IEEE Access
The journal for rapid open access publishing

Received July 13, 2015, accepted July 30, 2015, date of publication August 4, 2015, date of current version August 13, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2464352

Resource Allocation Optimization in Multi-User Multi-Cell Massive MIMO Networks Considering Pilot Contamination

TRI MINH NGUYEN, (Student Member, IEEE), VU NGUYEN HA, (Student Member, IEEE), AND LONG BAO LE, (Senior Member, IEEE)

Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC H5A 1K6, Canada

Corresponding author: V. N. Ha (hanguyen@emt.inrs.ca)

ABSTRACT In this paper, we study the joint pilot assignment and resource allocation for system energy efficiency (SEE) maximization in the multi-user and multi-cell massive multi-input multi-output network. We explicitly consider the pilot contamination effect during the channel estimation in the SEE maximization problem, which aims to optimize the power allocation, the number of activated antennas, and the pilot assignment. To tackle the SEE maximization problem, we transform it into a subtractive form, which can be solved more efficiently. In particular, we develop an iterative algorithm to solve the transformed problem where optimization of power allocation and number of antennas is performed, and then pilot assignment optimization is conducted sequentially in each iteration. To tackle the first sub-problem, we employ a successive convex approximation (SCA) technique to attain a solvable convex optimization problem. Moreover, we propose a novel iterative low-complexity algorithm based on the Hungarian method to solve the pilot assignment sub-problem. We also describe how the proposed solution approach can be useful to address the sum rate (SR) maximization problem. In addition to the algorithmic developments, we characterize the optimal structure of both SEE and SR maximization problems. The numerical studies are conducted to illustrate the convergence of the proposed algorithms, impacts of different parameters on the SR and SEE, and significant performance gains of the proposed solution compared the conventional design.

INDEX TERMS Energy efficiency, massive MIMO, multi-cell, pilot assignment, pilot contamination, power allocation.

I. INTRODUCTION

The evolution of wireless cellular networks has strongly relied on the advanced MIMO technology to fundamentally enhance the wireless capacity and communications reliability [1]. Recently, there has been an increasing interest in the large-scale MIMO or massive MIMO technology which is based on the deployment of a large number of antennas at the transmitter and/or receiver sides [2]. Massive MIMO has been shown to result in significant improvement in network throughput and energy efficiency with potentially low complexity deployment. In addition, time-division duplexing (TDD) has been mainly recommended for massive-MIMO systems since it can provide the required channel state information (CSI) with affordable cost for downlink beamforming and uplink detection via uplink pilot transmission and channel estimation.

It has been shown in [2]–[5] that pilot contamination is one of the major challenges in designing massive MIMO networks, which presents the fundamental performance bottleneck. Such problem comes from the fact that a finite number of pilot sequences must be reused over cells for the uplink channel estimation, which causes the estimated CSI in one cell to be contaminated by other neighboring cells. Pilot contamination indeed results in the network throughput saturation even in the large-scale antennas regime [2], [3]. This performance bottleneck occurs with both simple and sophisticated beamforming schemes [2]–[4], which reflects the severe impact of pilot contamination on the system performance. It was shown in [5] that under certain special structure of MIMO channels, employment of the minimum-mean-squared-error (MMSE) channel estimator and a simple pilot assignment scheme can efficiently mitigate the pilot contamination effect in the large antenna regime. Nonetheless, it is important to investigate the design of a general MC-MU MIMO network considering MC interference, pilot contamination, limited bandwidth, and power, which is the target of this paper.

In general, efficient allocation of radio resources (e.g., power, frequency, time, and antennas) plays a crucial role in improving wireless network performance. This has been indeed an active research area over the last decades [1], [6]–[8]. In particular, different radio resource allocation problems for the MU single-cell setting have been investigated for both single-input single-output (SISO) and MIMO systems considering various design objectives such as power minimization, SR, and utility maximization. Design of an optimal resource allocation algorithm can be challenging in general if the underlying resource allocation problems are non-convex. Different SCA techniques have been developed to resolve these challenges [9]–[11]. Moreover, various efficient resource allocation algorithms have been proposed to address joint allocation of subchannel and powers in multi-carrier wireless systems [12]–[16], which can typically achieve suboptimal solutions due to the mixed-integer and non-linear nature of underlying problems.

Resource allocation for EE maximization has been also an active area of research in recent years [17]–[19]. This is motivated by the need to reduce the network energy consumption, which helps lower operation costs for mobile network operators and contributes to decrease CO_2 emission of the communications and information technology sector. Various resource allocation algorithms for EE maximization have been proposed under different network settings. Power allocation for pilot and data symbols for the orthogonal frequency division multiple access (OFDMA) cellular network considering channel estimation errors and power constraints was considered in [20]. Optimal beamforming designs for maximizing the EE under power constraints in MU-MIMO systems were studied in [21] for the uplink and in [22] for the downlink. In addition, the EE maximization problem in the MU massive MIMO cellular network was considered in [23] where the authors proposed to impose certain interference limits to manage the pilot contamination and inter-user interference. In general, it is not easy to determine these interference limits that can guarantee good network performance.

To the best of our knowledge, there is no existing work that considers the joint design of pilot assignment and resource allocation in the MU-MC massive-MIMO environment accounting for the pilot contamination effect. This paper aims to fill this gap in the literature where we consider both SR and SEE optimization problems for this setting. Specifically, we make the following contributions.

- We derive the user signal-to-interference-plus-noise ratio (SINR) for a general reuse pattern of pilot sequences over cells (i.e., consideration of pilot contamination) under MMSE-based channel estimation and maximum ratio transmission (MRT) beamforming schemes. This derivation allows us to obtain a closed-form rate formula as a function of transmit power, number of activated antennas and pilot assignment pattern based on which we can formulate the SR and SEE maximization problems.
- We describe how to employ the fractional programming technique to transform the SEE maximization problem into the subtractive optimization form based on which efficient iterative algorithms can be developed. In addition, we characterize the optimal number of activated antennas for these two problems. Specifically, we prove that the SR maximization problem requires to activate all antennas at optimality while the optimal number of activated antennas for the SEE maximization problem is shown to be smaller than an upper bound, which does not scale as the number of available antennas tends to infinity.
- We develop an iterative algorithm to solve the transformed SEE maximization problems where we solve two sub-problems sequentially in each iteration. Specifically, the first sub-problem optimizes the number of activated antennas and power allocation while the second sub-problem optimizes the pilot assignment. We employ the SCA technique to develop an algorithm to solve the first sub-problem. Moreover, we propose a novel method to iteratively perform pilot assignments to users different cells. Specifically, we reformulate the pilot assignment problem for one cell in each iteration into the well-known job assignment problem and develop a modified Hungarian-based algorithm to achieve an efficient and feasible pilot assignment solution. We describe how the proposed approach can be employed to solve the SR maximization problem.
- Numerical studies confirm the convergence of our proposed iterative algorithms and their superior performance in terms of total SR and SEE compared to the optimal resource allocation with conventional pilot assignment. Specifically, the SEE gain becomes larger as the maximum transmit power $P_{\max}$, the number of users per cell or the required minimum rate increases. In particular, the SEE gain of about 100% can be achieved for the required minimum rate of 3 bits/s/Hz. Moreover, the SR gain becomes more significant as the number of available antennas is larger.

The remaining of this paper is organized as follows. In Section II, we present the system model. In Section III, we formulate the constrained SR and SEE optimization problems. Optimality characterization and algorithmic solutions for the SEE and SR maximization problems with detailed algorithm designs are presented in Section IV and Section V, respectively. Numerical results are described in Section VI followed by conclusion in Section VII. Some preliminary results of this paper have been published in [24]. Throughout the paper, we shall use the following notations. Boldface upper-case letters denote matrices or sets, boldface lower-case letters denote column vectors or sets, and plain lower-case letters denote scalars. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ denote the transpose, complex conjugate, and Hermitian transpose, respectively. We denote $\mathbf{I}_n$ as the $n \times n$ identity matrix, $\mathrm{Tr}\,(\cdot)$ as the trace operator, and $|\cdot|$ as the standard Euclidean norm.

II. SYSTEM MODEL

A. RECEIVED SIGNAL MODEL

We consider the downlink of the cellular network with L cells, i.e., $\{1, ..., L\}$. Each base station (BS) which is supported by a large number of antennas $M_{\max} \gg 0$ is assumed to serve its K single-antenna users. We will explicitly consider channel estimation effects on the precoding and system performance in a two-stage framework. In the first stage, each BS estimates the channel coefficients between itself and the users. The estimated CSI is then utilized to determine precoding vectors for transmitting users' data in the second stage. It is noted that employing all $M_{\max}$ antennas at each BS to serve the users is not always desirable because of the high energy consumption of radio chains deployed for activated antennas. Thus, in this paper, we are interested in optimizing the number of activated antennas out of $M_{\max}$ at each BS to optimize different design objectives.[1]

Let us denote $M_j^\star$ as the optimal number of activated antenna for considered problems, which will be stated shortly. In order to determine $M_j^\star$, we assume that each BS j activates M_j antennas to serve all associated users. We denote data symbols associated with the K users in cell j as $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jK}]^T$, where data symbols of each user have unit average power, e.g., $\mathbb{E}\left\{x_{jk}^* x_{jk}\right\} = 1$. We assume that data stream for the kth user in cell j is precoded with an $M_j \times 1$ precoding vector $\mathbf{w}_{jk}$ before transmission. Then, the received signal at the kth UE in cell j can be given as

$$y_{jk} = \sum_{l=1}^{L} \sqrt{\beta_{ljk}} \mathbf{h}_{ljk}^\dagger \sum_{i=1}^{K} \sqrt{P_{li}} \mathbf{w}_{li} x_{li} + n_{jk}, \tag{1}$$

where P_{jk} is the transmit power for the kth user in cell j, n_{jk} is the additive white Gaussian noise (AWGN) at the kth user in cell j, $\mathbf{h}_{ljk}$ and β_{ljk} are the $M_l \times 1$ small-scale fading vector and the large-scale channel coefficient for the channel between BS l and the kth user in cell j, respectively. Here, each fading vector is assumed to follow a complex Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_{ljk}$, i.e., $\mathbf{h}_{ljk} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{ljk}\right)$. We assume that the large-scale channel coefficient captures the path-loss effect with $\beta_{ljk} = d_{ljk}^{-\alpha}$ where d_{ljk} is the distance between BS l and the kth user in cell j, and α is the path-loss exponent. In addition, the noise power is assumed to be N_0.

B. MMSE CHANNEL ESTIMATION

The precoding vector is typically designed by using the estimated CSI available at each BS. We assume that each BS estimates the CSI for its users based on the pilot transmission in the uplink where we assume the TDD system (i.e., uplink and downlink CSI is the same). Assume that there is a same set of κ ($K \le \kappa$) orthogonal pilot sequences $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_\kappa\}$ that must be assigned to all users in the network where $\boldsymbol{\psi}_i = [\psi_{i,1}, \ldots, \psi_{i,\tau}]^T$, ($\kappa \le \tau$). Under the orthogonal condition, we have $\boldsymbol{\psi}_m^\dagger \boldsymbol{\psi}_i = 0, \forall m \ne i$ and $\boldsymbol{\psi}_i^\dagger \boldsymbol{\psi}_i = \left|\psi_{i,1}\right|^2 + \left|\psi_{i,2}\right|^2 + \cdots + \left|\psi_{i,\tau}\right|^2 = \pi$ where π is the pilot signal power.

