IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 July 2024, accepted 14 July 2024, date of publication 22 July 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432287

RESEARCH ARTICLE

Energy-Efficient Resource Allocation for Underlay Spectrum Sharing in Cell-Free Massive MIMO

ZAKIR HUSSAIN SHAIK[1], (Graduate Student Member, IEEE), RIMALAPUDI SARVENDRANATH[2], (Member, IEEE), AND ERIK G. LARSSON[1], (Fellow, IEEE)

[1]Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden
[2]Department of Electrical Engineering, Indian Institute of Technology Tirupati, Tirupati 517619, India

Corresponding author: Zakir Hussain Shaik (zakir.hussain.shaik@liu.se)

This work was supported in part by the REINDEER Project of European Union's Horizon 2020 Research and Innovation Program under Grant 101013425, and in part by the ELLIIT and KAW. The work of Rimalapudi Sarvendranath was supported in part by the Science and Engineering Research Board (SERB) Research under Grant SRG/2023/000070, and in part by the Department of Science and Technology (DST) INSPIRE Research Grant.

ABSTRACT Cell-free massive multiple-input-multiple-output (CF-mMIMO) networks incorporate a cell-free architecture with distributed antennas in a geographical area, and aim to deliver high data rates and support large numbers of users. It is crucial that such networks operate in an energy-efficient manner within the available spectrum. Thus, we focus on maximizing the energy efficiency (EE) of a CF-mMIMO network, which coexists with a collocated primary network in underlay mode. The EE maximization is a non-convex problem and to compute a power allocation policy efficiently, we propose a weighted minimum-mean-square-error (WMMSE) based Dinkelbach's algorithm. Besides this, we also provide a simplified algorithm for maximum-ratio precoding in which we approximate the non-convex EE objective function with a lower bound, transforming the non-convex EE problem into a convex problem. Subsequently, we propose a policy for downlink power allocation that maximizes the EE of the secondary CF-mMIMO network while adhering to power constraints at each access point and interference constraints at each primary user. We also compare with some heuristic power allocation policies. The results demonstrate that the proposed WMMSE based power allocation scheme outperforms the heuristic power allocation schemes to a significant degree.

INDEX TERMS Beyond 5G, cell-free massive MIMO, downlink, energy efficiency, spectrum sharing, optimization.

I. INTRODUCTION

Cell free massive multiple-input-multiple-output (CF-mMIMO) is a potential next-generation wireless technology wherein multiple access points (APs) distributed in a large geographical area jointly serve user equipment (UEs) in the region without cell-boundaries [2]. In other words, CF-mMIMO is a distributed implementation of massive MIMO, and variants of such technologies such as RadioWeaves have recently gained interest [3] because of their superior features such as high spectral efficiency (SE) and macro-diversity. Further, the interference management capability of the CF-mMIMO system makes it a suitable candidate to operate in spectrum sharing systems, which are essential to deploy new applications in the available limited spectrum [4]. In such applications, CF-mMIMO's ability to efficiently support large numbers of devices in the limited spectrum is crucial.

Accepting the need for such applications, spectrum regulators such as the federal communication commission (FCC) opened up the 3 GHz (3.55 - 3.70 GHz) and 6 GHz bands 5.925-7.125 GHz for shared spectrum operations [5], [6]. Furthermore, wireless standards such as long term evolution-license assisted access, MultiFire, citizen's broadband radio service, 5G new radio unlicensed, and IEEE 802.11be support spectrum sharing [7]. Among various

The associate editor coordinating the review of this manuscript and approving it for publication was Walid Al-Hussaibi.

spectrum sharing mechanisms, underlay spectrum sharing is one of the important approaches. In it, a secondary network (SN), which does not have the license to use the spectrum, operates concurrently with a primary network (PN) that owns the license [8]. So, it is crucial that the SN operates within the permissible levels of interference that it can incur to the PN. This implies that the SN should use its energy efficiently to serve its UEs alongside satisfying the interference constraint.

In communications system design, *energy efficiency (EE)* is an important metric that measures the number of information bits a system can reliably transmit per unit of energy. Maximizing EE involves balancing achievable SE and energy consumption. However, obtaining optimal power allocation to maximize EE often involves solving a non-convex problem, and good heuristic algorithms are required. The non-convexity arises due to the ratio of non-convex SE, a non-convex function of power allocation coefficients, and total power consumption; practical system constraints can also be non-convex depending on the system model, for instance, minimum rate constraint.

Several studies have examined *EE based power allocation in CF-mMIMO*, [9], [10], [11], [12], [13], [14], [15], [16], [17]. In [9], the authors proposed an algorithm based on successive-convex-approximation (SCA) to compute power coefficients that maximize the EE for a CF-mMIMO system with spatially uncorrelated channels and maximum-ratio (MR) precoding. In [10], the authors computed power coefficients that maximize the EE of a CF-mMIMO system by approximating the EE function to a convex function and then solved it using Dinkelbach's algorithm. This is done for a clustered channel model and zero forcing (ZF) precoder. In [11], EE-based power allocation using SCA was studied for layered-division multiplexing-based non-orthogonal multicast and unicast transmission systems, with MR precoding and a spatially uncorrelated channel model. In [12], a power allocation policy using Dinkelbach's algorithm was proposed for multi-cell mMIMO using max-min EE optimization; the authors considered Weichselberger's channel model with precoding based on statistical channel state information (CSI). Furthermore, the authors assume that UEs have access to perfect CSI. In [13], the authors proposed a joint power allocation and precoder design that maximize the EE in a distributed MIMO system that serves a single UE under the assumption of perfect CSI. In [14], the authors proposed an algorithm to compute power coefficients that maximize the EE through a block quadratic transformation for a CF-mMIMO system that jointly serves UEs and unmanned aerial vehicles (UAVs) with MR precoding. In [15], the authors proposed an iterative algorithm based on an accelerated projected gradient (APG) method to compute power coefficients that maximize the EE of a CF-mMIMO system. Further, the system model considers a spatially uncorrelated channel model and MR precoding. In [16], the authors proposed a power allocation algorithm that maximizes the ratio of the geometric mean of the UE rates (as proxy for sum-rate (SR)) and the power consumption. In [18], the authors developed a novel alternating optimization based algorithm to minimize battery energy usage in wireless-powered cell-free systems by leveraging self-energy recycling, which demonstrated significant improvement in energy efficiency and outage rates. In [19], the authors study the EE of coherent and non-coherent downlink data transmission strategies in cell-free systems under realistic fronthaul capacity constraints. The authors formulate novel EE maximization problems for both strategies and solve them using a framework that combines successive convex approximation and the Dinkelbach algorithm. In [20], the authors study the energy efficiency of cell-free massive MIMO systems with limited backhaul capacity, proposing an optimization scheme using distributed MR combining and the Bussgang theorem to model quantization errors. A non-convex energy efficiency maximization problem is solved through a successive convex approximation approach. In [21], the authors studied optimizing energy efficiency in cell-free massive MIMO systems by modeling the spatial distribution of APs as Poisson point processes. In the paper, they derived closed-form expressions for the optimal pilot reuse factor, AP density, number of antennas per AP, and number of users, providing practical design guidelines for achieving high energy and spectral efficiency. In [22], the authors investigated the uplink spectral and energy efficiency of cell-free massive MIMO systems with limited fronthaul capacity, using optimal uniform quantization. They compared different schemes for channel estimation and quantization, demonstrating that with optimal quantization strategies, these systems can achieve near-optimal performance with relatively few quantization bits. The paper concludes that cell-free massive MIMO systems are highly efficient and practical for future wireless networks even with fronthaul limitations.

The following recent studies have explored spectrum sharing in multi-cell MIMO and CF-mMIMO: [23], [24], [25], [26], [27], [28]. In [23], the authors studied a multi-objective power allocation policy that achieves max-min fairness for both PN and SN in both uplink and downlink with a spatially uncorrelated channel model and MR precoding. In [24], the authors proposed a power allocation policy in which each secondary AP allocates power to secondary UEs proportionally to their corresponding channel strength. Further, the channel is modeled to be spatially uncorrelated and MR precoding is used. In [27] the authors proposed a penalty dual decomposition-based gradient projection algorithm for maximizing achievable rates in intelligent reflecting surface assisted MIMO systems in an underlay spectrum sharing scenario, assuming perfect CSI is available. In [28], the authors considered a multi-cell multi-user setup and proposed an algorithm based on fractional programming and block coordinate descent to allocate resources (precoder design and spectrum allocation), assuming perfect CSI.

Focus and Contributions: We focus on an underlay spectrum sharing system with a secondary CF-mMIMO system operating concurrently with a primary collocated mMIMO system. Our proposed algorithm works for any precoder and further we consider spatially correlated channels, unlike [9], [10], [11], [12], [13], [14], [15], [16], [17], [23], [24], [25], [26], [27], [28] that studied only MR precoding and/or uncorrelated channels. We also consider imperfect CSI obtained using MMSE channel estimation rather than the perfect CSI assumption in [27] and [28]. For the above model, we focus on maximizing the EE of the secondary system, which is not considered in the literature to the best of our knowledge [1].

The specific contributions of this paper are as follows:

1) We develop a novel weighted-minimum-mean-square-error (WMMSE) combined with Dinkelbach's algorithm to compute power coefficients that maximize EE of the secondary CF-mMIMO system subject to an average power constraint and an interference constraint imposed by the primary mMIMO system. With this formulation, we exploit the dual relation between WMMSE and weighted SR to maximize the EE. This algorithm is also applicable to transmit precoding with spatially correlated channels.
2) For APs that employ MR precoding, we propose a minorization-maximization (MM) based Dinkelbach's algorithm that performs power allocation to maximize the EE of our secondary cell-free system. This algorithm, applicable with only MR precoding, has lower computational complexity than the WMMSE based approach, which applies to any precoder.
3) We also formulate an optimization problem with additional minimum rate or quality of service (QoS) constraints for the S-UEs. We study the impact of QoS constraints on the EE of the CF-mMIMO SN. We solve this problem using the proposed WMMSE based algorithm.
4) We also propose low-complexity heuristic power allocation policies and show through simulations that they perform better than simple equal power allocation. We also provide an algorithm that employs a projected gradient method and show through simulation that the results outperform commonly used heuristic power allocation policies.
5) In the simulations, we study the impact of different system parameters on the EE for the proposed WMMSE based and MM based power allocation algorithms. We benchmark them with the projected gradient and other heuristic power allocation methods. Our results show that the WMMSE algorithm outperforms the other methods. We also provide the results of proposed WMMSE based method with regularized-ZF (R-ZF) to illustrate the method is applicable to any precoder.
6) We provide insights into the similarities and differences between the SR and the EE maximization optimization problems. Additionally, we establish conditions under which solutions for maximizing the SR can perform similarly to directly maximizing the EE.

Items 1, 3, 4, 5, 6 are specific novel contributions over the conference paper (which was restricted to MR precoding [1]). Furthermore, our innovation lies in applying the novel WMMSE-Dinkelbach algorithm to maximize the EE in a spectrum-sharing scenario, rather than directly employing standard alternating optimization methods to tackle the non-convex EE problem. Moreover, the proposed algorithm outperforms the latter approach, where the alternating optimization method is directly used to solve the EE problem.

A. NOTATIONS

The notation $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ denotes that $\mathbf{z}$ is circularly symmetric Gaussian random vector with mean $\mathbf{0}$ and covariance $\mathbf{C}$. We denote $[M] = \{1, \ldots, M\}$. We denote $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ as mean and variance of the argument. We denote $\mathrm{Tr}(\cdot)$ as the trace of the argument.

B. PAPER OUTLINE

The remainder of the paper is organized as follows. Section II describes the system model for the primary and secondary networks in downlink. It includes the pilot sharing mechanism with the primary network, channel estimation, downlink payload transmission, interference to the primary network, and achievable rate for the secondary network. Section III presents WMMSE-Dinkelbach's algorithm, the main contribution, which maximizes the EE of the secondary network. In Section IV, we present the MM-based Dinkelbach's algorithm for computing power coefficients and several heuristic power allocation schemes. Finally, Section V presents numerical results demonstrating the proposed algorithm's performance and also discussion on incorporating a minimum rate constraint into the EE maximization problem. Appendices provide closed-form expressions useful in formulating the optimization problem and also includes a proof of convergence for the proposed algorithms.

II. SYSTEM MODEL

We consider a system where two networks share the spectrum, namely a PN that owns the spectrum license and a SN that does not own the spectrum license. The PN consists of a single mMIMO base station (P-BS) with M antennas serving K_p single-antenna primary users (P-UEs). The SN comprises a CF-mMIMO network with L secondary APs (S-APs), with N antennas at each S-AP jointly serving K_s single-antenna secondary users (S-UEs). As in [23], [24], [25], [26], [27], and [28], the two networks are assumed to be time and frequency synchronized such that they operate in the same coherence block for uplink and downlink operations. A pictorial depiction of the model is in Fig. 1.

