Garau Burguera, Pere; Al-Tous, Hanan; Tirkkonen, Olav

Distributed User-Centric Cell-Free Massive MIMO with Architectural Constraints

Published in:
2024 Joint European Conference on Networks and Communications and 6G Summit, EuCNC/6G Summit 2024

DOI:
10.1109/EuCNC/6GSummit60053.2024.10597006

Published: 01/01/2024

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Garau Burguera, P., Al-Tous, H., & Tirkkonen, O. (2024). Distributed User-Centric Cell-Free Massive MIMO with Architectural Constraints. In *2024 Joint European Conference on Networks and Communications and 6G Summit, EuCNC/6G Summit 2024* (pp. 541-546). (European Conference on Networks and Communications). IEEE. https://doi.org/10.1109/EuCNC/6GSummit60053.2024.10597006

Distributed User-Centric Cell-Free Massive MIMO with Architectural Constraints

Pere Garau Burguera, Hanan Al-Tous, and Olav Tirkkonen
Department of Information and Communications Engineering, Aalto University, Finland
Email:{pere.garauburguera, hanan.al-tous, olav.tirkkonen}@aalto.fi

***Abstract*—We consider a network-level coordination architecture operating on a disaggregated Radio Access Network (RAN). We introduce the concept of Remote Radio Head (RRH) multi-partitioning, which allows users to be served by clusters of RRHs in a user-centric manner. The network consists of a number of Centralized Units (CUs) connected to several Distributed Units (DUs), each controlling a priori a cluster of multiple-input multiple-output (MIMO) RRHs. Aiming at providing uniform coverage and user performance over the network area, RRHs at cluster boundaries are connected to two or more DUs, leading to a multi-partitioning architecture in which RRHs may belong to more than one cluster. Each DU, in turn, manages multiple overlapping clusters of RRHs. The total system bandwidth is divided into orthogonal resource parts, each of which is assigned to a different partition. We consider a coordination framework to allocate resources, where each DU, in a distributed way, assigns radio resources to the different RRH clusters that it manages. Each DU manages the scheduling of its users to their corresponding serving clusters. Simulation results show that the 5th percentile downlink user spectral efficiency improves by 45% when five RRH partitionings are used, compared to only one.**

***Index Terms*—Cell-free massive MIMO, user-centric networking, RRH multipartitioning, distributed cell-free, RAN disaggregation, functional splits, proportional fairness.**

I. Introduction

In cellular systems, cell-edge users experience the worst network performance, due to comparable received signal and interference levels. Cell-free distributed massive Multiple-input Multiple-output (mMIMO) communication is a promising technology to eliminate cell boundaries and provide reliable service for all users [1].

In a disaggregated Radio Access Network (RAN), functionalities are separated into different RAN entities. Corresponding splits are fundamental to the 5G New Radio (NR) RAN architecture [2], and are considered in the Open RAN (O-RAN) alliance for 6G [3]. In a typical functional split [4], analog radio functions are performed in Remote Radio Heads (RRHs), while digital baseband and Medium Access Control (MAC) functions are performed in Distributed Units (DUs). Packet handling in the user plane and higher layer control functionalities are performed in a Centralized Unit (CU). Processing efficiency is achieved by having a DU to manage multiple RRHs, and a CU to manage multiple DUs.

Ideal orchestrating of RRH actions distributed over a wide geographic area would enable ideal cell-free operation, with the whole network coordinated to serve a set of users. However, serving all users with all transmitters in a large region is impractical, due to the enormous overhead in terms of exchanging real-time Channel State Information (CSI), and computational load. Fully distributed operation in cell-free mMIMO, where no decisions are performed by a network-level centralized entity, however, remains an open challenge [5]. The objective of such an approach would be to provide a *cell-free experience* to each user, where the co-channel interference that users suffer from is small.

In the literature, user-specific clustering of RRHs is considered. As earlier considered for user-centric Coordinated Multipoint (CoMP) transmission [6], user-centric cell-free approaches are based on constructing a serving cluster of coordinated RRHs for each user [7]. Different levels of beamformer coordination have been considered in this context [8], [9], assuming that all transmissions happen in the same radio resource. In [10], a user-centric approach is addressed where there are multiple time slots, and a centralized controller schedules users to slots and finds the corresponding beamformers. From the perspective of disaggregated RAN, these works consider a network where a CU controls all the RRHs. There is no DU-layer, and there is only one CU in the whole network. Accordingly, these approaches do not solve the overhead problems in providing cell-free experience to users in wide-area networks.

