Cell-free Massive MIMO Infrastructure Sharing

F. Riera-Palou*† and G. Femenias *†
*SECOM Group and †Balearic Institute for Artificial Intelligence Research (IAIB)
Universitat de les Illes Balears - 07122 Mallorca, Spain, Email: {felip.riera,guillem.femenias}@uib.es

***Abstract*—This paper introduces the idea of infrastructure sharing between two separate cell-free massive MIMO (CF-mMIMO) networks providing service over a prescribed geographical area. It is argued that given the fully distributed nature of cell-free topologies, an attractive scheme to implement infrastructure sharing consists of allowing users from either network to switch to the alternative one, provided these changes bring along a measurable improvement. This effect is termed network diversity. Algorithms to govern this user swap are introduced and their effectiveness demonstrated through simulations. Interestingly, numerical results reveal that this technique is most effective when the two co-located networks have significantly different levels of densification.**

***Index Terms*—Cell-free massive MIMO, scalability, network sharing, network diversity.**

I. Introduction

The development of the sixth generation (6G) of mobile networks is rapidly accelerating with all intervening actors, academia, industry, and standardization bodies, delineating its potential capabilities while exploring the technological pillars it will rest upon [1]. In the 6G context, a new form of massive multiple-input multiple-output (mMIMO), called CF-mMIMO stands out as the new topology that will completely change the deployment of mobile networks. Contrary to the so far pervasive cellular architecture, a CF-mMIMO network consists of a multitude of access points (APs) scattered randomly throughout the coverage area and connected to a central processing unit (CPU) that takes care of most of the baseband processing [2]. This architecture brings the RF infrastructure closer to the users and effectively mitigates, through cooperative processing, the cell-edge effects that critically harm classical cellular deployments. Remarkably, and provided that each user is served from a sufficiently large number of APs, all the benefits of the mMIMO regime also apply to CF-mMIMO.

Despite its theoretical promises, CF-mMIMO brings along also a fair share of practical issues that need to be carefully handled for an effective implementation with most of these handicaps stemming from the required ultradense infrastructure deployment [3]. In particular, the need to install a huge number of geographically separated APs, all connected to the CPU, typically through fiber optic links, unavoidably entails large network roll-out costs [4]. Aiming at reducing them, alternative network structures to the prevailing CPU-centered star topology have been investigated in [5]–[7], demonstrating that bus-based architectures whereby only a subset of APs are connected to the CPU and provide service to the rest of APs, significantly reduce costs and fronthaul requirements. These proposals are very much related to the concept of radio stripes introduced in [8] where only one of the APs is connected to the CPU while the rest are strictly serially connected. Although these architectures help in cutting down the deployment costs, this comes at the cost of potential losses in spectral efficiency and a drastic reduction of the network robustness in case of eventual disruptions in a single node or the bus itself. An alternative approach to minimize cell-free deployment costs is to adopt a transitional approach from the cellular-based architecture to the cell-free topology as proposed in [9]–[11]. In this scenario, instead of deploying a multitude of APs throughout the whole coverage area, APs are only installed in the areas where the cellular coverage does not provide sufficient capacity with these typically corresponding to the boundaries among macro cells (i.e., cell-edges). Again, this progressive cell-free deployment comes at the cost of a spectral efficiency loss with respect to a fully distributed architecture. Anticipating the advent of cell-free schemes, ultradense networking was hailed in the 2010s as one of the pillars of the then forthcoming fifth generation (5G) of mobile networks [12]. Even back then, roll-out costs were already identified as an important hurdle of this trend prompting the proposal of network infrastructure sharing among different mobile network operators (MNOs) as a cost-effective method to densify a target coverage area by pooling the hardware resources to jointly serve all users [13], [14]. Recent standardization efforts have also been devoted to explore the possibilities infrastructure sharing offers [15].

The main objective of this paper is to explore the benefits infrastructure sharing can bring when applied to a context of various cell-free based MNOs serving a common coverage area. It is envisaged that this mechanism can result in substantial cost reductions for the operators when deploying and maintaining cell-free networks. The main contributions of this paper are:

1) The proposal of an architecture that allows operators to share spectral efficiency (SE) information of the mobile stations (MSs) each MNO is serving. This information then serves as the basis of an strategy to swap users from one MNO to another with the objective of improving the overall connectivity of the coverage area. Remarkably, this technique can be configured to target arbitrary rate-based quality of service (QoS) metrics (e.g. max-min, sum-rate) although at the cost of a considerable computational complexity and compromising scalability.

2) Addressing the aforementioned shortcomings, a low computational alternative is introduced that just relies on large-scale information and offers virtually the same performance at a negligible computational expenditure. Moreover, the resulting design can be shown to satisfy scalability requirements.

Note that other infrastructure sharing options are available in cellular-based architectures like for example allowing certain network elements (e.g., a base station) to be able to serve users from different operators (see [15] for a full discussion). However, such approaches require shared elements to be connected to the infrastructure of all the intervening MNOs. In the case of CF-mMIMO, the dense infrastructure with a multitude of APs makes such an approach impractical. Remarkably, the approach pursued here can be seen as a form of network diversity where users can be offered to operate through different MNOs infrastructures and are able to connect to the best one. Also notice that the problem considered here differs from the multi-CPU scenario (see for example [16], [17]) as in the latter case, each CPU controls APs that tend to cover spatially/geographically separated areas and inter-CPU operations are conducted to allow seamless transitions among neighbouring areas. In contrast, in this work, APs from different CPUs/MNOs have their controlled APs mingled over the same coverage area.

II. System model

This paper considers a geographical area where two distinct MNOs, tagged as MNO_i with $i \in \{1, 2\}$, co-exist to provide service to their respective users[1]. Denote by $\mathcal{K}_i^{(0)}$ the set of users served by MNO_i in the absence of any infrastructure sharing strategy and let $\mathcal{K} = \mathcal{K}_1^{(0)} \cup \mathcal{K}_2^{(0)}$ denote the set of all MSs in the coverage area. MNO_i operates a CF-mMIMO network consisting of M_i APs each equipped with N_i antennas and connected to a MNO-specific CPU through ideal fronthaul links. All users are assumed to be single-antenna. The i-th MNO operates on a bandwidth B_i at carrier frequency f_{ci}. The operational carriers of the two MNOs are assumed close enough so that their propagation conditions are similar (e.g., both networks operate at sub-6 GHz). Both CPUs are connected to a master operator (MOP) agreed by both MNOs that serves to coordinate the functions related to infrastructure sharing as described in [15]. This setup is shown diagrammatically in Fig. 1. At any given instant, $\mathcal{K}_i$ will denote the user set served by the i-th MNO, with $K_i = |\mathcal{K}_i|$, under the assumption that users may have been transferred from one MNO to another. Users in $\mathcal{K}_i^{(0)}$ will consider APs in $\mathcal{M}_i$ as *home* APs whereas those in $\mathcal{M}_j$ with $i \neq j$ will be deemed as *foreign* APs. Extensive networks may require of multi-CPU deployments, which nonetheless can be deployed in a scalable manner [16], [18].