For each S-UE, there is a channel to P-BS and also to every S-AP. We denote the channel between S-UE i and P-BS by $\mathbf{u}_{\text{sp-}i} \in \mathbb{C}^M$ and the channel from S-UE i to S-AP l by $\mathbf{h}_{il} \in \mathbb{C}^N \ \forall i \in [K_s], \ l \in [L]$. Similarly,

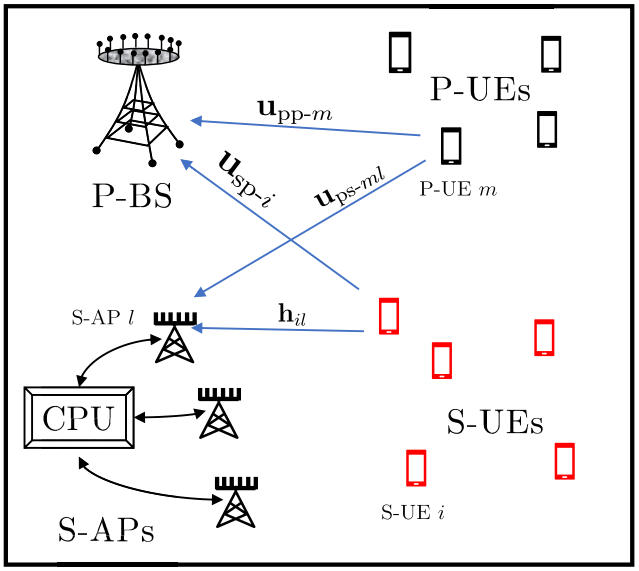


FIGURE 1. Network model.

for each P-UE there is a channel to P-BS and to every S-AP. We denote the channel between the P-UE m and P-BS by $\mathbf{u}_{\text{pp-}m} \in \mathbb{C}^M$ and the channel between P-UE m and S-AP l by $\mathbf{u}_{\text{ps-}ml} \in \mathbb{C}^N \ \forall m = 1, \ldots, K_p$. We consider a block fading channel model in which we assume the channel to be constant within a coherence block. We denote with τ_c the number of channel uses within a coherence block. In each coherence block, for both PN and SN, we consider that independent channel realizations are drawn from a correlated Rayleigh fading distribution as follows: $\mathbf{h}_{kl} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{kl})$, $\mathbf{u}_{\text{sp-}i} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_{\text{sp-}i})$, $\mathbf{u}_{\text{ps-}jl} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_{\text{ps-}jl})$ and $\mathbf{u}_{\text{pp-}m} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_{\text{pp-}m})$. The spatial correlation matrices $\mathbf{R}_{kl}$, $\mathbf{D}_{\text{sp-}i}$, $\mathbf{D}_{\text{ps-}jl}$, and $\mathbf{D}_{\text{pp-}m}$ capture the channel characteristics and large-scale fading of their corresponding channels. We assume all the channels are unknown a priori to both PN and SN. However, the channel statistics are known to both PN and SN. We consider that both PN and SN operate in time-division duplex (TDD) mode, where both networks allocate at most $\tau_p (< \tau_c)$ and $\tau_d = \tau_c - \tau_p$ channel uses for channel estimation and downlink payload transmission, respectively. In the next subsections, we will provide a detailed discussion of these two phases.

A. CHANNEL ESTIMATION

Both S-UEs and P-UEs transmit pilot symbols for the S-APs and P-BS, respectively, to estimate the corresponding channels. We assume that reciprocity of the channel holds, i.e., the estimates of channel realizations in the uplink remain the same during the data transmission phase in the downlink. We consider τ_p mutually orthogonal pilot vectors and represent them in matrix form as $\mathbf{\Phi} = \left[\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{\tau_p}\right] \in \mathbb{C}^{\tau_p \times \tau_p}$ such that $\|\boldsymbol{\phi}_t\| = \sqrt{\tau_p}$, for $t = 1, \ldots, \tau_p$. We divide these pilot vectors into three sets $\mathbf{\Phi} = \left[\mathbf{\Phi}_s, \mathbf{\Phi}_0, \mathbf{\Phi}_p\right]$ where $\mathbf{\Phi}_s \in \mathbb{C}^{\tau_p \times \tau_1}$ contains τ_1 pilot vectors that only SN uses, $\mathbf{\Phi}_0 \in \mathbb{C}^{\tau_p \times \tau_2}$ contains τ_2 pilot vectors that both PN and SN share, and $\mathbf{\Phi}_p \in \mathbb{C}^{\tau_p \times \tau_3}$ contains τ_3 pilot vectors that only PN uses, with $\tau_1 + \tau_2 + \tau_3 = \tau_p$.

To estimate the channels in SN, S-APs employ a minimum-mean-square-error (MMSE) estimator. Let $t_k \in \{1, \ldots, \tau_1 + \tau_2\}$ denote the index of the pilot vector that S-UE k uses. Furthermore, we consider $\mathcal{S}_k^s$ and $\mathcal{S}_k^p$ to be the subsets of S-UE indices, $\{1, \ldots, K_s\}$, and P-UE indices, $\{1, \ldots, K_p\}$, respectively, that share the same pilot vector as that of S-UE k. The signal that S-AP l receives in the first τ_p channel uses is

$$\mathbf{Y}_l^p = \sum_{i=1}^{K_s} \sqrt{\eta_{s\text{-}i}} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^H + \sum_{j=1}^{K_p} \sqrt{\eta_{p\text{-}j}} \mathbf{u}_{\text{ps-}jl} \boldsymbol{\phi}_{t_j}^H + \mathbf{N}_l, \quad (1)$$

where $\eta_{s\text{-}i} \geq 0$ and $\eta_{p\text{-}j} \geq 0$ are the power coefficients that S-UE i and P-UE j, respectively, assigns to their corresponding pilot signals, and $\mathbf{N}_l \in \mathbb{C}^{N \times \tau_p}$ is the noise at S-AP l with independent and identically distributed (i.i.d) entries drawn from $\mathcal{CN}(0, \varsigma^2)$ with ς^2 being the noise power.

The MMSE estimation of $\mathbf{h}_{kl}$ involves two steps. The first step is to compute a sufficient statistics by projecting the received pilot signal at AP l onto the pilot direction of S-UE k. The projected signal at AP l is $\mathbf{y}_{t_k l}^p$, given by

$$\begin{aligned} \mathbf{y}_{t_k l}^p &= \mathbf{Y}_l^p \boldsymbol{\phi}_{t_k} / \sqrt{\tau_p} \\ &= \sum_{i \in \mathcal{S}_k^s} \sqrt{\eta_{s\text{-}i} \tau_p} \mathbf{h}_{il} + \sum_{j \in \mathcal{S}_k^p} \sqrt{\eta_{p\text{-}j} \tau_p} \mathbf{u}_{\text{ps-}jl} + \mathbf{n}_{t_k l}, \end{aligned} \quad (2)$$

where $\mathbf{n}_{t_k l} = \mathbf{N}_l \boldsymbol{\phi}_{t_k}^H / \sqrt{\tau_p} \sim \mathcal{CN}(\mathbf{0}, \varsigma^2 \mathbf{I}_N)$. We note that the projected signal in (2) is a sufficient statistic of $\mathbf{h}_{il}$ $\forall i \in [K_s], l \in [L]$ as the pilots are orthogonal and $\mathbf{N}_l$ has i.i.d. Gaussian entries [29]. The second step is to compute the MMSE channel estimate $\widehat{\mathbf{h}}_{kl}$ using the projected received pilot signal $\mathbf{y}_{t_k l}^p$:

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\eta_{s\text{-}k} \tau_p} \mathbf{R}_{kl} \mathbf{\Psi}_{t_k l}^{-1} \mathbf{y}_{t_k l}^p, \quad (3)$$

where

$$\mathbf{\Psi}_{t_k l} = \sum_{i \in \mathcal{S}_k^s} \eta_{s\text{-}i} \tau_p \mathbf{R}_{il} + \sum_{j \in \mathcal{S}_k^p} \eta_{p\text{-}j} \tau_p \mathbf{D}_{\text{ps-}jl} + \varsigma^2 \mathbf{I}_N. \quad (4)$$

The channel estimate $\widehat{\mathbf{h}}_{kl}$ and channel estimation error $\widetilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \widehat{\mathbf{h}}_{kl}$ have Gaussian distributions, i.e., $\widehat{\mathbf{h}}_{kl} \sim \mathcal{CN}(\mathbf{0}, \widehat{\mathbf{R}}_{kl})$ and $\widetilde{\mathbf{h}}_{kl} \sim \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}}_{kl})$, where $\widehat{\mathbf{R}}_{kl} = \eta_{s\text{-}k} \tau_p \mathbf{R}_{kl} \mathbf{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl}$ and $\widetilde{\mathbf{R}}_{kl} = \mathbf{R}_{kl} - \widehat{\mathbf{R}}_{kl}$. From estimation theory, we know that $\widehat{\mathbf{h}}_{kl}$ and $\widetilde{\mathbf{h}}_{kl}$ are independent [30]. Furthermore, the channel estimates of S-UE k and S-UE i that use the same pilot vectors are in general correlated and their cross-correlation matrix is given by [31]

$$\breve{\mathbf{R}}_{kil} = \mathbb{E}\left\{\widehat{\mathbf{h}}_{kl} \widehat{\mathbf{h}}_{il}^H\right\} = \sqrt{\eta_{s\text{-}k} \eta_{s\text{-}i}} \tau_p \mathbf{R}_{kl} \mathbf{\Psi}_{t_k l}^{-1} \mathbf{R}_{il}, \ i \in \mathcal{S}_k^s. \quad (5)$$

B. DOWNLINK PAYLOAD TRANSMISSION

In the downlink, using the known channel statistics and channel estimates obtained in (3), S-AP l constructs the

precoders $\mathbf{w}_{il} \in \mathbb{C}^N$ for $\forall i \in [K_s]$, to transmit data. The transmit signal at each S-AP is given by

$$\mathbf{x}_l = \sum_{i=1}^{K_s} \sqrt{p_{il}} \mathbf{w}_{il} q_i, \quad \forall l \in [L], \tag{6}$$

where $p_{il} \geq 0$ is the transmit power that S-AP l assigns to transmit message symbol q_i of the S-UE i. We analyze a general precoder $\mathbf{w}_{il}$ here, and specialize to the MR precoder in Section IV. We assume that transmit symbols of different S-UEs are uncorrelated. Each S-AP has a maximum power budget, $P_{\max}$. The transmit symbols (with zero mean) and precoders have average unit power i.e., $\mathbb{E}\left\{|q_i|^2\right\} = 1$ and $\mathbb{E}\left\{\|\mathbf{w}_{il}\|^2\right\} = 1$, $\forall i \in [K_s], l \in [L]$. Therefore, from (6) the total transmit power at S-AP l is $\sum_{i=1}^{K_s} p_{il}$. The transmit power constraint, which limits the total transmit power at each AP to $P_{\max}$, can be written as

$$\sum_{i=1}^{K_s} p_{il} \leq P_{\max}, \ \forall l \in [L]. \tag{7}$$

The signal that S-UE k receives is given by

$$\begin{aligned} y_k &= \sum_{l=1}^{L} \mathbf{h}_{kl}^H \mathbf{x}_l + \mathbf{u}_{\text{sp-}k}^H \mathbf{s}_p + n_k \\ &= \sum_{l=1}^{L} \sqrt{p_{kl}} \mathbf{h}_{kl}^H \mathbf{w}_{kl} q_k + \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \sqrt{p_{il}} \mathbf{h}_{kl}^H \mathbf{w}_{il} q_i + d_k \\ &= \sum_{l=1}^{L} \sqrt{p_{kl}} g_{k_k l} q_k + \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \sqrt{p_{il}} g_{k_i l} q_i + d_k, \end{aligned} \tag{8}$$

where $\mathbf{s}_p$ is the transmit signal from P-BS to P-UEs with zero mean and covariance $\bar{\mathbf{D}}_{\text{p}}$, and n_k is the receiver noise at S-UE k with variance ς^2. Also, $d_k = \mathbf{u}_{\text{sp-}k}^H \mathbf{s}_p + n_k$ denotes the interference-plus-noise at the receiver, which has zero mean and variance $\sigma_k^2 = \text{Tr}\left(\bar{\mathbf{D}}_{\text{p}} \mathbf{D}_{\text{sp-}k}\right) + \varsigma^2$, and $g_{k_i l} = \mathbf{h}_{kl}^H \mathbf{w}_{il}$ is the effective channel that q_i sees when S-AP l transmits and S-UE k receives the signal.

However, the signal intended for S-UEs will be interfering with the signals to P-UEs. The interference signal y_m^{ps} that P-UE m receives is given by

$$y_m^{ps} = \sum_{l=1}^{L} \mathbf{u}_{\text{ps-}ml}^H \mathbf{x}_l. \tag{9}$$

Therefore, the S-APs should ensure that the average interference power is less than a threshold I_T i.e.,

$$\begin{aligned} &\mathbb{E}\left\{|y_m^{ps}|^2\right\} \\ &= \mathbb{E}\left\{\sum_{l=1}^{L}\sum_{l'=1}^{L}\sum_{i=1}^{K_s} \sqrt{p}_{il}\sqrt{p}_{il'} \mathbf{u}_{\text{ps-}ml}^H \mathbf{w}_{il} \mathbf{w}_{il'}^H \mathbf{u}_{\text{ps-}ml'} |q_i|^2\right\} \\ &= \sum_{l=1}^{L}\sum_{i=1}^{K_s} p_{il} \text{Tr}\left(\mathbb{E}\left\{\mathbf{w}_{il}\mathbf{w}_{il}^H\right\} \mathbb{E}\left\{\mathbf{u}_{\text{ps-}ml}\mathbf{u}_{\text{ps-}ml}^H\right\}\right) \leq I_T. \end{aligned} \tag{10}$$

In (10), we assumed that the S-APs precoders are independent of the channels between P-UEs and S-APs. Furthermore, we assumed that the channels between the same P-UE and different S-APs are uncorrelated.

