Rethinking Fronthaul Topologies
for Cell-Free 6G Networks

Max Franke*, Arash Pourdamghani*, Fabian Göttsch*†, Stefan Schmid*and Giuseppe Caire*†

*TU Berlin, †Massive Beams

Abstract—Due to significant progress in physical layer (PHY) technologies, future 6G networks are expected to feature much denser antenna deployments. Cell-free MIMO network designs are one of the most promising candidates to enable denser networks. While significant effort has been put into its PHY research, the challenges it brings to networking disciplines remain relatively unexplored. In this work, we propose various fronthaul network designs for cell-free MIMO. Our results show that while tree topologies may suffice for small-scale deployments, they become infeasible as the number of antennas increases. In contrast, with growing network size, the proposed Clos topology performs almost as well as the optimal topology.

Index Terms—Fronthaul, Cell-Free MIMO, Topologies, Optimization
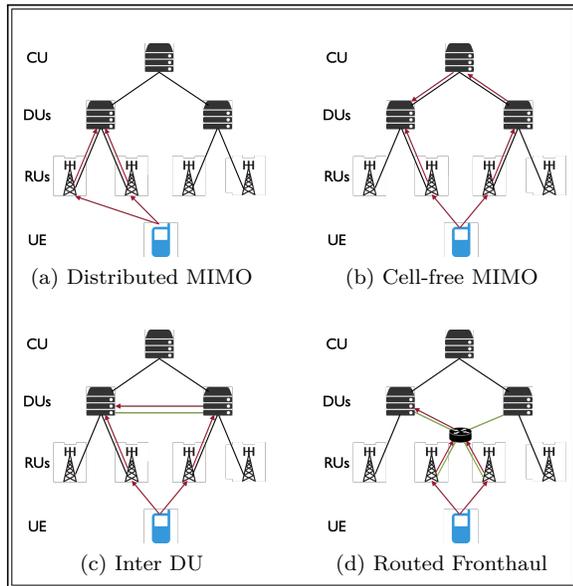
I. Introduction

Fig. 1. This figure depicts different radio access and fronthaul network designs.

The wireless communications community has long searched for new methods to tackle inter-cell interference. One approach is to distribute the antennas of a BS over the cell area, a so-called distributed MIMO system. A suitable architecture for distributed MIMO is proposed by the O-RAN alliance, where the monolithic BS is separated into antenna sites, i.e., radio units (RUs), and a distributed unit (DU). The RUs handle the radio frequency (RF) processing

and are typically located close to the antennas, while the DU acts as processing node responsible for network resource allocation and radio link control. In a distributed MIMO system with network-centric clusters, a UE can be connected to all RUs associated with the DU in a given cell, allowing for increased macrodiversity. Hence, even if the UE is located close to the cell-edge, it will maintain good channel gains with some of the RUs, such that the quality of service inside the cell is distributed more uniformly.

In an ideal *user-centric* cell-free network, each UE is connected to all surrounding RUs with good channel characteristics. Signals from other RUs not serving the UE are then typically weak and do not cause significant interference. This user-centric formation of RU clusters is the main difference and advantage compared to cell-based systems. The advantages of user-centric cell-free MIMO compared to cellular MIMO and small cell networks have been pointed out in several works [1], [2].

While user-centric cell-free MIMO obtains impressive results in the physical layer, it also introduces new challenges to the higher layers in terms of architecture and protocol design. In current deployments, DUs are only indirectly connected through centralized units (CUs), which sit one layer above the DUs. In today's O-RAN deployments, DUs communicate only via a centralized CU, which limits scalability (Fig. 1b).

To lower fronthaul load, we envision some local and cluster-level processing, meaning that some functions, such as channel estimation, are carried out at the RUs directly. Additionally, we propose a routed fronthaul network between RUs and DUs (as shown in Figs. 1c and 1d), avoiding CU bottlenecks.

A. Related Work

While the physical layer of cell-free networks has been extensively investigated in recent years (see [1] and references therein), the fronthaul is still very little studied. The vast majority of earlier works consider a centralized-RAN-type network, where the cluster-level processing of all UEs is done at a centralized unit [3], [4], which results in a non-scalable network. Only recently, a few works have been published that study the cluster processor placement at one of multiple DUs [5]–[7]. In [7], the objective of the proposed problem is to jointly minimize the fronthaul load and maximize the physical layer data rate of all UEs by optimizing the UE-RU associations. A routed fronthaul

network is considered in [5]–[7], where routing is done via dedicated routers placed in the fronthaul network [5] or routers co-located at DUs, utilizing inter-DU links [6], [7].

