An Overview of Network-Related Challenges in Cell-Free 6G

Max Franke[1], Fabian Göttsch[1], Ameya G Joshi[1], Furkan Özkoca[1], Abhishek Dandekar[1], Tugay Osman Başaran[1], Naveed Kaim Khani[1], Arash Pourdamghani[1], André Drummond[1], Julius Schulz-Zander[1], Thomas Bauschert[1], Stefan Schmid[1], Giuseppe Caire[1], Falko Dressler[1], and Sławomir Stańczak[1]

[1]Affiliation not available

October 08, 2025

Abstract

User-centric Cell-Free architectures are envisioned to be deployed in 6G networks to overcome the limitations inherent to the current cell-based architectures. While they offer drastic improvements on the physical layer, they bring new challenges to many other parts of the network. Due to their user-centric approach, many long standing principles of RAN design need to be reconsidered. This comes at an opportune time as there is currently a large push to break apart the monolithic base station based approaches and disaggregate functionalities. The O-RAN architecture in particular is a potentially good fit to accommodate cell-free networks. In this work, we highlight and discuss several major open issues related to architecture, protocols, and other network aspects encountered on the path to cell-free deployments. We give initial pointers on how the O-RAN architecture would need to evolve to enable cell-free use cases. Finally, we report first results on how different fronthaul topologies could overcome some of the cell-free knock-on effects. This work serves as a notice to the network community to make it aware of these challenges.

An Overview of Network-Related Challenges in Cell-Free 6G

Max Franke, Fabian Göttsch, Ameya G. Joshi, Furkan Özkoca, Abhishek Dandekar, Osman Tugay Başaran, Naveed Kaim Khani, Arash Pourdamghani, André Drummond, Julius Schulz-Zander, Thomas Bauschert, Stefan Schmid, Giuseppe Caire, Falko Dressler, Sławomir Stańczak

***Abstract*—User-centric Cell-Free architectures are envisioned to be deployed in 6G networks to overcome the limitations inherent to the current cell-based architectures. While they offer drastic improvements on the physical layer, they bring new challenges to many other parts of the network. Due to their user-centric approach, many long standing principles of RAN design need to be reconsidered. This comes at an opportune time as there is currently a large push to break apart the monolithic base station based approaches and disaggregate functionalities. The O-RAN architecture in particular is a potentially good fit to accommodate cell-free networks. In this work, we highlight and discuss several major open issues related to architecture, protocols, and other network aspects encountered on the path to cell-free deployments. We give initial pointers on how the O-RAN architecture would need to evolve to enable cell-free use cases. Finally, we report first results on how different fronthaul topologies could overcome some of the cell-free knock-on effects. This work serves as a notice to the network community to make it aware of these challenges.**

***Index Terms*—Cell-Free, Massive MIMO, 6G, Network planning, Fronthaul**

I. Introduction

While mobile networks have undergone major improvements over the past decades, their basic mode of operation has stayed the same. Each base station (BS) serves a fixed geographical area, called a cell, and all user equipments (UEs), e.g. phones, within that area connect to the same BS. This leads to several known problems, such as interference between neighboring cells and reduced performance at the cell edge, especially in dense deployments.

In future 6G networks, it is expected that much denser deployments of antennas will be used, especially in certain scenarios like in sports stadiums or on crowded city streets. This opens up the possibility for new ways of organizing the radio access network (RAN). One of the most promising ideas for dense antenna deployments is *cell-free massive MIMO*. Instead of users being assigned to a specific cell, in a cell-free setup each UE is served by a set of distributed radio units (RUs) that are selected based on channel quality. As the user moves, the set of serving RUs adapts. The goal is to provide more consistent quality of service and better interference management.

M. Franke, F. Göttsch, O. Başaran, N. Kaim Khani, A. Pourdamghani, S. Schmid, G. Caire and F. Dressler are with the School of Electrical Engineering and Computer Science, Technical University Berlin, Germany. A. Joshi, F. Özkoca and T. Bauschert are with the Chemnitz University of Technology, Germany. A. Drummond is with the Institute for Computer and Network Engineering, Technical University Braunschweig, Germany and the Department of Computer Science, University of Brasilia, Brazil. A. Dandekar, J. Schulz-Zander and S. Stańczak are with Fraunhofer HHI, Germany. S. Stańczak is also with Technical University Berlin.

The physical layer benefits of cell-free massive MIMO have been demonstrated in many recent works (see [1, 2] and references therein), but from a networking point of view, a lot of open issues remain.

