Cell-free mMIMO Support in the O-RAN Architecture: A PHY Layer Perspective for 5G and Beyond Networks

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

10-08-2021 / 14-08-2021

CITATION

Ranjbar, Vida; Girycki,, Adam; Rahman, Md Arifur; Pollin, Sofie; moonen, marc; Vinogradov, Evgenii (2021): Cell-free mMIMO Support in the O-RAN Architecture: A PHY Layer Perspective for 5G and Beyond Networks. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.15144516.v1>

DOI

[10.36227/techrxiv.15144516.v1](https://doi.org/10.36227/techrxiv.15144516.v1)

Cell-free mMIMO Support in the O-RAN Architecture: A PHY Layer Perspective for 5G and Beyond Networks

Vida Ranjbar Adam Girycki Md Arifur Rahman Sofie Pollin Marc Moonen Evgenii Vinogradov

Abstract—To keep supporting next-generation requirements, the radio access infrastructure will increasingly densify. Cell-free (CF) network architectures are emerging, combining dense deployments with extreme flexibility in allocating resources to users. In parallel, the Open Radio Access Networks (O-RAN) paradigm is transforming RAN towards an open, intelligent, virtualized, and fully interoperable architecture. This paradigm brings the needed flexibility and intelligent control opportunities for CF networking. In this paper, we document the current O-RAN terminology and contrast it with some common CF processing approaches. We then discuss the main O-RAN innovations and research challenges that remain to be solved.

Index Terms—Cell-free, O-RAN.

I. INTRODUCTION

Communication networks are the nervous system of modern society and the economy. Our world becomes more and more data-centric: the number of sensors is skyrocketing. On the other hand, productivity growth relies on creating customer-centric services and automation of manufacturing processes [1]. Society looks forward to exciting new experiences such as augmented/virtual reality or holographic calls/telepresence [1], [2]. 5G made significant progress toward accommodating all these needs and expectations. However, it is predicted that by 2030 this technology will not be able to satisfy the data rate requirements imposed by the fully digitalized world [1]. The above discussion has motivated academia and industry to start several advanced research activities in order to shape the vision of future 6G networks that are well summarized in [1], [2].

In this article, among all technological innovations pertaining to 6G, we would like to highlight novel network architectures such as cell-free (CF, or cell-less) massive multiple-input-multiple-output (mMIMO) since this architecture offers high Spectral Efficiency (SE) [3] required in 6G [1]. Moreover, the authors of [4] claimed that CF mMIMO networks offer the following benefits: i) large energy efficiency (i.e., reducing carbon footprint); ii) flexible and cost-efficient deployment; iii) uniform Quality of Service (due to absence of cell-edge

users and inter-cell interference); iv) favorable propagation conditions (lower path loss and blockage probability).

There are some fascinating attempts to analyze or even implement CF mMIMO networks using centralized processing (e.g., in [5] and [6], respectively). As a result of these works, it becomes clear that this novel communication architecture will need to rely on distributed processing to become scalable and, consequently, feasible [7]. This article goes even further in the practical domain. It presents several distributed CF mMIMO architectures aligned with the Open Radio Access Network (RAN) vision [8].

The Open RAN movement is primarily led by two co-operating industrial groups: i) Telecom Infra Project (TIP) focused on deployment and execution and ii) the O-RAN ALLIANCE focused on developing and driving standards to ensure that equipment from multiple vendors inter-operate with each other. The O-RAN ALLIANCE suggests splitting the RAN into smaller components beyond the radio and the baseband unit. Moreover, O-RAN will open the internal RAN interfaces and allow us to create an open ecosystem where there will be different vendors providing different solutions and generate a complete end-to-end ecosystem. This approach is exceptionally operator-friendly due to so-called disaggregation allowing to pick and choose different RAN components from different vendors. Potentially, it would bring more vendors into the wireless industry by allowing smaller companies to occupy their niche in the market. Moreover, the specialization would accelerate technological innovation by upgrading smaller RAN components without waiting for the entire radio or baseband unit to be upgraded.

The main goal of this paper is to introduce how the CF mMIMO architecture links to the O-RAN-specified architecture. We provide a brief overview of CF mMIMO literature and identify implications of the design choices on the O-RAN architecture. Next, the design trade-offs are discussed through a measurement-based performance study using a 64-antenna CF mMIMO deployment.

