification). A key technology for higher spectral efficiency is massive multiple-input multiple-output (MIMO) [1], where a base station is equipped with an antenna array for spatially multiplexing a small number of users on the same time-frequency resources.

One implication of network densification is the increased inter-cell interference that degrades performance. To combat this, several mitigation and cooperation techniques have been proposed, including coordinated multipoint beamforming [2] and joint transmission (CoMP-JT) [3]. In CoMP-JT, the cellular network is virtually divided into disjoint clusters of cooperating base stations that jointly transmit to users. In fact, CoMP-JT and concepts such as network MIMO [4] and distributed antenna system (DAS) [5] are examples of distributed massive MIMO [6] (i.e., spatially distributed antenna arrays). In [7], we have introduced dynamic DAS (dDAS) that enables dynamic allocation of remote radio units (RRUs) among the base stations to support changing traffic load conditions. These coordinated distributed architectures improve spectral efficiency and provide higher coverage probability, which enhances reliability for dependable wireless communications [8]. However, despite the performance gains, inter-cell interference still persists that is inherent in the cell-centric design of conventional networks [9].

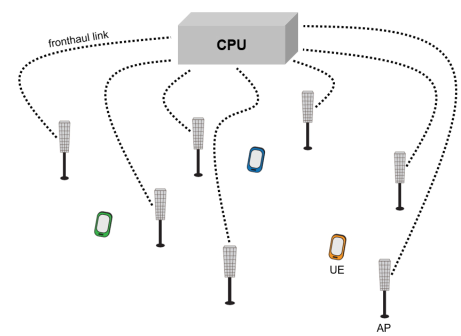

Fig. 1. Canonical cell-free massive MIMO network.

Cell-free massive MIMO is a novel concept introduced in [10] that suppresses inter-cell interference by removing the cell boundaries. It consists of a large number of geographically distributed access points (APs) simultaneously serving a smaller number of users (UEs) using the same resources. Fig. 1 illustrates its *canonical* form where each user is served by all the APs, and each AP is connected via the fronthaul link to a single centralized processing unit (CPU) that is mainly used for coordination and data processing. Following the assumptions in massive MIMO, the coherent joint transmission from multiple antennas enables good performance even with simple linear processing techniques (such as Maximum Ratio Transmission (MRT)) [1]. By bringing the APs closer to the users, the cell-free network can provide uniform performance throughout the coverage area [10]. This is particularly beneficial to cell-edge users who suffer from high levels of interference in a conventional cellular network. In [11], a performance comparison of centralized and distributed implementations is presented.

In this paper, we first discuss the *scalability problem* [9] to motivate the shift from canonical to user-centric [12] cell-free networks. In the latter architecture, user-centric clusters are formed such that only a subset of APs serves each user. Compared to prior studies [12] [13], we focus on the impact of cluster formation on practical aspects of the network, including fronthaul requirements and guaranteed quality of service (QoS) level by formulating corresponding optimization problems.

II. User-centric Cell-free Massive MIMO

The canonical form of cell-free massive MIMO assumes that all APs are connected to a single CPU. The number of fronthaul connections therefore increases proportionally as more APs are added to the network. Furthermore, with each AP serving all users, the overall fronthaul capacity requirement also increases. These factors give rise to scalability issues and thus, the canonical form is impractical and hardly feasible given the large number of APs ideally deployed in a cell-free network.

The scalability problem is addressed by forming what are known as *user-centric* clusters [12]. The idea is to have each user select the APs that give the best performance. That is, each user is jointly served by only a subset of nearby APs, depicted in Fig. 2. This is reasonable as the APs far away from the user have little or negligible impact on its performance. Such a user-centric approach directly translates to fronthaul savings since with fewer users being served by each AP, less information needs to be exchanged over the fronthaul links.

To make the cell-free network truly scalable, multiple interconnected CPUs may be employed [9]. Each CPU is assigned a number of APs, resulting in disjoint AP-clusters. This is illustrated in Fig. 2 where each color represents fronthaul connections to a particular CPU. Moreover, a user-centric cluster may consist of APs controlled by different CPUs. In this case, the CPUs must coordinate with each other via the backhaul link for the joint processing and transmission/reception to/from the user. Different degrees of interconnectivity among the CPUs are studied in [14].