From an algorithmic point of view, our contribution lies in the intersection of demand-aware network design [8] and traffic engineering [9]. In the context of network design, moving beyond simple topologies such as trees, while considering the potential costs of more complex configurations, has been examined in various settings, including data center networks [10], [11]. One of the most well-known examples is Clos topology [12] used in large-scale datacenter networks [13], [14] and adopted in this work. From traffic engineering point of view, prior work mostly either focused solely on unicast [15], [16] or multicast [17]. However, since DL traffic utilizes multicast, our model in this paper needs to support both unicast and multicast in the same instance.

B. Our Contributions

In this paper, we propose different fronthaul topologies and study their impact on fronthaul performance in terms of link load. To this end, we develop a realistic model for traffic demands and explore load optimization opportunities. Following this, we explore and contribute efficient algorithms for processing node assignment, highlighting the effect of the interplay between node assignment and topology. We then empirically evaluate our optimizations in different scenarios. The results give an insight into the optimal design of the fronthaul network architecture and show the impact of the processing node assignment algorithm on different metrics, such as maximum link traffic. While previous work has investigated different fronthaul topology desgins [18], this was done mostly from a technoeconomic standpoint.

Our main findings are the following.

1) Tree topologies, the current state-of-the-art solution, become unfeasible as the network size increases. In contrast, the performance of the proposed Clos topology approaches that of the optimal topology for large networks that we created using an MILP. Interestingly, the Clos topology results in maximum link loads of around 200 Gbps only, meaning off-the-shelf hardware can handle these demands.
2) Fronthaul traffic demands are predicted to increase further with the introduction of higher frequency bands, such as mmWave and sub-THz. As tree topologies are already reaching the limit of available link capacities, smarter fronthaul network designs, like the Clos topology, should be investigated further.
3) The choice of algorithm that assigns UEs to DUs can play a significant role, offering fronthaul traffic reductions of up to 20%.

II. Modeling Fronthaul Traffic

We consider a user-centric cell-free MIMO network consisting of the radio access network (RAN), i.e., the
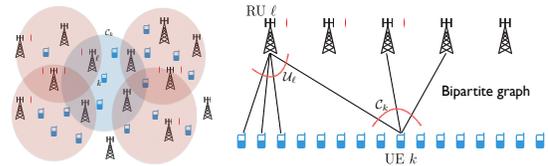

Fig. 2. An example of user-centric clusters and UE-RU association. The sets $\mathcal{U}_\ell$ of UEs associated to each RU ℓ and the sets $\mathcal{C}_k$ of RU serving UE k define a bipartite graph with RU and UE vertices and edges (k, ℓ) for all k and ℓ such that $k \in \mathcal{U}_\ell$ and $\ell \in \mathcal{C}_k$.

connections between the UEs and RUs, and a routed fronthaul network to enable the cluster-level signal processing at the DUs. The network consists of K single-antenna UEs, L RUs, Q routers, and N DUs. Each RU is equipped with M antennas, and the routers are located between RUs and DUs in the fronthaul network. The sets of UEs, RUs, routers, and DUs are denoted by $\mathcal{K} = \{1, \ldots, K\}$, $\mathcal{L} = \{1, \ldots, L\}$, $\mathcal{Q} = \{1, \ldots, Q\}$, and $\mathcal{N} = \{1, \ldots, N\}$, respectively. We follow closely the RAN model (and its implications for the fronthaul traffic) from [5], which will be briefly summarized in the following.

The wireless channel between RU ℓ and UE k is given by $\mathbf{h}_{\ell,k} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\ell,k})$, i.e., a correlated complex circularly symmetric Gaussian vector with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\ell,k} = \mathbb{E}[\mathbf{h}_{\ell,k}\mathbf{h}_{\ell,k}^{\mathsf{H}}]$, where $\boldsymbol{\Sigma}_{\ell,k}$ describes the spatial characteristics of the wireless channel. The corresponding large-scale fading coefficient (LSFC) is $\beta_{\ell,k} \triangleq \frac{1}{M}\mathrm{tr}(\boldsymbol{\Sigma}_{\ell,k})$ and captures the average signal attenuation between RU ℓ and user k due to distance and other macroscopic effects. The LSFCs are generated according to the 3GPP urban microcell street canyon pathloss model from [19, Table 7.4.1-1]. As shown in Fig. 2, each UE k is associated to a user-centric cluster of surrounding RUs $\mathcal{C}_k \subseteq \mathcal{L}$ (and each RU ℓ serves a set of UEs $\mathcal{U}_\ell \subseteq \mathcal{K}$) and an UL pilot for channel estimation based on $\{\boldsymbol{\Sigma}_{\ell,k}, \beta_{\ell,k}\}$. This is done according to the cluster formation scheme in [5].