[1]The scenario where more than two MNOs are active on a given coverage area is currently under investigation.

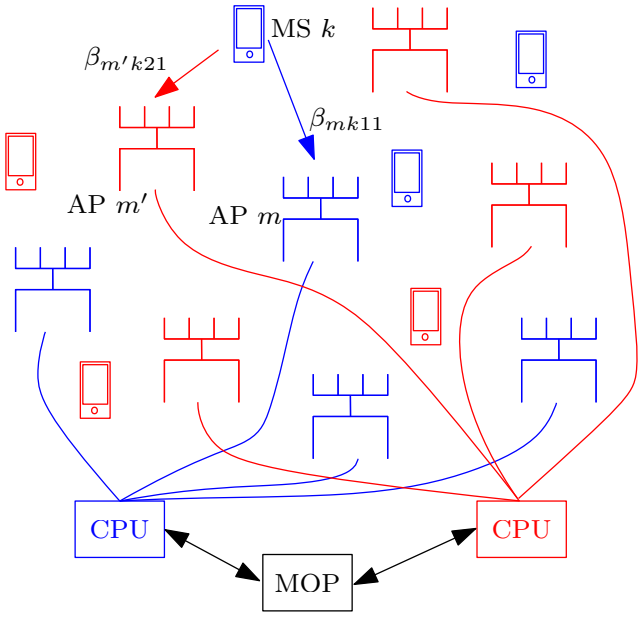

Fig. 1: Infrastructure sharing scenario between two CF-M-MIMO operators (operator 1 in blue, operator 2 in red). As an example, user $k \in \mathcal{K}_1$ can connect to its own network, MNO_1, or it can eventually switch to MNO_2.

A. Propagation environment

The propagation conditions encountered are defined in accordance to [19] for the specific case of outdoor users when operating at sub-6 GHz frequencies. In particular, our interest is in the urban microchannel (UMi) channel profile since this profile corresponds to densely populated urban environments, which are the most usual targets for CF-mMIMO deployments. Let us denote by d_{imk}^{2D} and d_{imk}^{3D} the 2D- and 3D-distances between user $k \in \mathcal{K}$ and AP $m \in \mathcal{M}_i$ with $i \in \{1, 2\}$.

It is now possible to characterize the probability of line-of-sight (LOS) propagation between user $k \in \mathcal{K}$ and AP $m \in \mathcal{M}_i$ as [19]

$$p_{imk}^{\text{LOS}} = \begin{cases} 1, & d_{imk}^{\text{2D}} \leq 10 \\ \exp\left(-\frac{d_{imk}^{\text{2D}} - 10}{1000}\right), & 10 < d_{imk}^{\text{2D}}. \end{cases} \tag{1}$$

The corresponding path loss in such environment is defined according to the link's LOS/non-line-of-sight (NLOS) character as

$$\text{PL}_{imk}^{\text{LOS}}[\text{dB}] = \begin{cases} 32.4 + 21\log_{10}(d_{imk}^{\text{3D}}) \\ \quad + 20\log_{10}(f_{ci}[\text{GHz}]) & , \text{if } d_{imk}^{\text{2D}} \leq d'_{BP} \\ 32.4 + 40\log_{10}(d_{imk}^{\text{3D}}) \\ \quad + 20\log_{10}(f_{ci}[\text{GHz}]) \\ \quad - \Delta'_{BP} & , \text{if } d_{imk}^{\text{2D}} > d'_{BP} \end{cases} \tag{2}$$

and

$$\text{PL}_{imk}^{\text{NLOS}}[\text{dB}] = \max\{\text{PL}_{imk}^{\text{LOS}}, \hat{\text{PL}}_{imk}^{\text{NLOS}}\}, \tag{3}$$

with

$$\hat{\text{PL}}_{imk}^{\text{NLOS}}[\text{dB}] = 35.3\log_{10}(d_{imk}^{\text{3D}}) + 22.4 + 21.3\log_{10}(f_{ci}[\text{GHz}]) - 0.3(h_{\text{MS}} - 1.5) \quad (4)$$

and where h_{MS} and h_{AP} denote the heights of the MSs and APs, respectively. Parameter $d'_{BP} = 4(h_{\text{MS}} - 1)(h_{\text{AP}} - 1)f_{ci}/(3 \times 10^8)$ defines a breakpoint distance and $\Delta'_{BP} = 9.5\log_{10}(d'^2_{BP} + (h_{\text{AP}} - h_{\text{MS}})^2)$. The other major large-scale effect, shadow fading, is modeled as a spatially correlated Normal random variable $\chi_{imk}[\text{dB}] \sim \mathcal{N}(0, \sigma_\chi^2)$ that in a UMi environment follows [19]

$$\sigma_\chi^2 = \begin{cases} 4, & \text{for LOS links} \\ 7.82, & \text{for NLOS links,} \end{cases} \quad (5)$$

with spatial autocorrelation $R_\chi(\Delta x) = \exp(|\Delta x|/d_{corr})$ with d_{corr} denoting the decorrelation distance (see [2, (54)-(55)]).

Taking into account all the aforementioned effects, the large-scale propagation gain between MS k in $\mathcal{K}$ and AP m in $\mathcal{M}_i$ is given by

$$\beta_{imk}[\text{dB}] = -\text{PL}_{imk}[\text{dB}] + \chi_{imk}[\text{dB}], \quad (6)$$

where $\text{PL}_{imk}[\text{dB}]$ will take either the form of (2) or (3) depending on the LOS/NLOS character of the link. As it is commonly assumed in the CF-mMIMO literature, large-scale propagation parameters are perfectly known and available wherever needed in the network. Furthermore, it is important to stress that APs are able to gather large-scale propagation information from all the MSs in the coverage area independently of at which network they are connected to. That is, AP m in $\mathcal{M}_i$ is able to perfectly estimate β_{imk} irrespective of whether $k \in \mathcal{K}_1$ or $\mathcal{K}_2$, and forward this information to the CPU. Toward this end, either MSs can transmit a sounding reference signal (SRS) on the foreign operator band or, alternatively, these foreign APs can *listen* to the MSs transmitting the SRS to their home APs.