A. System Model

We consider a downlink cell-free massive MIMO network with M APs, each having N_t antennas, and K single-antenna UEs ($MN_t \gg K$).

For the channel model, the macroscopic channel gain $g_{k,m}$ between AP m and UE k follows a distance-dependent path loss model with shadow fading as:

$$g_{k,m} = \left(\frac{\lambda_c}{4\pi}\right)^2 \left(\frac{1}{d_{k,m}}\right)^{n_c} s_{k,m}, \tag{1}$$

where λ_c denotes the wavelength of the carrier frequency, $d_{k,m}$ is the distance between UE k and AP m, n_c denotes the path loss exponent and $s_{k,m} \sim \mathcal{LN}(0, \sigma_c^2)$ is the random lognormally distributed shadow fading. We consider an independent and identically distributed (iid) Rayleigh fading channel between AP m and UE k as:

$$\mathbf{h}_{k,m} = \sqrt{g_{k,m}}\,\widetilde{\mathbf{h}}_{k,m} \in \mathbb{C}^{N_t \times 1}, \tag{2}$$

with $\widetilde{\mathbf{h}}_{k,m} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t})$ denoting the small-scale fading.

In the downlink data transmissions, we assume that each AP knows the channels to the UEs perfectly (i.e., uplink training for channel estimation is not considered in this work) and performs MRT, with $\mathbf{F}_m \in \mathbb{C}^{N_t \times K}$ being the precoder matrix of AP m and column $\mathbf{f}_{k,m} \in \mathbb{C}^{N_t \times 1}$ corresponding to the beamforming vector of UE k.

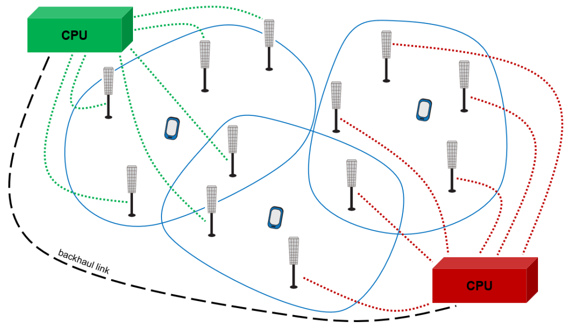

Fig. 2. User-centric cell-free network with multiple interconnected CPUs.

The input-output relationship of UE k is given by:

$$y_k = \sum_{m=1}^{M} \sqrt{P_{k,m}}\mathbf{h}_{k,m}^H \mathbf{F}_m \mathbf{s}_m \delta_{k,m} + z_k, \tag{3}$$

where $P_{k,m}$ denotes the power allocated to UE k by AP m, $\delta_{k,m}$ is a binary variable indicating whether an AP serves a user or not (i.e., $\delta_{k,m} = 1$ if UE k is served by AP m and $\delta_{k,m} = 0$ otherwise)[1], $\mathbf{s}_m \in \mathbb{C}^{K \times 1}$ includes the data symbols for the users such that $\mathbb{E}(\mathbf{s}_m \mathbf{s}_m^H) = \mathbf{I}_K$ and z_k denotes the receiver noise with variance σ_z^2.

The signal to interference and noise ratio (SINR) of UE k is expressed as:

$$\mathrm{SINR}_k = \frac{\left|\sum_{m=1}^{M} \sqrt{P_{k,m}}\mathbf{h}_{k,m}^H \mathbf{f}_{k,m} \delta_{k,m}\right|^2}{\sigma_z^2 + \sum_{\substack{j=1 \\ j \neq k}}^{K} \left|\sum_{m=1}^{M} \sqrt{P_{j,m}}\mathbf{h}_{k,m}^H \mathbf{f}_{j,m} \delta_{j,m}\right|^2}. \tag{4}$$

We calculate the signal power by coherently adding the contributions from the APs forming the user-centric cluster. The only sources of interference are the other users in the cell-free network. Finally, the spectral efficiency (SE) of UE k is:

$$\mathrm{SE}_k = \log_2(1 + \mathrm{SINR}_k). \tag{5}$$

B. Canonical versus User-centric

User-centric clustering is key to the realization of scalable cell-free massive MIMO networks. Here, we present a simple example to get an initial insight on the impact of such clustering on performance.

