Signal Processing 173 (2020) 107588

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Joint user association and power allocation for massive MIMO HetNets with imperfect CSI

Hao Li[a,b], Zhigang Wang[a,*], Houjun Wang[a]

[a] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] North Automatic Control Technology Institute, China North Industries Group Corporation, Taiyuan 030006, China

ARTICLE INFO

Article history:
Received 15 March 2019
Revised 24 December 2019
Accepted 16 March 2020
Available online 19 March 2020

Keywords:
Massive MIMO
HetNets
User association
Power allocation
Imperfect CSI

ABSTRACT

Since Massive multiple-input and multiple-output (MIMO) and heterogeneous networks (HetNets) have significantly improvement in spectrum efficiency, Massive MIMO enabled HetNets have emerged as a promising technique for the fifth-generation cellular networks. Most previous studies focus on energy-efficient resource allocation of Massive MIMO HetNets by assuming perfect channel state information (CSI). In this paper, we investigate the α-fairness network utility optimization problem of joint user association and power allocation for a downlink Massive MIMO HetNet with imperfect CSI. By utilizing zero-forcing beamforming, we show a new closed-form lower bound expression on the ergodic achievable rate. Furthermore, we formulate the optimization problem as a mixed-integer nonlinear programming problem, which is non-convex and NP-hard, to achieve the maximization of α-fairness network utility. Consequently, a joint iterative algorithm with respect to user association and power allocation is developed by decomposing the original problem and relaxing the constraints. Simulation results show that our proposed algorithm can yield much better network utility performance than the other algorithms.

1. Introduction

Due to the explosive growth in wireless data requirement and the number of wireless terminals, the fifth-generation (5G) cellular networks have gained substantial attention [31]. Also, 5G cellular networks are expected to improve spectrum efficiency and system capacity. Massive multiple-input and multiple-output (MIMO) and dense heterogeneous networks (HetNets) are considered as the candidate technology for the future 5G cellular networks, which can significantly enhance the spectrum efficiency [1–4]. By serving multiple users, Massive MIMO technology, where the base station (BS) is employing a large number of antennas, can achieve enormous improvement for spectrum efficiency. Meanwhile, in a HetNet, the small cells are deployed within a macrocell coverage, and the small cells can offload data traffic [4]; therefore, HetNets, comprised of massive MIMO base stations with different powers, numbers of antennas and multiplexing gain capabilities, will increase the system spectrum efficiency. One of the indispensable characteristic of HetNet is that a cell edge user can potentially be associated with either the macrocell BS or a small cell BS, resulting in different performance. Therefore, it is important to establish mechanisms for associating users to BSs in the dense HetNets, so that the available wireless infrastructure can be efficiently used. Since both Massive MIMO and HetNets simultaneously have superior performance benefits, studying Massive MIMO enabled HetNets has been considered as a key issue.

1.1. Related works and motivations

Recently, some existing papers have investigated the resource allocation problem of Massive MIMO systems [5–8]. For example, the authors in [5] focused on the energy efficiency optimization problem for a downlink single-cell Massive MIMO system under perfect channel state information (CSI), and a joint antenna selection and power allocation iterative algorithm was developed. In [6], the authors jointly optimized the power allocation and user association in multi-cell Massive MIMO downlink systems with imperfect CSI, and their goal was to minimize the transmit power consumption while maintaining the quality-of-service (QoS) constraints. In order to find appropriate QoS targets, [6] formulated a max-min fairness problem that maximizes the worst spectral efficiency among the users. However, they neglected the backhaul capacity constraint and the load constraint. Additionally, they do not studied the proportional fairness problem among the users. Various energy-efficient resource allocation in HetNets have been reported in [9–11]. The authors in [9] designed a user association

* Corresponding author.
E-mail addresses: haoli_uestc@163.com (H. Li), wangzhigang@uestc.edu.cn (Z. Wang), hjwang@uestc.edu.cn (H. Wang).

https://doi.org/10.1016/j.sigpro.2020.107588
0165-1684/

algorithm with lower-complexity to maximize the logarithmic utility. The authors of [10] proposed an iterative joint user association and power allocation algorithm to maximize the energy efficiency of HetNets. To maximize sum logarithmic user rate, a joint optimization scheme of cell association and bandwidth allocation in HetNets was studied in [11]. However, the above works investigated the resource allocation in HetNets by assuming perfect CSI.

A Massive MIMO enabled HetNet is potentially capable of improving the system capacity, compared to Massive MIMO systems or HetNets. Related works on energy-efficient resource allocation schemes in Massive MIMO enabled HetNets are briefly given below. In [12], the authors proposed an energy-efficient user association algorithm to maximize the network logarithmic utility in Massive MIMO enabled HetNets. To study the tradeoff between energy efficiency and spectral efficiency with proportional rate fairness, a joint user association and power coordination optimization problem was considered in [13] for Massive MIMO enabled HetNets. The authors in [14] constructed a nonconvex integer programming problem in Massive MIMO enabled HetNets, and developed a joint user association and resource allocation algorithm. However, the main distinction is that these aforementioned Massive MIMO HetNets contributions considered the case that all BSs and/or users know the perfect CSI, while in this paper we assume that all BSs and users simultaneously have imperfect CSI knowledge. By considering the case of imperfect CSI, the authors in [15] derived a closed-form expression for the spectral efficiency and designed a joint scheme of user association and power allocation with proportional fairness in Massive MIMO HetNets. Different from the previous works which considered the proportional fairness and/or the max-min fairness separately, the motivation of this paper is to study a general solution of α-fairness optimization problem under imperfect CSI.[1] Note that, in this paper, α-fairness problem includes the proportional fairness and the max-min fairness.

1.2. Contributions

To achieve the maximization of α-fairness network utility, we focus on the joint optimization problem of user association and power allocation in a two-tier downlink Massive MIMO HetNet with imperfect CSI. To the best of the authors' knowledge, this problem has not been studied in the existing literature. The main contributions of the paper are summarized as follows.

- We construct the two-tier downlink model of a Massive MIMO HetNet by assuming imperfect CSI of BSs and users. Using linear zero-forcing (ZF) beamforming, a new closed-form lower bound expression for the ergodic achievable rate is derived.
- We formulate a joint optimization problem of user association and power allocation to maximize the α-fairness network utility, under QoS requirements, backhaul capacity and load constraints. This problem is a non-convex mixed-integer nonlinear programming (MINLP) problem.
- We decompose the original non-convex problem into user association subproblem and power allocation subproblem, and develop a joint iterative sub-optimal algorithm. Specifically, based on Lagrangian dual decomposition (LDD) method, the optimal user association solution can be obtained; with a given user association, a gradient ascent method is used to find the solution of power allocation.
- We demonstrate the correctness of the derived closed-form expression, and the effectiveness of our proposed algorithm by extensive experiments. Numerical results provide some insights into the effect of the number of transmit antennas at MBS on the network utility, and show that the proposed algorithm achieves better network utility performance than the existing schemes.

[1] Fairness is an important consideration when designing wireless communication systems to provide uniformly great service for everyone. Furthermore, Fairness problem can also be applied the cooperative multiantenna relaying systems [16,17].

1.3. Paper organization

The rest of this paper is organized as follows. In Section 2, we construct the downlink system model of a Massive MIMO HetNet and formulate an optimization problem. Section 3 presents the joint algorithm design. Simulation results are shown in Section 4. Finally, we draw conclusions in Section 5.

1.4. Notation

The following notation is adopted throughout the paper. We use lowercase and uppercase boldface letters to denote the vectors and matrices, respectively. The transpose, the conjugate (Hermitian) transpose and the inverse are denoted by $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$, respectively. $\mathbb{E}\{\cdot\}$ is the expectation operation; tr $(\cdot)$ is the trace operation; $|\cdot|$ is the absolute value of a complex-valued scalar; $\|\cdot\|$ is the (Euclidean) vector 2-norm.

2. System model and formulation problem

2.1. System model

We establish a two-tier downlink HetNet composed of a Massive MIMO macro BS (MBS) and multiple pico BSs (PBSs). There are K single-antenna user terminals (UTs) that are uniformly distributed in this two-tier network. Without loss of generality, we denote the set of PBSs by $\mathcal{B}_p = \{1, \ldots, J\}$ and the set of all BSs by $\mathcal{B} = \mathcal{B}_p \cup \{0\}$, where the index 0 is introduced to indicate the MBS. We define M_j as the number of transmit antennas at the jth BS. We denote $k \in \mathcal{U}$ as the index for each UT, where $\mathcal{U} = \{1, \ldots, K\}$ is the set of UTs.

