IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 11, 2020, accepted November 27, 2020, date of publication December 28, 2020,
date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047720

A Green Downlink Power Allocation Scheme for Cell-Free Massive MIMO Systems

SEYYED SALEH HOSSEINI[1], (Graduate Student Member, IEEE), BENOIT CHAMPAGNE[1], (Senior Member, IEEE), AND XIAO-WEN CHANG[2]
[1]Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada
[2]School of Computer Science, McGill University, Montreal, QC H3A 2A7, Canada

Corresponding author: Seyyed Saleh Hosseini (seyyed.hosseini@mail.mcgill.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN2017-05138 and Grant RGPIN-2017-04223.

ABSTRACT In this article, we consider the problem of downlink power allocation in a cell-free massive multiple-input multiple-output (m-MIMO) communication system under spectral efficiency (SE) constraints for the users. From the perspective of green communications, the power allocation is formulated as an optimization problem where the aim is to maximize the sum SE as the objective function, while limiting the transmission power of each access point (AP) and imposing lower and upper bounds on the achievable SEs of different users. The resulting optimization problem is non-convex since the objective function is non-concave and the upper bounding constraints on user SEs are non-convex. To address these difficulties, we first derive a closed-form lower bound on the sum SE (objective function) and prove that it is a quasi-concave function. Then, we relax the unwieldy upper bounding constraints on the user SEs by replacing them with linear functions, which renders the optimization problem convex. An optimal solution to the relaxed problem is finally obtained by solving a sequence of convex feasibility programs. We evaluate the performance of the proposed downlink power allocation scheme through Monte Carlo simulations under the uncorrelated and correlated shadow fading models. The results show that for both models, the proposed algorithm can lead to a significant reduction in total power consumption compared to a benchmark approach, while accurately allocating power to the APs so that the SE constraint of each user is satisfied within the imposed bounds.

INDEX TERMS Cell-free massive MIMO, downlink power allocation, green communications, spectral efficiency.

I. INTRODUCTION

Cell-free massive multiple-input multiple-output (m-MIMO) systems have been recently proposed as a promising technology for the next generation of wireless communication systems [1]. This state-of-the-art technology employs a large number of access points (APs) which are distributed over a wide area without partitioning the latter into bounded cells. The APs are connected to a central processing unit (CPU) via a backhaul network of high-speed links. The CPU performs several key functions, including: downlink power allocation for the APs and pilot assignment to the users; information exchange with the APs, such as payload data and power control coefficients; use of APs received (and preprocessed) signals to perform detection in uplink transmissions [1].

Cell-free massive MIMO systems offer several appealing features which have attracted considerable attention in recent years. For example, using the decentralized transmission (reception) strategy, i.e., sending (receiving) information signals from a large number of different geographically-located APs, leads to a considerable shadowing diversity in the downlink (uplink) transmissions. More importantly, multi-user interferences are mitigated owing to the desirable phenomena of *channel hardening* and *favorable propagation* in the context of m-MIMO systems [1], [2]. The mutual coupling effect, which is a common issue in collocated m-MIMO systems [3]–[5], can also be alleviated since the large-scale array gain is achieved by a large number of separated APs which are normally equipped with only a small number of antennas.

The associate editor coordinating the review of this manuscript and approving it for publication was Yunlong Cai.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

The concept of cell-free m-MIMO systems was first introduced in [1], where it was proved that pilot contamination, similar to the collocated m-MIMO systems [6]–[10], is the fundamental performance limitation. Moreover, the max-min power allocation problems for both downlink and uplink transmissions were solved, leading to significant improvements in achievable spectral efficiency (SE). The problem of optimizing energy efficiency (EE) was formulated and solved in [11], where it was shown that a cell-free m-MIMO system with N single-antenna APs can be considerably more efficient than a collocated m-MIMO system with one AP and N transmit antennas. In order to further enhance the performance of max-min power allocation, the authors in [12] introduced a filtering block at the receiver side and jointly optimized the filter and power control coefficients. Considering that the per user downlink SE is expected to reach 30 bits/s/Hz in fifth generation (5G) wireless networks [13], the limited backhaul capacity represents a significant bottleneck for distributed applications. To address this issue in cell-free m-MIMO systems, the problem of optimal power allocation under the constraint of limited-capacity backhaul was formulated and solved using different objectives, including SE and EE maximization, in [14]–[19]. From the perspective of network physical layer security, cell-free m-MIMO systems are susceptible to pilot spoofing attacks during the uplink training phase. To mitigate these attacks, the problem of power control against active eavesdropping, with the aim of maximizing the legitimate user SE, was formulated and addressed in [20]–[24].

As seen from the above literature review, the problem of power allocation in cell-free m-MIMO systems has been studied from different angles. However, it has not been addressed yet in the presence of additional bounding constraints on the user SEs (either from below, above or both). The key reason for solving such a constrained power allocation problem is two-fold: *i*) In wireless applications, different users will typically have different SE requirements (target SEs), which justifies imposing a lower bound upon the user SEs [25]; *ii*) In a wireless network, the total power consumption is proportional to the user SEs and hence, can be reduced considerably by allocating a smaller portion of the available power to the users with low SE targets. This goal is not necessarily attainable by solving a power allocation problem in which only the user SEs are lower bounded since a user, who does not really require it, can achieve a high SE especially when its channel quality is good. Therefore, solving a power allocation problem with upper bounded user SEs can prevent unnecessary waste of power in a cell-free m-MIMO system. Note that the importance of saving power is more pronounced as the spectral bandwidth increases, since the backhaul power consumption, which is a major part of total power consumption, is proportional to this bandwidth [11]. Moreover, in some scenarios, by using SE bounding constraints, the quality of service (in terms of achievable SE) can be enhanced for users who require high SE but are located in periphery of the network coverage area.

To illustrate the last point, let us consider a scenario in which a group of users with good AP coverage is being served with low to moderate data-rate applications, e.g., voice calls or online gaming, while a second group of users located far away from the APs is being served with a high-data-rate service, e.g., video streaming or conferencing.[1] In this scenario, the first group of users might achieve needlessly high SEs as their received signals undergo a lower level of path loss, while the second group of users might only marginally achieve their SE targets, which is not desirable due to the possible outage. Moreover, the total power consumption might unnecessarily increase owing to the fact that a large amount of power is allocated to the first group of users.

Motivated by the above considerations, we address in this work the problem of downlink power allocation for cell-free m-MIMO systems in which the achievable user SEs are lower and upper bounded. This optimization problem can be labeled as a "green power allocation" for these systems since its solution leads to a total power consumption that is proportional to the users' data rate requirements. The main contributions of this article are as follows:

- We present a new formulation of the power allocation problem as a constrained optimization, where the aim is to maximize the sum SE as the objective function, while limiting the transmission power of each AP and imposing lower and upper bounds on the achievable SEs of different users. We show that the resulting optimization problem is non-convex, i.e., the objective function is non-concave and the upper bounding constraints on the user SEs are non-convex.
- To deal with non-convex difficulties, we first derive a closed-form lower bound on the sum SE (objective function) and show that it is quasi-concave. Then, we relax the unwieldy constraints on the user SEs by replacing them with linear functions, which renders the optimization problem convex. An optimal solution to the relaxed problem is finally obtained by solving a sequence of convex feasibility programs.
- The computational complexity of the proposed algorithms is analyzed in terms of the key system parameters. We also discuss possible mechanisms for informing the CPU about the lower and upper bounds on the achievable SE of each user in practical applications.
- We evaluate the performance of the proposed downlink power allocation scheme through Monte Carlo simulations under the uncorrelated and correlated shadow fading models. The results show that for both models, the proposed suboptimal algorithm can lead to a significant reduction in total power consumption compared to a benchmark approach, while accurately allocating power to the APs so that the SE constraint of each user is satisfied within the imposed bounds.

[1] In practice, this may occur if the second group of users is located at some distance from a campus or residential complex, where the deployment of APs is no longer possible due to the lack of infrastructure.

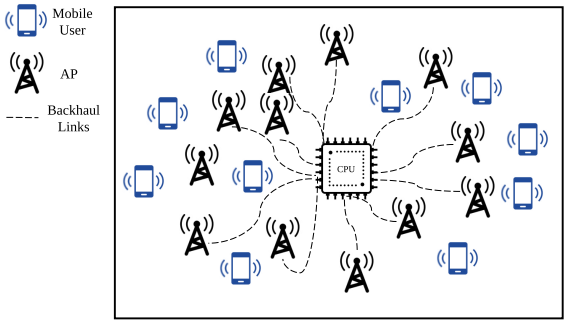


FIGURE 1. A cell-free massive MIMO system.

- Finally, it is shown that for a setting similar to the aforementioned scenario, imposing upper bounds upon the user SEs of the first group (with good AP coverage) brings two advantages: users in the second group (far away from the APs) achieves SEs that exceed the desired target by a sufficient margin, while the power consumed for the first group is substantially reduced.

The rest of this article is organized as follows: The detailed system model is presented in Section II. The problem of downlink power allocation with bounded user SEs is formulated in Section III, where the suboptimal problem and its algorithmic solution are developed. Simulation and numerical results are reported in Section IV. Finally, a conclusion is drawn in Section V.

Notation: The capital and small boldface letters indicate matrices and vectors, respectively. The real and circular complex Gaussian random variables x and z with means μ and variances σ^2 are represented by $x \sim \mathcal{N}(\mu, \sigma^2)$ and $z \sim \mathcal{CN}(\mu, \sigma^2)$, respectively. The circular complex Gaussian random vector x with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{R}$ is represented by $\mathsf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{R})$. For a random vector x, $\mathbb{E}_{\mathsf{x}}\{.\}$ denotes the expectation of its argument over the distribution of x. The symbols $\mathbf{0}_n$, $\mathbf{1}_n$, and $\mathbf{I}_n$ stand for the $n \times 1$ zero vector, the $n \times 1$ vector with all entries one, and the $n \times n$ identity matrix, respectively. We denote the ith standard basis vector in $\mathbb{R}^n$ by $\mathbf{e}_i$. For two $n \times n$ matrices (or vectors) $\mathbf{A}$ and $\mathbf{B}$, the Hadamard product and division are denoted by $\mathbf{A} \odot \mathbf{B}$ and $\mathbf{A} \oslash \mathbf{B}$, respectively. The operation diag(.) converts a vector or a list of matrices into a diagonal or a block diagonal matrix, respectively. The greatest (resp. smallest) integer less (resp. greater) than or equal to x is denoted by $\lfloor x \rfloor$ (resp. $\lceil x \rceil$).

II. SYSTEM MODEL

We consider a cell-free m-MIMO[2] system consisting of N APs, each connected to a CPU via backhaul links and together serving K users ($K < N$) across a relatively large geographical area (Fig.1). For simplicity, we assume that all APs are equipped with a single antenna while generalization of our approach to the case of multiple-antenna APs is possible. No direct cooperation is assumed between the APs while the CPU distributes the downlink payload data, pilot indices, and power control coefficients among APs through very high capacity backhaul links. In this work, we assume that there is no signal distortion due to quantization errors over backhaul links between the CPU and APs.