To optimize the pilot assignments, we introduce following variables $\boldsymbol{\rho}_j = [\rho_{j,1}, \ldots, \rho_{j,K}]^T$ for the K users in cell j and κ pilot sequences. Specifically, $\rho_{j,k} = m$ means that the kth user in cell j is assigned the mth pilot sequence in the channel estimation stage. To facilitate the algorithm development later, we also define a related notation $\tilde{\boldsymbol{\rho}}_j = [\tilde{\rho}_{j,1}, \ldots, \tilde{\rho}_{j,\kappa}]^T$ to present the reverse association between pilot sequences and users in cell j where $\tilde{\rho}_{j,m} = k$ means that the mth pilot sequence is assigned to the kth user in cell j, and $\tilde{\rho}_{l,m} = 0$ means that the mth pilot sequence is not used by any user in cell j. Then, during the CSI estimation stage, BS j receives the $M_j \times \tau$ signal matrix $\mathbf{Y}_j$ whose tth column ($1 \le t \le \tau$) can be expressed as

$$\mathbf{Y}_{j,t} = \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\beta_{jlk}} \mathbf{h}_{jlk} \boldsymbol{\psi}_{\rho_{l,k},t}^T + \mathbf{Z}_{j,t}, \tag{2}$$

where $\mathbf{Z}_{j,t} \in \mathbb{C}^{M_j \times 1}$ is the AWGN vector whose element is distributed according to $\mathcal{CN}\left(0, \sigma^2\right)$. Similar to [5], the MMSE estimator is employed for estimating channel fading vectors $\{\mathbf{h}_{jjk}\}$. Specifically, the estimated channel can be expressed as $\widehat{\mathbf{h}}_{jjk} = \sqrt{\beta_{jjk}} \mathbf{R}_{jjk} \mathbf{Q}_{jk} \mathbf{Y}_j \boldsymbol{\psi}_{\rho l,k}$, where

$$\mathbf{Q}_{jk} = \left(\sigma^2 \mathbf{I}_{M_j} + \pi \sum_{l=1}^{L} \beta_{jl\tilde{\rho}_{l,\rho_{j,k}}} \mathbf{R}_{jl\tilde{\rho}_{l,\rho_{j,k}}}\right)^{-1}. \tag{3}$$

Note that $\tilde{\rho}_{l,\rho_{j,k}}$ indicates that the user in cell l is assigned the same pilot sequences of the kth user in cell j. Moreover, we have $\beta_{jl\tilde{\rho}_{l,\rho_{j,k}}} = 0$ if $\tilde{\rho}_{l,\rho_{j,k}} = 0$. The actual channel can be written as the sum of the estimated channel and error term as $\mathbf{h}_{jjk} = \widehat{\mathbf{h}}_{jjk} + \widetilde{\mathbf{h}}_{jjk}$, where $\widehat{\mathbf{h}}_{jjk}$ is the Gaussian vector distributed according to $\mathcal{CN}\left(0, \boldsymbol{\Phi}_{jjk}\right)$ and $\widetilde{\mathbf{h}}_{jjk}$ is the Gaussian vector distributed according to $\mathcal{CN}\left(0, \mathbf{R}_{jjk} - \boldsymbol{\Phi}_{jjk}\right)$. In general, we have

$$\boldsymbol{\Phi}_{jjk} = \pi \beta_{jjk} \mathbf{R}_{jjk} \mathbf{Q}_{jk} \mathbf{R}_{jjk} = \beta_{jjk} \mathbf{R}_{jjk} \bar{\mathbf{Q}}_{jk} \mathbf{R}_{jjk}, \tag{4}$$

where $\bar{\mathbf{Q}}_{jk} = \left(1/\gamma_{\mathrm{p}} \mathbf{I}_{M_j} + \sum_{l=1}^{L} \beta_{jl\tilde{\rho}_{l,\rho_{j,k}}} \mathbf{R}_{jl\tilde{\rho}_{l,\rho_{j,k}}}\right)^{-1}$ and we denote $\gamma_{\mathrm{p}} = \pi/\sigma^2$ as the SNR of each pilot sequence.

C. ERGODIC ACHIEVABLE RATE

With the estimated CSI, we assume that the MRT beamforming is employed, i.e., $\mathbf{w}_{jk} = \sqrt{\xi_{jk}}\, \widehat{\mathbf{h}}_{jjk}$, where the parameter ξ_{jk} normalizes the average transmit power such that the term $P_{jk} \mathbb{E}\left\{\mathbf{w}_{jk}^\dagger \mathbf{w}_{jk}\right\}$ is equal to P_{jk}, i.e.,

$$\xi_{jk} = \frac{1}{\mathbb{E}\left\{\widehat{\mathbf{h}}_{jjk}^\dagger \widehat{\mathbf{h}}_{jjk}\right\}}. \tag{5}$$

[1] The general problem of optimizing the subset of activated antenna at each BS is outside the scope of this paper, which will be considered in our future works.

$$\gamma_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\frac{P_{jk}\beta_{jjk}A_{jjk}M_j}{P_{jk}\beta_{jjk}\left(A_{jjk}+1\right)+\sum\limits_{(l,i)\neq(j,k)}P_{li}\beta_{ljk}+\sum\limits_{l\neq j}P_{l\tilde{\rho}_{l,\rho_{j,k}}}\beta_{ljk}\left(B_{ljk}M_l+B_{ljk}-1\right)+N_0} \tag{10}$$

Then, the received signal at kth user in cell j can be rewritten as

$$\begin{aligned}y_{jk}&=\sum_{l=1}^{L}\sqrt{\beta_{ljk}}\mathbf{h}_{ljk}^{\dagger}\sum_{i=1}^{K}\sqrt{P_{li}}(\sqrt{\xi_{li}}\widehat{\mathbf{h}}_{lli})x_{li}+n_{jk}\\&=\sqrt{\xi_{jk}P_{jk}\beta_{jjk}}\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}x_{jk}\\&\quad+\sum_{(l,i)\neq(j,k)}\sqrt{\xi_{li}P_{li}\beta_{ljk}}\mathbf{h}_{ljk}^{\dagger}\widehat{\mathbf{h}}_{lli}x_{li}+n_{jk}.\end{aligned} \tag{6}$$

The ergodic rate achieved at the kth user in cell j with the given received signal from (6) can be written as $r_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\log\left[1+\gamma_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\right]$, where $\boldsymbol{\rho}=\text{vec}\left(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\ldots,\boldsymbol{\rho}_L\right)$ and $\gamma_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})$ can be interpreted as the achieved SINR associated with the achievable rate at the kth user in cell j, which is given by

$$\gamma_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\frac{\xi_{jk}P_{jk}\beta_{jjk}\left|\mathbb{E}\left\{\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right\}\right|^2}{\text{IN}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})}, \tag{7}$$

where the interference-plus-noise term can be presented as

$$\begin{aligned}\text{IN}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})&=\xi_{jk}P_{jk}\beta_{jjk}\text{Var}\left(\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right)\\&\quad+\sum_{(l,i)\neq(j,k)}\xi_{li}P_{li}\beta_{ljk}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\widehat{\mathbf{h}}_{lli}\right|^2\right\}+N_0,\end{aligned} \tag{8}$$

where we denote $\mathbf{P}_j=\left[P_{j1},P_{j2},\ldots,P_{jK}\right]^T$, $\mathbf{P}=\text{vec}\left(\mathbf{P}_1,\mathbf{P}_2,\ldots,\mathbf{P}_L\right)$ as the transmit powers from each BS to its users; and $\mathbf{M}=[M_1,M_2\ldots,M_L]^T$ describes the number of activated antennas at the BSs. In this paper, we mainly concentrate on the case where $\mathbf{R}_{ljk}=\mathbf{I}_{M_l}$ while the general case will be discussed at the end of Section IV. We now derive the received SINR and ergodic achievable rate in the following theorem.

Theorem 1: When the covariance matrix satisfies $\mathbf{R}_{ljk}=\mathbf{I}_{M_l}$, *the ergodic achievable rate of the* k*th user in cell* j *can be written as*

$$r_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\log_2\left[1+\gamma_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\right], \tag{9}$$

where $\gamma_{jm}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})$ *is given by (10) on top of this page and* $A_{jjk}=\frac{\beta_{ijk}}{1/\gamma_{\text{P}}+\sum_{l=1}^{L}\beta_{jl\tilde{\rho}_{l,\rho_{j,k}}}}$ *and* $B_{ljk}=\frac{\beta_{ljk}}{1/\gamma_{\text{P}}+\sum_{n=1}^{L}\beta_{ln\tilde{\rho}_{n,\rho_{j,k}}}}$.

Proof: The proof is provided in Appendix A. □

III. RESOURCE ALLOCATION PROBLEMS

A. SYSTEM ENERGY EFFICIENCY (SEE)

The total SR of the network can be expressed as

$$\mathcal{R}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\sum_{j=1}^{L}\sum_{k=1}^{K}r_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho}), \tag{11}$$

where $r_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})$ is the ergodic achievable rate of the kth user in cell j given by (9). To express the SEE metric, we assume the linear BS power consumption model [25] where the total power network consumption $\mathcal{P}(\mathbf{P},\mathbf{M})$ is

$$\mathcal{P}(\mathbf{P},\mathbf{M})=P_c\sum_{j=1}^{L}M_j+\sum_{j=1}^{L}\sum_{k=1}^{K}\nu P_{jk}+P_0, \tag{12}$$

where P_c is the constant circuit power consumption per antenna including the power dissipation in the transmitting filter, mixer, frequency synthesizer, and digital-to-analog converter which are independent from the transmit power. Thus, the first term in (12) represents the total circuit power consumption of all activated antennas at BSs. Moreover, $\nu\geq 1$ is a constant which accounts for the inefficiency of the power amplifier; P_0 is the basic power consumed at the BS independent of the number of active antennas. From this, the SEE, calculated in bits/Joule,[2] is defined as the ratio of SR and the total power consumption as follows:

$$\eta(\mathbf{P},\mathbf{M},\boldsymbol{\rho})=\frac{\mathcal{R}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})}{\mathcal{P}(\mathbf{P},\mathbf{M})}. \tag{13}$$

B. SR AND SEE MAXIMIZATION PROBLEMS

We consider the following two optimization problems. The first one aims at maximizing the SR $\mathcal{R}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})$ as follows:

$$\begin{aligned}&\underset{\mathbf{P},\mathbf{M},\boldsymbol{\rho}}{\text{maximize}}\ \mathcal{R}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\\&\text{subject to (C1)}: r_{jk}(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\geq R_{\min}\quad\forall j,k\\&\qquad\text{(C2)}:\sum_{k=1}^{K}P_{jk}\leq P_{\max};\quad\forall j,k\\&\qquad\text{(C3)}: P_{jk}>0,\quad\forall j,k\\&\qquad\text{(C4)}: 0<M_j\leq M_{\max}\quad\forall j,\ M_j\in\mathbb{Z}^+\\&\qquad\text{(C5)}:\boldsymbol{\rho}_j\in\{1,\ldots,K\},\quad\rho_j\neq\rho_i,\quad\forall i,j.\end{aligned} \tag{14}$$

The second problem maximizes the SEE $\eta(\mathbf{P},\mathbf{M},\boldsymbol{\rho})$ as follows:

$$\begin{aligned}&\underset{\mathbf{P},\mathbf{M},\boldsymbol{\rho}}{\text{maximize}}\ \eta(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\\&\text{subject to (C1), (C2), (C3), (C4), (C5).}\end{aligned} \tag{15}$$

The optimization variables for both problems are $\mathbf{P},\mathbf{M},\boldsymbol{\rho}$. Constraint (C1) imposes the minimum data rate requirements for all users; (C2) and (C3) describe the constraints for transmit powers of each BS where $P_{\max}$ is the BS maximum power. Moreover, (C4) captures the constraint for the

[2]Strictly speaking, it should be bits/Joule/Hz since the rate is calculated for one Hz of system bandwidth.

number of activated antennas and (C5) represents the integer pilot assignment constraint in each BS. Because of the non-concave objective functions and the mixed integer nature of optimization variables, these two problems are mixed integer non-convex programs, which are very hard to solve optimally. In this paper, we propose low-complexity algorithms which can achieve sub-optimal solutions for the considered problems.

Remark 1: It is worth noting that the SR and SEE maximization problems are not always feasible because of the constraints (C1) and (C2). Different practical methods can be employed to deal with this issue when these problems are not feasible. For example, one can adaptively adjust the required minimum rates for users or perform user removals (some users are removed) if the system is infeasible. In this paper, we assume that these problems are feasible, which allows us to focus on the joint pilot assignment and resource allocation design. We would like to address the feasibility problem in our future works.

IV. RESOURCE ALLOCATION ALGORITHMS FOR SEE MAXIMIZATION PROBLEM

In this section, we describe the transformation of the SEE maximization problem into a more solvable form. Then, we develop a novel approach to solve the transformed SEE maximization problems. This proposed solution approach can be used to solve the SR problem as discussed in the next section.

A. TRANSFORMATION OF SEE MAXIMIZATION PROBLEM

The objective of the SEE optimization problem (15) is in a fractional form, which is difficult to address. Therefore, we transform it into a subtractive form to aid the algorithm development. Toward this end, let us denote Ω as the set of feasible solutions $(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho})$ of (15), we can express the maximum SEE $\eta^\star$ as

$$\eta^\star(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star) = \max_{(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\in\Omega} \frac{\mathcal{R}(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho})}{\mathcal{P}(\mathbf{P}, \mathbf{M})}, \tag{16}$$

where $(\mathbf{P}^*, \mathbf{M}^*, \boldsymbol{\rho}^*)$ represents the optimal solution. The following theorem provides the foundation based on which the transformation of problem (15) can be performed.

Theorem 2: The maximum SEE $\eta^\star$ defined in (16) is achieved if and only if

$$\begin{aligned} F(\eta^\star) &= \max_{(\mathbf{P},\mathbf{M},\boldsymbol{\rho})\in\Omega} \mathcal{R}(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho}) - \eta^\star\mathcal{P}(\mathbf{P}, \mathbf{M}) \\ &= \mathcal{R}(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star) - \eta^\star\mathcal{P}(\mathbf{P}^\star, \mathbf{M}^\star) = 0, \end{aligned} \tag{17}$$

for $\mathcal{R}(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho}) \geq 0$ and $\mathcal{P}(\mathbf{P}^\star, \mathbf{M}^\star) > 0$.