C. ACHIEVABLE SPECTRAL EFFICIENCY

After receiving the signal (8), S-UE k decodes the signal, q_k, intended for it. We assume that all the S-UEs know the statistics of the effective channels. In order to apply the use-and-then-forget bound (also known as the hardening bound) [31], [32], we rewrite (8) in terms of known statistics as follows

$$\begin{aligned} y_k = &\sum_{l=1}^{L} \sqrt{p_{kl}} \mathbb{E}\left\{g_{k_k l}\right\} q_k + \sum_{\substack{i=1 \\ i \neq k}}^{K_s} \left(\sum_{l=1}^{L} \sqrt{p_{il}} g_{k_i l}\right) q_i \\ &+ \left(\sum_{l=1}^{L} \sqrt{p_{kl}} g_{k_k l} - \sum_{l=1}^{L} \sqrt{p_{kl}} \mathbb{E}\left\{g_{k_k l}\right\}\right) q_k + d_k. \end{aligned} \tag{11}$$

The achievable SE for S-UE k using the hardening bound is given by

$$\text{SE}_k = \frac{\tau_d}{\tau_c} \log_2(1 + \Gamma_k) \text{ bit/s/Hz}, \tag{12}$$

where Γ_k is the effective signal-to-interference-plus-noise-ratio (SINR) at S-UE k given by

$$\Gamma_k = \frac{|\text{DK}_k|^2}{\mathbb{E}\{|\text{DU}_k|^2\} + \mathbb{E}\{|\text{UI}_k|^2\} + \sigma_k^2}, \tag{13}$$

where

$$\text{DK}_k = \sum_{l=1}^{L} \sqrt{p_{kl}} \mathbb{E}\left\{g_{k_k l}\right\} \tag{14}$$

$$\text{DU}_k = \sum_{l=1}^{L} \sqrt{p_{kl}} g_{k_k l} q_k - \sum_{l=1}^{L} \sqrt{p_{kl}} \mathbb{E}\left\{g_{k_k l}\right\} q_k, \tag{15}$$

$$\text{UI}_k = \sum_{\substack{i=1 \\ i \neq k}}^{K_s} \left(\sum_{l=1}^{L} \sqrt{p_{il}} g_{k_i l}\right) q_i. \tag{16}$$

The above notations DU_k, UI_k can be interpreted as desired symbol q_k transmitted over unknown channel and inter-user interference from other S-UEs, respectively. σ_k^2 is the effective primary interference-plus-noise power at S-UE k. We assume that knowledge of only channel statistics is available at the S-UEs, which change at a much slower timescale than the instantaneous channel gains. These can obtained by the SN by exploiting the synchronous TDD operation of primary and secondary systems along with the known pilot sequences allotted to the primary UEs [33].

III. POWER ALLOCATION TO MAXIMIZE EE

We now introduce the mathematical formulation of our problem statement, and develop an efficient power allocation algorithm that maximizes the EE of an underlay CF-mMIMO system with any transmit precoding subject

to an average interference constraint imposed by a primary mMIMO system. Among different definitions of EE in the literature [34], we use the ratio of downlink throughput to total power consumed at all the S-APs as the definition of EE.[1] Mathematically, the EE can be written in units of [bit/J] as

$$\text{EE} = \frac{\sum_{i=1}^{K_s} B \cdot \text{SE}_i}{P_T}, \tag{17}$$

where B is the bandwidth of the SN and P_T is the total power consumed by all the S-APs during downlink transmission, which is given by [9] and [29]

$$P_T = \zeta \sum_{l=1}^{L}\sum_{i=1}^{K_s} p_{il} + \xi \sum_{i=1}^{K_s} B \cdot \text{SE}_i + \text{PC}, \tag{18}$$

where ζ is the inverse of the power amplifier efficiency, ξ is a front-haul power consumption factor that depends on the throughput, which has the unit Ws/bit or W/(bit/s) (Watt per unit rate), and PC is the total power that approximately accounts for the power consumption of the transceiver chains, channel estimation process, channel coding and decoding units, and linear processing at the S-APs.

We compute the power coefficients[2] that maximize the EE in (17) subject to the power constraint in (7) at each S-AP, and to the average interference constraint in (10) at each P-UE:

$$\underset{\mathbf{p}}{\text{maximize}} \quad \frac{\sum_{i=1}^{K_s} B \cdot \text{SE}_i(\mathbf{p})}{P_T(\mathbf{p})} \tag{19a}$$

$$\text{s.t.} \quad \sum_{i=1}^{K_s} p_{il} \leq P_{\max}, \ \forall l \in [L], \tag{19b}$$

$$p_{il} \geq 0, \ \forall i \in [K_s]; \forall l \in [L], \tag{19c}$$

$$\sum_{l=1}^{L}\sum_{i=1}^{K_s} p_{il} b_{i_m l} \leq I_T, \ \forall m \in [K_p], \tag{19d}$$

where $\mathbf{p} = [p_{11}, \ldots, p_{1L}, \ldots, p_{il}, \ldots, p_{K_s 1}, \ldots, p_{K_s L}]^T$ and $b_{i_m l} = \text{Tr}\left(\mathbb{E}\{\mathbf{w}_{il}\mathbf{w}_{il}^H\}\mathbb{E}\{\mathbf{u}_{\text{ps-}ml}\mathbf{u}_{\text{ps-}ml}^H\}\right)$. The problem in (19) is non-convex due to the non-convexity of the objective, although the constraints are convex. Besides the non-convexity of the problem, the total power consumption in (18) depends on the throughput, which further complicates the problem at hand. We note that the power allocation policy that maximizes (19) remains same when the throughput-dependent power term, $P_T(\mathbf{p})$, is replaced by a throughput-independent power term $\bar{P}_T(\mathbf{p})$

$$\bar{P}_T(\mathbf{p}) = \zeta \sum_{l=1}^{L}\sum_{i=1}^{K_s} p_{il} + \text{PC}. \tag{20}$$

A proof for this observation follows similar steps as described in [9, Appendix B]. However, the problem of non-convexity remains. To address it, we reformulate the problem in (19) and develop an algorithm based on WMMSE [36] and Dinkelbach's algorithm [37], [38].

A. REFORMULATION OF THE OPTIMIZATION PROBLEM

We introduce the following new notation to reformulate the problem in (19): $c_{il} = \sqrt{p_{il}}, \forall i \in [K_s], l \in [L]$. We further define the following notation: $\mathbf{c}_i = [c_{i1}, \ldots, c_{iL}]^T$ as the vector of power coefficients of S-UE i from all S-APs, $\mathbf{c} = [\mathbf{c}_1^T, \ldots, \mathbf{c}_{K_s}^T]^T$ as the vector of all the power coefficients of all S-UEs from all S-APs, and $\mathbf{g}_{ki} = [g_{k_i 1}, \ldots, g_{k_i L}]^T$ as the vector of effective channels of all S-APs whose lth entry corresponds to the channel component between $\mathbf{h}_{kl}$ and the precoder $\mathbf{w}_{il}$. Since there is dependency on both indices, we have adapted the above notation for convenience. We also let $\bar{\mathbf{g}}_{ki} = \mathbb{E}\{\mathbf{g}_{ki}\}$ denote the mean of the effective channel vector, $\mathbf{g}_{ki}^{\circ} = \mathbf{g}_{ki} - \bar{\mathbf{g}}_{ki}$ denote the error in the known effective channel vector, $\mathbf{A}_{kj} = \mathbb{E}\left\{\mathbf{g}_{kj}\mathbf{g}_{kj}^H\right\}$ denote the correlation matrix of the effective channel vector, and $\mathbf{b}_{im} = [b_{i_m l}, \ldots, b_{i_m L}]^T$ denote the vector of statistics of the interference seen by P-UE m due to the signal sent to the S-UE i from all the S-APs.

Using these notations, the received signal at S-UE k given in (11) can be written as follows:

$$y_k = \mathbf{c}_k^T \bar{\mathbf{g}}_{kk} q_k + \mathbf{c}_k^T \mathbf{g}_{kk}^{\circ} q_k + \sum_{i=1, i\neq k}^{K_s} \mathbf{c}_i^T \mathbf{g}_{ki} q_i + d_k. \tag{21}$$

Similarly, the effective SINR of S-UE k in (13) can be rewritten as

$$\begin{aligned}\Gamma_k(\mathbf{c}) &= \frac{|\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2}{\sum_{i=1, i\neq k}^{K_s} \mathbb{E}\left\{|\mathbf{c}_i^T \mathbf{g}_{ki}|^2\right\} + \mathbb{E}\left\{|\mathbf{c}_k^T \mathbf{g}_{kk}^{\circ}|^2\right\} + \sigma_k^2}, \\ &= \frac{|\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2}{\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki} \mathbf{c}_i - |\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2 + \sigma_k^2}.\end{aligned} \tag{22}$$

Finally, the optimization problem in (19) can be restated as follows in terms of the optimization vector $\mathbf{c}$:

$$\underset{\mathbf{c}}{\text{maximize}} \quad \frac{\frac{\tau_d B}{\tau_c \ln 2} \sum_{k=1}^{K_s} \ln\left(1 + \Gamma_k(\mathbf{c})\right)}{\zeta \|\mathbf{c}\|^2 + \text{PC}} \tag{23a}$$

$$\text{s.t.} \quad \sum_{i=1}^{K_s} c_{il}^2 \leq P_{\max}, \ \forall l \in [L], \tag{23b}$$

$$c_{il} \geq 0, \ \forall i \in [K_s]; \forall l \in [L], \tag{23c}$$

$$\sum_{l=1}^{L}\sum_{i=1}^{K_s} c_{il}^2 b_{i_m l} \leq I_T, \ \forall m \in [K_p]. \tag{23d}$$

[1] We want to remark that in mMIMO systems the gap between the AWGN Shannon capacity and that of a massive MIMO system with enough channel hardening and favorable propagation is very small. Moreover, with practical (adaptive) modulation and coding schemes one can achieve rates close to Shannon capacity [35] for given signal-to-noise ratio. Consequently, the definition of EE in our context will be a reasonable proxy for what can be achieved in a mMIMO system.

[2] Alternatively, one can optimize the precoders. However, the improved performance comes at the expense of increased complexity. Therefore, we focus on optimizing only power coefficients, which has lower complexity [31].

The optimization problem in (23) is still non-convex due to the non-convexity of the objective function. To deal with this non-convexity, we will adopt the methodology of Dinkelbach's algorithm [38]. For the discussions to follow we ignore the constant terms $\frac{\tau_d B}{\tau_c \ln 2}$ as these terms do not affect the optimal power coefficients. We study the impact of additional QoS constraints of (minimum S-UE rates) on EE in Section V-C.

B. WMMSE-BASED DINKELBACH'S ALGORITHM

In this section, we propose an iterative algorithm based on Dinkelbach's approach to solve the optimization problem in (23). In each iteration of Dinkelbach's algorithm, the sub-problem to be solved is non-convex. To tackle the non-convexity, we reformulate the non-convex objective function by making use of the duality between the SE and the WMMSE. This leads to a three-step iterative coordinate ascent algorithm within each iteration of Dinkelbach's algorithm.

1) DINKELBACH'S SUBPROBLEM

In order to use the Dinkelbach's approach for the fractional objective function in (23a), we define

$$f_1(\mathbf{c}) = \sum_{k=1}^{K_s} \ln\left(1+\Gamma_k(\mathbf{c})\right), \qquad f_2(\mathbf{c}) = \zeta \left\|\mathbf{c}\right\|^2 + \text{PC}, \tag{24}$$

and define the function that we need for Dinkelbach's subproblem,

$$F(\mathbf{c};\lambda_n) = f_1(\mathbf{c}) - \lambda_n f_2(\mathbf{c}), \tag{25}$$

where λ_n is the Dinkelbach's parameter in iteration n. Then for a given λ_n, we define the subproblem $\mathcal{P}_n$,

$$\begin{aligned} \mathcal{P}_n: \quad & \underset{\mathbf{c}}{\text{maximize}} \quad F(\mathbf{c};\lambda_n) \\ & \text{s.t.} \qquad (23b), (23c), \text{ and } (23d). \end{aligned} \tag{26}$$

Note that the function $f_2(\mathbf{c})$ in (25) is convex in $\mathbf{c}$ (ζ is positive constant). However, the function $f_1(\mathbf{c})$ is not concave. To tackle this, we resort to duality between sum SE and WMMSE.

