Improving Cell-Free Massive MIMO Networks Performance: A User Scheduling Approach

Juwendo Denis, *Member, IEEE*, and Mohamad Assaad, *Senior Member, IEEE*

***Abstract*— Cell-Free (CF) massive multiple-input multiple-output (MIMO) system is a distributed antenna system, wherein a large number of access points wish to simultaneously communicate with a relatively small number of users. Similar to co-located massive MIMO system, pilot contamination and multi-user interference are two major impediments to CF massive MIMO network performance improvement. One can mitigate the detrimental effect of pilot contamination and multi-user interference through judicious resource allocation. In this work, we formulate and investigate the problem of frequency assignment for a CF massive MIMO. The formulated optimization problem is proven to be generally NP-hard. Local optima can be found by approaches such as the Lagrangian method that however, has considerably slow rates of convergence. To circumvent this issue, we propose an alternative solution being of two-folds. Firstly, we reformulate the problem as a grouping strategy which enables to attenuate the effect of intra-group multi-user interference. In the second fold, frequencies are assigned in a non-overlapping fashion to each scheduled group to palliate the effect of inter-group interference and pilot contamination. To further improve the performance of the proposed approach, power coefficients are allocated to the users via a sequential convex approximation (SCA)-based framework. The effectiveness of the proposed algorithms is then verified through extensive numerical simulations which demonstrate a non-negligible improvement in the performance of the studied scenario.**

***Index Terms*— Massive multiple-input multiple-output (MIMO), users scheduling, semidefinite programming, pilot contamination, multi-user interference.**

I. Introduction

IN THE past decades, wireless communication systems have experienced incessant growth in terms of both mobile connected devices and mobile data traffic. This growth is foreseen to escalate during the years ahead [2]. In addition to that, there is an increasing demand from users for higher data rate. This implies that 5G and beyond related wireless networks should be able to accommodate the traffic growth, enable new use cases, provide seamless connectivity and ensure that users demands together with specific performance requirements are satisfied [3]. Massive multiple-input multiple-output (MIMO) emerges as an appealing technology for 5G physical-layer [4], [5] that enables to increase the network spectral efficiency [6].

Massive MIMO, which is often referred to as co-located massive MIMO, consists of deploying a combination of large-scale antenna arrays at the access-points (APs). Its counterpart known as distributed massive MIMO consists of disseminating the serving antennas over a relatively large geographical area. An instance of distributed massive MIMO is cell-free (CF) massive MIMO system which has attracted tremendous attention recently from the research community [7]–[10]. Cell-free massive MIMO systems are composed of very large number of APs that simultaneously and coherently serve, on the same time/frequency resource, a relatively small number of users [8]. This distributed configuration enables to achieve higher coverage probability, at the expense of higher back-haul overhead, than its counterpart co-located massive MIMO system [11].

In a CF massive MIMO network, all APs are communicating with the users using same time/frequency resource.The network will incur multi-user interference that limit the performance of the system. It was shown in [8] that multi-user interference can be eliminated under the assumption that the number of transmit antennas goes to infinity. However, this condition is very untenable for practical scenarios. In addition, the escalating number of mobile devices that are foreseen to request services from the APs will lead to unprecedented amount of interference that needs to be prudently managed in order to avoid severe performance deterioration. Therefore, novel and efficient approaches should be investigated given that traditional multi-user interference management methods deem to be infeasible or inefficient for 5G and beyond wireless networks in general [12], and CF massive MIMO systems in particular.

There exists a large body of literature focusing on the mitigation of multi-user interference for 5G-based networks [13]–[18] (and references therein). The issue of interference management for 5G heterogeneous networks was addressed in [17]. In [15], [16], [18], the authors leverage the spatial signature of the channel vectors, and invoke the principle of user scheduling, to attenuate the effect of multi-user interference and to reduce overhead signaling pertaining to channel state information (CSI) acquisition. A common theme in existing schemes [14]–[16], [18] is that the APs

Manuscript received February 5, 2020; revised September 15, 2020 and March 5, 2021; accepted May 14, 2021. Date of publication May 31, 2021; date of current version November 11, 2021. This article was presented at the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 2018 [1]. The associate editor coordinating the review of this article and approving it for publication was S. K. Mohammed. *(Corresponding author: Juwendo Denis.)*

Juwendo Denis is with the School of Engineering and Applied Sciences (SEAS), Harvard University, Cambridge, MA 02138 USA (e-mail: jdenis@seas.harvard.edu).

Mohamad Assaad is with the Laboratoire des Signaux et Systems (L2S, CNRS), CentraleSupelec, University Paris-Saclay, 91190 Gif-sur-Yvette, France (e-mail: mohamad.assaad@centralesupelec.fr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3083139.

Digital Object Identifier 10.1109/TWC.2021.3083139

are equipped with antennas in the form of uniform linear array (ULA) and the covariance matrix of the user's channel follows a one-ring local scattering model. Although, these findings cannot directly apply to CF massive MIMO networks, the standpoint of user grouping can be exploited to design algorithms to efficiently combat multi-user interference in cell-free systems.

As in co-located massive MIMO, cell-free massive MIMO networks can achieve good performance at the expense of perfect CSI acquisition at the APs. For time-division duplex (TDD) cell-free massive MIMO networks, the APs can rely on the assumption of uplink/downlink channel reciprocity to estimate the downlink CSI [8] from known pilot signaling of the users. Considering the huge number of users that will be active in 5G networks, non-orthogonal training sequences must be utilized during the uplink training stage. Consequently, the network will incur pilot contamination, a major limiting factor of the performance of massive MIMO network [6]. It has been argued in [19] that the capacity increases without bound under the condition that the covariance matrices of the users are asymptotically linearly independent and the precoder is the centralized minimum mean-square error (MMSE). Although this result is theoretically achievable, it is unattainable for practical implementations. In fact, the MMSE precoder requires the computation of the inverse of a matrix that scales with the number of APs and antennas. On top of that, the acquisition of the covariance matrix is also impeded by pilot contamination during the training phase [20]. Therefore, there is still a need to provide efficient and practical (in terms of computation) schemes to combat pilot contamination. One possible way to do so is to resort to resource allocation such as pilot assignment and power control for cell-free massive MIMO [8], [10], [21]. In this work, we advocate that the impact of pilot contamination can be circumvented through radio resource allocation.

Existing literature about performance evaluation for CF massive MIMO networks has been done considering single frequency system. However, for dense networks such as CF massive MIMO network single frequency system implies aggressive frequency reuse that leads to co-channel multi-user interference that deteriorate the quality of service (QoS) of the users [22]. This might be avoided by leveraging the fragmentation of the frequency band along with proper management of the resulting sub-frequency with the goal of achieving performance enhancement.

Motivated by the above discussion, we investigate the impact of sub-frequency bands allocation on the issues of pilot contamination and multi-user interference which are two major limiting factors for the performance gains of a CF massive MIMO system. Despite the great deal of existing works on CF massive MIMO [7]–[10], [21](and references therein), no prior work has attempted to address these issues from the perspective of radio resource assignment. In this paper, we consider a dynamic radio resource allocation approach and evaluate its impact on the performance of TDD CF massive MIMO networks.

Different from [1] where we solely proposed an algorithm to improve favorable propagation for CF massive MIMO, the key contributions of this paper are as follows:

1) We investigate the benefits of dynamic scheduling of users for a TDD CF massive MIMO network by optimizing the total downlink achievable rate subject to each user's QoS requirement. The problem is formulated as a binary optimization problem which we prove to be an NP-hard problem. In addition, a local solution using the Lagrangian method is provided. This Lagrangian-based solution, which is obtained by means of gradient-descent approach, has however a very low rates of converge in general.
2) For the purpose of providing faster convergence-based and more efficient approaches, we establish an alternative design optimization problem from the perspective of user grouping. Firstly, we design a framework that partition users into group according to the level of their *mutual* interference. Such paradigm can palliate intra-group multi-user interference. Secondly, groups are scheduled on different radio resources to mitigate the both the effect of inter-group interference as well as the impact of interference due to pilot contamination.
3) We invoke the semidefinite relaxation (SDR) [23] method to design a polynomial time solvable randomized procedure in the same vein of Gaussian randomization [23] to efficiently solve the formulated user grouping optimization problem.
4) In addition, we design a polynomial time solvable algorithm based on sequential convex approximation (SCA) to find an efficient local solution to the problem of frequency assignment for the purpose of group scheduling. We establish the convergence to a stationary point of the proposed SCA-based algorithm. Moreover, we consider the power control optimization problem which is efficiently solved by using an SCA-based approach.

The rest of this work is structured as follows. In Section II, the CF massive MIMO system model is presented. The resource management problem formulation is given in Section III together with the Lagragian-based solution. The user grouping approach is introduced in Section IV. In Section V, the description of our proposed SDR-based algorithm is provided followed by the associated sub-frequency band assignment as well as the problem power control allocation. In section VI, simulation results demonstrating the effectiveness of our proposed methods are provided. And we conclude the paper in Section VII.

Notation: In this paper, $x \sim \mathcal{CN}(\mu, \sigma^2)$ represents a circularly symmetric complex random variable drawn from a Gaussian distribution with mean μ and variance σ^2. The natural logarithmic function is denoted as $\log(\cdot)$. Given a nonempty set $\mathcal{A}$, $|\mathcal{A}|$ stands for the cardinal of $\mathcal{A}$. $[x]_a^b$ is the Euclidean projection of x onto the interval$[a, b]$ while $(x)^+ \triangleq \max(0, x)$.

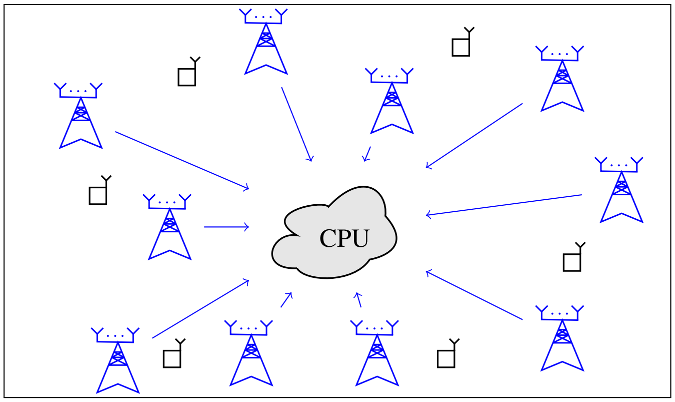


Fig. 1. Illustration of a Cell-Free massive MIMO network.

II. System Model And Preliminaries

In this work, we consider a cell-free massive MIMO system that consists of M access points (APs) that serve simultaneously on a fully-cooperative fashion K mobiles users (UEs), equipped each with a single omni-directional antenna as depicted in Figure 1. The APs are assumed to have each N omni-directional antennas. Furthermore, we suppose that the number of APs is much higher than the number of UEs i.e. $K \ll M$. All users are arbitrarily distributed over a relatively large geographical area. The APs and the UEs are assumed to be perfectly synchronized in both time and frequency. Moreover, we suppose that the APs are managed by a central processing unit (CPU) to which they are connected through a *error-free* back-haul link. The CPU can handle part the physical layer information process such as data coding and decoding.

The channel of a user is considered to be frequency-flat slow fading on each orthogonal frequency division multiplexing (OFDM) subcarrier. In the sequel, the subcarrier index will be omitted for the sake of simplicity. Let $\boldsymbol{g}_{mk} \in \mathbb{C}^{N\times 1}$, denote the complex channel vector between the kth user and the mth AP. It can be modeled as follows

$$\boldsymbol{g}_{mk} = \sqrt{\beta_{mk}}\boldsymbol{h}_{mk} \tag{1}$$

where $\boldsymbol{h}_{mk} \sim \mathcal{CN}(0, \mathbf{I}_{\mathrm{N}})$ with $\mathbf{I}_{\mathrm{N}}$ being the $N \times N$ identity matrix, $m = 1, \cdots, M, \quad k = 1, \cdots, K$, denote the small-scale fading coefficients which are independent and identically distributed (i.i.d) while $\beta_{mk}, m = 1, \cdots, M, \quad k = 1, \cdots, K$, the large scale fading coefficients that include path-loss and shadowing.