Proof: Theorem 2 can be proved by following the similar approach in [26]. □

The results of Theorem 2 imply that the optimal value of $\eta^\star \geq 0$ must satisfy $F(\eta^\star) = 0$. Thus, to deal with the constrained optimization problem (15), we can consider the following optimization problem

$$\begin{aligned} &\underset{\mathbf{P},\mathbf{M},\boldsymbol{\rho}}{\text{maximize}}\ \mathcal{R}(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho}) - \eta\mathcal{P}(\mathbf{P}, \mathbf{M}) \\ &\text{subject to (C1), (C2), (C3), (C4), (C5).} \end{aligned} \tag{18}$$

Then we can address problem (15) by iteratively solving (18) for a current value of η and updating η until we reach the optimal $\eta^\star \geq 0$ satisfying $F(\eta^\star) = 0$ [26]. This problem is a mixed integer non-convex optimization problem, which is still very difficult to solve.

B. OPTIMALITY CHARACTERIZATION

We now describe the optimal structure of the SEE maximization problem in the following theorem with the support from the following proposition.

Proposition 1: For a given fixed pilot assignment $\boldsymbol{\rho}$, consider the following optimization problem

$$\begin{aligned} &\underset{\mathbf{P},\mathbf{M}}{\text{minimize}}\ \max_{j=1,\ldots,L} M_j \\ &\text{subject to (C1), (C2), (C3), (C4).} \end{aligned} \tag{19}$$

Suppose there exists a feasible solution of (19) when $M_{\max} \to \infty$, then there exists a set of thresholds $\mathbf{M}^{th} = \{M_1^{th}, \ldots, M_j^{th}, \ldots, M_L^{th}\}$ such that

(i) *when $\mathbf{M}_{\max} \preceq \mathbf{M}^{th}$, problem (19) is infeasible[3] and when $\mathbf{M}_{\max} \succ \mathbf{M}^{th}$, problem (19) is feasible, where $\mathbf{M}_{\max} = \{M_{\max}, \ldots, M_{\max}\}$.*

(ii) *when $\mathbf{M}_{\max} \succ \mathbf{M}^{th}$, the optimal solution $\mathbf{M}^\star$ is independent of $M_{\max}$.*

Proof: The proof is provided in Appendix B. □

Theorem 3: If there exists a feasible solution for problem (19) then the optimal number of activated antennas $M_j^\star$ at any BS j for problem (18) is upper bounded by a fixed value M_0 independent of $M_{\max}$ when $M_{\max}$ is sufficiently large.

Proof: The proof is provided in Appendix C. □

The results in this theorem provide interesting insights. Specifically, from the SEE maximization perspective it is more beneficial to utilize a moderate number of antennas at each BS even if the number of available antennas is very large.

In the remaining of this section, we describe the proposed algorithm to solve the transformed SEE maximization problem (18). Since we have to optimize over combinatorial sets for $\boldsymbol{\rho}$ as well as real $\mathbf{P}$ and integer variables $\mathbf{M}$, solving this joint problem is very challenging. To overcome this difficulty, we propose to decompose it into two sub-problems and solve them sequentially in each iteration, i.e., optimization over $\mathbf{P}$, $\mathbf{M}$ for given $\boldsymbol{\rho}$ and optimization over $\boldsymbol{\rho}$ for given $\mathbf{P}$, $\mathbf{M}$.

C. SCA-BASED ALGORITHM

For a fixed $\boldsymbol{\rho}$, we solve the sub-problem corresponding to (18) over variables $\mathbf{P}$ and $\mathbf{M}$. We can observe that the objective function in (18) is a non-concave function and the optimization variables M_j, $j = 1, \ldots, L$ take integer values. To address the second challenge, we relax M_j to real variables and obtain a sub-optimal solution of $\breve{M}_j^\star$ by rounding it as $\breve{M}_j^\star = \lceil M_j^\star \rceil$. For the first challenge, we employ the SCA technique proposed in [9] to approximate the objective function based on the following result

$$\log\left(1 + \omega_{jk}\right) \geq f(\omega_{jk}, a_{jk}, b_{jk}) = a_{jk}\omega_{jk} + b_{jk}, \tag{20}$$

[3]Here, $\preceq$ denotes the component-wise inequality.

Algorithm 1 SCA-Based Algorithm to Optimize $\mathbf{P}, \mathbf{M}$

Input: Pilot assignment $\boldsymbol{\rho}$, path-loss coefficients β_{ljk}
Output: Power and number of antenna update $(\mathbf{P}^\star, \mathbf{M}^\star)$.

1: Initialize with a feasible point $\mathbf{P}^{(0)}, \mathbf{M}^{(0)}$ and set $\mathbf{a}^{(0)} = 1,\ \mathbf{b}^{(0)} = 0,\ t_1 = 1$.
2: **repeat**
3: Solve (22) to obtain the solution $\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}$.
4: Compute $\gamma_{jk}\left(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}\right),\ \forall j, k$ and update $\mathbf{a}^{(t_1)},\ \mathbf{b}^{(t_1)}$ by using (21).
5: $t_1 = t_1 + 1$
6: **until** Convergence of $\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}$.
7: **return** $(\mathbf{P}^\star, \mathbf{M}^\star) = \left(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}\right)$.

for some a_{jk}, b_{jk} that are adaptively calculated as a function of of each point ω_{jk} to achieve the tightest lower bound. In particular, for a particular value $\omega_{jk} = \tilde{\omega}_{jk}$, the parameters of $f(\cdot)$ can be chosen as

$$a_{jk} = \frac{\tilde{\omega}_{jk}}{1+\tilde{\omega}_{jk}}; \quad b_{jk} = \log\left(1+\tilde{\omega}_{jk}\right) - \frac{\tilde{\omega}_{jk}\log\omega_{jk}}{1+\tilde{\omega}_{jk}}. \tag{21}$$

Motivated by the the above convexity approximation, we employ this lower bound to approximate $r_{jk} = \log\left(1+\gamma_{jk}\right)$ where ω_{jk} corresponds to γ_{jk}. Then, we apply the following changes of variables $\hat{\mathbf{P}} = \log\mathbf{P}$ and $\hat{\mathbf{M}} = \log\mathbf{M}$. Finally, we arrive at the following approximated optimization problem

$$\begin{aligned}
&\underset{\hat{\mathbf{P}},\hat{\mathbf{M}},\mathbf{a},\mathbf{b}}{\text{maximize}}\ \hat{\mathcal{R}}(\hat{\mathbf{P}},\hat{\mathbf{M}},\mathbf{a},\mathbf{b}) - \eta\mathcal{P}(\hat{\mathbf{P}},\hat{\mathbf{M}}) \\
&\text{subject to (C1}'\text{)}: \ \hat{r}_{jk}(\hat{\mathbf{P}},\hat{\mathbf{M}},a_{jk},b_{jk}) \geq R_{\min}, \quad \forall j,k, \\
&\qquad\qquad \text{(C2}'\text{)}: \sum_{k=1}^{K} e^{\hat{P}_{jk}} \leq P_{\max}\forall j, \\
&\qquad\qquad \text{(C3}'\text{)}: \hat{M}_j \leq \log M_{\max}, \quad \forall j,
\end{aligned} \tag{22}$$

where $\hat{r}_{jk}(\hat{\mathbf{P}},\hat{\mathbf{M}},a_{jk},b_{jk}) = f(\gamma_{jk}(\hat{\mathbf{P}},\hat{\mathbf{M}}),a_{jk},b_{jk})$ is the corresponding approximated rate lower-bound obtained by using (20) and $\hat{\mathcal{R}}(\hat{\mathbf{P}},\hat{\mathbf{M}},\mathbf{a},\mathbf{b}) = \sum_j\sum_k \hat{r}_{jk}(\hat{\mathbf{P}},\hat{\mathbf{M}},a_{jk},b_{jk})$. Since the objective function of (22) is concave, the problem (22) is a convex optimization problem.

Note that we only maximize a lower bound of the objective function of (18) in the approximated problem (22). To obtain an efficient solution for problem (18), we tighten the bound by iteratively updating $a_{jk}^{(t_1)}$ and $b_{jk}^{(t_1)}$ as follows. At any iteration $t_1 > 0$, we update these parameters according to (21) using $\tilde{\omega}_{jk}^{(t_1)} = \gamma_{jk}\left(\mathbf{P}^{(t_1)},\mathbf{M}^{(t_1)}\right)$, where $\mathbf{P}^{(t_1)},\mathbf{M}^{(t_1)}$ is the optimal solution of (22). The algorithm to find $\mathbf{P}$ and $\mathbf{M}$ by iteratively solving problem (22) is presented in Algorithm 1 where the iteration index is denoted as t_1. We state the monotone increasing property of the objective function achieved by this algorithm in the following proposition. This property together with the upper-boundedness of the objective function guarantee the convergence of Algorithm 1.

Proposition 2: For a given $\boldsymbol{\rho}$, Algorithm 1 generates a sequence of feasible solutions $\mathbf{P}^{(t_1)},\mathbf{M}^{(t_1)}$ with increasing objective value for problem (18).

Proof: The proof is given in Appendix D. □

Even through Algorithm 1 converges to a stationary point, the resulting stationary point at convergence may not satisfy the Karush-Kuhn-Tucker (KKT) conditions (i.e., necessary conditions for optimality) of the original problem (18) optimized over variables $\mathbf{P}$ and $\mathbf{M}$ for a fixed $\boldsymbol{\rho}$.

D. PILOT ASSIGNMENT ALGORITHM

For fixed $\mathbf{P}$ and $\mathbf{M}$, we notice that the SR function $\mathcal{R}(\boldsymbol{\rho})$ can be rewritten as

$$\begin{aligned}
\mathcal{R}(\boldsymbol{\rho}) = &\underbrace{\sum_{j=1}^{L} r_{j,\tilde{\rho}_{j,1}}(\tilde{\rho}_{1,1},\tilde{\rho}_{2,1},\ldots,\tilde{\rho}_{L,1})}_{\text{weight w.r.t users assigned to the 1}st\text{ pilot}} + \ldots \\
&+ \underbrace{\sum_{j=1}^{L} r_{j,\tilde{\rho}_{j,\kappa}}(\tilde{\rho}_{1,\kappa},\tilde{\rho}_{2,\kappa},\ldots,\tilde{\rho}_{L,\kappa})}_{\text{weight w.r.t users assigned to the }\kappa th\text{ pilot}},
\end{aligned} \tag{23}$$

where there are κ terms in this expression, each of which corresponds to the rate weight achieved by assigning users in different cells to the mth pilot, which concerns variables $\tilde{\rho}_{1,m},\tilde{\rho}_{2,m},\ldots,\tilde{\rho}_{L,m}$. Thus, the underlying pilot assignment problem requires to determine an optimal matching (between users and pilots) that achieves the maximum total weight. In general, this problem is known as the 3-index matching optimization problem, which is known to be NP-hard.

In what follow, we propose to decouple the original assignment problem into sub-problems where in each sub-problem, we aim at optimizing the assignment of K users to κ pilot sequences in only one particular cell. In particular, assuming that we desire to obtain the optimal pilot assignment to users in cell j, i.e., $\rho_{j,k},\ k = 1,\ldots,K$, we fix the remaining pilot assignment variables in other $L-1$ cells and transform the reformulated sub-problem into the standard one-to-one matching problem (says between "users" and "pilots"). To facilitate the description of the solution approach for the pilot assignment sub-problem in each cell, we provide the following definition related to the one-to-one matching problem.

Definition 1 (Matching Problem): Assuming that there are κ pilots that are to be assigned to K users ($\kappa \geq K$). The matching rule is that every user is assigned one pilot sequence and every pilot sequence is assigned to at most one user. Each possible assignment between mth pilot and kth user is associated with a cost $c_{k,m} > 0$, which is given in Table 1. Then, the matching problem can be represented by

TABLE 1. Cost table of assigning pilot to users.

User \ Pilot	1	2	3	...	κ
1	$c_{1,1}$	$c_{1,2}$	$c_{1,3}$	...	$c_{1,\kappa}$
2	$c_{2,1}$	$c_{2,2}$	$c_{2,3}$	...	$c_{2,\kappa}$
3	$c_{3,1}$	$c_{3,2}$	$c_{3,3}$	...	$c_{3,\kappa}$
⋮	⋮	⋮	⋮	⋱	⋮
K	$c_{K,1}$	$c_{K,2}$	$c_{K,3}$	...	$c_{K,\kappa}$

the following optimization problem:

$$\begin{aligned}&\max_{\{x_{k,m}\}} \sum_{(k,m)} x_{k,m}c_{k,m}\\ &s.t. \sum_{k=1}^{K} x_{k,m} \le 1, \quad \forall\, m,\\ &\quad \sum_{m=1}^{\kappa} x_{k,m} = 1, \quad \forall\, k,\\ &\quad x_{k,m} \in \{0,1\}, \quad \forall\, k, m, \end{aligned} \tag{24}$$

where $x_{k,m} \in \{0,1\}$ is the binary assignment variable where $x_{k,m} = 1$ means that the pilot sequence m is assigned to user k, and $x_{k,m} = 0$, otherwise.

