On the Energy and Spectral Efficiency Tradeoff in Massive MIMO Enabled HetNets with Capacity-Constrained Backhaul Links

Yuanyuan Hao, Qiang Ni, *Senior Member, IEEE*, Hai Li, *Member, IEEE*, and Shujuan Hou

***Abstract*—In this paper, we propose a general framework to study the tradeoff between energy efficiency (EE) and spectral efficiency (SE) in massive MIMO enabled HetNets while ensuring proportional rate fairness among users and taking into account the backhaul capacity constraint. We aim at jointly optimizing user association, spectrum allocation, power coordination, and the number of activated antennas, which is formulated as a multi-objective optimization problem maximizing EE and SE simultaneously. With the help of weighted Tchebycheff method, it is then transformed into a single-objective optimization problem, which is a mixed-integer non-convex problem and requires unaffordable computational complexity to find the optimum. Hence, a low-complexity effective algorithm is developed based on primal decomposition, where we solve the power coordination and number of antenna optimization problem and the user association and spectrum allocation problem separately. Both theoretical analysis and numerical results demonstrate that our proposed algorithm can fast converge within several iterations and significantly improve both the EE-SE tradeoff performance and rate fairness among users compared to other algorithms.**

***Index Terms*—Energy efficiency, HetNets, massive MIMO, power coordination, proportional fairness, spectral efficiency, spectrum allocation, user association.**

I. Introduction

OWING to the spectrum scarcity and the explosive growth of mobile data traffic demand for multimedia applications, there is an urgent need to significantly improve spectral efficiency (SE) [1]. Meanwhile, with steadily rising energy costs and increasing environmental concerns, energy efficiency (EE) is becoming increasingly important for wireless communications and has caught more and more attention [2]. Therefore, the joint maximization of EE and SE has turned into one of the main goals for the future fifth-generation (5G) cellular communication networks. Also, it has been widely recognized that massive multiple-input-multiple-output (MIMO) and multi-tier heterogenous networks (HetNets) are two promising techniques to achieve this goal for 5G networks [3]. As massive MIMO base stations (BS) are capable of communicating with multiple single-antenna users over the same time-frequency slot, the system SE can be greatly improved [4]. Besides, massive MIMO provides a high power gain, which helps to reduce the transmit power and further obtain a higher EE. On the other hand, since multiple small cells are overlaid on the coverage of each macrocell in the HetNets, the spectrum is densely reused over a geographical area, which makes BSs closer to users. Consequently, this topology can significantly improve the system EE and SE [5], and also enhance the performance of cell-edge users [6].

Although HetNets show great promises to obtain higher SE by universal frequency reuse, this infrastructure also introduces new challenges. Specifically, since the transmit powers can be much diverse for different kinds of BSs in HetNets, the conventional user association scheme which is determined according to signal-to-interference-plus-noise ratio (SINR), will cause heavy load imbalance and further damage the system SE and EE [7]. Therefore, the joint user association and spectrum allocation is needed to improve the network SE. In addition, as multiple small cells coexist with the macro cell, the power coordination is indispensable to mitigate the inter-tier and intra-tier interference. As for massive MIMO, although equipping large number of antennas can improve the system SE, it also causes higher circuit power consumption when more antennas are deployed, which may degrade the performance of EE. Furthermore, these problems are closely coupling, influencing the system EE and SE together. Hence, intelligent joint resource optimization is one of crucial issues for massive MIMO enabled HetNets.

The energy efficient resource optimization for HetNets has been investigated extensively [8]-[15]. In [8], energy efficient user association is optimized to minimize the total power consumption. The work in [9] investigates both the noncooperative and cooperative energy-efficient power control, where all tiers selfishly or cooperatively choose their transmit power to optimize their network EE. Similarly, the authors in [10] and [11] provide insight on energy efficient power coordination, where a new adjustable utility function is adopted to jointly optimize SE and EE of each BS in HetNets. In [12], the energy efficiency issue is formulated to minimize the total power consumption of all pico BSs under the average delay constraint. Nevertheless, the studies for the EE optimization in massive MIMO enabled HetNets are still limited. In [13], the authors analyze the impact of massive MIMO on the SE and EE of K-tier HetNets by employing a stochastic geometry method. In [14], the summation of users' EE is maximized with the optimization of user association. Recently, the tradeoff

Manuscript received January 1, 2017; revised April 19, 2017; accepted July 13, 2017. This work was supported in part by the China Scholarship Council and EU CROWN project under grant PIRSES-GA-2013-610524. The editor coordinating the review of this paper and approving it for publication was Tony Q.S. Quek. *(Corresponding Author: Hai Li.)*

Y. Hao, H. Li, and S. Hou are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (email: {tracyhao, haili, shujuanhou}@bit.edu.cn).