We consider a downlink scenario with 50 APs, each having 10 antennas, and 10 single-antenna UEs, uniformly distributed in a square area of 100x100 m^2. In Fig. 3, we compare the achievable spectral efficiency per user for the canonical and user-centric cases. We also employ different cluster sizes for the latter. We form the clusters by having each UE select X APs with the strongest channel. For instance, UC 25% implies that each cluster consists of 25% of the total number of APs in

[1] When UE k is not served by AP m, the corresponding power $P_{k,m}$ and beamforming vector $\mathbf{f}_{k,m}$ are set to 0.

the network. Note that a different criterion may be used in the cluster formation. However, for our purpose, the AP selection is solely based on the macroscopic channel gain.

In this section, we specifically assume equal power allocation among the UEs served by an AP (i.e., power is only shared among the UEs with $\delta_{k,m} = 1$). The sum of UE powers is equal to the maximum transmit power per AP as:

$$\sum_{k=1}^{K} P_{k,m}\delta_{k,m} = P_{Tx,m}, \tag{6}$$

with $P_{Tx,m}$ being the total transmit power of AP m. More sophisticated power allocation schemes exist that may result to different performance plots. However, the goal of this example is to show that we can get comparable performance even when each user is served by only a subset of APs and thus, further motivate us to adopt the user-centric approach.

In Fig. 3, we show the empirical cumulative distribution function (ECDF) of the per-user spectral efficiency and observe that user-centric clustering outperforms the canonical case. In fact, as we make the clusters smaller, we see further improvement on performance. A smaller cluster size means that fewer APs make up the individual clusters and therefore, each AP is likely to serve a smaller number of UEs. Assuming equal power allocation, this implies that the total transmit power per AP is shared among fewer UEs, which contributes to the increase in signal power and SINR improvement for those users. However, using optimized power allocation strategies, the canonical form is more likely to perform better than the user-centric approach.

We also investigate the single-AP case where each UE only connects to the best AP in the network. The effect of random positions is more pronounced in this case since the performance highly depends on the distance between the UE and the selected AP. In contrast, the achievable spectral efficiency for the other cases does not vary that much, which suggests that different users experience a more uniform performance. This is not true for the single-AP case, where some users experience good or bad service depending on their position. This highlights the benefits of coherent joint transmission from multiple APs when compared to being served by only the AP with the best channel [10].

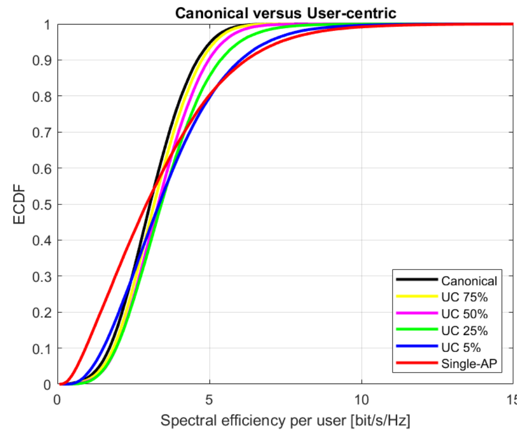

Fig. 3. ECDF of the per-user spectral efficiency for the canonical and user-centric cases assuming equal power allocation.

III. Practical Aspects of a Cell-free Network

We previously saw that applying the principle of user-centric clustering can lead to uniformly good performance. In this section, we focus on understanding how the cluster formation relates to other aspects of the cell-free massive MIMO network, namely the overall fronthaul requirement and the guaranteed QoS level.

A. Fronthaul Optimization

One of the limiting factors in the realization of coordinated distributed systems is the fronthaul capacity. As such, we want to minimize the total fronthaul requirement in a cell-free network by formulating the following optimization problem:

$$\begin{aligned} \min \quad & \sum_{k=1}^{K}\sum_{m=1}^{M} \delta_{k,m} \qquad (7)\\ \text{w.r.t.} \quad & \delta_{k,m} \in \{0,1\} \\ & P_{k,m} \geq 0 \\ \text{s.t.} \quad & \sum_{k=1}^{K} P_{k,m}\delta_{k,m} \leq P_{Tx,m}, \forall m \in \{1,\ldots,M\} \\ & \text{SINR}_k \geq \text{SINR}_{min}, \forall k \in \{1,\ldots,K\}. \end{aligned}$$

The goal is to minimize the sum of the elements of a matrix consisting of the indicator variables $\delta_{k,m}$, which in effect, minimizes the fronthaul requirements. Note that in this formulation, we consider neither the number of bits nor the type of information communicated via the fronthaul links, but rather, we count the number of AP-UE connections. This is reasonable as the fronthaul requirement increases with the number of data streams that needs to be transmitted using the same resources. Moreover, this relates to the cluster formation since with fewer AP-UE connections, smaller clusters are formed and therefore, less information needs to be exchanged by the APs and the CPU over the fronthaul links.