2) REFORMULATION OF $F(\mathbf{C};\lambda_N)$ IN TERMS OF WMMSE

Given a linear combiner μ_k of the intended data symbol q_k at S-UE k, the MSE is

$$\begin{aligned} & e_k(\mu_k,\mathbf{c}) \\ &= \mathbb{E}\left\{\left|\mu_k^* y_k - q_k\right|^2\right\} \\ &= |\mu_k|^2\left(\sum_{i=1}^{K} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i + \sigma_k^2\right) - 2\mathcal{R}e\left\{\mu_k(\bar{\mathbf{g}}_{kk}^H)\right\}\mathbf{c}_k + 1. \end{aligned} \tag{27}$$

We can compute the linear-MMSE (LMMSE) estimator μ_k^L in closed form:

$$\mu_k^L = \mathbf{c}_k^T \bar{\mathbf{g}}_{kk}\left(\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i + \sigma_k^2\right)^{-1}, \forall k \in [K_s]. \tag{28}$$

The corresponding MMSE $e_k^{\min}(\mathbf{c})$ is

$$\begin{aligned} e_k^{\min}(\mathbf{c}) &= 1 - \frac{|\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2}{\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i + \sigma_k^2} \\ &= \frac{\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i - |\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2 + \sigma_k^2}{\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i - |\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2 + \sigma_k^2 + |\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2} \\ &= \left(1 + \frac{|\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2}{\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i - |\mathbf{c}_k^T \bar{\mathbf{g}}_{kk}|^2 + \sigma_k^2}\right)^{-1} \\ &= \left(1+\Gamma_k(\mathbf{c})\right)^{-1}. \end{aligned} \tag{29}$$

The last step in (29) establishes the exact relation between the MMSE and the SINR of S-UE k. Therefore, solving the following WMMSE problem is equivalent to maximizing the SE:

$$\underset{\mathbf{c},v_k,\mu_k}{\text{maximize}} \quad -v_k e_k(\mu_k,\mathbf{c}) + \ln v_k. \tag{30}$$

This can be seen as follows: the optimal $\{\mu_k\}$ for given $\mathbf{c}$, $\{v_k\}$ are given in (28), the optimal $v_k = 1/e_k$ for given $\mathbf{c}$, $\{\mu_k\}$ and substituting the optimal v_k and μ_k, the problem in (30) reduces to solving

$$\underset{\mathbf{c}}{\text{maximize}} \quad \ln\left(e_k^{\min}(\mathbf{c})^{-1}\right), \tag{31}$$

which is equivalent to maximizing the SE. We now extend this framework and generalize it to solve the subproblem in (26)

$$\underset{\mathbf{c},\mathbf{v},\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{i=1}^{K_s} -v_i e_i(\mu_i,\mathbf{c}) + \ln v_i - \lambda_n\left(\zeta\left\|\mathbf{c}\right\|^2 + \text{PC}\right) \tag{32a}$$

$$\text{s.t.} \quad \sum_{i=1}^{K_s} c_{il}^2 \le P_{\max}, \ \forall l \in [L], \tag{32b}$$

$$c_{il} \ge 0, \ \forall i \in [K_s]; \forall l \in [L], \tag{32c}$$

$$\sum_{l=1}^{L}\sum_{i=1}^{K_s} c_{il}^2 b_{i_m l} \le I_T, \ \forall m \in [K_p], \tag{32d}$$

where $\mathbf{v} = [v_1,\ldots,v_{K_s}]$ and $\boldsymbol{\mu} = [\mu_1,\ldots,\mu_{K_s}]$.

3) COORDINATE ASCENT ALGORITHM

The advantage of the formulation in (32) is that the problem is now convex in each of the optimization variables when the other two variables are fixed. So the steps involved in solving the problem are:

1) The optimal solution of $\boldsymbol{\mu}$ for a given $\mathbf{c}$ and $\mathbf{v}$ is

$$\mu_k = \mathbf{c}_k^T \bar{\mathbf{g}}_{kk}\left(\sum_{i=1}^{K_s} \mathbf{c}_i^T \mathbf{A}_{ki}\mathbf{c}_i + \sigma_k^2\right)^{-1}, \forall k \in [K_s]. \tag{33}$$

2) The optimal solution of $\mathbf{v}$ for given $\mathbf{c}$ and $\boldsymbol{\mu}$ is,

$$v_k = e_k(\mu_k, \mathbf{c})^{-1}, \forall k \in [K_s]. \tag{34}$$

3) To obtain the solution of $\mathbf{c}$ for a given $\boldsymbol{\mu}$ and $\mathbf{v}$, we solve a quadratically-constrained-quadratic-programming (QCQP) problem. In order to formulate the QCQP problem, we let $\mathbf{Q}_0 \in \mathbb{C}^{K_sL \times K_sL}$, $\bar{\mathbf{Q}}_l \in \mathbb{C}^{K_sL \times K_sL}$ and $\widetilde{\mathbf{Q}}_m \in \mathbb{C}^{K_sL \times K_sL}$ be block diagonal matrices. We define the k^{th} block of these matrices as follows:

$$\begin{aligned}
[\mathbf{Q}_0]_{kk} &= \sum_{j=1}^{K_s} v_j |\mu_j|^2 \mathbf{A}_{jk} + \lambda_n \zeta \mathbf{I}, \\
[\bar{\mathbf{Q}}_l]_{kk} &= \text{diag}(0, \cdots, 0, 1, 0, \ldots, 0), \\
[\widetilde{\mathbf{Q}}_m]_{kk} &= \text{diag}(b_{k_m 1}, \cdots, b_{k_m L}),
\end{aligned} \tag{35}$$

where $\text{diag}(\cdot)$ is a diagonal matrix whose diagonal entries are the elements of the argument, and $[\bar{\mathbf{Q}}_l]_{kk}$ is a diagonal matrix with only one nonzero entry at index (l, l). Also, let $\mathbf{r}_0 \in \mathbb{R}^{K_sL}$ be a block vector whose k^{th} block entry equals $v_k \mathcal{R}\{\mu_k \bar{\mathbf{g}}_{kk}^*\}$. Then the solution of $\mathbf{c}$ for a given $\boldsymbol{\mu}$ and $\mathbf{v}$ is obtained as the solution of the following QCQP problem:

$$\underset{\mathbf{c}}{\text{maximize}} \quad -\mathbf{c}^T \mathbf{Q}_0 \mathbf{c} + 2\mathbf{r}_0^T \mathbf{c} \tag{36a}$$

$$\text{s.t.} \quad \mathbf{c}^T \bar{\mathbf{Q}}_l \mathbf{c} \le P_{\max}, \ \forall l \in [L], \tag{36b}$$

$$c_{li} \ge 0, \ \forall i \in [K_s]; \forall l \in [L], \tag{36c}$$

$$\mathbf{c}^T \widetilde{\mathbf{Q}}_m \mathbf{c} \le I_T, \ \forall m \in [K_p]. \tag{36d}$$

We can solve the QCQP problem in (36) efficiently using convex optimization algorithms, for example, an interior-point method.

The overall WMMSE-based Dinkelbach's algorithm to solve (32) is given in Algorithm 1. We note that the objective function in (32) converges as it produces a sequence of increasing function values with each iteration and the objective is upper bounded. For proof of the convergence, refer to Appendix B.

Algorithm 1 Sequential WMMSE-Dinkelbach's Algorithm

Initialize $\epsilon > 0$; $n = 0$; $\lambda_n = \frac{f_1(\mathbf{c}_n)}{f_2(\mathbf{c}_n)}$; where $\mathbf{c}_n$ is any feasible point.

Repeat

(i). For given λ_n, solve the subproblem $\mathcal{P}_n$ in (32) iteratively using (33), (34), and (36), until convergence;
Let the solution be $\mathbf{c}_n^*$;

(ii). $\lambda_{(n+1)} = \dfrac{f_1(\mathbf{c}_n^*)}{f_2(\mathbf{c}_n^*)}$;

(iii). n = n+1;

Until $F(\mathbf{c}_n^*; \lambda_n) \le \epsilon$

The Dinkelbach's WMMSE algorithm can be applied to any precoder. For MR precoding, $\mathbf{w}_{il} = \widehat{\mathbf{h}}_{il}/\sqrt{\mathbb{E}\{\|\widehat{\mathbf{h}}_{il}\|^2\}}$, we can compute the entries of the matrix $\{\mathbf{A}_{ki}\}, \forall k, i \in [K_s]$ in closed form; the details are in Appendix A. Furthermore, the interference coefficients are given by $b_{i_m l} = \text{Tr}(\widehat{\mathbf{R}}_{il}\mathbf{D}_{\text{ps-}ml})/\text{Tr}(\widehat{\mathbf{R}}_{il})$.

IV. POWER CONTROL FOR MR PRECODING AND OTHER ALTERNATIVE METHODS

In this section, we focus on MR pecoding and develop a power allocation algorithm with lower complexity. We also introduce the heuristic power allocation policies. Finally, we also provide the computational complexity of all the algorithms.

A. POWER CONTROL FOR MR PRECODING

We start by an algorithm that is specifically applicable to MR precoding (this algorithm was also described in [1]). We define the following two quantities:

$$\begin{aligned}
a_{kil} &= \mathbb{E}\left\{\mathbf{h}_{kl}^H \mathbf{w}_{il}\right\} = \frac{\text{Tr}(\widehat{\mathbf{R}}_{il}\mathbf{R}_{kl})}{\text{Tr}(\widehat{\mathbf{R}}_{il})}, \\
b_{kil} &= \mathbb{V}\left\{\mathbf{h}_{kl}^H \mathbf{w}_{il}\right\} = \frac{\text{Tr}(\breve{\mathbf{R}}_{ikl})}{\sqrt{\text{Tr}(\widehat{\mathbf{R}}_{il})}},
\end{aligned} \tag{37}$$

where $\breve{\mathbf{R}}_{ikl} = \mathbb{E}\left\{\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{kl}^H\right\}$. Using this notation, we can rewrite the numerator of the EE i.e, the sum SE, as the difference between two concave functions, i.e., $\sum_{k=1}^{K_s} \text{SE}_k = f_{1a}(\mathbf{p}) - f_{1b}(\mathbf{p})$ where

$$f_{1a}(\mathbf{p}) = \sum_{k=1}^{K_s} \log_2 \left[\sum_{i=1}^{K_s} \sum_{l=1}^{L} p_{il} b_{kil} + \sum_{i=1}^{K_s} \left| \sum_{l=1}^{L} \sqrt{p_{il}} a_{kil} \right|^2 + \sigma_k^2 \right], \tag{38}$$

and

$$f_{1b}(\mathbf{p}) = \sum_{k=1}^{K_s} \log_2 \left[\sum_{i=1}^{K_s} \sum_{l=1}^{L} p_{il} b_{kil} + \sum_{i \ne k}^{K_s} \left| \sum_{l=1}^{L} \sqrt{p_{il}} a_{kil} \right|^2 + \sigma_k^2 \right]. \tag{39}$$

We note that the sum SE is not a concave function of $\mathbf{p}$; however, the functions f_{1a} and f_{1b} are concave for MR precoding. To approximate the sum SE with a concave function, we start by upper bounding $f_{1b}(\mathbf{p})$ by the function $\bar{f}_{1b}(\mathbf{p})$ to obtain a concave lower bound on the sum SE:

$$\sum_{k=1}^{K_s} \text{SE}_k = \ f_{1a}(\mathbf{p}) - f_{1b}(\mathbf{p}) \ge f_{1a}(\mathbf{p}) - \bar{f}_{1b}(\mathbf{p}; \mathbf{p}_0), \tag{40}$$

where $\bar{f}_{1b}(\mathbf{p})$ is the first-order Taylor series expansion of $f_{1b}(\mathbf{p})$ evaluated at the point $\mathbf{p}_0$.

Now to solve the modified EE problem, we define the following two functions required to define Dinkelbach's subproblem:

$$\begin{aligned}
f_1(\mathbf{p}; \mathbf{p}_0) &= f_{1a}(\mathbf{p}) - \bar{f}_{1b}(\mathbf{p}; \mathbf{p}_0), \\
f_2(\mathbf{p}) &= \zeta \sum_{i=1}^{K_s} \sum_{l=1}^{L} p_{il} + \text{PC}.
\end{aligned} \tag{41}$$

Thus, the function that we need for Dinkelbach's subproblem is:

$$F(\mathbf{p}; \mathbf{p}_0, \lambda_n) = f_1(\mathbf{p}; \mathbf{p}_0) - \lambda_n f_2(\mathbf{p}). \tag{42}$$

In the case of MR precoding, the function (42) is concave [1], however, this is not true in general for other precoders. The overall optimization problem that we are solving is to maximize a lower bound on the EE (42). In this approach, we propose an algorithm that uses the MM approach to obtain a concave lower bound on the sum SE, and then uses Dinkelbach's approach to solve a fractional problem. Therefore, we will call the algorithm in this paper the MM-based Dinkelbach's algorithm.

Using a similar argument as for the WMMSE-based Dinkelbach approach, we conclude that with the MM-based Dinkelbach algorithm the objective of the optimization problem converges.

B. HEURISTIC METHODS

In this section, we present some alternative heuristic power allocation policies. First, we note that all constraints in (19) are linear. We represent the linear constraint in matrix/vector notation as $\mathbf{Tp} \leq \mathbf{s}$, where $\mathbf{T}$ and $\mathbf{s}$ contain the scaling coefficients and thresholds, respectively. We now present three different methods.

1) POWER ADAPTATION USING PROJECTION/SCALING

We can adapt any existing heuristic power allocation policy to the system model discussed in this paper using a projection or scaling approach. One way to implement a projection-based method is by projecting the power coefficient vector onto the feasible set defined by the constraints in (19). Mathematically, this projection for some given $\bar{\mathbf{p}}$ can be expressed as:

$$\mathbf{p} = \underset{\mathbf{p}}{\operatorname{argmin}} \quad \|\mathbf{p} - \bar{\mathbf{p}}\|^2, \text{ s.t.} \quad \mathbf{Tp} \leq \mathbf{s}. \tag{43}$$

Alternatively, we can adopt a scaling-based power allocation policy within the system model. This approach involves iteratively scaling the power coefficient vector with a scaling factor $0 < \eta < 1$ to meet the constraints, that is, $\mathbf{Tp} \leq \mathbf{s}$. This can be achieved by iterating the following assignment:

$$\mathbf{p} := \eta \mathbf{p}. \tag{44}$$

The scaling process is repeated until the resulting power coefficient vector satisfies the given constraints. This approach enables the incorporation of any existing heuristic power allocation policy into the system model, without having to solve any optimization problem.