The system is operating according to a TDD protocol. Each coherence interval T_c is divided between uplink training and downlink data transmission. By exploiting uplink/downlink channel reciprocity [6], the downlink CSI can be estimated through uplink training. During the uplink training phase, each user is assigned a training sequence that spans $\tau < T_c$ channel uses. The pilot sequences used in the channel estimation phase can be represented by a matrix $\mathbf{\Phi} \in \mathbb{C}^{\tau\times K}$. The kth column of the matrix denoted by $\boldsymbol{\phi}_k \in \mathbb{C}^{\tau\times 1}$ accounts for the pilot sequence used by the kth UE. Each element of $\boldsymbol{\phi}_k$ is of unit-magnitude so that it has a constant power level $\|\boldsymbol{\phi}_k\|^2 = \tau$. The channel estimate can be obtained in a decentralized fashion at each AP m. Define $\boldsymbol{G}_m \triangleq [\boldsymbol{g}_{m1}, \cdots, \boldsymbol{g}_{mk}]$, the $N \times K$ received pilot matrix signal at the mth AP is expressed as

$$\boldsymbol{Y}_{m,p} = \sqrt{\tau\rho_p}\boldsymbol{G}_m\mathbf{\Phi}^\dagger + \boldsymbol{N}_{m,p} \tag{2}$$

where ρ_p is the transmit power during the training phase, and $\boldsymbol{N}_{m,p} \in \mathbb{C}^{N\times K}$ is the complex additive Gaussian noise matrix at the mth AP. The elements of $\boldsymbol{N}_{m,p}$ are i.i.d. random variables that follow a standard normal distribution. The mth AP can perform MMSE [24] estimation of $\boldsymbol{g}_{mk}$ to obtain the corresponding channel estimate, $\widehat{\boldsymbol{g}}_{mk}$ which is given by [10]

$$\widehat{\boldsymbol{g}}_{mk} = \frac{\sqrt{\tau\rho_p}\beta_{mk}}{\rho_p\tau\sum_{j=1}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2 + 1} \times \left(\sqrt{\tau\rho_p}\boldsymbol{g}_{mk} + \sqrt{\tau\rho_p}\sum_{j=1, j\neq k}^{K}\boldsymbol{g}_{mj}\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j + \boldsymbol{w}_{mk}\right) \tag{3}$$

where $\boldsymbol{w}_{mk} \sim \mathcal{CN}(0,1)$ is the kth column of the matrix $\boldsymbol{N}_{m,p}\mathbf{\Phi}$. The quality of the channel estimation is captured through the mean-squared error (MSE) $\mathbb{E}\{\|\boldsymbol{g}_{mk} - \widehat{\boldsymbol{g}}_{mk}\|^2\}$. A good estimation quality is usually represented by a small MSE. Also, the channel estimate in (3) is corrupted by pilot signals sent by other users, in case $K > \tau$, leading to pilot contamination which degrades user's achievable rate [6]. It is worth mentioning that, we only assume a given transmit power ρ_p during each training phase as this paper does not focus on the impact of adjusting ρ_p on the quality of channel estimates. Similar to [25], this paper assumes imperfect channel reciprocity. Denote $\widetilde{\boldsymbol{g}}_{mk} \triangleq \boldsymbol{g}_{mk} - \widehat{\boldsymbol{g}}_{mk}$ as the channel estimation error. By considering random realizations of the MMSE channel estimate and the channel estimation error during an arbitrary coherence block, it holds true that $\widehat{\boldsymbol{g}}_{mk}$ and $\widetilde{\boldsymbol{g}}_{mk}$ are uncorrelated and are respectively distributed as [7]

$$\begin{aligned}\widehat{\boldsymbol{g}}_{mk} &\sim \mathcal{CN}\left(0, \frac{\tau\rho_p\beta_{mk}^2}{\rho_p\tau\sum_{j=1}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2 + 1}\mathbf{I}_{\mathrm{N}}\right), \\ \widetilde{\boldsymbol{g}}_{mk} &\sim \mathcal{CN}\left(0, \left(\beta_{mk} - \frac{\tau\rho_p\beta_{mk}^2}{\rho_p\tau\sum_{j=1}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2 + 1}\right)\mathbf{I}_{\mathrm{N}}\right)\end{aligned} \tag{4}$$

A. Downlink Payload

The downlink signal intended to each UE is precoded at the APs with conjugate beamforming. Accordingly, the transmit signal of the mth AP to all users is

$$\boldsymbol{y}_{m,d} = \sqrt{\rho_d}\sum_{k=1}^{K}\sqrt{\eta_{mk}}\widehat{\boldsymbol{g}}_{mk}^* d_k \tag{5}$$

where d_k with $\mathbb{E}\{|d_k|^2\} = 1$, denotes the data symbol intended for user k, ρ_d accounts for the downlink transmit power and η_{mk}, the power coefficient between from the mth AP to user k

chosen to satisfy $\mathbb{E}\{\|\boldsymbol{y}_{m,d}\|^2\} \leq \rho_d$ at each AP. The received signal at the kth user is given by

$$r_k = \sqrt{\rho_d}\left(\sum_{m=1}^{M}\sum_{j=1}^{K}\sqrt{\eta_{mj}}\boldsymbol{g}_{mk}^{\top}\widehat{\boldsymbol{g}}_{mj}^{*}d_j\right) + n_k, \tag{6}$$

where $n_k \sim \mathcal{CN}(0,1)$ denotes the additive white Gaussian noise at user k. Similarly to [7], [10], we assume that UEs have only statistical knowledge of the channel estimate. Accordingly, the net downlink rate of the kth user is given by [10]

$$R_k = \mathcal{B}\left(1-\frac{\tau}{T_c}\right)\times\log_2\left(1+\frac{\rho_d\left(\sum_{m=1}^{M}\sqrt{\eta_{mk}}N\nu_{mk}\right)^2}{N\rho_d\sum_{m=1}^{M}\eta_{mk}\beta_{mk}\nu_{mk}+\rho_d N\sum_{j\neq k}^{K}\mathrm{I}_{kj}+1}\right) \tag{7}$$

where $\mathcal{B}$ is the spectral bandwidth, $\left(1-\frac{\tau}{T_c}\right)$ is the fraction of the channel uses spent for downlink data transmission and

$$\begin{aligned}\nu_{mk} &\triangleq \frac{\tau\rho_p\beta_{mk}^2}{\rho_p\tau\sum_{j=1}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^{H}\boldsymbol{\phi}_j\right|^2+1},\\ \mathrm{I}_{kj} &\triangleq N\left(\sum_{m=1}^{M}\sqrt{\eta_{mj}}\frac{\nu_{mj}}{\beta_{mj}}\beta_{mk}\right)^2\left|\boldsymbol{\phi}_j^{\top}\boldsymbol{\phi}_k^{*}\right|^2\\ &\quad+\sum_{m=1}^{M}\eta_{mj}\beta_{mk}\nu_{mj}\end{aligned} \tag{8}$$

B. Dynamic Radio Resource Assignment

For 5G and beyond cells structured networks, radio resources are intended to be allocated in terms of resource-block (RB) on a per-cell basis [26]. For instance, given the signal bearer is modulated using orthogonal frequency division multiple access (OFDMA), each base station assigns each RB to only one user at a given time. The same RBs can be reused simultaneously within different cells which will create inter-cell interference. Contrarily, the users in cell-free systems do not experience cells boundaries [27]. In a TDD cell-free network, all APs coherently communicate with all users on the same radio resource at the same time. The network may incur substantial amount of multi-user interference, on top of unavoidable pilot contamination, which will severely affect the performance of the system. In this work, to palliate the damaging effect of both interference and pilot contamination, we consider dynamic assignment of bandwidth for the users. Specifically, the available bandwidth is divided equally into L orthogonal parts, which we refer to as *sub-frequency bands* that are allocated to the users in order to diminish the effect of interference. In other words, a user may be scheduled only on few sub-frequency bands in contrast to existing literature [7]–[10] where a user is scheduled over the entire bandwidth.

III. Resource Allocation Problem Formulation

Before we proceed to formulate the radio resource allocation problem, let us define the binary variable $\vartheta_k^{\ell} \in \{0,1\}$, and $\boldsymbol{\vartheta} = \{\{\vartheta_k^{\ell}\}_{k=1}^{K}\}_{\ell=1}^{L}$. The value of ϑ_k^{ℓ} takes one if the APs can communicate with the k UE on the ℓth sub-frequency band, otherwise it takes value 0. Moreover, the power coefficient allocation from one AP to any use on each sub-frequency band is assumed to be fixed. Our goal is to dynamically assign the available radio resources for the purpose of maximizing the total throughput of the network subject to QoS constraint. The optimization problem is mathematically formulated as follow

$$\max_{\boldsymbol{\vartheta}\in\{0,1\}^{(K\times L)}} f(\boldsymbol{\vartheta}) \triangleq \sum_{k=1}^{K}\sum_{\ell=1}^{L}\vartheta_k^{\ell}\widehat{R}_k^{\ell} \tag{9a}$$

$$\text{s.t. } \overline{R}_k \leq \sum_{\ell=1}^{L}\vartheta_k^{\ell}\widehat{R}_k^{\ell},\ \forall\ell,\ \forall k \tag{9b}$$

where $\overline{R}_k$ is the minimum rate requirement for user k and $\widehat{R}_k^{\ell}$ in units of bits/s/Hz, the net downlink throughput, on the lth sub-frequency band.

$$\widehat{R}_k^{\ell} = \frac{\mathcal{B}}{L}\left(1-\frac{\tau}{T_c}\right)\times\log_2\left(1+\frac{\rho_d\left(\sum_{m=1}^{M}N\sqrt{\eta_{mk}^{\ell}}\nu_{mk}^{\ell}\right)^2}{N\rho_d\sum_{m=1}^{M}\eta_{mk}^{\ell}\beta_{mk}\nu_{mk}^{\ell}+\rho_d N\sum_{j\neq k}^{K}\mathrm{I}_{kj}^{\ell}\vartheta_j^{\ell}+1}\right) \tag{10}$$

where $\frac{\mathcal{B}}{L}$ is the width of a sub-frequency band, η_{mk}^{ℓ} denotes the power coefficient on the lth frequency sub-band and

$$\begin{aligned}\mathrm{I}_{kj}^{\ell} &\triangleq N\left(\sum_{m=1}^{M}\sqrt{\eta_{mk}^{\ell}}\frac{\nu_{mj}^{\ell}}{\beta_{mj}}\beta_{mk}\right)^2\left|\boldsymbol{\phi}_j^{\top}\boldsymbol{\phi}_k^{*}\right|^2\\ &\quad+\sum_{m=1}^{M}\eta_{mj}^{\ell}\beta_{mk}\nu_{mj}^{\ell}\\ \nu_{mk}^{\ell} &\triangleq \frac{\tau\rho_p\beta_{mk}^2}{\tau\rho_p\beta_{mk}+\rho_p\tau\sum_{j\neq k}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^{H}\boldsymbol{\phi}_j\right|^2\vartheta_j^{\ell}+1}\end{aligned} \tag{11}$$

Remark 1: From the expression of ν_{mk}^{ℓ} in (11)*, one may notice that for a given sub-frequency band l and two users k, k_1, if $\vartheta_k^{\ell} = 1$ while $\vartheta_{k_1}^{\ell} = 0$, it can be observed that the channel estimate of user k on sub-frequency band l will not be affected by pilot contamination from user k_1 regardless that both users are assigned the same pilot vector. This leads to believe that dynamic sub-frequency band allocation can potentially palliate the effect of pilot contamination.*

Problem (9) is a mixed-integer programming which is more likely difficult to be solved in the global optimum [28]. Therefore, it of interest to us to investigate its complexity status.

Proposition 1: The optimization problem formulated in (9) *is NP-hard.*

Proof: The proof of Proposition 1 is relegated to Appendix A ■

Proposition 1 advocates that it is impossible to find a polynomial time solvable algorithm capable of seeking the global optimum for problem (9) unless P = NP. Nevertheless, a local solution can be found by mean of classical approaches such that Lagrangian approach.