In [4], user-centric disaggregated RAN operation is considered, where user association and RRH clustering are centrally optimized, constrained to each RRH belonging to only one RRH cluster. This solution, however, does not provide a cell-free experience to all users. Users in boundary regions between RRH clusters will suffer from heavy co-channel interference. Existing solutions like dual connectivity allowed by the Packet Data Convergence Protocol (PDCP) of Fifth Generation (5G) RANs [11] offer only a partial solution to this. While a dual-connected User Equipment (UE) may receive packets from multiple RRH clusters connected to multiple DUs, the root cause of poor performance at the cluster edge, the interference from other clusters, is not removed.

In this paper, we consider a disaggregated RAN with multiple DUs. In contrast to the literature, we allow RRHs to be connected to multiple DUs, which makes it possible to provide a cell-free experience to users. For users that are in the boundary region between two RRH clusters, using a single partitioning would lead to poor service due to inter-cluster interference. Accordingly, instead of one clustering of RRHs to DUs, we perform *multipartitioning*, where each DU has a

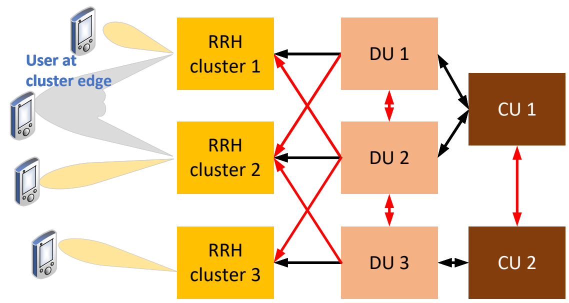

Fig. 1: Logical RAN architecture, consisting of CUs, DUs, and clusters of RRHs. Certain users may be served by RRHs from multiple clusters. For this, RRH resources are allocated between DUs.

Fig. 2: Two basic clusters are shown, both in the same resource partition. A user located at the border between them will always suffer from strong interference. Being served by a cluster in a different partition (in gray) mitigates the cluster-edge problem.

different cluster of RRHs in each partitioning. To improve the situation of users at the cluster edge, we consider an architecture where RRHs close to cluster boundaries are connected to multiple DUs. The DUs negotiate over an interface for the division of resources between the RRHs that they share, using a related application protocol. Thus, specific frequency domain Physical Resource Blocks (PRBs) in a given RRH are allocated to a specific DU, on a time scale longer than the scheduling interval. On the sub-millisecond MAC time scale, a DU may then independently schedule users to be served by different groups of RRHs for different sets of PRBs. We divide the total bandwidth into orthogonal resources, such that each partitioning of the RRHs has an orthogonal resource. Real-time user scheduling and beamformer selection is performed by the DUs, implicitly solving the user-centric selection of an active set of RRHs to serve a user. Resource allocation to different partitionings is coordinated by the DUs and CUs in a *distributed* manner.

The remainder of this paper is organized as follows: In Section II, the system model is introduced. In Section III, the resource allocation problem and the solution approach are presented. Simulation results are presented and discussed in Section IV. Finally, conclusions are drawn in Section V.

II. System Model

We consider a disaggregated RAN, with the logical architecture depicted in Fig. 1. RRHs are clustered so that each cluster is managed by a DU. There is a front-haul network connecting the DUs to the RRHs, and a back-haul network connecting the CUs to the DUs. We consider an architectural limitation in front- and back-haul connectivity; not all DUs are connected to all RRHs, nor are all CUs connected to all DUs.

For concreteness, we consider the downlink transmission in a disaggregated RAN, where A RRHs, each with N antennas, are deployed in the area to serve single antenna UEs. Each cluster of RRHs is connected to a DU via a front-haul connection. D DUs are deployed in the area, which are connected to one CU. UE u will be served by at least one RRH cluster, managed by a DU.

A. RRH and UE Clustering and Multipartitioning Principle

In the literature, several approaches are proposed to cluster the UEs and the RRHs [9], [12]. For simplicity, we will assume that the RRHs are clustered based on their physical locations into D clusters, referred to as *basic* clusters. The number of basic clusters equals the number of DUs. Each cluster of RRHs is connected to a DU. Each user is assigned to the DU managing the RRH with which it has the greatest average channel gain, based on the basic clusters.

Considering the basic clusters only, the resource allocation becomes a user scheduling problem. Such a scheduling suffers from a cluster-edge problem, i.e., the signal level of a user at the boundary between two basic clusters is comparable to the interference. To handle the cluster-edge problem, we consider a multipartitioning principle. The main idea is to divide the total bandwidth into orthogonal partitions and create multiple overlapping clusters of RRHs connected to a DU. Each resource division is assigned to a partition. Fig. 2 illustrates the multipartitioning principle. The idea is that no user will be found at cluster edges in all RRH partitionings. That is, there is always at least one cluster for each user where it does not suffer from strong inter-cluster interference.