II. CELL-FREE NETWORKS

CF mMIMO network refers to a network with many distributed Access Points (APs) cooperatively serving User Equipment units (UEs) through coherent joint transmission and reception [3] using the same time-frequency resources. Consequently, the concept of cells is eliminated, motivating the name. The APs are connected via fronthaul links to

Vida Ranjbar, Sofie Pollin, Marc Moonen, and Evgenii Vinogradov, are with Department of Electrical Engineering, KU Leuven, Belgium.

Adam Girycki and Md Arifur Rahman are with IS-Wireless, Pulawska Plaza, ul. Pulawska 45b, 05-500 Piaseczno, Poland.

This research has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017171 (MARSAL project).

(Vida Ranjbar and Adam Girycki are co-first authors; Corresponding author: Vida Ranjbar vida.ranjbar@kuleuven.be)

central processing units (CPUs) responsible for the coordination. The CPUs are interconnected by backhaul links. Precoding/combining operations can be performed locally at each AP, centrally at the CPU, or the processing can be distributed between APs and CPUs.

A. Distributed Processing

Initial works (e.g., [3] by Ngo et al.) assumed a single CPU operating in a Network MIMO fashion (i.e., a centralized approach in which the CPU is responsible for coordinating and processing the signals of all UEs). In this case, APs are quite simple since they only need to receive the signal and transfer it to the central entity (CPU). This approach demonstrated very high SE in simulations [3]. Moreover, Wang et al. [6] put impressive efforts into implementing a cloud-based CF mMIMO network working in a centralized manner. In their computationally advanced testbed, 32 general-purpose servers (with four 18-core processors each) constructed a CPU that can serve up to 16 eight-antenna UEs with 16 eight-antenna APs.

However, the implementation feasibility is questionable for wider-scale networks with more APs and UEs. Indeed, the centralized approach implies that the fronthaul capacity and computational resources, required for each AP to process and share the data signals related to all UEs, grow linearly (or faster) with the number of UEs [7]. In other words, the original form of Cell-Free Massive MIMO was unscalable.

Björnsson and Sanguinetti [7] suggested a decentralized network architecture in which each (more advanced) AP locally estimates the channels of its associated UEs and uses this information to locally process data signals. In this way, achievable SE is lower but expensive fronthaul, and computational resources are utilized more rationally. In our work, we adopt this idea. However, we also assess the performance of several architectures using centralized processing.

B. Inter-CPU Coordination and Cluster Formation

Another popular but clearly unrealistic assumption (also made in [3], [4], [6], [7]) is utilization of a single CPU. A large-scale CF mMIMO network consisting of APs connected to several CPUs is investigated in [9].

Inter-CPU coordination is tightly linked to the idea of AP clustering described in [7]. In this approach, each UE is served by a subset of APs called "cluster" instead of the full set of APs in the network. By this, the scalability is improved by confining the signal co-processing within the cluster. Two distinctive clustering approaches may be identified: i) network-centric and ii) user-centric.

Network-centric clustering (NCC) consists of deploying fixed disjoint clusters of APs serving only the UEs residing in their coverage area. Each network-centric cluster is connected to a single CPU. In this configuration, the clusters mutually interfere unless they cooperate (e.g., through coherent transmission which negatively affects the network scalability). Non-cooperative NCC is fully scalable, but it suffers from poor performance as we demonstrate later.

Fig. 1. Cell-free system with user- and network-centric clustering and multiple interconnected CPUs. Two types of users are depicted: *local users* (UE_{2,3}) and *cluster-edge users* (UE₁) served by a single or multiple network-centric clusters, respectively. The alternative naming of the network nodes (in brackets) is aligned with O-RAN specifications.

In **user-centric clustering (UCC)**, dynamic clusters of APs are formed based on the needs of each UE. Usually, the user is served by nearby APs offering the best channel (e.g., lowest path loss). Clusters formed in this way may i) partly overlap and ii) consist of APs connected to several CPUs. UCC results in lower interference; thus it offers better SE than the NCC. However, this concept requires more control signaling to dynamically coordinate several clusters. Moreover, when the number of users (and, consequently, clusters) grows, we may face the need for very intensive inter-CPU coordination when many UEs are served by APs not connected by fronthaul to the same CPU.