A. Lagrangian-Based Solution

Before proceeding in describing the Lagrangian-based frequency band allocation solution, we rewrite the binary constraint $\vartheta_k^\ell \in \{0,1\}$as the difference of two concave-functions

$$\begin{aligned} &0 \le \vartheta_k^\ell \le 1 \\ &\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(\vartheta_k^\ell - \left(\vartheta_k^\ell\right)^2\right) \le 0 \end{aligned} \tag{12}$$

and define the following set:

$$\mathcal{D} \triangleq \left\{ \boldsymbol{\vartheta} : \begin{cases} \text{(9b)} \\ 0 \le \boldsymbol{\vartheta} \le 1 \end{cases} \right\}$$

Problem (9) can be equivalently recast as

$$\begin{aligned} &\min_{\boldsymbol{\vartheta}\in\mathcal{D}} \; -f(\boldsymbol{\vartheta}) \\ &\text{s.t.} \; \sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(\vartheta_k^\ell - \left(\vartheta_k^\ell\right)^2\right) \le 0 \;\; \forall k, \forall \ell \end{aligned} \tag{13}$$

We have the following proposition

Proposition 2: For large value of $\varsigma \gg 1$, the optimization problem (13) *is equivalent to*

$$\begin{aligned} &\min_{\boldsymbol{\vartheta}} \; \mathcal{L}(\boldsymbol{\vartheta},\varsigma) \triangleq -f(\boldsymbol{\vartheta}) + \varsigma\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(\vartheta_k^\ell - \left(\vartheta_k^\ell\right)^2\right) \\ &\textit{s.t. } \boldsymbol{\vartheta} \in \mathcal{D} \end{aligned} \tag{14}$$

The proof of Proposition 2 can be derived in a similar fashion as the proof of [29, Proposition 1] or [30, Proposition 2]. It can be inferred from Proposition 2 that strong duality holds. Consequently, our goal is to solve the optimization problem (14) which we rewrite as

$$\begin{aligned} &\max_{0\le\boldsymbol{\vartheta}\le 1} \; \sum_{k=1}^{K}\sum_{\ell=1}^{L}\vartheta_k^\ell \widehat{R}_k^\ell - \varsigma\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(\vartheta_k^\ell - \left(\vartheta_k^\ell\right)^2\right) \\ &\text{s.t. } \overline{R}_k \le \sum_{\ell=1}^{L}\vartheta_k^\ell \widehat{R}_k^\ell, \; \forall k \\ &\quad 1 \le \sum_{\ell=1}^{L}\vartheta_k^\ell, \; \forall k \end{aligned} \tag{15}$$

The Lagrangian associated with the optimization problem (15) is formulated as

$$\begin{aligned} \overline{\mathcal{L}}(\boldsymbol{\vartheta},\boldsymbol{\gamma},\boldsymbol{\upsilon}) = &\sum_{k=1}^{K}\gamma_k\left(\overline{R}_k - \sum_{\ell=1}^{L}\vartheta_k^\ell \widehat{R}_k^\ell\right) \\ &+\sum_{k=1}^{K}\upsilon_k\left(1 - \sum_{\ell=1}^{L}\vartheta_k^\ell\right) \\ &-\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(\vartheta_k^l \widehat{R}_k^\ell - \varsigma\left(\vartheta_k^\ell - \left(\vartheta_k^\ell\right)^2\right)\right) \end{aligned} \tag{16}$$

where $\boldsymbol{\gamma} \ge 0, \boldsymbol{\upsilon} \ge 0$ are the Lagrange multipliers. Define

$$\begin{aligned} \frac{\partial\overline{\mathcal{L}}(\boldsymbol{\vartheta},\boldsymbol{\gamma},\boldsymbol{\upsilon})}{\partial\vartheta_k^\ell} = &-\sum_{j\neq k}^{K}\vartheta_j^\ell\gamma_j\frac{\partial\widehat{R}_j^\ell}{\partial\vartheta_k^\ell} - \gamma_k\left(\vartheta_k^\ell\frac{\partial\widehat{R}_k^\ell}{\partial\vartheta_k^\ell} + \widehat{R}_k^\ell\right) \\ &-\widehat{R}_k^\ell - \vartheta_k^\ell\frac{\partial\widehat{R}_k^\ell}{\partial\vartheta_k^\ell} - \sum_{j\neq k}^{K}\vartheta_j^\ell\frac{\partial\widehat{R}_j^\ell}{\partial\vartheta_k^\ell} \\ &+\varsigma\left(1 - 2\vartheta_k^\ell\right) \end{aligned} \tag{17}$$

The Karush-Kuhn-Tucker (KKT) conditions [31] are given by

$$\begin{aligned} \frac{\partial\overline{\mathcal{L}}(\boldsymbol{\vartheta}^\star,\boldsymbol{\gamma}^\star,\boldsymbol{\upsilon}^\star)}{\partial{\vartheta_k^\ell}^\star} &= 0, \quad \forall\ell, \forall k \\ \upsilon_k^\star\left(1 - \sum_{\ell=1}^{L}{\vartheta_k^\ell}^\star\right) &= 0, \quad \forall k \\ {\gamma_k}^\star\left(\overline{R}_k - \sum_{\ell=1}^{L}\vartheta_k^\ell\widehat{R}_k^\ell\right) &= 0, \quad \forall\ell, \forall k \end{aligned}$$

Considering that it is extremely complicated to find a closed form solution from (17), a stationary point $\boldsymbol{\vartheta}^\star$ can be found by invoking the gradient descent method [32] where at the nth iteration.

$$\begin{aligned} \vartheta_k^\ell[n] = \Bigg[&\vartheta_k^\ell[n-1] \\ &-\delta_\vartheta\frac{\partial\overline{\mathcal{L}}(\boldsymbol{\vartheta},\boldsymbol{\gamma}[n-1],\boldsymbol{\upsilon}[n-1])}{\partial\vartheta_k^\ell}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}[n-1]}\Bigg]_0^1 \forall\ell, \quad \forall k \end{aligned} \tag{18}$$

where $\delta_\vartheta > 0$ is a sufficiently small positive number. The Lagrange multipliers can be updated by the sub-gradient approach [32, Chapter 6]. The Lagrangian-based method is summarize in Algorithm 1 where $\delta_\upsilon > 0$, δ_γ is a sufficiently small positive number. Let κ_1 denotes the time complexity required to compute the gradient of ϑ_k^ℓ, with a solution accuracy of ϵ_1. The worst case computational complexity for the gradient descent is $\mathcal{O}(KL\kappa_1\log(\epsilon_1^{-1}))$. Since the objective function is differentiable, the complexity of the sub-gradient approach to calculate the Lagrange multipliers is computing in a similar fashion as the computational complexity of the gradient method [33]. Then, the overall computational complexity of the Lagrangian-based solution to find a stationary point to problem (15) is given by $\mathcal{O}(K\kappa_2\log(\epsilon_2^{-1}) + K\kappa_3\log(\epsilon_3^{-1}))\cdot\mathcal{O}(KL\kappa_1\log(\epsilon_1^{-1}))$ where $\mathcal{O}(\kappa_2\log(\epsilon_2^{-1}))$ and $\mathcal{O}(\kappa_3\log(\epsilon_3^{-1}))$ are the computational complexity of the Lagrangian multiplier $\boldsymbol{\gamma}$ and $\boldsymbol{\upsilon}$, respectively. ϵ_2, ϵ_3 are the solution accuracy for the sub-gradient method to compute $\boldsymbol{\gamma}$ and $\boldsymbol{\upsilon}$, respectively. The convergences condition can be chosen to be $(\overline{\mathcal{L}}(\boldsymbol{\vartheta}[n],\boldsymbol{\gamma}[n],\boldsymbol{\upsilon}[n]) - \overline{\mathcal{L}}(\boldsymbol{\vartheta}[n-1],\boldsymbol{\gamma}[n-1],\boldsymbol{\upsilon}[n-1]))/\overline{\mathcal{L}}(\boldsymbol{\vartheta}[n-1],\boldsymbol{\gamma}[n-1],\boldsymbol{\upsilon}[n-1]) < \epsilon_1$. It is worth noticing that the convergence of Algorithm 1 can be extremely slow especially when higher accuracy is required since the local solution is computed by means of gradient descent and sub-gradient approaches.

Although the Lagrangian-based scheme can find a local solution to problem (9), this method has rather low-rate of convergence. This motivates us to provide an alternative approach with higher rates of convergence and that can enable the attenuation of the effect of multi-user interference and

Algorithm 1 Lagrangian-Based Approach to Solve Problem (15)

Input A feasible $\boldsymbol{\vartheta}[0]$ to problem (15) and a solution accuracy $\epsilon_1 > 0$.
Set $n \leftarrow 0$;
repeat
  Initialize $\boldsymbol{\gamma}[0], \boldsymbol{\upsilon}[0]$ and set $m \leftarrow 0$;
  repeat
    $m \leftarrow m+1$;
    Update $\boldsymbol{\gamma}, \boldsymbol{\upsilon}$ as

$$\begin{aligned}\upsilon_k[m] &\leftarrow \left(\upsilon_k[m-1] + \delta_\upsilon \left(1 - \sum_{\ell=1}^{L} \vartheta_k^\ell[n]\right)\right)^+, \forall k\\ \gamma_k[m] &\leftarrow \Bigg(\gamma_k[m-1] + \delta_\gamma \\ &\quad \left(\overline{R}_k - \sum_{\ell=1}^{L} \vartheta_k^\ell[n] \widehat{R}_k^\ell\right)\Bigg)^+, \forall k \end{aligned} \tag{19}$$

  until the subgradient approach converges
  $n \leftarrow n+1$;
  Update $\boldsymbol{\vartheta}[n]$ by $\vartheta_k^\ell[n] \leftarrow \left[\vartheta_k^\ell[n-1] - \delta_\vartheta \frac{\partial \overline{\mathcal{L}}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}[m], \boldsymbol{\upsilon}[m])}{\partial \vartheta_k^\ell}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}[n-1]}\right]_0^1, \ \forall \ell, \ \forall k.$
until the gradient descent method converges.
Output the stationary solutions $\boldsymbol{\vartheta}[n]$.

pilot contamination. This is done from the standpoint of user grouping.

IV. Scheduling-Based Resource Allocation

Our proposed scheduling approach aims at clustering users to palliate the effect of multi-user interference. The motivation lies in the fact that by grouping together users that cause small amount of interference to each other, the intra-group interference can be made as small as possible. Regarding inter-group interference and pilot contamination interference, their effect can be alleviate by scheduling each group on different sub-frequency bands.

Our scheduling procedure is executed at the CPU side and is involved large scale fading of the users' channel and some other parameters. The group coefficients are sent to the access points for downlink data payload. As discussed subsequently, the users will be partitioned into groups from the perspective of graph theory. To be more specific, a weighted graph will be constructed with the weight being the mutual interference, which needs to be defined, between two users. Towards this end, we consider the interference between two users j and k (we assume that both users are assigned same pilot sequence) in the absence of all the other users. It is written as

$$\begin{aligned}\widehat{\mathrm{I}}_{kj} &= N\left(\sum_{m=1}^{M} \frac{\sqrt{\eta_{mk}}\tau\rho_p\beta_{mj}\beta_{mk}}{\tau\rho_p\beta_{mj} + \rho_p\tau\beta_{mk} + 1}\right)^2 \\ &\quad + \sum_{m=1}^{M} \frac{\eta_{mj}\tau\rho_p\beta_{mk}\beta_{mj}^2}{\tau\rho_p\beta_{mj} + \rho_p\tau\beta_{mk} + 1} \end{aligned} \tag{20}$$

and we define the mutual interference weight as follow

$$\omega_{kj} = \omega_{jk} \triangleq \frac{\widehat{\mathrm{I}}_{kj} + \widehat{\mathrm{I}}_{jk}}{2} \tag{21}$$

A. Graphical Modeling

We start by constructing a spatial correlation graph that captures the level of interference for a set of users which are active simultaneously. More specifically, we design an undirected interference graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. The set of vertices $\mathcal{V}$, with $|\mathcal{V}| = K$, of the considered graph accounts for the users in the coverage area while each edge $e_{kj} \in \mathcal{E}$ is associated with a weight ω_{kj} defined in (21) set to capture the interference level. Using the constructed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, our goal is to formulate an optimization problem to partition users into groups. The considered optimization framework aims at minimizing the mutual interference weight between the users that are assigned to the same group.