In a downlink scenario, each AP divides its power among the UEs it serves. The per-AP power constraint ensures that the sum of the non-negative UE powers does not exceed the maximum transmit power of the AP, which we assume to be the same for all APs for simplicity. Furthermore, we impose a per-UE SINR constraint where we define a threshold SINR_{min}. This constraint allows us to check if all users in the network experience good performance as defined by the threshold.

The SINR constraint in Problem (7) can be reformulated following a similar approach as in [7]. We can therefore solve it using the CVX framework [15] and solvers such as MOSEK [16]. However, this problem formulation does not scale well with the number of APs and users. As the network gets larger, the optimization problem becomes more difficult to solve. The original mixed-integer linear programming (MILP) problem is in fact NP-hard and thus, we apply linear programming (LP) relaxation to make the problem solvable in polynomial time. We remove the integer requirement for the binary variable $\delta_{k,m}$ and allow continuous values such that $\delta_{k,m} \in [0,1]$. The resulting convex problem is readily solvable even for a large scenario with hundreds of APs and users.

We relate the LP relaxation solution back to the original MILP problem by utilizing a threshold $\alpha \in [0,1]$ as:

$$\delta_{k,m}^{MILP} = \begin{cases} 1 & \delta_{k,m}^{LP} \geq \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We obtain $\delta_{k,m}^{LP}$ by solving the relaxed problem. Applying (8) gives $\delta_{k,m}^{MILP}$ that we use to derive a suboptimal solution to the original problem. When $\delta_{k,m}^{LP}$ is at least α, we decide that UE k is served by AP m and set $\delta_{k,m}^{MILP}$ to 1. Otherwise, we set it to 0 implying that no connection between UE k and AP m is established. Note that for a minimization problem, the relaxation solution gives a lower bound on the solution to the original problem.

When interpreting $\delta_{k,m}^{LP}$, it is possible to get an infeasible solution to the original problem, which means one or all constraints are violated. It may also be the case that we get a feasible solution but not the optimal one. Thus, we want the α that gives us not only a feasible solution but also the smallest value for the objective function. One way to determine the best α is to sweep over its possible values [17]. We initially set α to 0, then we iterate over its possible values, each time incrementing it by a predefined step size until we reach α equals 1. For every iteration, we interpret the LP relaxation solution based on the current α following (8). The resulting $\delta_{k,m}^{MILP}$ is used to compute the corresponding value of the objective function and check if the SINR constraint is satisfied for all users. After all iterations, we know which α values result to a feasible solution. From these, we select the one that gives the optimal value for the objective function (or the smallest fronthaul requirement). Algorithm 1 summarizes the steps to approximately solve Problem (7).

Algorithm 1 Generic approach to solve Problem (7)

```
Solve the LP relaxation problem
for (α = 0; α ≤ 1; α = α + step_size) do
    if δ^{LP} ≥ α then
        δ^{MILP} ← 1
    else
        δ^{MILP} ← 0
    end if
    Compute objective function using δ^{MILP}
    Compute SINR(k) ∀k ∈ {1, ..., K} using δ^{MILP}
    SINR_check ← max_k(SINR_min − SINR)
end for
Find feasible points (SINR_check ≤ 0)
Select the point with smallest objective
```

B. Max-min SINR Optimization with Cluster Size Constraint

The main selling point of the cell-free massive MIMO concept is its claim of being able to provide uniformly good service to everyone regardless of their geographical location. Related to this, we are interested in finding the highest QoS level that a cell-free network can guarantee to all users by formulating the following maximization problem:

$$\begin{aligned} \max \quad & \min_k \mathrm{SINR}_k \\ \text{w.r.t.} \quad & \delta_{k,m} \in \{0,1\} \\ & P_{k,m} \geq 0 \\ \text{s.t.} \quad & \sum_{k=1}^{K} P_{k,m}\delta_{k,m} \leq P_{Tx,m}, \forall m \in \{1,\ldots,M\} \\ & \sum_{m=1}^{M} \delta_{k,m} \leq csize_{max}, \forall k \in \{1,\ldots,K\}. \end{aligned} \tag{9}$$