2) PROJECTED GRADIENT METHOD FOR MAXIMIZING EE

In this section, we introduce the projected gradient ascent (PGA) method for solving the non-linear EE problem in (19). PGA is an iterative algorithm designed to solve (19). In iteration k, the following steps are performed:

1) Compute an intermediate update for the power coefficient vector as:

$$\bar{\mathbf{p}}_k = \mathbf{p}_{k-1} + \alpha_k \nabla \text{EE}(\mathbf{p}_{k-1}). \tag{45}$$

2) To compute the power coefficient vector $\mathbf{p}_k$ we project onto the constraint set by solving the following optimization problem:

$$\mathbf{p}_k = \underset{\mathbf{p}}{\operatorname{argmin}} \quad \|\mathbf{p} - \bar{\mathbf{p}}_k\|^2, \text{ s.t.} \quad \mathbf{Tp} \leq \mathbf{s}. \tag{46}$$

In the above equations, $\mathbf{p}_k$ represents the power coefficient vector at iteration k. The initial point $\bar{\mathbf{p}}_0$ for the algorithm is chosen from the feasible set. The step size α_k is determined using a backtracking algorithm [39]. The gradient of EE($\mathbf{p}$) evaluated at the point $\bar{\mathbf{p}}_k$ is denoted by ∇EE($\bar{\mathbf{p}}_k$). The optimization problem in (46) aims to select the power coefficient vector from the feasible set that is closest in Euclidean distance to the power vector obtained using the gradient ascent update in (45). We repeat these steps until the objective converges.

Remark on complexity: Solving the convex optimization problems in (23) and (42) is the major contributor to the complexity of the proposed WMMSE based and MM methods. When interior point method is employed the complexity is of the order of $\mathcal{O}(\gamma_1^{0.5}(\gamma_1 + \gamma_2)\gamma_2^2)$, where $\gamma_1 = (1 + K_s)L + K_p$ and $\gamma_2 = LK_s$ [40]. We note that the WMMSE-based method has additional complexity of outer iterations to solve for the auxiliary optimization variables, $\{\mu_k, v_k\}$ when compared to the MM based method. The number of required iterations depends on the initial point. Similarly, the projection-based method, which involves iteratively solving convex quadratic problems, also has the same complexity order as the WMMSE based approach.

V. NUMERICAL RESULTS

We evaluate the performance of the proposed power allocation schemes: the sequential WMMSE-based Dinkelbach algorithm, and the sequential MM-based Dinkelbach's algorithm, through numerical simulations. We also compare the performance with other heuristic power allocation policies suitable for CF-mMIMO and adapt them to satisfy the constraints. We consider a secondary CF-mMIMO network in an outdoor square area of 125 $m \times 125$ m, and a PN with users distributed in an outdoor area of 100 $m \times 100$ m. The S-APs are placed on the border of the service area with equal spacing, and at a vertical height of 5 m above the ground. The S-UEs are uniformly distributed in the square area. Similarly, P-UEs are uniformly distributed in the PN area and the PN base station is located in the center of this area with a vertical height of 5 m above the ground. The SN and PN are placed diagonally opposite in overall square setup of 250 $m \times 250$ m and one instance of the simulation setup is shown in Fig. 2. The thermal noise variance and path-loss between S-APs and S-UEs are modeled as $\varsigma^2 = -174 + 10\log_{10}(B) + \text{NF}$ dBm and the large-scale fading coefficient is modeled $\beta_{kl} = -30.5 + 36.7\log_{10}(d_{kl}/1\ m) + 10\log_{10}(z_{kl})$ dB, respectively,

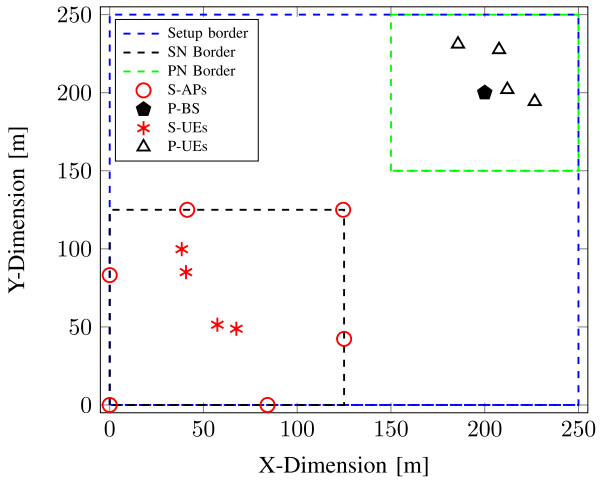


FIGURE 2. An instance of the simulation setup.

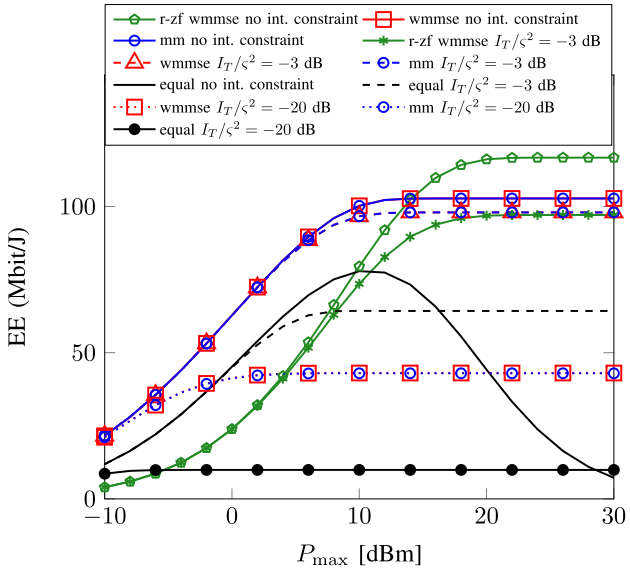


FIGURE 3. Comparison of proposed and equal power allocation schemes with EE versus $P_{\max}$ at S-APs for different normalized interference thresholds I_T/ς^2.

where $B = 20$ MHz, noise figure NF $= 9$ dB and z_{kl} captures the log-normal shadowing and we chose the standard deviation to be 4 dB. The same path loss model is considered between P-UEs and S-APs. We consider $\tau_c = 2000$, $\tau_p = 8$, $\xi = 0.25$ W/(Gbit/s), $\zeta = 1.4$, and PC $= 1$ W. We set $L = 6, N = 4, K_s = 4, M = 5$, and $K_p = 4$.

In the plots, we denote the WMMSE-based Dinkelbach's algorithm by *wmmse*, the MM-based Dinkelbach's algorithm by *mm*, equal power allocation by *equal*, and results generated without interference constraints by *no int. constraint*. The equal power allocation policy we compare with is as follows:

$$p_{il} = \min\left\{\frac{P_{\max}}{K_s}, \frac{I_T}{V_1}, \ldots, \frac{I_T}{V_{K_p}}\right\}, \; V_m \triangleq \sum_{l=1}^{L}\sum_{i=1}^{K_s} b_{iml}. \tag{47}$$

We also assume that the P-BS employs MR precoding and MMSE channel estimation $\{\widehat{\mathbf{u}}_{\text{pp-}i}, \; i = 1, \ldots, K_p\}$. Then $\bar{\mathbf{D}}_{\text{p}}$ can be computed as follows,

$$\bar{\mathbf{D}}_{\text{p}} = \sum_{i=1}^{K_p} \frac{\widehat{\mathbf{D}}_{\text{pp-}i}}{\text{Tr}\left(\widehat{\mathbf{D}}_{\text{pp-}i}\right)}, \tag{48}$$

where $\widehat{\mathbf{D}}_{\text{pp-}i} = \mathbb{E}\{\widehat{\mathbf{u}}_{\text{pp-}i}\widehat{\mathbf{u}}^H_{\text{pp-}i}\}$ and $\widehat{\mathbf{u}}_{\text{pp-}i}$ is the MMSE channel estimate of $\mathbf{u}_{\text{pp-}i}$ at the P-BS.

To study the affect of M on the performance of the S-UEs: note that M only affects the interference component, say $\bar{\sigma}_k^2$ of σ_k^2, in the optimization problem. From Section II-B, this is given by

$$\bar{\sigma}_k^2 = \text{Tr}\left(\bar{\mathbf{D}}_{\text{p}}\mathbf{D}_{\text{sp-}k}\right). \tag{49}$$

With the uncorrelated channel model and MR precoding, we have $\bar{\mathbf{D}}_{\text{p}} = \frac{K_{\text{p}}}{M}\mathbf{I}_M$ (as $\widehat{\mathbf{D}}_{\text{pp-}i} = \widehat{\beta}_{\text{pp-}i}\mathbf{I}_M$, with $\widehat{\beta}_{\text{pp-}i}$ being the variance of each element of the corresponding channel estimate vector $\widehat{\mathbf{u}}_{\text{pp-}i}$) and $\mathbf{D}_{\text{sp-}k} = \beta_{\text{sp-}k}\mathbf{I}_M$, where $\beta_{\text{sp-}k}$ is the corresponding path loss coefficient. This implies that

$$\bar{\sigma}_k^2 = K_{\text{p}}\beta_{\text{sp-}k} \tag{50}$$

holds. Therefore, the average interference posed by P-BS to S-UEs is independent of M. Even with correlated channel models and other precoding schemes, we have seen through simulations that M does not cause any noticeable change in the performance.

In Fig. 3, we report the results of EE versus $P_{\max}$ for different interference threshold values normalized by the noise power I_T/ς^2. The performances of both proposed methods, i.e., the WMMSE- and MM-based Dinkelbach algorithms, are plotted along with that of the equal power allocation scheme. For the proposed algorithms, the EE increases as $P_{\max}$ increases for small values of $P_{\max}$ and saturates for large $P_{\max}$. The EE versus $P_{\max}$ curve has three regions of operation: *power constrained region*, *interference constrained region* and *EE constrained region*. In the power constrained region, the EE is constrained by $P_{\max}$ and increases with increasing $P_{\max}$; in other words, the power constraint is active. In the EE constrained region, the transmission with higher power is limited by the EE, though the power and interference constraints are inactive. We can observe this with the performance of the proposed algorithms with no interference constraint. We see that equal power allocation decreases after a certain $P_{\max}$ because the SE increases logarithmically and the sum power increases linearly, i.e., the gain in sum SE is much less than the increase in $P_{\max}$.

In Fig. 3, we observe that the proposed WMMSE- and MM-based algorithms offer superior performance over equal power allocation. Especially, the performance gap is larger in the EE-constrained region. This is because the algorithms allocate power coefficients that maximize EE and hence after a certain threshold, the algorithms do not allocate any further power as that would decrease the EE. On the other hand,

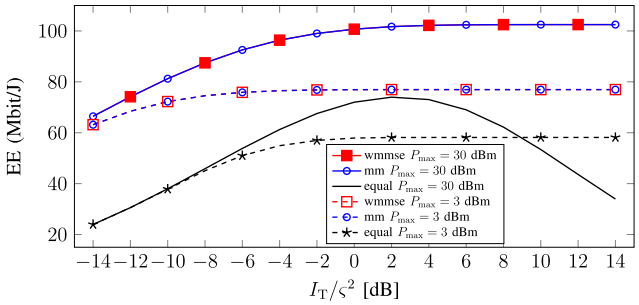


FIGURE 4. Comparison of proposed and equal power allocation schemes with EE versus I_T/ς^2 at S-APs for different power budgets $P_{\max}$.

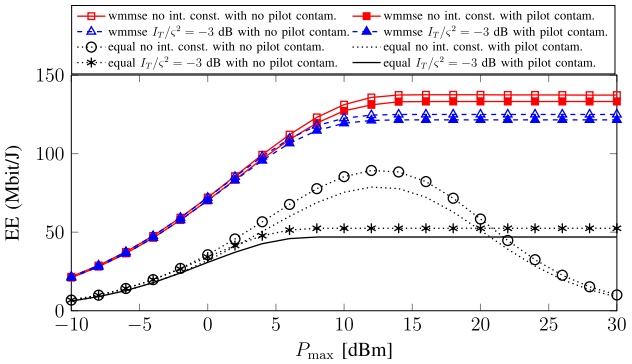


FIGURE 5. EE versus $P_{\max}$ at S-APs for different I_T/ς^2 with pilot contamination.

the equal power allocation in the EE-constrained region has degrading performance; for instance, this behavior can be seen in the plot with no interference constraint. We observe that both the proposed methods have similar performance. Furthermore, they perform significantly better than the equal power allocation. The figure also shows the performance with regularized R-ZF (denoted by r-zf in Fig. 3) given by

$$\mathbf{w}_{il} = \frac{\bar{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_{il}\|^2\right\}}} \tag{51}$$

where $\bar{\mathbf{w}}_{kl}$ is the kth column of the matrix

$$\bar{\mathbf{W}}_l = \widehat{\mathbf{H}}_l \left(\widehat{\mathbf{H}}_l^H \widehat{\mathbf{H}}_l + \alpha \mathbf{I}\right)^{-1} \tag{52}$$

and $\widehat{\mathbf{H}}_{kl} \triangleq [\widehat{\mathbf{h}}_{1l}, \ldots, \widehat{\mathbf{h}}_{K_s l}]$, where $\alpha \geq 0$ is regularizing parameter. For the simulation, we set $\alpha = 5$. We observe that at low $P_{\max}$, the performance of R-ZF is inferior to that of the MR approach. However, as $P_{\max}$ increases, the performance of R-ZF improves rapidly. Specifically, in the unconstrained case, there is a clear EE gain at high SNR levels. Conversely, with an interference constraint of $I_T/\varsigma^2 = -3$ [dB], the increase in EE is bounded.