Regarding the fronthaul, note that the user-centric cluster of RUs of each UE is associated to one of the DUs, which is responsible for the cluster-level processing. The DU is thus the *cluster processor* of the user-centric cluster. In order to enable cluster-level processing, data must be exchanged between the RUs serving a given user and the corresponding cluster processor.

We use a standard block-fading model (e.g., see [2], [20], [21] and references therein), where the UL and DL channel coefficients remain constant (channel reciprocity) over blocks of T signal dimensions (time-frequency channel uses). We will refer to one block of T signal dimensions with constant channel coefficients as "coherence block" in the following. One coherence block consists of τ_p pilot symbols for channel estimation and $T - \tau_p$ payload data symbols. The channel estimate is obtained from the τ_p pilots using subspace-based channel estimation and used to compute the precoding and receiving vector [22]. The network operates the UL and DL in time division duplex (TDD), such that

disjoint fractions of the T signal dimensions with constant channel coefficients are used for UL and DL, respectively. We denote by $\gamma_{\mathrm{DL}} \in (0,1)$ the resource fraction allocated to the DL, and $(1 - \gamma_{\mathrm{DL}})$ is allocated to the UL.

A. Physical Layer Rates

All UEs transmit with the same average energy per symbol P^{ue} and we define the system parameter $\mathsf{SNR} \triangleq \frac{P^{\mathrm{ue}}}{N_0}$, where N_0 denotes the complex baseband noise power spectral density. Following [5], we let $\alpha_{\ell,k}$ and $\widehat{\sigma}_{\ell,k}^2$ the fronthaul quantization factor and error (to be defined later). The resulting UL Signal to Interference plus Noise Ratio (SINR) of user k is given by [5]

$$\mathsf{SINR}_k^{\mathrm{ul}} = \frac{\mathsf{SNR} \left| \sum_{\ell \in \mathcal{C}_k} \widetilde{g}_{\ell,k,k} \right|^2}{\sum_{\ell \in \mathcal{C}_k} \widetilde{d}_{\ell,k} + \mathsf{SNR} \sum_{i \neq k} \left| \sum_{\ell \in \mathcal{C}_k} \widetilde{g}_{\ell,k,i} \right|^2},$$

where $\widetilde{g}_{\ell,k,i} \triangleq w_{\ell,k}^* \alpha_{\ell,k} \mathbf{v}_{\ell,k}^{\mathsf{H}} \mathbf{h}_{\ell,i}$, $\widetilde{d}_{\ell,k} \triangleq |w_{\ell,k}|^2 \left(\alpha_{\ell,k}^2 \|\mathbf{v}_{\ell,k}\|^2 + \widehat{\sigma}_{\ell,k}^2 \right)$ and $w_{\ell,k}$ and $\mathbf{v}_{\ell,k}$ are the cluster-level combining coefficient[1] and the receiving vector of (ℓ, k), respectively. The receiving/combining vectors are computed based on a linear Minimum Mean-Square Error (LMMSE) principle as in [5]. As a performance measure of the UL physical layer, we consider the so-called *Optimistic Ergodic Rate* (OER) [23] given by

$$R_k^{\mathrm{ul}} = \mathbb{E}[\log(1 + \mathsf{SINR}_k^{\mathrm{ul}})], \tag{1}$$

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

For the downlink (DL) precoding vectors, we use the approximate UL-DL reciprocity of [22] and let $\mathbf{u}_{\ell,k} \propto w_{\ell,k}^0 \mathbf{v}_{\ell,k}$, where $w_{\ell,k}^0$ are the cluster-level combining coefficients for the case of zero quantization distortion, i.e., $\alpha_{\ell,k} = 1$ and $\widehat{\sigma}_{\ell,k}^2 = 0$. The DL SINR is then given by

$$\mathsf{SINR}_k^{\mathrm{dl}} = \frac{| \mathbb{h}_k^{\mathsf{H}} \mathbb{u}_k|^2 q_k}{\mathsf{SNR}^{-1} + \sum_{j \neq k} | \mathbb{h}_k^{\mathsf{H}} \mathbb{u}_j|^2 q_j}, \tag{2}$$

