Minimizing Power Consumption under SINR Constraints for Cell-Free Massive MIMO in O-RAN

Vaishnavi Kasuluru, Luis Blanco, Miguel Ángel Vázquez, Cristian J. Vaca-Rubio, Engin Zeydan

Centre Technologic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Barcelona, Spain, 08860.

Emails: {vkasuluru, lblanco, mavazquez, cvaca, ezeydan}@cttc.es

Abstract—This paper deals with the problem of energy consumption minimization in Open RAN cell-free (CF) massive Multiple-Input Multiple-Output (mMIMO) systems under minimum per-user signal-to-noise-plus-interference ratio (SINR) constraints. Considering that several access points (APs) are deployed with multiple antennas, and they jointly serve multiple users on the same time-frequency resources, we design the precoding vectors that minimize the system power consumption, while preserving a minimum SINR for each user. We use a simple, yet representative, power consumption model, which consists of a fixed term that models the power consumption due to activation of the AP and a variable one that depends on the transmitted power. The mentioned problem boils down to a binary-constrained quadratic optimization problem, which is strongly non-convex. In order to solve this problem, we resort to a novel approach, which is based on the penalized convex-concave procedure. The proposed approach can be implemented in an O-RAN cell-free mMIMO system as an xApp in the near-real time RIC (RAN intelligent Controller). Numerical results show the potential of this approach for dealing with joint precoding optimization and AP selection.

Index Terms—Open RAN, Cell-free, SCA, Optimization

I. INTRODUCTION

Cell-free (CF) massive multiple-input-multiple-output (mMIMO) systems is a key enabling technology for 6G networks [1] seemingly merging and pushing forward two key techniques already in use in 5G, namely, network densification and massive MIMO antenna arrays. In a nutshell, CF-mMIMO takes advantage of network MIMO by assuming a large number of access points (APs) deployed over the coverage area but with joint baseband processing done by a central processing unit [2]. In fact, CF-mMIMO is often described as a conventional mMIMO network whereby the RF heads are pulled apart from the BS and randomly (and densely) scattered throughout the coverage area. The theoretical foundations underpinning mMIMO mostly apply to the CF-mMIMO scenario. In particular, the so-called channel hardening that virtually eliminates the fast fading effects also applies to the CF-mMIMO scenario [3]. Furthermore, its highly distributed topology brings the RF infrastructure geographically closer to the users, thus effectively reducing the propagation losses. This theoretical gain is obtained at the expense of costly and power-hungry equipment that connects APs in-phase and quadrature (IQ) signals to the central control unit and the requirement to maintain strict synchronization among all APs.

Aligned with the Green Deal initiative promoted by the European Union, energy efficiency will again be a fundamental metric in the forthcoming wireless generation [4]. While power consumption in mMIMO mostly addressed issues on radiofrequency and baseband processing units [3], CF-mMIMO must consider new energy-related aspects such as the presence of a dedicated fronthaul connection linking the APs to the CPU and the fixed power expenditure each individual AP entails. In practical networks, these fixed power terms represent the most significant factor in total energy expenditure. In this context, to reduce global power consumption, it might be wise to just activate a few APs in order to attend to specific user data rate demands. Techniques to dynamically switch on/off APs in an attempt to maximize energy efficiency have been proposed in [5] whereby a zero-forcing (ZF) precoder is combined with the heuristic power allocation. Unfortunately, the maximization of energy efficiency is carried out without providing the users with any performance guarantees. Most previous works on CF-mMIMO separate the precoder design from the power allocation strategy. While the first one is carried out relying on instantaneous channel state information (CSI), the second one typically targets a specific performance objective such as maximizing the minimum user signal-to-noise-plus-interference ratio (SINR) (Max-Min) or maximizing the network sum rate.

It has been recently proposed the use of CF-mMIMO within the Open Radio Access Network (O-RAN) architecture [6], representing a groundbreaking convergence of two disruptive technologies. Open RAN's disaggregated framework fosters network operator inter-operability, driving down costs and unlocking vendor diversity. When coupled with CF-mMIMO, this synergy amplifies the benefits by breaking down traditional cell boundaries and leveraging distributed antennas, maximizing spectral efficiency, extending coverage, and minimizing interference. This potent combination not only optimizes resource utilization but also accelerates the deployment of advanced wireless networks, empowering operators to meet the ever-growing demands for connectivity.