The goal is to maximize the threshold SINR_{min}, which is the worst SINR value for any user in the network. The larger the value of this threshold, the higher is the guaranteed QoS. For this problem, we also impose a constraint on the cluster size. Specifically, $csize_{max}$ is the maximum number of APs available to each user. By doing so, we can study the impact of cluster formation on the achievable SINR values. The motivation for this is while we ideally want users to experience the highest possible QoS level, we are often restricted by the limited fronthaul capacity in practical deployments.

Problem (9) can be written as:

$$\begin{aligned} \max \quad & \mathrm{SINR}_{min} \\ \text{w.r.t.} \quad & \delta_{k,m} \in \{0,1\} \\ & P_{k,m} \geq 0 \\ \text{s.t.} \quad & \sum_{k=1}^{K} P_{k,m}\delta_{k,m} \leq P_{Tx,m}, \forall m \in \{1,\ldots,M\} \\ & \sum_{m=1}^{M} \delta_{k,m} \leq csize_{max}, \forall k \in \{1,\ldots,K\} \\ & \mathrm{SINR}_k \geq \mathrm{SINR}_{min}, \forall k \in \{1,\ldots,K\}. \end{aligned} \tag{10}$$

Problem (10) can be reformulated following a similar approach as in [7]. We can then solve it using the bisection method, where at each step, we solve a feasibility problem [17] as indicated in lines 1 to 10 of Algorithm 2.

We again apply LP relaxation such that $\delta_{k,m} \in [0,1]$. The solution to the convex relaxed problem gives $\delta_{k,m}^{LP}$, interpreted as the probability that AP m serves UE k. Thus, the larger the value of $\delta_{k,m}^{LP}$, the more likely an AP-UE connection is established. Since the cluster size constraint dictates the number of APs forming each user-centric cluster, we use this probability as a mechanism to select which APs to activate when going back to the original problem. This is done systematically by sorting the $\delta_{k,m}^{LP}$ values in descending order for each user, then activating one AP at a time starting from those with the highest probabilities until the number of activated APs reaches

$csize_{max}$. The resulting $\delta^{MILP}_{k,m}$ is set to 1 to denote that AP m is activated with respect to UE k. Otherwise, it is set to 0. We then use the $\delta^{MILP}_{k,m}$ values in determining the maximum SINR level guaranteed to all users by the cell-free network. This procedure corresponds to lines 11 to 21 of Algorithm 2. We also note that for a maximization problem, the relaxation solution gives an upper bound on the solution to the original problem.

Algorithm 2 Generic approach to solve Problem (10)

1: Initialize t_{min} and t_{max} (range of values for SINR_{min}) and bisection tolerance $\epsilon > 0$
2: **while** $t_{max} - t_{min} \geq \epsilon$ **do**
3: $\quad t \leftarrow \frac{t_{max}+t_{min}}{2}$
4: $\quad$ Perform feasibility test
5: $\quad$ **if** feasible **then**
6: $\quad\quad t_{min} \leftarrow t$
7: $\quad$ **else**
8: $\quad\quad t_{max} \leftarrow t$
9: $\quad$ **end if**
10: **end while**
11: **for** $k = 1$ to K **do**
12: $\quad$ Sort $\delta^{LP}(k,:)$ in descending order
13: $\quad$ Store in $[val, ap_{idx}]$
14: $\quad$ **for** $m = 1$ to M **do**
15: $\quad\quad$ **if** $sum(\delta^{MILP}(k,:)) < csize_{max}$ **then**
16: $\quad\quad\quad \delta^{MILP}(k, ap_{idx}(m)) \leftarrow 1$
17: $\quad\quad$ **end if**
18: $\quad$ **end for**
19: **end for**
20: Compute $\text{SINR}(k)\ \forall k \in \{1, \ldots, K\}$ using δ^{MILP}
21: Max $\text{SINR}_{min} \leftarrow \min_k(\text{SINR})$

IV. Simulations

In our simulations, we consider a downlink cell-free massive MIMO network with fixed number of $M = 50$ APs, each having $N_t = 10$ antennas. We vary the number of users $K \in \{2, \ldots, 15\}$. The APs and UEs are uniformly distributed in a square area of 100x100 m^2. We summarize the simulation parameters in Table I and select the values to ensure the feasibility of the optimization problems. We average the results over 100 random realizations of AP and UE positions.