This one-to-one matching problem can be solved optimally in polynomial time by applying the well-known Hungarian algorithm (i.e., [27, Algorithm 14.2.3]). Motivated by this matching solution with the notice that if we aim at optimizing the assignment of κ pilot sequences to K users in one particular cell j while fixing the pilot assignments for users from other $L-1$ cells, then the problem degenerates to a standard matching problem, which can be solved by using this well-known method. Specifically, by fixing the pilot assignments for $L-1$ cells except cell j, the pilot assignment problem for cell j is[4]

$$\begin{aligned}&\underset{\boldsymbol{\rho}_j=\rho_{j,1},\ldots,\rho_{j,K}}{\text{maximize}} \ \mathcal{R}(\boldsymbol{\rho}_{-j}, \boldsymbol{\rho}_j)\\ &\text{subject to (C1}'')\text{:}\ \ r_{jk}(\boldsymbol{\rho}_l) \ge R_{\min}, \quad \forall\, j, k, \end{aligned} \tag{25}$$

where $\boldsymbol{\rho}_{-j}$ describes the pilot assignment decisions for all cells $l \neq j$, i.e., variables $\boldsymbol{\rho}_l$ for $l \neq j$. It is noticed that without constraint (C1″), this maximization problem can be solved optimally by applying the Hungarian method to achieve the maximum sum-weight subject to the constraints that each user is associated with one pilot. In order to circumvent the rate constraint (C1″) at each user, we propose a modified Hungarian method to maintain the QoS constraints (C1″). In particular, we define the weight information for user k and cell j for the mth pilot as

$$w^{j}_{k,m} = \sum_{l=1}^{L} r_{l\tilde{\rho}_{l,m}}. \tag{26}$$

This weight represents the total rate achieved by users in all cells who are assigned mth pilot. Then, we can verify whether if each term in each $w^{j}_{k,m}$ satisfies the constraints (C1″) or not. If any of them violates this constraint, we replace $w^{j}_{k,m}$ by a sufficiently small positive value ≈ 0. This is to avoid performing infeasible pilot assignments. Finally, we apply the Hungarian method on the modified weights for cell j as $c_{k,m} = w^{j}_{k,m}, \forall (k,m)$. The iterative pilot assignment algorithm is summarized in Algorithm 2, which can be explained for each iteration t_2 as follows. At iteration t_2, upon finding the solution for cell j, we update value of $\rho^{(t_2)}_{j,1}, \ldots, \rho^{(t_2)}_{j,K}$ by the achieved value of $\rho^{\star}_{j,1}, \ldots, \rho^{\star}_{j,K}$. Then, we move to the next cell and optimize its pilot assignment variables while fixing the variables in other cells. We state the convergence of this algorithm in the following proposition.

[4]For a fixed solution of **P** and **M**, optimization of the objective function of (18) is equivalent to optimizing the first term of the objective function (i.e., the SR) since the second term related to total consumed power is fixed.

Algorithm 2 Pilot Assignment Algorithm

Input: **P**, **M**, $\boldsymbol{\rho}$, Path-loss coefficients β_{ljm}
Output: Pilot assignment $\boldsymbol{\rho}^{\star}$

```
Initialize t2 = 1
repeat
  for j = 1, ..., L do
    Apply modified Hungarian method to solve subproblem-j to obtain (ρ*_{j,1}, ..., ρ*_{j,K}).
    Replace (ρ^{(t2)}_{j,1}, ..., ρ^{(t2)}_{j,K}) = (ρ*_{j,1}, ..., ρ*_{j,K})
  end for
  t2 = t2 + 1
until Convergence of ρ
return Pilot assignment ρ*
```

Proposition 3: For given **P**, **M**, *Algorithm 2 converges after a finite number of iterations.*

Proof: The proposition can be proved as follows. In each iteration, the pilot assignment is determined according to the Hungarian method, which results in the global optimal solution of the corresponding matching problem. Since the total weight, which is the system SR, is maximized in this optimization, the objective function of problem (25) increases over each iteration. Therefore, Algorithm 2 converges since the objection function of problem (25) is bounded from above. □

While Proposition 3 establishes the convergence of Algorithm 2, the resulting stationary pilot assignment solution at convergence may not necessarily satisfy the KKT conditions of problem (18). Therefore, Algorithm 2 only attains a sub-optimal solution of problem (18) in general.

Given the proposed Algorithms 1 and 2, we then utilize them to construct the main algorithm, namely Algorithm 3, which attains a sub-optimal joint pilot assignment and resource allocation solution for the SEE maximization problem. Specifically, for a given value of η, we sequentially optimize the resource allocation and pilot assignments in Algorithm 3 (steps 3-8). Since Algorithm 3 is constructed based on the Dinkelbach method [23], the value of η is updated (in steps 10-18) by using the achieved solution from steps 3-8.

Remark 2: Convergence of an iterative algorithm based on the Dinkelbach's method can be attained if the joint pilot assignment and resource allocation problem (18) *for a given value of η can be solved optimally. For the design considered in our paper, problem* (18) *is a large-scale mixed-integer and non-convex problem, which cannot be solved optimally by algorithms with polynomial complexity algorithm. This is the reason why we develop the sub-optimal Algorithm 3 in which we sequentially optimize the resource allocation*

Algorithm 3 Main Algorithm

Input: Maximum iteration number N, tolerance ϵ

Output: $\eta^\star$ and $(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star)$

Initialize arbitrary pilot assignment $\boldsymbol{\rho}^{(0)}$, $\eta^{(0)} = 0$, BOOL = **false** and $t = 1$.
repeat
  repeat
    Apply Algorithm 1 to solve (22) to obtain $(\mathbf{P}^\star, \mathbf{M}^\star)$
    Set $(\mathbf{P}, \mathbf{M}) = (\mathbf{P}^\star, \mathbf{M}^\star)$
    Apply Algorithm 2 to solve (25) to obtain $\boldsymbol{\rho}^\star$
    Set $\boldsymbol{\rho} = \boldsymbol{\rho}^\star$
  until Convergence of $(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho})$
  Replace $\left(\mathbf{P}^{(t)}, \mathbf{M}^{(t)}, \boldsymbol{\rho}^{(t)}\right) = (\mathbf{P}, \mathbf{M}, \boldsymbol{\rho})$
  if $\hat{\mathcal{R}}\left(\mathbf{P}^{(t)}, \mathbf{M}^{(t)}, \boldsymbol{\rho}^{(t)}\right) - \eta \mathcal{P}(\mathbf{P}^{(t)}, \mathbf{M}^{(t)}) < \epsilon$ **then**
    Set BOOL = **true**
    return $\eta^\star = \frac{\hat{\mathcal{R}}(\mathbf{P}^{(t)}, \mathbf{M}^{(t)}, \boldsymbol{\rho}^{(t)})}{\mathcal{P}(\mathbf{P}^{(t)}, \mathbf{M}^{(t)})}$
    $(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star) := (\mathbf{P}^{(t)}, \mathbf{M}^{(t)}, \boldsymbol{\rho}^{(t)})$
  else
    BOOL = **false**
    return $\eta^{(t)} = \frac{\hat{\mathcal{R}}(\mathbf{P}^{(t)}, \mathbf{M}^{(t)}, \boldsymbol{\rho}^{(t)})}{\mathcal{P}(\mathbf{P}^{(t)}, \mathbf{M}^{(t)})}$
    $t := t + 1$
  end if
until BOOL = **true** or $t = N$

(i.e., optimization of $\mathbf{P}, \mathbf{M}$ *using Algorithm 1) and pilot assignment (i.e., optimization of* ρ *using Algorithm 2). Such solution approach can only achieve a sub-optimal solution for problem* (18) *in general. Therefore, we cannot strictly guarantee the convergence of Algorithm 3 according to the Dinkelbach's method. However, we have performed extensive numerical studies and the convergence of Algorithm 3 is always achieved. The convergence of Algorithm 3 will be demonstrated in Section VI.*

Remark 3: Algorithm 3 can be implemented in a centralized manner via a control unit which has knowledge of the fixed system parameters including required minimum rate $R_{\min}$*, maximum transmit power* $P_{\max}$*, number of antennas* $M_{\max}$ *and must collect dynamic CSI to perform the optimization. The control unit sends the obtained solution* $(\mathbf{P}, \mathbf{M}, \boldsymbol{\rho})$ *to all BSs for implementation upon completing the computation. While the fixed system parameters would not change over the short time period, the required CSI for the optimization process involves only long-term CSI (i.e., long-term channel gains due to path loss* β_{ljk} *as can been seen in the SINR expression in (10)). Therefore, relatively slow estimations and updates are needed to make such reliable long-term CSI available at the control unit. Note that even though pilot assignments are updated iteratively for each cell according to Algorithm 2 to achieve affordable computational complexity, these updates are all performed by the control unit.*

E. FURTHER EXTENSION

We now discuss the scenario with general channel covariance matrices, namely $\mathbf{R}_{ljk} = \mathbf{R}$, where $\mathbf{R} \in \mathbb{R}^{N \times N}$ is a real symmetric positive definite matrix. This general model for the channel covariance matrix reflects the potential correlation among antennas. For this general model of the channel covariance matrices, the SINR expression becomes much more complicated which in turn renders the design of efficient resource allocation algorithms more challenging. To obtain further insights, we consider the case where γ_{p} is sufficiently large, in which the pilot interference dominates the noise power in the CSI estimation process. For the case $1/\gamma_{\mathsf{p}} \ll 1$, the following approximation can be applied to any covariance matrix $\boldsymbol{\Phi}_{jjk}$

$$\begin{aligned}
&\mathrm{Tr}\left(\boldsymbol{\Phi}_{jjk}\right) \\
&= \mathrm{Tr}\left(\beta_{jjk}\mathbf{R}\left(1/\gamma_{\mathsf{p}}\mathbf{I} + \sum_{l=1}^{L}\beta_{jlk}\mathbf{R}\right)^{-1}\mathbf{R}\right) \\
&\simeq \mathrm{Tr}\left(\beta_{jjk}\mathbf{R}\left(\sum_{l=1}^{L}\beta_{jlk}\right)^{-1}\mathbf{R}^{-1}\mathbf{R}\right) = \frac{\beta_{jjk}\mathrm{Tr}\left(\mathbf{R}\right)}{\sum_{l=1}^{L}\beta_{jlk}}. \qquad (27)
\end{aligned}$$

Using this approximation, we can characterize the optimal number of activated antennas for the SR maximization problem in the following theorem.

Theorem 4: When $1/\gamma_{\mathsf{p}} \ll 1$*, the optimal numbers of activated antennas in the SR maximization problem in* (14) *satisfy* $M_j = M_{\max}, \forall j$.

Proof: We prove Theorem 4 by first showing that $\mathrm{Tr}(\mathbf{R})$ increases with the number of activated antennas. Let us consider $M_1, M_2 \in \mathbb{Z}^+$ such that $M_1 < M_2$. Denote $\mathbf{R}_1$ and $\mathbf{R}_2$ as the corresponding $M_1 \times M_1$ and $M_2 \times M_2$ symmetric positive definite matrices, respectively, where $\mathbf{R}_2$ are built by adding more non-negative columns and rows to matrix $\mathbf{R}_1$. Then, we have $\mathrm{Tr}(\mathbf{R}_2) = \sum_{i=1}^{M_2} r_{ii} > \sum_{i=1}^{M_1} r_{ii} = \mathrm{Tr}(\mathbf{R}_1)$. Therefore, $\mathrm{Tr}(\mathbf{R})$ increases with the number of activated antennas.

We now consider the SR maximization problem (14). Since we can approximate $\mathrm{Tr}(\boldsymbol{\Phi})$ as a function of $\mathrm{Tr}(\mathbf{R})$ for $1/\gamma_{\mathsf{p}} \simeq 0$ as in (27), we can conclude that $\mathrm{Tr}(\boldsymbol{\Phi})$ increases with the number of activated antennas. By repeating the same steps as in the proof of Theorem 5, we can finally prove that the optimal number of antennas at any BS is equal to $M_{\max}$. □

Therefore, when $1/\gamma_{\mathsf{p}} \ll 1$ we only need to optimize the power allocation and pilot assignment for the SR optimization problem. In addition, since we have $M_j = M_{\max}, \forall j$ at optimality, we can re-write the SINR expression (7) in terms of transmit powers and pilot assignment variables $\boldsymbol{\rho}$ as in (28) on bottom of the next page. where $\tilde{A}_{jjk} = \frac{\beta_{jjk}}{\sum_{l=1}^{L}\beta_{jl}\tilde{\rho}_{l,\rho_{j,k}}}$ and $\tilde{B}_{ljk} = \frac{\beta_{ljk}}{\sum_{n=1}^{L}\beta_{ln\tilde{\rho}_{n,\rho_{j,k}}}}$. Then, the proposed algorithms in the previous sub-sections can still be employed to solve the SR maximization problem using this SINR expression.