It is assumed that different mobile users are being provided with particular services requiring different data transmission rates [25]. For example, a user might be watching a full high-definition YouTube video on her iPhone while a roaming user is watching a 4K high-dynamic-range video in his car through an installed Apple TV. We assume that the system employs the time-division duplexing (TDD) technique for data transmission so that channel reciprocity holds true, which is valid when the duplexing period is smaller than the channel coherence time and the radio frequency chains at the transmitter and receiver sides are accurately calibrated [26].

We use the complex-valued baseband model for the representation of either narrowband or wideband signal transmission over the wireless channel. For this model, we let T_c and B_c denote the channel coherence time and coherence bandwidth, respectively. Using these parameters, the time-frequency plane is partitioned into a rectangular grid of so-called *coherence elements*, each corresponding to a time duration T_c and frequency bandwidth B_c. According to the sampling theorem, each coherence element can be used to represent $\tau_c = \lfloor T_c B_c \rfloor$ independent complex-valued signal samples; hence, we refer to τ_c has the *dimension* of a coherence element [27, p. 22].[3] The complex channel gain affecting the time-frequency samples contained in any given coherence element can be decomposed into a product of two components, corresponding respectively to small-scale and large-scale fading. The small-scale fading component is assumed to remain constant over the coherence element and change independently from one coherence element to the next. The large-scale fading component, which includes path loss and shadow fading, is assumed to change slowly over the time and frequency domains, and therefore remains nearly constant across several contiguous coherence elements. For instance, in [1], it is assumed that the large-scale fading component remains constant for about $40\tau_c$ in a high-mobility environment. Hence, the complex channel gain between the nth AP and the kth user within any coherence element can be written as follows:

$$g_{nk} = \beta_{nk}^{\frac{1}{2}} h_{nk}, \quad n \in \mathcal{N},\ k \in \mathcal{K} \tag{1}$$

where $\beta_{nk}^{\frac{1}{2}}$ and $h_{nk} \sim \mathcal{CN}(0, 1)$ are the large-scale and small-scale fading components, respectively, and the sets $\mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{K} = \{1, \ldots, K\}$. The coefficient β_{nk} is, in turn, defined as $\beta_{nk} = a_{nk} 10^{\frac{\sigma_{\text{sh}} z_{nk}}{10}}$ where a_{nk} represents the path loss, $z_{nk} \sim \mathcal{N}(0, 1)$ is the shadow fading magnitude in dB, and σ_{sh}^2 is the shadowing variance.

[2] In this work, the term massive MIMO refers to the use of a large number of APs which are distributed over the network, as opposed to the use of a large number of antennas at one or more APs.

[3] In the literature on m-MIMO, the terms coherence interval and length of coherence interval are also used, respectively, to denote a coherence element and its dimension, as defined here.

As in [1], [11], [21], [28], we assume that the parameters β_{nk} for all n, k are available at the CPU and APs for the purpose of pilot assignment to users, downlink power allocation, channel estimation, and precoding. We also assume channel gains g_{nk} corresponding to different APs or users are independent.

For a cell-free m-MIMO system, the data transmission in the downlink is divided into two phases [1], [11]. In the first phase, referred to as uplink training, the users simultaneously send their preassigned pilot signals to all APs. Each AP then applies a suboptimal minimum mean squared error (MMSE) method to the received pilot signals to estimate the user channels, as needed for the implementation of a precoding scheme in the next phase. In the second phase, referred to as downlink payload data transmission, the APs precode the payload signals using the conjugate beamforming (CB) scheme and transmit the resultant signals to all users. These two phases are explained in further detail below.

A. UPLINK TRAINING PHASE

Let $\boldsymbol{\phi}_k \in \mathbb{C}^{\tau_p}$ denote the vector of τ_p pilot samples transmitted by the kth user, where $\|\boldsymbol{\phi}_k\|_2^2 = 1$ and $\tau_p < \tau_c$. Then, the received pilot vector by the nth AP can be written as follows:

$$\mathsf{y}_{p,n} = \sqrt{\rho_p \tau_p} \sum_{k=1}^{K} g_{nk} \boldsymbol{\phi}_k + \mathsf{v}_{p,n}, \tag{2}$$

where ρ_p denotes the pilot transmit signal-to-noise ratio (SNR) and $\mathsf{v}_{p,n} \sim \mathcal{CN}(\mathbf{0}_{\tau_p}, \mathbf{I}_{\tau_p})$ is an additive complex circular Gaussian noise vector. By applying a suboptimal MMSE-based method to the preprocessed signal $\boldsymbol{\phi}_k^H \mathsf{y}_{p,n}$ and assuming that the large-scale fading components β_{nk} for all n, k are known at the APs, the conditional estimation of g_{nk} at the nth AP can be obtained as follows [1], [11], [12], [14]–[17]:

$$\hat{g}_{nk} = \frac{1}{\beta_{nk}\sqrt{\rho_p \tau_p}} \gamma_{nk} \boldsymbol{\phi}_k^H \mathsf{y}_{p,n}, \tag{3}$$

where

$$\begin{aligned} \gamma_{nk} &= \mathbb{E}_{h,\mathsf{v}}\left\{ |\hat{g}_{nk}|^2 \middle| \beta_{n,k} \right\} \\ &= \frac{\rho_p \tau_p \beta_{nk}^2}{\rho_p \tau_p \sum_{\ell=1}^{K} \beta_{n\ell} |\boldsymbol{\phi}_\ell^H \boldsymbol{\phi}_k|^2 + 1}, \end{aligned} \tag{4}$$

and the expectation is taken over the channel and noise distributions.

B. DOWNLINK PAYLOAD DATA TRANSMISSION PHASE

In this phase, the CPU first distributes the downlink payload signals among all N APs via backhaul links. Each AP treats the channel gain estimates, given by (3), as true channels and applies the CB to precode the payload signals. Then, the precoded signals are transmitted simultaneously from all APs to all users. Hence, the received signal by the kth user can be expressed as follows:

$$\begin{aligned} y_{d,k} &= \sum_{n=1}^{N} \underbrace{\sqrt{\rho_d} \eta_{nk}^{\frac{1}{2}} g_{nk} \hat{g}_{nk}^*}_{\equiv g_{d,nk}^{\text{eff}}} q_k \\ &\quad + \sqrt{\rho_d} \sum_{n=1}^{N} \sum_{\ell \neq k}^{K} \eta_{n\ell}^{\frac{1}{2}} g_{nk} \hat{g}_{n\ell}^* q_\ell + v_{d,k}, \end{aligned} \tag{5a}$$

$$= g_{d,k}^{\text{eff}} q_k + I_{d,k} + v_{d,k}, \tag{5b}$$

where ρ_d is the downlink transmit SNR, q_k denotes the signal sample intended for the kth user, with zero-mean and variance $\mathbb{E}\{|q_k|^2\} = 1$, and $v_{d,k} \sim \mathcal{CN}(0, 1)$ is an additive complex circular Gaussian noise. In (5a), the parameters η_{nk} represent power control coefficients which must satisfy the following constraints:

$$\eta_{nk} \geq 0, \quad n \in \mathcal{N}, \ k \in \mathcal{K} \tag{6a}$$

$$\sum_{k=1}^{K} \gamma_{nk} \eta_{nk} \leq 1, \quad n \in \mathcal{N} \tag{6b}$$

where the latter inequality simply reflects the fact that the total average transmission power of each AP is limited to ρ_d. Referring to (5b), one can interpret the coefficient of q_k, defined as $g_{d,k}^{\text{eff}}$, as the overall *effective channel* of the kth user, while the second term $I_{d,k}$ represents the total interference caused by the signals intended for all the other users, i.e. with $\ell \neq k$.

In order to detect signal q_k from $y_{d,k}$ in (5) while avoiding the need for a downlink training phase, the kth user adopts a channel-hardening approach [1], [11]. To this end, it is assumed that the user knows the expected value of its effective channel, i.e., $\bar{g}_{d,k}^{\text{eff}} = \mathbb{E}_{h,\mathsf{v}}\{g_{d,k}^{\text{eff}}\}$ and uses it merely, instead of the true value of $g_{d,k}^{\text{eff}}$, to detect q_k via, e.g., an MMSE method [1], [11]. This assumption is plausible since in a cell-free m-MIMO system, the number N of APs is large and the channel gains $g_{d,nk}^{\text{eff}}$ are nearly independent, so that according to the law of large numbers [29, p. 185], the overall effective channel $g_{d,k}^{\text{eff}}$ is close to its mean $\bar{g}_{d,k}^{\text{eff}}$.[4] In light of the channel-hardening approach for detection, it is convenient to rewrite (5) as follows:

$$y_{d,k} = \bar{g}_{d,k}^{\text{eff}} q_k + v_{d,k}^{\text{eff}}, \tag{7}$$

where

$$v_{d,k}^{\text{eff}} = \left(g_{d,k}^{\text{eff}} - \bar{g}_{d,k}^{\text{eff}} \right) q_k + I_{d,k} + v_{d,k},$$

can be construed as the *effective noise* corrupting the received signal at the kth user. The latter consists of three main sources of error: the difference between the expectation and true values of the effective channel, which is referred to as beamforming gain uncertainty [1], the multi-user interference mentioned previously, and the additive receiver noise. All these sources of error are taken into consideration in our subsequent analysis in Section III.

[4] Let $\mathbb{E}\{g_{d,nk}^{\text{eff}}\} < \infty$ and $\text{Var}\{g_{d,nk}^{\text{eff}}\} < \infty \ \forall n$. Then for any $\epsilon > 0$, $\lim_{N \to +\infty} \mathbb{P}\left(|\frac{1}{N} g_{d,k}^{\text{eff}} - \bar{g}_{d,k}^{\text{eff}}| > \epsilon \right) = 0$.

III. POWER ALLOCATION WITH BOUNDED USER SEs

In this section, we first formulate the problem of downlink power allocation in a cell-free m-MIMO system by maximizing the sum SE while limiting the transmission power of the APs and imposing lower and upper bounds on the achievable SEs of different users. Unfortunately, the objective function is non-concave and the upper bounding constraints on user SEs are non-convex. To address these difficulties, we then derive a closed-form lower bound on the sum SE and prove that it is quasi-concave; we also apply linear relaxation to the upper bounding constraints on the user SEs. An optimal solution to the resulting convex problem is then obtained by solving a sequence of convex feasibility programs. Finally, we analyze the computational complexity of the given algorithm and discuss practical mechanisms for informing the CPU about the lower and upper bounds on the SE of each user.

A. PROBLEM FORMULATION

Let R_k denote the achievable downlink SE of the kth user in a cell-free m-MIMO system with N APs and K users. For the signal model given in (7), it is shown in [1] that R_k can be expressed as follows:

$$R_k = \log_2\left(1 + \frac{N_k}{D_k}\right), \tag{8}$$

where

$$N_k = \rho_{\mathrm{d}}\left(\sum_{n=1}^{N}\gamma_{nk}\eta_{nk}^{\frac{1}{2}}\right)^2, \tag{9}$$

$$D_k = \rho_{\mathrm{d}}\sum_{\ell\neq k}^{K}|\boldsymbol{\phi}_\ell^H\boldsymbol{\phi}_k|^2\left(\sum_{n=1}^{N}\gamma_{n\ell}\frac{\beta_{nk}}{\beta_{n\ell}}\eta_{n\ell}^{\frac{1}{2}}\right)^2 + \rho_{\mathrm{d}}\sum_{\ell=1}^{K}\sum_{n=1}^{N}\gamma_{n\ell}\beta_{nk}\eta_{n\ell} + 1. \tag{10}$$

In (8), N_k and D_k respectively denote the average power of the desired signal and that of the effective noise $v_{\mathrm{d},k}^{\mathrm{eff}}$, as previously defined in (7).