The set of orthogonal resource partitions is denoted as $\mathcal{K}$. We assume that each DU manages $|\mathcal{K}|$ RRH partitions, each tied to a resource division. The network physical limits, such as front-haul connections, affect the possible partitionings. The choice of the optimal number of partitions and the assignment of RRHs to partitions are left for future work.

The CU finds the optimal amount of resources to allocate to each partition, and w_k denotes the amount of resources allocated to resource k. We assume that the resources are infinitely divisible. The total amount of system resources W is divided into parts as

$$\sum_{k\in\mathcal{K}} w_k = W,\ w_k \geq 0. \tag{1}$$

Users are served by RRH clusters managed by the DU to which they are connected. Each DU manages the scheduling of its users into the resource partitions. The scheduling weight of user u in partition k is denoted by $\alpha_{k,u}$. The set of DUs is denoted by $\mathcal{D}$, with $|\mathcal{D}| = D$, and the set of UEs assigned to DU d is denoted by $\mathcal{U}_d$. The total number of users in the network is denoted by U', i.e., $\sum_{d \in \mathcal{D}} |\mathcal{U}_d| = U'$. In each cluster managed by a DU, U users are simultaneously served with a zero-forcing beamformer, giving

$$\sum_{u \in \mathcal{U}_d} \alpha_{k,u} = U, \ \alpha_{k,u} \geq 0. \tag{2}$$

B. Functional Split and Front-Haul Connection

We consider two RRH-DU functional splits. First, we consider Common Public Radio Interface (CPRI) split E [13], in which low physical layer functionalities are performed by the DUs. The corresponding 3rd Generation Partnership Project (3GPP) split is option 8 [2]. The second split considers an intra-physical layer split, where some functions are performed by the RRHs. Concretely, it corresponds to Enhanced CPRI (eCPRI) split $\mathrm{II_D}$ [13], and to O-RAN option 7.2x and option 7 in 3GPP [2].

Four front-haul links span from each DU, in the 4 directions. Each RRH is connected by following the path through other RRHs in a tree-like manner. The front-haul has the minimum length possible, and its traffic is aggregated. Each front-haul link spanning from a DU then has to support the traffic to a quarter of the RRHs connected to it. In the case of CPRI split E, the front-haul carries time-domain samples; in the case of eCPRI split $\mathrm{II_D}$, it carries frequency-domain samples.

C. Channel Model

The channel gain between RRH a and UE u is denoted as

$$\mathbf{g}_{a,u} = \sqrt{\beta_{a,u}}\mathbf{h}_{a,u}, \tag{3}$$

where $\mathbf{g}_{a,u} \in \mathbb{C}^{N \times 1}$. $\beta_{a,u}$ is the large scale coefficient, represents the path-loss and shadow fading, i.e., $\beta_{a,u} = L_{a,u}S_{a,u}$, where $L_{a,u} = \frac{\rho_0}{d_{a,u}^{\nu}}$ is the path loss with ρ_0 the path-loss at the reference distance of 1 m, and ν the path loss exponent, $d_{a,u}$ is the distance between RRH a and UE u. The shadow fading $S_{a,u}$ expressed in dB is modelled as a zero mean Gaussian random variable with standard deviation of σ_{S}. The small scale fading is $\mathbf{h}_{a,u} \sim \mathcal{NC}(\mathbf{0}_N, \mathbf{I}_N)$, its elements are complex circularly symmetric normal random variables.

D. Rate Calculation

We assume that a beamforming vector applies to all RRHs in a cluster to serve U UEs in that cluster. Given the orthogonality of the partitions, the following model is formulated for each partition separately, i.e., there is no interference between partitions. To simplify the notation, the partition index is not used. The combined multiuser channel matrix in cluster c with A_c RRHs is:

$$\mathbf{G}^{(c)} = \begin{bmatrix} \mathbf{g}_{1,1}^{\mathrm{H}} & \cdots & \mathbf{g}_{A_c,1}^{\mathrm{H}} \\ \vdots & \vdots & \vdots \\ \mathbf{g}_{1,U}^{\mathrm{H}} & \cdots & \mathbf{g}_{A_c,U}^{\mathrm{H}} \end{bmatrix}, \tag{4}$$