where $\mathbb{h}_k = \left[\mathbf{h}_{1,k}^{\mathsf{T}} \ldots \mathbf{h}_{L,k}^{\mathsf{T}} \right]^{\mathsf{T}}$ and $\mathbb{u}_k = \left[\mathbf{u}_{1,k}^{\mathsf{T}} \ldots \mathbf{u}_{L,k}^{\mathsf{T}} \right]^{\mathsf{T}}$. Note that $\mathbf{u}_{\ell,k} \propto w_{\ell,k}^0 \mathbf{v}_{\ell,k}$ for all $\ell \in \mathcal{C}_k$ and $\mathbf{u}_{\ell,k} = \mathbf{0}$ if $\ell \notin \mathcal{C}_k$. Further, we normalize each $\mathbb{u}_k$ to have unit norm. For simplicity, in this paper we choose uniform power allocation to all data streams, i.e., $q_k = 1$ for all k [22]. The corresponding OER for the DL is given by

$$R_k^{\mathrm{dl}} = \mathbb{E}[\log(1 + \mathsf{SINR}_k^{\mathrm{dl}})]. \tag{3}$$

B. Fronthaul Network Load

The fronthaul load in the UL for UE k between a RU $\ell \in \mathcal{C}_k$ and the DU hosting $\mathcal{C}_k$ is given by

$$B_{\ell,k} = \max \left\{ \log_2 \frac{\sigma_{\ell,k}^2}{D}, 0 \right\}, \tag{4}$$

[1] The cluster-level combining coefficients $w_{\ell,k}$ for all (ℓ, k) are chosen to maximize a nominal SINR expression of the channel given the local knowledge of the cluster processor $\mathcal{C}_k$ (see [5] for details). Further, the coefficients are such that each $\mathbb{v}_k = \left[w_{1,k} \mathbf{v}_{1,k}^{\mathsf{T}} \ldots w_{L,k} \mathbf{v}_{L,k}^{\mathsf{T}} \right]^{\mathsf{T}}$ has unit norm.
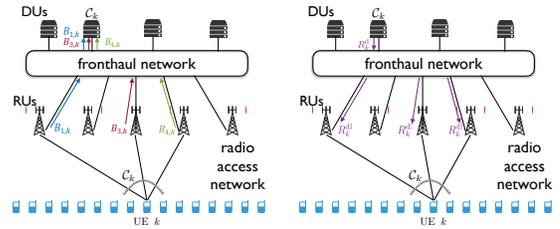

Fig. 3. Example of the fronthaul data exchange between RUs $\ell \in \mathcal{C}_k$ and the DU hosting $\mathcal{C}_k$ in the UL (left) and DL (right).

where $\sigma_{\ell,k}^2$ and D are the received signal strength after LMMSE combining at RU ℓ of UE k's UL data signal and the quantization distortion level in the fronthaul, respectively [5]. Note that these two quantities define the quantization factor $\alpha_{\ell,k} = \frac{\sigma_{\ell,k}^2 - D}{\sigma_{\ell,k}^2}$ and error $\widehat{\sigma}_{\ell,k}^2 = (1 - D/\sigma_{\ell,k}^2)D$. In the DL, all RUs in $\mathcal{C}_k$ transmit the same symbols to UE k. We assume that modulation and precoding are done at the RUs, such that the DU hosting cluster $\mathcal{C}_k$ sends the same information bits to the RUs in $\mathcal{C}_k$. The number of information bits per channel use necessary to encode the DL signal for user k at each RU $\ell \in \mathcal{C}_k$ is equal to R_k^{dl} [5]. The UL and DL fronthaul loads are depicted in Fig. 3.

III. Modeling the Network

A. Abstractions and Assumptions

As mentioned, current state of the art for fronthaul networks are passive optical networks. These do not require any routing and are thus not susceptible to any issues related to it, such as delays caused by processing or queueing. However, since routing will be necessary in cell-free networks, we are considering only routed fronthaul networks in our evaluation. This means that at the very least, there will be a router co-located with each DU. Additional routers may also be added, depending on the given topology. It should be noted that it is beyond the scope of this paper to investigate how to facilitate such a routed fronthaul.

As the processing delay is usually only in the range of a few microseconds [24], the more critical delay component added by routers is usually related to queueing. Such delay is highly dependent on metrics like congestion, buffer size etc. Emerging technologies such as optical switches can prevent queueing altogether and achieve submillisecond total switching times [25]. For these reasons, we choose to abstract delays and instead introduce a maximum hop count as a constraint to prevent unfeasibly deep networks.