In this work, we focus on the O-RAN architecture, which is suitable for implementing cell-free massive MIMO [1]. It is not yet clear how to efficiently implement "O-RAN cell-free massive MIMO", how to deal with the resulting control signaling overhead, and what kind of fronthaul topologies are needed to support RU cooperation. Handling mobility, especially fast-moving users who trigger frequent re-clustering events, also introduces new challenges. Moreover, realistic end-to-end simulations and testbed implementations are still rare, and there is a lack of standardized terminology across protocol layers. We focus on these challenges and discuss how the network architecture and control mechanisms will need to evolve in order to support scalable user-centric cell-free massive MIMO.

Before we discuss the challenges in detail, we will provide a brief background on cell-free massive MIMO. For more in-depth information, we refer the reader to one of the various surveys on the topic, such as [1, 3]. These surveys indicate the breadth to which cell-free MIMO has been studied, while little attention has been paid to the issues discussed in our work.

II. Background

As mentioned, the advantages of cell-free massive MIMO compared to conventional massive MIMO and small cell networks have been discussed in several works. In a well-designed cell-free massive MIMO system, each user is connected to all RUs that provide a significant channel gain, forming the *user-centric* cluster, which is handled by its associated cluster processor (CP). The downlink (DL) and uplink (UL) signals from and to the RUs that are not part of the cluster are considered weak enough not to cause substantial interference. This design, combined with intelligent cluster-level signal processing, such as power control, UL combining and DL precoding, results in a network with little (or in the ideal case virtually no) interference. In addition, the spatial distribution of RUs increases macrodiversity, i.e., diversity from widely separated antennas that improves reliability against signal fluctuations, such that users should experience a high quality of service over the entire network area. In cell-based networks, on the contrary, users located at the cell edges suffer from severe interference. Hence, larger physical layer (PHY) data

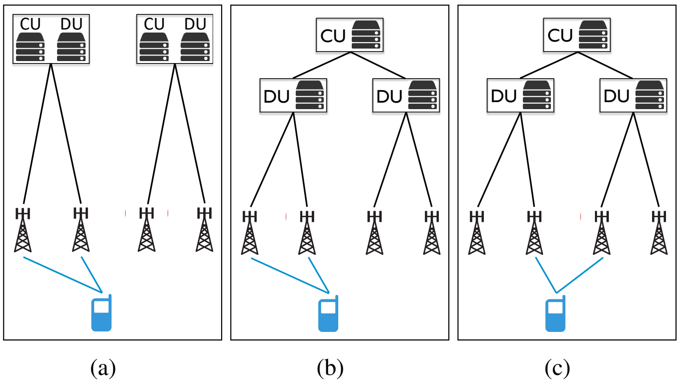


Figure 1: Different distributed MIMO RAN architectures: (a) CRAN, (b) O-RAN, (c) user-centric cell-free O-RAN. While the monolithic base station design of CRAN does not fit well for a cell-free network, O-RAN is well suited due to its disaggregated architecture.

rates and higher reliability can be achieved with a cell-free network design compared to collocated massive MIMO and small cell networks.

It is important to point out the differences to similar (previous) technologies. Coordinated multipoint (CoMP) follows a network-centric clustering approach [4], so that a more uniform distribution of the quality of service is achieved within the cluster, but not over the entire network area. In a cloud (or centralized) radio access network (CRAN), UL signals (same for the DL) received at distributed RUs from the UEs are jointly processed at a centralized unit (CU). Note that CRAN is a network architecture and CoMP is a LTE feature. There is no major conceptual difference between CRAN and CoMP in a fully centralized approach from a theoretical point of view. The reasons why CoMP and CRAN have not been successful in practice include imperfect synchronization between RUs and the large fronthaul load caused by the exchange of channel state information (CSI) and user data [4]. While some challenges remain (e.g., RU synchronization), user-centric cell-free massive MIMO networks are designed to address the issues of CoMP and CRAN, i.e., very large fronthaul traffic and a high computational load at the CU. Therefore, a decentralized cell-free massive MIMO approach foresees some local signal processing at the RUs and DUs, avoiding the need to exchange CSI and user data through a CU. We notice that Open RAN (O-RAN) with separated CU and DU is an excellent architecture for decentralized processing [1, 5]. However, the current O-RAN implementation considers fixed RU-DU associations and is not ready for user-centric cell-free massive MIMO.