Remark 4: For a general value of γ_{p}*, it is not tractable to express the SR as an explicit function of the number of activated antennas. This renders the development of efficient resource allocation algorithms for the SR and SEE optimization problems difficult. One approach to overcome*

Algorithm 4 Sum Rate Maximization Algorithm

Input: Maximum iteration number N, tolerance ϵ
Output: $(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star)$

1: By Theorem 5, we set $M_j^\star = M_{\max},\ \forall j$.
2: Initialize arbitrary pilot assignment $\boldsymbol{\rho}$ and $t = 1$.
3: **repeat**
4: Apply Algorithm 1 to solve (22) to obtain $\mathbf{P}$ where $\eta = 0$ and $M_j^\star = M_{\max},\ \forall j$, then update $\mathbf{P}$.
5: Apply Algorithm 2 to solve (25) to obtain and update $\boldsymbol{\rho}$.
6: Set $t = t + 1$.
7: **until** Convergence of $(\mathbf{P}, \boldsymbol{\rho})$ or $t = N$.
8: **return** $(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star) = (\mathbf{P}, \mathbf{M}_{\max}, \boldsymbol{\rho})$.

this challenge is to upper- and lower-bound Tr $(\boldsymbol{\Phi})$ *by corresponding functions of the number of antennas. Then, efficient algorithms to solve SR and EE optimization problems can be developed based on these bounds. Due to the space constraint, we leave this design for our future works.*

V. RESOURCE ALLOCATION ALGORITHMS FOR SR MAXIMIZATION PROBLEM

In this section, we discuss the optimality characterization and low-complexity algorithms to solve problem (14). Specifically, we characterize the optimal number of activated antennas at the BSs for problem (14) in the following theorem.

Theorem 5: The optimal values $M_j^\star$ of the SR optimization problem in (14) *satisfy $M_j^\star = M_{\max},\ \forall j$.*

Proof: The proof is provided in Appendix E. □

This theorem means that it is always optimal in terms of system SR if we utilize all available antennas for transmission. Due to the results of Theorem 5, one can simply set $\mathbf{M}^\star$ equal to $\mathbf{M}_{\max}$ for the SR maximization problem and one only needs to optimize the pilot assignment and power allocation. As can be observed, the problem (14) is equivalent to the problem (18) for $\eta = 0$. Therefore, the SR optimization problem can be solved by Algorithm 3 using steps 3-8 only with $\eta = 0$. We present an algorithm to solve the SR maximization problem in Algorithm 4.

VI. NUMERICAL RESULTS

We evaluate the performance of the proposed algorithms in terms of achieved SEE and SR. Simulation parameters are chosen as in Table 2, which are used to obtain following numerical results unless stated otherwise. We will compare the performance of the propose algorithm and a "conventional" scheme where the conventional scheme simply allocates the kth pilot sequence to the kth user for $k = 1, 2, \ldots, K$ in each cell and it optimizes the power allocation and number of activated antennas by using Algorithm 1. The results corresponding to the proposed algorithm and the conventional scheme are indicated as "Proposed" and "Conventional" in the following figures, respectively. The obtained SEE in all figures are normalized to one Hz of system bandwidth.

TABLE 2. System parameters.

Parameter	Description	Value
K	Number of user per cell	5
L	Number of cells	3
N_0	Noise power	-120 dBm
$M_{\max}$	Number of available antennas	100
$P_{\max}$	Maximum transmit power	0 dBm
P_c	Circuit power per antenna	30 dBm
P_0	Static basic power	40 dBm
$R_{\min}$	Minimum required rate	2 bits/s/Hz
ν	Power inefficiency factor	5
κ	Number of pilots	5
γ_p	Pilot SNR	10 dB

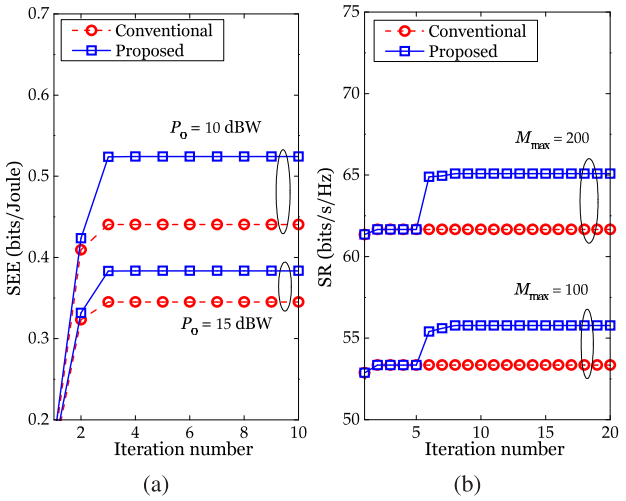

FIGURE 1. Convergence of the proposed algorithms for (a) SEE at different circuit power values P_0; (b) SR at different maximum number of antennas $M_{\max}$.

Fig. 1a shows the evolution of SEE under the proposed algorithm over iterations for $M_{\max} = 100$ antennas, $P_0 = 10, 15$ dBW. It can be observed that the SEE converges to a fixed value after just three iterations under the proposed iterative algorithm. In addition, the proposed algorithm achieves noticeably higher SEE compared to that of the conventional pilot assignment scheme where the SEE gains about 20% and 10% are achieved by the proposed algorithm over

$$\gamma_{jm}(\mathbf{P}, \boldsymbol{\rho}) = \frac{P_{jk}\beta_{jjk}\tilde{A}_{jjk}\mathrm{Tr}\left(\mathbf{R}\right)}{\frac{P_{jk}\beta_{jjk}\left(\tilde{A}_{jjk}+1\right)\mathrm{Tr}(\mathbf{R}^2)}{\mathrm{Tr}(\mathbf{R})} + \frac{\sum_{(l,i)\neq(j,k)}P_{li}\beta_{ljk}\mathrm{Tr}(\mathbf{R}^2)}{\mathrm{Tr}(\mathbf{R})} + \sum_{l\neq j}P_{l\tilde{\rho}_{l,\rho_{j,k}}}\beta_{ljk}\left(\tilde{B}_{ljk}\frac{\mathrm{Tr}(\mathbf{R}^2)}{\mathrm{Tr}(\mathbf{R})} + \tilde{B}_{ljk}\mathrm{Tr}\left(\mathbf{R}\right) - 1\right) + N_0} \tag{28}$$

the conventional scheme for $P_0 = 10, 15$ dBW, respectively. The performance gap between the two schemes reduces as the fixed power consumption P_0 increases. Similarly, Fig. 1b shows the variations of the SRs over iterations due to the two schemes for $P_{\max} = 0$ dB and $M_{\max} = 100, 200$ antennas. For each value of $M_{\max}$, we observe that the proposed pilot assignment scheme achieves better SR than that of the conventional pilot assignment scheme. In addition, both schemes achieve higher SR at higher value of $M_{\max} = 200$ compared to the case with $M_{\max} = 100$.

In Fig. 2 and Fig. 3, we compare the SR achieved by the two schemes versus the maximum number of antennas $M_{\max}$ and number of users per cell K, respectively. Fig. 2 shows that the SR at $P_{\max} = 0$ dBW due to either scheme is better than the corresponding SR at $P_{\max} = -50$ dBW. In addition, the relative SR improvement

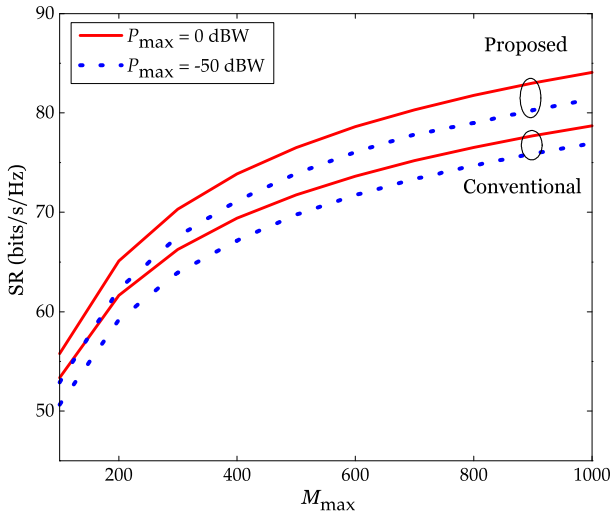

FIGURE 2. Achieved SR versus maximum number of antennas $M_{\max}$ for maximum powers $P_{\max} = -50, 0$ dBW.

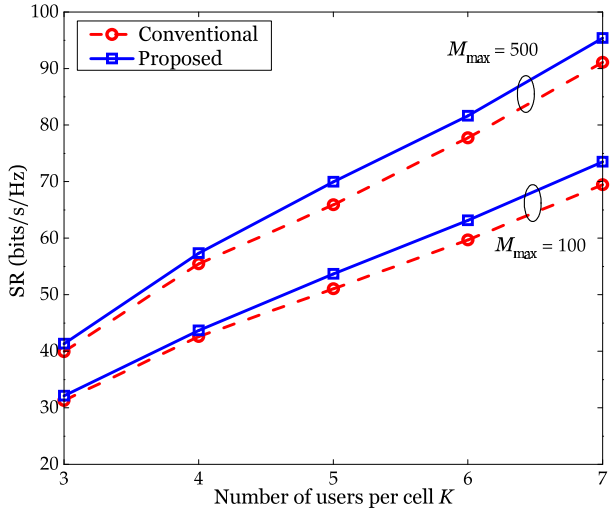

FIGURE 3. Achieved SR versus K for different maximum number of antennas $M_{\max} = 100, 500$.

as the maximum power increases from $P_{\max} = 0$ dBW to $P_{\max} = -50$ dBW is quite moderate in the high $M_{\max}$ regime. This is because if $M_{\max}$ is sufficiently large, BSs tend to utilize less power than their maximum budget $P_{\max}$ to attain the maximum SR. Moreover, this figure shows that the SR gap between two schemes becomes larger as $M_{\max}$ increases. In addition, Fig. 3 suggests that the SR increases with the larger number of users per cell and the performance gap between the two schemes also becomes larger as K grows. This confirms the effectiveness of the proposed algorithm.

In Fig. 4, we show the SEE achieved by the proposed algorithm when it is used to solve the SEE and SR maximization problems, whose results are indicated as "max SEE" (Algorithm 3) and "max SR" (only apply step 3 to step 8 of Algorithm 3), respectively. It is shown that the achieved SEE of the "max SEE" scheme always dominates that of the "max SR" scheme under both "Conventional" and "Proposed" pilot assignment strategies. This is expected since maximization of the system SR may indeed hurt the SEE due to the excessive power consumption. It can also be observed that when the maximum transmit power $P_{\max}$ is sufficiently large, the SEE of the "max SEE" scheme approaches a constant. In fact, limited transmit power and number of activated antennas are required to maximize the SEE; therefore, large BS power does not help improve the SEE in the large power regime. In contrast, the "max SR" scheme utilizes more power and number of antennas to maximize the SR when the maximum power $P_{\max}$ is larger. This results in the decrease of SEE for the "max SR" scheme as $P_{\max}$ becomes sufficiently large.

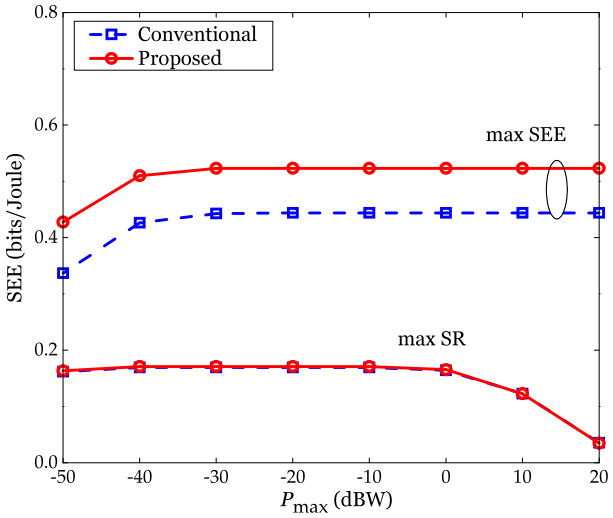

FIGURE 4. Achieved SEE versus maximum power $P_{\max}$ for $M_{\max} = 100$.

Fig. 5 shows the SEE versus the required minimum rate $R_{\min}$. We can see from this figure that the SEE of both proposed and conventional schemes reduces as $R_{\min}$ increases. This is because as $R_{\min}$ increases, larger power and number of antennas are required to maintain the QoS constraints, which results in the decrease of SEE.

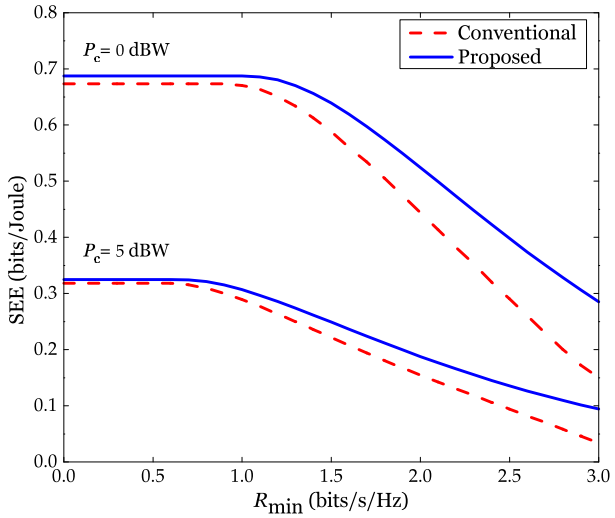

FIGURE 5. Achieved SEE versus required minimum rate $R_{\min}$ for circuit power $P_c = 0, 5$ dBW.