In this paper, we focus on the optimization problem of minimizing the power consumption of CF-mMIMO systems within O-RAN architectures, under minimum quality-of-

service (QoS) requirements per user in the form of a minimum SINR. Indeed, relying on the seminal work in [7], we revisit the idea of minimizing the transmit power under minimum SINR [7] for CF-mMIMO considering the fixed power consumption term. In our case, and similar to what happens in [7], the precoder design and power allocation steps merge, and they are both carried out taking into account instantaneous CSI, thus allowing the satisfaction of instantaneous SINR constraints (rather than average). When only the transmit power consumption is assumed, the optimization problem has a convex reformulation [7] that permits a closed-form solution. However, incorporating the fixed term transforms the problem into a highly non-convex problem. In order to tackle this new optimization problem, binary decision variables are introduced that decide which APs are active, along with the derivation of the corresponding precoding vectors. We propose a convex relaxation based on the Penalty Convex-Concave Procedure (PCCP) [8] able to handle the described problem. The numerical results show the benefits of this technique for joint precoding design and AP subset selection, which minimize global power consumption while satisfying user demands.

The paper's outline is as follows: Section II introduces the architecture of the Cell-Free O-RAN system, followed by the System model in Section III. Section IV describes the considered problem statement. Section V and Section VI present the numerical results and conclusions, respectively.

II. Cell-Free O-RAN Architecture

Traditional RANs, with their hardware dependency, have long posed challenges for network providers. These systems often lead to vendor lock-in issues, significantly increasing CAPEX and OPEX costs [9]. The struggle to incorporate intelligence and build a collaborative, reliable network environment has been a major hurdle [10]. The promise of next-generation RAN solutions, which can manage the network in real-time and respond quickly to scenarios using software-defined vendor-neutral technologies, is a hope in this landscape [11].

With Virtualization and RAN disaggregation being the cornerstone, O-RAN technology aims to provide openness and intelligence in network management [12]. One of the latest essential foundations is the modernization of O-RAN architecture by integrating it with cell-free technology [6]. This helps in improving the scalability and flexibility of the network. Figure 1 shows the architecture of O-RAN Cell-free technologies [6], incorporating power-saving eXtended application (xApp). It consists of control, management, and data layers. In the Management plane, there is a Service Management and Orchestration (SMO) layer equipped with Non-Real Time RAN Intelligent Controller (RIC). The Non-RT RIC constitutes a fundamental component of the Open RAN architecture, specifically designed to manage RAN operations with control loops exceeding one second. It supports third-party applications, referred to as radio App (rApp)s, which facilitate the optimization and management of RAN functions.

As an integral part of the SMO framework within the O-RAN architecture, the Non-RT RIC connects to other network elements through A1, O1, and O2 interfaces.

There is also a Near-Real Time RIC, which helps in intelligent system management through xApps. In this work, these xApps are considered to use the information from Open-Radio Unit (O-RU)s to make decisions regarding power saving by using a sophisticated optimization algorithm. The xApps primary objective is to promote power efficiency within the network. Moreover, The O-RUs are designed to support a cell-free architecture for serving multiple User Equipments (UEs) in a seamless manner. The Near-Real Time RIC interfaces via A1 and E2 interfaces with other network elements.

Combining O-RAN and cell-free technologies provides various benefits. It reduces signal interference and improves the efficient usage of power. As the APs cooperatively serve multiple UEs, the traffic demands can be handled dynamically, which in turn enhances capacity and coverage [6]. Nevertheless, it also helps ensure network services' consistent availability to improve user experience and handle scalability and agility cost-effectively in the long run [3]. The main components are:

- O-RU, Open-Distribution Unit (O-DU), and Open-Central Unit (O-CU) whose functionalities are similar to that in 5G dis-aggregated RAN except with added support of O-RAN based specifications and interface. Without loss of generality, throughout this paper, we will refer to the O-RUs as cell-free APs [13].
- Near-Real Time RIC to control/optimize RAN elements and resources based on fine-grained data using online Artificial Intelligence (AI)/Machine Learning (ML) based services. It is suitable for applications with latency requirements between 10 ms and 1s [14]. Herein, the power saving xApp is placed in Near-Real Time RIC, selecting the best subset of APs to be active at a particular time to handle the service requirements of users efficiently.
- Non-Real Time RIC to control/optimize RAN elements and resources based on coarse-grained data using online AI/ML services [14]. It is suitable for applications with latency requirements greater than 1s. It also provides policy-based guidance to near-real Time RIC.

III. System Model

Consider a cell-free scenario with K User terminals / User Equipments (UEs) and N Access Points (APs). In the following, we assume that APs are equipped with $N_{\mathrm{AP}} > 1$ antennas. APs are connected through a high-speed link to a central controller. We assume that UEs are equipped with a single antenna.