We assume that the whole system is operating in time-division duplexing (TDD) mode, this means that the uplink channel gain and the downlink channel gain are assumed to be same. Since the path loss and shadow fading are slowly varying under a block fading channel, the large-scale fading coefficients can be estimated perfectly by the BS [18]. For a given resource block (RB), let $S_j (1 \le S_j \le M_j)$ be the maximum number of downlink data streams that BS j can transmit simultaneously. For analytical tractability, we assume that each BS can connect to more than one UT and the maximum number of UTs is S_j, while each UT is associated with one BS at a time[2] Thus, the received signal of the kth UT associated with the jth BS can be formulated as

$$y_{j,k} = \sqrt{\frac{P_j}{S_j}} \mathbf{g}_{j,k}^H \mathbf{a}_{j,k} q_{j,k} + \sum_{j' \in \mathcal{B} \setminus j} \sum_{k' \in S_{j'}} \sqrt{\frac{P_{j'}}{S_{j'}}} \mathbf{g}_{j',k}^H \mathbf{a}_{j',k'} q_{j',k'} + n_{j,k} \tag{1}$$

where P_j is the transmit power from the jth BS, and $j = 0$ denotes the transmit power from MBS; $\mathbf{g}_{j,k} \in \mathbb{C}^{M_j \times 1}$ denotes the downlink channel gain between the jth BS and the kth UT; $\mathbf{a}_{j,k} \in \mathbb{C}^{M_j \times 1}$ denotes the precoding beamforming vector; $q_{j,k}$ is the transmitted information symbol from BS j to UT k, and it satisfies $\mathbb{E}\{|q_{j,k}|^2\} = 1$; the noise term $n_{j,k}$ is a zero-mean complex additive white Gaussian noise (AWGN) with unit variance.

[2] For simplicity, we assume that each UT is served by one BS through a subchannel with unit bandwidth.

In the Massive MIMO HetNets, obtaining the perfect knowledge of CSI is a challenging for both BSs and UTs. In our work, we investigate the problem of joint user association and power allocation by assuming that all BSs and UTs have imperfect CSI. In (1), the second term is the interference signal of the kth UT associated with the jth BS from remaining $k-1$ UTs, and this interference signal increases the difficulty of the desired signal extraction for the receiver side. Here, we assume that each user detects its desired signal from the received signal of (1) by using the successive interference cancellation technique [19]. However, for any user, the current precoded channel information is unknown in the downlink payload data transmission. In this case, we will utilize the knowledge of the channel statistics to approximate the exact precoded channel. More precisely, the realistic channel in signal detection can be treated as the mean of the channel gain, which is the consequence of the law of large numbers [20]. Therefore, when all UTs have imperfect CSI, we can decompose (1) as [21]

$$y_{j,k} = \sqrt{\frac{P_j}{S_j}}\mathbb{E}\left\{\mathbf{g}_{j,k}^H\mathbf{a}_{j,k}\right\}q_{j,k} + \sqrt{\frac{P_j}{S_j}}\left(\mathbf{g}_{j,k}^H\mathbf{a}_{j,k} - \mathbb{E}\left\{\mathbf{g}_{j,k}^H\mathbf{a}_{j,k}\right\}\right)q_{j,k} + \sum_{j'\in\mathcal{B}\setminus j}\sum_{k'\in S_{j'}}\sqrt{\frac{P_{j'}}{S_{j'}}}\mathbf{g}_{j',k}^H\mathbf{a}_{j',k'}q_{j',k'} + n_{j,k} \tag{2}$$

where the first term is the received desired signal over the deterministic average precoded channel from BS j, while the remaining three terms can be treated as the system interference in the signal extraction.

2.2. Channel model

As shown in [2], the channel gain vector can be modeled as $\mathbf{g}_{j,k} = \sqrt{\beta_{j,k}}\mathbf{h}_{j,k}$, where $\beta_{j,k}$ and $\mathbf{h}_{j,k} \in \mathbb{C}^{M_j\times 1}$ are the large-scale fading and small-scale fading between the jth BS to the kth UT, respectively. Note that $\beta_{j,k}$ is approximated as a constant; $\mathbf{h}_{j,k}$ is also known as Rayleigh fading and therefore we know that the elements of $\mathbf{h}_{j,k}$ are independent and identically distributed (i.i.d) random variables with zero mean and unit variance. With the assumption of the imperfect CSI at BS, the channel gains have to be estimated. According to [2,6], we know that a standard way of channel estimation is to employ uplink pilots. Specifically, all users simultaneously transmit $\tau = K$ mutually orthogonal pilot sequences to the BS in a coherence interval of length T, then the BS exploit these pilots to estimate the uplink channel matrix. This process is recognized as the *uplink training phase*. Since the system operates in TDD mode, the uplink channel and the downlink channel are assumed to be reciprocal. Based on reciprocity, the BS can utilize the estimated downlink channel response to achieve the downlink data transmission in the remaining coherence interval of length $T-\tau$. This process is known as the *data transmission phase*. However, different from [2,6], we assume the large-scale fading coefficient is known and only the small-scale fading channel is estimated at the BS. We mainly focus on the effects of estimation error on downlink transmission, and therefore we omit the uplink channel estimation process. Note that, a similar channel estimation model has been shown in [15,22]. Based on the minimum mean square error (MMSE) channel estimation error model [23,24], the Rayleigh fading coefficients can be expressed as

$$\mathbf{h}_{j,k} = \hat{\mathbf{h}}_{j,k} + \mathbf{e}_{j,k} \tag{3}$$

where $\mathbf{h}_{j,k}$ denotes the actual Rayleigh fading channel vector; $\hat{\mathbf{h}}_{j,k}$ is the estimated channel vector; $\mathbf{e}_{j,k}$ is the estimated error vector, and its elements follow complex Gaussian distribution with mean zero and the variance σ_e^2. Accordingly, we can know that the elements of $\hat{\mathbf{h}}_{j,k}$ can be distributed as $\mathcal{CN}(0, 1-\sigma_e^2)$. With the orthogonality property of the MMSE estimation, $\hat{\mathbf{h}}_{j,k}$ and $\mathbf{e}_{j,k}$ are uncorrelated. Further, we can obtain the estimated downlink channel matrix of jth BS as $\hat{\mathbf{G}}_j = [\hat{\mathbf{g}}_{j,1}, \ldots, \hat{\mathbf{g}}_{j,K}]$, where the kth column of $\hat{\mathbf{G}}_j$ is denoted by $\hat{\mathbf{g}}_{j,k} = \sqrt{\beta_{j,k}}\hat{\mathbf{h}}_{j,k}$. Denote $\hat{\mathbf{H}}_j$ as the estimated Rayleigh fading channel matrix.

2.3. Ergodic achievable rate

Owing to the superior performance of linear ZF beamforming, we employ the lower-complexity ZF to precode payload data before the downlink transmission of each BS, so as to eliminate the intra-cell interference. With the normalized ZF, we have $\mathbf{a}_{j,k} = \frac{\mathbf{g}_{j,k}^{\dagger}}{\sqrt{\mathbb{E}\{\|\mathbf{g}_{j,k}^{\dagger}\|^2\}}}$, where $\mathbf{g}_{j,k}^{\dagger}$ is the kth column of ZF precoding matrix $\mathbf{G}_j^{\dagger} = \hat{\mathbf{G}}_j(\hat{\mathbf{G}}_j^H\hat{\mathbf{G}}_j)^{-1}$.