Interestingly, the gap between the two schemes increases quite significantly as $R_{\min}$ increases. Specifically, as $R_{\min}$ reaches 3bits/s/Hz, the SEE gain of the proposed scheme over the conventional pilot assignment scheme is about 100%. This confirms the significance of the pilot assignment optimization in mitigating the pilot contamination effect.

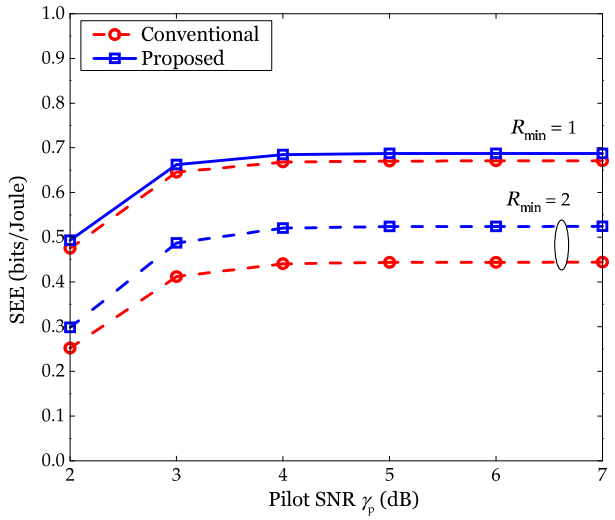

FIGURE 6. Achieved SEE versus pilot SNR γ_p for different $R_{\min} = 1, 2$ bits/s/Hz.

In Fig. 6 and Fig. 7, we show the variations of SEE and SR versus the pilot SNR γ_p, respectively. In both figures, we observe that the gap between the "Conventional" and "Proposed" schemes are fairly small at low γ_p and becomes a constant at sufficiently high pilot SNR value, i.e., $\gamma_p \geq 5$ dB. Fig. 6 illustrates the SEE at different values of $R_{\min} = 1$, 2 bits/s/Hz where increasing the pilot SNR provides larger SEE gain, which is more significant for higher $R_{\min}$. This is consistent with the results demonstrated in Fig. 5. In addition, the similar trend in the SR can be observed in Fig. 7 and higher SR gain is achieved for higher values of $M_{\max}$.

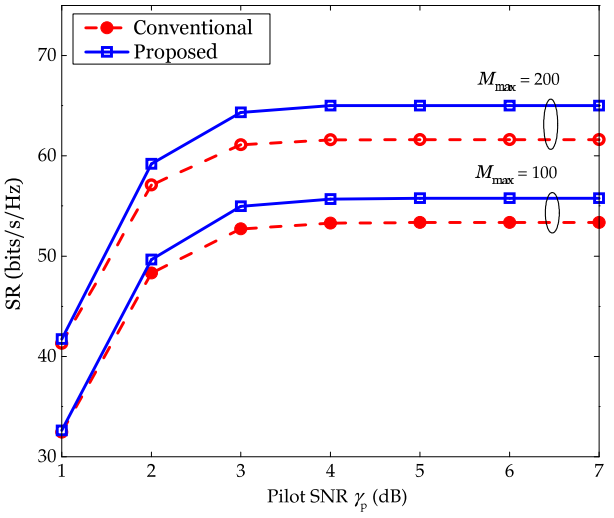

FIGURE 7. Achieved SR versus pilot SNR γ_p for different $M_{\max} = 100, 200$.

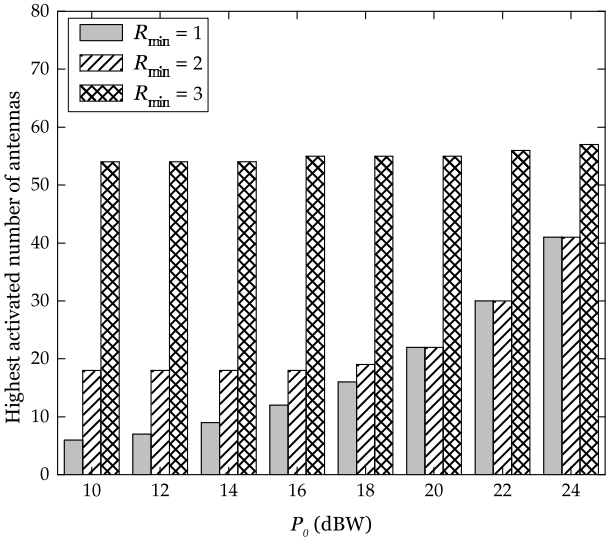

FIGURE 8. Highest number of activated antennas among BSs versus circuit power P_0 for different $R_{\min} = 1, 2, 3$ bits/s/Hz.

In Fig. 8, we show the highest number of activated antennas among BSs versus the basic power P_0 for different values of $R_{\min}$ under the SEE maximization. It can be observed that the highest number of activated antennas tends to increase as P_0 increases. This can be interpreted as follows. When the basic power P_0 is larger, more radio resources in terms of transmit power and number of activated antennas should be utilized to maximize the SEE. The increase in the highest number of activated antennas among BSs is more significant for smaller values of $R_{\min}$. Moreover, when $R_{\min}$ increases, more BS antennas should be employed to meet the required minimum rate.

In Fig. 9, we show the SR when the channel covariance matrix $\mathbf{R}_{ljm} = \mathbf{R}$ is modeled as $\mathbf{R} = [H_{kl}]$ where $H_{kl} = \lambda^{|k-l|}$, $0 \leq \lambda \leq 1$ where λ is the covariance factor capturing antenna correlation. It can be observed that the achieved SR decreases as λ increases. This can be interpreted as follows. As the covariance factor λ increases,

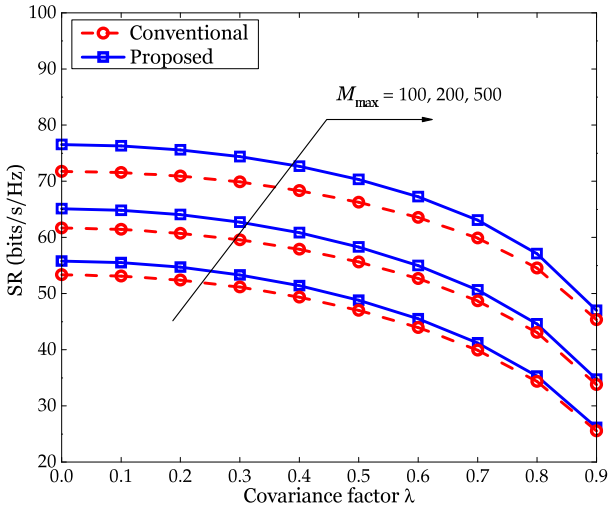

FIGURE 9. Achieved SR versus covariance factor λ at different $M_{max} = 100, 200, 300$.

adjacent antennas in the array are more correlated, which in turns increases the intra-cell and inter-cell interference and degrades the achieved SR performance. In addition, as λ increases, the SR gap between the "Proposed" and "Conventional" schemes decreases and approaches zero at sufficiently high λ.

VII. CONCLUSION

In this paper, we have studied the SR and SEE optimization problems for the MU-MC massive MIMO network. We have proposed efficient algorithms to optimize the number of activated antennas together with power allocation and pilot assignment. Specifically, we have developed a novel iterative algorithm to solve the considered problems where we sequentially solve two sub-problems in each iteration, namely the first sub-problem optimizes the power allocation and number of activated antennas and the second sub-problem optimizes the pilot assignments. We have developed an SCA-based algorithm to solve the first sub-problem for a given pilot assignment solution. Also, we have proposed to address the pilot assignment sub-problem through solving a related weight-based assignment problem. Numerical results have confirmed the convergence of the proposed algorithms and that the proposed algorithm can achieve much better performance in terms of total SR and SEE than that due to the conventional pilot assignment scheme.

APPENDIX A
PROOF OF THEOREM 1

We adopt the derivation in [3] to prove Theorem 1. Toward this end, we first recall some results on the product of two Gaussian vectors in [28], which are given in the following proposition.

Proposition 4: Denote $\mathbf{x} \in \mathbb{C}^{N\times 1}$ *as a complex normal random vector with zero mean and real, symmetry and positive-definite covariance matrix* $\mathbf{\Sigma}$, *e.g.,* $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$. *Then, if* $\mathbf{A}$ *and* $\mathbf{B}$ *are two real non-stochastic* $\mathbb{R}^{N\times N}$ *square matrices, we have the following expression*

$$\mathbb{E}\left\{\left(\mathbf{x}^{\dagger}\mathbf{A}\mathbf{x}\right)\left(\mathbf{x}^{\dagger}\mathbf{B}\mathbf{x}\right)\right\} = 2Tr\left(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\right) + Tr\left(\mathbf{A}\mathbf{\Sigma}\right)Tr\left(\mathbf{B}\mathbf{\Sigma}\right). \tag{29}$$

In general, the parameter ξ_{jk} can be expressed as

$$\xi_{jk} = \frac{1}{\mathbb{E}\left\{\widehat{\mathbf{h}}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right\}} = \frac{1}{\mathrm{Tr}\left(\mathbf{\Phi}_{jjk}\right)}. \tag{30}$$

Thus, by applying the expression of $\xi_{jk}, \forall j, k$, we can manipulate the numerator of (7) as follows:

$$\begin{aligned}\xi_{jk}&\left|\mathbb{E}\left\{\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right\}\right|^2 \\ &= \frac{1}{\mathrm{Tr}\left(\mathbf{\Phi}_{jjk}\right)}\left|\mathrm{Tr}\left(\mathbf{\Phi}_{jjk}\right)\right|^2 \\ &= \left(\frac{M_j\beta_{jjk}}{1/\gamma_{\mathrm{p}} + \sum_{l=1}^{L}\beta_{jl}\tilde{\rho}_{l,\rho_{j,k}}}\right) = M_j A_{jjk}.\end{aligned} \tag{31}$$

In addition, we can manipulate the first term in (8) as follows:

$$\begin{aligned}&\xi_{jk}\mathrm{Var}\left(\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right) \\ &= \xi_{jk}\left(\mathbb{E}\left\{\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\widehat{\mathbf{h}}_{jjk}^{\dagger}\mathbf{h}_{jjk}\right\} - \mathbb{E}\left\{\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right\}^2\right) \\ &= \xi_{jk}\left(\mathbb{E}\left\{\left|\widehat{\mathbf{h}}_{jjk}\right|^4\right\} + \mathbb{E}\left\{\widetilde{\mathbf{h}}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\widehat{\mathbf{h}}_{jjk}^{\dagger}\widetilde{\mathbf{h}}_{jjk}\right\} - \mathrm{Tr}\left(\mathbf{\Phi}_{jjk}\right)^2\right) \\ &= 2\xi_{jk}\mathrm{Tr}\left(\mathbf{\Phi}_{jjk}^2\right) + \xi_{jk}\mathrm{Tr}\left((\mathbf{R}_{jjk} - \mathbf{\Phi}_{jjk})\mathbf{\Phi}_{jjk}\right).\end{aligned} \tag{32}$$

Then, if $\mathbf{R}_{jjk} = \mathbf{I}_{M_l}$, we have

$$\xi_{jk}\mathrm{Var}\left(\mathbf{h}_{jjk}^{\dagger}\widehat{\mathbf{h}}_{jjk}\right) = \frac{\beta_{jjk}}{1/\gamma_{\mathrm{p}} + \sum_{l=1}^{L}\beta_{jl}\tilde{\rho}_{l,\rho_{j,k}}} + 1 = A_{jjk} + 1. \tag{33}$$

Similarly, the second term of (8) can be expressed as

$$\begin{aligned}&\xi_{li}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\widehat{\mathbf{h}}_{lli}\right|^2\right\} \\ &= \xi_{li}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\sqrt{\beta_{lli}}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\mathbf{Y}_l\boldsymbol{\psi}_{\rho_{l,i}}\right|^2\right\} \\ &= \begin{cases}\xi_{li}\mathrm{Tr}\left(\mathbf{R}_{ljk}\mathbf{\Phi}_{lli}\right), & \text{if} \quad \rho_{l,i} \neq \rho_{j,k} \\ \xi_{li}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\sqrt{\beta_{lli}}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\sqrt{\beta_{ljk}}\mathbf{h}_{ljk}\right|^2\right\} + \xi_{li}\mathrm{Tr}\left(\mathbf{R}_{ljk}\mathbf{\Phi}_{lli}\right) \\ -\xi_{li}\mathrm{Tr}\left(\sqrt{\beta_{lli}\beta_{ljk}}\mathbf{R}_{ljk}\mathbf{\Phi}_{ljk}\bar{\mathbf{Q}}_{li}\mathbf{R}_{lli}\right), & \text{if} \quad \rho_{l,i} = \rho_{j,k}.\end{cases}\end{aligned} \tag{34}$$