To express the numerator N_k in a more compact form, we define

$$\boldsymbol{\gamma}_k = [\gamma_{1k}, \gamma_{2k}, \ldots, \gamma_{Nk}]^T \in \mathbb{R}^N,$$

$$\boldsymbol{\varsigma}_k = [\varsigma_{1k}, \varsigma_{2k}, \ldots, \varsigma_{Nk}]^T \in \mathbb{R}^N, \varsigma_{nk} = \eta_{nk}^{\frac{1}{2}}$$

and rewrite (9) as follows:

$$N_k = \rho_d(\boldsymbol{\gamma}_k^T\boldsymbol{\varsigma}_k)^2. \tag{11}$$

Also, by defining

$$\phi_{\ell k} = |\boldsymbol{\phi}_\ell^H\boldsymbol{\phi}_k|,\ \boldsymbol{\beta}_k = [\beta_{1k}, \beta_{2k}, \ldots, \beta_{Nk}]^T \in \mathbb{R}^N,$$
$$\boldsymbol{\varsigma} = [\boldsymbol{\varsigma}_1^T, \boldsymbol{\varsigma}_2^T, \ldots, \boldsymbol{\varsigma}_K^T]^T \in \mathbb{R}^{NK}$$

the denominator D_k can be rewritten as follows:

$$\begin{aligned}
D_k &= \rho_d\sum_{\ell=1,\ell\neq k}^{K}\phi_{lk}^2\left(\sum_{n=1}^{N}(\beta_{nk}\gamma_{n\ell}\beta_{n\ell}^{-1})\cdot\varsigma_{n\ell}\right)^2 + \rho_d\sum_{\ell=1}^{K}\sum_{n=1}^{N}\varsigma_{n\ell}\cdot(\beta_{nk}\gamma_{n\ell})\cdot\varsigma_{n\ell} + 1\\
&= \rho_d\sum_{\ell=1,\ell\neq k}^{K}\phi_{\ell k}^2\left((\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell\oslash\boldsymbol{\beta}_\ell)^T\boldsymbol{\varsigma}_\ell\right)^2 + \rho_d\sum_{\ell=1}^{K}\boldsymbol{\varsigma}_\ell^T\mathrm{diag}(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell)\boldsymbol{\varsigma}_\ell + 1\\
&= \rho_d\sum_{\ell=1,\ell\neq k}^{K}\boldsymbol{\varsigma}_\ell^T\phi_{\ell k}^2(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell\oslash\boldsymbol{\beta}_\ell)(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell\oslash\boldsymbol{\beta}_\ell)^T\boldsymbol{\varsigma}_\ell + \rho_d\sum_{\ell=1}^{K}\boldsymbol{\varsigma}_\ell^T\mathrm{diag}(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell)\boldsymbol{\varsigma}_\ell + 1\\
&= \rho_d\,\boldsymbol{\varsigma}^T\boldsymbol{\Omega}_k\boldsymbol{\varsigma} + 1.
\end{aligned} \tag{12}$$

where

$$\begin{aligned}
\boldsymbol{\Omega}_k &= \mathrm{diag}(\boldsymbol{\Omega}_{1,k}, \ldots, \boldsymbol{\Omega}_{K,k}) \in \mathbb{R}^{NK\times NK},\\
\boldsymbol{\Omega}_{k,k} &= \mathrm{diag}(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_k) \in \mathbb{R}^{N\times N},\\
\boldsymbol{\Omega}_{\ell,k} &= \phi_{\ell k}^2(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell\oslash\boldsymbol{\beta}_\ell)(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell\oslash\boldsymbol{\beta}_\ell)^T + \mathrm{diag}(\boldsymbol{\beta}_k\odot\boldsymbol{\gamma}_\ell) \in \mathbb{R}^{N\times N}.\ell\neq k
\end{aligned}$$

We assume that for each user k, R_k is bounded from below and above, as expressed by $0 < R_k^{\min} \leq R_k \leq R_k^{\max}$, where the lower and upper limits $R_k^{\min}$ and $R_k^{\max}$, respectively are dependent upon the service provided to the kth user. By considering these bounds together with the constraints given in (6), we formulate the downlink power allocation problem in cell-free m-MIMO with constraints on the user SEs as follows:

$$\mathcal{P}_1:\begin{cases}
\max\limits_{\boldsymbol{\varsigma}}\ \sum\limits_{k=1}^{K}\log_2\left(1+\dfrac{(\boldsymbol{\gamma}_k^T\boldsymbol{\varsigma}_k)^2}{\boldsymbol{\varsigma}^T\boldsymbol{\Omega}_k\boldsymbol{\varsigma}+\rho_d^{-1}}\right) & \text{(13a)}\\
\text{s.t. } R_k^{\min} \leq \log_2\left(1+\dfrac{(\boldsymbol{\gamma}_k^T\boldsymbol{\varsigma}_k)^2}{\boldsymbol{\varsigma}^T\boldsymbol{\Omega}_k\boldsymbol{\varsigma}+\rho_d^{-1}}\right) \leq R_k^{\max}, \forall k & \text{(13b)}\\
\boldsymbol{\varsigma}^T\boldsymbol{\Lambda}_n\boldsymbol{\varsigma} \leq 1,\ \ \forall n & \text{(13c)}\\
0 \leq \boldsymbol{\varsigma} & \text{(13d)}
\end{cases}$$

where $\boldsymbol{\Lambda}_n = \mathrm{diag}\big((\mathbf{e}_n^T\boldsymbol{\gamma}_1)\mathbf{e}_n\mathbf{e}_n^T, (\mathbf{e}_n^T\boldsymbol{\gamma}_2)\mathbf{e}_n\mathbf{e}_n^T, \ldots, (\mathbf{e}_n^T\boldsymbol{\gamma}_K)\mathbf{e}_n\mathbf{e}_n^T\big)$.

Note that (13c) is the compact form of the total power constraint in (6b). In the literature a common choice of the objective for maximizing the SE of MIMO systems is the minimum SE over the users, see, e.g., [1], [12], [28]. But here we use the sum of all SEs as the objective function in $\mathcal{P}_1$ to maximize the overall performance. Note that the lower bound

$R_k^{\min}$ that we imposed on each R_k can avoid the risk of using this objective function, i.e., the resulting R_k being too small for some k.

Let us ignore constraints (13b) in problem $\mathcal{P}_1$ for the moment. The resulting problem is essentially a fractional programming which cannot be solved straightforwardly as its objective function is non-concave. To tackle the non-concavity of problem $\mathcal{P}_1$ without considering constraints (13b), one can directly apply the quadratic transformation technique proposed in [30]. By doing so, the objective function converts into an equivalent function with some auxiliary variables, which is concave when these variables are fixed. The equivalent problem can be solved via a block coordinate ascent method as suggested in [30]. Although this approach overcomes the non-concavity difficulty of problem $\mathcal{P}_1$ without constraints (13b), it may not produce the desired solution since it only guarantees the convergence to a stationary point.

In order to address the non-concavity of problem $\mathcal{P}_1$, we first derive a lower bound for its objective function over the feasible set and show that the obtained lower bound is a quasi-concave function. Then, we replace the objective function in $\mathcal{P}_1$ with the derived lower bound and formulate a new problem where the objective function is a quasi-concave function. The resulting problem cannot be directly solved as the second inequality of constraint (13b) defines a non-convex set. To overcome this issue, we shall relax the unwieldy constraint by a linear function and solve the relaxed problem by a sequence of convex feasibility problems, as explained below.

B. LOWER BOUND DERIVATION AND OPTIMIZATION ALGORITHM

We first present the following theorem, which states the existence of a lower bound for the objective function of problem $\mathcal{P}_1$ and also shows that the derived lower bound is a quasi-concave function over its feasible set.

Theorem 1: Let R_k be bounded from below and above by $R_k^{\min}$ and $R_k^{\max}$, respectively where $R_k^{\min} \geq 1$. Then, the objective function of problem $\mathcal{P}_1$ can be lower bounded over its feasible set by

$$C \log_2 \left(1 + \frac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}}} \right), \tag{14}$$

where C is a positive constant independent of optimization variables. Moreover, the derived lower bound in (14) is a quasi-concave function of $\boldsymbol{\varsigma}$ over an arbitrary convex set $\mathbb{S}$ where $\mathbb{S} \subset \mathbb{R}_+^{NK}$.

Proof: To simplify notation, we introduce

$$\widetilde{\boldsymbol{\Omega}}_k = \begin{bmatrix} \boldsymbol{\Omega}_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad \widetilde{\boldsymbol{\varsigma}} = \begin{bmatrix} \boldsymbol{\varsigma} \\ \rho_d^{-1/2} \end{bmatrix}.$$

Let $\mathbf{L}_k$ be the Cholesky factor of $\widetilde{\boldsymbol{\Omega}}_k$, which is symmetric positive definite. Define the norm

$$\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} = \sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}} = \|\mathbf{L}_k^T \widetilde{\boldsymbol{\varsigma}}\|_2. \tag{15}$$

By recalling the objective function of problem $\mathcal{P}_1$, we have the lower bound in (16) which is presented at the bottom of the next page. In (16), (a) directly results from the assumption $R_k^{\min} \geq 1$ in Theorem 1 and (b) follows from the well-known *log-sum inequality*.[5] Now, we define

$$k_{\mathrm{m}} = \arg\min_k \left[(1+\alpha_k) \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} \right],$$
$$k_{\mathrm{M}} = \arg\max_k \left[(1+\alpha_k) \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} \right]$$

where $\alpha_k = \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}}$. Then, we can derive a lower bound for the second factor on the last line in (16) as follows:

$$\begin{aligned} \frac{\min_k \left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} \right)}{\max_k \left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} \right)} &= \frac{(1+\alpha_{k_{\mathrm{m}}}) \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_{k_{\mathrm{m}}}}}{(1+\alpha_{k_{\mathrm{M}}}) \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_{k_{\mathrm{M}}}}} \\ &\overset{(a)}{\geq} \frac{1+\sqrt{2^{R_{k_{\mathrm{m}}}^{\min}} - 1}}{1+\sqrt{2^{R_{k_{\mathrm{M}}}^{\max}} - 1}} \frac{\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_{k_{\mathrm{m}}}}}{\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_{k_{\mathrm{M}}}}} \\ &\overset{(b)}{\geq} \frac{1+\sqrt{2^{R_{k_{\mathrm{m}}}^{\min}} - 1}}{1+\sqrt{2^{R_{k_{\mathrm{M}}}^{\max}} - 1}} \frac{\sigma_{\min}(\mathbf{L}_{k_{\mathrm{m}}}^T)}{\sigma_{\max}(\mathbf{L}_{k_{\mathrm{M}}}^T)}, \end{aligned} \tag{17}$$

where (a) results from $R_k^{\min} \leq R_k = \log_2(1+\alpha_k^2) \leq R_k^{\max}$, (b) is due to the fact that for given matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and vector $\mathbf{x} \neq 0 \in \mathbb{R}^n$, we have the inequality $\sigma_{\min}(\mathbf{A}) \leq \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \sigma_{\max}(\mathbf{A})$ where $\sigma_{\min}(\mathbf{A})$ and $\sigma_{\max}(\mathbf{A})$ are the minimum and maximum singular values of matrix $\mathbf{A}$. As seen from (17), the obtained expression is independent of the optimization variables and can be defined as the constant C/K.