TABLE I
Simulation Parameters

Parameter	Value
Carrier frequency	2 GHz
Transmit power per AP	1 W
Path loss exponent	2
Shadow fading variance	6
Noise variance	10^{-5}

A. Total Fronthaul Requirement

In Fig. 4, we show the results of the fronthaul optimization, where the dashed curve corresponds to the LP relaxation solution while the solid curve represents the solution to the original problem (following Algorithm 1). We first study the impact of SINR_{min} to the total fronthaul requirement of the network. Recall that SINR_{min} denotes the threshold value that must be satisfied by every user. Increasing SINR_{min} leads to a larger number of AP-UE connections as users may require more APs to serve them in order to meet the stricter per-UE QoS constraint. Consequently, with some APs now serving more users, bigger clusters are formed and greater amount of information needs to be transmitted over the fronthaul links. Therefore, increasing SINR_{min} implies higher fronthaul capacity requirement.

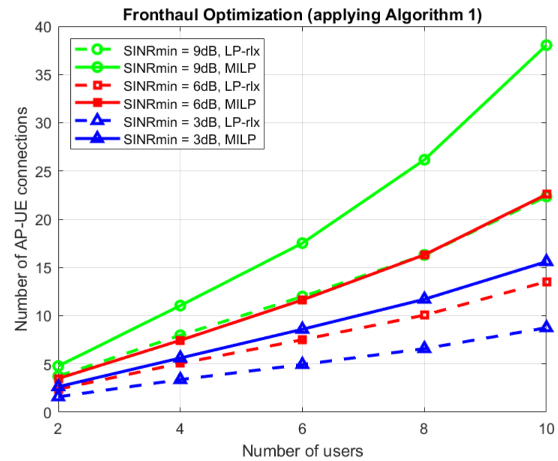

Fig. 4. Total fronthaul requirement with varying number of users and guaranteed QoS level (applying Algorithm 1).

We also investigate the impact of the number of users for a given value of SINR_{min}. In Fig. 4, we observe that increasing the number of users naturally increases the fronthaul requirement. For every user that enters the network, we add a minimum of one AP-UE connection since at least one AP needs to serve this newcomer. In fact, as the cell-free architecture benefits from the coherent joint transmission of multiple APs, in many cases, each user-centric cluster consists of more than one AP. Moreover, the increase in the fronthaul requirement depends on the size of the clusters formed by individual users. We point out that the cluster size varies for different users. Some require more APs to serve them to satisfy the per-UE SINR constraint, while others need less to achieve the same performance. Such variation is attributed to the distance between the AP and the user that affects the macroscopic path loss and SINR calculation.

B. Guaranteed Quality of Service

In Fig. 5, we present the results for the max-min SINR optimization. In particular, we study the impact of the maximum cluster size $csize_{max}$ (i.e., maximum number of APs available to each user) on the maximum SINR_{min}, which is the highest achievable QoS guaranteed to all users.

The red dashed curve represents the LP relaxation solution for a scenario of $K = 10$ users. We observe that increasing

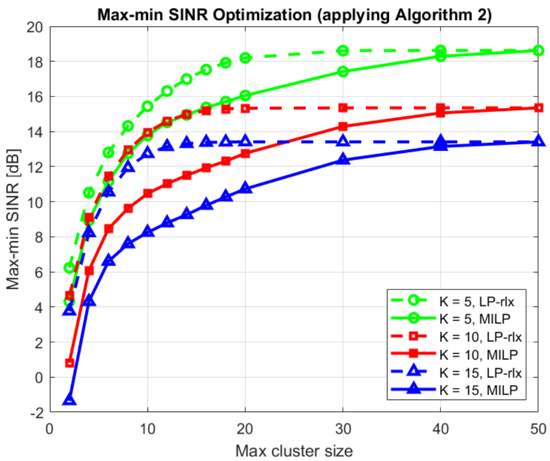

Fig. 5. Maximum guaranteed QoS with varying number of users and cluster size (applying Algorithm 2).