By applying the results in Proposition 4, we can rewrite the term $\xi_{li}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\sqrt{\beta_{lli}}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\sqrt{\beta_{ljk}}\mathbf{h}_{ljk}\right|^2\right\}$ in (34) as

$$\begin{aligned}&\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\sqrt{\beta_{lli}}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\sqrt{\beta_{ljk}}\mathbf{h}_{ljk}\right|^2\right\} \\ &= \mathbb{E}\left\{\beta_{lli}\beta_{ljk}\mathbf{h}_{ljk}^{\dagger}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\mathbf{h}_{ljk}\mathbf{h}_{ljk}^{\dagger}\bar{\mathbf{Q}}_{li}^{\dagger}\mathbf{R}_{lli}^{\dagger}\mathbf{h}_{ljk}\right\} \\ &= 2\beta_{lli}\beta_{ljk}\mathrm{Tr}\left(\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\mathbf{R}_{ljk}\bar{\mathbf{Q}}_{li}^{\dagger}\mathbf{R}_{lli}^{\dagger}\mathbf{R}_{ljk}\right) + \left|\mathrm{Tr}\left(\mathbf{\Phi}_{ljk}\right)\right|^2 \\ &= 2\mathrm{Tr}\left(\sqrt{\beta_{lli}\beta_{ljk}}\mathbf{R}_{ljk}\mathbf{\Phi}_{ljk}\bar{\mathbf{Q}}_{li}^{\dagger}\mathbf{R}_{lli}^{\dagger}\right) + \left|\mathrm{Tr}\left(\mathbf{\Phi}_{ljk}\right)\right|^2,\end{aligned} \tag{35}$$

$$\sum_{(l,i)\neq(j,k)} \xi_{li}P_{li}\beta_{ljk}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\widehat{\mathbf{h}}_{lli}\right|^2\right\} = \sum_{(l,i)\neq(j,k)} \delta(\rho_{l,i}-\rho_{j,k})\xi_{li}P_{li}\beta_{ljk}\left[\mathrm{Tr}\left(\sqrt{\beta_{lli}\beta_{ljk}}\mathbf{R}_{ljk}\mathbf{\Phi}_{ljk}\bar{\mathbf{Q}}_{li}\mathbf{R}_{lli}\right)+\left|\mathrm{Tr}\left(\mathbf{\Phi}_{ljk}\right)\right|^2\right] + \sum_{(l,i)\neq(j,k)}\left[1-\delta(\rho_{l,i}-\rho_{j,k})\right]\xi_{li}P_{li}\beta_{ljk}\mathrm{Tr}\left(\mathbf{R}_{ljk}\mathbf{\Phi}_{lli}\right) \tag{36}$$

$$\sum_{(l,i)\neq(j,k)} \xi_{li}P_{li}\beta_{ljk}\mathbb{E}\left\{\left|\mathbf{h}_{ljk}^{\dagger}\widehat{\mathbf{h}}_{lli}\right|^2\right\} = \sum_{l\neq j}\frac{P_{l\tilde{\rho}_{l,\rho_{j,k}}}\beta_{ljk}\beta_{ljk}(M_l+1)}{1/\gamma_{\mathsf{p}}+\sum_{n=1}^{L}\beta_{ln\tilde{\rho}_{n,\rho_{j,k}}}} + \sum_{(l,i)\neq(j,k)}\left[1-\delta(\rho_{l,i}-\rho_{j,k})\right]P_{li}\beta_{ljk} = \sum_{l\neq j}P_{l\tilde{\rho}_{l,\rho_{j,k}}}\beta_{ljk}(B_{ljk}M_l+B_{ljk}-1)+\sum_{(l,i)\neq(j,k)}P_{li}\beta_{ljk} \tag{37}$$

where we have denoted $\mathbf{\Phi}_{ljk} = \sqrt{\beta_{lli}\beta_{ljk}}\mathbf{R}_{lli}\bar{\mathbf{Q}}_{li}\mathbf{R}_{ljk}$. Since $\mathbf{R}_{lli}$ is a real symmetric matrix, we have $\mathbf{R}_{lli}^{\dagger} = \mathbf{R}_{lli}$ and $\bar{\mathbf{Q}}_{li}^{\dagger} = \bar{\mathbf{Q}}_{li}$. Using the results in (35), the second term of (8) can be expressed explicitly as (36), as shown at the top of this page, where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. Then, if $\mathbf{R}_{jjk} = \mathbf{I}_{M_l}$, we have the result as in (37), as shown at the top of this page. By replacing the terms (31), (33), and (37) into (8), we complete the proof.

APPENDIX B
PROOF OF PROPOSITION 1

We prove part (i) of Proposition 1 by first noting that for $M_{\max} = 0$, problem (19) is infeasible or the feasible set of $\mathbf{M}$ is $\mathcal{M}_0 = \emptyset$. Therefore, we are only interested in the system where there exists a feasible solution when $M_{\max}$ is sufficiently large. By denoting the feasible set of $\mathbf{M}$ as $M_{\max} \to \infty$ as $\mathcal{M}_\infty$ then obviously there exists a set of thresholds denoted by $\mathbf{M}^{th}$ that satisfies statement (i) of Proposition 1.

We prove part (ii) of Proposition 1 by proving that if there exists two optimal solutions in the number of antenna $\mathbf{M}^{\star,1}$ and $\mathbf{M}^{\star,2}$ of problem (19) corresponding to two different value $M_{\max,1}$ and $M_{\max,2}$, where $\mathbf{M}_{\max,2} \succ \mathbf{M}_{\max,1} \succ \mathbf{M}^{th}$, then we must have $\mathbf{M}^{\star,1} = \mathbf{M}^{\star,2}$. The proof is completed as follows.

Suppose that $\mathbf{M}^{\star,1} \neq \mathbf{M}^{\star,2}$. For convenience, we denote $\mathcal{M}_1$ and $\mathcal{M}_2$ as the feasible sets of $\mathbf{M}$ corresponding to two values $M_{\max,1}$ and $M_{\max,2}$, respectively. If $\mathbf{M}^{\star,2} \in \mathcal{M}_1$, we see that $\mathbf{M}^{\star,2}$ is the feasible solution of (19) in set $\mathcal{M}_1$. Since $\mathbf{M}^{\star,1}$ is the optimal solution of (19) in set $\mathcal{M}_1$, we must have $\max_j M_j^{\star,1} \leq \max_j M_j^{\star,2}$. However, since $\mathbf{M}^{\star 2} \in \mathcal{M}_2$, $\mathbf{M}^{\star 2}$ is the optimal solution of (19) in set $\mathcal{M}_2$ and $\mathbf{M}^{\star 1}$ is the feasible solution of (19) in set $\mathcal{M}_2$, we have $\max_j M_j^{\star 2} \leq \max_j M_j^{\star 1}$. Thus, $\mathbf{M}^{\star,2} = \mathbf{M}^{\star,1}$.

If $\mathbf{M}^{\star 2} \in \mathcal{M}_2 \setminus \mathcal{M}_1$, we always have $\max_j M_j^{\star 1} \leq \max_j M_j^{\star 2}$. And since $\mathbf{M}^{\star 2} \in \mathcal{M}_2$, $\mathbf{M}^{\star 2}$ is the optimal solution of (19) in set $\mathcal{M}_2$ and $\mathbf{M}^{\star 1}$ is the feasible solution of (19) in set $\mathcal{M}_2$, we have $\max_j M_j^{\star 2} \leq \max_j M_j^{\star 1}$. Thus, $\mathbf{M}^{\star 2} = \mathbf{M}^{\star 1}$. This completes the proof.

APPENDIX C
PROOF OF THEOREM 3

To proceed, we first introduce the following proposition.

Proposition 5: Consider function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sum_{i=1}^{L}\log(1+a_i x) - \theta b x$. If $\theta < \frac{\sum_i a_i}{b}$, then $f(x)$ has the following properties:

(i) *$f(x)$ achieves its maximum positive value at certain $x > 0$.*

(ii) *There exists a unique $x > 0$ that satisfies $f(x) = 0$.*

Proof: Consider the function $f(x) = \sum_{i=1}^{L}\log(1+a_i x) - \theta b x$. The first derivative of $f(x)$ can be expressed as

$$\frac{\partial f(x)}{\partial x} = \sum_{i=1}^{L}\frac{a_i}{1+a_i x} - \theta b. \tag{38}$$

Evaluating this derivative at $x = 0$, we have

$$\left.\frac{\partial f(x)}{\partial x}\right|_{x=0} = \sum_{i=1}^{L} a_i - \theta b. \tag{39}$$

Since we have assumed $\theta < \frac{\sum_i a_i}{b}$, the right-hand-side (RHS) of (39) is positive at $x = 0$. However, as $x \to \infty$, the RHS of (39) is negative. Thus, there exists a solution $x_0 > 0$ that satisfies $\partial f(x)/\partial x = 0$ at $x = x_0$. Since $\partial^2 f(x)/\partial x^2 = \sum_{i=1}^{L}\frac{-a_i^2}{(1+a_i x)^2} < 0$, $f(x)$ is a concave function that takes the maximum value at $x_0 > 0$. This complete the proof of part (i) of Proposition 5.

Since $f(x_0) > 0$ (since $f(0) = 0$) and $\lim_{x\to\infty} f(x) < 0$, there exists a solution x_1 that satisfies $f(x_1) = 0$. This completes the proof for part (ii) of Proposition 5. □

Recall that the objective function (18) is equal to 0 at the optimal solution $\mathbf{P}^\star$, $\mathbf{M}^\star$ and $\eta^\star$. We prove Theorem 3 by showing that if there is at least one optimal $M_j^\star$ equal to $M_{\max}$ when $M_{\max} \gg 1$, the objective function of (18) is different from 0. Let us star by giving the upper bound the objective function of (18) as

$$g(\mathbf{M}^\star) = \mathcal{R}(\mathbf{P}^\star, \mathbf{M}^\star, \boldsymbol{\rho}^\star) - \eta^\star\mathcal{P}(\mathbf{P}^\star, \mathbf{M}^\star) < \sum_{j,k}\log\left[1+\frac{P_{jk}^\star\beta_{jjk}A_{jjk}M_j^\star}{N_0}\right] - \eta^\star P_c\sum_{l=1}^{L}M_l^\star. \tag{40}$$

By denoting $\bar{M} = \max_{j=1,\ldots,L} M_j^\star$, which represents the highest value of the optimal numbers of antennas at all BSs,

it follows that

$$\begin{aligned} g(\mathbf{M}^\star) &< \sum_{j,k} \log\left[1+\frac{P^\star_{jk}\beta_{jjk}A_{jjk}M^\star}{N_0}\right] - \eta^\star P_c \sum_{l=1}^{L} M^\star_l \\ &< \sum_{j,k} \log\left[1+\frac{P^\star_{jk}\beta_{jjk}A_{jjk}\bar{M}}{N_0}\right] - \eta^\star P_c \bar{M} \\ &< \sum_{j,k} \log\left[1+\frac{P^\star_{jk}\beta_{jjk}A_{jjk}\bar{M}}{N_0}\right] - \tilde{\eta} P_c \bar{M}, \end{aligned} \tag{41}$$

where $\tilde{\eta} \leq \eta^\star$ is a feasible SEE achieved at some feasible solution of $\mathbf{P}, \mathbf{M}$. In fact, by applying the results of Proposition 1, we can see that the feasible solution sets of (19) and (18) are the same. Moreover, the optimal solution of problem (19) can be chosen as the feasible solution corresponding to the SEE of $\tilde{\eta}$ and this solution is also independent of $M_{\max}$. Moreover, since $\tilde{\eta} > 0$, there exists a sufficiently small $\eta_0 < \tilde{\eta}$ such that $\eta_0 < \frac{\sum_{j,k} P_{\max}\beta_{jjk}/N_0}{P_c}$.

By denoting $f(\bar{M}) = \sum_{j,k} \log\left[1+\frac{P_{\max}\beta_{jjk}\bar{M}}{N_0}\right] - \eta_0 P_c \bar{M}$, we can see that $f(\bar{M})$ has the same properties of the function $f(x)$ described in Proposition 5 with $a_j = P_{\max}\beta_{jjk}/N_0$, $b = P_c$ and $\theta = \eta_0$. Thus, $f(\bar{M}) < 0$ when $\bar{M} > M_0$, where M_0 is the solution of $f(\bar{M}) = 0$ according to Proposition 5. In addition, $f(\bar{M})$ is a concave function and positive in $0 < \bar{M} < M_0$.