B. Time division duplex (TDD) frame format

CF-mMIMO relies on a time-division duplex (TDD) transmission protocol whereby frames are divided into an uplink (UL) training phase (of size τ_p samples), an UL payload data transmission phase (of size τ_u samples), and a downlink (DL) payload data transmission phase (of size τ_d samples). The transmission frame length is then given by $\tau_f = \tau_p + \tau_u + \tau_d$. It must be ensured that the frame length τ_f does not exceed the environment-dependent coherence block $\tau_c = B_c T_c$ with B_c and T_c denoting the frequency and time coherence sizes, respectively. Since the two considered CF-mMIMO networks target the same propagation environment, it is safe to assume they endure similar propagation conditions (i.e., comparable B_c, T_c) and thus both networks can fix a common frame length τ_f. In both networks, the effective data transmission time is equally split between UL and DL, that is, $\tau_u = \tau_d$. For conciseness, this paper focuses on the UL side notwithstanding that many of the issues discussed here equally apply to the DL.

C. Channel model and estimation

This paper considers a spatially correlated Ricean fading channel model with random phase shifts [20]. In this context, the vector describing the channel between MS $k \in \mathcal{K}$ and AP $m \in \mathcal{M}_i$ follows

$$\boldsymbol{g}_{imk} = \boldsymbol{g}_{imk}^{\text{LOS}} + \boldsymbol{g}_{imk}^{\text{NLOS}}, \quad (7)$$

where $\boldsymbol{g}_{imk}^{\text{LOS}}$ and $\boldsymbol{g}_{imk}^{\text{NLOS}}$ denote the LOS and NLOS components of the channel, respectively. The LOS and NLOS components can, in turn, be expressed as

$$\boldsymbol{g}_{imk}^{\text{LOS}} = \sqrt{\frac{\beta_{imk}K_{imk}}{K_{imk}+1}} e^{j\zeta_{imk}} \boldsymbol{a}_{imk}^{\varphi_{imk},\vartheta_{imk}}, \quad (8)$$

and

$$\boldsymbol{g}_{imk}^{\text{NLOS}} = \sqrt{\frac{\beta_{imk}}{K_{imk}+1}} \boldsymbol{F}_{imk}^{1/2} \boldsymbol{q}_{imk}, \quad (9)$$

where K_{imk} is the so-called Ricean K-factor holding $10\log_{10} K_{imk} \sim \mathcal{N}(\mu_K, \sigma_K^2)$, $\zeta_{imk} \sim \mathcal{U}(-\pi, \pi)$ is the random phase of the LOS component, $\boldsymbol{a}_{imk}^{\varphi_{imk},\vartheta_{imk}}$ represents the array response vector at AP $m \in \mathcal{M}_i$ with respect to MS k at azimuth and elevation angles φ_{imk} and ϑ_{imk}, respectively, and holding $\left\|\boldsymbol{a}_{imk}^{\varphi_{imk},\vartheta_{imk}}\right\|^2 = N$, $\boldsymbol{F}_{imk}$ represents the positive semi-definite spatial correlation matrix at AP $m \in \mathcal{M}_i$ as seen from MS k, holding that $\text{tr}(\boldsymbol{F}_{imk}) = N$, and finally, $\boldsymbol{q}_{imk} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$. The spatial correlation matrix characterizing the channel vector $\boldsymbol{g}_{imk}$ can be straightforwardly obtained as

$$\boldsymbol{R}_{imk} = \frac{\beta_{imk}K_{imk}}{K_{imk}+1} \boldsymbol{a}_{imk}^{\varphi_{imk},\vartheta_{imk}} (\boldsymbol{a}_{imk}^{\varphi_{imk},\vartheta_{imk}})^H + \frac{\beta_{imk}}{K_{imk}+1} \boldsymbol{F}_{imk}, \quad (10)$$

thus holding that $\text{tr}(\boldsymbol{R}_{imk}) = N\beta_{imk}$.

A set of τ_p mutually orthogonal pilot sequences are used during the UL training phase. These pilot sequences are allocated to MSs in a deterministic way, and whenever $K > \tau_p$ the so-called pilot contamination effect arises. The $N \times \tau_p$ matrix of received samples at AP m managed by MNO i during the training phase can be written as

$$\boldsymbol{Y}_{im} = \sum_{k \in \mathcal{K}_i} \sqrt{P_p \tau_p} \boldsymbol{g}_{imk} \boldsymbol{\phi}_{ik}^T + \boldsymbol{n}_{imk}, \quad (11)$$

where $\boldsymbol{\phi}_{ik} \in \mathbb{C}^{\tau_p \times 1}$ is the pilot assigned to user $k \in \mathcal{K}_i$, P_p is the transmit per-pilot symbol at the MSs and $\boldsymbol{n}_{imk} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_v^2 \boldsymbol{I}_N)$. Owing to the orthogonality of the pilot sequences, projecting $\boldsymbol{Y}_{mi}$ onto a specific pilot $\boldsymbol{\phi}_{ik}$ constitutes a sufficient statistic towards the estimation of $\boldsymbol{g}_{imk}$. Hence, denoting by $\mathcal{P}_{ik}$ the set of MSs in $\mathcal{K}_i$ using the same pilot sequence as MS k (including itself), let us define

$$\boldsymbol{y}_{imk} = \boldsymbol{Y}_{im}\boldsymbol{\phi}_{ik}^* = \sqrt{P_p\tau_p}\boldsymbol{g}_{imk} + \sqrt{P_p\tau_p} \sum_{k' \in \mathcal{P}_{ik} \setminus k \in \mathcal{K}_i} \boldsymbol{g}_{imk'} + \boldsymbol{n}_{imk}\boldsymbol{\phi}_{ik}^*. \quad (12)$$