$csize_{max}$ leads to an increase in the maximum SINR_{min}. A larger $csize_{max}$ implies a higher number of allowable AP-UE connections per user. Consequently, the coherent transmission from possibly more APs leads to an improvement in UE performance that explains the higher maximum SINR_{min} values. However, after some point, allowing more APs to join the user-centric cluster does not improve the guaranteed QoS anymore. When $csize_{max}$ is at least 20, we observe a saturated (flat) behavior. From this point on, the maximum SINR_{min} is just above 15 dB regardless of $csize_{max}$ value. In contrast, we do not see the same trend with the red solid curve, corresponding to our solution to the original problem (following Algorithm 2). However, notice when $csize_{max}$ is at least 20, the increase in the maximum SINR_{min} is not that much compared when $csize_{max}$ is below 20. In fact, when the cluster consists of the first 10 APs, the network is able to guarantee almost 11 dB SINR level for all users. Adding another 10 APs only improves the performance by 2 dB. Furthermore, even in the extreme case of each user being served by all 50 APs (canonical), the gap is approximately 5 dB while the additional 400 AP-UE connections (40 per user) results to a significant increase in the overall fronthaul capacity requirement. This suggests that the new APs joining the cluster are already too far from the user that they have little impact on performance. In this scenario, these new APs do not contribute as much as the first $\sim$10 APs. In this sense, user-centric outperforms canonical, since our results show that we do not need as much AP-UE connections (and thus, lower fronthaul requirement) to achieve more or less the same maximum guaranteed QoS.

We also investigate the impact of the number of users in Fig. 5. Increasing the number of users to $K = 15$ decreases the maximum achievable SINR_{min}. This is because more users need to satisfy the per-UE SINR constraint while maintaining the same number of APs and transmit antennas. On the other hand, with less users admitted to the network, it becomes possible to guarantee higher QoS levels.

V. Conclusion

Cell-free massive MIMO is a key enabler for next-generation mobile communications [18]. Scalable cell-free networks can only be realized through user-centric clustering. Motivated by this, we investigated how cluster formation impacts performance. Using our proposed algorithms, we showed that good QoS can be guaranteed even with relatively small cluster sizes that require lower fronthaul capacity. The insights from this paper can be used in future work to design an intelligent approach to AP selection and dynamic cluster formation given varying network load conditions and practical fronthaul limitations.

Acknowledgment: The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," in *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[2] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," in *IEEE Access*, vol. 2, pp. 930-946, 2014.

[3] R. Irmer et al., "Coordinated multipoint: concepts, performance, and field trial results," in *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102-111, Feb. 2011.

[4] S. Venkatesan, A. Lozano and R. Valenzuela, "Network MIMO: overcoming intercell interference in indoor wireless systems," *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 83-87, 2007.

[5] S. Schwarz, R. W. Heath and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," *2012 IEEE Globecom Workshops*, pp. 546-551, 2012.

[6] U. Madhow, D. R. Brown, S. Dasgupta and R. Mudumbai, "Distributed massive MIMO: algorithms, architectures and concept systems," *2014 Information Theory and Applications Workshop (ITA)*, pp. 1-7, 2014.

[7] S. Schwarz, "Remote radio head assignment and beamforming in dynamic distributed antenna systems," *2018 IEEE International Conference on Communications (ICC)*, pp. 1-6, 2018.

[8] S. Schwarz et al., "Dependable wireless connectivity: insights and methods for 5G and beyond," *e & i Elektrotechnik und Informationstechnik* 135, pp. 449–455, 2018.

[9] G. Interdonato, P. Frenger and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," *2019 IEEE International Conference on Communications (ICC)*, pp. 1-6, 2019.

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

[11] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77-90, Jan. 2020.

[12] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: user-centric approach," in *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706-709, Dec. 2017.

[13] F. Riera-Palou, G. Femenias, A. G. Armada and A. Pérez-Neira, "Clustered cell-free massive MIMO," *2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6, 2018.

[14] F. Riera-Palou and G. Femenias, "Decentralization issues in cell-free massive MIMO networks with zero-forcing precoding," *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 521-527, 2019.

[15] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.2*, http://cvxr.com/cvx, Jan. 2020.

[16] M. ApS, *MOSEK optimization suite, version 8.1.0.82*, https://www.mosek.com/documentation, Oct. 2019.

[17] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2014.

[18] N. Rajatheva et al., "White paper on broadband connectivity in 6G", Apr. 2020.