Now we prove that the derived lower bound is a quasi-concave function. From the definition of a quasi-concave function, the t-superlevel set of $\log_2 \left(1 + \frac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}} \right)$, i.e.,

$$\mathcal{A} = \left\{ \boldsymbol{\varsigma} \in \mathbb{S} \middle| \sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k \geq (2^t - 1) \sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} \right\}, \tag{18}$$

must be convex for any $t \in \mathbb{R}_+$. Taking any $\boldsymbol{\varsigma}^{(1)}$ and $\boldsymbol{\varsigma}^{(2)}$ inside the set $\mathcal{A}$, we show that the line segment between $\boldsymbol{\varsigma}^{(1)}$ and $\boldsymbol{\varsigma}^{(2)}$ lies inside the set $\mathcal{A}$. In fact, for $0 \leq \lambda \leq 1$,

$$\begin{aligned} &\frac{1}{(2^t-1)} \sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \left[\lambda \boldsymbol{\varsigma}_k^{(1)} + (1-\lambda) \boldsymbol{\varsigma}_k^{(2)} \right] \\ &= \frac{\lambda}{(2^t-1)} \sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k^{(1)} + \frac{(1-\lambda)}{(2^t-1)} \sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k^{(2)} \\ &\overset{(a)}{\geq} \lambda \sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}^{(1)}\|_{\widetilde{\boldsymbol{\Omega}}_k} + (1-\lambda) \sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}^{(2)}\|_{\widetilde{\boldsymbol{\Omega}}_k} \end{aligned}$$

[5] For nonnegative numbers, $a_1, \ldots, a_K$ and $b_1, \ldots, b_K$, we have $\sum_{k=1}^{K} a_k \log \frac{a_k}{b_k} \geq \left(\sum_{k=1}^{K} a_k \right) \log \frac{\sum_{k=1}^{K} a_k}{\sum_{k=1}^{K} b_k}$, with equality if and only if $\frac{a_k}{b_k} = c$ for for all k's [31, p. 31].

$$\overset{(b)}{\geq} \sum_{k=1}^{K} \|\lambda\widetilde{\boldsymbol{\varsigma}}^{(1)} + (1-\lambda)\widetilde{\boldsymbol{\varsigma}}^{(2)}\|_{\widetilde{\boldsymbol{\Omega}}_k}$$

$$= \sum_{k=1}^{K} \left\| \mathbf{L}_k^T \begin{bmatrix} \lambda\boldsymbol{\varsigma}^{(1)} + (1-\lambda)\boldsymbol{\varsigma}^{(2)} \\ \rho_d^{-1/2} \end{bmatrix} \right\|_2,$$

where $\widetilde{\boldsymbol{\varsigma}}^{(i)} = \begin{bmatrix} \boldsymbol{\varsigma}^{(i)} \\ \rho_d^{-1/2} \end{bmatrix}$ for $i = 1, 2$, the inequality (a) follows from the assumption that $\boldsymbol{\varsigma}^{(1)}$ and $\boldsymbol{\varsigma}^{(2)}$ belong to $\mathcal{A}$ and (b) results from the triangle inequality. Hence, $\lambda\boldsymbol{\varsigma}^{(1)}+(1-\lambda)\boldsymbol{\varsigma}^{(2)}$ belongs to set $\mathcal{A}$ which completes the proof. □

Remark 1: The lower bound given in (14) is a quasi-concave function which can be used as the objective function of $\mathcal{P}_1$. At first glance, the constraint $R_k^{\min} \geq 1$ (bits/s/Hz) seems to limit the applicability of the derived bound to problems where the target SEs are greater than one for all users. However, having SEs greater than one for all users is only a sufficient condition since the first inequality in (16) can be fulfilled even if only a subset of users, rather than all users, satisfies $R_k^{\min} \geq 1$. More precisely, we conduct a numerical analysis in Subsection IV-A where it is shown that if approximately a quarter of the users have a target SE meeting the above condition, the lower bound still holds with a probability which is very close to one. For 5G applications, it is very likely that such a fraction of users will run high-data-rate applications which require an SE of greater than one. Therefore, the lower bound given in (14) can safely substitute for the sum SE in $\mathcal{P}_1$. Moreover, we solve in Subsection IV-C power allocation problems whose objective functions are the derived lower bound and the values of $R_k^{\max}$ are assumed to be much less than one (between 0.1 and 0.2 bits/s/Hz) for all users. The results suggest that the derived lower bound can also substitute for the objective function of $\mathcal{P}_1$ under the worst-case condition, i.e., when $R_k^{\max} < 1$ for all users.

Considering the discussion in Remark 1, we formulate an alternative problem to $\mathcal{P}_1$ as follows:

$$\mathcal{P}_2 : \begin{cases} \max\limits_{\boldsymbol{\varsigma}} \ \log_2\left(1 + \dfrac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \|\mathbf{L}_k^T \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\|_2}\right) & \text{(19a)} \\ \text{s.t.} \quad \|\mathbf{L}_k^T \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\|_2 \geq \dfrac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}, \forall k & \text{(19b)} \\ \|\mathbf{L}_k^T \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\|_2 \leq \dfrac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\min}} - 1}}, \ \forall k & \text{(19c)} \\ \text{(13c)} - \text{(13d)} & \text{(19d)} \end{cases}$$

where we use (15) to rewrite (14) and constraints (13b). Note that $\mathcal{P}_2$ is still a non-concave problem as the inequality (19b) is non-convex, which is problematic. To overcome this difficulty, we replace (19b) with the following linear inequality:

$$\boldsymbol{\psi}_k^T \mathbf{L}_k \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T \geq \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}, \tag{20}$$

where $\boldsymbol{\psi}_k \in \mathbb{R}^{NK+1}$. By doing so, we can write the relaxed version of $\mathcal{P}_2$ for a given $\boldsymbol{\psi}_k$ as follows:

$$\mathcal{P}_2' : \begin{cases} \max\limits_{\boldsymbol{\varsigma}} \ \log_2\left(1 + \dfrac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \|\mathbf{L}_k^T \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\|_2}\right) & \text{(21a)} \\ \text{s.t.} \quad \boldsymbol{\psi}_k^T \left(\mathbf{L}_k \left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\right) \geq \dfrac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}, \forall k & \text{(21b)} \\ \text{(19c)} - \text{(19d)} & \text{(21c)} \end{cases}$$

Note that replacing constraint (19b) with the linear inequality in (20) does not affect the existence of the lower bound (14) over the feasible set of $\mathcal{P}_2'$. To see this, we have

$$\sum_{k=1}^{K} \log_2\left(1 + \frac{\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k\right)^2}{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}}\right) = \sum_{k=1}^{K} \log_2\left(1 + \frac{\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k\right)^2}{\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}^2}\right)$$

$$\overset{(a)}{\geq} \sum_{k=1}^{K} \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}}{\max_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)} \times \log_2\left(1 + \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k / \max_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)}{\|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k} / \max_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)}\right)$$

$$\overset{(b)}{\geq} \left(\sum_{k=1}^{K} \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}}{\max_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)}\right) \log_2\left(1 + \frac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}}\right)$$

$$\geq K \frac{\min_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)}{\max_k\left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}\right)} \log_2\left(1 + \frac{\sum_{k=1}^{K} \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^{K} \|\widetilde{\boldsymbol{\varsigma}}\|_{\widetilde{\boldsymbol{\Omega}}_k}}\right), \tag{16}$$

from (21b) that

$$\frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\|\boldsymbol{\psi}_k\|_2 \|\mathbf{L}_k\big[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\big]^T\|_2} \leq \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\boldsymbol{\psi}_k^T \left(\mathbf{L}_k\big[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\big]^T\right)} \leq \sqrt{2^{R_k^{\max}} - 1}. \tag{22}$$

Therefore, one can substitute the term $\sqrt{2^{R_k^{\max}} - 1}$ with $\sqrt{2^{R_{k_{\mathrm{M}}}^{\max}} - 1}\|\boldsymbol{\psi}_{k_{\mathrm{M}}}\|_2$ in the first inequality of (17) and re-derive the constant term of Theorem 1 for a given $\boldsymbol{\psi}_{k_{\mathrm{M}}}$. Problem $\mathcal{P}_2'$ is solvable since the objective function is quasi-concave and all constraints are either linear or convex. The optimal solution of $\mathcal{P}_2'$ can be obtained by applying a bi-section algorithm which solves a convex feasibility problem at each step. For a given value of the bi-section parameter t, the corresponding convex feasibility problem can be formulated as follows [32, p. 145]:

$$\mathcal{P}_2'': \begin{cases} \min\limits_{\boldsymbol{\varsigma}} 0 & \text{(23a)} \\ \text{s.t. } \log_2\left(1 + \dfrac{\sum_{k=1}^K \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sum_{k=1}^K \|\mathbf{L}_k^T \big[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\big]^T\|_2}\right) \geq \log_2(1+t) & \text{(23b)} \\ \boldsymbol{\psi}_k^T \left(\mathbf{L}_k\big[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\big]^T\right) \geq \dfrac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}, \forall k & \text{(23c)} \\ \text{(21c)} & \text{(23d)} \end{cases}$$

Problem $\mathcal{P}_2''$ can be solved by using an optimization package such as CVX [33].