The minimum mean square error (MMSE) channel estimate of $\boldsymbol{g}_{imk}$ can be obtained as

$$\hat{\boldsymbol{g}}_{imk} = \sqrt{P_p \tau_p} \boldsymbol{R}_{imk} \boldsymbol{\Psi}_{imk}^{-1} \boldsymbol{y}_{imk}, \tag{13}$$

where

$$\boldsymbol{\Psi}_{imk} = \mathbb{E}\left\{\boldsymbol{y}_{imk} \boldsymbol{y}_{imk}^H\right\} = P_p \tau_p \sum_{k' \in \mathcal{P}_{ik}} \boldsymbol{R}_{imk'} + \sigma_i^2 \boldsymbol{I}_N. \tag{14}$$

The channel estimation error $\tilde{\boldsymbol{g}}_{imk} = \boldsymbol{g}_{imk} - \hat{\boldsymbol{g}}_{imk}$ is independent of the channel estimate and distributed as $\tilde{\boldsymbol{g}}_{imk} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_{imk})$, where

$$\boldsymbol{C}_{imk} = \boldsymbol{R}_{imk} - P_p \tau_p \boldsymbol{R}_{imk} \boldsymbol{\Psi}_{imk}^{-1} \boldsymbol{R}_{imk}. \tag{15}$$

For later convenience, we define at this point the compound user channel $\boldsymbol{g}_{ik} = [\boldsymbol{g}_{i1k}^T \dots \boldsymbol{g}_{iM_ik}^T]^T$, compound user channel estimate $\hat{\boldsymbol{g}}_{ik} = [\hat{\boldsymbol{g}}_{i1k}^T \dots \hat{\boldsymbol{g}}_{iM_ik}^T]^T$, compound user channel covariance matrix $\boldsymbol{R}_{ik} = \text{diag}(\boldsymbol{R}_{i1k}, \dots, \boldsymbol{R}_{iM_ik})$, and compound estimation error user covariance matrix $\boldsymbol{C}_{ik} = \text{diag}(\boldsymbol{C}_{i1k}, \dots, \boldsymbol{C}_{iM_ik})$.

D. Combiner design and power allocation

During the UL payload transmission phase, all MSs served by MNO i transmit their respective information signals resulting in the $N \times 1$ reception vector at AP $m \in \mathcal{M}_i$

$$\boldsymbol{r}_{im} = \sum_{k \in \mathcal{K}_i} \boldsymbol{g}_{imk} \sqrt{p_{ik}} s_{ik} + \boldsymbol{v}_{im}, \tag{16}$$

where p_{ik} and s_{ik}, with $E\{|s_{ik}|^2\} = 1$, are the UL transmission power and information symbol for user k in $\mathcal{K}_i$, respectively, and $\boldsymbol{v}_{im} \sim \mathcal{CN}(0, \sigma_v^2 \boldsymbol{I}_N)$. It is well known that centralized processing, whereby the samples received at every AP are forwarded to the CPU for joint MMSE processing, leads to best performance [21]. Importantly, and in order to keep the system architecture scalable, dynamic cooperation cluster (DCC) [22] is applied to determine which APs serve each MS and also to perform the pilot allocation to users.

Assuming the use of a centralized combiner, the SE of MS $k \in \mathcal{K}_i$ takes the form of an achievable rate (measured in bit/s/Hz) as [22, Section 5.1]

$$\eta_{ik} = \frac{1}{2}\left(1 - \frac{\tau_u}{\tau_f}\right) E\left\{\log_2\left(1 + \text{SINR}_{ik}\right)\right\}, \tag{17}$$

where the expectation is taken with respect to the compound channel and the instantaneous effective signal-to-interference-plus-noise ratio (SINR) is given by

$$\text{SINR}_{ik} = \frac{p_{ik} \left|\boldsymbol{v}_{ik}^H \boldsymbol{D}_{ik} \hat{\boldsymbol{g}}_{ik}\right|^2}{\sum\limits_{k' \in \mathcal{K}_i, k' \neq k} p_{ik'} \left|\boldsymbol{v}_{ik}^H \boldsymbol{D}_{ik'} \hat{\boldsymbol{g}}_{ik'}\right|^2 + \boldsymbol{v}_{ik}^H \boldsymbol{Z}_{ik} \boldsymbol{v}_{ik} + \tilde{\sigma}_{ik}^2}, \tag{18}$$

where $\tilde{\sigma}_{ik}^2 = \sigma_v^2 \left\|\boldsymbol{D}_{ik} \boldsymbol{v}_{ik}\right\|^2$ and the vector $\boldsymbol{v}_{ik}$ represents the part of the centralized combiner implemented at the CPU of MNO i to detect MS k in $\mathcal{K}_i$ and

$$\boldsymbol{Z}_{ik} = \sum_{k' \in \mathcal{K}_i} p_{ik'} \boldsymbol{D}_{ik} \boldsymbol{C}_{ik'} \boldsymbol{D}_{ik}.$$

Matrices $\boldsymbol{D}_{ik} = \text{diag}(\boldsymbol{D}_{i1k}, \dots, \boldsymbol{D}_{iM_ik})$ actually implement the MS-AP connectivity dictated by the DCC by defining [22]

$$\boldsymbol{D}_{imk} = \begin{cases} \boldsymbol{I}_{N_i}, & \text{if AP } m \text{ serves MS } k \text{ on MNO } i \\ \boldsymbol{0}_{N_i}, & \text{otherwise.} \end{cases} \tag{19}$$

It is assumed here the use of the improved partial MMSE (IP-MMSE) combiner that it is known to outperform the classical partial MMSE while preserving network scalability [23]. This combiner takes the form of

$$\boldsymbol{v}_{ik} = \left(\sum_{k' \in \mathcal{S}_{ik}} p_{ik'} \boldsymbol{D}_{ik} \hat{\boldsymbol{g}}_{ik'} \hat{\boldsymbol{g}}_{ik'}^H \boldsymbol{D}_{ik} + \boldsymbol{Z}_{\mathcal{S}_{ik}} + \sigma_v^2 \boldsymbol{I}_{MN}\right)^{-1} \times p_{ik} \boldsymbol{D}_{ik} \hat{\boldsymbol{g}}_{ik}, \tag{20}$$

where $\mathcal{S}_{ik}$ denotes the set formed by user $k \in \mathcal{K}_i$ and all the MSs in $\mathcal{K}_i$ sharing at least one AP with user k, and

$$\boldsymbol{Z}_{\mathcal{S}_{ik}} = \sum_{k' \in \mathcal{S}_{ik}} p_{ik'} \boldsymbol{D}_{ik} \boldsymbol{C}_{ik'} \boldsymbol{D}_{ik} + \sum_{k' \in \mathcal{K}_i \setminus \mathcal{S}_{ik}} p_{ik'} \boldsymbol{D}_{ik} \boldsymbol{R}_{ik'} \boldsymbol{D}_{ik}.$$