Fig. 1 illustrates different distributed MIMO RAN architectures, where a UE is served by multiple RUs at the same time. The current technologies CRAN and O-RAN shown in Figs. 1a and 1b, respectively, support network-centric clusters. As already discussed, the fronthaul traffic in CRAN is substantial. The O-RAN architecture can alleviate this issue by splitting the functions in CU, DU and RU. Regarding the PHY performance, the set of RUs in the network-centric clusters are not always optimal. In a user-centric cell-free massive MIMO system however, the UE can connect to the optimal set of RUs, as depicted in Fig. 1c. While user-centric sets of RUs are optimal from the PHY perspective, they introduce the necessity of a fronthaul network to connect RUs and DUs, as well as challenges to its realization. Although user-centric cell-free massive MIMO systems introduce challenges compared to cellular networks, they also solve problems such as inter-cell interference and frequency reuse planning, which are rooted in the rigid cell-centric architecture of cellular networks .

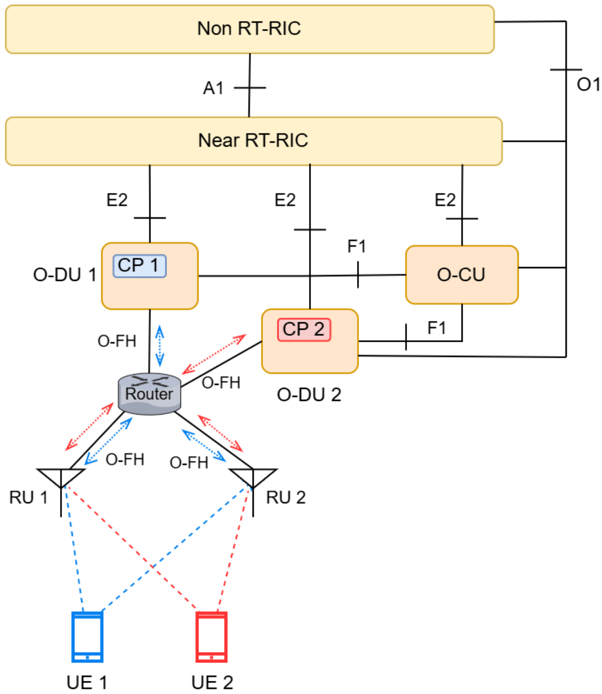


Figure 2: This figure shows how a possible extension to the O-RAN architecture could look like to facilitate a cell-free design. In particular, the addition of routing capabilities to the fronthaul will play a key role.

III. Challenges

In this section, we will highlight some of the major open challenges that a cell-free approach would bring to the higher layers of the network stack beyond the physical layer.

A. *High-level architecture*

In the O-RAN architecture, RUs are controlled and coordinated by a cloud-native infrastructure composed of the DUs and CUs, with orchestration and management provided by the Near-Real-Time Radio Intelligent Controller (near-RT RIC) and the Non-Real-Time RIC (non-RT RIC). O-RAN standards support two splits in the DL - 7.2a where the precoding operation is done at the DU, and 7.2b where the precoding is done at the RU. For a cell-free system, the 7.2b split option requires less bandwidth in the fronthaul than 7.2a [6]. Hence, from a practical deployment perspective 7.2b might be preferable to 7.2a. In addition, 3GPP proposed split option 7.3 for the DL, which, in contrast to split options 7.2a and 7.2b, places the modulation (from information bits to symbols) at the RU [7].

This also reduces the overall required fronthaul bandwidth. Given that the fronthaul network is often already considered the bottleneck [8] and that in cell-free MIMO multiple streams of the same data must be transmitted to the RUs in a user-centric cluster, minimizing the required fronthaul resources becomes imperative.

The UL 7.2b split considers UL beamforming at the RU to reduce the fronthaul load compared to forwarding the received signal at each RU antenna to the DU, but places most other functions at the DU. For further performance improvement and fronthaul bandwidth reduction, the UL 7.2b split may need to be modified by moving the channel estimation and equalization functions from DU to RU, as recently proposed in [9] as "7.2x Cat-B ULPI-A". This allows local processing of data at the RUs, i.e., each RU in a user-centric cluster computes a local estimate of the user's UL data symbol and sends it to the CP, which is the entity processing the higher PHY layers located at some DU site. The CP then computes a cluster-level symbol estimate based on the local RU symbol estimates. In order to enable cluster-level processing in a user-centric cell-free MIMO system, the O-RAN fronthaul requires traffic of a given UE to reach its assigned cluster processor irrespective of which RU it is connected to. This requires moving away from point to point RU-DU links to a more dynamic routed fronthaul architecture, where traffic between RUs and DUs can be dynamically routed both in UL and DL direction. This could be enabled by using routed fronthaul as seen in Fig. 2. This also is the most significant change that would be required to the existing O-RAN architecture. In the remainder of this paper, we discuss the challenges of cell-free massive MIMO from a network perspective considering the use of split option 7.3 in the DL and 7.2x Cat-B ULPI-A in the UL (as in the recent works [1], [10]).