Since the solution of $\mathcal{P}_2'$ does not necessarily satisfy the constraint (19b) in $\mathcal{P}_2$, a recursive algorithm which includes the bi-section algorithm as a procedure while updating the value of $\boldsymbol{\psi}_k$ at the end of each iteration can address this issue. To this end, the value of $\boldsymbol{\psi}_k$ can be updated for the $(i+1)$th iteration of the recursive algorithm as follows:

$$\boldsymbol{\psi}_k^{(i+1)} = \frac{\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right)^T, \rho_d^{-\frac{1}{2}}\big]^T}{\|\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right)^T, \rho_d^{-\frac{1}{2}}\big]^T\|_2}, \tag{24}$$

where $i \geq 0$ and $\boldsymbol{\varsigma}^{(i)}$ is the solution of the bi-section algorithm at the ith iteration. To see why the above update guarantees that constraint (19b) will be satisfied by the solution of the bi-section algorithm at the ith iteration, we can use the following inequalities:

$$\begin{aligned} \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k^{(i)}}{\sqrt{2^{R_k^{\max}} - 1}} &\leq \left(\boldsymbol{\psi}_k^{(i)}\right)^T \left(\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right)^T, \rho_d^{-\frac{1}{2}}\big]^T\right) \\ &\leq \|\boldsymbol{\psi}_k^{(i)}\|_2 \, \|\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right), \rho_d^{-\frac{1}{2}}\big]^T\|_2 \\ &\leq \|\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right), \rho_d^{-\frac{1}{2}}\big]^T\|_2, \end{aligned} \tag{25}$$

where $i \geq 1$ and the last inequality results from (24) and the fact that $\|\boldsymbol{\psi}_k^{(i)}\|_2 \leq 1$. The details of this proposed recursive algorithm are presented in Algorithm 1. To terminate the execution, two stopping criteria are integrated within Algorithm 1: *i*) reaching a preset maximum number of iteration, and *ii*) solution not improving significantly over two successive iterations. It can be understood from Algorithm 1 that the termination depends on the feasibility of $\mathcal{P}_2''$ which in turn, is dependent on parameters t, $R_k^{\min}$, $R_k^{\max}$, and $\boldsymbol{\psi}_k^{(0)}$. Below, we elaborate on the individual effect of each parameter over the feasibility of $\mathcal{P}_2''$, assuming other parameters are appropriately selected:

- The value of t depends on $t_{\min}$ and $t_{\max}$. To maximize a quasi-concave function, it is usual to choose $t_{\max}$ as a large number to avoid trapping the algorithm into a locally maximum point [32, p. 145]. However, based on our experiences, the optimal solution of $\mathcal{P}_2''$ is found with a small value of t for the various cases studied in this article. This turns out to be an advantage in terms of computational complexity for Algorithm 1 since one can choose a small value of $t_{\max}$, hence allowing to find the

Algorithm 1: Proposed Recursive Algorithm for Solving $\mathcal{P}_2$

Initialization: Set tolerances ϵ_1 and ϵ_2, number of iterations N_I, parameters $E > \epsilon_1$ and $i = 0$.
Choose appropriately $t_{\min} > 0$, $t_{\max} > t_{\min}$, and $\boldsymbol{\psi}_k^{(0)}$.
while $E \geq \epsilon_1$ and $i < N_I$ **do**
  Bi-section Algorithm:
  while $t_{\max} - t_{\min} \geq \epsilon_2$ **do**
    Set $t = \frac{t_{\min}+t_{\max}}{2}$ and solve the following convex feasibility problem:

$$\mathcal{P}_2'': \begin{cases} \min\limits_{\boldsymbol{\varsigma}} 0 \\ \text{s.t. } \sum_{k=1}^K \|\mathbf{L}_k^T \big[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\big]^T\|_2 \leq \frac{1}{t} \sum_{k=1}^K \boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k \\ \left(\boldsymbol{\psi}_k^{(i)}\right)^T \left(\mathbf{L}_k\big[\boldsymbol{\varsigma}^T \rho_d^{-\frac{1}{2}}\big]^T\right) \geq \dfrac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}, \forall k \\ \text{(21c)} \end{cases}$$

    if $\mathcal{P}_2''$ *is feasible* **then**
      $t_{\min} \leftarrow t$
      Save $\boldsymbol{\varsigma}^{(i)}$
    else
      $t_{\max} \leftarrow t$
    end
  end
  $E = \|\boldsymbol{\varsigma}^{(i)} - \boldsymbol{\varsigma}^{(i-1)}\|_2$
  $\boldsymbol{\psi}_k^{(i+1)} = \frac{\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right)^T, \rho_d^{-\frac{1}{2}}\big]^T}{\|\mathbf{L}_k\big[\left(\boldsymbol{\varsigma}^{(i)}\right)^T, \rho_d^{-\frac{1}{2}}\big]^T\|_2}$
  $i \leftarrow i + 1$.
end

optimal solution of $\mathcal{P}_2''$ by running the bi-section algorithm with fewer steps. The expression of the number of steps that a bi-section algorithm requires to compute an optimal solution in terms of $t_{\min}$, $t_{\max}$, and ϵ_2 is given in the next subsection.

- As long as $R_k^{\min} < R_k^{\max}$ and the value of $R_k^{\min}$ is not chosen larger than the maximum achievable SE (or the channel capacity), the problem is feasible and the bi-section algorithm can generate a solution. Note that $\mathcal{P}_2''$ will not be a convex feasibility problem if $R_k^{\min} = R_k^{\max}$ for the kth user as inequality constraints (19b) and (19c) will merge into a nonlinear equality constraint.
- If Algorithm 1 is initialized with a value of $\boldsymbol{\psi}_k^{(0)}$ which satisfies $\|\boldsymbol{\psi}_k^{(0)}\|_2 \leq 1$ and makes $\mathcal{P}_2''$ feasible, then, $\boldsymbol{\varsigma}^{(0)}$ will satisfy the constraint (19b). However, it is observed in practice that finding such an initial vector is a difficult task. To address this issue, we relax the constraint $\|\boldsymbol{\psi}_k^{(0)}\|_2 \leq 1$ and search for a $\boldsymbol{\psi}_k^{(0)}$ that makes $\mathcal{P}_2''$ feasible not only at the initial iteration but also at iteration $i = 1$. This condition is necessary since the updated vector $\boldsymbol{\psi}_k^{(1)}$ might not make $\mathcal{P}_2''$ feasible at iteration $i = 1$ for arbitrarily selected $\boldsymbol{\psi}_k^{(0)}$. Once the algorithm becomes feasible at iteration $i = 1$, the feasibility of $\mathcal{P}_2''$ will be guaranteed at the next iterations. To see this fact for iteration $i = 2$, it suffices to show that $\boldsymbol{\varsigma}^{(1)}$ satisfies the second inequality constraint of $\mathcal{P}_2''$. To this end, we use (24) to rewrite this inequality constraint as follows:

$$\left(\boldsymbol{\psi}_k^{(2)}\right)^T \left(\mathbf{L}_k\left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\right) = \left(\frac{\mathbf{L}_k\left[\left(\boldsymbol{\varsigma}^{(1)}\right)^T, \rho_d^{-\frac{1}{2}}\right]^T}{\|\mathbf{L}_k\left[\left(\boldsymbol{\varsigma}^{(1)}\right)^T, \rho_d^{-\frac{1}{2}}\right]^T\|_2}\right)^T \times \left(\mathbf{L}_k\left[\boldsymbol{\varsigma}^T, \rho_d^{-\frac{1}{2}}\right]^T\right) \geq \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{2^{R_k^{\max}} - 1}}. \quad (26)$$

By comparing (26) with (25) for $i = 1$, it will be revealed that $\boldsymbol{\varsigma}^{(1)}$ satisfies (26), and hence, $\mathcal{P}_2''$ is feasible in iteration $i = 2$. By using the mathematical induction, one can show that $\mathcal{P}_2''$ is feasible for $i \geq 2$.

C. PRACTICAL CONSIDERATIONS

The significant portion for the computational complexity of Algorithm 1 results from running the bi-section algorithm to solve the convex feasibility problem $\mathcal{P}_2''$ at each iteration. The required number of steps for the execution of the bi-section algorithm is equal to $\lceil \log_2(\frac{t_{\max}-t_{\min}}{\epsilon_2}) \rceil$, where ϵ_2 is the desired accuracy [32]. Hence, the total number of steps by which the solution of Algorithm 1 will be obtained, is equal to $N_{\mathrm{I}} \lceil \log_2(\frac{t_{\max}-t_{\min}}{\epsilon_2}) \rceil$, where N_{I} is the required number of outer *while* loops. Note that N_{I} generally is a random number between 2 and N_I since we use the norm of the difference between two successive solutions as a stopping criterion.

From the perspective of green power allocation in this work, one might naturally ask how the CPU, where the power allocation is performed (see Fig. 1), can be informed of the desired bounds on user SEs. This can be managed in two ways: *i*) subscription of different users to the different levels of services [25], or *ii*) dynamically informing the CPU about the approximate range of required SE from the users. The latter option can be realized by classifying required SEs of mobile applications into different levels and sending the index of the desired level (or application) from the user to the CPU via APs and backhaul links. We believe that the second option is more useful since nowadays, a great number of mobile users are demanding to subscribe to the Internet which in turn, involves a wide range of services with different required SEs such as web browsing, online gaming, multimedia messaging, video streaming and downloading. Moreover, the media-providers are currently offering services with different levels of quality. This includes for example video streaming with the option of setting to different video resolutions associated with different data usages. Such level-based preferences allow a user, which has a limitation on budget and/or device energy, to select her/his preferred option for the video streaming.

IV. NUMERICAL AND SIMULATION RESULTS

In this section, we first carry out a numerical analysis to show how the bound given in Theorem 1 is still satisfied if a subset of users do not meet the condition $R_k^{\min} \geq 1$. Then, we introduce the methodology used in this work for the simulation of cell-free m-MIMO systems. Finally, we present the simulation results obtained by executing the power allocation Algorithm 1 for difference cases.

A. NUMERICAL ANALYSIS ON THE LOWER BOUND

It can be observed from the proof of Theorem 1 that the condition $R_k^{\min} \geq 1$ for all $k \in \mathcal{K}$ is employed to justify the inequality

$$\sum_{k=1}^{K} \log_2(1 + \alpha_k^2) \geq \sum_{k=1}^{K} \log_2(1 + \alpha_k)^{u_k}, \quad (27)$$

where

$$\alpha_k = \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k}{\sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}}}$$

and

$$u_k = \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}}}{\max_k \left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}}\right)} \leq 1.$$

However, the inequality (27) also be met if $R_k^{\min} \geq 1$ (or equivalently $\alpha_k \geq 1$) for some but not all users. To illustrate this fact, we consider a cell-free m-MIMO system with K users operating over a bandwidth of 20MHz. The parameters τ_c and τ_p are fixed to 200 and 20, respectively. For this setting, it is assumed that $K - j$ users achieve SEs which randomly

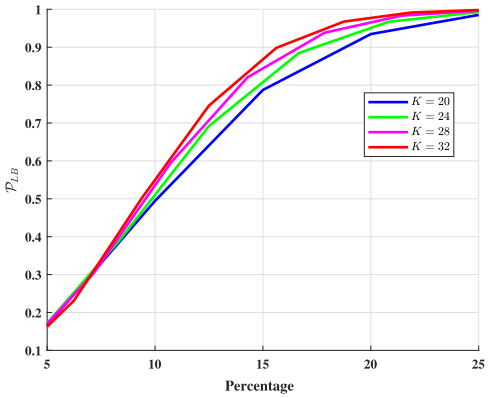


FIGURE 2. Probability $\mathcal{P}_{\text{LB}}$ versus the percentage of users with $R_k^{\min} \geq 1$ for different number of users.

lie within the interval $\left[\frac{1}{9}, \frac{2}{9}\right]$ [6] where $1 \leq j \leq \frac{K}{4}$, while the remaining j users achieve SEs which randomly lie within the interval $\left[\frac{9}{9}, \frac{10}{9}\right]$. We also assume that u_k is a random variable which is uniformly distributed over the interval [0, 1] and replace randomly one of the values u_k with one since we know that $u_{k_{\text{M}}} = 1$ where

$$k_{\text{M}} = \arg\max_{k} \left(\boldsymbol{\gamma}_k^T \boldsymbol{\varsigma}_k + \sqrt{\boldsymbol{\varsigma}^T \boldsymbol{\Omega}_k \boldsymbol{\varsigma} + \rho_d^{-1}} \right).$$

The reason we choose u_k to be a uniform random variable over the interval [0, 1] is that finding a tighter lower or upper bound for this quantity is too difficult. Fig. 2 plots the probability $\mathcal{P}_{\text{LB}} = \mathbb{P}\left\{ \frac{\sum_{k=1}^{K} \log_2(1+\alpha_k^2)}{\sum_{k=1}^{K} \log_2(1+\alpha_k)^{u_k}} \geq 1 \right\}$ that the inequality (27) is satisfied versus the percentage of users having a maximum SE $R_k^{\min} \geq 1$. As seen from this figure, inequality (27) is satisfied with a probability very close to one if approximately $\frac{1}{4}$ of all users have an SE greater than one. Hence, the sufficient condition $R_k^{\min} \geq 1$ for all users, which is applied in Theorem 1, can be relaxed in practice as $R_k^{\min} \geq 1$ for $\frac{K}{4}$ users. Simulation results in Subsection IV-C further confirm this conclusion.