As a last design step, power coefficients p_{ik} need to be fixed. Here we resort to fractional power allocation (FPA) [22] which has proved to be a simple and flexible technique to select the power allocation coefficients to approximate different metrics, most notably, max-min rate and sum-rate while remaining scalable. Under FPA, power weights for $k \in \mathcal{K}_i$ are set as

$$p_{ik} = P_{\text{MS}} \frac{\left(\sum_{l \in \mathcal{M}_{ik}} \beta_{ilk}\right)^{\upsilon}}{\max_{k' \in \mathcal{S}_{ik}} \left(\sum_{l \in \mathcal{M}_{ik'}} \beta_{ilk'}\right)^{\upsilon}}, \tag{21}$$

where P_{MS} is the maximum transmit power per MS, $\mathcal{M}_{ik}$ is the set of all APs in $\mathcal{M}_i$ serving user k and $\upsilon \in [-1, 1]$ is a parameter that controls the rate policy: positive values strive for sum-rate performance and negative values approach the max-min criterion. It is worth noting that the choice of $\upsilon = -0.5$ is most common in the CF-mMIMO literature as it is more closely aligned with the philosophy of this architecture, that is, to provide uniform QoS throughout the coverage area.

III. CF-mMIMO INFRASTRUCTURE SHARING

In trying to improve the overall performance across both networks, MNO_1 and MNO_2 are allowed to pursue the transfer of MSs from one network to another. Two approaches are introduced next, one based on the users' SE, and another one that simply relies on large-scale propagation information. Note that an important aspect of this user transfer concerns how these can be managed in a private and accountable manner using, for instance, a blockchain infrastructure. This a topic of current research but it is out of the scope of the current work.

A. SE-based infrastructure sharing optimization

SE-related metrics are often the prime QoS indicators for both users and operators. Consequently, a target objective function can be generically defined as

$$J_i = f\left(\{\eta_{ik}\}_{k \in \mathcal{K}_i}\right) \tag{22}$$

where the function $f(\cdot)$ serves to target different objectives (e.g., minimum user rate, average user rate, prescribed percentile of user rate) usually implemented through suitable heuristics for power and/or pilot allocation. Let us assume that both MNOs estimate their target function, J_1 and J_2 (typically at their corresponding CPUs) when supporting sets $\mathcal{K}_1$ and $\mathcal{K}_2$ of MSs, respectively. The following lemma paves the way to devise a strategy to drive the inter-MNO user transfer.

Lemma 1: If a user $k \in \mathcal{K}_i$ is transferred from MNO$_i$ to MNO$_j$ with $i \neq j$ and the target metrics for both networks are re-evaluated (denoted by J_1' and J_2') then it holds:

1) $J_i' \geq J_i$
2) $J_j' \leq J_j$.

Proof. After transferring user $k \in \mathcal{K}_i$, MNO$_i$ is supporting $K_i - 1$ MSs. This implies equal or less (never superior) pilot contamination and fewer constraints when designing the IP-MMSE combiner, thus leading to a higher target metric J_i'. In contrast, MNO$_j$ has to endure with an equal or higher (never inferior) level of pilot contamination and has to face further constraints when designing the combiner as the stream for the newly served user k needs to be isolated, thus resulting in a lower target metric J_j'. □

Under the assumption that the two MNOs agree on cooperating, it is possible to define an overall metric, J_O, that encompasses all MSs in the coverage area

$$J_O = f\left(\{\eta_{ik}\}_{k\in\mathcal{K}}\right). \tag{23}$$

The optimization of this global metric can now be pursued by the master operator (MOP). Mathematically, the problem to be solved is

$$\begin{aligned} \mathcal{K}_1^*, \mathcal{K}_2^* = &\arg\max_{\mathcal{K}_1,\mathcal{K}_2} (J_O - \Delta) \\ &\text{subject to} \\ &\mathcal{K}_1 \cup \mathcal{K}_2 = \mathcal{K}_1^{(0)} \cup \mathcal{K}_2^{(0)} \\ &\mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset, \end{aligned} \tag{24}$$

where the first restriction ensures that no user is left unserviced and the second restriction ensures that all MSs are assigned to either MNO$_1$ or MNO$_2$ but not both. Parameter $\Delta = (|K_1^{(0)} - K_1| + |K_2^{(0)} - K_2|)\delta$ with δ a suitable constant, is a dimensionless penalty term associated to the cost of transferring users from home to foreign networks. The particular case of $\delta = 0$ reflects the case where no such penalty applies and users can be swapped across networks at no cost.

The solution to (24) requires an exhaustive search and evaluation over all of the 2^K possible user groupings, a figure that becomes challenging even for modest K. Consequently, the SE greedy strategy depicted in Algorithm 1 is proposed. This algorithm progresses iteratively by first identifying which of the two MNOs performs worse regarding the local objective metric. It then proceeds to swap the user with the lowest spectral efficiency of the worst network to the alternative MNO and checks whether the user transfer improves the overall penalized target metric. This procedure is repeated until no further improvements are registered. Experimental results have shown that the number of iterations Algorithm 1 executes is related to how unequal the conditions of the two networks are: when the two intervening networks have similar levels of infrastructure and load, only a few iterations are required (e.g. 1 to 3), whereas when the conditions are very disparate, the number of required iterations is proportional to their imbalance.

just a small number of iterations is required for the algorithm to converge to its final solution.

Algorithm 1 Greedy SE-based infrastructure sharing.