B. Clustering and cluster processor placement

In a cell-free network, each UE is served by a subset of all RUs in the network [5]. We consider the user-centric case, where the cluster of RUs is determined for each UE individually. The objective of clustering is to form a cluster for each UE that includes all RUs with a potentially significant channel gain. In this way, the RUs in the cluster contribute to the useful DL signal, while the signals transmitted by the remaining RUs are weak enough to not cause significant interference (analogously for the UL). In the O-RAN architecture, some cooperation between the RUs and DUs using the Near-RT RIC is required to define the user-centric cluster of the UE and to find an available pilot sequence for channel estimation at the RUs in its cluster. However, the data exchange between RUs, DUs and a more centralized unit (such as the Near-RT RIC) necessary for said cluster/pilot assignment has been very little studied for cell-free massive MIMO. We note that it is a common approach to assign clusters and pilots together due to pilot reusability constraints, i.e., a RU can obtain an accurate channel estimate of a given UE only if the pilot sequence is not transmitted by another UE with a large channel gain to the RU [1]. Furthermore, pilots and clusters are typically assigned based on second-order channel statistics, i.e., the channel covariance matrix describing the channel gain and directivity [1].

Regarding cluster formation and cluster processor placement, and the roles of RUs, DUs and the near-RT RIC therein, we will first discuss one approach from [5], followed by a more general discussion. The cluster formation procedure in [5] considers that each UE has a primary RU associated to the primary DU. The primary DU is the unit that controls the primary RU. It automatically takes on the role of cluster processor and forms a cluster centered on the primary RU. This results in a user-centric cluster consisting of the primary RU (linked to the primary DU) and a set of secondary RUs, which may be associated with the same DU or with one or more secondary DUs. In the UL (DL data transmission is not considered in [5]), each of the DUs computes one symbol estimate for their associated RUs in the cluster (based on local RU estimates) and sends it to the primary DU. The primary DU then computes the "final" cluster-level symbol estimate using its own local estimate and the local estimates of the secondary DUs. Note that this UL split is different from our preferred option, where only one DU is involved in the cluster-level signal processing. The approach in [5] could be modified to match our option by directly sending the local RU estimates to the primary DU, while the secondary DUs (or collocated routers) are only used to forward packets.

In an alternative cluster formation approach, there is no fixed RU-DU association, and the location of the cluster processor (CP) can be freely selected for each individual UE. In this setup, RUs and DUs communicate via a routed fronthaul network, which may include routers located at the DUs or between RUs and DUs (we will further discuss a routed fronthaul network in section III-E). The CP for each UE's cluster can be dynamically placed at one of the DUs, optimized according to metrics such as load, latency, or fronthaul capacity. The RUs serving the UE can then send their local symbol estimates over the routed fronthaul to the DU hosting the CP of that UE. In both this and the fixed-association case, a routed fronthaul network is required.

Assuming RUs, DUs, and the near-RT RIC can all exchange information, the clustering decision itself (i.e., selecting which RUs belong to a UE's cluster and assigning a pilot) should ideally be made from a global network perspective. This responsibility can lie either with the near-RT RIC or be delegated to individual DUs. For example, one method for clustering and pilot assignment would be to let the Near-RT RIC gather information on channel gains from all requested UE-RU associations and on the pilot availability at the RUs. Then, it could assign the cluster and pilot together with the cluster processor placement at one of the DUs. Alternatively, the Near-RT RIC could define a DU as the cluster processor for a UE and communicate the set of potential RUs for the cluster and their pilot availability. The DU could then assign a pilot and RUs to the user-centric cluster, which would result in a more distributed way of cluster and pilot assignment.

It remains to be studied what exact roles the different network entities will have for pilot and cluster assignments and for cluster-level signal processing. While fixed RU-DU associations result in a simpler fronthaul load routing problem (since the CPs

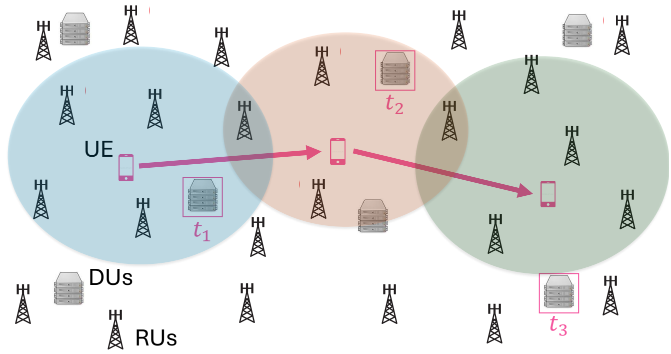


Figure 3: Example reclustering of a given mobile user in time instances t_1, t_2 and t_3. The DU, at which the CP of the user is located in the different time instances, is marked by a box.

are assigned by the primary RU), dynamic RU-DU associations may lead to a more balanced distribution of the fronthaul load (on the fronthaul links) and the cluster processors (at the DUs).