A. Channel Model

Let us denote by $\beta_k^{[n]}$ the large-scale propagation losses (i.e., path loss and shadowing) of the link joining $\mathrm{AP}n$ and UE k, and which can be expressed as $\beta_k^{[n]} = \zeta_k^{[n]} \chi_k^{[n]}$ with $\zeta_k^{[n]}$ representing the distance-dependent path loss

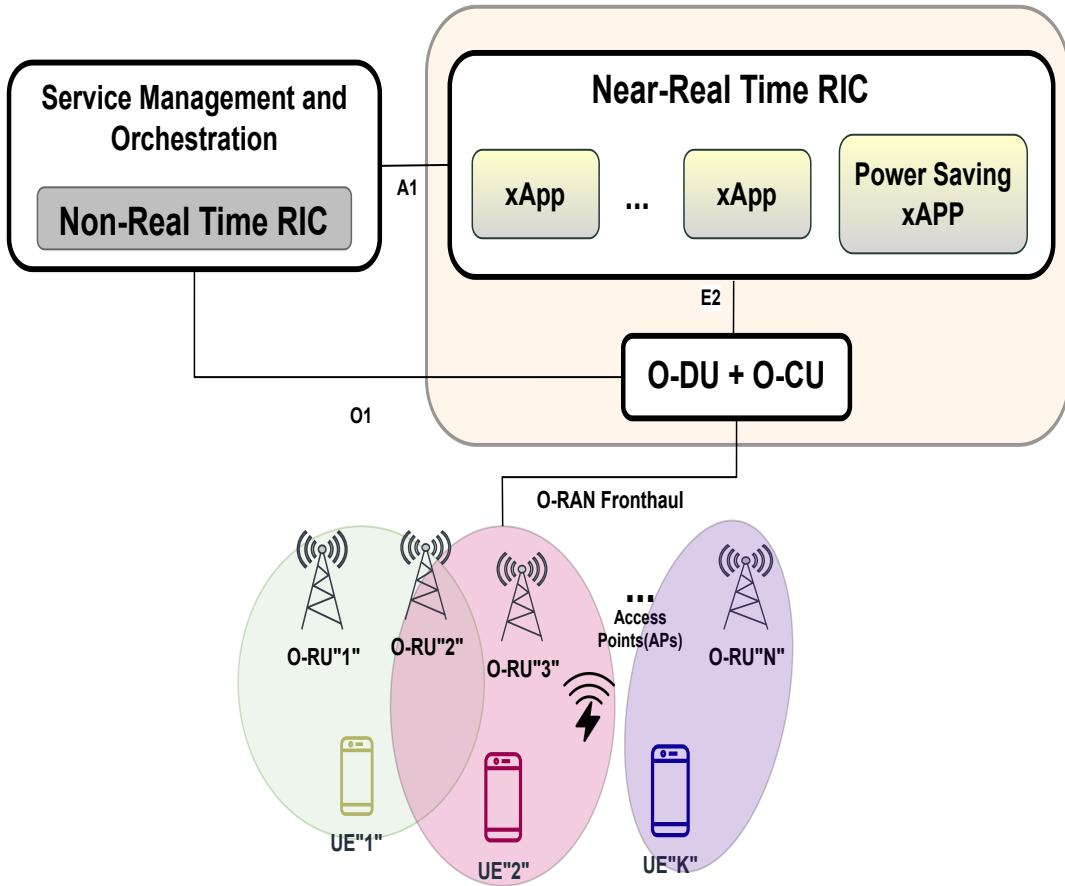$$\zeta_k^{[n]}[\mathrm{dB}] = \zeta_0 + 10\alpha \log_{10}\left(d_k^{[n]}\right), \tag{1}$$

Fig. 1. Cell-free O-RAN Architecture.

where ζ_0 is the path loss at a reference distance of 1 m, $d_k^{[n]}$ is the distance from AP n to UE k and α is the path loss exponent. The term $\chi_k^{[n]}$ corresponds to the shadow fading component modeled as a spatially correlated log-normal random variable with zero mean and variance σ_χ^2 whose spatial correlation model is described in [[2], (54)-(55)]. The resulting downlink channel vector $h_k^{[n]} \in \mathbb{C}^{N_{AP} \times 1}$ from the k-th UE to the n-th AP (including both large-scale and small-scale fading) can then be generically characterized as a correlated Rayleigh fading channel $h_k^{[n]} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{R}_{nk}\right)$ with $\boldsymbol{R}_{nk}$ denoting the $N_{AP} \times N_{AP}$ spatial channel correlation matrix subject to the constraint $\mathrm{Tr}\left(\boldsymbol{R}_{nk}\right) = \beta_k^{[n]}/N^{AP}$. For later convenience, we define $h_k = \left[h_k^{[1],T}, \cdots, h_k^{[n],T}\right]^T$ as the $N_{AP} \times 1$ vector collecting the overall channel responses from the N APs to user k.