Inputs: $\mathcal{K}_1^{(0)}$, $\mathcal{K}_2^{(0)}$ and δ.
Initialization:
$J_O^{(0)} = 0,\ \Delta^{(0)} = 0,\ l = 1, \mathcal{K}_1^{(l)} = \mathcal{K}_1^{(0)},\ \mathcal{K}_2^{(l)} = \mathcal{K}_2^{(0)}$
$J_O^{(l)} = f\left(\{\eta_{1k}^{(l)}\}_{k\in\mathcal{K}_1^{(l)}}, \{\eta_{2k}^{(l)}\}_{k\in\mathcal{K}_2^{(l)}}\right), \Delta^{(l)} = 0$
User allocation refinement:
while $(J_O^{(l)} - \Delta^{(l)}) \geq (J_O^{(l-1)} - \Delta^{(l-1)})$ **do**
$l = l + 1$
$\bar{i} = \arg\min_{i\in\{1,2\}} f\left(\{\eta_{ik}^{(l-1)}\}_{k\in\mathcal{K}_i^{(l-1)}}\right)$, $\bar{j} = \{1,2\} \setminus \bar{i}$
$k^* = \arg\min_{k\in\mathcal{K}_{\bar{i}}^{(l-1)}} \eta_{ik}^{(l-1)}$
$\mathcal{K}_{\bar{i}}^{(l)} = \mathcal{K}_{\bar{i}}^{(l-1)} \setminus \{k^*\}$, $\mathcal{K}_{\bar{j}}^{(l)} = \mathcal{K}_{\bar{j}}^{(l-1)} \cup \{k^*\}$
$J_O^{(l)} = f\left(\{\eta_{1k}{}^{(l)}\}_{k\in\mathcal{K}_1^{(l)}}, \{\eta_{2k}{}^{(l)}\}_{k\in\mathcal{K}_2^{(l)}}\right)$
$\Delta^{(l)} = (|K_1^{(0)} - K_1^{(l)}| + |K_2^{(0)} - K_2^{(l)}|)\delta$
end
Output: $\hat{\mathcal{K}}_1^* = \mathcal{K}_1^{(l-1)}, \hat{\mathcal{K}}_2^* = \mathcal{K}_2^{(l-1)}$.

B. Low-complexity infrastructure sharing

The greedy optimization introduced in the previous subsection relies on the users' SEs to transfer MSs among the two MNOs. Unfortunately, in general, no closed-form expressions exist to estimate the SEs. In particular, the expectation operator in (17) needs to be conducted by averaging over multiple fast fadings. Also note that any user transfer checked implies the recalculation of the channel estimates and combiners. These two traits result in a considerable computational complexity further exaggerated by the iterative nature of Algorithm 1. Another handicap of the procedure in Section III-A is that scalability is compromised as SEs of a given user are likely to be affected by all the other MSs in the network, even by very distant users possibly controlled by another CPU. Consequently, a practically-deployable low-cost alternative becomes necessary. In trying to fulfill this goal, a new algorithm is proposed that just relies on performance metrics derived from the large-scale propagation gains. A new user-based metric is defined for MS k of MNO i with respect to the home APs as

$$T_{ki} = g(\tilde{\beta}_{i1k}, \ldots, \tilde{\beta}_{iM_ik}), \tag{25}$$

where $g(\cdot)$ denotes a function such as the sum or the minimum value of the evaluated elements, and

$$\tilde{\beta}_{imk} = \begin{cases} \beta_{imk} & \text{if } \boldsymbol{D}_{imk} = \boldsymbol{I}_N \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Algorithm 2 Low-complexity greedy infrastructure sharing.

Inputs: $\mathcal{K}_1^{(0)}$, $\mathcal{K}_2^{(0)}$, δ and $\beta_{imk}\ \forall i,m,k$.
Initialization:
$T_O^{(0)}=0,\ \Delta^{(0)}=0,\ l=1,\ \mathcal{K}_1^{(l)}=\mathcal{K}_1^{(0)},\ \mathcal{K}_2^{(l)}=\mathcal{K}_2^{(0)}$
$T_{k_i}^{(l)}=g(\tilde{\beta}_{i1k},\ldots,\tilde{\beta}_{iM_ik})\ \forall k\in\mathcal{K}_i^{(l)},\forall i$
$T_O^{(l)}=f\left(\{T_{k_1}^{(l)}\}_{k\in\mathcal{K}_1},\{T_{k_2}^{(l)}\}_{k\in\mathcal{K}_2}\right)$
User allocation refinement:
while $(T_O^{(l)}-\Delta^{(l)})\geq(T_O^{(l-1)}-\Delta^{(l-1)})$ **do**
$l=l+1$
$\tilde{\mathcal{K}}_i^{(l)}=\{k\in\mathcal{K}_i^{(l)}\setminus T_{k_i}^{(l)}<T_{k_j}^{(l)}\},\ \forall i,j$
$\bar{i}=\arg\min_{i\in\{1,2\}} f\left(\{T_{k_i}^{(l)}\}_{k\in\tilde{\mathcal{K}}_i^{(l)}}\right),\ \bar{j}=\{1,2\}\setminus\bar{i}$
$k^*=\arg\min_{k\in\mathcal{K}_{\bar{i}}^{(l-1)}}\left(\{T_{k_{\bar{i}}}^{(l)}\}_{k\in\mathcal{K}_{\bar{i}}}\right)$
$\mathcal{K}_{\bar{i}}^{(l)}=\mathcal{K}_{\bar{i}}^{(l-1)}\setminus\{k^*\},\ \mathcal{K}_{\bar{j}}^{(l)}=\mathcal{K}_{\bar{j}}^{(l-1)}\cup\{k^*\}$
$T_{k_i}^{(l)}=g(\tilde{\beta}_{i1k},\ldots,\tilde{\beta}_{iM_ik})\ \forall k\in\mathcal{K}_i^{(l)},\forall i$
$T_O^{(l)}=f\left(\{T_{k_1}^{(l)}\}_{k\in\mathcal{K}_1},\{T_{k_2}^{(l)}\}_{k\in\mathcal{K}_2}\right)$
$\Delta^{(l)}=(|K_1^{(0)}-K_1^{(l)}|+|K_2^{(0)}-K_2^{(l)}|)\delta$
end
Output: $\hat{\mathcal{K}}_1^*=\mathcal{K}_1^{(l-1)},\hat{\mathcal{K}}_2^*=\mathcal{K}_2^{(l-1)}$.

Thanks to the availability of the large-scale propagation parameters between MS $k\in\mathcal{K}_i$ and APs in $\mathcal{M}_j$, the user-based metric with respect to the foreign APs can also be found as

$$T_{kj}=g(\tilde{\beta}_{j1k},\ldots,\tilde{\beta}_{jM_jk}). \tag{27}$$

Note that T_{kj} actually takes into account the connectivity MS k would obtain once transferred to MNO$_j$ in light of the active users in the alternative network $\mathcal{K}_j$.