Similar to [21], we can obtain the following lower bound of the ergodic achievable rate, for the kth UT associated with the jth BS, as

$$R_{j,k} = a_{j,k}\log_2\left(1+\Gamma_{j,k}\right) \tag{4}$$

where $a_{j,k}(0 \le a_{j,k} \le 1)$ is the proportion of RBs that BS j uses for downlink payload data transmission to UT k; and the signal-to-interference-plus-noise ratio (SINR) is

$$\begin{aligned}\Gamma_{j,k} &= \frac{\frac{P_j}{S_j}\left|\mathbb{E}\left\{\hat{\mathbf{g}}_{j,k}^H\mathbf{a}_{j,k}\right\}\right|^2}{\frac{P_j}{S_j}\left(\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j,k}^H\mathbf{a}_{j,k}\right|^2\right\} - \left|\mathbb{E}\left\{\hat{\mathbf{g}}_{j,k}^H\mathbf{a}_{j,k}\right\}\right|^2\right) + \sum_{j'\in\mathcal{B}\setminus j}\sum_{k'\in S_{j'}}\frac{P_{j'}}{S_{j'}}\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j',k}^H\mathbf{a}_{j',k'}\right|^2\right\}+1} \\ &\overset{(a)}{=} \frac{\frac{P_j}{S_j}\frac{\left|\mathbb{E}\left\{\hat{\mathbf{g}}_{j,k}^H\mathbf{g}_{j,k}^{\dagger}\right\}\right|^2}{\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\}}}{\frac{P_j}{S_j}\frac{\left(\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j,k}^H\mathbf{g}_{j,k}^{\dagger}\right|^2\right\}-\left|\mathbb{E}\left\{\hat{\mathbf{g}}_{j,k}^H\mathbf{g}_{j,k}^{\dagger}\right\}\right|^2\right)}{\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\}} + \sum_{j'\in\mathcal{B}\setminus j}\sum_{k'\in S_{j'}}\frac{P_{j'}}{S_{j'}}\frac{\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right|^2\right\}}{\mathbb{E}\left\{\left\|\mathbf{g}_{j',k'}^{\dagger}\right\|^2\right\}}+1} \\ &\overset{(b)}{=} \frac{\frac{P_j}{S_j}}{\sum_{j'\in\mathcal{B}\setminus j}\sum_{k'\in S_{j'}}\frac{P_{j'}}{S_{j'}}\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right|^2\right\}+\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\}}\end{aligned} \tag{5}$$

where (a) is obtained by letting $\mathbf{a}_{j,k} = \frac{\mathbf{g}_{j,k}^{\dagger}}{\sqrt{\mathbb{E}\{\|\mathbf{g}_{j,k}^{\dagger}\|^2\}}}$; (b) is because $\mathbf{G}_j^{\dagger} = \hat{\mathbf{G}}_j(\hat{\mathbf{G}}_j^H\hat{\mathbf{G}}_j)^{-1}$, i.e., $\hat{\mathbf{g}}_{j,k}^H\mathbf{g}_{j,k}^{\dagger} = 1$ and $\hat{\mathbf{g}}_{j,k}^H\mathbf{g}_{j,i}^{\dagger} = 0$, $\forall j \in \mathcal{B}$, $1 \le k \le S_j$, $k \ne i$, and $\|\mathbf{g}_{j,k}^{\dagger}\|^2 = \|\mathbf{g}_{j',k'}^{\dagger}\|^2$.

We next state a Proposition 1, which derive an analytical closed-form expression for the proposed lower bound of the ergodic achievable rate.[3]

Proposition 1. *With ZF, $M_j > S_j$, an exact closed-form lower bound on the achievable rate of the kth UT associated with the jth BS can be expressed as*

$$R_{j,k} = a_{j,k}\log_2\left(1+\frac{P_j\beta_{j,k}\left(M_j-S_j\right)\left(1-\sigma_e^2\right)}{S_j\left(\sum_{j'\in\mathcal{B}\setminus j}P_{j'}\beta_{j',k}\left(1-\sigma_e^2\right)+1\right)}\right). \tag{6}$$

Proof. See Appendix A. □

[3] Different from [6,25], we consider total transmit power of each BS as optimization variables, then the transmit power of each BS is equally allocated to all its associating UTs.

2.4. Formulation problem

Based on the regime that each UT can only be associated with one BS, we introduce a matrix $\mathbf{X} = [x_{j,k}]$ to describe the user allocation indicator, where $x_{j,k}$ is the binary index variable, which satisfies : $x_{j,k} = 1$ if and only if UT k is associated with BS j, and $x_{j,k} = 0$ otherwise. Accordingly, each user throughput from the jth BS to the kth UT is given by

$$r_{j,k} = x_{j,k} R_{j,k}. \tag{7}$$

In this paper, our aim is to achieve the maximization of network utility by jointly optimizing user association and power allocation for a downlink Massive MIMO HetNet network with imperfect CSI. We define the α-fairness utility function of the system as [26]

$$U_\alpha(x) = \begin{cases} \log x & \alpha = 1 \\ \dfrac{x^{1-\alpha}}{1-\alpha} & \alpha > 0, \alpha \neq 1 \end{cases} \tag{8}$$

where α denotes a fairness parameter, which will determine the level of fairness among UTs. Specifically, when $\alpha = 1$, the utility function is the logarithm of the objective variable (which is known as proportional fairness); when $\alpha \to \infty$, the utility function is the worst-case of the objective variable (which is known as max-min fairness). Therefore, the optimization problem, which is to maximize the total network utility with respect to $r_{j,k}$, can be formulated as

$$\max_{\mathbf{X},\mathbf{P}} \sum_{k\in\mathcal{U}} \sum_{j\in\mathcal{B}} U_\alpha\left(r_{j,k}\right) \tag{9a}$$

$$s.t. 0 \le r_{j,k} \le R_{j,k}, \quad \forall\ j \in \mathcal{B}, k \in \mathcal{U} \tag{9b}$$

$$\sum_{k\in\mathcal{U}} R_{j,k} \le C_j, \quad \forall\ j \in \mathcal{B} \tag{9c}$$

$$0 \le P_j \le P_{max,j}, \quad \forall\ j \in \mathcal{B} \tag{9d}$$

$$\sum_{k\in\mathcal{U}} x_{j,k} \le S_j, \quad \forall\ j \in \mathcal{B} \tag{9e}$$

$$\sum_{j\in\mathcal{B}} x_{j,k} = 1, \quad \forall\ k \in \mathcal{U} \tag{9f}$$

$$x_{j,k} = \{0, 1\}, \quad \forall\ j \in \mathcal{B}, k \in \mathcal{U} \tag{9g}$$

where $\mathbf{P} = [P_1, \ldots, P_{J+1}]$ is the transmit power vector of all BSs; (9c) is the backhaul constraint, and where C_j denotes the maximum backhaul capacity of BS j; (9d) is the transmit power constraint, and where $P_{max,j}$ is the maximum transmit power of BS j; (9e) is the load constraint of BS j; (9b), (9f) and (9g) illustrate that each UT is only associate with one BS.

3. Solution to the optimization problem

The maximization problem in (9) is a non-convex MINLP due to the involvement of the binary variable $x_{j,k}$. Theoretically, the global optimal solution for user association and power allocation is obtained by using the exhaustive search method. However, this method has huge computational complexity in practice. Therefore, to solve the proposed non-convex problem efficiently, we design a sub-optimal optimization algorithm by employing LDD [27]. By introducing an auxiliary constraint, we first transform the original problem (9) as

$$\max_{\mathbf{X},\mathbf{P}} \sum_{k\in\mathcal{U}} \sum_{j\in\mathcal{B}} U_\alpha\left(r_{j,k}\right) \tag{10a}$$

$$s.t. \quad (9b) - (9g)$$
$$0 \le a_{j,k} \le x_{j,k}, \quad \forall\ j \in \mathcal{B}, k \in \mathcal{U}. \tag{10b}$$

In (10), the constraints (9b)-(9g) will not be tightened by introducing constraint (10b), this is because $a_{j,k} \le x_{j,k} = 1$ if UT k is associated with BS j, while $a_{j,k} = x_{j,k} = 0$ if UT k is not associated with BS j. It is worth noting that, for parameter $a_{j,k}$ at BS j, we will adopt the equal allocation scheme [9], i.e., $a_{j,k} = \frac{1}{\sum_{k\in\mathcal{U}} x_{j,k}}$.