B. Uplink training and channel estimation

During the UL training phase, all K UEs simultaneously transmit pilot sequences of τ_p samples to the APs and thus, the $N_{AP} \times \tau_p$ received UL signal matrix at the n-th active AP is given by

$$\boldsymbol{Y}_p^{[n]} = \sqrt{\tau_p P_p^{AP}} \sum_{k=1}^{K} h_k^{[n]} \boldsymbol{\varphi}_k^T + \boldsymbol{N}_p^{[n]}, \tag{2}$$

where P_p^{AP} is the available pilot symbol power, $\boldsymbol{\varphi}_k$ denotes the $\tau_p \times 1$ training sequence assigned to UE k, with $\|\boldsymbol{\varphi}_k\|^2 = 1$, and $N_p^{[n]} \in \mathbb{C}^{N_{AP} \times \tau_p}$ is a matrix of independent identically distributed (iid) zero-mean circularly symmetric Gaussian random variables with standard deviation σ_u. Define now the projection of $\boldsymbol{Y}_p^{[n]}$ on the k-th training sequence

$$\breve{\boldsymbol{y}}_{pk}^{[n]} = \boldsymbol{Y}_p^{[n]} \boldsymbol{\varphi}_k^* = \sum_{k'=1}^{K} \sqrt{\tau_p P_p^{AP}} h_{k'}^{[n]} \boldsymbol{\varphi}_{k'}^T \boldsymbol{\varphi}_k^* + \boldsymbol{N}_p^{[n]} \boldsymbol{\varphi}_k^*. \tag{3}$$

Standard results from estimation theory allow the MMSE channel estimate between of $h_k^{[n]}$ to be expressed as

$$\hat{h}_k^{[n]} = \left(\sqrt{\tau_p \left(P_p^{AP}\right)}/\sigma_u^2\right) \boldsymbol{R}_{nk} \boldsymbol{\Psi}_{nk}^{-1} \breve{\boldsymbol{y}}_{pk}^{[n]}, \tag{4}$$

where

$$\boldsymbol{\Psi}_{nk} = \tau_p \left(P_p^{AP}/\sigma_u^2\right) \sum_{k'=1}^{K} \boldsymbol{R}_{nk'} \left|\boldsymbol{\varphi}_{k'}^H \boldsymbol{\varphi}_k\right|^2 + \boldsymbol{I}_N. \tag{5}$$

The channel estimate $\hat{h}_{k^{[n]}}^{[n]}$ and the MMSE channel estimation error $\tilde{h}_k^{[n]} = h_k^{[n]} - \hat{h}_k^{[n]}$ are uncorrelated random vectors distributed as $\hat{h}_k^{[n]} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Gamma}_{mk}\right)$, and $\tilde{h}_k^{[n]} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{A}_{mk}\right)$, respectively, where

$$\boldsymbol{\Gamma}_{nk} = \tau_p \left(P_p^{\mathrm{AP}} / \sigma_u^2 \right) \boldsymbol{R}_{nk} \boldsymbol{\Psi}_{nk}^{-1} \boldsymbol{R}_{nk}^H \qquad (6)$$

and

$$\boldsymbol{A}_{nk} = \mathbb{E} \left\{ \tilde{\boldsymbol{h}}_k^{[n]} \tilde{\boldsymbol{h}}_k^{[n],H} \right\} = \boldsymbol{R}_{nk} - \boldsymbol{\Gamma}_{nk}. \qquad (7)$$

C. SINR computation

We aim to optimize the downlink transmission, assuming perfect channel state information. This is, the downlink transmission is performed by a central controller which has a perfect estimation of $\mathbf{h}_k \in \mathbb{C}^{N N_{\mathrm{AP}} \times 1}$ for $k = 1, \ldots, K$. Vector $\mathbf{h}_k$ collapses the channel vector of each AP

$$\mathbf{h}_k = \left(\mathbf{h}_k^{[1],T}, \ldots, \mathbf{h}_k^{[N],T} \right)^T, \qquad (8)$$

where $\mathbf{h}_k^{[n],T} \in \mathbb{C}^{N_{\mathrm{AP}} \times 1}$ is the channel vector between the n-th AP and the k-th UE.

The received signal by the k-th UE can be written as

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \sum_{j \neq k}^K \mathbf{h}_k^H \mathbf{w}_j s_j + n_k, \qquad (9)$$

where s_k is the unit norm zero mean symbol of the k-th UE and $\mathbf{w}_k \in \mathbb{C}^{N N_{\mathrm{AP}} \times 1}$ the transmit precoding vector. Note that transmission takes place in all APs simultaneously so that the n-th AP is transmitting

$$\mathbf{t}^{[n]} = \sum_{k=1}^K \mathbf{w}_k^{[n]} s_k, \qquad (10)$$

where $\mathbf{w}_k^{[n]} \in \mathbb{C}^{N_{\mathrm{AP}} \times 1}$ is the precoding vector at the n-th AP transmitting the symbol to the k-th UE. Remarkably

$$\mathbf{w}_k = \left(\mathbf{w}_k^{[1],T}, \ldots, \mathbf{w}_k^{[N],T} \right)^T. \qquad (11)$$