Q. Ni is with the School of Computing and Communications, Lancaster University, LA1 4WA, UK (email: q.ni@lancaster.ac.uk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

between EE and SE is studied in our earlier work [15], where the user association and power coordination are jointly optimized.

However, in the aforementioned studies [8]–[15], the constraint of backhaul capacity between BSs and the core network is not considered. With the employment of dense small cells and large-scale antenna arrays in macro BSs, there will be enormous data transmissions over backhaul links, while current backhaul solutions can not provide sufficient capacity [16], [17] and has become one of crucial performance constraints [18], [19]. Hence, it is indispensable to take the backhaul constraint into account when optimizing EE and SE in massive MIMO enabled HetNets. The backhaul bottleneck of HetNets has been investigated to some extent in the literature [20]–[22]. The authors in [20] propose a distributed load-balancing user association algorithm. In [21], the in-band wireless backhaul in massive MIMO enabled HetNets is investigated, where the downlink user association and backhaul bandwidth allocation are jointly optimized to maximize the sum logarithmic user rate. Similarly, the work in [22] considers the joint optimization of user association and resource allocation in hybrid energy powered HetNets, which aims at maximizing the network utility under the proportional fairness subject to both the backhaul and energy constraints. Nevertheless, the studies in [20]–[22] only involve SE and rate fairness, and the performance of EE is totally ignored.

To the best of our knowledge, there is no existing work that studies the tradeoff between EE and SE in massive MIMO enabled HetNets while taking into consideration the backhaul capacity constraint. Motivated by such observations, this paper studies the tradeoff between EE and SE in massive MIMO enabled HetNets from an optimization perspective under the proportional fairness criterion and the constraint of backhaul capacity, where we comprehensively consider the joint optimization of user association, spectrum allocation, power coordination, and the number of activated antennas. To be more specific, the contributions of this paper are summarized in the following:

- A general framework is proposed to investigate the tradeoff between the system EE and SE while ensuring proportional rate fairness among users and taking into account the backhaul capacity constraint. Specifically, this problem is first formulated as an multi-objective optimization (MOO) problem maximizing the sum log-utility and minimizing the total power consumption simultaneously. The weighted Tchebycheff method is then employed to transform the MOO problem into a single-objective optimization (SOO) problem, which is necessary and sufficient for Pareto optimality and thus provides the complete Pareto optimal set with the variation of the weighting parameter.
- We exploit four degrees of freedom in resource optimization for massive MIMO enabled HetNets, which is the first time in the literature to jointly optimize user association, power coordination, spectrum allocation, and the number of activated antennas simultaneously. Note that these optimization items are closely coupling in terms of EE and SE. To reduce computational complexity, an effective algorithm is proposed to solve this mixed-integer and non-convex problem, where primal decomposition is used to separate the original problem into two levels. Specifically, the power coordination and antenna number optimization problem is solved via successive convex programming (SCP), which is transformed into a series of convex problem and thus requires only polynomial complexity. For the other problem, we develop a distributed user association and spectrum allocation algorithm via Lagrange dual decomposition, which reduces both the computational complexity and information exchange.
- The optimality, convergence, and complexity of the proposed algorithm are analyzed. It is first demonstrated theoretically that the proposed power coordination and antenna number optimization algorithm is guaranteed to converge to the solution satisfying KKT conditions. Then, we study the optimality of the proposed distributed user association and spectrum allocation algorithm, and prove that the dual gap between the primal problem and its dual problem is bounded. In addition, the overall computational complexity of the proposed algorithm is analyzed in detail. Furthermore, the convergence property of the proposed algorithm is testified by simulation results, including both the iterations of SCP algorithm and the overall iterations between two subproblems. Also, the effectiveness of the proposed algorithm is demonstrated numerically compared to other algorithms in terms of the EE-SE tradeoff and rate fairness.
- Particularly, the characteristics about the optimal number of activated antennas are investigated theoretically and numerically. We find that the optimal number of activated antennas is always the number of equipped antennas when the system SE is maximized. By contrast, the optimal number of activated antennas corresponding to the maximum EE diminishes with the increase of the circuit power consumption per antenna, when the number of equipped antennas is sufficiently large. The impact of the backhaul capacity constraint on the system EE and SE is also captured via numerical results, which indicates that the backhaul capacity is indeed a bottleneck for network performance.

The rest of this paper is organized as follows. We first describe the system model in Section II. Then, in Section III, a multi-objective optimization problem is formulated to maximize SE and EE while ensuring proportional rate fairness, which is further solved in Section IV. Finally, simulation results are presented in Section V, which is followed by conclusions in Section VI.