Regarding link capacities, PONs only reach 50 Gbps, with the next generation likely moving up to 100 Gbps [26].

B. Inputs & Objective

We consider the input to be in the form of $\langle N, Q, \mathbf{D}_F \rangle$ in which N indicates number of DUs, Q indicates number of routers, and $\mathbf{D}_F$ is a *fronthaul demand* matrix, indicating
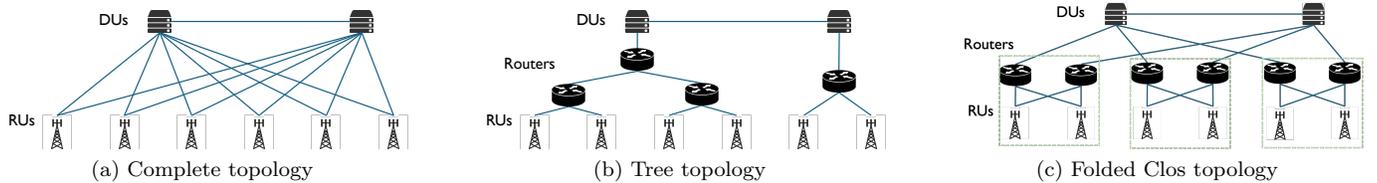
Fig. 4. A schematic view of topologies considered in this paper. In Figure 4c, the dotted green boxes show the groups containing the RUs.

fronthaul traffic (bandwidth) resulting from the physical layer communication between each associated UE-RU pair. Furthermore, we assume a set of constraints for each network element. In particular Δ^{DU}, Δ^{RU} and Δ^{R}, representing the maximum number of connections of DUs, RUs and routers respectively. Additionally, we consider β as the maximum number of UEs that a DU can handle.

Our algorithms aim to optimize for two metrics: (1) *Link utilization* (the maximum amount of bandwidth that goes through a link, in either Uplink or Downlink direction), (2) *Delay* (the maximum number of hops between UEs and their assigned DUs).

C. Topology Design

Our graph model, based on the system model, considers four types of nodes: (1) UEs (2) RUs, (3) routers and (4) DUs, as part of an underlying graph G. We consider three different topology types for the underlying graph.

Complete Topology. Here, we consider that RUs are directly connected to all DUs. We consider this type of graph as a baseline. See an example in Figure 4a.

Tree Topology. The previously discussed PON networks are known to be easily deployable. In their basic shape, they have a tree like structure. In our work, we adhere to the following steps to build a tree topology: Initially, we connect all DUs together, considering a ring topology. Next, based on the number of DUs, number of RUs and the varying maximum degrees, we determine how routers can be put in different levels. In the end, we connect the layers in a tree like topology. See an example in Figure 4b, in which $\Delta^{\mathrm{DU}} = 2$ and $\Delta^R = 3$.

Clos Topology. A *Clos Topology* is a graph in which nodes are put into three layers, and each layer is connected to the layer before and after it, in a more dense manner compared to a Tree structure (for more general description of Clos topologies, refer to [9]). In our context, layers of a Clos topology come in the form of the RUs, DUs and an intermediate layer of routers, respectively from top to bottom (at the bottom layer, RUs are connected to UEs as usual). We partition the first layer into groups of RUs with size p. Then for each group, we add $q = \frac{Q}{N}$ routers, and connect these routers to all the corresponding RUs in that group. Lastly, we connect the i-th DU to the i-th router in each group. See an example of Clos topology shown in Figure 4c.

D. Cluster placement algorithms

We consider the following heuristics for the assignment of each user-centric cluster to one of the DUs. We propose two simple (but fast) demand-oblivious and two demand-aware assignment algorithms.

Random. This algorithm assigns UE cluster to DUs uniformly at random.

Fixed. This algorithm divides the UEs equally among DUs, hoping that demand of UEs would also be distributed equally among DUs. This algorithm is designed to work well in scenarios where UEs are distributed uniformly.

Aware. This algorithm first computes the shortest path from each UE to all DUs, and sorts UEs by the sum of their uplink demands. Then, it goes over the sorted list of UEs, and selects a valid DU (that has not reached its capacity) that minimizes increase in link utilization.

Hybrid. This is an extension of the AWARE algorithm, that among the remaining valid DUs computes a *hybrid score* for each path based on both its length and current link utilization.