3.1. User association

We now study the user association subproblem under a given transmit power $\mathbf{P}$. Based on the LDD approach, we can relax the coupled constraint (10b) by introducing the non-negative dual variables $\boldsymbol{\psi} = [\psi_{j,k}]$. Therefore, we can obtain a partial Lagrangian dual problem associated with (10) as

$$\min_{\boldsymbol{\psi}\ge 0} D\left(\boldsymbol{\psi}\right) = F\left(\boldsymbol{\psi}\right) + G\left(\boldsymbol{\psi}\right) \tag{11}$$

where

$$F\left(\boldsymbol{\psi}\right) = \begin{cases} \max\limits_{\mathbf{P}} & \sum_{k\in\mathcal{U}} \sum_{j\in\mathcal{B}} \left(U_\alpha\left(r_{j,k}\right) - \psi_{j,k} a_{j,k}\right) \\ s.t. & (9b) - (9d) \\ & a_{j,k} \ge 0 \end{cases} \tag{12}$$

and

$$G\left(\boldsymbol{\psi}\right) = \begin{cases} \max\limits_{\mathbf{X}} & \sum_{k\in\mathcal{U}} \sum_{j\in\mathcal{B}} \psi_{j,k} x_{j,k} \\ s.t. & (9e) - (9g). \end{cases} \tag{13}$$

We can know from (13) that, when the optimal value $\boldsymbol{\psi}^*$ of (11) is obtained, a local optimal user association solution can be expressed as

$$x_{j,k}^* = \begin{cases} 1 & \text{if} \quad j = \arg\max\limits_{j\in\mathbf{B}} \ \psi_{j,k}^* \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

In general, we can use a sub-gradient method to alternatively update the dual variables [28], so as to obtain an optimal $\boldsymbol{\psi}^*$ of dual problem (11). However, in each update iteration, we have to compute the sub-gradient of $D(\boldsymbol{\psi})$, as well solve the constrained optimization problem (12) and (13). Thus, the computational cost of the above method is usually high. Here, we provide an efficient method to find the optimal solution of problem (11) indirectly, as shown in Theorem 1.

Theorem 1. *The optimal solution* $\{\boldsymbol{\varpi}^*, \boldsymbol{\beta}^*, \boldsymbol{\varphi}^*, \boldsymbol{\eta}^*, \boldsymbol{\psi}^*\}$ *of the convex optimization problem (15) is such that* $\boldsymbol{\psi}^*$ *is the optimal solution of problem (11).*

$$\begin{aligned} \min_{\{\boldsymbol{\varpi},\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\eta},\boldsymbol{\psi}\}\succ \mathbf{0}} & \sum_{k\in\mathcal{U}} \sum_{j\in\mathcal{B}} V_\rho\left(\varpi_{j,k}\right) + \sum_{j\in\mathcal{B}} \beta_j C_j + \sum_{j\in\mathcal{B}} \varphi_j S_j + \sum_{k\in\mathcal{U}} \eta_k \\ s.t. \quad & \left(\varpi_{j,k} - \beta_j\right)\bar{R}_{j,k} - \psi_{j,k} \le 0, \quad \forall\ j \in \mathcal{B}, k \in \mathcal{U} \\ & \psi_{j,k} - \varphi_j - \eta_k \le 0, \quad \forall\ j \in \mathcal{B}, k \in \mathcal{U} \end{aligned} \tag{15}$$

where $\{\boldsymbol{\varpi}, \boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\eta}, \boldsymbol{\psi}\}\succ\mathbf{0}$ denotes all elements of $\boldsymbol{\varpi}, \boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\eta}, \boldsymbol{\psi}$ are greater than zero, respectively, i.e., $\varpi_{j,k} > 0$, $\beta_j > 0$, $\varphi_j > 0$, $\eta_k > 0$, $\psi_{j,k} > 0$, $\forall\ j \in \mathcal{B},\ k \in \mathcal{U}$; $\bar{R}_{j,k} = R_{j,k}/a_{j,k}$; the function $V_\rho(x)$ is defined as

$$V_\rho(x) = \begin{cases} -\log(x) - 1 & \rho = 1 \\ \dfrac{x^{1-\rho}}{\rho - 1} & \rho > 0, \rho \neq 1 \end{cases} \tag{16}$$

and where $\rho = 1/\alpha$.

Proof. See Appendix B. □

In (16), the objective function is differentiable and all of constraints are linear, thereby, the optimal set $\{\boldsymbol{\varpi}^*, \boldsymbol{\beta}^*, \boldsymbol{\varphi}^*, \boldsymbol{\eta}^*, \boldsymbol{\psi}^*\}$ can be solved by using the CVX tool [29]. Once $\boldsymbol{\psi}^* = [\psi^*_{j,k}]$ is known, the user association matrix $\mathbf{X}^* = [x^*_{j,k}]$ is obtained according to (14), which can be used for power allocation strategy in next subsection.

3.2. Power allocation

With given $\mathbf{X}^* = [x^*_{j,k}]$, the optimization subproblem on power allocation can be formulated as

$$\begin{aligned} &\max_{\mathbf{P}} \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} U_\alpha\left(r_{j,k}\right) \\ &\quad s.t. \quad \sum_{k\in\mathcal{U}} R_{j,k} \le C_j, \quad \forall\ j\in\mathcal{B} \\ &\quad 0 \le P_j \le P_{max,j}, \quad \forall\ j\in\mathcal{B}. \end{aligned} \tag{17}$$

The direct solution to problem (17) is complicated, because the objective function has two-layer logarithmic function when $\alpha = 1$, or it will be the product of multiple logarithmic functions when $\alpha > 0$, $\alpha \ne 1$. By utilizing the convexity of the function $R_{j,k}$ in P_j, an equivalent optimization problem of (17) is expressed as

$$\begin{aligned} &\max_{\mathbf{R}} \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} U_\alpha\left(x^*_{j,k} R_{j,k}\right) \\ &s.t. \quad \sum_{k\in\mathcal{U}} R_{j,k} \le C_j, \quad \forall\ j\in\mathcal{B} \\ &\quad 0 \le \frac{2^{R_{j,k}/a_{j,k}} - 1}{\Delta_{j,k}} \le P_{max,j}, \quad \forall\ j\in\mathcal{B} \end{aligned} \tag{18}$$

where $\Delta_{j,k} = \frac{\beta_{j,k}(M_j - S_j)(1-\sigma_e^2)}{S_j\left(\sum_{j'\in\mathcal{B}\backslash j} P_{j'}\beta_{j',k}(1-\sigma_e^2)+1\right)}$, and $\mathbf{R} = [R_{j,k}]$. Since (18) is a convex optimization problem in $R_{j,k}$, we can obtain the optimal solution $\mathbf{R}^* = [R^*_{j,k}]$. After some algebraic manipulation, the optimal power allocation $\mathbf{P}^* = [P^*_j]$ to problem (17) can be obtained. In this work, we employ the gradient assisted binary search (GABS) method to solve problem (18) [30], as shown in Algorithm 1. We denote $U(\mathbf{R}) = \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} U_\alpha(x^*_{j,k}R_{j,k})$.

Algorithm 1 GABS-based power allocation algorithm.

Initialization: $\mathbf{R}_0$, maximum tolerance ε, iteration number $t = 0$ and the maximum number of the iterations $t_{\max}$.
repeat
  Use the GABS to find the optimal step size $\delta^{(t+1)}$,
  $\mathbf{R}^{(t+1)} = \left[\mathbf{R}^{(t)} + \delta^{(t+1)}\nabla U\left(\mathbf{R}^{(t)}\right)\right]^+$.
  if $U\left(\mathbf{R}^{(t+1)}\right) - U\left(\mathbf{R}^{(t)}\right) < \varepsilon$ **then**
    $\mathbf{R}^* = \mathbf{R}^{(t+1)}$.
  else
    $\mathbf{R}^{(t)} = \mathbf{R}^{(t+1)}$, $t = t+1$.
  end if
until $t \le t_{\max}$
Compute the optimal power allocation $P^*_j = \dfrac{2^{R^*_{j,k}/a_{j,k}} - 1}{\Delta_{j,k}}$, $\forall\ j \in \mathcal{B}$.

3.3. Joint algorithm development

Based on the user association and power allocation analyzed above, we develop a feasible joint two-layer iterative algorithm as shown in Algorithm 2. For the inner layer, we compute the optimal user association $\mathbf{X}^*$ with a given transmit power. Once the inner iteration achieves convergence, we use Algorithm 1 to obtain the optimal power allocation $\mathbf{P}^*$ in the outer layer. Thus, when outer iteration achieves convergence, the optimal solution of joint user association and power allocation can be obtained.

Algorithm 2 Joint two-layer iterative algorithm.

Initialization: a feasible transmit power $\mathbf{P}_0$, iteration number $t = 0$ and the maximum number of the iterations $t_{\max}$.
repeat
  for all $j\in\mathcal{B}, k\in\mathcal{U}$ **do**
    Solve problem (15) to find the optimal $\boldsymbol{\psi}^* = \left[\psi^*_{j,k}\right]$ with fixed $\mathbf{P}$.
  end for
  Calculate $\mathbf{X}^* = \left[x^*_{j,k}\right]$ according to (14).
  With specific $\mathbf{X}^* = \left[x^*_{j,k}\right]$, use Algorithm 1 to find optimal transmit power vector $\mathbf{P}^* = \left[P^*_j\right]$.
until $t \le t_{\max}$
return $\mathbf{X}^*$, $\mathbf{P}^*$.