where $\mathbf{G}^{(c)} \in \mathbb{C}^{U \times NA_c}$, and $U < NA_c$. The set of RRHs $\mathcal{A}_c$ in cluster c aplies a zero-forcing precoding matrix $\mathbf{F}^{(c)}$ based on the pseduo-inverse of channel matrix $\mathbf{G}^{(c)}$, then normalizing the columns. The multiuser precoders for users in other clusters are similarly defined. The received signal at UE u in cluster c is

$$y_{c,u} = \sqrt{p}\,\mathbf{g}_u^{(c)}\mathbf{f}_u^{(c)}x_u + \sum_{c',\, c' \neq c} \sqrt{p}\,\bar{\mathbf{g}}_u^{(c')}\mathbf{F}^{(c')}\mathbf{x}^{(c')} + n_u, \tag{5}$$

where $\mathbf{g}_u^{(c)}$ is the uth row of $\mathbf{G}^{(c)}$, $\mathbf{f}_u^{(c)}$ is the uth column of the precoder matrix $\mathbf{F}^{(c)}$, n_u is additive white Gaussian noise with variance σ^2, p is the transmitted power, and x_u denotes the symbol intended for UE u transmitted from cluster c with $\mathbb{E}\{|x_u|^2\} = 1$. The interference channel to UE u from cluster c' is denoted by $\bar{\mathbf{g}}_u^{(c')}$, the zero-forcing precoder at cluster c' is denoted by $\mathbf{F}^{(c')}$ and the transmitted symbols from cluster c' are denoted by $\mathbf{x}^{(c')}$.

We consider a sum power constraint for each cluster to simplify the analysis. Per RRH power constraints can be considered based on [14]. However, this will complicate the analysis. The ergodic rate for UE u, in cluster c is given as [15]:

$$R_{c,u} = \\ \mathbb{E}\log\left(1 + \frac{\frac{p}{U\sigma^2}|\mathbf{g}_u^{(c)}\mathbf{f}_u^{(c)}|^2}{\frac{p}{U\sigma^2}\sum_{c',\, c' \neq c} \bar{\mathbf{g}}^{(c')}{}_u\mathbf{F}^{(c')}\mathbf{F}^{(c')\mathrm{H}}\bar{\mathbf{g}}_u^{(c')\mathrm{H}} + 1}\right), \tag{6}$$

where the expectation is with respect to the small scale fading.

The ergodic rate lacks a closed form solution; we adopt the moment matched Gamma approximation as in [15] to get a closed form approximation. The signal term is approximated by a Gamma distribution $\Gamma(\kappa_x, \theta_x)$ with finite shape parameter κ_x and finite scale parameter θ_x. The mean is $\kappa_x\theta_x$ and the variance is $\kappa_x\theta_x^2$. The parameters are selected as [15]:

$$\kappa_x = A_cN - U + 1, \text{ and } \theta_x = \frac{\mathbb{E}\{\mathbf{g}_u^{(c)}\mathbf{g}_u^{(c)\mathrm{H}}\}}{A_cNU\sigma^2}. \tag{7}$$

Similarly, each of the interference terms is approximated by a Gamma distribution $\Gamma(\kappa_{c'}, \theta_{c'})$ with parameters

$$\kappa_{c'} = \frac{U\mu^2}{\sigma_v^2}, \text{ and } \theta_{c'} = \frac{\sigma_v^2}{\mu}, \tag{8}$$

where μ and σ_v^2 are computed based on the large scale coefficients of the channel to interfering cluster c' as in [15]. The total interference then is approximated with a Gamma distribution $\Gamma(\kappa_y, \theta_y)$ with

$$\kappa_y = \frac{\left(\sum_{c',\, c' \neq c} \kappa_{c'}\theta_{c'}\right)^2}{\sum_{c',\, c' \neq c} \kappa_{c'}\theta_{c'}^2}, \text{ and } \theta_y = \frac{\sum_{c',\, c' \neq c} \kappa_{c'}\theta_{c'}^2}{\sum_{c',\, c' \neq c} \kappa_{c'}\theta_{c'}}. \tag{9}$$