Define the following variable

$$x_{k,c} = \begin{cases} 1 & \text{if user } k \text{ is allocated to the } c\text{-th group} \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

The user's scheduling problem is formulated as the following combinatorial optimization problem

$$\begin{aligned} &\max_{x_{k,c} \in \{0,1\}, \forall c, \forall k} \sum_{c=1}^{C} \sum_{k \in \mathcal{V}} \sum_{j \in \mathcal{V}, j \neq k} \omega_{kj}(1 - x_{k,c}) x_{j,c} \\ &\text{s.t. } 1 \leq \sum_{c=1}^{C} x_{k,c} \leq \alpha, \forall k \in \mathcal{V} \\ &\qquad \sum_{k \in \mathcal{V}} x_{k,c} \leq \tau, \forall c = 1, \ldots, C, \end{aligned} \tag{23}$$

where C denote the total number of groups and α, the maximum number of groups a user is permitted to be assigned to concurrently. We also refrain each group to have a maximum number of users, i.e τ, to avoid intra-group pilot contamination. Problem (23) is a combinatorial optimization problem and therefore very difficult to solve. In the next section, we propose an algorithm to seek an efficient solution to problem (23).

V. Semidefinite Relaxation-Based Approach

In this section, we turn our attention in developing a polynomial-time solvable algorithm to find an efficient sub-optimal solution to problem (23). However, let us first rewrite the optimization problem (23) in a more compact way. Towards this end, we define the following change of variables

$$\begin{aligned} \mathbf{x}_c &\triangleq (x_{1,c} \cdots, x_{K,c})^\top, \ \mathbf{y}_c \triangleq 2\mathbf{x}_c - \mathbf{1}_K, \\ \mathbf{W} &\triangleq \begin{pmatrix} 0 & w_{21} & \cdots & w_{K1} \\ w_{12} & 0 & \cdots & w_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1K} & w_{2K} & \cdots & 0 \end{pmatrix} \end{aligned} \tag{24}$$

where $\mathbf{1}_K$ is a column vector which entries are 1. Using (24), problem (23) can be equivalently reformulated as

$$\begin{aligned}
&\min \ \frac{1}{4}\sum_{c=1}^{C}\mathbf{y}_c^\top \mathbf{W}\mathbf{y}_c\\
&\text{s.t. } (2-C)\mathbf{1}_K \le \sum_{c=1}^{C}\mathbf{y}_c \le \overline{\alpha}\mathbf{1}_K\\
&\qquad \operatorname{Tr}\left(\operatorname{diag}\left(\mathbf{y}_c\right)\right) \le \overline{\tau},\ \forall c\\
&\qquad \mathbf{y}_c \in \{-1,1\}^K, \forall c
\end{aligned} \tag{25}$$

where $\overline{\alpha} = 2\alpha - C$ and $\overline{\tau} = 2\tau - K$. Furthermore, we want to express the variables in problem (25) to have only quadratic terms which have nicer property. To do so, let us define the following change of variables

$$\begin{aligned}
\mathbf{z}_c &\triangleq \left(\mathbf{y}_c^\top, 1\right)^\top,\ \ \widetilde{\mathbf{W}} \triangleq \begin{pmatrix}\mathbf{W} & \mathbf{0}_{K\times 1}\\ \mathbf{0}_{1\times K} & 0\end{pmatrix}\\
\mathbf{W}_1 &\triangleq \begin{pmatrix}\mathbf{0}_{K\times K} & \boldsymbol{\pi}\\ \boldsymbol{\pi}^\top & 0\end{pmatrix},\ \ \mathbf{Q}_k \triangleq \begin{pmatrix}\mathbf{0}_{K\times K} & \widetilde{\mathbf{p}}\\ \widetilde{\mathbf{p}}^\top & 0\end{pmatrix}
\end{aligned} \tag{26}$$

where $\boldsymbol{\pi} \triangleq \frac{1}{2}\mathbf{1}_{K\times 1}$, and $\widetilde{\mathbf{p}}$ is a $K\times 1$ vector with entries zeros except the kth entry which is $\frac{1}{2}$. Therefore, by combining (26) and (25), the combinatorial optimization problem (23) can be compactly written as

$$\begin{aligned}
&\min \ \frac{1}{4}\sum_{c=1}^{C}\mathbf{z}_c^\top \widetilde{\mathbf{W}}\mathbf{z}_c\\
&\text{s.t. } 2-C \le \sum_{c=1}^{C}\mathbf{z}_c^\top \mathbf{Q}_k\mathbf{z}_c \le \overline{\alpha},\ \forall k\\
&\qquad \mathbf{z}_c^\top \mathbf{W}_1\mathbf{z}_c \le \overline{\tau},\ \forall c\\
&\qquad \mathbf{z}_c \in \{-1,1\}^{K+1},\ \mathbf{z}_{c,K+1} = 1,\ \forall c
\end{aligned} \tag{27}$$

To find an efficient sub-optimal solution to problem (27), we resort to the semidefinite relaxation (SDR) method [23] which has been widely known as a powerful tools capable of handling combinatorial optimization problems [34]. The application of SDR method comes after a key observation namely, $\mathbf{z}_c^\top\mathbf{Q}_k\mathbf{z}_c = \operatorname{Tr}\left(\mathbf{z}_c^\top\mathbf{Q}_k\mathbf{z}_c\right) = \operatorname{Tr}\left(\mathbf{Q}_k\mathbf{z}_c\mathbf{z}_c^\top\right)$. Following [23], SDR approach consists of making the change of variables $\mathrm{Z}_c = \mathbf{z}_c\mathbf{z}_c^\top$. This change of variable is equivalent to substituting $\mathrm{Z}_c = \mathbf{z}_c\mathbf{z}_c^\top$ by using a positive semidefinite (PSD) matrix Z together with an additional rank one constraint. Consequently, the combinatorial problem (27) can be equivalently recast as

$$\begin{aligned}
&\min \ \frac{1}{4}\sum_{c=1}^{C}\operatorname{Tr}\left(\widetilde{\mathbf{W}}\mathbf{Z}_c\right)\\
&\text{s.t. } 2-C \le \sum_{c=1}^{C}\operatorname{Tr}\left(\mathbf{Q}_k\mathbf{Z}_c\right) \le \overline{\alpha},\ \forall k\\
&\qquad \operatorname{Tr}\left(\mathbf{W}_1\mathbf{Z}_c\right) \le \overline{\tau},\ \forall c\\
&\qquad \mathbf{Z}_c \succeq 0,\ \operatorname{diag}\left(\mathrm{Z}_c\right) = \mathbf{1}_{K+1},\ \forall c\\
&\qquad \operatorname{rank}\left(\mathrm{Z}_c\right) = 1,\ \forall c
\end{aligned} \tag{28}$$

The optimization problem (29) is a non-convex problem because the convexity of the constraint $\operatorname{rank}(\mathrm{Z}) = 1$. Therefore, it is as difficult to finding the optimal solution to the optimization problem (28) than solving its counterpart combinatorial problem (27). However, dropping the rank one constraint, i.e. $\operatorname{rank}(\mathrm{Z}) = 1$, the optimization problem (28) can be relaxed as

$$\begin{aligned}
&\min \ \frac{1}{4}\sum_{c=1}^{C}\operatorname{Tr}\left(\widetilde{\mathbf{W}}\mathbf{Z}_c\right)\\
&\text{s.t. } 2-C \le \sum_{c=1}^{C}\operatorname{Tr}\left(\mathbf{Q}_k\mathbf{Z}_c\right) \le \overline{\alpha},\ \forall k\\
&\qquad \operatorname{Tr}\left(\mathbf{W}_1\mathbf{Z}_c\right) \le \overline{\tau},\ \forall c\\
&\qquad \mathbf{Z}_c \succeq 0,\ \operatorname{diag}\left(\mathrm{Z}_c\right) = \mathbf{1}_{K+1},\ \forall c
\end{aligned} \tag{29}$$

Problem (29) is a standard convex optimization problem. Therefore, we can invoke interior-point based solvers such as CVX [35] to efficiently finding its global optimal solution with a worst case complexity of $\mathcal{O}\left((K+1)^{3.5}\log(\epsilon_4^{-1})\right)$ [36] where ϵ_4 is the solution accuracy. If the optimal solution obtained in solving problem (29) is rank one, it is therefore also the optimal solution to the optimization problem (28). However, it is not generally the case since problem (29) is a relaxed instance of problem (28). Consequently, solving (29) cannot not always guarantee to yield an optimal solution Z_c satisfying the rank one constraint. However, the optimal solution $\mathrm{Z}_c^\star \succeq 0, \forall c$ to problem (29) can serve as a basis for finding an approximate solution to the optimization problem (28).

One suggestion for obtaining an approximate solution is to resort to powerful randomization approaches [37] such as the Gaussian randomization [23]. We propose a randomized procedure in the same vein as the randomized technique in [37] that takes as input the optimal solution to problem (29) and outputs a feasible solution to problem (28). Given $\mathrm{Z}_c^\star \succeq 0, \forall c$, an optimal solution to problem (29), it can be decomposed using Cholesky factorization as $\mathrm{Z}_c^\star = \mathrm{V}_c^\top\mathrm{V}_c, \forall c$ with $\mathrm{V}_c = (\mathbf{v}_{c,1}, \cdots, \mathbf{v}_{c,K+1})$. The proposed randomized scheme is summarized below.

Algorithm 2 A Randomized Algorithm to Solve Problem (25)

input an optimal solution $\mathrm{Z}_c^\star, \forall c$ to problem (29).
Generate $\overline{L}$ vector samples $\boldsymbol{\xi}_c^\ell$, $\ell, = 1, \cdots, \overline{L}$ uniformly distributed on the unit sphere;
Generate $\boldsymbol{\zeta}_c^\ell, = (\zeta_{1,c}^\ell, \cdots, \zeta_{K+1,c}^\ell,)^\top$ according to

$$\zeta_{k,c}^\ell, = \begin{cases}1 & \text{if } \mathbf{v}_{c,k}^\top\boldsymbol{\xi}_c{}^\ell, \ge 0\\ -1 & \text{otherwise}\end{cases} \tag{30}$$

Verify that $\boldsymbol{\zeta}_c^\ell$, meet the constraint of (27).
Compute $\ell,^\star = \arg\min_{\ell,=1,\cdots,\overline{L}} \frac{1}{4}\sum_{c=1}^{C}\left(\boldsymbol{\zeta}_c^\ell,\right)^\top \mathbf{W}\boldsymbol{\zeta}_c^\ell,$;
output the solution $\widehat{\mathbf{z}}_c = \boldsymbol{\zeta}_c^{\ell,^\star}, \forall c$.

A. Disjoint Allocation of Sub-Frequency Band

The proposed grouping strategy enables to partition together users that cause less interference to each other. In other words, the proposed user grouping policy only tackles intra-group interference. We now turn our attention to investigate solutions to palliate inter-group interference. We advocate that inter-group interference can be eliminated by scheduling the groups on different sub-frequency bands. The problem of group scheduling is formulated as

$$\max_{\theta_c^\ell \in\{0,1\}, \forall \ell, \forall c} \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^\ell \widetilde{R}_{k,c}^\ell$$
$$\text{s.t. } \overline{R}_k \leq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\theta_c^\ell \widetilde{R}_{k,c}^\ell, \forall k$$
$$\theta_c^\ell \theta_{\tilde{c}}^\ell = 0, \forall c \neq \tilde{c},\ \forall \ell \tag{31}$$

where $\theta_c^\ell = 1$ if group c is scheduled for downlink transmission on the lth sub-frequency band, otherwise $\theta_c^\ell = 0$. Moreover, $\widetilde{R}_{k,c}^\ell$ is the rate of the kth user within group c on the lth sub-frequency band. It can be written as (32) (see bottom of the page), where $\eta_{mk}^{\ell,c}$ denotes the power coefficient from the mth AP to user k within group c and on the lth sub-frequency band, $x_{k,c}^\star$ is solution to problem (23), and the prelog factor $1 - \frac{|\Gamma(c)|}{T_c}$ accounts for the number of channel uses destined for downlink transmission on a given sub-frequency band. The power coefficients are assumed to be fixed during the sub-frequency band allocation. Moreover, a maximum of $|\Gamma(c)| \leq \tau$ users is scheduled within a given group which is exclusively assigned a given sub-frequency band. Therefore, during the uplink training stage on that particular sub-frequency band, the length of pilot sequences assigned to the users can simply be $|\Gamma(c)|$ which explains why the fraction of channels uses dedicated to downlink transmission on that particular sub-frequency band is $\frac{|\Gamma(c)|}{T_c}$. Accordingly, the interference term $\mathrm{I}_{kj,c}$ is given by $\mathrm{I}_{kj,c} = \sum_{m=1}^{M}\eta_{mj}^{\ell,c}\beta_{mk}\nu_{mj,c} + 1$. The pilot contamination term that appears initially in the interference I_{kj} in (32) vanishes thanks to our grouping strategy that allocates orthogonal pilot vector to users that belong to the same group.