In Algorithm 2, the computational complexity is mainly determined by solving problem (15) and exploiting Algorithm 1. We define the number of inner and outer iterations is t_1 and t_2, respectively. Therefore, the overall computational complexity of the proposed joint iterative algorithm is $\mathcal{O}(t_1((3J+5)K)^3 + t_2(J+1)(2K+1))$. As a comparison, we consider the complexity of global optimal scheme using exhaustive search, which is $\mathcal{O}right)$, where T is the cardinal number of power set. Since t_1 and t_2 are finite in the proposed joint algorithm, the complexity of Algorithm 2 is less than the optimal exhaustive search.

4. Simulation results

In this section, simulation results are provided to assess the performance and effectiveness of the proposed joint resource allocation algorithm for the Massive MIMO HetNet system. In our simulations, we consider one MBS located in the center of cell coverage area with radius of 500 m, and J PBSs are uniformly distributed in the macrocell. The minimum distance between MBS and PBS is 50 m, and the minimum distance between BS and UT is 40 m. We set the number of PBSs as $J = 10$ and the number of UTs as $K = 20$. Based on the 3GPP LTE standard [6], the large-scale fading coefficients between MBS and UT is modeled as $-128.1 - 37.6\log_{10} d + z_{j,k}$, and the large-scale fading coefficients between PBS and UT is modeled as $-148.1 - 36.7\log_{10} d + z_{j,k}$, where d is the distance between each BS and UT in kilometers; $z_{j,k} \sim \mathcal{CN}(0,\sigma^2)$ is a log-normal Gaussian distribution with standard deviation $\sigma = 8$ dB. The maximum transmit powers for MBS and PBS are set to $P_{max,0} = 40$ dBm and $P_{max,j} = 30$ dBm, $\forall j\in\mathcal{B}_p$, respectively. The maximum number of downlink data streams is $S_j = 4$, $\forall j\in\mathcal{B}$. The maximum backhaul capacity of BS j is 50 Mbps, $\forall j\in\mathcal{B}$.

In order to demonstrate the tightness of the derived closed-form lower bound on the achievable rate, Fig. 1 shows the achievable rate with different variances of estimated error versus the number of transmit antennas at MBS. In this simulation, we compare the proposed closed-form lower bound of the achievable rate with the original ergodic achievable rate in (4), with the closed-form expression in [13]. In this figure, we set $P_0 = 20$ dBm and $P_j = 10$ dBm, $\forall j\in\mathcal{B}_p$. The proportion of RB at MBS is assumed to be $a_{0,k} = 0.25$. It is shown that the achievable rate of the kth UT increases as the number of transmit antennas at MBS grows. From Fig. 1, we can observe that our proposed closed-form expression for the achievable rate is tight. Additionally, we see that there is

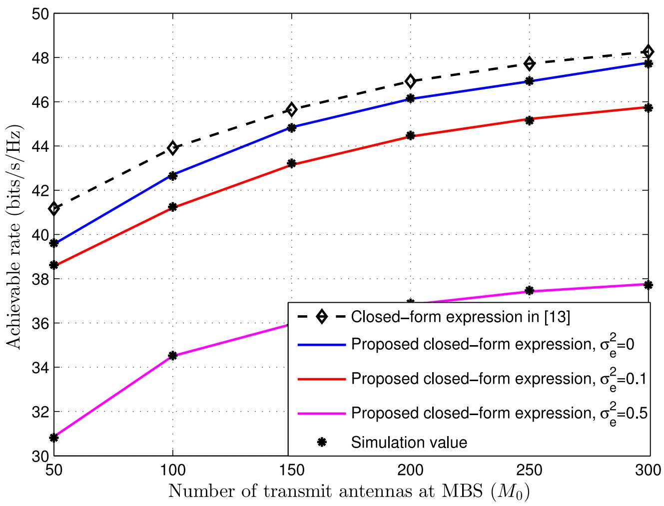

Fig. 1. Achievable rate versus the number of transmit antennas at MBS.

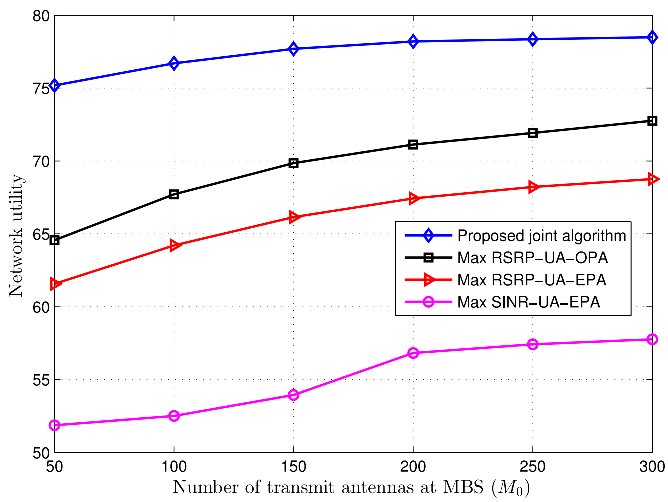

Fig. 2. Network utility with $\alpha = 1$ versus the number of transmit antennas at MBS.

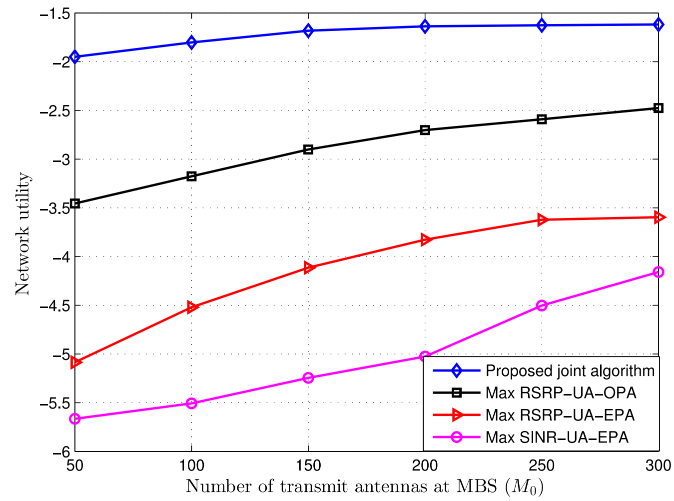

Fig. 3. Network utility with $\alpha = 2$ versus the number of transmit antennas at MBS.

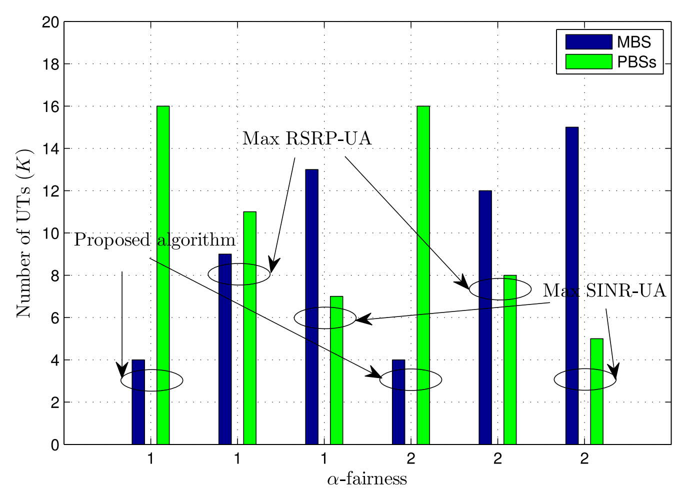

Fig. 4. Number of UTs versus α-fairness parameter for different algorithms.

a gap between the proposed closed-form expression and the existing closed-form expression in [13]. That is because the proposed closed-form expression is derived under imperfect CSI, while the existing closed-form expression is derived under perfect CSI. In other words, these gaps are caused by the estimation error at BSs and the approximation at UTs.