Interdonato et al. [9] presented a hybrid approach where the CF mMIMO network was shaped by several network-centric clusters. However, each user also forms a cluster in a user-centric manner (APs can belong to different network-centric clusters). Specifically, for a given UE

- 1) A set of serving APs is identified in a user-centric fashion: they are selected according to some metric (e.g., distance or channel quality).
- 2) The clusters involved by the selected APs define the serving network-centric clusters.
- 3) The data to the UE is distributed only to the involved CPUs.

In our work, to avoid all the signaling overhead associated with UCC mentioned earlier, only to serve the users at the edge of the NCC clusters, the clusters will cooperate by exchanging local estimates of cluster-edge users. A similar approach can be found in [10].

C. Key Design Choices in CF mMIMO Network Design

Summarizing our reasoning above, we can emphasize the importance of two key points:

- 1) Which network unit performs precoding/combining (AP vs. CPU)?
- 2) Do we consider multiple CPUs and what level of Inter-CPU coordination is acceptable?

Answers to these questions define fronthaul and backhaul requirements, computational resources available at the network elements (i.e., APs and CPUs), and the achievable network performance.

III. CELL-FREE O-RAN SUPPORT

As it follows from the above reasoning, CF mMIMO architectures require fine granularity of the processing options (processing at AP vs CPU, Inter-CPU coordination), which can be offered by hardware and software supporting the Open RAN disaggregation concept. In order to put the CF mMIMO idea out of the academic bubble and make this concept clearer for industry actors, we aim to align the CF mMIMO terminology with the O-RAN alliance vision on communication networks.

Note that "Open RAN" is the movement to disaggregate hardware and software and to create open interfaces between them. "O-RAN" refers to the O-RAN Alliance, which publishes new RAN specifications, releases open software for the RAN, and supports its members in integration and testing of their implementations.

A. Disaggregation and Terminology Alignment

An essential part of the NG-RAN disaggregation is already defined in the 3GPP TS 38.401 [11], where the gNB is split into Centralized Unit Control Plane (CU-CP), CU - User Plane (CU-UP), and Distributed Unit (DU) nodes connected by corresponding interfaces. The DU takes over the data-processing responsibilities while the CU handles the control and data plane connections to the core network.

The O-RAN standard extends the 3GPP disaggregated NG-RAN by the introduction of the Near-Real-Time RAN Intelligent Controller (Near-RT RIC). According to O-RAN, the nodes of the disaggregated NG-RAN are named O-CU-CP, O-CU-UP and O-DU, respectively. The O-RAN specifications also bring the next level of NG-RAN disaggregation by splitting the 3GPP defined DU functionalities into O-DU and O-RU and introducing the open fronthaul interface between them.

Consequently, now we can translate the CF mMIMO network components presented in the previous section to the O-RAN terminology. The CPU is disaggregated into O-CU-CP, O-CU-UP, and O-DU while the AP is represented by Radio Unit (O-RU). Note that in the following, we will use this terminology instead of the one used in [3]–[7], [9].

B. O-RAN Disaggregation of the PHY Layer

The O-RAN is working on several so-called "functional splits" creating a division somewhere amid the many steps

between a signal's arrival from the core network and its wireless transmission to a UE. By applying different splits, it is possible to adjust the distribution of the processing between different parts of the network (e.g., O-CU, O-DU, and O-RU) and optimize the CAPEX/OPEX according to the operator's strategy.

The choice of how to split the PHY layer depends on factors related to radio network deployment scenarios, constraints, and intended supported services for example:

- Support of specific QoS per offered services (e.g., low latency, high throughput);
- Support of specific O-RU/UE densities and load demand per given geographical area;
- Availability of transport networks, their capacity, and density;
- The service level agreement requirements of the users.

The most interesting options for the CF mMIMO network are Split 8 and Split 7.2x. In Split 8 (used in [6]), O-RUs are responsible only for converting signals while even the beamforming is shifted to the O-DU. This split requires very high fronthaul capacity. On the other hand, it can be attractive for a network where a vast number of (low price) O-RUs have to be deployed. Note that despite its very high level of centralization, Split 8 is not the only option for designing centralized CF mMIMO networks.

In Split 7.2x, O-DUs are responsible for signal encoding and decoding, while O-RUs perform some light-weight processing (see Fig. 2). In this approach, O-RUs are slightly more complex but the fronthaul requirements are lower. Additionally, for the flexible deployment of PHY layer functionalities, the O-RAN suggested Category A and Category B O-RUs [8].