We consider total power constraint

$$\sum_{k=1}^K \left\| \mathbf{w}_k^{[n]} \right\|^2 \leq P_n, \quad n = 1, \ldots, N \qquad (12)$$

where P_n is the available power at the n-th AP. The attainable rate of the k-th user is

$$R_k = \log_2 \left(1 + \mathrm{SINR}_k \right), \qquad (13)$$

where

$$\mathrm{SINR}_k = \frac{\left| \mathbf{h}_k^H \mathbf{w}_k \right|^2}{\sum_{j \neq i}^K \left| \mathbf{h}_k^H \mathbf{w}_j \right|^2 + \sigma^2}. \qquad (14)$$

IV. PROBLEM STATEMENT

In this scenario, we consider minimizing the global power consumption guaranteeing a certain minimum SINR, γ_0 to all terminals. We model the Total Power Consumption (TPC) as follows

$$\mathrm{TPC} = \frac{1}{\eta} \sum_{k=1}^K \| \mathbf{w}_k \|^2 + \sum_{n=1}^N P_{\mathrm{FIX},n} \qquad (15)$$

where η is the high power amplifier efficiency, $\eta \in (0, 1]$ assumed to be the same for all APs and all antennas. The term $P_{\mathrm{FIX},n}$ denotes the fixed consumed power due to backhauling, circuitry power consumption, and control signaling of the n-th AP in case it is active. This power consumption model aligns with other related approaches to the energy consumption of CF-mMIMO systems [4], [15].

Mathematically, this can be written as

$$\underset{\{\mathbf{w}_k\}_{k=1}^K, \mathcal{U}}{\mathrm{minimize}} \frac{1}{\eta} \sum_{k=1}^K \| \mathbf{w}_k \|^2 + P_{\mathrm{FIX}} |\mathcal{U}|$$

subject to

$$\sum_{k=1}^K \left\| \mathbf{w}_k^{[n]} \right\|^2 \leq P_n \quad n = 1, \ldots, N \qquad (16)$$

$$\mathrm{SINR}_k \geq \gamma_0 \quad k = 1, \ldots, K$$

$$\mathcal{U} \subseteq \mathcal{V},$$

where $\mathcal{V}$ denotes the set with all the APs and $|\cdot|$ denotes the cardinality of the set. We assume all APs have the same fixed power consumption P_{FIX}. The aforementioned cardinality problem can be re-written as follows

$$\underset{\{\mathbf{w}_k\}_{k=1}^K, \{b_n\}_{n=1}^N}{\mathrm{minimize}} \frac{1}{\eta} \sum_{k=1}^K \| \mathbf{w}_k \|^2 + P_{\mathrm{FIX}} \sum_{n=1}^N b_n$$

subject to

$$C1 : \sum_{k=1}^K \left\| \mathbf{w}_k^{[n]} \right\|^2 \leq b_n P_n \quad n = 1, \ldots, N \qquad (17)$$

$$C2 : \mathrm{SINR}_k \geq \gamma_0 \quad k = 1, \ldots, K$$

$$C3 : b_n \in \{0, 1\} \quad n = 1, \ldots, N.$$

This latter optimization problem is non-convex due to constraints $C2$ and $C3$. Let us focus on how to deal with these non-convex constraints. As reported in [14], constraint $C1$ can be rewritten as

$$\frac{1}{\gamma_0} \left| \mathbf{h}_k^H \mathbf{w}_k \right|^2 \geq \sum_{j \neq k}^K \left| \mathbf{h}_k^H \mathbf{w}_j \right|^2 + \sigma^2. \qquad (18)$$

This last expression shows that an arbitrary phase shift does not affect the constraint. Therefore, we can consider a real solution such that its imaginary part $\Im \left(\mathbf{h}_k^H \mathbf{w}_k \right) = 0$. In this context, we can transform the original constraint into

$$\frac{1}{\sqrt{\gamma_0}} \Re\left(\mathbf{h}_k^H \mathbf{w}_k\right) \geq \sqrt{\sum_{j \neq k}^{K} \left|\mathbf{h}_k^H \mathbf{w}_j\right|^2 + \sigma^2}, \qquad (19)$$

which is a convex constraint. Alternatively, $C3$ is a binary constraint whose quadratic definition, as reported in [16], can be written as

$$\mathbf{b}^T \mathbf{E}_i \mathbf{b} - \mathbf{b}^T \mathbf{e}_i = 0. \qquad (20)$$

where $\mathbf{E}_i$ is a zero matrix whose i-th diagonal element is equal to 1 and $\mathbf{e}_i$ is a zero vector whose i-th entry is equal to one. This equality constraint can be divided into two different constraints, namely

$$\mathbf{b}^T \mathbf{E}_i \mathbf{b} - \mathbf{b}^T \mathbf{e}_i \leq 0 \qquad (21)$$
$$\mathbf{b}^T \mathbf{E}_i \mathbf{b} - \mathbf{b}^T \mathbf{e}_i \geq 0. \qquad (22)$$