In Fig. 2, the total network utility is evaluated versus the number of transmit antennas at MBS. We set fairness parameter as $\alpha = 1$, and the variance of estimated error is 0.1. We assume that each PBS has the same number of transmit antennas, i.e., $M_j = 30$, $\forall j \in \mathcal{B}_p$. Here, we provide the performance improvement of the proposed joint iterative algorithm in comparison with several reference algorithms, i.e., max RSRP-UA-OPA, max RSRP-UA-EPA, max SINR-UA-EPA.[4] In Fig. 2, the network utility of the system for all schemes simultaneously increases as the number of transmit antennas at MBS, but the growth rate of the curve becomes slower. As observed from Fig. 2, our proposed joint algorithm can achieve higher network utility than the max RSRP-UA-OPA and the max RSRP-UA-EPA, as well as the max SINR-UA-EPA. For example, when the number of transmit antennas at MBS is 150, the network utility of the proposed joint algorithm is about 11.2% more than that of the max RSRP-UA-OPA, is about 17.4% more than that of the max RSRP-UA-EPA, and is about 44.1% more than that of the max SINR-UA-EPA.

Different from the simulation conditions of Figs. 2, 3 shows the network utility with $\alpha = 2$ versus the number of transmit antennas at MBS. It can be observed that the total network utility still increases with the number of transmit antennas at MBS. We can see clearly from Fig. 3 that the performance of the proposed joint algorithm is much better than the existing reference algorithms in terms of total network utility. Among all the algorithms studied, max SINR-UA-EPA scheme performs the worst. For example, when the number of transmit antennas at MBS is 150, the network utility of the proposed joint algorithm is about 41.6% more than that of the max RSRP-UA-OPA, is about 56.1% more than that of the max RSRP-UA-EPA, and is about 63.9% more than that of the max SINR-UA-EPA.

Corresponding to Figs. 2 and 3, 4 plots the load distribution of UTs under the case of $M_0 = 100$, for $\alpha = 1$ and $\alpha = 2$. We can see that, regardless of the value of α, the number of UTs associated with MBS for the proposed algorithm is the least compared to the other algorithms, which will lead to a better load balancing. This

[4] In max RSRP-UA-OPA, each user associates the BS with the maximum reference signal received power (max RSRP) and the optimal power allocation is found by exhaustive search. In max RSRP-UA-EPA, each user associates the BS with the maximum RSRP and all BSs adopt the equal power allocation method. In max SINR-UA-EPA, each user associates the BS with the maximum SINR and all BSs equally allocate power.

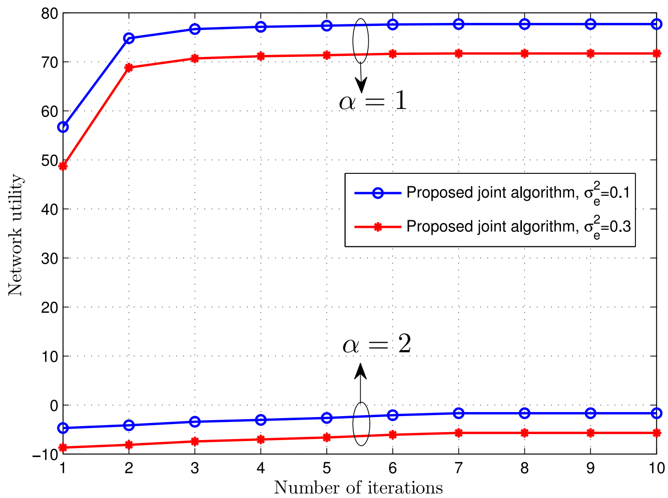

Fig. 5. Network utility versus the number of iterations.

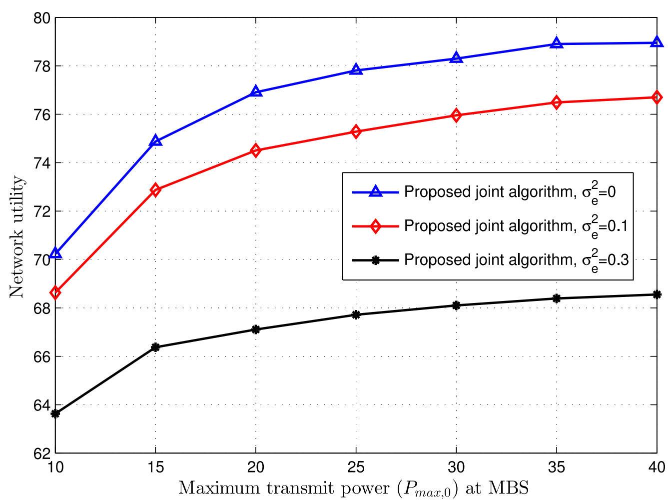

Fig. 6. Network utility with $\alpha = 1$ versus the maximum transmit power at MBS.

reveals that we can reasonably assign some UTs at PBS to avoid MBS overload, and further improve the network utility.

Fig. 5 evaluates the network utility performance with different α-fairness versus the number of iterations of Algorithm 2. We set the number of transmit antennas at MBS as $M_0 = 150$, and the number of transmit antennas at PBS as $M_j = 30,\ \forall j \in \mathcal{B}_p$. In our simulations, the convergence of the proposed joint iterative algorithm is also presented with different variance of estimated error. It can be observed that the proposed joint iterative algorithm converges after about 7 iterations. We can also observe that the system network utility performance will deteriorate as the error variance increases.

In Fig. 6, we exhibit the network utility with different estimation error variances when the maximum transmit power $P_{max,0}$ of MBS is increased from 10 dBm to 40 dBm. In this simulation, we set fairness parameter as $\alpha = 1$. The number of transmit antennas at MBS is $M_0 = 100$, and the number of transmit antennas at PBS is $M_j = 30,\ \forall j \in \mathcal{B}_p$. The maximum transmit power of PBS is $P_{max,j} = 30$ dBm, $\forall j \in \mathcal{B}_p$. As observed in Fig. 6, the network utility increases with the maximum power constraint because the optimization region of P_0 is enlarged. But the network utility increases slowly as the maximum transmit power constraint grows larger. This is because the closed-form expression of the achievable rate in (6) has the Shannon's capacity bound. As expected, we can also observe that the network utility decreases when the error variance increases. For example, when the maximum transmit power is 20 dBm, the network utility of the proposed joint algorithm with $\sigma_e^2 = 0$ is about 3.2% more than that with $\sigma_e^2 = 0.1$, and is about 14.6% more than that with $\sigma_e^2 = 0.3$.

5. Conclusions

For the Massive MIMO HetNets with imperfect CSI, we studied the optimization problem of joint user association and power allocation to maximize the system network utility. Under the backhaul capacity and load constraints, the original non-convex optimization problem was decomposed into user association subproblem and power allocation subproblem. An optimal user association solution was proposed by using the Lagrangian dual method. Given the user association scheme, we solved the power allocation subproblem by employing GABS approach. Finally, we designed a feasible joint two-layer iterative algorithm to achieve the maximum network utility. The effectiveness and convergence of our proposed joint iterative algorithm was illustrated, and the superior performance of the proposed algorithm was also shown in comparison with several reference algorithms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

To obtain a closed-form of $R_{j,k}$ in (4), we will calculate the following expectation. Substituting $\mathbf{G}_j^{\dagger} = \hat{\mathbf{G}}_j(\hat{\mathbf{G}}_j^H\hat{\mathbf{G}}_j)^{-1}$ into $\mathbb{E}\{\|\mathbf{g}_{j,k}^{\dagger}\|^2\}$, we have

$$\begin{aligned}\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\} &= \mathbb{E}\left\{\mathbf{g}_{j,k}^{\dagger H}\mathbf{g}_{j,k}^{\dagger}\right\} = \mathbb{E}\left\{\left[\mathbf{G}_j^{\dagger H}\mathbf{G}_j^{\dagger}\right]_{kk}\right\} \\ &= \mathbb{E}\left\{\left[\left(\hat{\mathbf{G}}_j^H\hat{\mathbf{G}}_j\right)^{-1}\right]_{kk}\right\} = \mathbb{E}\left\{\left(\hat{\mathbf{g}}_{j,k}^H\hat{\mathbf{g}}_{j,k}\right)^{-1}\right\} \\ &= \frac{1}{\beta_{j,k}}\mathbb{E}\left\{\left(\hat{\mathbf{h}}_{j,k}^H\hat{\mathbf{h}}_{j,k}\right)^{-1}\right\} = \frac{1}{\beta_{j,k}}\mathbb{E}\left\{\left[\left(\hat{\mathbf{H}}_j^H\hat{\mathbf{H}}_j\right)^{-1}\right]_{kk}\right\} \\ &= \frac{1}{\beta_{j,k}S_j}\mathbb{E}\left\{\mathrm{tr}\left(\hat{\mathbf{H}}_j^H\hat{\mathbf{H}}_j\right)^{-1}\right\} \overset{(a)}{=} \frac{1}{\beta_{j,k}\left(M_j - S_j\right)\left(1-\sigma_e^2\right)}\end{aligned} \tag{19}$$

where (a) is obtained by using the property of Wishart matrix [6], i.e., $\mathbb{E}\{\mathrm{tr}(\hat{\mathbf{H}}_j^H\hat{\mathbf{H}}_j)^{-1}\} = \frac{S_j}{(M_j - S_j)(1-\sigma_e^2)}$.