In Fig. 4, we report the results of EE versus I_T/ς^2 for different values of $P_{\max}$. We observe that with an increase in the interference threshold I_T/ς^2 there is a significant gain in the performance over the equal power allocation policy. We also note from the plot that the EE-constrained region is observed only with the equal power allocation policy when the interference constraint is inactive. This plot provides complementary insights into performance compared to Figure 3.

In Fig. 5, we report the EE versus $P_{\max}$ for different values of I_T/ς^2 under pilot contamination. For this specific plot, to obtain substantial impact of pilot contamination on the performance, we set the simulation parameters to $K_s = 10$ and $K_p = 8$; this amounts to a total of 60 optimization variables and 78 constraints (power budget per S-AP, interference threshold at each P-UE and non-negativity of the power coefficients) and as such the system is highly loaded. For the case of pilot contamination, we considered $\tau_1 = 5$, $\tau_2 = 2, \tau_3 = 6$. Further, pilots assignment to P-UEs and S-UEs are done as described in [41]. We observe that the pilot contamination degrades the S-UEs performance, as expected. However, the relative performance of the methods remains the same.

A. COMPARISON WITH STATE-OF-THE-ART POWER ALLOCATION SCHEMES

We will now compare the performance of the WMMSE-based algorithm with state-of-the-art power allocation policies. The first one is the scalable distributed power allocation for CF-mMIMO networks described in [42]. It is given by

$$p_{kl} = P_{\max} \frac{(\beta_{kl})^{\delta}}{\sum_{i=1}^{K_s} (\beta_{il})^{\delta}} \tag{53}$$

where the parameter $\delta \in [-1, 1]$ controls the power allocation behavior. We will use $\delta = 0.5$ as it is shown in [42] that the power allocation with $\delta = 0.5$ is a good heuristic for maximizing the SR. When $\delta = 0$, we get the equal power allocation scheme.

Although the power allocation policy in (53) satisfies the power constraint (19b), it may not satisfy the interference constraints. In order to adapt any power allocation policy to satisfy the constraints in (19), we use the projection and scaling approach discussed in (43) and (44). We use the scaling factor $\eta = 0.99$ in the simulation results.

The next algorithm we compare with is the PGA method discussed in Section IV-B2. The final algorithm that we will compare with is the MM-based approach without an interference constraint. The so-obtained power coefficients are adapted to satisfy the interference constraint by scaling and projecting the power coefficients to the constraint space. The algorithm is analogous to the coordinate descent-based algorithm presented in [10].

For the simulation plot in Fig. 6, we will label the power allocation policy (53) by projecting and scaling as *cf power (proj)* and *cf power (scaling)*, respectively. We will use *pga* for the PGA method. Finally, we denote the MM-based algorithm without interference constraints and adapting using projection by *modified mm (proj)*, and adapting by scaling by *modified mm (scaling)*. The reason for the label *modified mm* is that the power coefficients are computed using

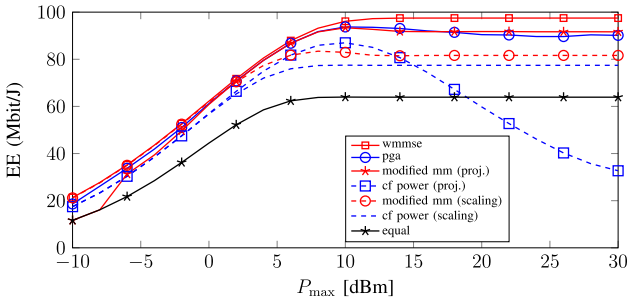


FIGURE 6. Comparison of different power allocation schemes with the proposed WMMSE algorithm through EE versus $P_{\max}$ at S-APs for $I_T/\varsigma^2 = -3$ dB, $\eta = 0.99$.

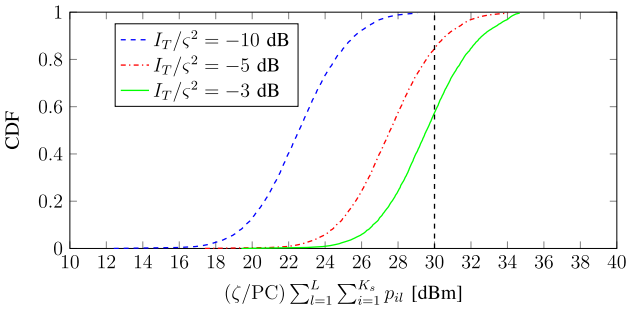


FIGURE 7. Maximum total power (scaled by (ζ/PC)) allocated by the WMMSE method under different interference thresholds I_T/ς^2 and $P_{\max} = 30$ dBm.

the MM-based Dinkelbach's algorithm without interference constraints.

The performance of any other competing algorithm cannot always be guaranteed to be optimal due to the non-convex nature of the problem, unless a method to compute the global optimum exists of which to the best of our knowledge we are unaware. Furthermore, the proposed method depends only on statistics, which means the optimization is carried out once or very infrequently, as long as the statistics of the channels do not change substantially. This optimization can be performed at the CPU and the optimized power coefficient can then be distributed to the APs of the secondary network. Therefore, our approach is suitable in applications and scenarios where the assumption of relatively constant channel statistics holds true rendering continuous resource-intensive calculations not a bottleneck.

In Fig. 6, we plot the EE versus $P_{\max}$ and compare the performance of the various methods. We first note that the WMMSE-based approach performs best for all values of $P_{\max}$. The best among the alternative algorithms is the PGA method. For a fair comparison, the initial starting point of both the WMMSE-based algorithm and the PGA method are set to the same. The advantage of the PGA method is that we only need to compute the gradient and avoid second-order derivatives. However, the technique requires projection onto the constraint space in every iteration of the algorithm.

Furthermore, as the objective is non-convex, the PGA algorithm may not converge for some initial points. The other good heuristic power allocation policy (53) that we adapted by projecting onto the constraint space (43) has a significant performance gain over the equal power allocation policy. The reason is embedded in the observation that this scheme maximizes the sum SE [31] and thus improves EE. We notice that heuristic power allocation schemes adapted by scaling the power coefficients have inferior performance compared to their projection-based counterparts, albeit better performance than that of the equal power allocation policy. The advantage of the scaling-based algorithms is that we avoid solving the optimization problem in (43), which otherwise is required for the projection. We also compare the same methods as in Fig. 6, with $I_T/\varsigma^2 = -20$ dB. We observe that the relative performance remains the same. However, the performance gap to the WMMSE-based algorithm is relatively larger than when the interference constraint is stringent. We also note that the performance of the scaling-based schemes is inferior to that of the projection-based counterparts.

B. COMPARISON OF SUM-RATE MAXIMIZATION AND EE FORMULATION

In this section, we discuss a few interesting parallels between EE maximization and SR maximization, which is a widely used utility function for communication systems. By SR we mean bandwidth times sum SE. The first important feature to note is how both cost functions, i.e., EE and SR, behave with an increase in $P_{\max}$ at each AP. The SR increases monotonically with an increase in the total per-AP power budget, $P_{\max}$. However, the EE increases with an increase in total power only up to a certain threshold. After that, the gain in SR is minimal, with a significant increase in power causing the EE to decrease.

With very strict interference constraints (or when the power budget is very low), the power allocated by the S-APs to the S-UEs reduces, and when $(\zeta/\mathrm{PC})\,\|\mathbf{c}\|^2 \ll 1$, the total power consumed in (23) becomes $\bar{P}_T(\mathbf{p}) \approx \mathrm{PC}$, which is approximately independent of the power coefficients. In this situation, the optimal power coefficients that maximize the SR also maximize the EE. Thus, depending on the interference threshold one can reduce the complexity of solving for the power coefficients by solving a sum SE maximization problem instead. We can solve for the SR as a special case of methods already proposed for EE maximization, i.e., the WMMSE- and MM-based Dinkelbach algorithms by substituting $\zeta = 0$ and $\mathrm{PC} = 1$.

In Fig. 7, we validate that under tighter interference constraints, the total power allocated by all the S-APs is very small. The plots in figure demonstrate that with $I_T/\varsigma^2 = -10$ dB, almost 100% of the setups have a total allocated fractional power of less than unity i.e., $(\zeta/\mathrm{PC})\,\|\mathbf{c}\|^2 \ll 1$. Similarly, with $I_T/\varsigma^2 = -5$ dB, we observe that 85% of the setups satisfy the condition. Under such operating regimes, the solution of the SR maximization will have an EE comparable to that of the proposed EE maximization algorithms, which can be seen in Fig. 8.

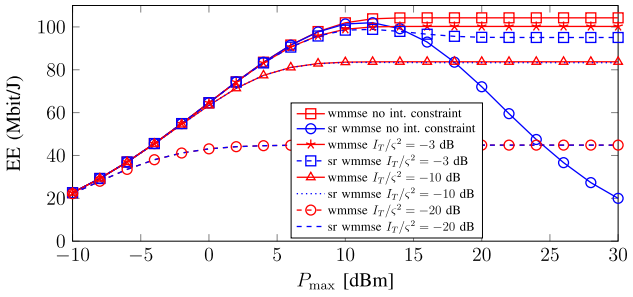


FIGURE 8. EE versus $P_{\max}$ at S-APs for different interference thresholds I_T/ς^2 to compare power coefficients of the WMMSE based algorithm and those obtained through SR maximization.

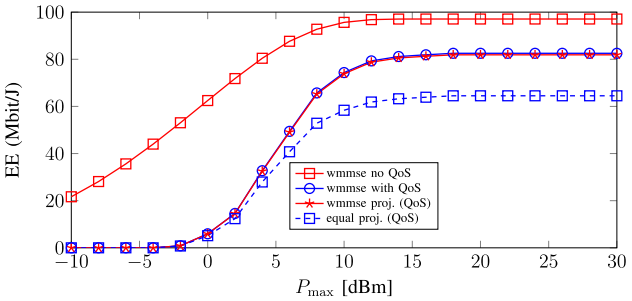


FIGURE 9. EE versus $P_{\max}$ at S-APs for $I_T/\varsigma^2 = -3$ dB with $\mathrm{SE}_k^\circ = 1$ bpcu.

In Fig. 8, we plot the EE versus $P_{\max}$ to compare the power coefficients computed through SR maximization with the EE maximization. For SR maximization, we use the proposed WMMSE based Dinkelbach's algorithm and set $\zeta = 0$ and PC $= 1$; we denote this method by *sr wmmse* in the plot. When $P_{\max}$ is very small then performances with the power coefficients obtained through maximizing SR and EE are comparable, as the condition $(\zeta/\mathrm{PC})\,\|\mathbf{c}\|^2 \ll 1$ is satisfied. After certain $P_{\max}$, with no interference constraint the performance of power coefficients obtained by maximizing SR degrades in terms of EE because SE increases logarithmically with $P_{\max}$ whereas the total power consumption increases linearly. With $I_T/\varsigma^2 = -3$ dB, the EE with SR maximizing coefficients reaches a peak and then decreases as $P_{\max}$ increases due to transmission with higher power. A further increase in $P_{\max}$ saturates the EE due to limitations in the power imposed by the interference constraint. However, with $I_T/\varsigma^2 = -10$ or -20 dB, the condition $(\zeta/\mathrm{PC})\,\|\mathbf{c}\|^2 \ll 1$ is satisfied for all values of $P_{\max}$ that are of interest. Therefore, power allocation obtained to maximize EE and SR performs similarly.

C. QUALITY-OF-SERVICE CONSTRAINT

The main challenge for an underlay SN is to ensure a minimal QoS to S-UEs while maintaining the interference to PN within permissible levels. To provide minimum QoS to S-UEs, we can impose a minimum SE constraint as an additional constraint to (19). We can model the minimum SE constraint, SE_k°, as a conic constraint as

$$\begin{aligned} \mathrm{SE}_k^\circ &\le \frac{\tau_d}{\tau_c}\log_2\left(1+\Gamma_k(\mathbf{c})\right) \\ \rho_k &\le \frac{|\mathbf{c}_k^T\bar{\mathbf{g}}_{kk}|^2}{\sum_{i=1}^{K_s}\mathbf{c}_i^T\mathbf{A}_{ki}\mathbf{c}_i - |\mathbf{c}_k^T\mathbf{g}_{kk}| + \sigma_k^2}, \\ \left\|\begin{bmatrix}\mathbf{B}_k^{1/2}\mathbf{c}\\ \sigma_k\end{bmatrix}\right\|_2 &\le \sqrt{\frac{1+\rho_k}{\rho_k}}\mathbf{c}_k^T\bar{\mathbf{g}}_{kk}, \forall k \in [K_s], \end{aligned} \tag{54}$$

where

$$\begin{aligned} \rho_k &\triangleq 2^{\frac{\tau_c SE_k^\circ}{\tau_d}} - 1 \ge 0, \\ \mathbf{B}_k^{1/2} &= \mathrm{blkdiag}(\mathbf{A}_{k1}^{1/2}, \cdots, \mathbf{A}_{kK_s}^{1/2}). \end{aligned} \tag{55}$$

We note that each constraint in (54) is a second-order cone and hence is convex [43]. These constraints in particular affect Step 3 of the coordinate ascent algorithm in Section III-B3. Moreover, since these additional constraints are convex, the optimization problem in (36) is convex. Furthermore, from [43], we can cast a QCQP problem into a second-order conic programming (SOCP). Thus, the overall optimization problem in (36) along with the constraints in (54) is a SOCP problem and thus, can be efficiently solved using interior-point methods for conic problems. A proof of convergence is given in Appendix B (in the proof $\mathcal{Z}$ represents the feasible set defined by the convex constraints defined in (32) and (54)).