Remark 2: In (32), $\nu_{mk,c}$ *is given by* $\nu_{mk,c} = \frac{\tau\rho_p\beta_{mk}^2}{\tau\rho_p\beta_{mk}+\rho_p\tau\sum_{j\neq k, j\notin\Gamma(c)}^{K}\beta_{mj}\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2+1}$. *The denominator excludes users in same group because our user partition strategy restricted the maximum number of users in each group to be τ in problem* (23)*. Consequently, the users that belong to the same group are assigned pair-wisely orthogonal training vector. In addition, the group scheduling problem* (31) *prevents two different groups to be scheduled on the same sub-frequency band. By combining our proposed user partition strategy and proposed group scheduling approach, the pilot contamination in the denominator of $\nu_{mk,c}$ will vanish. Therefore, the combination of both approaches will mitigate the effect of pilot contamination.*

The optimization problem (31) can be recast as

$$\max_{\theta_c^\ell \in\{0,1\}, \forall \ell, \forall c} \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^\ell \widetilde{R}_{k,c}^\ell \tag{33a}$$
$$\text{s.t. } \overline{R}_k \leq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\theta_c^\ell \widetilde{R}_{k,c}^\ell, \forall k \tag{33b}$$
$$\theta_c^\ell + \theta_{\tilde{c}}^\ell \leq 1, \forall c \neq \tilde{c},\ \forall \ell \tag{33c}$$

Let $\boldsymbol{\theta} \triangleq \{\{\theta_c^\ell\}_{\ell=1}^L\}_{c=1}^C$ and define the following set

$$\mathcal{X} \triangleq \left\{\boldsymbol{\theta} : \begin{cases}(33\text{b}), (33\text{c}) \\ 0 \leq \boldsymbol{\theta} \leq 1\end{cases}\right\}$$

It is straightforward to demonstrate that the set $\mathcal{X}$ is a convex set. Using Proposition 2, problem (33) can be reformulated as

$$\max_{\boldsymbol{\theta}\in\mathcal{X}} \psi(\boldsymbol{\theta}) \triangleq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^\ell \widetilde{R}_{k,c}^\ell - \varsigma_1 \sum_{\ell=1}^{L}\sum_{c=1}^{C}\left(\theta_c^\ell - (\theta_c^\ell)^2\right) \tag{34}$$

The optimization problem (34) is a non-convex optimization problem because the objective function is not a concave function. In the subsequent subsection, we propose an algorithm to efficiently solve problem (34).

B. Sequential Convex Approximation-Based Solution

The proposed approach is motivated by the sequential parametric convex approximation (SCA) method [38]. Let $\overline{\boldsymbol{\theta}}$ be a feasible point to problem (34). Our goal is to conservatively approximate the objective function of problem (34) at the point $\overline{\boldsymbol{\theta}}$ by invoking first order convex condition [31]. More specifically,

$$(\theta_c^\ell)^2 \geq (\overline{\theta_c}^\ell)^2 + 2\overline{\theta_c}^\ell\left(\theta_c^\ell - \overline{\theta_c}^\ell\right) \tag{35}$$

By using (35), the optimization problem (34) can be conservatively approximated by

$$\max_{\boldsymbol{\theta}\in\mathcal{X}} \overline{\psi}(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}) \triangleq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^\ell \widetilde{R}_{k,c}^\ell - \varsigma_1 \sum_{\ell=1}^{L}\sum_{c=1}^{C}\left(\theta_c^\ell - (\overline{\theta_c}^\ell)^2 - 2\overline{\theta_c}^\ell\left(\theta_c^\ell - \overline{\theta_c}^\ell\right)\right) \tag{36}$$

The optimization problem (36) is a convex problem and therefore can be solved via interior-point based solvers such as CVX [35]. It is worth recalling that we obtained problem (36) by a restrictive approximation of the non convex optimization problem (34) around the feasible point $\overline{\boldsymbol{\theta}}$. Intuitively, the

$$\widetilde{R}_{k,c}^\ell \triangleq x_{k,c}^\star \mathcal{B}\left(1 - \frac{|\Gamma(c)|}{T_c}\right)\log_2\left(1 + \frac{\rho_d\left(\sum_{m=1}^{M} N\sqrt{\eta_{mk}^{\ell,c}}\nu_{mk,c}\right)^2}{N\rho_d\sum_{m=1}^{M}\eta_{mk}^{\ell,c}\beta_{mk}\nu_{mk,c} + \rho_d N\sum_{j\in\Gamma(c), j\neq k}^{K}\mathrm{I}_{kj,c} + 1}\right) \tag{32}$$

performance of this approximation can be further enhanced by simply using the optimal solution obtained from previous iteration and by sequentially solving the standard convex problem (36). To be more specific, denote $\widehat{\boldsymbol{\theta}}[n]$ the solution of the problem (36) at the nth iteration and set $\overline{\boldsymbol{\theta}}[n] = \widehat{\boldsymbol{\theta}}[n]$. We solve the following convex optimization problem at the $(n+1)$-iteration

$$\widehat{\boldsymbol{\theta}}[n+1] = \arg\max_{\boldsymbol{\theta}\in\mathcal{X}} \overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}[n]\right) \tag{37}$$

The proposed SCA-based algorithm to obtain solution to problem (34) is summarized in Algorithm 3.

Algorithm 3 SCA-Based Algorithm to Solve Problem (34)

1: **Input** A solution accuracy $\epsilon_5 > 0$ and a feasible point $\overline{\boldsymbol{\theta}}[0]$ for problem (34).
2: Set $n = 0$;
3: **Repeat**
4: $n = n + 1$;
5: Compute $\widehat{\boldsymbol{\theta}}[n]$ by solving problem (37) using CVX [35];
6: Update $\overline{\boldsymbol{\theta}}[n] = \widehat{\boldsymbol{\theta}}[n]$.
7: **Until** a predefined stopping criterion.
8: **Output** the approximated solutions $\widehat{\boldsymbol{\theta}}[n]$.

There are CL variables in the optimization problem (36) and $K + C(C-1)L$ linear and convex constraints. Therefore, the worst-case computational complexity of the SCA-based Algorithm 3 is $\mathcal{O}\left((CL)^3\left(K + C(C-1)L\right)\right)$. The stopping criterion can be chosen to be $(\overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}[n]\right) - \overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}[n-1]\right))/(\overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}[n-1]\right)) > \epsilon_5$. To further address the efficiency of the proposed SCA-based Algorithm 3, it is important to investigate its convergence analysis. This is done through next subsection.

C. Convergence Analysis

Theorem 1: The sequence of objective function $\left\{\overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n];\overline{\boldsymbol{\theta}}[n]\right)\right\}_{n=1}^{+\infty}$ *generated by the proposed SCA Algorithm 3 is a convergent sequence.*

Proof: At $(n+1)$-th iteration, the solution $\widehat{\boldsymbol{\theta}}[n]$ is used as an iterate value (ref. step 6) to problem (37). Therefore

$$\begin{aligned}\overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n+1];\overline{\boldsymbol{\theta}}[n]\right) &= \overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n+1];\widehat{\boldsymbol{\theta}}[n]\right)\\ &\geq \overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n];\widehat{\boldsymbol{\theta}}[n]\right)\\ &= \psi\left(\widehat{\boldsymbol{\theta}}[n]\right)\\ &\overset{(\bar{a})}{\geq} \overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n];\overline{\boldsymbol{\theta}}[n-1]\right)\end{aligned}$$

where $(\bar{a})$ follows from (35). The objective function is non-decreasing at each iteration. Moreover, $\boldsymbol{\theta}$ is bounded above. Therefore, the sequence $\left\{\overline{\psi}\left(\widehat{\boldsymbol{\theta}}[n];\overline{\boldsymbol{\theta}}[n]\right)\right\}_{n=1}^{+\infty}$ converges.

Corollary 1: Any limit point of sequence $\{\widehat{\mathbf{y}}[n]\}_{n=1}^{+\infty}$ *for problem* (37) *generated by the proposed Algorithm 3 is a stationary point to the optimization problem* (34).

Proof: Firstly, it is worth noticing that the function $\overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}\right)$ is a continuous function of both $\boldsymbol{\theta}$ and $\overline{\boldsymbol{\theta}}$. Secondly, the function $\overline{\psi}\left(\boldsymbol{\theta};\overline{\boldsymbol{\theta}}\right)$ is a tight lower bound to $\psi\left(\boldsymbol{\theta}\right)$, that is $\overline{\psi}\left(\overline{\boldsymbol{\theta}};\overline{\boldsymbol{\theta}}\right) = \psi\left(\overline{\boldsymbol{\theta}}\right)$. Thirdly, it holds true that $\frac{\partial\overline{\psi}(\boldsymbol{\theta};\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}\to\overline{\boldsymbol{\theta}}} = \frac{\partial\psi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}\to\overline{\boldsymbol{\theta}}}$. Hence, our proposed Algorithm 3 is special instance of the successive-upper-bound minimization (SUM) algorithm [39]. Based on [39, Theorem 1], any limit point to problem (37) generated by the SCA-based Algorithm 3 is a stationary point to problem (34). ■

D. Power Control

Once the solution of the sub-frequency allocation problem is found, the performance of our proposed scheme can be further improved by reducing the intra-group interference through power control. The power control optimization problem is formulated as

$$\begin{aligned}&\max_{\{\eta_{mk}^{\ell,c}\}} \ \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^{\ell\star}\widetilde{R}_{k,c}^{\ell}\\ &\text{s.t. } \overline{R}_k \leq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\theta_c^{\ell\star}\widetilde{R}_{k,c}^{\ell}, \forall k\\ &\quad N\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{c=1}^{C} x_{k,c}^{\star}\theta_c^{\ell\star}\eta_{mk}^{\ell,c}\nu_{mk,c} \leq 1, \ \forall m\\ &\quad \eta_{mk}^{\ell,c} \geq 0, \forall m, \forall k, \forall \ell, \forall c\end{aligned} \tag{38}$$

where the second constraint is the total power budget constraint at each AP. Define the following change of variables:

$$\begin{aligned}t_{mk}^{\ell,c} &\triangleq \sqrt{\eta_{mk}^{\ell,c}}\\ \gamma_{k,c}^{\ell} &\triangleq \frac{\rho_d N^2\left(\sum_{m=1}^{M} t_{mk}^{\ell,c}\nu_{mk,c}\right)^2}{N\rho_d\sum_{j\in\Gamma(c)}\sum_{m=1}^{M}\left(t_{mj}^{\ell,c}\right)^2\beta_{mk}\nu_{mj,c} + 1}\\ \Delta_{k,c} &= x_{k,c}^{\star}\mathcal{B}\left(1 - \frac{|\Gamma(c)|}{T_c}\right)\end{aligned} \tag{39}$$