IV. EMPIRICAL EVALUATION

A. Setup

Input generation. We consider the number of RUs in the range $L = [20, 100]$, increasing in steps of 10, where the number of RU antennas is $M = 10$. For our first set of experiments, we have $K = 3.5L$ UEs, so that K is slightly less than $LM/2$. With user load $K \approx LM/2$, the network operates at optimal SE [27]. To study how different user loads and spatial distributions (hotspots instead of uniform placement) impact the fronthaul traffic, in a second set of experiments we consider $K = 5L$ and $K = 6.5L$. We consider 10 instances of each scenario, and report the average demands.

The UE locations are generated randomly for each instance of L, such that in each instance we have different user-centric clusters and fronthaul requirements. In particular, let us consider UE k for a given instance. In the UL, each RU $\ell \in \mathcal{C}_k$ must send $(1 - \gamma_{\mathrm{DL}})(T - \tau_p)WB_{\ell,k}$ bits through the fronthaul to the DU running the cluster processor of $\mathcal{C}_k$, where W is the physical layer channel bandwidth (i.e., the used spectrum on the PHY). In the DL, the DU must send $\gamma_{\mathrm{DL}}(T - \tau_p)WR_k^{\mathrm{dl}}$ bits to each RU $\ell \in \mathcal{C}_k$. The input is thus given by the RAN graph and the corresponding UL and DL fronthaul requirements

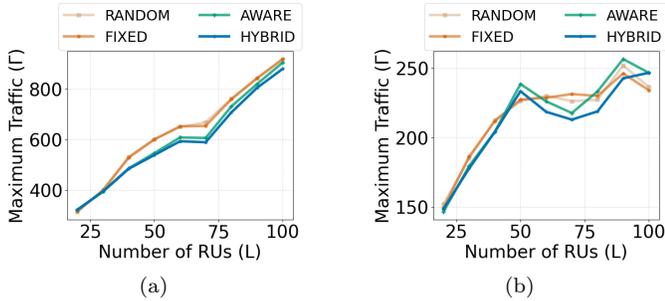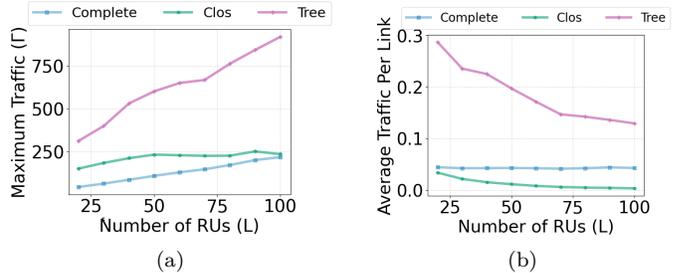Fig. 5. Maximum traffic vs. number of RUs. The underlying topology is (a) tree and (b) Clos, respectively.

Fig. 6. Maximum traffic for various topologies (a) and average maximum link traffic (b) for a varying number of RUs. The underlying algorithm in both figures is the HYBRID algorithm.

$\{(1 - \gamma_{\mathrm{DL}})(T - \tau_p)WB_{\ell,k}, \gamma_{\mathrm{DL}}(T - \tau_p)WR_k^{\mathrm{dl}}\}$ for each RU-UE association (ℓ, k), forming the demand matrix $\mathbf{D}_F$. As the UE locations are generated randomly for each instance, the inputs of each scenario are different.

We notice that the maximum channel bandwidth is $W = 100$ MHz in frequency range 1 (i.e., sub-6 GHz systems with carrier frequency $f_c = [0.41, 7.125]$ GHz) and $W = 400$ MHz in frequency range 2 (i.e., mmWave systems with $f_c = [24.25, 52.6]$ GHz) [28]. In this work, we consider a sub-6 GHz system. At a higher carrier frequency such as in mmWave systems, more spectrum is available to support wider channel bandwidths, which in turn allows more symbols to be transmitted per second. For mmWave systems, it is thus expected that the fronthaul requirements will be larger due to the wider channel bandwidth.

Simulation parameters. We set the maximum outgoing degrees of network elements $\Delta^{\mathrm{DU}} = \Delta^{\mathrm{RU}} = \Delta^{\mathrm{R}} = 10$. We then consider up to 10 DUs, and set $Q = 1.5L$. Furthermore, the maximum number of user-centric clusters that can be supported by each DU is $\beta = \frac{K}{10}$.

B. Results

We simulate our algorithms in python 3.10, benefitting from the Matplotlib [29] library. The following results report the maximum per link traffic Γ in Gbps. The possible minor irregularities are due to the realistic nature of inputs, even when averaging over several iterations. This first set of experiments used $K = 3.5L$ for all instances with a uniform user location distribution over the area.