B. SIMULATION METHODOLOGY

We consider a cell-free m-MIMO system in which N single-antenna APs are serving K single-antenna users within a rectangular area of $D \times D$ km^2. Both the APs and users are randomly uniformly distributed over the given area. For a fixed distance d_{nk} (in m) between the nth AP and the kth user, the path loss coefficient a_{nk} (in dB) follows the Hata-COST231 propagation model [1], [12], [28]:

$$a_{nk} = \begin{cases} -L - 35\log_{10}(d_{nk}), & d_{nk} > d_2 \\ -L - 20\log_{10}(d_{nk}) - 15\log_{10}(d_2), & d_1 < d_{nk} \leq d_2 \\ -L - 15\log_{10}(d_2) - 20\log_{10}(d_1), & d_{nk} \leq d_1 \end{cases}$$

where

$$L = 46.3 + 33.9\log_{10}(f_{\text{c}}) - 13.82\log_{10}(h_{\text{AP}}) - (1.1\log_{10}(f_{\text{c}}) - 0.7)h_{\text{u}} + (1.56\log_{10}(f_{\text{c}}) - 0.8)$$

[6]For the studied setting, this SE interval is equivalent to the net data-rate interval of [1, 2]Mbits/s since "net data-rate $= \frac{1}{2}B\left(\frac{\tau_c - \tau_p}{\tau_c}\right) \times$ SE".

is the path loss (in dB) at the reference distance 1km and f_{c} is the carrier frequency (in MHz). Besides, d_0 and d_1 are the break point distances where the path-loss exponent or slope changes [34], while h_{AP} and h_{u} are the antenna heights of the APs and the users (in m), respectively. Moreover, the correlated shadowing is modeled as follows [1], [34], [35]:

$$z_{nk} = \sqrt{\delta} r_n + \sqrt{1-\delta} s_k, \tag{28}$$

where $r_n \sim \mathcal{N}(0, 1)$, $s_k \sim \mathcal{N}(0, 1)$, and $0 \leq \delta \leq 1$. The covariance functions of r_n and s_k are expressed as

$$\mathbb{E}\{r_n r_{n'}\} = 2^{-\frac{d_{\text{AP}}(n,n')}{d_{\text{decorr}}}}, \quad \mathbb{E}\{s_k s_{k'}\} = 2^{-\frac{d_{\text{u}}(k,k')}{d_{\text{decorr}}}}, \tag{29}$$

where $d_{\text{AP}}(n, n')$ is the distance between the nth and n'th APs, $d_{\text{u}}(k, k')$ is the distance between the kth and k'th users, and d_{decorr} is the decorrelation distance which is an environmental parameter typically varying between 20 and 200 meters.

The total power consumption is modeled as [11]:

$$P_{\text{T}} = \sum_{n=1}^{N} \left[P_{\text{tc},n} + \frac{1}{\xi_n} \rho_{\text{d}} P_{\text{N}} \left(\sum_{k=1}^{K} \eta_{nk} \gamma_{nk} \right) \right] + \sum_{n=1}^{N} \left[P_{0,n} + \left(\frac{\tau_{\text{c}} - \tau_{\text{p}}}{2\tau_{\text{c}}} B \sum_{k=1}^{K} R_k \right) P_{\text{bt},n} \right]. \tag{30}$$

In the first term of (30), which gives the total consumption by all N APs, γ_{nk} is defined as in (4), η_{nk} is the power control coefficient corresponding to the nth AP and the kth user, ξ_n is the power amplifier efficiency of the nth AP, P_{N} is the noise power (in Watt), ρ_{d} is the downlink SNR, and $P_{\text{tc},n}$ is the required internal power for running the circuits at the nth AP. The noise power can be calculated as

$$P_{\text{N}} = Bk_{\text{B}}T_0 F_{\text{N}},$$

where k_{B} is the Boltzmann constant (in Joule/Kelvin), T_0 is the noise temperature (in Kelvin), F_{N} is the noise figure, and B is the system bandwidth. In the second term of (30), which represents the total backhaul power consumption, $P_{\text{bt},n}$ is the traffic-dependent backhaul power consumption by the nth AP (in Watt/(bits/s)), R_k is the achievable SE of the kth user, τ_{c} is the dimension of coherence elements, τ_{p} is the length of pilot sequence in terms of coherence elements, and $P_{0,n}$ is the traffic-independent power consumption by the nth AP.

The SNRs ρ_{d} and ρ_{p} are computed as $\rho_{\text{d}} = \frac{\tilde{\rho}_{\text{d}}}{P_{\text{N}}}$ and $\rho_{\text{p}} = \frac{\tilde{\rho}_{\text{d}}}{P_{\text{N}}}$, respectively where $\tilde{\rho}_{\text{d}}$ and $\tilde{\rho}_{\text{p}}$ are the transmission powers of downlink and pilot phases, respectively. The numerical values of most parameters are taken from [1], [11] and listed in Table 1. The range of bandwidth is selected between 20 and 100MHz which is supported by LTE and LTE-A standards [36]. We choose values of N, K, and D such that the density of APs and users are the same as cases considered in [1], [11], [12], [28].

The curves related to the solutions of power allocation problems are obtained by averaging over 100 independent random realizations of user and AP locations, as well as shadow-fading coefficients. For the proposed approach, CVX [33] is employed to solve the optimization problem $\mathcal{P}_2''$.

TABLE 1. System specifications.

Parameter	Symbol	Value
Number of APs	N	$\{50, 60\}$
Number of users	K	$\{20, 24\}$
Side of area	D	$\frac{1}{\sqrt{2}}$km
Carrier frequency	f_{c}	1.9GHz
Dimension of coherence element	τ_{c}	200
Length of pilot sequence	τ_{p}	$[20, 40]$
Hight of the AP antenna (for all APs)	h_{AP}	15m
Hight of the user antenna (for all users)	h_{u}	1.65m
First break point distance	d_1	10m
Second break point distance	d_2	50m
Decorrelation distance	d_{decorr}	100m
Downlink transmission power	$\tilde{\rho}_{\mathrm{d}}$	$[0.2, 1]$W
Pilot training transmission power	$\tilde{\rho}_{\mathrm{p}}$	$[0.1, 0.2]$W
Bandwidth	B	$[20, 100]$MHz
Power amplifier efficiency	$\xi_n, \forall n$	0.4
Internal power consumption	$P_{\mathrm{tc},n}, \forall n$	0.2W
Traffic-independent backhaul power	$P_{0,n}, \forall n$	0.825W
Traffic-dependent backhaul power	$P_{\mathrm{bt},n}, \forall n$	$[0.25, 2]$W/(Gbits/s)
Boltzmann constant	k_{B}	1.381×10^{-23}Joule/Kelvin
Noise temperature	T_0	290Kelvin
Noise figure	F_{N}	9dB
Shadowing variance	σ_{sh}	8dB

To initialize Algorithm 1, we use $\boldsymbol{\psi}_k^{(0)} = c\,\mathbf{1}_{NK+1}$ where $c \in [0.01, 0.2]$, while N_I is set to 2. Based on our experience, $c = 0.1$ works well for most simulation cases considered in this article. We compare the achievable sum rate obtained by maximizing the proposed lower bound with that achieved by the *max-min* power allocation approach, which maximizes the minimum SE of the K users [1] when no upper bounding constraints are imposed on the achievable user SEs, i.e., $R_k^{\max} = \infty$ for all users. Note that we use (11) and (12) to formulate the optimization problem of max-min power allocation, as it requires less number of optimization variables compared to the formulation used in [1] which involves some extra slack variables.

C. RESULTS AND DISCUSSION

In this subsection, we study the achievable rate performance of cell-free m-MIMO by solving the power allocation problem $\mathcal{P}_2$ under various parametric settings for both uncorrelated and correlated shadowing effects. Specifically, we first run Algorithm 1 by considering different practical cases for parameters $R_k^{\min}$ and $R_k^{\max}$. Next, we use Algorithm 1 to find the optimal solutions of two extreme scenarios similar to the one described in Section I. Finally, we compare the achievable sum rate performance obtained by the proposed Algorithm 1 and the max-min power allocation method.

We first study the behavior of the proposed power allocation Algorithm 1 for different values of the minimum and maximum user SEs. To this end, we partition the K users into four groups G_i, $i \in \{1, 2, 3, 4\}$, and for each group set common values of the minimum and maximum SEs, i.e., $R_k^{\min}$ and $R_k^{\max}$. Hence, each group corresponds in effect to a different type of service. In our simulations, we consider three different cases of assignment of the SE range for the different groups, as summarized in Table 2. The results of executing Algorithm 1 for four cases are provided in Fig. 3, Fig. 4 and Table 3. The detailed discussions on the results of each case are as follows:

TABLE 2. Different cases for executing power allocation Algorithm 1.