C. Reclustering and cluster processor migration

In a practical user-centric cell-free massive MIMO network, it will be necessary to update the user-centric RU clusters and the CP placement. Reasons for this include UE mobility, pilot reuse, overload of the fronthaul links or DU computational capacity, increased latency for data exchange between UEs and the CP, and changing UE traffic demand. We notice that reclustering triggers based on channel gains alone are insufficient to address the challenges of cell-free networks. Reclustering may be necessary even with strong signals when RUs or DUs are overloaded or fronthaul links become congested. Meeting diverse and dynamic QoS requirements also demands flexible migration mechanisms. This area remains underexplored, and new reclustering triggers could significantly improve network efficiency and adaptability.

Under the assumption that fronthaul networks are planned for high user loads, we consider that reclustering and cluster processor migration is rarely triggered by an overload of the fronthaul links or DU computational capacity. Therefore, in the following, we focus on the case where the reclustering is triggered by UE mobility, causing changes of the channel gains.

The user-centric cluster is updated by removing RUs from it and adding new ones. A scheme to drop RUs can be implemented in a simple way by letting the responsible entity of the cluster (CP or Near-RT RIC) remove a RU from the cluster when a given channel gain threshold is reached. Adding new RUs is more challenging because of pilot availability at the RUs. In general, we can assume that the UEs receive some sort of pilot signal from all RUs (located closely) to monitor their channel gains. If the channel gain of a given UE is above some threshold, the UE can request the RUs to join the user-centric cluster. The RUs in turn would then request at the CP or near-RT RIC to join the cluster of the given UE. Assume that a UE has a given pilot sequence and moves towards a RU that is not part of its cluster, such that the channel gain meets the threshold. If the RU does not use the pilot for another UE, it can be added to the cluster (after some communication with the cluster processor or near-RT RIC). In the other case, a mechanism is needed to determine how the cluster and pilot assignment for a UE should evolve when additional RUs are to be included in the cluster. This process involves coordination between the cluster processor and the Near-RT RIC, both of which have information about pilot availability at the RUs.

The migration of the cluster processor from one DU to another DU is a similar problem. When the user-centric cluster evolves, the cluster processor may be located quite far from the RUs currently serving the UE. This may result in many hops between the RU and the cluster processor, and can significantly increase the RU-CP latency. Therefore, it may be necessary to move the cluster processor to another DU to reduce the RU-CP latency, as illustrated in Fig. 3. To enable the DU migration of the cluster processor, the near-RT RIC should be aware of the fronthaul link loads, the RU-CP latency of each user and the processing loads of the DUs. Based on this information, the near-RT RIC should then make decisions (based, e.g., on thresholds for the DU processing load or for the RU-CP latency) on cluster processor migration and CP-RU traffic rerouting.

We notice that it may not always be possible to include all RUs with a significant channel gain in the user-centric clusters, which comes at the cost of reduced PHY performance. Here, an alternative solution would be to schedule the users that use the same pilot or fronthaul links on different time-frequency resources. While not degrading the PHY performance, this would require a scheduler that considers conflicts of users that shall not be scheduled together on the same time-frequency resource, as considered, e.g., in [11].

D. Overhead calculation

One of the main requirements to conduct proper network planning for a cell-free network, i.e. deciding on what kind of fronthaul transport technologies to use, how much computing DUs need etc., is a good understanding of the datarates that will be present in the network. These not only depend on the amount of user data, but also largely on the overhead created by the functions related to the physical layer. We will now discuss why this overhead is hard to estimate. Here, we consider overhead from the point of view of the user plane IP traffic for the fronthaul network design. This is because the control plane operations, such as the coordination of pilot and cluster assignments do not significantly increase the fronthaul load compared to the fronthaul traffic associated with the data plane. The information used for cluster and pilot assignment (e.g., the channel gains and the availability of pilots) is small compared to the user plane payload data such as video. In addition, the time scale on which the second-order statistics change is typically on the order of a few seconds, while the payload transmission slots have a duration of up to one millisecond in 5G. As a result, cluster and pilot allocation requires much less frequent information exchange and smaller packets on the fronthaul in relation to the payload data.