Effect of Algorithms. First, we focus on the performance of suggested algorithms, in terms of maximum link traffic. Our simulation results show that the maximum link traffic increases linearly with an increase in the number of RUs. In particular, we observe that on average, AWARE and HYBRID algorithms outperform the RANDOM and FIXED algorithms. The difference between different algorithms can be up to 20%, as Figs. 5a and 5b show for the Tree and Clos topology, respectively.

Effect of Topology. Secondly, we compare the difference in maximum link traffic, considering various topologies. Here, our simulation results indicate that the Clos topology can significantly decrease the traffic in the fronthaul network compared to more traditional tree topologies by around 75%. We observe that the Clos topology's performance gets much closer to the ideal complete topology, as the number of RUs grows. Fig. 6a shows a comparison of different topologies with increasing number of RUs. Furthermore, we report that in terms of average link traffic, the Clos topology is the best suited among the topologies that we have considered, see Fig. 6b.

For our second set of experiments, we modified the average number of UEs per RU. In addition, we also distributed the UEs in different patterns to emulate "hotspots" in the network area where many users congregate. The results for $K = 5L$ are shown in Fig. 7a and for $K = 6.5L$ in Fig. 7b. We observe similar behavior for the different fronthaul topologies as in the scenario with uniform distribution of user locations.

C. Discussion

Our results confirm our assumptions with regard to the feasibility of tree-like topologies for the fronthaul of future cell-free RANs. With growing number of RUs, they quickly reach a point where existing optical transport technologies can no longer support them.

Keeping with the assumption made of using 100 GBit links, even for small amounts of RUs, e.g. $L = 20$, the required capacities already reach over 300 GBit, tripling the available capacities. Furthermore, they breach even the currently highest available point-to-point capacities of 800 GBit starting at around 90 RUs. Higher carrier frequencies, such as mmWave or sub-THz, would push the demands up by at least another magnitude, with some researchers predicting data rates of up to 1 TBit per RU [30]. This highlights that it is not the radio/physical layer that is posing bandwidth and capacity limitations in cell-free MIMO, but instead the fronthaul network. It is thus up to network designers and planners to come up with new architectures and approaches to overcome these challenges. We showed that fast heuristics can have an impact on the link capacity requirements. While it is not yet clear which algorithm is ideal, traffic reductions of up to 20% are observable in certain scenarios.
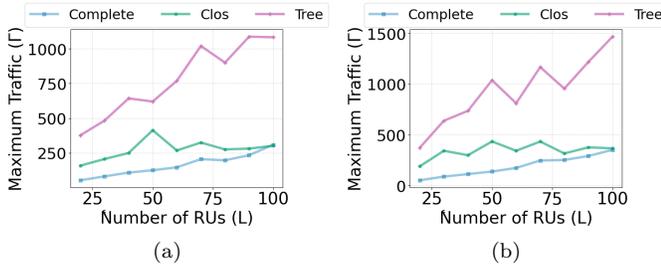
Fig. 7. Maximum traffic for a hotspot distribution with (a) $K = 5L$ and (b) $K = 6.5L$.

Finally, we aim to find topologies that improve performance while allowing for feasible deployments. Using complete graphs as a baseline, we show that Clos topologies offer almost on par performance, especially as the number of RUs grows. As shown, while overall maximum traffic approaches that of an ideal complete graph, the average link traffic actually drops below, indicating even better load balancing of traffic across different links. We also note that our results hold for various distributions of UE locations in the mobile network area.

V. Conclusion

Our findings show that cell-free 6G networks will not be able to live up to their full potential if the current state-of-the-art passive optical networks with tree topologies are used for the fronthaul. Instead, rethinking the architecture and topologies of fronthaul networks will be required to deal with the new demands and requirements imposed by the physical layer. It is up to the network community to design topologies in such a way that they offer a good trade-off between cost, complexity and efficiency. We showed another example of Clos topologies versatility here, outside of their current use cases in datacenter, enterprise networks, etc.

Acknowledgment

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] Ö. T. Demir, E. Björnson, L. Sanguinetti *et al.*, "Foundations of User-Centric Cell-Free Massive MIMO," *Found. Trends Signal Process.*, 2021.

[3] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-Free Massive MIMO with Limited Backhaul," in *IEEE ICC*, 2018.