By using (39), the optimization problem (38) can be equivalently reformulated as:

$$\max_{\{t_{mk}^{\ell,c}\},\{\gamma_k^{\ell,c}\}} \ \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^{\ell\star}\Delta_{k,c}\log_2\left(1 + \gamma_{k,c}^{\ell}\right) \tag{40a}$$

$$\text{s.t. } N\rho_d\sum_{j\in\Gamma(c)}\sum_{m=1}^{M}\left(t_{mj}^{\ell,c}\right)^2\beta_{mk}\nu_{mj,c} + 1 \leq \frac{\rho_d N^2\left(\sum_{m=1}^{M} t_{mk}^{\ell,c}\nu_{mk,c}\right)^2}{\gamma_{k,c}^{\ell}}, \forall k, \forall \ell, \forall c \tag{40b}$$

$$\overline{R}_k \leq \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\theta_c^{\ell\star}\Delta_{k,c}\log_2\left(1 + \gamma_{k,c}^{\ell}\right), \forall k \tag{40c}$$

$$N\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{c=1}^{C} x_{k,c}^{\star}\theta_c^{\ell\star}\left(t_{mk}^{\ell,c}\right)^2\nu_{mk,c} \leq 1, \ \forall m \tag{40d}$$

$$t_{mk}^{\ell,c} \geq 0, \forall m, \forall k, \forall \ell, \forall c \tag{40e}$$

$$\gamma_k^{\ell,c} \geq 0, \forall k, \forall \ell, \forall c \tag{40f}$$

One can notice that the objective function in (40a) is a concave function and that all constraints from (40c) to (40f) are convex constraints. However, problem (40) is a non-convex problem due to constraint (40b). The right hand side of (40b) possesses some nice properties which we can leverage to solve problem (40). Indeed, it is a convex function. This can be proven by deriving the Hessian. Therefore, problem (40) can also be solved by invoking the principle of SCA. This is done by conservatively approximate the right hand side of (40b) via the first order Taylor expansion. Accordingly, the optimization problem (40) can be approximated as

$$\begin{aligned}
&\max_{\{t_{mk}^{\ell,c}\},\{\gamma_k^{\ell,c}\}} \frac{1}{L}\sum_{\ell=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)} \theta_c^{\ell\star}\Delta_{k,c}\log_2\left(1+\gamma_{k,c}^{\ell}\right)\\
&\text{s.t. } N\rho_d\sum_{j\in\Gamma(c)}\sum_{m=1}^{M}\left(t_{mj}^{\ell,c}\right)^2\beta_{mk}\nu_{mj,c}+1\\
&\le \rho_d N^2\left(\frac{\left(\sum_{m=1}^{M}\overline{t_{mk}^{\ell,c}}\nu_{mk,c}\right)^2}{\overline{\gamma_{k,c}^{\ell}}}+\right.\\
&\frac{2\sum_{m=1}^{M}\overline{t_{mk}^{\ell,c}}\nu_{mk,c}\left(t_{mk}^{\ell,c}-\overline{t_{mk}^{\ell,c}}\right)}{\overline{\gamma_{k,c}^{\ell}}}\\
&\left.-\left(\frac{\sum_{m=1}^{M}\overline{t_{mk}^{\ell,c}}\nu_{mk,c}}{\overline{\gamma_{k,c}^{\ell}}}\right)^2\left(\gamma_{k,c}^{\ell}-\overline{\gamma_{k,c}^{\ell}}\right)\right),\\
&\forall k,\ \forall\ell,\ \forall c (40c)-(40f) \qquad (41)
\end{aligned}$$

At each iteration of the SCA approach, the closed-form expression for the solutions of problem (41) which can be found by solving the KKT conditions is given by

$$\begin{aligned}
\gamma_{k,c}^{\ell} &= \left[\frac{\Delta_{k,c}\theta_c^{\ell\star}\left(1+\omega_k\right)}{\mu_{k,c}^{l}L\ln(2)\rho_d N^2\left(\frac{\sum_{m=1}^{M}\overline{t_{mk}^{\ell,c}}\nu_{mk,c}}{\overline{\gamma_{k,c}^{\ell}}}\right)^2}-1\right]^{+}\\
t_{mk}^{\ell,c} &= \frac{\rho_d N\mu_{k,c}^{l}\overline{t_{mk}^{\ell,c}}}{\overline{\gamma_{k,c}^{\ell}}\left(\rho_d\sum_{j\in\Gamma(c)}\mu_{j,c}^{l}\beta_{mj}+\lambda_m x_{k,c}^{\star}\theta_c^{\ell\star}\right)} \qquad (42)
\end{aligned}$$

where $\{\mu_{k,c}^{l}\}$, $\{\omega_k\}$ and $\{\lambda_m\}$ which can be solved using the subgradient approach [31], are Lagrange coefficients associated respectively with first, second and third constraint of the optimization problem (41). The convergence of the SCA-based algorithm for the power control problem can be derived in a similar fashion as in [38].

E. Max-Min Fairness

Most existing works on cell-free massive MIMO have considered the problem of max-min fairness [7]–[9], [21]. To be able to provide a fair comparison between our proposed approach and existing literature in Section VI, we now consider the max-min optimization problem which is implemented during the group scheduling stage.

TABLE I
SIMULATION PARAMETERS

Noise Variance	-196 dBm	Bandwidth	20 MHz
Noise figure	5 dB	Training duration: τ	20
N	10 antennas	ϵ_1, ϵ_2, ϵ_3, ϵ_5	10^{-3}
T_c, $\overline{R}_k$	200, 1bit/s/Hz	ρ_p	1, W

$$\begin{aligned}
&\max_{k}\ \min_{\varrho_c^l\in\{0,1\},\forall l,\forall c}\ \frac{1}{L}\sum_{l=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)}\varrho_c^l\widetilde{R}_{k,c}\\
&\text{s.t. } \varrho_c^{\ell}+\varrho_{\tilde{c}}^{\ell}\le 1,\forall c\ne\tilde{c}\,\forall\ell \qquad (43)
\end{aligned}$$

Using similar arguments as the ones adopted to solve problem (33), i.e the result of Proposition 2 and the first order convex condition [31], we can apply the SCA algorithm principle to the following conservative approximation of problem (43)

$$\begin{aligned}
&\max_{\hat{\zeta}\ge 0, 0\le\varrho_c^l\le 1,\forall l,\forall c}\\
&\hat{\zeta}-\varsigma_2\sum_{\ell=1}^{L}\sum_{c=1}^{C}\left(\varrho_c^{\ell}-(\overline{\varrho_c}^{\ell})^2-2\overline{\varrho_c}^{\ell}\left(\varrho_c^{\ell}-\overline{\varrho_c}^{\ell}\right)\right)\\
&\text{s.t. } \hat{\zeta}\le\frac{1}{L}\sum_{l=1}^{L}\sum_{c=1}^{C}\sum_{k\in\Gamma(c)}\varrho_c^l\widetilde{R}_{k,c},\forall k\\
&\varrho_c^{\ell}+\varrho_{\tilde{c}}^{\ell}\le 1,\forall c\ne\tilde{c}\,\forall\ell \qquad (44)
\end{aligned}$$

VI. NUMERICAL RESULTS

In this section, we provide numerical results in order to evaluate the performance of the proposed algorithms. All results are obtained using Monte Carlo simulations by averaging, unless stated otherwise, over 500 realizations for the position of the APs and the users. All APs and users are uniformly distributed over a square region of width $1\,\mathrm{Km}$. The large-scale fading coefficient is modeled as [8]

$$\beta_{mk}[\text{dB}] = \begin{cases} -35.7-35\log_{10}\left(\frac{r_{mk}}{1m}\right)+z_{mk} & r_{mk}\ge 50m\\ -61.2-20\log_{10}\left(\frac{r_{mk}}{1m}\right) & 10m\le r_{mk}<50m\\ -81.2 & \text{otherwise}\end{cases} \qquad (45)$$

where r_{mk} is the distance from the mth AP to the kth user. z_{mk} is the log-normal shadowing with standard deviation of 8 dB. All other parameters are summarized in table I.

For cases without power allocation, the APs transmit with uniform power allocation. That is (i) $\eta_{mk}=\frac{1}{N\sum_k^K\nu_{mk}}$ for canonical CF, (ii) $\eta_{mk}^{\ell}=\frac{1}{LN\sum_k^K\nu_{mk}}$ the power on each sub-frequency band and (ii) $\eta_{mk}^{\ell,c}=\frac{1}{LNC\sum_k^K\nu_{mk,c}}$ for each sub-frequency band and for each group. For our proposed scheme, we implement the randomization Algorithm 2 and SCA-based Algorithm 3 which we term as Proposed SDR + SCA. Whenever power control is performed alongside the grouping and frequency allocation, we term it as Proposed SDR + SCA + Power. We first compare the cumulative distribution of the

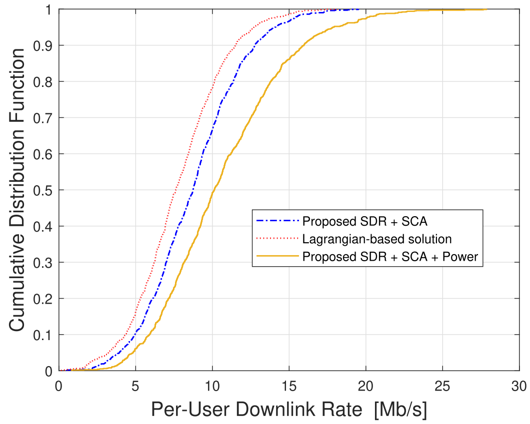
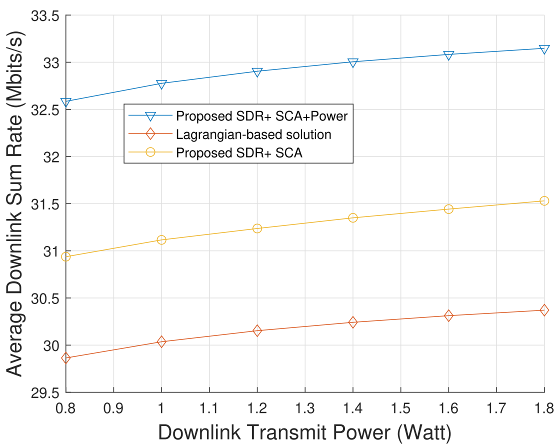
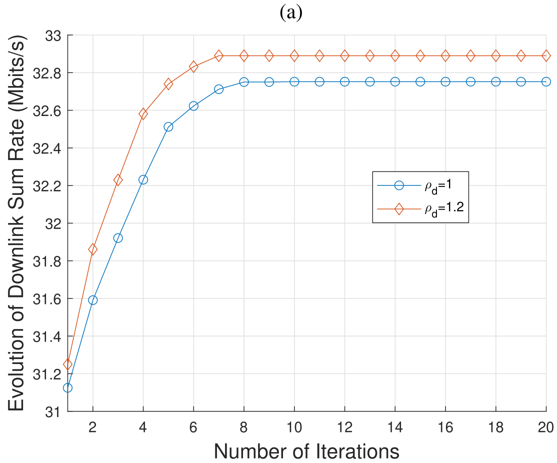


Fig. 2. Cumulative distribution of the per-user downlink achievable rate for $K = 20,\ M = 50,\ L = 10, \rho_d = 1,\ \alpha = 4$ and $C = 5$.

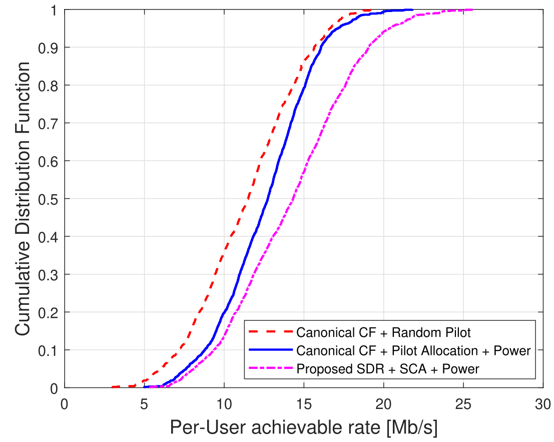


(a)

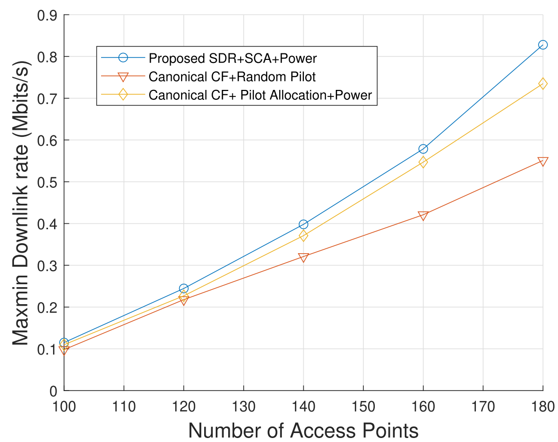


(b)

Fig. 3. (a) Average Sum of Downlink Achievable Rate versus Downlink Transmit Power for $K = 20,\ M = 50,\ L = 10, \rho_d = 1,\ \alpha = 4$ and $C = 5$. (b) Evolution of Sum of Downlink Achievable Rate $K = 20,\ M = 50,\ L = 10,\ \alpha = 4$ and $C = 5$.