As before, an overall metric T_O can now be defined as

$$T_O=f(\{T_{k1}\}_{k\in\mathcal{K}_1},\{T_{k2}\}_{k\in\mathcal{K}_2}), \tag{28}$$

where, as in the SE-based case, $f(\cdot)$ can target the minimum value, the mean or any other percentile. These two metrics drive the procedure shown in Algorithm 2, which should be understood as a low-complexity variant of Algorithm 1. In this case, what the algorithm pursues, again in a greedy fashion (one candidate at a time), is to identify the worst-performing user who would see his performance improved by switching to the alternative MNO and this in turn will improve the overall performance of the two networks. In particular, it analyzes the switch of MSs in $\mathcal{K}_i$ for whom $T_{kj}>T_{ki}$ by checking whether this change improves the overall metric T_O. It is worth pointing out that in Algorithm 2, the required minimization searches to determine $\bar{i}$ and k^* are conducted over metrics solely relying on large-scale propagation information that any given CPU has readily available from the MSs it is controlling.

IV. Numerical results

A square-shaped coverage area is considered with side length $L=1000$ m where wrap-around is applied to eliminate boundary effects from the simulation results. This area is served by two MNOs, MNO$_1$ and MNO$_2$, operating at frequencies $f_{c_1}=3.5$ and $f_{c_2}=3.55$ GHz, respectively, exploiting non-overlapping bandwidths $B_1=B_2=20$ Mhz. Every AP, irrespective of the MNO, is equipped with a uniform linear array (ULA) of $N_1=N_2=4$ antenna elements. The spatial correlation matrices of the NLOS component have been generated according to [22] with angular standard deviations denoted by σ_ζ^A and σ_ζ^E. All MSs and APs are assumed to be uniformly distributed throughout the coverage area and thus, the positions of APs and MSs managed by each of the MNOs are uncorrelated. For the results shown here, the target metric J_i aims at maximizing the worst 10% of users' rate, a criterion well-aligned with the cell-free philosophy of ensuring acceptable performance to even the worst users in the network [2]. Table I lists the rest of common parameters used to generate the results shown next which have been selected in line with most of the cell-free literature (see for example [22]).

TABLE I: Summary of default simulation parameters

Parameter (symbol)	Value
AP maximum power (P_{AP})	200 mW
MS maximum power (P_{MS})	100 mW
Fractional power allocation factor (υ)	-0.5
AP / MS antenna height ($h_{\text{AP}}/h_{\text{MS}}$)	10 m / 1.65 m
Frame length (τ_f)	200 samples
Training phase length (τ_p)	15 samples
Shadow fading decorrelation distance d_{dcorr}	15 m
Shadow fading correlation among APs	0.5
Rician K-factor parameters (μ_K/σ_K^2)	9 / 3.5
Azimuth Angular standard deviations (σ_ζ^A)	15°
Elevation Angular standard deviations (σ_ζ^E)	10°
Sharing transfer penalty factor (δ)	0
Number of realizations per simulation	100

A balanced scenario is first considered where the coverage area is served by MNOs with the same degree of densification ($M_1=M_2=50$ APs) while serving equal network loads ($K_1=K_2=15$ MSs). Fig. 2 depicts the cumulative distribution functions (CDFs) of the per-user SE considering the K users in the coverage area. It can be appreciated how the two infrastructure sharing schemes, both, the greedy optimal and the greedy low computational perform similarly and offer a gain of around 0.6 bits/s/Hz for the worst-10% of users. Despite this relatively modest gain of around 20% achieved through sharing, note that it comes without any extra infrastructure inversion from any of the MNOs with the benefit being solely attributed to the network diversity effect, that is, the fact that the worst users in each MNO user set have the possibility to swap to the other network. While on some individual realizations, imbalances have been observed with one MNO serving a higher number of users than the other, as expected, on average, both MNOs ended up supporting the same network load with an average number of users transferred from one MNO to the other of little more than 2 users.

Unbalanced scenarios offer far more potential for the proposed infrastructure sharing techniques to shine. Toward this end, an scenario is considered where one MNO has a dense infrastructure ($M_1=50$ APs) while the other one has a significantly lower level of densification ($M_2=20$ APs). Critically, both aim at serving at identical network loads

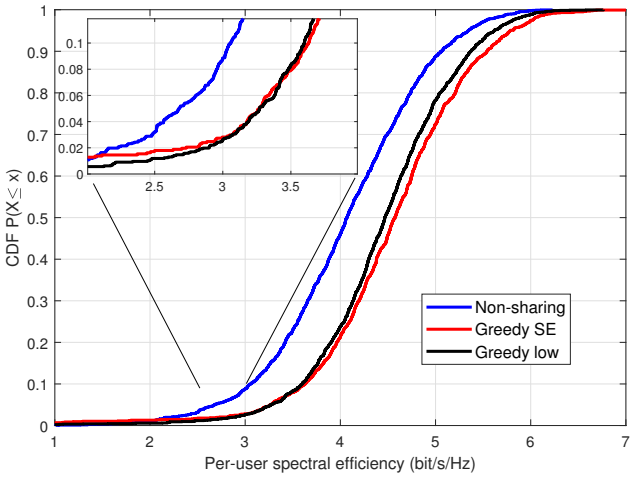

Fig. 2: Balanced scenario. Sharing vs non-sharing SEs for $M_1 = M_2 = 50$ APs supporting $K_1 = K_2 = 15$ MSs.

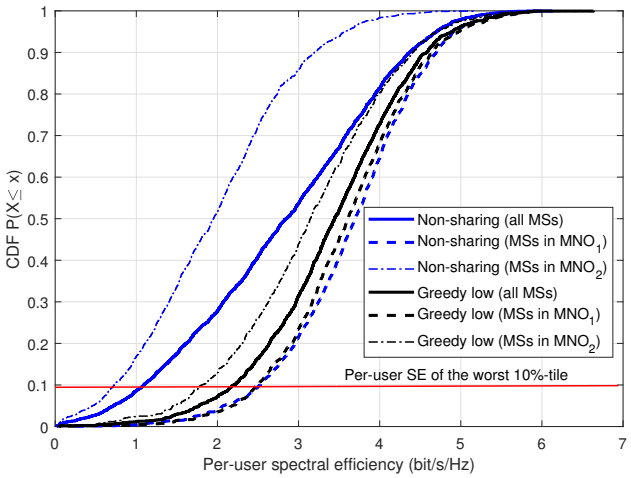

Fig. 3: Unbalanced scenario. Sharing vs non-sharing SEs for $M_1 = 50, M_2 = 20$ APs supporting $K_1 = K_2 = 25$ MS. The red line denotes the 10%-tile of worst user rates.