Equipped with Gamma descriptions of the signal and interference terms, and under the high signal to interference ratio assumption, i.e., $\mathbb{E}_{X,Y}\log_2(1 + \frac{x}{y+1}) \approx \mathbb{E}_{X,Y}\log_2(x + y) - \mathbb{E}_Y\log_2(y)$, the ergodic rate is approximated as

$$R_{c,u} \approx \log_2\left(\frac{\kappa_{xy}\theta_{xy}}{\kappa_y\theta_y}\right) + \frac{\log_2 e}{\kappa_{xy}} - \frac{\log_2 e}{\kappa_y}, \tag{10}$$

where e is Euler's number, and

$$\kappa_{xy} = \frac{(\kappa_x\theta_x + \kappa_y\theta_y)^2}{\kappa_x\theta_x^2 + \kappa_y\theta_y^2}, \text{ and } \theta_{xy} = \frac{\kappa_x\theta_x^2 + \kappa_y\theta_y^2}{\kappa_x\theta_x + \kappa_y\theta_y}. \tag{11}$$

III. Problem Formulation and Solution

We are interested in allocating the network resources to serve U' users in a downlink scenario, where the number of users is much larger than the number of distributed antennas in the system, i.e., $U' \gg AN$. We consider a coordinated approach to allocate the resources between different partitions, and a distributed approach for scheduling the users.

We consider $K+1$ resource partitions; i.e., the basic cluster partition and K additional clusters, giving $|\mathcal{K}| = K+1$.

We have two interacting resource allocation problems. The main problem of interest is the resource allocation between the $K+1$ partitions. To solve this problem, we need to have a model of MAC scheduling. For this, we consider an abstraction of network performance based on Network Utility Maximization (NUM). The network utility is modeled as the sum of the utilities of all users in the system. We consider a proportionally fair utility function for resource allocation among the users [10], giving the network utility

$$f_d(\mathbf{w}, \boldsymbol{\alpha}) = \sum_{u \in \mathcal{U}_d} \log \left(\sum_{k \in \mathcal{K}} w_k \, \alpha_{k,u} \, R_{k,u} \right), \text{ for } d \in \mathcal{D},$$

where $R_{k,u}$ is the ergodic rate of UE u in partition k, $\boldsymbol{\alpha}$ represents the user scheduling vector and $\mathbf{w}$ is the resource allocation vector. The joint optimization problem is not convex. Additionally, centrally solving the MAC scheduling and resource allocation problem requires a centralized knowledge (at the CU[1]) of the large scale fading coefficients of all RRH-user pairs. To avoid that, the MAC scheduling is performed by the DUs in a distributed way. The DUs then, for a fixed $\mathbf{w}$, solve

$$\max_{\boldsymbol{\alpha}} f_d(\mathbf{w}) \tag{12a}$$

$$\text{subject to: (2) for } d \in \mathcal{D}. \tag{12b}$$

For a given resource allocation $\mathbf{w}$, the user scheduling problem (12) is convex. The CU, for a given $\boldsymbol{\alpha}$, finds the optimal allocation of resources to partitions, by solving

$$\max_{\mathbf{w}} \sum_{d \in \mathcal{D}} f_d(\boldsymbol{\alpha}) \tag{13a}$$

$$\text{subject to: (1) for } k \in \mathcal{K}. \tag{13b}$$

For a given user scheduling $\boldsymbol{\alpha}$, the resource allocation problem (13) is convex. To reduce the back-haul overhead caused by the DUs sending their MAC scheduling weights to the CU, it is sufficient if the DUs compute the gradient of their utility functions with respect to $\mathbf{w}$, and send those to the CU. The CU sums the gradient from all DUs and then performs a gradient projection to update $\mathbf{w}$. The DUs and the CU solve their respective problems on different time scales. On the sub-millisecond scale, the DUs schedule users to partitions. On a longer time scale, the CU updates the resource allocation to partitions and sends the updated values to the DUs. This approach is independent of the utility function used.

[1] In the case of multiple CUs, inter-CU coordination is needed (as illustrated in Fig. 1), in order to eliminate the CU edges and provide a cell-free experience to all users, regardless of their physical location.

In the case of two partitions, the user scheduling problem at DU d for a given resource allocation $\mathbf{w} = [w_o, w_1]^{\mathrm{T}}$ is

$$\max_{\boldsymbol{\alpha}} \sum_{u \in \mathcal{U}_d} \log \left(w_0 \, \alpha_{0,u} \, R_{0,u} + w_1 \, \alpha_{1,u} \, R_{1,u} \right), \tag{14a}$$

$$\sum_{u \in \mathcal{U}_d} \alpha_{k,u} = U, \text{ for } k = 0, \; 1. \tag{14b}$$

Formulating the Lagrangian function, taking the derivative with respect to $\alpha_{0,u}$, and setting to zero, results in

$$\frac{w_0 \, R_{0,u}}{w_0 \, \alpha_{0,u} \, R_{0,u} + w_1 \, \alpha_{1,u} \, R_{1,u}} = \lambda_0, \tag{15}$$

where λ_0 is the Lagrange multiplier associated with total user scheduling in resource w_0. The Lagrangian multiplier λ_1 associated with total user scheduling in resource w_1 is found analogously.