While the first constraint is convex, the latter is concave. For this last constraint, we resort to the Concave-Convex Procedure (CCP), which sequentially approximates the concave parts of the problem by its first order Taylor expansion. Starting from an initial point $\mathbf{z}$, CCP algorithm linearizes the non-convex part, by linearizing it by its affine approximation, leading into the following constraint

$$\mathbf{b}^T \mathbf{e}_i + \mathbf{z}^T \mathbf{E}_i \mathbf{z} - 2\Re\left(\mathbf{z}^T \mathbf{E}_i \mathbf{b}\right) \leq 0. \qquad (23)$$

Bearing in mind the above discussion and the use of slack variables $\mathbf{s}$, the original problem at the t-th step becomes

$$\underset{\{\mathbf{w}_k\}_{k=1}^K, \{b_n\}_{n=1}^N, \{s_n\}_{n=1}^N}{\text{minimize}} \quad \frac{1}{\eta} \sum_{k=1}^K \|\mathbf{w}_k\|^2 + P_{\text{FIX}} \sum_{b=1}^N b_n + \lambda \sum_{n=1}^N s_n$$

subject to

$$C1: \quad \sum_{k=1}^K \left\|\mathbf{w}_k^{[n]}\right\|^2 \leq b_n P_n, n = 1, \ldots, N,$$

$$C2.1: \quad \frac{1}{\sqrt{\gamma_0}} \Re\left(\mathbf{h}_k^H \mathbf{w}_k\right) \geq \sqrt{\sum_{j \neq k}^K \left|\mathbf{h}_k^H \mathbf{w}_j\right|^2 + \sigma^2},$$

$$k = 1, \ldots, K,$$

$$C2.2: \quad \Im\left(\mathbf{h}_k^H \mathbf{w}_k\right) = 0, k = 1, \ldots, K,$$

$$C3.1: \quad \mathbf{b}^T \mathbf{E}_i \mathbf{b} - \mathbf{b}^T \mathbf{e}_i \leq 0, n = 1, \ldots, N,$$

$$C3.2: \quad \mathbf{b}^T \mathbf{e}_i + \mathbf{z}_t^T \mathbf{E}_i \mathbf{z}_t - 2\Re\left(\mathbf{z}_t^T \mathbf{E}_i \mathbf{b}\right) \leq s_i. \qquad (24)$$

The penalized concave-convex procedure is summarized in Algorithm 1. Note that slack variables s_i are introduced to avoid the need for an initial feasible solution. In our case, we consider as initial solution the one that uses all APs.

Algorithm 1 PCCP optimization for On/Off Optimization

1: **Initialization** of $z^{(0)}, b^{(0)}, w^{(0)}$ to random initial points.

2: **Set** $t = 0$ and define the values of $\psi, T_{max}, \omega, \lambda_{\max}, \rho$ and $\lambda^{(0)}$

3: **while** $\sum_{m=1}^{2N+1} s_m \leq \psi$ and $\|b^{(n)} - b^{(n-1)}\| \geq \omega$ **do**
4:    **if** $t < T_{\max}$ **then**
5:       Compute $b^{(n)}$ according to (24).
6:       $z^{(n+1)} \leftarrow b^{(n)}$;
7:       $\lambda^{(n+1)} \leftarrow \max(\lambda^{(n)}\rho, \lambda_{\max})$;
8:       $t \leftarrow t + 1$;
9:    **else**
10:       $t \leftarrow 0$;
11:       Initialize with a new random value $z^{(0)}$;
12:       Set up $\lambda^{(0)}$ again;
13: **Output** the final solution $\mathbf{w}^*, \mathbf{b}^*$;

V. NUMERICAL RESULTS

This section quantifies the performance of the joint AP selection and power consumption minimization algorithm in a cell-free mMIMO O-RAN scenario.