As discussed in Section III-A, we consider a 7.3 split for the DL and a 7.2x Cat-B ULPI-A in the UL. With split option 7.3 for the DL, information bits are transmitted from the DU to the respective RUs. Apart from protocol overheads, these include Forward Error Correction (FEC) bits which are added to the overhead calculations. A coding rate of m/n means each n-bit codeword contains m information bits and $n-m$ redundant bits. If channel conditions between a UE and it's cluster of serving RUs are good, a higher coding rate can be used, which results in a lower overhead while poorer conditions result in the transmission of more error correction bits and a higher overhead. In 5G NR, coding rates typically range from approximately 0.0292 to 0.9258, meaning that at the lowest rate, less than 3% of the transmitted bits carry user data [12].

With the UL split option 7.2x Cat-B ULPI-A, for each user UL symbol received at a given RU, the RU sends an in-phase and quadrature-phase (IQ) sample describing the local RU symbol estimate to the DU, at which the cluster processor of the corresponding UE is located. Here, the overhead depends on the applied Modulation and Coding Scheme (MCS) i.e. the coding rate as well as the modulation order. Let us consider a given number of bits a user wants to transmit in the UL. In the case of good channel conditions, a higher modulation order can be used, which allows more information bits to be sent per symbol. In general, this reduces the required overhead since less symbols must be sent in total. Poor channel conditions result in a lower modulation order and therefore a higher overhead. In 5G NR, modulation schemes range from QPSK (2 bits per symbol) to 256-QAM (8 bits per symbol). The overhead calculation for split option 7.2x in a cellular O-RAN network is done in [13] assuming a given load per cell, while [14] shows that the 7.2x split creates almost 9 times more traffic compared to split 7.3 for lower modulation levels.

The UL overhead not only depends on the chosen split option, but also heavily on the quantization method. A common approach is the midrise uniform quantizer function with a fixed number of quantization bits. An alternative approach is to define a common distortion level for the whole network and let the quantization rate be dependent on the received signal strength at the RU, leading to quantization rates close to the rate-distortion limit [1], [10]. The latter scheme has the advantage that the fronthaul load is dependent on the received signal power, i.e., "strong" signal samples are quantized with a larger precision. In the DL with split option 7.3, no quantization is required since the information bits are sent from the DU to RUs, which by definition is the most efficient way.

It is important to point out the difference between UL and DL with these split options for a cell-free network. In the DL, the same information bits are sent from the cluster processor to the RUs. This can be realized either through multicast or unicast. While multicast support is widely available in most vendors devices, it still might not be straightforward to deploy and use in a fronthaul setting. If the fronthaul network does support multicast transmissions, the routers can forward the same data to different RUs. However, in the case that there is only support for unicast, the CP must send multiple copies of the same data, one for each RU. This increases computing overhead for both the CP as well as a higher bandwidth demand on the fronthaul links. In the UL, the exchanged data is unique for each UE-RU pair. For each user, the samples from each RU in the user-centric cluster must be sent individually through the fronthaul network to the CP responsible for processing the given user-centric cluster.

All of this has a direct impact on network planning for the fronthaul, and an appropriate model is necessary to properly estimate the incurred overhead for transmitting a given amount of application layer traffic in order to correctly dimension the network resources. Given the large number of parameters and variables involved, developing and verifying an accurate model remains difficult without first collecting real-world data. Even simple testbed setups, which are already under development at several institutions, will be valuable to estimate the actual overhead levels.