All users scheduled at DU d need to satisfy the above equation (for λ_0 and λ_1). Assuming that each user can be scheduled to both the 0 and 1 resources, then by taking the ratio between λ_0 and λ_1, we get

$$\frac{w_0 \, R_{0,u}}{w_1 \, R_{1,u}} = \frac{\lambda_0}{\lambda_1},$$

this is impossible to satisfy by all the users. This means that a user will be scheduled either to resource 0 or 1 and then it will receive a constant scheduling weight. Maximally one user can be scheduled to both resources.The solution approach is called two-band partition.

Consider the case of two partitions and applying the two-band partition for user scheduling. Formulating the Lagrangian, differentiating with respect to w_0 and w_1 and setting to zeros, gives

$$\lambda_w = (1-j) \frac{\sum_{d \in \mathcal{D}} (M^{(d)} - 1)}{w_0} + j \frac{\sum_{d \in \mathcal{D}} |\mathcal{U}_d| - M^{(d)}}{w_1} + \sum_{d \in \mathcal{D}} \frac{\alpha_{j,M^{(d)}} \, R_{j,M^{(d)}}}{w_0 \, \alpha_{0,M^{(d)}} \, R_{0,M^{(d)}} + w_1 \, \alpha_{1,M^{(d)}} \, R_{1,M^{(d)}}}, \tag{16}$$

for $j = 0, 1$, respectively. $M^{(d)} - 1$ are the scheduled users to resource 0 at DU d. As the number of users scheduled to both resources in the system is very small (i.e., $|\mathcal{D}| \ll U'$), ignoring the last term in (16), assigning these users to resource 0 and using the total bandwidth constraint (1), results in

$$w_0 = \frac{W \sum_{d \in \mathcal{D}} M^{(d)}}{\sum_{d \in \mathcal{D}} |\mathcal{U}_d|}, \; w_1 = \frac{W \sum_{d \in \mathcal{D}} |\mathcal{U}_d| - M^{(d)}}{\sum_{d \in \mathcal{D}} |\mathcal{U}_d|}. \tag{17}$$

This means that the bandwidth in each partition is allocated proportionally to the number of served users in that partition. This allows a simple approach for updating $\mathbf{w}$ at the CU, without the need for computing the gradients, only with the number of users scheduled in each partition.

IV. Simulation Results

We consider a square layout of L^2 km^2, divided in D square clusters of RRHs, each managed by a different DU. We simulate a scenario where the RRHs are placed in a regular grid-like manner. The users are uniformly placed at random.

TABLE I: Simulation parameters.

Parameter	Value
Bandwidth	200 MHz
Tx power	30 dBm
Noise power	-174 dBm/Hz
User noise figure	9 dB
Path-loss	$13.54 + 39.08 \log_{10}(d)$ dB
Shadow fading	$\mathcal{N}(0, 4^2)$ dB
Shadow fading correlation ρ	0.5
Size of area L^2	2.25 km^2
Inter-RRH distance	50 m
Min. UE-RRH distance	5 m
No. of RRHs A	900
No. of DUs D	{9, 16, 25}
Shift size s	{1, 2, 3}
No. of antennas at RRH N	4
No. of served UEs U	$\lfloor 450/D \rfloor$
Total no. of UEs U'	5000

To form the partitions, the basic clusters are considered as a starting point. The clusters that are connected to a DU are formed based on a shifting principle, where the centers of the basic clusters are shifted in a certain direction. As a result, RRHs located at the borders of the basic clusters are also connected to the neighboring DUs. In particular, the RRHs in the s closest rows/columns to the DU borders, based on the basic clusters. The sizes of the resulting clusters are almost the same as the size of the basic clusters. Extra front-haul connections are needed. To simplify the discussion, we consider four shift directions; left, right, up, and down, attempting to eliminate the edges of the square DU areas. The front-haul connections resulting from the multipartitioning principle only they depend on the value of s and remain fixed.

Interference is simulated with a wrap-around geometry. The simulation parameters, path-loss, and the shadow fading models can be found in Table I.

We evaluate the average Spectrum Efficiency of user u in [bits/s/Hz] computed as

$$\eta_u = \frac{\sum_{k \in \mathcal{K}} w_k \, \alpha_{k,u} \, R_{k,u}}{W},$$

where $\mathbf{w}$ and $\boldsymbol{\alpha}$ are obtained by solving the resource allocation and scheduling problems, and $R_{k,u}$ is computed with (10). The average and the 5th percentile of the spectral efficiency in the area are used to evaluate the performance.