The power saving xApp is placed in Near-RT-RIC, as shown in Figure 1. We consider N=15 APs with four antennas each. The channels are generated following a complex normal distribution of zero mean and unit variance. The number of users K is considered to be 10. The maximum transmit power is fixed to 1W, maximum value of penalty parameter λ_{max} is fixed to 10^4, maximum allowed constraints violation ψ is 10^{-5}, initial value $\lambda^{(0)}$ is set as $5 \cdot 10^{-2}$, multiplicative update value ρ is fixed to 2, maximum number of iterations T_{max} is 10^2 and finally 10^{-5} is assigned to maximum allowed constraints violation ω. All the experiments are run for 100 Monte Carlo runs.
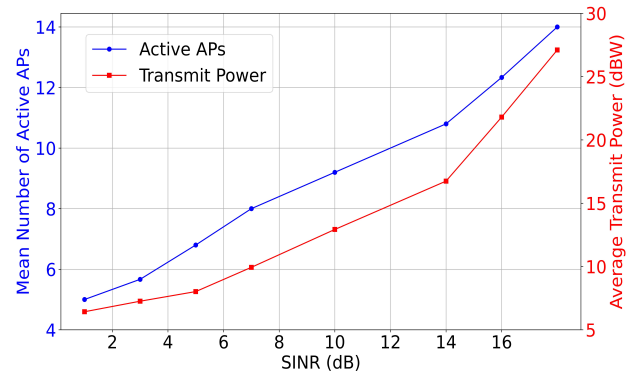

Fig. 2. Mean number active APs and average transmit power as a function of SINR.

Fig. 2 helps understand the relationship between SINR, number of active APs, and power consumption. Our work aims to jointly optimize total power consumption and beamforming

design. A direct consequence of solving the problem (24) is to select the best set of access points to be activated. The results show that the mean number of active access points monotonically increases with SINR. An increase in SINR signifies better signal quality at the UEs, requesting demands for more active access points, leading to a gradual increase in power consumption. The results show that up to 7dB to fulfill the SINR constraints implies a slight increase in the total power. On the contrary, for values higher than 15dB, there is a huge increase in total transmit power, around 9dB, to guarantee the quality of service. This analysis highlights the significant tradeoff between network performance in terms of UE's quality of service and energy efficiency, which, as a consequence, results in increasing active APs.
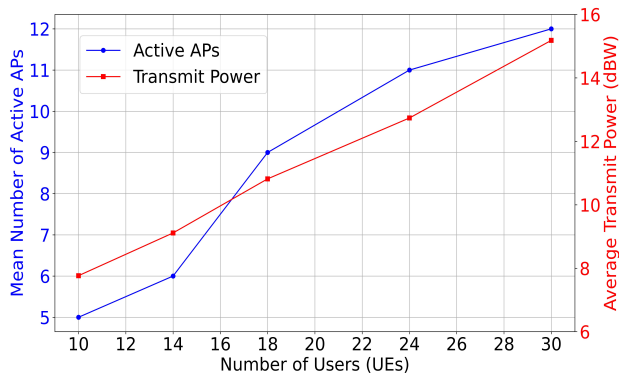

Fig. 4. Empirical Cumulative Distribution Function of the number of active APs for different SINRs.

Fig. 3. Effect of changing the number of users on active APs and transmit power.

Fig. 5. Feasibility of finding an optimal solution for various SINRs.

Fig. 3 shows the impact of increasing the number of UEs on the average transmit power and the mean number of active access points. For this analysis, the SINR value was fixed to 4dB. As the total number of UEs rises, there is a consistent increase in the number of APs activated to serve these users. In the same way, the transmit power is also increased. To maintain the quality of service, in terms of SINR at the terminals, a higher number of active points needs to be increased, having an impact on the total transmit power in the system.

The Fig. 4 shows the Empirical Cumulative Distribution Function (ECDF) of the actual number of active APs corresponding to 100 Montecarlo runs for SINRs equal to 4dB and 7dB, the number of available APs $N = 10$, and the number of UEs $K = 21$. More access points need to be activated as the SINR constraint increases, shifting the ECDFs of required APs further to the right. This is because to find a feasible solution in higher SINR regimes, a larger minimum amount of APs is needed, as shown in the 7dB case. Around 90% of cases require 8 APs to fulfill the constraints. Furthermore, the figure shows that for lower values of SINR, i.e., 4dB, broader solutions in terms of active APs are found. For this case, the total amount of APs required ranges from 6 to 10.