Category B O-RUs are more advanced and able to perform precoding locally which is needed for architectures relying on local processing (e.g., similar to [7]). Category A O-RUs [8] can be used for centralized CF mMIMO networks while having lower (than Split 8) fronthaul requirements. A more detailed description of the operations performed by O-RUs of different categories (and their O-DUs) is shown in Fig. 2.

C. Cell-free Deployment Options Based on O-RAN Split 7.2x

Motivated by the flexibility offered by Split 7.2x, we think that it will enable practical implementations of future CF mMIMO network architectures as shown in Fig. 3.

Considering the key design choices presented in the previous section (i.e., centralized/local processing and absence/presence of Inter-DU cooperation), it is possible to define the following implementation options:

- *Option 1:* Local precoding/combining (requiring Category B O-RU) and no coordination between O-DUs.
- *Option 2:* Local precoding/combining; O-DUs cooperate to serve cluster-edge users.
- *Option 3:* Precoding/combining is performed by each O-DU based on the signal received by all O-RUs constructing its network-centric cluster. Low complexity O-RU of Category A can be used; O-DUs do not cooperate.
- *Option 4:* Same as Option 3 but O-DUs exchange their estimates for cluster-edge users.

Fig. 2. Functionality of O-RUs. Category A O-RUs are suitable for centralized CF mMIMO processing (i.e., at the DU). Category B O-RUs are able to perform precoding locally.

Fig. 3. Cell-free O-RAN architecture supporting both centralized and local processing (depending on the used O-RUs). When a UE is served by O-RUs connected to different O-DUs, Inter-DU coordination may be enabled in order to boost the performance.

- *Option 5*: Signals from all O-RUs connected to all O-DUs are processed centrally. This may require appointing a master O-DU which would gather all O-RU signals relayed by their O-DUs (replicas). Next, the primary O-DU computes precoding/combining vectors.

D. O-RAN Cell-free Random Access Procedure

In a real-life network, we will have multiple DUs in charge of a few O-RUs. In this part, we try to demonstrate a CF mMIMO architecture using a hybrid version of the 7.2x split in which the O-RUs are clustered between O-DUs assuming

Fig. 4. User categorization during random access procedure.

each O-RU is connected to one specific regional O-DU. In this scenario, the users will be divided into two main categories: cluster-edge users and local users. Local users will be served by the O-RUs connected to one O-DU (i.e., O-RUs belong to one network-centric cluster) while the cluster-edge users should be served collaboratively by the O-RUs of two (or more) O-DUs. The decision that whether a user will be treated as a local or a cluster-edge user is done into three phases as below.

The 3GPP TS 38.321 [12] describes the random access procedure which is controlled by the MAC protocol. The first step of the random access procedure is sending the random access preamble by the user MAC and delivering it to the MAC in the O-DU. This step can be divided into the following three phases:

- 1) User sends the random access preamble with a specific random access preamble ID.
- 2) The random access preamble ID can be detected by several O-RUs. Each O-RU sends the received random access preamble ID, together with its received signal strength level, to its O-DU.
- 3) In case of the NCC, each O-DU may inform other O-DUs about the fact of reception of the random access preamble ID including its signal strength. If the random access was received by more than one O-DU and the difference between the signal strength measured by different O-DUs is less than some configured value (e.g. 9 dB), then the user is considered as the *cluster-edge user*. Otherwise, the user is categorized as the *local user*, which is illustrated in Fig. 4. A similar algorithm could be used to define user-centric clusters in the UCC solution.

In this approach, the network determines the user category (local or cluster-edge). The user does not impact the categorization process.

E. Near Real-Time RAN Intelligent Controller

In order to optimize the network capacity and user data transfer performance (e.g., data rate, block error rate, or latency), the deployment of the cell-free networks require the following actions:

- Collection of the Channel State Information (CSI) for each O-RU–user association.

- Based on the collected CSI:
 - In case of the NCC: determine the user category (i.e. local user or cluster-edge user). The initial category of the user was determined during the random access process, but the user category may change, e.g. as a result of user movement or system load.
 - In case of the UCC: selection of the O-RUs serving each user.

Consequently, the cooperating O-DUs may change or even the O-DU serving the user may change.

- Upon selection of the above-mentioned O-RUs serving the user, selection of the precoder for each O-RU–user association.
- Frequent update of the user category (in case of NCC) or serving O-RUs (in case of UCC) as well as precoders, to follow changing radio environment and system load.