We add these constraints in (54) to the optimization problem in (23) and solve it using the proposed WMMSE based approach. We illustrate the effect of the constraint (54) on the performance through Fig. 9 and Fig. 10 for which we set $\mathrm{SE}_k^\circ = 1$ bpcu (bit/channel use), $k \in [K_s]$. In these plots, curves labeled by *wmmse no QoS* show the performance of the power allocation obtained by solving (23) without the QoS constraint (54); curves labeled by *wmmse with QoS* show the performance when we solve (23) with QoS constraint using WMMSE based algorithm; curves labeled by *wmmse proj. (QoS)* is the approach in which the solution of (23) is projected on to the constraint space including (54); and similarly *equal proj. (QoS)* represents the projection of equal power allocation on the constraint space. In Fig. 9, we plot EE versus $P_{\max}$, and observe that for lower values of $P_{\max}$, the system is unable to provide minimum QoS to all S-UEs. At high $P_{\max}$, there is a approximately a 15% decrease in the performance with a QoS constraint compared to the case with no QoS. Further, we observe that *wmmse with QoS* is comparable to *wmmse proj. QoS* and both approaches perform better than *equal proj. (QoS)*.

In Fig. 10, we plot a CDF of the achievable EE with and without a QoS constraint and randomness is over different user locations. We note that the QoS constraint has a profound affect on the algorithms feasibility; specifically we note that approximately 18% of setups are in outage when there is a QoS constraint compared to the case with no QoS.

In Fig. 11, we plot Jain's fairness for the per-user SE for SN (average over setups),

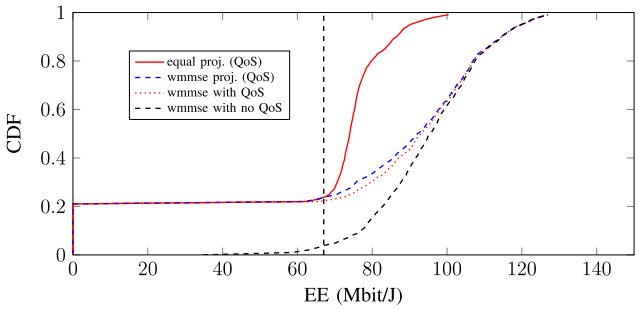


FIGURE 10. CDF of EE for $I_T/\varsigma^2 = -3$ dB with $\mathrm{SE}_k^\circ = 1$ bpcu.

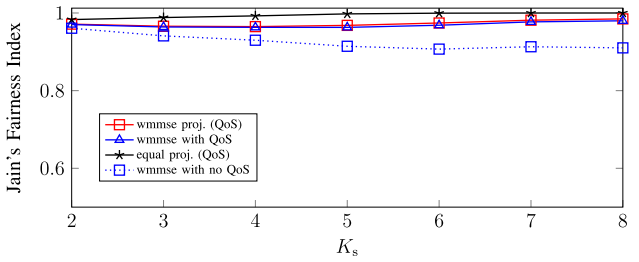


FIGURE 11. Comparing SE fairness among S-UEs using Jain's fairness index.

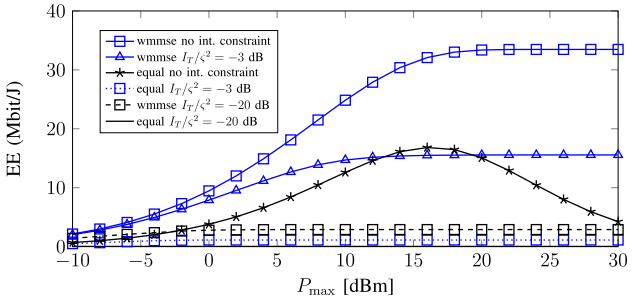


FIGURE 12. EE versus $P_{\max}$ at S-APs for different $I_T/\varsigma^2 = -3$ dB with the mixed setup.

$\mathbb{E}\{(\sum_{k=1}^{K_s} \mathrm{SE}_k)^2/(\sum_{k=1}^{K_s} \mathrm{SE}_k^2)\}$. We observe that the proposed *wmmse method with QoS* provides better fairness among users compared to methods that have no QoS constraint and to that obtained by projecting onto the constraint space. On the other hand equal power allocation with projection achieves the best fairness among S-UEs; however this comes at the cost of a performance loss.

D. MIXED LOCATION SCENARIO

Finally, in Fig. 12, we demonstrate that the methodology developed in the paper can be applied to any configurations of PN and SN network locations. We plot the average EE versus $P_{\max}$ with locations of PN and SN networks overlapping and P-UEs/S-UEs are uniformly distributed in a 250 $m \times 250$ m area, while all other parameters remain the same. While the quantitative results may vary, the relative performance remains similar. It may be required to adjust certain parameters such as thresholds and maximum power budgets to get reasonable performance.

VI. CONCLUSION

In this paper, we proposed power allocation policies that maximize the EE of a secondary CF-mMIMO network in a spectrum sharing scenario under constraints on the transmit powers, on the average received interference power, and on minimum per-user rates. Specifically, we developed two iterative algorithms, namely, a WMMSE-based Dinkelbach algorithm that is applicable to any transmit precoder, and an MM-based Dinkelbach algorithm for MR precoding that has less complexity. We also provided computationally simple, heuristic power allocation policies for the problem at hand. Our numerical results demonstrated that the two proposed iterative algorithms achieve a significant performance gain over the other methods. The results also showed that having minimum-rate constraints reduces the maximum EE but on other hand improves fairness among users.

APPENDIX A: CLOSED-FORM TERMS FOR THE MR PRECODER

With MR processing, we can derive close form expressions for $\mathbf{A}_{ki} = \mathbb{E}\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\}$. Recall that $\breve{\mathbf{R}}_{ikl} = \mathbb{E}\left\{\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{kl}^H\right\} = \sqrt{\eta_{s\text{-}k}\eta_{s\text{-}i}}\tau_p\mathbf{R}_{il}\boldsymbol{\Psi}_{t_kl}^{-1}\mathbf{R}_{kl} = \breve{\mathbf{R}}_{kil}^H$, also note that $\breve{\mathbf{R}}_{ikl} = \widehat{\mathbf{R}}_{kl}$ for $k = i$. The parameters of the WMMSE can be computed from parameters of the MM method

$$\mathbf{A}_{ki} = \mathrm{diag}(\mathbf{b}_{ki}) + \mathbf{a}_{ki}\mathbf{a}_{ki}^H, \forall k, i = 1, \ldots, K_s \tag{56}$$

where $\mathrm{diag}(\mathbf{b}_{ki})$ is a diagonal matrix with (r, r) entry being $b_{kir}, \forall r \in [K_s]$ and $\mathbf{a}_{ki}$ is a vector with jth entry being $a_{kij}, \forall j \in [K_s]$.

Now the other term required for simulation is $\bar{\mathbf{g}}_{ki}$ which is given as $\bar{\mathbf{g}}_{kil} = \frac{\mathrm{Tr}(\breve{\mathbf{R}}_{ikl})}{\sqrt{\mathrm{Tr}(\widehat{\mathbf{R}}_{il})}}\mathcal{I}_{i\in\mathcal{S}_k^p}$ for $k \neq i$ and $\bar{\mathbf{g}}_{iil} = \mathrm{Tr}(\widehat{\mathbf{R}}_{il})^{1/2}$ for $k = i$, where $\mathcal{I}_{i\in\mathcal{S}_k^p}$ is an indicator function which is equal to one when $i \in \mathcal{S}_k^p$ otherwise it is zero.

APPENDIX B: CONVERGENCE PROOF FOR THE WMMSE-BASED DINKELBACH'S ALGORITHM

The sequence of objective function values in Algorithm 1 converges to a limit, since this sequence is non-decreasing, and uniformly bounded from above by a constant.

In more detail, we explain this as follows. Consider the optimization problem in (32). Let $\mathbf{z} = \{\mathbf{c}, \mathbf{v}, \boldsymbol{\mu}\} \in \mathcal{Z}$ represent single vector encompassing all the optimization variables, where $\mathcal{Z}$ is the feasibility set defined by the convex constraints in (32). Also, note that the power coefficients vector $\mathbf{c}$ is bounded. Let

$$f(\mathbf{z}) = \sum_{i=1}^{K_s} -v_i e_i(\mu_i, \mathbf{c}) + \ln v_i + K_s, \tag{57}$$

and recall that

$$f_2(\mathbf{z}) = \zeta \|\mathbf{c}\|^2 + \mathrm{PC}, \ \mathrm{PC} \geq 0. \tag{58}$$

Note that for $\boldsymbol{\mu}$ given in (33) and $\mathbf{v}$ given in (34), $f(\mathbf{z}) = f_1(\mathbf{z})$ (i.e., sum SE) given in (24). Also let,

$$\bar{F}(\mathbf{z}; \lambda) = f(\mathbf{z}) - \lambda f_2(\mathbf{z}) \tag{59}$$

The algorithm has two loops: the outer loop, which we will call Dinkelbach's part, iterates over λ, and the inner loop, which we will call the WMMSE part, uses a coordinate ascent method for a given λ. In the outer loop, only λ is updated, while in the inner loop, the optimization vector $\mathbf{z}$ is updated. We denote $\widetilde{\mathbf{z}}_n^m \in \mathcal{Z}$ as the value of $\mathbf{z}$ during the outer loop iteration n and the inner loop iteration m. Similarly, we denote λ_n as the value of λ during the outer loop iteration n. The algorithm is initiated with some $\widetilde{\mathbf{z}}_0^0 \in \mathcal{Z}$. Moreover, we denote $\mathbf{z}_n$ as the value of $\widetilde{\mathbf{z}}_n^m$ when the WMMSE part converges, and then we set $\lambda_{(n+1)} = \frac{f(\mathbf{z}_n)}{f_2(\mathbf{z}_n)} \geq 0$ for the outer loop. To prove the convergence of the proposed algorithm, we first prove the convergence of the WMMSE part and then the convergence of Dinkelbach's part.

CONVERGENCE OF INTERNAL LOOP: WMMSE PART

For a given outer loop iteration n, the algorithm involves an inner loop comprising three sub-steps. For ease of notation, we shall drop the outer loop index n. Let $\widetilde{\mathbf{z}}^m = \{\widetilde{\mathbf{c}}^m, \widetilde{\mathbf{v}}^m, \widetilde{\boldsymbol{\mu}}^m\}$ for the inner loop iteration m. The first sub-step for a given $\{\widetilde{\mathbf{c}}^m, \widetilde{\mathbf{v}}^m\}$ is:

$$\boldsymbol{\mu}_\circ^m = \underset{\boldsymbol{\mu}}{\operatorname{argmax}}\ \bar{F}(\widetilde{\mathbf{c}}^m, \widetilde{\mathbf{v}}^m, \boldsymbol{\mu}; \lambda). \tag{60}$$

In the second step for a given $\left\{\widetilde{\mathbf{c}}^m, \boldsymbol{\mu}_\circ^m\right\}$ is

$$\mathbf{v}_\circ^m = \underset{\mathbf{v}}{\operatorname{argmax}}\ \bar{F}(\widetilde{\mathbf{c}}^m, \mathbf{v}, \boldsymbol{\mu}_\circ^m; \lambda), \tag{61}$$

and in the third step for a given $\left\{\mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m\right\}$, we update $\mathbf{c}$ by

$$\mathbf{c}_\circ^m = \underset{\mathbf{c}\in\mathcal{Z}}{\operatorname{argmax}}\ \bar{F}(\mathbf{c}, \mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m; \lambda). \tag{62}$$

Thus, the following holds,

$$\begin{aligned}\bar{F}(\widetilde{\mathbf{c}}^m, \widetilde{\mathbf{v}}^m, \widetilde{\boldsymbol{\mu}}^m; \lambda) &\leq \bar{F}(\widetilde{\mathbf{c}}^m, \widetilde{\mathbf{v}}^m, \boldsymbol{\mu}_\circ^m; \lambda) \leq \bar{F}(\widetilde{\mathbf{c}}^m, \mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m; \lambda) \\ &\leq \bar{F}(\mathbf{c}_\circ^m, \mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m; \lambda) \\ &= \bar{F}(\widetilde{\mathbf{z}}_\circ^m; \lambda),\end{aligned} \tag{63}$$

where, $\widetilde{\mathbf{z}}_\circ^m = \{\mathbf{c}_\circ^m, \mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m\}$.