	G_1	G_2	G_3	G_4
Case I	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{1}{9}, \frac{2}{9}]$
Case II	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{1}{9}, \frac{2}{9}]$	$[\frac{9}{9}, \frac{10}{9}]$
Case III	$[\frac{3}{9}, \frac{6}{9}]$	$[\frac{9}{9}, \frac{12}{9}]$	$[\frac{6}{9}, \frac{9}{9}]$	$[\frac{12}{9}, \frac{14}{9}]$
Case IV	$[\frac{1}{9}, \infty]$	$[\frac{1}{9}, \infty]$	$[\frac{1}{9}, \infty]$	$[\frac{1}{9}, \infty]$

- Case I: As a worst-case scenario, we set $R_k^{\min} = \frac{1}{9}$ and $R_k^{\max} = \frac{2}{9}$ for all users to study how maximizing the proposed bound works even if the condition $R_k^{\min} \geq 1$ is not satisfied for a quarter of users (see Remark 1 and the discussion in Subsection IV-A). As seen from Fig. 3, the 95% per user achievable rates are approximately 1.78 and 1.30 Mbits/s for the uncorrelated and correlated shadowing, respectively. Moreover, median

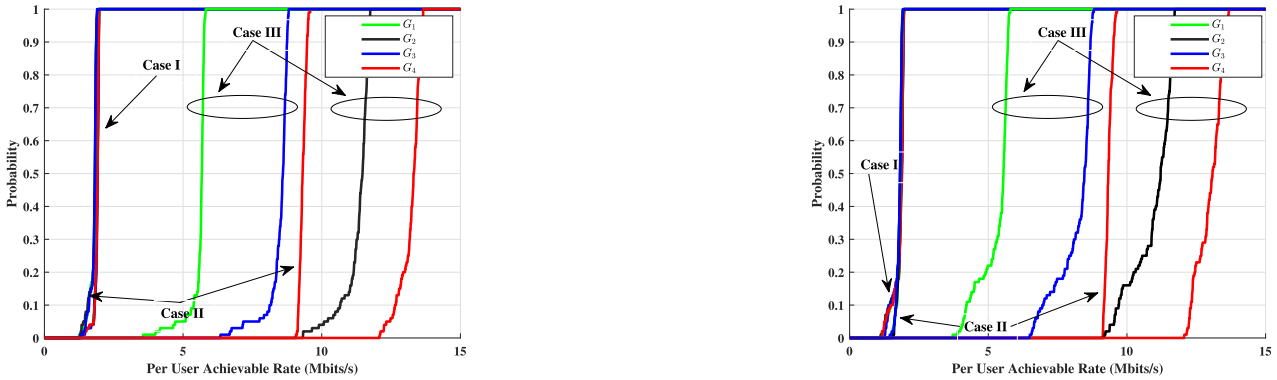


FIGURE 3. CDF of average per user achievable rate for the cell-free m-MIMO system with powers are allocated using Algorithm 1: (a) uncorrelated shadowing, (b) correlated shadowing; $N = 50$, $K = 20$, $\tilde{\rho}_{\mathbf{d}} = 5\tilde{\rho}_{\mathbf{p}} = 1$W, $B = 20$MHz, and $\tau_{\mathbf{p}} = 20$.

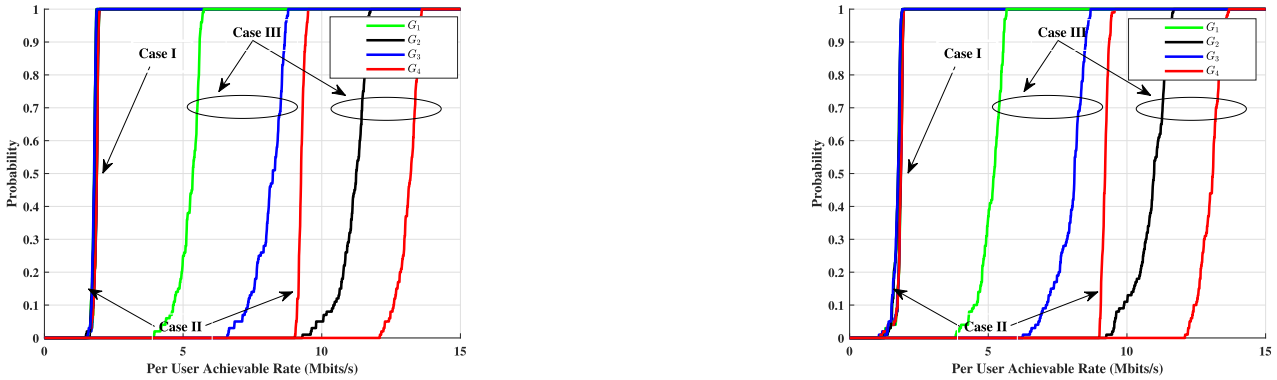


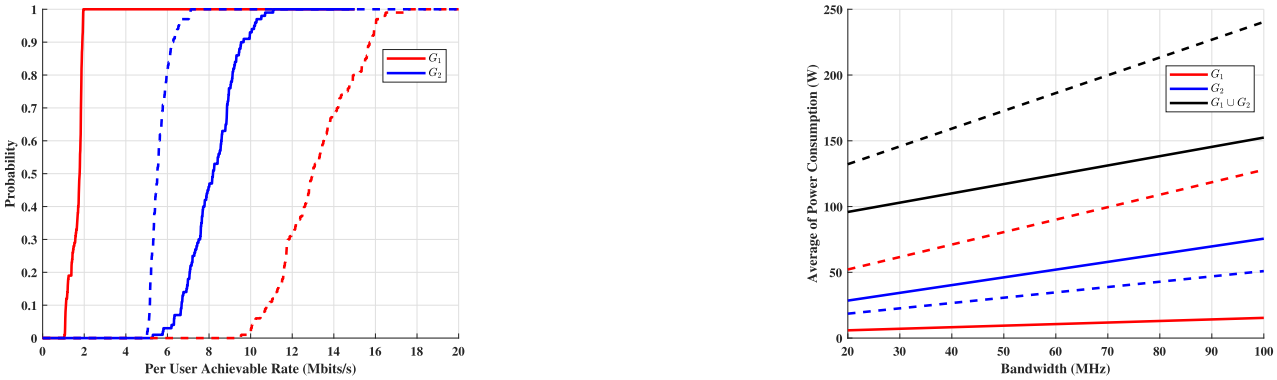
FIGURE 4. CDF of average per user achievable rate for the cell-free m-MIMO system with powers are allocated using Algorithm 1: (a) uncorrelated shadowing, (b) correlated shadowing; $N = 60$, $K = 24$, $\tilde{\rho}_{\mathbf{d}} = 5\tilde{\rho}_{\mathbf{p}} = 1$W, $B = 20$MHz, and $\tau_{\mathbf{p}} = 20$.

achievable rates of 1.92 and 1.88 Mbits/s are obtained for the uncorrelated and correlated scenarios. Regarding the discussion in Subsection IV-A, inequality (27) is met 14% and 19% of the realizations for the uncorrelated and correlated shadowing, respectively.

- Case II: We change $R_k^{\min}$ and $R_k^{\max}$ to values of $\frac{9}{9}$ and $\frac{10}{9}$, respectively for G_4 while keeping the same values as in Case I for the other three groups, so that a quarter of users satisfy the condition $R_k^{\min} \geq 1$. In this case, it is observed that inequality (27) is satisfied for all realizations, which corroborates the conclusion of our numerical analysis in Subsection IV-A. Regarding the per user achievable rates of users in G_1-G_3, at least 1.5 and 1.62 Mbits/s are achieved with probability of 95% for the uncorrelated and correlated scenarios, respectively. The corresponding values for G_4 are equal to 9.14Mbits/s for both uncorrelated and correlated shadowing. As seen from Fig. 3, the difference between the 95% and median achievable rates are small for G_4 which means that the given values are observed almost always.
- Case III: In this case, we study a more practical situation where the four different groups have different rate requirements. Similar to Case II, the inequality in (27) is met in all realizations for both the uncorrelated and correlated shadowing. As seen from Fig. 3, the per user achievable rates of G_1-G_4 are 5.09, 10.20, 7.75, and 12.47 Mbits/s, respectively, when the shadow fading is uncorrelated. The corresponding values for correlated shadowing are 4.10, 9.55, 6.69, and 12.24 Mbits/s. Similar results can be observed from Fig. 4 where we increase the density of APs and users by 20%.
- Case IV: In this scenario, in order to demonstrate the benefits of imposing upper bounding constraints upon the SEs of different users on the average total power consumption, we set $R_k^{\min} = 1$ and $R_k^{\max} = \infty$, which amounts to relaxing such constraints. The power consumption results from Algorithm 1 for Cases I-IV are presented in Table 3 for different network configurations and shadowing conditions. It can be observed that compared to Case IV, imposing upper bounding SE constraints in Cases I-III, leads to significant reductions in power consumption for all combinations of network and shadowing conditions. For example, when comparing Case III and Case IV, the power consumption is reduced

TABLE 3. The average of total power consumption in Watt for different cases and types of shadow fading; $P_{\mathrm{bt},n} = 0.25$ $\forall n$.

	$N = 50, K = 20$		$N = 60, K = 24$	
	Uncorrelated	Correlated	Uncorrelated	Correlated
Case I	52.95	53.57	63.34	63.63
Case II	54.77	59.57	65.24	66.10
Case III	60.66	63.11	73.72	75.84
Case IV	86.18	87.23	101.08	101.75



(a) (b)

FIGURE 5. (a) Per user achievable rate performance of the cell-free m-MIMO achieved in the extreme scenario; (b) Average of power consumption (in Watt) versus the bandwidth (in MHz); bounded G_1 (solid line), unbounded G_1 (dash line), uncorrelated shadowing; $N = 60$, $K = 24$, $\tau_{\mathrm{p}} = 20$, $\tilde{\rho}_{\mathrm{d}} = 5\tilde{\rho}_{\mathrm{p}} = 1$W, and $P_{\mathrm{bt},n} = 2$ $\forall n$.

by about 30% and 28% for uncorrelated and correlated shadowing, respectively when $N = 50$ and $K = 20$, and by 27% and 25% when $N = 60$ and $K = 24$.

Next, we consider the extreme scenario in which the first group of $\frac{K}{2}$ users, say G_1, is located in the sub-area $[0, D/2] \times [0, D/2]$ while the second group of $\frac{K}{2}$ users, say G_2, is being served within the sub-area $[3D/4, D] \times [3D/4, D]$. We further assume that the users of group G_1 are being served with low-speed services whose required data rates lie in the interval [1, 2]Mbits/s, while those of group G_2 are being served with high-speed services requiring data rates higher than 5Mbits/s. The APs are uniformly distributed over the region $[0, D] \times [0, D/2] \cup [0, D/2] \times [D/2, D]$. Moreover, we consider two different cases with *unbounded* and *bounded* intervals $[1, \infty)$ and $[1, 3]$ for the achievable rates of group G_1 while those of G_2 are fixed to the unbounded interval $[5, \infty)$. Fig. 5a depicts the CDF of per user achievable rate of both groups G_1 and G_2 when $B = 20$MHz for both unbounded and bounded cases. As seen from the figure, the 95% per user achievable rate of group G_1 is 10.20Mbits/s for the unbounded case and 1.08Mbits/s for the bounded case. Moreover, the per user achievable rate of group G_2 increases by 1.23Mbits/s from 5.09Mbits/s to 6.32Mbits/s as we impose the upper bound on the achievable data rate of group G_1.

To see the impact of bandwidth on the power consumption, Fig. 5b illustrates the average of power consumption by group G_1, group G_2, and $G_1 \cup G_2$ (all users) versus the bandwidth. Note that we excluded the summation terms involving $P_{\mathrm{tc},n}$ and $P_{0,n}$ in (30) from the power measurement of groups G_1 and G_2 but not from that of $G_1 \cup G_2$. The average power consumption of group G_1 is 52.15W for the unbounded scenario and 5.89W for the bounded scenario when $B = 20$MHz. This clearly shows that the average of power consumption related to group G_1 is significantly reduced (by 89%) by imposing an upper bound upon the achievable rates of group G_1. Similar results also hold true for other bandwidth values. For group G_2, the average of power expectedly increases as the achievable data rates of group G_2 increases when the achievable data rates of group G_1 are upper bounded. However, the total power consumption still decreases from 132W to 95.9W and from 240W to 152W at the extreme frequency points of 20MHz and 100MHz, respectively. As observed from Fig. 5b, the saving in total power consumption increases from 27% to 37% as the bandwidth increases from 20 to 100MHz.