(a)



(b)

Fig. 4. (a)CDF of per-user achievable downlink rate for $M = 100,\ K = 30,\ L = 10,\ \rho_d = 1,\ C = 5$ and $\alpha = 4$. (b) Comparison of Max-min downlink rate for $K = 30,\ L = 10,\ \rho_d = 1,\ C = 5$ and $\alpha = 4$.

per-user downlink achievable rate for the Proposed SDR + SCA + Power and the Lagrangian-based approach in Figure 2. It can be seen an improvement of 15% between the Proposed SDR+SCA and Lagrangian-based solution. This improvement increases to 42% if power control is performed. Moreover, there is a gain of 21% between the proposed approach with power allocation and without power allocation. This indicates that proper power control can further improve the performance of the system.

We compare the performance of the Lagrangian-based solution Algorithm 1 with the proposed SDR + SCA in terms of sum of downlink achievable rate versus downlink transmit power in Figure 3 (a). From Figure 3 (a), it can be observed that the proposed scheduling-based approach outperforms the Lagrangian-based solution. In fact, there is a performance gain varying from 9.1% to 11.23% between our the alternative proposed SDR+SCA and the Lagrangian-based scheme without power allocation and the gain increases from 24.08% to 27.31% when power control is taking into account.

Now, we numerically evaluate the convergence of the SCA approach of the power control problem. The performance is given in terms of evolution of the sum downlink rate versus the number of iterations in Figure 3 (b). It can be concluded by observing 3 (b) that the proposed SCA-based approach to solve the power control problem converges.

In Section III, we stated that the Lagrangian-based Algorithm 1 is generally very slow in terms of convergence speed. We verify this assertion by mean of simulations. It is worth pointing out that simulations were running on a Windows desktop with 4 Intel i7 cores and 8GB of RAM. Given the same error of tolerance, the simulations revealed that there is a computation time gain of 85% of the proposed SDR+SCA as compared to the Lagrangian-based approach. Algorithm 1 converges on average 7 times slower than the Proposed SDR+SCA.

We proceed to compare our approach with existing scheme [8], that is the canonical form of representing CF massive MIMO where all users are scheduled simultaneously over the available bandwidth. We implement the random pilot allocation scheme as well as [8, Algorithm 1] which we term as : Canonical CF+ Pilot allocation and implement the power allocation given in [8, Algorithm 2]. Firstly, we compare the CDF of the per-user downlink rate in Figure 4 (a). It can be seen from Figure 4 (a) that our proposed approach outperforms the existing scheme [8]. Secondly, We provide comparison in terms of maximum minimum downlink rate versus the number of APs. It can be observed from Figure 4(b) that our proposed scheme outperforms existing approaches. There is a performance gain varying from 9.1% to 27% between our proposed scheduling-based solution SDR + SCA + Power and the considered baseline [8]. This comparison leads us to believe that partitioning users into groups followed by non-overlapping group scheduling can palliate the effect of pilot contamination.

VII. Conclusion

In this work, we advocated that efficient radio resource allocation can lead to diminution of the effect of pilot contamination and can mitigate multi-user interference for a Cell-Free massive MIMO network. We investigated the problem of sub-frequency bands allocation. Firstly, we locally solved the NP-hard problem formulation by mean of Lagrangian-based method which has substantially low convergence speed. We also proposed a more appealing alternative method that consists of addressing the problem from the perspective of users partitioning. We proposed a semidefinite relaxation based algorithm to efficiently solve the problem. In addition to that, we designed a successive convex approximation algorithm to find a stationary point to the problem of sub-frequency bands allocation as well as the power control problem. The simulation results demonstrated that the alternative proposed scheme outperforms the Lagrangian-based solution. Moreover, the simulation results showed that the proposed scheduling-based method enables to attain a performance gain compared to existing schemes.

Appendix A
Proof of Proposition 1

To prove that problem (9) is NP-hard, we consider investigating the NP-hardness of a special case. More specifically, we assume that $L = 1$ and suppose that the number of active users is constrained by the number of orthogonal pilot sequences, i.e $K \leq \tau$. Before proceeding with the proof, define the following changes of variables:

$$\begin{aligned}
a_k &\triangleq \left(\frac{MN\tau\rho_p\beta_{mk}^2}{\rho_p\beta_{mk}+1}\right)^2, \; b_k \triangleq \frac{N\rho_d M\tau\rho_p\beta_{mk}^3}{\rho_p\beta_{mk}+1}+1,\\
\Delta_{kj} &\triangleq \rho_d NM\beta_{mk}\frac{\tau\rho_p\beta_{mj}^2}{\rho_p\beta_{mj}+1}, \; \tilde{b}_k \triangleq \frac{b_k}{a_k}, \; \widetilde{\Delta}_{kj} \triangleq \frac{\Delta_{kj}}{a_k}
\end{aligned}$$

Consequently, the optimization problem (9) is reduced to

$$\begin{aligned}
&\max_{\boldsymbol{\vartheta}\in\{0,1\}^K} \sum_{k=1}^{K} \vartheta_k \log_2\left(1+\frac{1}{\tilde{b}_k+\sum_{j\neq k}^{K}\widetilde{\Delta}_{kj}\vartheta_j}\right)\\
&\quad \text{s.t. } \overline{R}_k \leq \vartheta_k \log_2\left(1+\frac{1}{\tilde{b}_k+\sum_{j\neq k}^{K}\widetilde{\Delta}_{kj}\vartheta_j}\right), \forall k \quad \text{(A.1)}
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\max_{\boldsymbol{\vartheta}\in\{0,1\}^K} \sum_{k=1}^{K} \log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\neq k}^{K}\widetilde{\Delta}_{kj}\vartheta_j}\right)\\
&\quad \text{s.t. } \overline{R}_k \leq \log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\neq k}^{K}\widetilde{\Delta}_{kj}\vartheta_j}\right), \forall k \quad \text{(A.2)}
\end{aligned}$$

To prove the NP-hardness, our goal is to introduce a polynomial time reduction in order to transform the vertex packing problem into the optimization problem (A.2). To proceed with the proof, we write the following definition

Definition 1: Consider an undirected graph $G = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{1, \cdots, |\mathcal{V}|\}$ is the set of vertices in G while $\mathcal{E} = \{(j,k) : j \in \mathcal{V}\ k \in \mathcal{V}\}$ denotes the set of edges in G. An independent set of G is defined as a subset, say $\widetilde{G} \subset G$, where no two nodes in $\widetilde{G}$ are connected. That is, $\forall\, v_j,\, v_k \in \widetilde{G}$ implies that $(v_j, v_k) \notin \mathcal{E}$. The vertex packing problem refers as the problem of finding an independent set with a predefined size.

Let us construct an undirected with $|\mathcal{V}| = K$ vertices. For any couple $(v_j, v_k) \in \widetilde{G}$, let

$$\widetilde{\Delta}_{kj} = \begin{cases} 1, & \text{if } v_j \text{ and } v_k \text{ are connected.} \\ 0, & \text{otherwise.} \end{cases}$$

and let $\tilde{b}_k = \frac{1}{2^{|\mathcal{V}|}}$, $\overline{R}_k = R \leq \log_2\left(1+2^{|\mathcal{V}|}\right)$, $\forall k$. The transformation of the maximum independent set problem into problem (A.2) is given through the following Lemma.

Lemma 1: Solving the special instance optimization problem (A.2) *is equivalent to finding a maximum independent set $\widetilde{G}$ of size $|\widetilde{G}|$ of the undirected graph G. Moreover, the optimal solution $s^\star$ to problem* (A.2) *satisfies* $\log_2\left(1+2^{|\mathcal{V}|}\right)$ $|\widetilde{G}| \leq s^\star$.

Proof of Lemma 1 is given in Appendix B. Given that the vertex packing problem is an NP-hard problem, it follows that the optimization (A.2) is NP-hard. This concludes our proof. ■

Appendix B
Proof of Lemma 1

We proceed with the proof of Lemma 1 by firstly demonstrating the necessary condition. Suppose that there exists an independent set $\widetilde{G}$ of size $|\widetilde{G}|$. Define $\vartheta_k,\ k=1,\cdots,K$ as follow:

$$\vartheta_k=\begin{cases}1, & \text{if } k\in\widetilde{G}\ .\\ 0, & \text{otherwise.}\end{cases} \tag{A.3}$$

Since no two nodes in $\widetilde{G}$ are connected, that is $\widetilde{\Delta}_{kj}=0,\ \forall(k,j)\in\widetilde{G}$ with $k\neq j$. It follows that $\overline{R}_k\leq\log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\neq k}^{K}\widetilde{\Delta}_{kj}\vartheta_j}\right)=\log_2\left(1+2^{|\mathcal{V}|}\right)$. Hence, solution provided in (A.3) is feasible for the optimization problem (A.2) and the corresponding objective function is

$$\begin{aligned} s &= \sum_{k\in\widetilde{G}}\log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\vartheta_j}\right)\\ &= \sum_{k\in\widetilde{G}}\log_2\left(1+2^{|\mathcal{V}|}\right)=|\widetilde{G}|\log_2\left(1+2^{|\mathcal{V}|}\right) \end{aligned} \tag{A.4}$$

with $s\leq s^\star$. To prove the sufficient condition, we suppose that problem (A.2) possesses an optimal solution. Denote the optimal objective function value $s^\star$ and the optimal set as $\mathcal{D}^\star$. Define the following set

$$\widetilde{G}=\{v_k|\vartheta_k^\star=1,\ 1\leq k\leq K\}\subseteq\mathcal{V} \tag{A.5}$$

It can be inferred from (A.5) that set $\widetilde{G}$ encompasses all nodes or elements of $\mathcal{D}^\star$ that have value 1. Our goal is to demonstrate that the set $\widetilde{G}$ is an independent set. Before starting with the proof, we have the following claim

Claim 1: The graph characterized by combining all nodes in $\widetilde{G}$ is an incomplete graph.

Claim 1 is proven in Appendix C. Given from Claim 1 that the graph formed by the nodes in $\widetilde{G}$ is not be complete graph, we need to show that no two nodes in the optimal set are connected. This is done by contradiction. Suppose not, namely there exists a subset $\mathcal{S}\subset\widetilde{G}$ with at least two nodes that are connected. Under this consideration, the optimal objective function value $s^\star$ is given by

$$\begin{aligned} s^\star &= \sum_{k\in\mathcal{D}^\star}\log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\in\mathcal{D}^\star,j\neq k}\widetilde{\Delta}_{kj}\vartheta_j}\right)\\ &= \sum_{k\in\widetilde{G}}\log_2\left(1+\frac{1}{\tilde{b}_k+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &= \sum_{k\in\mathcal{S}}\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\mathcal{S},j\neq k}1\cdot 1}\right)\\ &\quad+\sum_{k\in\widetilde{G}/\mathcal{S}}\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &\overset{(a)}{\leq} |\mathcal{S}|\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+1}\right)+\sum_{k\in\widetilde{G}/\mathcal{S}}\log_2\left(1\right.\\ &\quad\left.+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &< |\mathcal{S}|+\sum_{k\in\widetilde{G}/\mathcal{S}}\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right) \end{aligned} \tag{A.6}$$

where (a) follows from the fact reducing the interference leads to higher rate. We now consider a feasible solution $\widehat{\mathcal{D}}$ for problem (A.2), where $\widehat{\mathcal{D}}$ is constructing as follow $\widehat{\mathcal{D}}=(\mathcal{D}^\star/\mathcal{S})\cup\widehat{\mathcal{S}}$, $\widehat{\mathcal{S}}$ includes the same index elements as $\mathcal{S}$, with only one element being 1 and all other elements 0. Without loss of generality, suppose that the only element with 1 is k_0. The corresponding objective function attained with the feasible set $\widehat{\mathcal{D}}$ is

$$\begin{aligned} s_{\widehat{\mathcal{D}}} &= \sum_{k\in\widehat{\mathcal{D}}}\log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum\limits_{j\in\widehat{\mathcal{D}},j\neq k}\widetilde{\Delta}_{kj}\vartheta_j}\right)=\sum_{k\in(\widetilde{G}/\mathcal{S})\cup\{k_0\}}\\ &\quad\log_2\left(1+\frac{1}{\tilde{b}_k+\sum\limits_{j\in(\widetilde{G}/\mathcal{S})\cup\{k_0\},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &= \log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}}\right)+\sum_{k\in\widetilde{G}/\mathcal{S}}\\ &\quad\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in(\widetilde{G}/\mathcal{S})\cup\{k_0\},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &\overset{(b)}{>} \log_2\left(2^{|\mathcal{V}|}\right)+\sum_{k\in\widetilde{G}/\mathcal{S}}\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &> |\mathcal{S}|+\sum_{k\in\widetilde{G}/\mathcal{S}}\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &> s^\star \end{aligned} \tag{A.7}$$