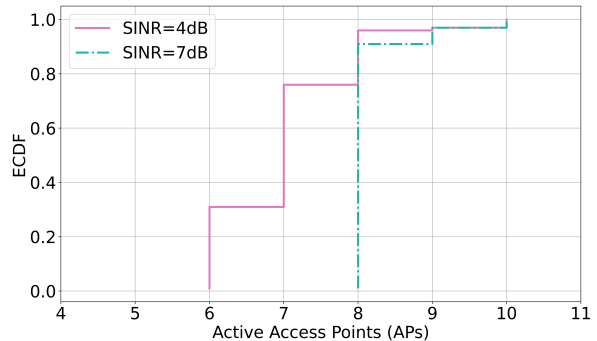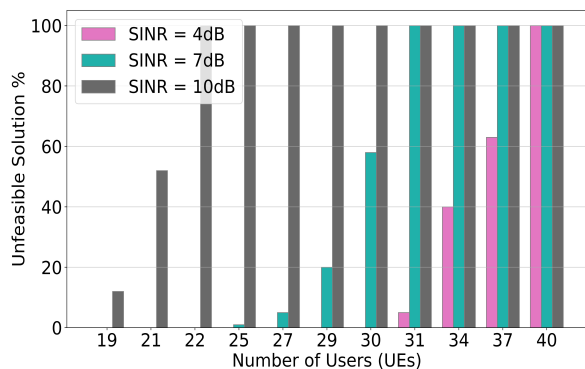Fig. 5 investigates the effect of increasing the total number of users on the feasibility of finding optimal solutions to problem (24). Having this aim in mind, different SINR values were evaluated, namely 4dB, 7dB, and 10dB, with the number of available APs $N = 10$. In the case of 4dB SINR, when the number of users in the system is higher than 30, in some instances the problem becomes unfeasible. Similarly, for cases of 7dB and 10dB, this unfeasibility effect starts earlier when the number of users is 25 and 19, respectively. The rationale behind this behavior is that for the given number of APs and antennas, $N = 10$ and $K = 4$, respectively, fulfilling the quality of service constraints becomes unfeasible for a higher number of users. This effect is consistently increased, and when the number of users reaches 40, the optimization problem completely fails in finding a feasible solution for all the listed SINRs. Note that increasing the number of APs and/or antennas would allow for a higher number of users to be served before reaching unfeasibility.

VI. CONCLUSIONS

This paper presented a novel approach for addressing the energy consumption minimization problem in Open RAN CF-mMIMO systems while guaranteeing the quality of service requirements per user. This problem involves the joint precoding design and the AP selection, with a minimum SINR requirement for each user. The combinatorial nature

of the problem makes it computationally demanding, and a sub-optimal method was presented, based on the penalized convex-concave procedure. Our technique offered a promising solution, demonstrating its efficacy through numerical simulations. Furthermore, the proposed approach, which can be implemented as an xApp in the near-time RIC, holds significant potential for practical implementation of O-RAN-based CF-mMIMO systems.

REFERENCES

[1] M. Matthaiou, O. Yurduseven, H. Q. Ngo, D. Morales-Jimenez, S. L. Cotton, and V. F. Fusco, "The road to 6g: Ten physical layer challenges for communications engineers," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 64–69, 2021.

[2] 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.

[3] R. Beerten, V. Ranjbar, A. P. Guevara, and S. Pollin, "Cell-free massive mimo in the o-ran architecture: Cluster and handover strategies," in *IEE GLOBECOM*, pp. 5943–5948, 2023.

[4] A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Towards optimal energy efficiency in cell-free massive mimo systems," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 816–831, 2021.

[5] G. Femenias, N. Lassoued, and F. Riera-Palou, "Access point switch on/off strategies for green cell-free massive mimo networking," *IEEE access*, vol. 8, pp. 21788–21803, 2020.

[6] Ö. T. Demir, M. Masoudi, E. Björnson, and C. Cavdar, "Cell-free massive mimo in o-ran: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE Journal on Selected Areas in Communications*, 2024.

[7] A. Zhou, J. Wu, E. G. Larsson, and P. Fan, "Max-min optimal beamforming for cell-free massive mimo," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2344–2348, 2020.

[8] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, pp. 263–287, 2016.

[9] V. Kasuluru, L. Blanco, and E. Zeydan, "On the use of probabilistic forecasting for network analysis in open ran," in *2023 IEEE Meditcom*, pp. 258–263, 2023.

[10] M. Polese *et al.*, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, 2023.

[11] A. Kliks, M. Dryjanski, V. Ram, L. Wong, and P. Harvey, "Towards autonomous open radio access networks," *ITU Journal on Future and Evolving Technologies*, vol. 4, no. 2, pp. 251–268, 2023.

[12] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in o-ran for data-driven nextg cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021.

[13] M. S. Oh, A. B. Das, S. Hosseinalipour, T. Kim, D. J. Love, and C. G. Brinton, "A decentralized pilot assignment algorithm for scalable o-ran cell-free massive mimo," *IEEE Journal on Selected Areas in Communications*, 2023.

[14] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5g: When cellular networks meet artificial intelligence," *IEEE Wireless communications*, vol. 24, no. 5, pp. 175–183, 2017.

[15] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive mimo," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, 2017.

[16] M. A. Vázquez, L. Blanco, and A. I. Pérez-Neira, "Hybrid analog–digital transmit beamforming for spectrum sharing backhaul networks," *IEEE transactions on signal processing*, vol. 66, no. 9, pp. 2273–2285, 2018.