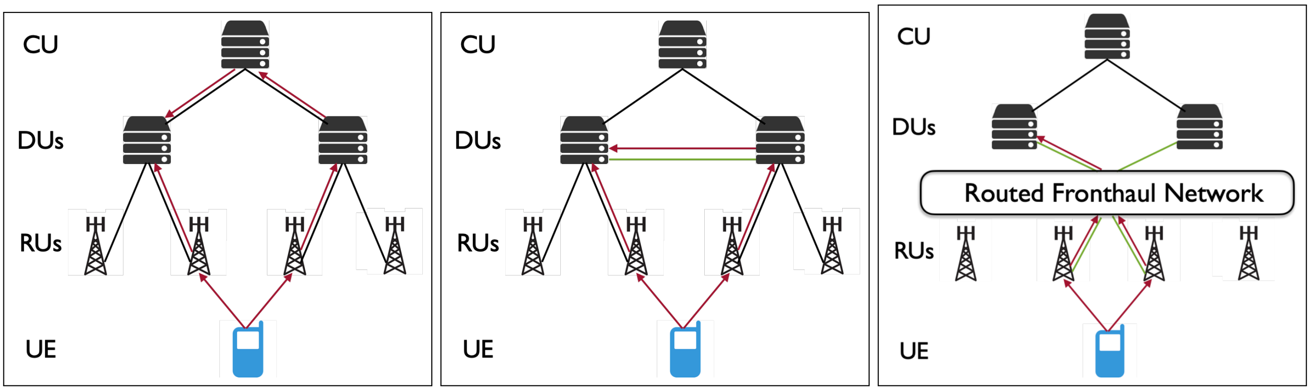
E. Fronthaul topologies

As mentioned in the previous sections, a cluster of RUs is serving a given UE in cell-free networks and must communicate with the CP. As a result, there is now a requirement to bring traffic that originates at one RU to a DU that is not connected directly. Fig. 4 shows different fronthaul network designs supporting user-centric clusters. The current O-RAN design, as shown in Fig. 4a, would involve data exchange through the CU, resulting in longer latencies and high bandwidth demands on the DU-CU links. Different designs, using either inter-DU links as shown in Fig. 4b, or even a fully routed fronthaul in 4c, could alleviate these issues.

Major challenges and research questions arise due to the requirement of having to route data to the CP. On one hand, it will be imperative to find good topologies that can accommodate this type of traffic. Here, it will be important to strike a balance between the complexity (and thus deployment cost), scalability (mmWave frequencies offer upwards of 1 Gbps of user data per UE) and energy efficiency (electrical switching will increase power consumption). On the other hand, the selection of which UE is assigned to which DU is no longer trivial. Here, as this problem is computationally hard, fast heuristics will be required to balance the additional fronthaul network overhead that suboptimal assignments bring with computational overhead caused by frequent reassignments and migrations. Finally, current protocols used in fronthaul networks, in particular eCPRI, do not contain any fields in their headers that could be used for addressing or routing. In particular, it is unclear where the association that maps a UE to a CP could be encoded. It will need to be investigated, how feasible relying entirely on the MAC layer, e.g. utilizing VLAN-tags in this regard will be. Future research will also need to focus on what type of routing will best work in this setting. The available options here reach from simple shortest path all the way to a full software defined network.

To further show the need for a new approach and suggest one possible solution, we use the physical layer model described in [10] to generate fronthaul data demands to determine how current topologies such as PONs with added inter-DU links perform in such scenarios.

Our results are shown in Fig. 5. It shows the required bandwidth on the link with the highest amount of traffic depending on the number of RUs in the topology. We investigated two different types of topologies here. On the one hand there are



(a) (b) (c)

Figure 4: This figure depicts different radio access network design types. In 4a, the cell-free paradigm is used to allow the UE to connect to optimal RUs. However, the topology only allows for very inefficient routes, i.e. having to route through the CU. 4b and 4c extend the current state of the art networks by either adding links between DUs or routers in the fronthaul to allow for more efficient routes. While 4b is easier to deploy, it is not clear how scalable such a topology is compared to 4c.

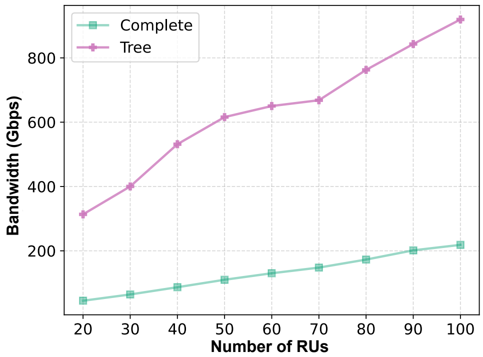


Figure 5: Figure showing the bandwidth demand in Gbps for the link with the most amount of traffic for different number of RUs.

the tree-like topologies of passive optical networks where each RU is connected to a DU. If the data needs to be forwarded to a different DU as a UEs CP is located there, inter-DU links need to be used. These inter-DU links are also the bottlenecks and show the highest bandwidth requirement. The second topology we investigated are complete graphs, where each RU can be directly connected to multiple DUs as required. While these are completely unfeasible in practice, we used them as a baseline for comparison. Our results indicate that, as expected, these topologies quickly reach traffic demands that exceed the capacities of currently available technologies such as 50G-PON. It should be noted that the results shown in Fig. 5 are obtained in a sub-6 GHz system. For higher carrier frequency regimes in 6G, larger channel bandwidths will be used, resulting in even further increased fronthaul traffic demands.