The second expectation of the denominator in (5) can be simplified as

$$\begin{aligned}\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right|^2\right\} &= \mathbb{E}\left\{\mathbf{g}_{j',k'}^{\dagger H}\hat{\mathbf{g}}_{j',k}\hat{\mathbf{g}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right\} \\ &= \beta_{j',k}\mathbb{E}\left\{\mathbf{g}_{j',k'}^{\dagger H}\hat{\mathbf{h}}_{j',k}\hat{\mathbf{h}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right\} \\ &= \beta_{j',k}\left(1-\sigma_e^2\right)\mathbb{E}\left\{\mathbf{g}_{j',k'}^{\dagger H}\mathbf{g}_{j',k'}^{\dagger}\right\} \\ &= \beta_{j',k}\left(1-\sigma_e^2\right)\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\}\mathbb{E}\left\{\mathbf{a}_{j',k'}^H\mathbf{a}_{j',k'}\right\} \\ &= \beta_{j',k}\left(1-\sigma_e^2\right)\mathbb{E}\left\{\left\|\mathbf{g}_{j,k}^{\dagger}\right\|^2\right\}.\end{aligned} \tag{20}$$

Substituting (19) into (20), we obtain

$$\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{j',k}^H\mathbf{g}_{j',k'}^{\dagger}\right|^2\right\} = \frac{\beta_{j',k}}{\beta_{j,k}\left(M_j - S_j\right)}. \tag{21}$$

The expected results in (6) are obtained by plugging (19) and (21) into (5).

Appendix B

We consider the optimization problem $G(\boldsymbol{\psi})$ in (13), and relax $x_{j,k}$ to be a continuous variable. As a result, the problem (13) can be transformed into

$$H(\boldsymbol{\psi}) = \begin{cases} \max\limits_{\mathbf{X}} & \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \psi_{j,k}x_{j,k} \\ s.t. & (9e)-(9f) \\ & x_{j,k}\geq 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U}. \end{cases} \tag{22}$$

Therefore, the dual problem (11) can be written as

$$\min_{\boldsymbol{\psi}\geq 0} D(\boldsymbol{\psi}) = F(\boldsymbol{\psi}) + H(\boldsymbol{\psi}). \tag{23}$$

Note that, given $\mathbf{P}$, (23) is a partial Lagrangian dual problem of the following convex optimization problem

$$\begin{aligned} \max_{\mathbf{X}}\ & \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} U_\alpha\left(r_{j,k}\right) \\ s.t.\ & (9b), (9c), (9e), (9f), (10b) \\ & x_{j,k}\geq 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U}. \end{aligned} \tag{24}$$

For problem (24), we can obtain its Lagrangian function as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\vartheta},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\eta},\boldsymbol{\psi},\boldsymbol{\phi},\boldsymbol{\theta}) \\ &= \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} U_\alpha\left(r_{j,k}\right) + \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \vartheta_{j,k}\left(R_{j,k}-r_{j,k}\right) + \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \lambda_{j,k}r_{j,k} \\ &+ \sum_{j\in\mathcal{B}} \beta_j\left(C_j - \sum_{k\in\mathcal{U}} R_{j,k}\right) + \sum_{j\in\mathcal{B}} \varphi_j\left(S_j - \sum_{k\in\mathcal{U}} x_{j,k}\right) \\ &+ \sum_{k\in\mathcal{U}} \eta_k\left(1 - \sum_{j\in\mathcal{B}} x_{j,k}\right) \\ &+ \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \psi_{j,k}\left(x_{j,k}-a_{j,k}\right) + \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \phi_{j,k}a_{j,k} + \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \theta_{j,k}x_{j,k} \end{aligned} \tag{25}$$

where $\boldsymbol{\vartheta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\eta}$, $\boldsymbol{\psi}$, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ are the vectors of the dual variables. We can rewrite the Lagrangian function in (25) with the decomposition-based method [28], and it can be expressed as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\vartheta},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\eta},\boldsymbol{\psi},\boldsymbol{\phi},\boldsymbol{\theta}) \\ &= \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} L_{j,k}(r_{j,k}) + \sum_{j\in\mathcal{B}} \beta_jC_j + \sum_{j\in\mathcal{B}} \varphi_jS_j + \sum_{k\in\mathcal{U}} \eta_k \\ &+ \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \left(\vartheta_{j,k}-\beta_j\right)R_{j,k} + \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \left(\phi_{j,k}-\psi_{j,k}\right)a_{j,k} \\ &+ \sum_{j\in\mathcal{B}} \left(\psi_{j,k}+\theta_{j,k}-\varphi_j-\eta_k\right)x_{j,k} \end{aligned} \tag{26}$$

where $L_{j,k}(r_{j,k}) = U_\alpha(r_{j,k}) - (\vartheta_{j,k}-\lambda_{j,k})r_{j,k}$. Thus, the full Lagrangian dual problem of (24) is given by

$$\begin{aligned} \min_{(\boldsymbol{\vartheta},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\eta},\boldsymbol{\psi},\boldsymbol{\phi},\boldsymbol{\theta})\succ\mathbf{0}}\ & \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} \left(\max_{r_{j,k}} L_{j,k}\left(r_{j,k}\right)\right) \\ & + \sum_{j\in\mathcal{B}} \beta_jC_j + \sum_{j\in\mathcal{B}} \varphi_jS_j + \sum_{k\in\mathcal{U}} \eta_k \\ s.t.\ & \left(\vartheta_{j,k}-\beta_j\right)\bar{R}_{j,k} + \phi_{j,k} - \psi_{j,k} = 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U} \\ & \psi_{j,k} + \theta_{j,k} - \varphi_j - \eta_k = 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U}. \end{aligned} \tag{27}$$

Based on the objective function of (27), we have $\max\limits_{r_{j,k}} L_{j,k}(r_{j,k}) = V_\rho(\vartheta_{j,k}-\lambda_{j,k})$. For simplicity, let $\varpi_{j,k} = \vartheta_{j,k}-\lambda_{j,k}$. Since $\phi_{j,k} > 0$ and $\theta_{j,k} > 0$, problem (27) can be written as

$$\begin{aligned} \min_{(\boldsymbol{\varpi},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{\eta},\boldsymbol{\psi})\succ\mathbf{0}}\ & \sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{B}} V_\rho\left(\varpi_{j,k}\right) + \sum_{j\in\mathcal{B}} \beta_jC_j + \sum_{j\in\mathcal{B}} \varphi_jS_j + \sum_{k\in\mathcal{U}} \eta_k \\ s.t.\ & \left(\varpi_{j,k}+\lambda_{j,k}-\beta_j\right)\bar{R}_{j,k} - \psi_{j,k} \leq 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U} \\ & \psi_{j,k} - \varphi_j - \eta_k \leq 0, \quad \forall\ j\in\mathcal{B}, k\in\mathcal{U}. \end{aligned} \tag{28}$$

In (28), the minimum value can be found when $\boldsymbol{\lambda} = 0$, for a fixed vector $\{\boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\eta}, \boldsymbol{\psi}\}$. Consequently, we can obtain the minimization problem (15). Next, we introduce the following Lemma to prove that the optimal value $\boldsymbol{\psi}^*$ of (15) will be the optimal solution to problem (11) [14].