Next, we prove that there is no element $M^\star_j$ in the optimal solution of (18) exceeding M_0 by contradiction. Assume that there is at least one $M^\star_j > M_0$. From the optimality property, we see that the function takes the value of 0 at optimal $\mathbf{P}^\star, \mathbf{M}^\star$. However, we have $f(M^\star_j) < 0$ for $M^\star_j > M_0$ due to the properties of function $f(.)$. This contradicts to the assumption that $f(\bar{M}) > g(\mathbf{M}^\star) = 0$. Thus, there not exists any $M^\star_j$ such that $M^\star_j > M_0$.

In summary, we can state that each element $M^\star_j$ of the optimal solution $\mathbf{M}^\star$ is upper bounded by M_0 which is independent of $M_{\max}$. This completes the proof of the theorem.

APPENDIX D
PROOF OF PROPOSITION 2

Assume that the optimal solutions for problem (22) at t_1th step are $(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)})$. We will prove that the value of the objective function in (22) evaluated at $\left(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}\right)$ is smaller than that evaluated at $\left(\mathbf{P}^{(t_1+1)}, \mathbf{M}^{(t_1+1)}\right)$ as

$$\begin{aligned} &\mathcal{R}_{jk}\left(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}\right) - \eta \mathcal{P}\left(\mathbf{P}^{(t_1)}, \mathbf{M}^{(t_1)}\right) \\ &= \hat{\mathcal{R}}_{jk}\left(\hat{\mathbf{P}}^{(t_1)}, \hat{\mathbf{M}}^{(t_1)}, \mathbf{a}^{(t_1+1)}, \mathbf{b}^{(t_1+1)}\right) - \eta \mathcal{P}\left(\hat{\mathbf{P}}^{(t_1)}, \hat{\mathbf{M}}^{(t_1)}\right) \\ &\leq \hat{\mathcal{R}}_{jk}\left(\hat{\mathbf{P}}^{(t_1+1)}, \hat{\mathbf{M}}^{(t_1+1)}, \mathbf{a}^{(t_1+1)}, \mathbf{b}^{(t_1+1)}\right) \\ &\quad - \eta \mathcal{P}\left(\hat{\mathbf{P}}^{(t_1+1)}, \hat{\mathbf{M}}^{(t_1+1)}\right) \\ &\leq \mathcal{R}_{jk}\left(\mathbf{P}^{(t_1+1)}, \mathbf{M}^{(t_1+1)}\right) - \eta \mathcal{P}\left(\mathbf{P}^{(t_1+1)}, \mathbf{M}^{(t_1+1)}\right), \end{aligned} \tag{42}$$

where the first equality holds due to the fact that the approximation is tight at the current value of $\left\{\hat{\mathbf{P}}^{(t_1)}, \hat{\mathbf{M}}^{(t_1)}, \mathbf{a}^{(t_1+1)}, \mathbf{b}^{(t_1+1)}\right\}$; the first inequality holds because the SCA-based algorithm achieves the optimal solution at $\left\{\hat{\mathbf{P}}^{(t_1+1)}, \hat{\mathbf{M}}^{(t_1+1)}\right\}$. Moreover, the second inequality follows from the nature of the logarithmic approximation. Thus, we have completed the proof.

APPENDIX E
PROOF OF THEOREM 5

We prove Theorem 5 by using the contradiction method. For a given fixed $\boldsymbol{\rho}$, let the optimal solution of (14) be $\mathbf{P}^\star, \mathbf{M}^\star$. Suppose that there is at least one element of $\mathbf{M}^\star$ to be smaller than $M_{\max}$. We will prove that there exists another feasible solution $\mathbf{P}', \mathbf{M}'$ that improves the objective function, so that the solution $\mathbf{P}^\star, \mathbf{M}^\star$ where there is at least one element of $\mathbf{M}^\star$ smaller than $M_{\max}$ is not the optimal solution.

Let us denote the set of BS indices whose number of antennas according to $\mathbf{M}^\star$ is equal to $M_{\max}$ as $\mathcal{J}_1$ and the set of other BS indices as $\mathcal{J}_2$. Consider the following feasible power allocation solution $P'_{jk} = P^\star_{jk}, \forall j \in \mathcal{J}_1$ and $P'_{li} = \zeta_l P^\star_{li}, \forall l \in \mathcal{J}_2$, where $\zeta_l = \frac{M^\star_l}{M_{\max}} < 1$. By denoting $\mathbf{M}_{\max} = [M_{\max}, \ldots, M_{\max}]^T$, let us consider the potential solution $\mathbf{P}', \mathbf{M}_{\max}$. It can be easily verified that $\mathbf{P}', \mathbf{M}_{\max}$ satisfies the constraints (C3) and (C4). Since $P'_{jk} \leq P^\star_{jk}, \forall j, k$, it also satisfies constraint (C2).

We evaluate the achieved rate of the kth user in the jth cell for $j \in \mathcal{J}_1$ as

$$\log\left[1+\gamma_{jk}(\mathbf{P}', \mathbf{M}_{\max})\right] = \log\left[1+\frac{P^\star_{jk}M_{\max}A_{jk}}{F^\star_{\mathcal{J}_1} + F'_{\mathcal{J}_2} + N_0}\right], \tag{43}$$

where $F^\star_{\mathcal{J}_1} = \sum_{l\in\mathcal{J}_1}\left(\sum_i P^\star_{li}B_{li} + P^\star_{l\tilde{\rho}_{l,\rho_{j,k}}} M_{\max} C_{li}\right)$ and $F'_{\mathcal{J}_2} = \sum_{l\in\mathcal{J}_2}\left(\sum_i P'_{li}B_{li} + P'_{l\tilde{\rho}_{l,\rho_{j,k}}} M_{\max} C_{li}\right)$. Since $P'_{li} = \zeta_l P^\star_{li}, \forall l \in \mathcal{J}_2$ and $M^\star_l = \zeta_l M_{\max}$, we have $F'_{\mathcal{J}_2} < F^\star_{\mathcal{J}_2} = \sum_{l\in\mathcal{J}_2}\left(\sum_i P^\star_{li}B_{li} + P^\star_{l\tilde{\rho}_{l,\rho_{j,k}}} M^\star_l C_{li}\right)$. Therefore, we achieve

$$\begin{aligned} &\log\left[1+\gamma_{jk}(\mathbf{P}', \mathbf{M}_{\max})\right] \\ &\quad > \log\left[1+\frac{P^\star_{jk}M_{\max}A_{jk}}{F^\star_{\mathcal{J}_1} + F^\star_{\mathcal{J}_2} + N_0}\right] \\ &\quad = \log\left[1+\gamma_{jm}(\mathbf{P}^\star, \mathbf{M}^\star)\right] \geq R_{\min}, \quad j \in \mathcal{J}_1. \end{aligned} \tag{44}$$

Similarly, the rate for kth user in jth for $j \in \mathcal{J}_2$ can be given as

$$\log\left[1+\gamma_{jk}(\mathbf{P}', \mathbf{M}_{\max})\right] = \log\left[1+\frac{P'_{jk}M_{\max}A_{jk}}{F^\star_{\mathcal{J}_1} + F'_{\mathcal{J}_2} + N_0}\right]. \tag{45}$$

Since $\zeta_l < 1$, we have $P'_{jk}M_{\max}A_{jk} = P^{\star}_{jk}M^{\star}_{l}A_{jk}$ and $F'_{\mathcal{J}_2} < F^{\star}_{\mathcal{J}_2}$. Then, we have the same behavior as

$$\log\left[1+\gamma_{jk}(\mathbf{P}', \mathbf{M}_{\max})\right] > \log\left[1+\frac{P^{\star}_{jk}M^{\star}_{j}A_{jk}}{F^{\star}_{\mathcal{J}_1}+F^{\star}_{\mathcal{J}_2}+N_0}\right] = \log\left[1+\gamma_{jk}(\mathbf{P}^{\star}, \mathbf{M}^{\star})\right] \geq R_{\min}, \quad j \in \mathcal{J}_2. \tag{46}$$

The results in (44) and (46) also mean that $\mathbf{P}', \mathbf{M}_{\max}$ satisfies constraint (C1) since it results in improved rate for each user in any cell. In addition, the SR is also increased under $\mathbf{P}', \mathbf{M}_{\max}$, which contradicts the assumption that $\mathbf{P}^{\star}, \mathbf{M}^{\star}$ is the optimal solution. Therefore, we have completed the proof.

REFERENCES

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[2] 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.

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[4] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[5] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[6] C. W. Tan, M. Chiang, and R. Srikant, "Fast algorithms and performance bounds for sum rate maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 706–719, Jun. 2013.

[7] J. Kim and D.-H. Cho, "A joint power and subchannel allocation scheme maximizing system capacity in indoor dense mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4340–4353, Nov. 2010.

[8] Y.-B. Lin, T.-H. Chiu, and Y. T. Su, "Optimal and near-optimal resource allocation algorithms for OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4066–4077, Aug. 2009.

[9] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.

[10] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 510–515, Feb. 2012.

[11] H. Al-Shatri and T. Weber, "Achieving the maximum sum rate using D.C. programming in cellular networks," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1331–1341, Mar. 2012.

[12] T. Wang and L. Vandendorpe, "On the SCALE algorithm for multiuser multicarrier power spectrum management," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4992–4998, Sep. 2012.

[13] E. Hossain, L. B. Le, and D. Niyato, *Radio Resource Management in Multi-Tier Cellular Wireless Networks*. New York, NY, USA: Wiley, 2013.

[14] V. N. Ha and L. B. Le, "Fair resource allocation for OFDMA femtocell networks with macrocell protection," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1388–1401, Mar. 2014.

[15] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 342–355, Jan. 2014.

[16] H. Zhang *et al.*, "Joint subchannel and power allocation in interference-limited OFDMA femtocells with heterogeneous QoS guarantee," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 4572–4577.

[17] J. B. Rao and A. O. Fapojuwo, "A survey of energy efficient resource management techniques for multicell cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 154–180, May 2014.

[18] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 1–14, Jan. 2015.

[19] L. B. Le, D. Niyato, E. Hossain, D. I. Kim, and D. T. Hoang, "QoS-aware and energy-efficient resource management in OFDMA femtocells," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 180–194, Jan. 2013.

[20] Z. Xu, G. Y. Li, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient power allocation for pilots in training-based downlink OFDMA systems," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3047–3058, Oct. 2012.

[21] G. Miao and J. Zhang, "On optimal energy-efficient multi-user MIMO," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–6.

[22] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.

[23] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.

[24] T. M. Nguyen and L. B. Le, "Joint pilot assignment and resource allocation in multicell massive MIMO network: Throughput and energy efficiency maximization," in *Proc. IEEE WCNC*, Mar. 2015, pp. 393–398.

[25] M. Imran *et al.*, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *EARTH Project Del. D*, vol. 2, Jan. 2012.

[26] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[27] D. Jungnickel, *Graphs, Networks and Algorithms*. Berlin, Germany: Springer-Verlag, 2008.

[28] Y. Bao and A. Ullah, "Expectation of quadratic forms in normal and nonnormal variables with econometric applications," Dept. Econ., Univ. California, Riverside, CA, USA, Working Papers, Apr. 2009.

TRI MINH NGUYEN (S'10) received the B.Eng. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2009, and the M.E. degree in electrical engineering from Kyung Hee University, Yongin, Korea, in 2012. His current research interests include interference management in heterogeneous networks, wireless cooperative localization, and massive multiple-input and multiple-output.

VU NGUYEN HA (S'11) received the B.Eng. degree from the French Training Program for Excellent Engineers, Ho Chi Minh City University of Technology, Vietnam, and the Addendum degree from the École Nationale Supérieure des Télécommunications de Bretagne–Groupe des École des Télécommunications, Bretagne, France, in 2007. He is currently pursuing the Ph.D. degree with the Institut National de la Recherche Scientifique–Énergie, Matériaux et Télécommunications, Université du Québec, Montréal, QC, Canada. From 2008 to 2011, he was a Research Assistant with the School of Electrical Engineering, University of Ulsan, Ulsan, Korea. His research interests include radio resource management and emerging enabling technologies for 5G wireless systems with a special emphasis on heterogeneous small-cell networks, cloud RAN, and multiple-input and multiple-output transmission.

LONG BAO LE (S'04–M'07–SM'12) received the B.Eng. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 1999, the M.Eng. degree in telecommunications from the Asian Institute of Technology, Thailand, in 2002, and the Ph.D. degree in electrical engineering from the University of Manitoba, Canada, in 2007. He was a Post-Doctoral Researcher with the Massachusetts Institute of Technology (2008–2010) and the University of Waterloo (2007–2008). Since 2010, he has been with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC, Canada, where he is currently an Associate Professor. He has co-authored a book entitled *Radio Resource Management in Multi-Tier Cellular Wireless Networks* (Wiley, 2013). His current research interests include smart grids, cognitive radio, radio resource management, network control and optimization, and emerging enabling technologies for 5G wireless systems. He is a member of the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served as the Technical Program Committee Chair/Co-Chair for several IEEE conferences, including the IEEE WCNC, the IEEE VTC, and the IEEE PIMRC.