The gap between PON and complete graph indicates that there will likely be a good middleground that is feasible to deploy. As a result, a routed fronthaul network will likely become a necessity. This requires significant changes to how we design fronthaul networks, demanding new topologies, technologies and protocols.

IV. Conclusion

Table I: Cell-Free 6G: Networking challenges and solution directions.

Challenge	Solution Directions
Fronthaul bottlenecks from multi-RU service	Use efficient splits (7.3 DL, 7.2x Cat-B ULPI-A UL), local RU processing, and routed fronthaul
Signaling overhead from dynamic clustering	Near-RT RIC or DU coordination; joint pilot/cluster assignment; flexible CP placement
Frequent reclustering due to mobility/overload	Smarter triggers; CP migration to reduce RU–CP latency; pilot-aware scheduling
Uncertain overhead from coding and quantization	Develop models; adaptive quantization (rate–distortion); validate in testbeds
Rigid tree-like fronthaul topologies	Inter-DU links or routed/SDN fronthaul; balance scalability and cost; extend beyond eCPRI

In this paper, we gave an overview of some of the most pressing network related challenges that stand in the way of realizing an actual cell-free deployment. We covered aspects related to architecture, topologies, clustering and migration, which are also summarized in table I. Overall, it should become clear that there will be significant research required to overcome these challenges. As the physical layer community is now beginning to reach a consensus on split options and the first network components supporting cell-free MIMO are being manufactured, we will soon be able to gain experimental data to validate possible models and algorithms against. In particular, this will give us a better understanding of the overhead. Using this, we will be able to design feasible cell-free networks and create full end-to-end simulators and demonstrators. These will ultimately show how serious the issues discussed here are and how feasible a path to cell-free deployments will be.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Research, Technology and Space in the programme of "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification numbers: 16KISK020K, 16KISK030, 16KISK031 and 16KISK032.

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*, 2024.

[2] L. Miretti, R. L. Cavalcante, and S. Stańczak, "Two-timescale joint power control and beamforming design with applications to cell-free massive MIMO," *IEEE Transactions on Wireless Communications*, 2025.

[3] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive mimo networks: A survey of opportunities, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611–652, 2021.

[4] R. Fantini, W. Zirwas, L. Thiele, D. Aziz, P. Baracca, A. Osseiran, J. Monserrat, and P. Marsch, "Coordinated multi-point transmission in 5G," *5G Mobile and Wireless Communications Technology*, pp. 248–276, 2016.

[5] R. Beerten, V. Ranjbar, A. P. Guevara, and S. Pollin, "Mobile Cell-Free Massive MIMO: A Practical O-RAN Based Approach," *IEEE Open Journal of the Communications Society*, 2024.

[6] MARSAL Consortium, "Report on testing and integration of the du-ru and cu-near-rt-ric," deliverable d3.5, version v1, MARSAL Project (EU H2020), Sept. 2024. Public deliverable of the MARSAL project.

[7] 3GPP, "Study on Central Unit (CU) - Distributed Unit (DU) lower layer split for NR (Release 15)," Technical Report TR 38.816, 3rd Generation Partnership Project, Dec. 2017.

[8] M. Waqar, A. Kim, and P. K. Cho, "A study of fronthaul networks in crans-requirements and recent advancements," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 10, pp. 4618–4639, 2018.

[9] Ericsson, "Driving Open RAN forward – An improved open fronthaul interface bringing performance to Open RAN," tech. rep., 2023.

[10] Z. Li, F. Göttsch, S. Li, M. Chen, and G. Caire, "Joint Fronthaul Load Balancing and Computation Resource Allocation in Cell-Free User-Centric Massive MIMO Networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14125–14139, 2024.

[11] F. Göttsch, N. Osawa, I. Kanno, T. Ohseki, and G. Caire, "Fairness Scheduling in User-Centric Cell-Free Massive MIMO Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 9, pp. 11942–11957, 2024.

[12] "5G NR; Physical layer procedures for data (3gpp ts 38.214 version 17.1.0 release 17)," Tech. Rep. TS 38.214, 3rd Generation Partnership Project (3GPP), 2024. Available online at ETSI.

[13] A. Ndao, X. Lagrange, N. Huin, G. Texier, and L. Nuaymi, "Optimal placement of virtualized DUs in O-RAN architecture," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–6, IEEE, 2023.

[14] S. Bidkar, P. Dom, R. Bonk, and T. Pfeiffer, "Mobile Xhaul traffic modelling for high-speed TDM-PON," in *2020 European Conference on Optical Communications (ECOC)*, pp. 1–4, IEEE, 2020.