The inequality (b) comes from the monotonicity of the logarithm function, that is $\log_2\left(2^{|\mathcal{V}|}\right)<\log_2\left(1+2^{|\mathcal{V}|}\right)$. Equation (A.7) leads to $s_{\widehat{\mathcal{D}}}>s^\star$. This result contradicts with the fact that $s^\star$ is the global optimum for problem (A.2). Therefore, it does not exist subsets in $\widetilde{G}$ with a least two adjacent nodes which implies that $\widetilde{G}$ is an independent set. ■

Appendix C
Proof of Claim 1

Suppose that the graph formed by all nodes in $\widetilde{G}$ is complete. Then, the optimal objective value function $s^\star$ is given by

$$\begin{aligned} s^\star &= \sum_{k\in\mathcal{D}^\star}\log_2\left(1+\frac{\vartheta_k}{\tilde{b}_k+\sum_{j\in\mathcal{D}^\star,j\neq k}\widetilde{\Delta}_{kj}\vartheta_j}\right)\\ &= \sum_{k\in\widetilde{G}}\log_2\left(1+\frac{1}{\tilde{b}_k+\sum_{j\in\widetilde{G},j\neq k}\widetilde{\Delta}_{kj}\cdot 1}\right)\\ &= |\widetilde{G}|\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+(|\widetilde{G}|-1)2^{|\mathcal{V}|}}\right) \end{aligned} \tag{A.8}$$

Constructing a feasible set $\widetilde{\mathcal{D}}$ constituting only of one active node, say $k_1\in\widetilde{G}$, i.e $\widetilde{\mathcal{D}}=(\mathcal{D}^\star/\widetilde{G})\cup k_1$. The associated objective function $\tilde{s}$ is given by $\tilde{s}=\log_2\left(1+\frac{1}{\frac{1}{2^{|\mathcal{V}|}}}\right)=\log_2\left(1+2^{|\mathcal{V}|}\right)$. Let us verify the sign of $\tilde{s}-s^\star=\log_2\left(1+2^{|\mathcal{V}|}\right)-|\widetilde{G}|\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+(|\widetilde{G}|-1)2^{|\mathcal{V}|}}\right)$. Towards that end, define the functions $\tilde{f}(x) = x\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+(x-1)2^{|\mathcal{V}|}}\right)$ and $\tilde{f}_1(x) = \log_2\left(1+2^{|\mathcal{V}|}\right) - \tilde{f}(x)$ for $2 \leq x \leq |\mathcal{V}|$. The derivative of $\tilde{f}(x)$ with respect to x is given by

$$\begin{aligned}\frac{\partial \tilde{f}(x)}{\partial x} &= \log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+(x-1)2^{|\mathcal{V}|}}\right) \\ &\quad +\frac{1}{\log(2)}\frac{\frac{-x2^{2|\mathcal{V}|}}{\left(1+(x-1)2^{|\mathcal{V}|}\right)^2}}{1+\frac{2^{|\mathcal{V}|}}{1+(x-1)2^{|\mathcal{V}|}}} \\ &= \log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+(x-1)2^{|\mathcal{V}|}}\right) \\ &\quad -\frac{1}{\log(2)}\frac{x2^{2|\mathcal{V}|}}{\left(1+x2^{|\mathcal{V}|}\right)\left(1+(x-1)2^{|\mathcal{V}|}\right)}\end{aligned}$$

The second derivative of $\tilde{f}(x)$ with respect to x can be found as follow

$$\begin{aligned}\frac{\partial^2 \tilde{f}(x)}{\partial x^2} &= \frac{1}{\log(2)}\frac{\frac{-2^{2|\mathcal{V}|}}{\left(1+(x-1)2^{|\mathcal{V}|}\right)^2}}{1+\frac{2^{|\mathcal{V}|}}{1+(x-1)2^{|\mathcal{V}|}}} \\ &\quad +\frac{1}{\log(2)}\frac{2^{2|\mathcal{V}|}(2^{|\mathcal{V}|}-1+x^2 2^{2|\mathcal{V}|})}{\left(1+x2^{|\mathcal{V}|}\right)^2\left(1+(x-1)2^{|\mathcal{V}|}\right)^2} \\ &= -\frac{1}{\log(2)}\frac{2^{2|\mathcal{V}|}}{\left(1+x2^{|\mathcal{V}|}\right)\left(1+(x-1)2^{|\mathcal{V}|}\right)} \\ &\quad +\frac{1}{\log(2)}\frac{2^{2|\mathcal{V}|}(2^{|\mathcal{V}|}-1+x^2 2^{2|\mathcal{V}|})}{\left(1+x2^{|\mathcal{V}|}\right)^2\left(1+(x-1)2^{|\mathcal{V}|}\right)^2} \\ &= \frac{1}{\log(2)}\frac{2^{2|\mathcal{V}|}(2^{1+|\mathcal{V}|}-2-x2^{|\mathcal{V}|})}{\left(1+x2^{|\mathcal{V}|}\right)^2\left(1+(x-1)2^{|\mathcal{V}|}\right)^2}\end{aligned}$$

By evaluating the sign of the second derivative, it is straightforward demonstrating that it is always negative for $x \geq 2-\frac{2}{2^{|\mathcal{V}|}}$. Therefore, the maximum value of $\frac{\partial \tilde{f}(x)}{\partial x}$ is attained at $x=2$ and is equal to

$$\begin{aligned}\max_{2\leq x\leq|\mathcal{V}|}\frac{\partial \tilde{f}(x)}{\partial x} &= \frac{\partial \tilde{f}(x)}{\partial x}|_{x=2}\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+2^{|\mathcal{V}|}}\right) \\ &\quad -\frac{1}{\log(2)}\frac{2^{(1+2|\mathcal{V}|)}}{((1+2^{(1+|\mathcal{V}|)})(1+2^{|\mathcal{V}|})}\end{aligned} \tag{A.9}$$

Define $\tilde{f}_2(y) = \ln(1+\frac{2^y}{1+2^y}) - \frac{2^{(1+2y)}}{(1+2^{(1+y)})(1+2^y)}, \quad y \geq 2$. The derivative of $\tilde{f}_2(y)$ is computed as

$$\begin{aligned}\frac{\partial \tilde{f}_2(y)}{\partial y} &= \frac{2^y\log(2)}{\left(1+2^{1+y}\right)\left(1+2^y\right)} - \frac{2^{2y+1}\log(2)\left(3\cdot 2^y+2\right)}{\left(1+2^y\right)^2\left(1+2^{1+y}\right)^2} \\ &= \frac{\log(2)2^y\left(1+2^y-2^{y+1}-2^{2y+2}\right)}{\left(1+2^y\right)^2\left(1+2^{1+y}\right)^2} \leq 0\end{aligned}$$

Hence, $\tilde{f}_2(y)$ is a decreasing function of y and its maximum value is attained at $y=2$, i.e

$$\max_{2\leq y}\tilde{f}_2(y) = \log_2\left(1+\frac{4}{5}\right) - \frac{1}{\log(2)}\frac{32}{45} = -0.1779 < 0 \tag{A.10}$$

Therefore, by combining (A.9) and (A.10), it can be concluded that $\tilde{f}(x)$ is a decreasing function of x. Its maximum value is given by $2\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+2^{|\mathcal{V}|}}\right)$. Consequently, $\tilde{f}_1(x)$ is an increasing function of x and its minimum is attained at

$$\begin{aligned}&\log_2\left(1+2^{|\mathcal{V}|}\right) - 2\log_2\left(1+\frac{2^{|\mathcal{V}|}}{1+2^{|\mathcal{V}|}}\right) \\ &\quad = \log_2\left(\frac{1+3\cdot 2^{|\mathcal{V}|}+2^{2+|\mathcal{V}|}+2^{3|\mathcal{V}|}}{1+2^{2+|\mathcal{V}|}+2^{2+2|\mathcal{V}|}}\right) > 0\end{aligned} \tag{A.11}$$

which means that $\tilde{s} - s^\star > 0$. This leads to a contradiction since $s^\star$ is the global maximum. Therefore, the graph formed by the nodes in $\widetilde{G}$ cannot be complete.

References

[1] S. E. Hajri, J. Denis, and M. Assaad, "Enhancing favorable propagation in cell-free massive MIMO through spatial user grouping," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[2] Cisco White Paper. (Feb. 2019). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf

[3] Ericsson White Paper. (2014). *5G Radio Access*. [Online]. Available: https://pdfs.semanticscholar.org/9a06/bcadf0f4e3770260e0193746d5365b6c9114.pdf

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[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] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2015, pp. 695–699.

[8] 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.

[9] 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.

[10] 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, Nov. 2017.

[11] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 108–113, Mar. 2003.

[12] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 118–127, Jun. 2014.

[13] R. F. Ustok, P. A. Dmochowski, P. J. Smith, and M. Shafi, "Interference cancellation with jointly optimized transceivers in multiuser multicellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7219–7229, Aug. 2018.

[14] J. Ma, S. Zhang, H. Li, N. Zhao, and V. C. M. Leung, "Interference-alignment and soft-space-reuse based cooperative transmission for multi-cell massive MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1907–1922, Mar. 2018.

[15] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[16] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[17] L. Sanguinetti, A. L. Moustakas, and M. Debbah, "Interference management in 5G reverse TDD HetNets with wireless backhaul: A large system analysis," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1187–1200, Jun. 2015.

[18] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[19] E. Bjornson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[20] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.

[21] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11264–11268, Nov. 2018.

[22] G. T. version Release 12. (2014). *LTE; Scenarios and Requirements for Small Cell Enhancements for E-UTRA and E-UTRAN*. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/136900_136999/136932/12.01.00_60/tr_136932v120100p.pdf

[23] Z.-Q. Luo, W.-K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[25] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[26] E. Hossain, L. B. Le, and D. Niyato, *Radio Resource Management in Multi-Tier Cellular Wireless Networks*. Hoboken, NJ, USA: Wiley.

[27] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, Nov. 2019.

[28] M. R. Garey and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman, 1990.

[29] A. Khalili, S. Zarandi, and M. Rasti, "Joint resource allocation and offloading decision in mobile edge computing," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 684–687, Apr. 2019.

[30] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5481–5495, Oct. 2014.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[32] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[33] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Stanford Univ., CA, USA, Tech. Rep., 2004. [Online]. Available: https://web.stanford.edu/class/ee392o/subgrad_method.pdf

[34] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[35] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx

[36] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM J. Optim.*, vol. 6, no. 2, pp. 342–361, 1996.

[37] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[38] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[39] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.

Juwendo Denis (Member, IEEE) received the B.E. degree in electrical engineering and the M.Sc. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2011 and 2013, respectively, and the Ph.D. degree in communications engineering from the CEDRIC Research Laboratory, Conservatoire National des Arts et Métiers, Paris, France, in 2016. He was a Postdoctoral Researcher with CentraleSupelec, Gif-sur-Yvette, France and a Former Post-Doctoral Research Fellow with the Electrical and Computer Engineering Department, University of Illinois at Chicago. He is currently a Post-Doctoral Fellow with the School of Engineering and Applied Sciences (SEAS), Harvard University. His current research interests include cross-layer optimization, resource allocation, interference management, channel estimation and users scheduling for 5G, and beyond wireless communications systems

Mohamad Assaad (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications from Telecom ParisTech, France, in 2002 and 2006, respectively. Since 2006, he has been with the Telecommunications Department, CentraleSupélec, where he is currently a Professor. He is also a Researcher with the Laboratoire des Signaux et Systèmes (L2S, CNRS). He has co-authored one book and more than 120 publications in journals and conference proceedings. He has given in the past successful tutorials on 5G systems at various conferences including IEEE ISWCS'15 and IEEE WCNC'16 conferences. His research interests include 5G and beyond systems, fundamental networking aspects of wireless systems, MIMO systems, resource optimization, and machine learning in wireless networks. He also serves regularly as a TPC member or the TPC co-chair for top-tier international conferences. He is currently an Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS and the *Journal of Communications and Information Networks*.