In the O-RAN architecture, the above actions are controlled by the Near RT-RIC node, which is connected to neighboring O-DUs by the E2 interface, see Fig. 3. The number of O-DUs, which can be connected to one Near-RT RIC, depends on the E2 interface link capacity and the E2 interface signaling load.

The controlling algorithms located in the Near-RT RIC are referred to as the *xApps*. In order to achieve system coordination and optimum performance, a *Cell-free xApp* will collect the CSI from all users and, with the use of its intelligent algorithms, select the O-RUs serving each user together with the precoder for each O-RU. The precoders are next communicated over the E2 interface to the O-DU serving the user.

The MAC protocol in the O-DU uses the precoders for real-time scheduling, which is done every Transmission Time Interval (TTI) (e.g. 0.5 ms for 30 kHz subcarrier spacing) until a new precoder is signaled by the Near-RT RIC. The Near-RT RIC to O-DU control loop is specified by the O-RAN standard [13] from 10 ms to 1 s. Thus, the decision of what precoder to use is taken centrally by the Near-RT RIC and may be updated frequently enough to follow user movement or load change, but the actual precoding process is the physical layer functionality located either in O-DU (in case of O-RUs Category A) or O-RU (in case of O-RU Category B).

F. Cell-Free Enabling Modifications of O-RAN Specifications

According to the 3GPP standard Release 16 [11], the DUs are not interconnected (there is no interface between DUs). The O-RAN standard does not introduce the additional O-DU to O-DU signaling either. Thus, the O-DUs coordination necessary for Option 2 and 4, are not supported by the current version of the O-RAN standard and must be introduced as the proprietary O-DU vendor solution.

The Near-RT RIC uses E2 interface to control *E2 nodes* (e.g., O-DU or O-CU). This control is done with the use of so-called *RAN Functions*. In the current O-RAN specification release, the MAC scheduler is not standardized (in other words, the corresponding RAN Function is not specified). Consequently, the E2 interface must be modified by the introduction of a proprietary RAN Function.

In Option 1, to support uplink transmission, the combining of the uplink signals is done at the O-RUs and next the local O-RU estimates are averaged at the O-DU. The O-DU averaging of the local estimates received from O-RUs is a non standard process. To support the O-DU averaging internal modifications in the O-DU software are necessary.

Considering the above limitations, the current versions of the 3GPP and O-RAN standards do not allow for a straightforward deployment of the cell-free networks:

- Addition of a new E2 interface RAN Function can be seen as relatively uncomplicated tasks. The O-RAN E2AP protocol allows for the addition of a new RAN function, but such a new RAN Function would have to be supported by both the Near-RT RIC and the O-DU.
- Support of the local estimates averaging, which is necessary for Option 1,2, and 4, is an internal operation of the O-DU and can be implemented without modification of any interface.
- However, the inter-O-DU coordination, which is necessary for Option 2 and 4, is a challenging task because it requires the development of a new interface.
- Also the strategy of the precoders delivery from the Near-RT RIC to the O-DUs needs to be thought out.

IV. EXPERIMENTAL PERFORMANCE ESTIMATION OF THE O-RAN CELL-FREE DEPLOYMENT OPTIONS

In this section, we quantify the performance of various deployment options discussed above. We base the analysis on CSI obtained in a dense CF mMIMO deployment [14]. The dataset is obtained with the KU Leuven testbed deployed in an indoor environment. The central frequency is 2.6 GHz, system bandwidth is 20MHz.

Fig. 1 represents the scenario: we deployed 8 O-RUs with 8 antennas each (64 antennas in total). We assume two O-DUs each of which serves 4 O-RUs so that we have two network-centric clusters. This CF mMIMO network is used to serve 10 UEs (both local and cluster-edge ones). Note that for sake of results tractability, we apply a simple clustering algorithm: local users served by 4 O-RUs (their network-centric cluster), whereas cluster-edge users served by 8 O-RUs. More advanced clustering algorithms can be applied and analyzed in the future.

To illustrate the performance of different levels of cooperation between DUs, we investigate the five deployment options described in the previous section. Note that we use Local Minimum Mean Square Error (LMMSE) method to serve users in Options 1 and 2. Respectively, Centralized Minimum Mean Square Error (CMMSE) is used in Options 3-5. Refer [7] for more details on these two methods. For benchmarking, we consider Option 5, where all O-RUs from the two clusters collaborate in CMMSE fashion to serve all UEs. This option offers optimal performance but it is computationally expensive and it imposes a lot of fronthaul signaling on the whole system.