Lemma : Consider the following convex optimization problem

$$\begin{aligned} &\max_{\mathbf{y}} f(\mathbf{y}) \\ &\quad s.t.\ \ g_m(\mathbf{y}) \leq 0, \quad m = 1,\ldots,M \\ &\quad h_n(\mathbf{y}) = 0, \quad n = 1,\ldots,N \end{aligned} \tag{29}$$

where $g_m(\mathbf{y})$ and $h_n(\mathbf{y})$ are linear constraints. We assume that a partial Lagrangian dual problem of (29) is

$$\begin{aligned} &\min_{\boldsymbol{\xi}\geq 0}\max_{\mathbf{y}}\left(f(\mathbf{y}) - \sum_{m=1}^{L}\bar{\xi}_m g_m(\mathbf{y})\right) \\ &\quad s.t.\ \ g_m(\mathbf{y}) \leq 0, \quad m = L+1,\ldots,M \\ &\quad h_n(\mathbf{y}) = 0, \quad n = 1,\ldots,N. \end{aligned} \tag{30}$$

The full Lagrangian dual problem of (29) is expressed as

$$\begin{aligned} \min_{\boldsymbol{\xi}\geq 0,\boldsymbol{\zeta}\geq 0,\boldsymbol{\varsigma}\geq 0}\max_{\mathbf{y}}\Bigg(& f(\mathbf{y}) - \sum_{m=1}^{L}\xi_m g_m(\mathbf{y}) \\ & - \sum_{m=L+1}^{M}\zeta_m g_m(\mathbf{y}) + \sum_{n=1}^{N}\varsigma_n h_n(\mathbf{y})\Bigg). \end{aligned} \tag{31}$$

If $\{\boldsymbol{\xi}^*, \boldsymbol{\zeta}^*, \boldsymbol{\varsigma}^*\}$ is the optimal solution of (31), then we have that $\boldsymbol{\xi}^*$ is optimal to (30).

Based on the above discussion, we know that problem (24) is a convex optimization problem with feasible set, and problem (23) is a partial Lagrangian dual problem of (24), while problem (28) is the full Lagrangian dual of (24) and $\{\boldsymbol{\varpi}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}^*, \boldsymbol{\varphi}^*, \boldsymbol{\eta}^*, \boldsymbol{\psi}^*\}$ is an optimal solution to (28). According to *Lemma*, we can obtain the conclusion that $\boldsymbol{\psi}^*$ is the optimal solution to problem (11) or problem (23).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sigpro.2020.107588.

CRediT authorship contribution statement

Hao Li: Conceptualization, Methodology, Formal analysis. **Zhigang Wang:** Data curation, Writing - original draft. **Houjun Wang:** Supervision, Writing - review & editing.

References

[1] T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas, IEEE Trans. Wireless Commun. 9 (11) (Nov. 2010) 3590–3600.

[2] H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems, IEEE Trans. Commun. 61 (4) (Apr. 2013) 1436–1449.

[3] J.B. Rao, A.O. Fapojuwo, A survey of energy efficient resource management techniques for multicell cellular networks, IEEE Commun. Surveys & Tuts. 16 (1) (1st Quart. 2014) 154–180.

[4] H. Zhang, X. Chu, W. Guo, S. Wang, Coexistence of wi-Fi and heterogeneous small cell networks sharing unlicensed spectrum, IEEE Commun. Mag. 53 (3) (Mar. 2015) 158–164.

[5] H. Li, J. Cheng, Z. Wang, H. Wang, Joint antenna selection and power allocation for an energy-efficient massive MIMO system, IEEE Wireless Commun. Lett. 8 (1) (Feb. 2019) 257–260.

[6] T. Van Chien, E. Björnson, E.G. Larsson, Joint power allocation and user association optimization for massive MIMO systems, IEEE Trans. Wireless Commun. 15 (9) (Jun. 2016) 6384–6399.
[7] D. Bethanabhotla, O.Y. Bursalioglu, H.C. Papadopoulos, G. Caire, Optimal user–cell association for massive MIMO wireless networks, IEEE Trans. Wireless Commun. 15 (3) (Mar. 2016) 1835–1850.
[8] R. Hamdi, E. Driouch, W. Ajib, Resource allocation in downlink large-scale MIMO systems, IEEE Access 4 (Nov. 2016) 8303–8316.
[9] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, J.G. Andrews, User association for load balancing in heterogeneous cellular networks, IEEE Trans. Wireless Commun. 12 (6) (Jun. 2013) 2706–2716.
[10] G. Ye, H. Zhang, H. Liu, J. Cheng, V.C. Leung, Energy efficient joint user association and power allocation in a two-tier heterogeneous network, in: IEEE GLOBECOM, Dec. 2016, pp. 1–5.
[11] N. Wang, E. Hossain, V.K. Bhargava, Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays, IEEE Trans. Wireless Commun. 15 (5) (May 2016) 3251–3268.
[12] D. Liu, L. Wang, Y. Chen, T. Zhang, K.K. Chai, M. Elkashlan, Distributed energy efficient fair user association in massive MIMO enabled HetNets, IEEE Commun. Lett. 19 (10) (Oct. 2015) 1770–1773.
[13] Y. Hao, Q. Ni, H. Li, S. Hou, Energy and spectral efficiency tradeoff with user association and power coordination in massive MIMO enabled HetNets, IEEE Commun. Lett. 20 (10) (Oct. 2016) 2091–2094.
[14] H. Ma, H. Zhang, X. Wang, J. Cheng, Backhaul-aware user association and resource allocation for massive MIMO-enabled hetnets, IEEE Commun. Lett. 21 (12) (Sep. 2017) 2710–2713.
[15] Y. Lin, Y. Wang, C. Li, Y. Huang, L. Yang, Joint design of user association and power allocation with proportional fairness in massive MIMO HetNets, IEEE Access 5 (May 2017) 6560–6569.
[16] C. Li, H.J. Yang, S. Fan, J. Cioffi, L. Yang, Multi-user overhearing for cooperative two-way multi-antenna relays, IEEE Trans. Veh. Technol. 65 (5) (Jun. 2016) 3796–3802.
[17] C. Li, S. Zhang, P. Liu, F. Sun, J.M. Cioffi, L. Yang, Overhearing protocol design exploiting intercell interference in cooperative green networks, IEEE Trans. Veh. Technol. 65 (1) (Jan. 2016) 441–446.
[18] S.S. Ikki, S. Aissa, Two-way amplify-and-forward relaying with gaussian imperfect channel estimations, IEEE Commun. Lett. 16 (7) (Jul. 2012) 956–959.
[19] J. Li, E. Björnson, T. Svensson, T. Eriksson, M. Debbah, Joint precoding and load balancing optimization for energy-efficient heterogeneous networks, IEEE Trans. Wireless Commun. 14 (10) (Jun. 2015) 5810–5822.
[20] H.Q. Ngo, L.-N. Tran, T.Q. Duong, M. Matthaiou, E.G. Larsson, On the total energy efficiency of cell-free massive MIMO, IEEE Trans. Green Commun. Netw. 2 (1) (Nov. 2018) 25–39.
[21] J. Jose, A. Ashikhmin, T.L. Marzetta, S. Vishwanath, Pilot contamination and precoding in multi-cell TDD systems, IEEE Trans. Wireless Commun. 10 (8) (Aug. 2011) 2640–2651.
[22] F. Fang, H. Zhang, J. Cheng, S. Roy, V.C. Leung, Joint user scheduling and power allocation optimization for energy efficient NOMA systems with imperfect CSI, IEEE J. Sel. Areas Commun. 35 (12) (Nov. 2017) 2874–2885.
[23] X. Wang, F. Zheng, P. Zhu, X. You, Energy-efficient resource allocation in coordinated downlink multi-cell ofdma systems, IEEE Trans. Veh. Technol. 65 (3) (Mar. 2016). 1–1.
[24] D.W.K. Ng, E.S. Lo, R. Schober, Energy-efficient resource allocation in ofdma systems with large numbers of base station antennas, IEEE Trans. Wireless Commun. 11 (9) (Sep. 2012) 3292–3304.
[25] T.L. Marzetta, Fundamentals of Massive MIMO, Cambridge University Press, 2016.
[26] D. Bethanabhotla, O.Y. Bursalioglu, H.C. Papadopoulos, G. Caire, User association and load balancing for cellular massive MIMO, in: Inf. Theory and Appl. Workshop (ITA), Feb. 2014, pp. 1–10.
[27] S.H. Low, D.E. Lapsley, Optimization flow controli: basic algorithm and convergence, IEEE/ACM Trans. Netw. 7 (6) (Dec. 1999) 861–874.
[28] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge Univ. Press, New Yourk, NY, USA, 2004.
[29] M. Grant, S. Boyd, Y. Ye, CVX: Matlab software for disciplined convex programming, 2008.
[30] G. Miao, N. Himayat, G.Y. Li, Energy-efficient link adaptation in frequency-selective channels, IEEE Trans. Commun. 58 (2) (Feb. 2010).
[31] Su J., Z. Sheng, V. C.M. Leung, Y. Chen, Energy efficient tag identification algorithms for RFID: survey, motivation and new design, IEEE Wireless Commun. 26 (3) (Feb. 2019) 118–124.