Finally, we compare in Fig. 6, the achievable sum rate of cell-free m-MIMO as obtained by two different power allocation methods: *i*) the max-min algorithm, and *ii*) maximizing the lower bound in (14) by executing Algorithm 1 when $R_k^{\min} = 1$ and $R_k^{\max} = \infty$ for all users. Note that Algorithm 1 will then be simplified into a bi-section algorithm and require neither the outer while loop nor the initial point $\boldsymbol{\psi}_k^{(0)}$. As seen from the figure, maximizing the lower bound in (14) leads

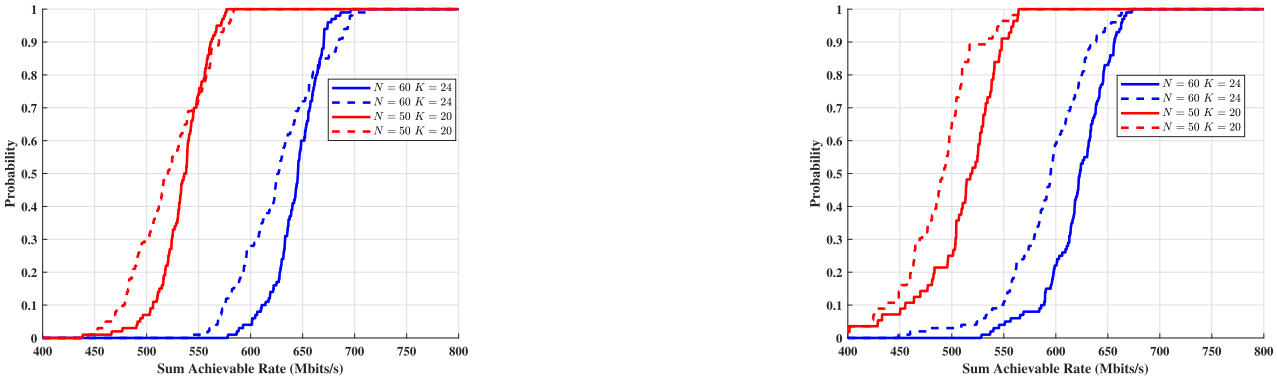


FIGURE 6. Achievable sum rate performance of the cell-free m-MIMO achieved by two power allocation approaches *i*) max-min algorithm (dash lines), and *ii*) Algorithm 1 (solid lines) (a) uncorrelated shadowing, (b) correlated shadowing; $R_k^{\min} = 1$, $R_k^{\max} = \infty$, $B = 40$MHz, $\tilde{\rho}_d = 5\tilde{\rho}_p = 1$W and $\tau_p = 20$.

to a higher achievable sum rate compared with the max-min approach in both the uncorrelated and correlated scenarios. Specifically, the 95% per user achievable rate under the uncorrelated shadowing scenario increases by 1.32Mb/s for $(N = 50, K = 20)$ and by 1.62Mb/s for $(N = 60, K = 24)$. For the correlated shadowing both power allocation methods achieve approximately the same 95% achievable rate performance for $(N = 50, K = 20)$. However, there still exists an increase of 1.14Mb/s for the median per user achievable rate when the lower bound (14) is maximized. As the numbers of APs and users increase by 20% (to $N = 60$ and $K = 24$), maximizing the lower bound (14) achieves a higher 95% per user achievable rate compared to the max-min approach.

V. CONCLUSION

In this article, we considered the problem of downlink power allocation in a cell-free m-MIMO system under SE lower and upper bounding constraints for the users. We observed that the resulting optimization problem was non-convex since the objective function was non-concave and the upper bounding constraints on user SEs were non-convex. To overcome these difficulties, we first derived a closed-form lower bound on the sum SE (objective function) and proved that it was a quasi-concave function. Then, we relaxed the unwieldy upper bounding constraints on the user SEs by replacing them with linear functions. An optimal solution to the relaxed problem was finally obtained by solving a sequence of convex feasibility programs. We evaluated the performance of the proposed downlink power allocation scheme through Monte Carlo simulations under the uncorrelated and correlated shadow fading models. The results showed that for both models, the proposed suboptimal algorithm can lead to a significant reduction in total power consumption compared to a benchmark approach.

REFERENCES

[1] 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.

[2] Z. Chen and E. Bjornson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, Nov. 2018.

[3] X. Chen, S. Zhang, and Q. Li, "A review of mutual coupling in MIMO systems," *IEEE Access*, vol. 6, pp. 24706–24719, May 2018.

[4] I. Nadeem and D.-Y. Choi, "Study on mutual coupling reduction technique for MIMO antennas," *IEEE Access*, vol. 7, pp. 563–586, Jan. 2019.

[5] X. Chen, M. Abdullah, Q. Li, J. Li, A. Zhang, and T. Svensson, "Characterizations of mutual coupling effects on switch-based phased array antennas for 5G millimeter-wave mobile communications," *IEEE Access*, vol. 7, pp. 31376–31384, Mar. 2019.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] A. Morsali, S. S. Hosseini, B. Champagne, and X.-W. Chang, "Design criteria for omnidirectional STBC in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1435–1439, Oct. 2019.

[8] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, "Iterative algorithm induced deep-unfolding neural networks: Precoding design for multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, early access, Oct. 30, 2020, doi: 10.1109/TWC.2020.3033334.

[9] S. S. Hosseini, J. Abouei, and M. Uysal, "Fast-decodable MIMO HARQ systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2827–2840, May 2015.

[10] N. Kaur, S. S. Hosseini, and B. Champagne, "Enhanced channel tracking in THz beamspace massive MIMO: A deep CNN approach," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Auckland, New Zealand, Dec. 2020, pp. 76–81.

[11] 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.

[12] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max–min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.

[13] *Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)*, Int. Telecommun. Union, Geneva, Switzerland, Nov. 2016.

[14] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–7.

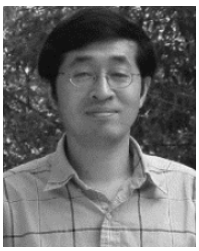
[15] M. Bashar, H. Q. Ngo, A. G. Burr, D. Maryopi, K. Cumanan, and E. G. Larsson, "On the performance of backhaul constrained cell-free massive MIMO with linear receivers," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 624–628.

[16] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 971–987, Dec. 2019.

[17] 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. ICC-IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–7.
[18] M. Bashar, A. Akbari, K. Cumanan, H. Q. Ngo, A. G. Burr, P. Xiao, and M. Debbah, "Deep learning-aided finite-capacity fronthaul cell-free massive MIMO with zero forcing," in *Proc. ICC-IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
[19] P. Parida, H. S. Dhillon, and A. F. Molisch, "Downlink performance analysis of cell-free massive MIMO with finite fronthaul capacity," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–6.
[20] S. Timilsina, D. Kudathanthirige, and G. Amarasuriya, "Physical layer security in Cell- free massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–7.
[21] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, Oct. 2018.
[22] X. Zhang, D. Guo, K. An, Z. Ding, and B. Zhang, "Secrecy analysis and active pilot spoofing attack detection for multigroup multicasting cell-free massive MIMO systems," *IEEE Access*, vol. 7, pp. 57332–57340, Apr. 2019.
[23] X. Zhang, D. Guo, and K. An, "Secure communication in multigroup multicasting cell-free massive MIMO networks with active spoofing attack," *Electron. Lett.*, vol. 55, no. 2, pp. 96–98, Jan. 2019.
[24] X. Zhang, D. Guo, K. An, and B. Zhang, "Secure communications over cell-free massive MIMO networks with hardware impairments," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1909–1920, Jun. 2020.
[25] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.
[26] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. IEEE Future Netw. Mobile Summit*, Florence, Italy, Jun. 2010, pp. 1–10.
[27] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[28] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.
[29] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester, *A Modern Introduction to Probability and Statistics: Understanding Why and How*. London, U.K.: Springer-Verlag, 2005.
[30] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[33] M. Grant and S. Boyd. (Jan. 2020). *CVX: MATLAB Software For Disciplined Convex Programming, Version 2.2*. [Online]. Available: http://cvxr.com/cvx
[34] M. Karlsson, E. Bjornson, and E. G. Larsson, "Techniques for system information broadcast in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 244–257, Jan. 2019.
[35] Z. Wang, E. K. Tameh, and A. R. Nix, "Joint shadowing process in urban peer-to-peer radio channels," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 52–64, Jan. 2008.
[36] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*. Hoboken, NJ, USA: Wiley, 2012.

SEYYED SALEH HOSSEINI (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in communication systems engineering from the Shahid Bahonar University of Kerman, Kerman, Iran, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, McGill University, QC, Canada. From 2013 to 2017, he was a Faculty Member with the Department of Electrical Engineering, Azad University, Iran. His research interests are in general areas of communication and estimation theory with reference to massive MIMO, FSO communications, channel estimation, and convex optimization. He has served as a reviewer for several IEEE conferences and international journals, including the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE SYSTEMS JOURNAL, the IEEE WIRELESS COMMUNICATIONS LETTERS, the *EURASIP Journal on Wireless Communications and Networking*, and *IET Communications*.

BENOIT CHAMPAGNE (Senior Member, IEEE) received the B.Ing. degree in engineering physics from the École Polytechnique de Montréal in 1983, the M.Sc. degree in physics from the Université de Montréal in 1985, and the Ph.D. degree in electrical engineering from the University of Toronto in 1990. From 1990 to 1999, he was an Assistant and then an Associate Professor with the INRS-Telecommunications, Université du Quebec, Montreal. In1999, he joined McGill University, Montreal, where he is currently a Full Professor with the Department of Electrical and Computer Engineering. He served as an Associate Chair of graduate studies with the Department of Electrical and Computer Engineering from 2004 to 2007. His research focuses on the study of advanced algorithms for the processing of information bearing signals by digital means. His interests span many areas of statistical signal processing, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband communications and audio processing. He has co-authored more than 300 refereed publications in these areas. His research has been funded by the Natural Sciences and Engineering Research Council of Canada, the Fonds de Recherche sur la Nature et les Technologies from the Government of Quebec, and some major industrial sponsors, including Nortel Networks, Bell Canada, InterDigital, and Microsemi. He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *EURASIP Journal on Applied Signal Processing*. He has also served on the technical committees of several international conferences in the fields of communications and signal processing.



XIAO-WEN CHANG received the B.S. and M.S. degrees in computational mathematics from Nanjing University, Nanjing, China, in 1986 and 1989, respectively, and the Ph.D. degree (Dean's Honor List) in computer science from McGill University, Montreal, QC, Canada, in 1997. He is currently an Associate Professor with the School of Computer Science, McGill University. His research interests are in the area of scientific computing, with the emphasis on numerical linear algebra and its applications. He has authored or coauthored more than 40 articles in refereed journals.