In Fig. 5, the five possible options are compared in terms of their SE. The horizontal line, "5% outage", shows the smallest rate among 95% of the best users [15]. More detailed comparison can be found in Table I. In the last column of

Fig. 5. Cumulative Distribution Function of spectral efficiency

this table, the 5% outage improvement of different options in comparison with option 1 is shown.

TABLE I
PERFORMANCE OF O-RAN CF mMIMO NETWORKS DEPENDING ON THE ARCHITECTURE

Option	Precoding/ combining approach	Inter-DU cooperation	5% outage improvement
Option 1	LMMSE	No	-
Option 2	LMMSE	Cluster-edge users	1.8-fold
Option 3	CMMSE	No	2.5-fold
Option 4	CMMSE	Cluster-edge users	3.8-fold
Option 5	CMMSE	All users	7.6-fold

For a small scenario with 10 UEs, we already see a very large benefit of central processing and cooperation. Central processing, or Option 3 versus Option 1, already gives a benefit of factor 2.5 in terms of 5% outage. On the other hand, an improvement can be achieved by using cooperating O-DUs. By combining scalable central processing and O-DUs cooperation, we can further enhance the performance (Option 4).

V. CONCLUSIONS

This paper has introduced an alignment of CF mMIMO and O-RAN terminology. Moreover, it explained multiple options to implement CF mMIMO in current or possible future O-RAN generations. The limitations of the current O-RAN architecture are discussed, and potential performance gains when adding more coordination are quantified based on an experimental channel database.

The SE of O-RAN compliant CF mMIMO systems can be further improved by using the UCC paradigm (the price to pay is more intensive Inter-DU signaling load). Beyond SE, the current simulation results should be extended to consider also implementation cost, energy consumption, processing latency, and reliability. More distributed processing can significantly

delay the decision and hence add latency in the physical layer. Optimal near-RT RIC strategies should allocate UE to O-RU and O-DU not only based on SE but considering the complete picture, which can be different for diverse UE.

Future work should focus on extending the current spatial scheduling work to joint time/frequency and spatial domain scheduling strategies. The impact of the allocation of O-RU to UE for different scheduling strategies should be investigated. Also, this paper gave some very rough approaches to the initial access and procedures to decide on local or edge UE categories. A more thorough investigation is needed to come to solid and practical solutions for practical cell-free systems.

Finally, the presented work should be extended to multi-band approaches: the UE–O-RU association might be frequency-dependent. Multiple O-DU coordination approaches, with different levels of coordination for different frequency bands, also require more study. The degrees of freedom for future, open, disaggregated RAN are immense. Intelligent and data-driven methods will be needed to tune all parameters and achieve the best possible configuration of the network resources at each time, for each UE and each different application.

REFERENCES

- [1] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, pp. 20–29, 2019.
- [2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.
- [3] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [4] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-Free Massive MIMO: A New Next-Generation Paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.
- [5] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink Spectral and Energy Efficiency of Cell-Free Massive MIMO With Optimal Uniform Quantization," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 223–245, 2021.
- [6] D. Wang, C. Zhang, Y. Du, J. Zhao, M. Jiang, and X. You, "Implementation of a Cloud-Based Cell-Free Distributed Massive MIMO System," *IEEE Communications Magazine*, vol. 58, no. 8, pp. 61–67, 2020.
- [7] E. Björnson and L. Sanguinetti, "Scalable Cell-Free Massive MIMO Systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.
- [8] O-RAN Alliance, "Control, user and synchronization plane specification," Technical Specification (TS) O-RAN.WG4.CUS.0-v04.00, November 2020.
- [9] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability Aspects of Cell-Free Massive MIMO," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.
- [10] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1910–1921, 2009.
- [11] 3GPP, "NG-RAN; Architecture description," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.401, January 2020, version 16.0.0.
- [12] 3gpp, "NR; Medium Access Control (MAC) protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.321, April 2020, version 16.0.0.
- [13] O-RAN Alliance, "O-ran architecture description," Technical Specification (TS) O-RAN.WG1.O-RAN-Architecture-Description-v04.00, March 2021.
- [14] S. De Bast and S. Pollin, "Ultra Dense Indoor MaMIMO CSI Dataset," 2021. [Online]. Available: <https://dx.doi.org/10.21227/nr6k-8r78>
- [15] E. Nayeibi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 203–207.