For the multipartitioning approach, 5 partitions are considered; each DU considers its basic cluster and 4 more clusters obtained via shifting to the 4 directions. We compare the RRH multipartitioning approach performance to the following benchmark schemes:

- Upper bound of interference-free cell-free MIMO in an isolated universe; all the RRHs are connected to one DU implementing a zero-forcing precoder, and there are no other users of the radio resources. The number of simultaneously served users is UD. Each user is scheduled a fraction UD/U' of the resources. In this scenario, there is no inter-cluster co-channel interference.
- Distributed MIMO (D-MIMO) per DU [15]; a cellular system with D DUs, where only the basic cluster is considered and there are no partitions. Each DU simultaneously serves U users with zero-forcing. A user in cluster d is scheduled to $\frac{U}{|\mathcal{U}_d|}$ fraction of the resource.
- PDCP dual connectivity [11]; users may be connected to two D-MIMO DUs. Each DU implements a zero-forcing precoder to serve U users. The user scheduling problem is formulated to achieve proportional fairness at the network level. The optimization problem is convex.
- D-MIMO per CU, where all A RRHs are coordinated by one CU and serve UD users simultaneously with zero-forcing. The functions are centralized at the CU. The users suffer from interference from neighboring CUs operating in the same way.
- Collocated Massive MIMO (mMIMO) per DU [15]; a single RRH with $\frac{AN}{D}$ antennas placed at the center of each DU area, each simultaneously serving U users.

Optimization problems (12) and (13) are solved in an iterative alternating way. The scheduling vector $\boldsymbol{\alpha}$ is found by solving the convex optimization problem using CVX MATLAB toolbox [16] with a pre-defined resource allocation $\mathbf{w}$, which is then updated proportionally to the number of users allocated in each partition.

The empirical Cumulative Distribution Functions (CDFs) of the user system spectral efficiency can be found in Fig. 3, for the case $D=9$ and $s=2$. The mean and the 5th percentile of the system spectral efficiency are shown in Fig. 4, for the case $D = 9$. The results show that the average spectral efficiency achieved at the users is larger in the multipartitioning case, compared to the benchmark cases. The reason is that users are served from RRH clusters in which inter-cluster interference is lower, keeping the interfering RRHs far, as was illustrated in Fig. 2. Users at the cluster edge obtain the greatest benefit of multipartitioning. Table II details the spectral efficiency gains for the users in the 5th percentile. They correspond to the multipartitioning gains as compared to the single partitioning solution (D-MIMO per DU). By allowing different clusters of RRHs to serve users, fewer users are located at the edges of their serving clusters. They consequently suffer from less interference, and the impact of the cluster-edge problem is mitigated. The gains are better when the DU areas are larger (smaller values of D), as the total number of edge users is lower. The best result is for $s = 2$. Solutions with $s > 2$ are impractical due to their marginal improvement, with the incurred extra cost of a longer total front-haul length. This is because the RRHs that contribute the most to inter-cluster interference are those located nearby. In the D-MIMO per CU case, the users near the CU edges still suffer from strong interference from the neighboring CU. Lastly, dual connectivity improves performance only marginally when compared to single connectivity (D-MIMO per DU), as PDCP does not manage physical layer interference.

Table III summarizes the maximum front-haul load, in samples per symbol period. The most relevant result is that multipartitioning does not increase the front-haul load when compared to the D-MIMO per DU approach, if eCPRI with functional split II_D is used. This is due to each resource oc-

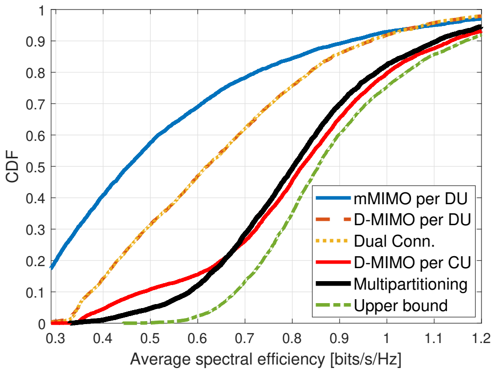

Fig. 3: The empirical CDF of the spectrum efficiency, comparing RRH multiclustering with the benchmark scenarios.