($K_1 = K_2 = 25$ MSs), noting in this case that since $\tau_p = 15$ samples, pilot contamination appears. Fig. 3 details in blue the curves corresponding to the performance achieved without infrastructure sharing. The solid blue curve denotes the CDF of the per-user SE of the whole set of users in the coverage area. To gain further insight, the performance achieved by each MNO is displayed separately. The dashed blue curve corresponds to the CDF of the SE for the users managed by MNO_1, which can be seen to more than treble that of the users managed by MNO_2 (dot-dashed blue curve) when focusing on the worst 10% of users of each network (from 0.8 bits/s/Hz to 2.5 bits/s/Hz). This result is to be expected since MNO_2 has a poor infrastructure to support such a large user demand. Curves in black depict the performance achieved when allowing infrastructure sharing (low computational technique in this case). When examining the overall results (solid black curve), it is noticeable that overall performance, when accounting for the 50 users operating in the coverage area, sharing results in an improvement exceeding 100% (from 1 bit/s/Hz to 2.15 bit/s/Hz). Examining the results achieved by each MNO under sharing is revealing: the proposed technique allows to (mostly) transfer users from MNO_2 to MNO_1 so that the remaining users in MNO_2 see their rates greatly enhanced (black dash-dot curve) while the MNO_1 is forced to serve a larger number of users but with only a small performance penalty (dashed black curve). In particular, and comparing the two dashed curves, note how the worst 10% users in MNO_2 do no experience any service degradation and it is only the users that are better off who pay a small performance penalty. As expected in such an unbalanced situation, on average, MNO_1 ends up supporting 31 users approximately while MNO_2 keeps providing service to an average of 19 users. On average, in this specific setup, the average number of iterations in Algorithm 2 was found to be 8.4. Qualitatively identical results are obtained when two networks with the same densification level serve distinct network loads.

V. Conclusion

This paper has proposed a scheme that allows two independent CF-mMIMO networks serving a common area to cooperate by allowing the interchange of MSs. User swaps are carried out whenever they bring along an overall improvement of the performance of all users in the area. Two greedy techniques have been proposed to guide the user grouping. One targets the maximization of an arbitrary percentile of the users' SE and, despite its effectiveness, its non-scalability precludes its practical implementation. A second approach is then presented that relies only on large-scale propagation parameters that can be acquired and utilized in a scalable manner. Simulation results reveal that both techniques perform similarly while they are effective in exploiting network diversity.

Acknowledgements

We acknowledge support from grants GERMINAL (TED2021-131624B-I00, funded by MICIU/AEI/10.13039/501100011033, Spain, and by European Union "NextGenerationEU"/PRTR) and SOFIA-WIND (PID2023-147305OB-C33, funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU).

References

[1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

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

[3] I. Kanno, K. Yamazaki, Y. Kishi, and S. Konishi, "A survey on research activities for deploying cell free massive MIMO towards beyond 5G," *IEICE Transactions on Communications*, p. 2021MEI0001, 2022.

[4] Y. Xiao, P. Mähönen, and L. Simić, "System cost analysis of scalable cell-free massive MIMO architectures for 6G networks," in *2022 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2022, pp. 310–316.

[5] A. L. Fernandes, D. D. Souza, C. Natalino, F. Tonini, A. M. Cavalcante, P. Monti, and J. C. Costa, "A cost assessment methodology for user-centric distributed massive MIMO architectures," *IEEE Open Journal of the Communications Society*, 2024.

[6] 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 *2022 16th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2022, pp. 1–5.
[7] J. Gao, Q. Zhang, J. Zhang, and S. Jin, "A stochastic-geometry analysis on scaling up cell-free massive MIMO systems with a cost-efficient topology," *IEEE Communications Letters*, 2025.
[8] Z. H. Shaik, E. Björnson, and E. G. Larsson, "MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7775–7789, 2021.
[9] G. Femenias and F. Riera-Palou, "From cells to freedom: 6G's evolutionary shift with cell-free massive MIMO," *IEEE Transactions on Mobile Computing*, vol. 24, no. 2, pp. 812–829, 2025.
[10] T. Kim, H. Kim, S. Choi, and D. Hong, "How will cell-free systems be deployed?" *IEEE Commun. Mag.*, vol. 60, no. 4, pp. 46–51, 2022.
[11] S. Buzzi, C. D´ Andrea, L. Wang, A. Hasim Gokceoglu, and G. Peters, "Co-existing/cooperating multicell massive MIMO and cell-free massive MIMO deployments: Heuristic designs and performance analysis," *IEEE Open Jour. of the Commun. Society*, vol. 5, pp. 6180–6200, 2024.
[12] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
[13] L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando, "On optimal infrastructure sharing strategies in mobile radio networks," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 5, pp. 3003–3016, 2017.
[14] T. Sanguanpuak, S. Guruacharya, E. Hossain, N. Rajatheva, and M. Latva-aho, "Infrastructure sharing for mobile network operators: Analysis of trade-offs and market," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2804–2817, 2018.
[15] 3GPP, "Technical specification group services and system aspects; telecommunication management;network sharing;concepts and requirements," 3GPP TR 32.130 (version 19.0.0), March 2025.
[16] F. Riera-Palou and G. Femenias, "Decentralization issues in cell-free massive MIMO networks with zero-forcing precoding," in *57th IEEE Allerton Conf.*, 2019, pp. 521–527.
[17] S. Kim, S. Ahn, J. Park, J. Youn, Y. Kwon, and S. Cho, "Revisiting the coverage boundary of multi-CPU cell-free massive MIMO: CPU cooperation aspect," in *ICC 2023-IEEE International Conference on Communications*. IEEE, 2023, pp. 1022–1028.
[18] Y. Guo, D. Wang, X. Xia, Z. Zhang, J. Li, P. Zhu, and X. You, "Stochastic geometry analysis of scalable cell-free RAN with dynamic association and deployment," *IEEE J. of Sel. Topics in Sig. Proc.*, 2025.
[19] 3GPP, "Study on channel model for frequencies from 0.5 to 100 ghz (Release 18)," 3GPP TR 38.901 (version 18.0.0), April 2024.
[20] Ö. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive MIMO with Rician fading and phase shifts," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5299–5315, 2019.
[21] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 1, pp. 77–90, 2019.
[22] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, 2021.
[23] G. Femenias, F. Riera-Palou, and E. Björnson, "Another twist to the scalability in cell-free massive MIMO networks," *IEEE Trans. on Communications*, vol. 71, no. 11, pp. 6793–6804, 2023.