[4] H. Masoumi and M. J. Emadi, "Performance Analysis of Cell-Free Massive MIMO System With Limited Fronthaul Capacity and Hardware Impairments," *IEEE TWC*, 2019.

[5] 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 TWC*, 2024.

[6] A. G. Joshi, N. T. Khai, and T. Bauschert, "Fronthaul Resource Optimization for Cell-Free Massive MIMO Networks," in *IEEE RNDM*, 2024.

[7] P. Lai, W. Xiang, W. D. Lukito, K. T. Phan, P. Cheng, C. Liu, and G. Mao, "Hybrid Network-and User-Centric Scalable Cell-Free Massive MIMO for Fronthaul Signaling Minimization," *IEEE Trans. Veh. Technol.*, 2024.

[8] C. Avin and S. Schmid, "Toward demand-aware networking: a theory for self-adjusting networks," *ACM CCR*, 2018.

[9] M. Chiesa, G. Kindler, and M. Schapira, "Traffic engineering with equal-cost-multipath: An algorithmic perspective," *IEEE/ACM TON*, 2017.

[10] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in *ACM CCR*, 2008.

[11] A. Pourdamghani, C. Avin, R. Sama, and S. Schmid, "Seedtree: A dynamically optimal and local self-adjusting tree," in *IEEE INFOCOM*, 2023.

[12] C. Clos, "A study of non-blocking switching networks," *Bell System Technical J.*, 1953.

[13] A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead *et al.*, "Jupiter rising: A decade of clos topologies and centralized control in google's datacenter network," in *ACM SIGCOMM*, 2015.

[14] H. Ballani, P. Costa, R. Behrendt, D. Cletheroe, I. Haller *et al.*, "Sirius: A flat datacenter network with nanosecond optical switching," in *ACM SIGCOMM*, 2020.

[15] S. Agarwal, M. S. Kodialam, and T. V. Lakshman, "Traffic engineering in software defined networks," in *IEEE INFOCOM*, 2013.

[16] M. Henzinger, A. Paz, A. Pourdamghani, and S. Schmid, "The augmentation-speed tradeoff for consistent network updates," in *ACM SOSR*, 2022.

[17] M. Rost and S. Schmid, "Virtucast: Multicast and aggregation with in-network processing: An exact single-commodity algorithm," in *OPODIS*. Springer, 2013.

[18] L. Furtado, A. Fernandes, A. Ohashi, F. Farias, A. Cavalcante, and J. Costa, "Cell-free massive mimo deployments: Fronthaul topology options and techno-economic aspects," in *IEEE EuCAP*, 2022.

[19] 3GPP, "Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 16) V16.1.0," Tech. Rep. Rep. TR 38.901, 2020.

[20] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE TWC*, 2010.

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

[22] F. Göttsch, N. Osawa, T. Ohseki, K. Yamazaki, and G. Caire, "Subspace-Based Pilot Decontamination in User-Centric Scalable Cell-Free Wireless Networks," *IEEE TWC*, 2023.

[23] G. Caire, "On the Ergodic Rate Lower Bounds with Applications to Massive MIMO," *IEEE TWC*, 2018.

[24] L. Angrisani, L. Peluso, A. Tedesco, and G. Ventre, "Measurement of processing and queuing delays introduced by a software router in a single-hop network," in *IEEE I2MTC*, 2005.

[25] M. Calvarese, P. Paiè, F. Ceccarelli *et al.*, "Strategies for improved temporal response of glass-based optical switches," *Scientific Reports*, vol. 12, p. 239, 2022.

[26] S. Cwalina, S. K. Mohammadi, K. Habel, V. Jungnickel, and R. Freund, "Demonstration of a real-time 100g ethernet space division multiplexing pon using a weakly coupled multicore fiber," in *IEEE FOAN*, 2023, pp. 32–35.

[27] F. Göttsch, N. Osawa, I. Kanno, T. Ohseki, and G. Caire, "Fairness Scheduling in User-Centric Cell-Free Massive MIMO Wireless Networks," *IEEE TWC*, 2024.

[28] 3GPP, "Base Station (BS) radio transmission and reception (Release 19) V19.0.0," Tech. Rep. Rep. TR 38.104, 2025.

[29] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Comput. Sci. Eng.*, 2007.

[30] Y. Corre, G. Gougeon, J.-B. Doré, S. Bicaïs, B. Miscopein, E. Faussurier, M. Saad, J. Palicot, and F. Bader, "Sub-thz spectrum as enabler for 6g wireless communications up to 1 tbit/s," in *6G Wireless Summit*, 2019.