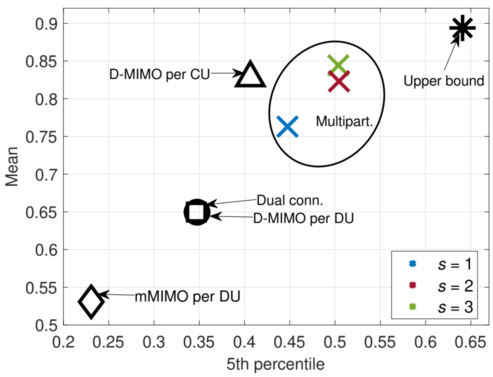

Fig. 4: Mean and fifth percentile of the average spectral efficiency of users in [bits/s/Hz], for the different approaches.

cupying a fraction of the total available bandwidth. However, multipartitioning increases the front-haul load for CPRI split E, because of the additional RRHs connected to each DU.

V. Conclusions

In this paper, we have considered resource allocation and user scheduling in a disaggregated RAN. The network consists of several DUs, each controlling several MIMO RRHs. These RRHs are grouped, and radio resources of RRHs at cluster boundaries are shared among multiple DUs. We have considered a distributed coordination framework for sharing RRH resources among DUs, and optimized the radio resource usage at the network level to achieve proportional fairness.

We have shown that the proposed architecture, in which DUs manage different RRH partitionings, each tied to a resource part, results in an increase of the user average spectral efficiency, when compared to D-MIMO systems with no resource partitioning, since users are located closer to the centers of their serving RRH clusters. The user spectral efficiency improvement can be achieved without increasing the front-haul load, if eCPRI split $\mathrm{II_D}$ is used. Our results clearly show that higher-layer solutions, such as PDCP dual connectivity, are not sufficient for eliminating the user cell-edge problem in a mobile communication system.

TABLE II: Multipartitioning 5th percentile spectral efficiency gain [%].

D \ s	1	2	3
9	28.73	**45.11**	44.89
16	26.21	36.67	36.28
25	28.76	34.71	33.25

TABLE III: Maximum front-haul load [time/frequency domain samples].

Approach	CPRI (split E)	eCPRI (split $\mathrm{II_D}$)
D-MIMO per DU	$\frac{A\,N}{4\,D}$	$\frac{A\,N}{4\,D}$
Multipartitioning	$\frac{A\,N}{4\,D}+s\sqrt{\frac{A\,N}{D}}$	$\mathbf{\frac{A\,N}{4\,D}}$
D-MIMO per CU	$\frac{A\,N}{4}$	$\frac{A\,N}{4}$

Acknowledgment

This work has been partly funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101095759 (Hexa-X-II).

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, 2017.

[2] 3GPP, "Study on new radio access technology: Radio access architecture and interfaces," 3rd Generation Partnership Project, Technical Report (TR) 38.801, 2017.

[3] J. S. Vardakas *et al.*, "Machine learning-based cell-free support in the O-RAN architecture: An innovative converged optical-wireless solution toward 6G networks," *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 20–26, 2022.

[4] C.-Y. Chang, N. Nikaein, T. Spyropoulos, and K. De Schepper, "FlexDRAN: flexible centralization in disaggregated radio access networks," *IEEE Access*, vol. 10, pp. 11 789–11 808, 2022.

[5] J. Zheng *et al.*, "Mobile cell-free massive MIMO: Challenges, solutions, and future directions," *IEEE Wireless Commun.*, 2024, early access.

[6] W. Zirwas, W. Mennerich, and A. Khan, "Main enablers for advanced interference mitigation," *Trans. Emerging Tel. Tech.*, vol. 24, pp. 18–31, 2013.

[7] 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.

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

[9] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Downlink resource allocation in multiuser cell-free MIMO networks with user-centric clustering," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1482–1497, 2022.

[10] S. Wu, Y. Wei, S. Zhang, and W. Meng, "Proportional-fair resource allocation for user-centric networks," *IEEE Trans. Veh. Commun.*, vol. 71, no. 2, pp. 1549–1561, 2022.

[11] 3GPP, "Evolved universal terrestrial radio access (E-UTRA) and NR; multi-connectivity," 3rd Generation Partnership Project, Technical Specification (TS) 37.340, 2020.

[12] R. Wang, M. Shen, Y. He, and X. Liu, "Performance of cell-free massive MIMO with joint user clustering and access point selection," *IEEE Access*, vol. 9, pp. 40 860–40 870, 2021.

[13] "Common Public Radio Interface: eCPRI interface specification," 2019.

[14] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.

[15] R. Heath, T. Wu, Y. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4885–4899, 2011.

[16] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, 2014.