Moreover, using (27)-(31), we have $\bar{F}(\widetilde{\mathbf{c}}^m, \mathbf{v}_\circ^m, \boldsymbol{\mu}_\circ^m; \lambda) = f_1(\widetilde{\mathbf{c}}^m) - \lambda f_2(\widetilde{\mathbf{c}}^m)$. This implies, using (62), that we maximize the difference of the sum SE, and term that is which is directly proportional to the square-norm of the transmit power coefficients vector. Since, the power coefficients are bounded, $\bar{F}(\widetilde{\mathbf{z}}_\circ^m; \lambda) \leq B_0$ is also upper bounded by some positive constant B_0. At iteration, $m+1$, we set $\widetilde{\mathbf{z}}^{(m+1)} = \widetilde{\mathbf{z}}_\circ^m$. Thus, we obtain a monotonically increasing sequence $\{\bar{F}(\widetilde{\mathbf{z}}^m; \lambda)\}$ and at iteration $m+1$ the following holds

$$\bar{F}(\widetilde{\mathbf{z}}^m; \lambda) \leq \bar{F}(\widetilde{\mathbf{z}}^{(m+1)}; \lambda) \leq B_0. \tag{64}$$

Therefore, the internal loop converges and we terminate when $\bar{F}(\widetilde{\mathbf{z}}^{(m+1)}; \lambda) - \bar{F}(\widetilde{\mathbf{z}}^m; \lambda) \leq \epsilon$ for some $\epsilon > 0$.

CONVERGENCE OF OUTER LOOP: DINKEBACH'S PART

For the convergence of WMMSE part, we adapt the methodology presented in [38]. For the outer loop iteration n

$$\bar{F}(\mathbf{z}_n; \lambda_n) = f(\mathbf{z}_n) - \lambda_n f_2(\mathbf{z}_n). \tag{65}$$

Now consider two important properties of $\bar{F}(\mathbf{z}_n; \lambda_n)$ that we will use for the convergence proof.

1) $\bar{F}(\lambda)$ is continuous in λ, for a given $\mathbf{z}$.
2) For $\lambda_2 \geq \lambda_1$, $\bar{F}(\lambda_2) \leq \bar{F}(\lambda_1)$. Thus, $\bar{F}(\lambda)$ is a decreasing function of λ.
3) The function in the outer loop, $\bar{F}(\mathbf{z}_n; \lambda_n)$ produces nonnegative sequence of values because

$$\bar{F}(\mathbf{z}_n; \lambda_n) \geq \bar{F}(\mathbf{z}_{(n-1)}; \lambda_n) = 0. \tag{66}$$

First inequality is due to the maximization of the internal loop for given λ_n and $\mathbf{z}_{(n-1)}$. The second equality is because $\lambda_n = \frac{f(\mathbf{z}_{(n-1)})}{f_2(\mathbf{z}_{(n-1)})}$.

Now consider the update rule of λ:

$$\lambda_{(n+1)} = \frac{f(\mathbf{z}_n)}{f_2(\mathbf{z}_n)} = \lambda_n + \frac{f(\mathbf{z}_n) - \lambda_n f_2(\mathbf{z}_n)}{f_2(\mathbf{z}_n)}. \tag{67}$$

From (67) and (66), the following holds

$$(\lambda_{(n+1)} - \lambda_n) f_2(\mathbf{z}_n) = \bar{F}(\mathbf{z}_n; \lambda_n) \geq 0. \tag{68}$$

Since $f_2(\mathbf{z}_n) > 0$, this implies that $\lambda_{(n+1)} > \lambda_n$ (unless the outer loop has converged). Hence, $\{\lambda_n\}$ is a non-decreasing sequence with each outer iteration. This implies that the outer loop produces non-increasing valued sequence of $\bar{F}(\lambda_n)$ and further this sequence is lower bounded by zero. From the fact that this function is continuous in λ and lower bounded, the outer loop converges and we terminate the outer loop when $\bar{F}(\mathbf{z}_n; \lambda_n) \leq \epsilon$, for some $\epsilon > 0$.

ACKNOWLEDGMENT

An earlier version of this paper was presented in part at IEEE ICC 2022 [DOI: 10.1109/ICC45855.2022.9838491].

REFERENCES

[1] Z. H. Shaik, R. Sarvendranath, and E. G. Larsson, "Energy-efficient power allocation for an underlay spectrum sharing RadioWeaves network," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 799–804.

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

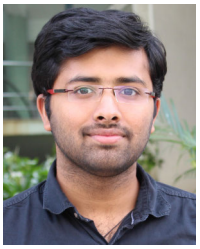
[3] L. Van der Perre, E. G. Larsson, F. Tufvesson, L. D. Strycker, E. Björnson, and O. Edfors, "RadioWeaves for efficient connectivity: Analysis and impact of constraints in actual deployments," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 15–22.

[4] W. S. H. M. W. Ahmad, N. A. M. Radzi, F. S. Samidi, A. Ismail, F. Abdullah, M. Z. Jamaludin, and M. N. Zakaria, "5G technology: Towards dynamic spectrum sharing using cognitive radio networks," *IEEE Access*, vol. 8, pp. 14460–14488, 2020.

[5] *Amendment of the Commission's Rules With Regard to Commercial Operations in the 3550–3650 MHz Band*, document FCC-15-47, Federal Commun. Commission (FCC), 2015.

[6] *Unlicensed Use of the 6 GHz Band; Expanding Flexible Use in Mid-band Spectrum Between 3.7 and 24 GHz*, document FCC-20-51, 2020.

[7] E. Au, "IEEE 802.11be: Extremely high throughput [Standards]," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 138–140, Sep. 2019.
[8] L. Giupponi and C. Ibars, "Distributed cooperation in cognitive radio networks: Overlay versus underlay paradigm," in *Proc. VTC Spring-IEEE 69th Veh. Technol. Conf.*, Apr. 2009, pp. 1–6.
[9] 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 Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.
[10] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sep. 2019.
[11] F. Tan, P. Wu, Y.-C. Wu, and M. Xia, "Energy-efficient non-orthogonal multicast and unicast transmission of cell-free massive MIMO systems with SWIPT," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 949–968, Apr. 2021.
[12] L. You, Y. Huang, D. Zhang, Z. Chang, W. Wang, and X. Gao, "Energy efficiency optimization for multi-cell massive MIMO: Centralized and distributed power allocation algorithms," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5228–5242, Aug. 2021.
[13] X. Yu, W. Xu, S.-H. Leung, Q. Shi, and J. Chu, "Power allocation for energy efficient optimization of distributed MIMO system with beamforming," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8966–8981, Sep. 2019.
[14] V. Tentu, E. Sharma, D. N. Amudala, and R. Budhiraja, "UAV-enabled hardware-impaired spatially correlated cell-free massive MIMO systems: Analysis and energy efficiency optimization," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2722–2741, Apr. 2022.
[15] T. C. Mai, H. Q. Ngo, and L.-N. Tran, "Energy efficiency maximization in large-scale cell-free massive MIMO: A projected gradient approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6357–6371, Aug. 2022.
[16] H. D. Tuan, A. A. Nasir, H. Q. Ngo, E. Dutkiewicz, and H. V. Poor, "Scalable user rate and energy-efficiency optimization in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6050–6065, Sep. 2022.
[17] S. Chen, J. Zhang, E. Björnson, Ö. T. Demir, and B. Ai, "Energy-efficient cell-free massive MIMO through sparse large-scale fading processing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 1–19, Dec. 2023.
[18] I. M. Braga, R. P. Antonioli, G. Fodor, Y. C. B. Silva, and W. C. Freitas, "Efficient battery usage in wireless-powered cell-free systems with self-energy recycling," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6856–6861, May 2023.
[19] R. Pinto Antonioli, I. M. Braga, G. Fodor, Y. C. B. Silva, A. L. F. de Almeida, and W. C. Freitas, "On the energy efficiency of cell-free systems with limited fronthauls: Is coherent transmission always the best alternative?" *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8729–8743, Oct. 2022.
[20] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "On the energy efficiency of limited-backhaul cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.
[21] A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Towards optimal energy efficiency in cell-free massive MIMO systems," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 2, pp. 816–831, Jun. 2021.
[22] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink spectral and energy efficiency of cell-free massive MIMO with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 223–245, Jan. 2021.
[23] D. L. Galappaththige and G. A. A. Baduge, "Exploiting underlay spectrum sharing in cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7470–7488, Nov. 2021.
[24] F. Rezaei, C. Tellambura, and A. Tadaion, "Rate enhancement for distributed massive MIMO systems with underlay spectrum sharing," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–5.
[25] F. Rezaei, A. R. Heidarpour, C. Tellambura, and A. Tadaion, "Underlaid spectrum sharing for cell-free massive MIMO-NOMA," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 907–911, Apr. 2020.
[26] R. Saif, Z. Pourgharehkhan, S. ShahbazPanahi, M. Bavand, and G. Boudreau, "Underlay spectrum sharing in massive MIMO systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 9, no. 3, pp. 1–14, Jun. 2023.
[27] V. Kumar, M. F. Flanagan, R. Zhang, and L.-N. Tran, "Achievable rate maximization for underlay spectrum sharing MIMO system with intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1758–1762, Aug. 2022.
[28] T. Wang and R. Adve, "Fair licensed spectrum sharing between two MNOs using resource optimization in multi-cell multi-user MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6714–6730, Aug. 2022.
[29] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.
[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[31] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, 2021.
[32] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[33] P. K. Sangdeh, H. Pirayesh, A. Quadri, and H. Zeng, "A practical spectrum sharing scheme for cognitive radio networks: Design and experiments," *IEEE/ACM Trans. Netw.*, vol. 28, no. 4, pp. 1818–1831, Aug. 2020.
[34] D. Lopez-Perez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 653–697, 1st Quart., 2022.
[35] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.
[36] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
[37] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.
[38] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[39] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[40] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 1994.
[41] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.
[42] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. ICC-IEEE Int. Conf. Commun.*, May 2019, pp. 1–6.
[43] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, nos. 1–3, pp. 193–228, Nov. 1998.



ZAKIR HUSSAIN SHAIK (Graduate Student Member, IEEE) received the Bachelor of Technology (B.Tech.) degree in electronics and communication engineering from Jawaharlal Nehru Technological University (JNTU), Kakinada, India, in 2014, and the Master of Science (Research) degree in electronics and communication engineering from the Signal Processing and Communication Research Centre (SPCRC), International Institute of Information Technology (IIIT), Hyderabad, India, in 2019. He is currently pursuing the Ph.D. degree with the Division of Communication Systems, Department of Electrical Engineering, Linköping University, Sweden. From 2014 to 2016, he was an Assistant Systems Engineer with Tata Consultancy Services (TCS), Bengaluru, India. His research interests include cell-free massive MIMO, radio stripes, distributed processing, and spectrum sharing.

RIMALAPUDI SARVENDRANATH (Member, IEEE) received the Bachelor of Technology degree in electrical and electronics engineering from the National Institute of Technology Karnataka, Surathkal, in 2009, and the Master of Engineering and Ph.D. degrees from the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India, in 2012 and 2020, respectively. From 2012 to 2016, he was with Broadcom Communications Technologies, Bengaluru, where he worked on the development and implementation of algorithms for LTE and IEEE 802.11ac wireless standards. In 2021, he was a Postdoctoral Researcher with the Department of Electrical Engineering, Linköping University, Sweden. He was an Assistant Professor with the Electronics and Electrical Engineering Department, Indian Institute of Technology Guwahati, from January 2022 to June 2023. He is currently an Assistant Professor with the Department of Electrical Engineering, Indian Institute of Technology Tirupati. His research interests include machine learning for wireless communication, multiple antenna techniques, spectrum sharing, and next-generation wireless standards.

ERIK G. LARSSON (Fellow, IEEE) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 2002. He is currently Professor of Communication Systems at Linköping University (LiU) in Linköping, Sweden. He was with the KTH Royal Institute of Technology in Stockholm, Sweden, the George Washington University, USA, the University of Florida, USA, and Ericsson Research, Sweden. His main professional interests are within the areas of wireless communications and signal processing. He co-authored *Space-Time Block Coding for Wireless Communications* (Cambridge University Press, 2003) and *Fundamentals of Massive MIMO* (Cambridge University Press, 2016).

He served as chair of the IEEE Signal Processing Society SPCOM technical committee (2015–2016), chair of the IEEE Wireless Communications Letters steering committee (2014–2015), member of the IEEE Transactions on Wireless Communications steering committee (2019–2022), General and Technical Chair of the Asilomar SSC conference (2015, 2012), technical co-chair of the IEEE Communication Theory Workshop (2019), and member of the IEEE Signal Processing Society Awards Board (2017–2019). He was Associate Editor for, among others, the IEEE Transactions on Communications (2010–2014), the IEEE Transactions on Signal Processing (2006–2010), and the IEEE *Signal Processing Magazine* (2018–2022).

He received the IEEE Signal Processing Magazine Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, the IEEE ComSoc Best Tutorial Paper Award in 2018, the IEEE ComSoc Fred W. Ellersick Prize in 2019, and the IEEE SPS Donald G. Fink Overview Paper Award in 2023. He is a Fellow of the IEEE, a member of the Swedish Royal Academy of Sciences (KVA), and Highly Cited according to ISI Web of Science.