II. System Model

As presented in Fig. 1, we consider the downlink transmission in a two-tier HetNet consisting of a macro BS (MBS), $I-1$ single-antenna pico BSs (PBS), and J single-antenna users. Let $i \in \{1, 2, \cdots, I\}$ be the index of BSs, where we use $i = 1$ to represent the MBS, and the others are PBSs. The indexes of users are denoted by $j \in \{1, 2, \cdots, J\}$.

Besides, a time-division duplexing (TDD) scheme is considered with perfect channel state information (CSI). It is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

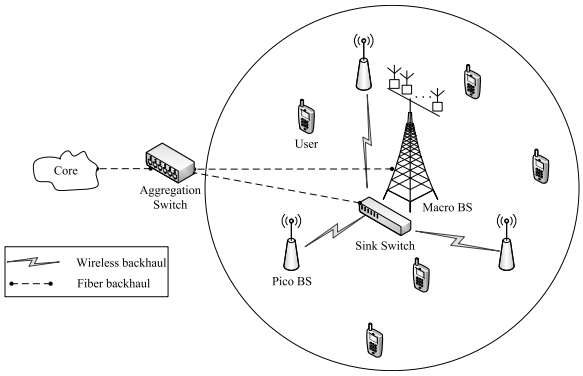

Fig. 1. System model.

also assumed that all BSs share the same frequency, and the orthogonal frequency-division multiple access (OFDMA) scheme is applied among users associated with the same BS. Specifically, the MBS is equipped with a large number of antennas, and let M represent the number of active antennas. We also assume that the MBS can transmit at most N ($N \ll M$) downlink data streams simultaneously over the same frequency band, i.e., the beamforming group size is N, and linear zero-forcing beamforming (LZFBF) is adopted for the downlink transmission. As shown in Fig. 1, wireless backhaul is assumed for the information exchange between PBSs and the sink switch, and the data transmission between MBS and the core network is based on the fiber backhaul.

A. Achievable Data Rate and Power Consumption

In this paper, we assume 'favorable propagation' between the MBS and users, i.e., the effects of small-scale fading can be eliminated completely by using a large number of antennas, which is derived based on the asymptotic random matrix theory [23]. $\mathbf{p} = [p_1, p_2, \cdots p_I]$ is defined as the transmit power vector of all BSs. Here we adopt the data rate formula in [21] and [24] for the LZFBF precoding. Thus, when $M, N \to \infty$ and $\frac{N}{M} \ll 1$ is fixed, the achievable normalized downlink data rate of user j associated with the MBS converges to the following deterministic expression

$$r_{ij} = N\log_2\left(1 + \mathrm{SINR}_{ij}\left(\mathbf{p}, M\right)\right),\ i = 1, \tag{1}$$

where the signal-to-interference-and-noise ratio (SINR) is calculated by

$$\mathrm{SINR}_{ij}\left(\mathbf{p}, M\right) = \frac{M - N + 1}{N} \cdot \frac{p_i g_{ij}}{\sum\limits_{l \neq i} p_l g_{lj} + \sigma_j^2},\ i = 1. \tag{2}$$

Here g_{ij} is the large-scale fading between the i-th BS and the j-th user, and σ_j^2 represents the noise power at user j. Note that this large-system asymptotic is demonstrated to be accurate for practical values of the antenna array size M and the beamforming group size N in [24] and the references therein.

Generally, the user association period is sufficiently longer than the period of small-scale fading so that the impact of small-scale fading can be averaged out in the channel measurement [12], [21]. Hence, the long-term transmission data rate for users associated with PBSs is considered, i.e.,

$$r_{ij} = \log_2\left(1 + \mathrm{SINR}_{ij}\right),\ i > 1, \tag{3}$$

where the long-term SINR is presented as

$$\mathrm{SINR}_{ij} = \frac{p_i g_{ij}}{\sum\limits_{l \neq i} p_l g_{lj} + \sigma_j^2},\ i > 1. \tag{4}$$

Assume that each user can be only associated with one of BSs, and the binary variable matrix $\mathbf{X} = [x_{ij}]$ is introduced to describe the user association status, where

$$x_{ij} = \begin{cases} 1, & \text{if user } j \text{ associated with BS } i, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

As the system generally occupies several hundreds of resource blocks (RBs), we normalize this number to be 1 and define the resource allocation variable $y_{ij}, 0 \leq y_{ij} \leq 1$, indicating the fraction of RBs used for the communication between user j and the i-th BS. Thus, the transmission data rate of user j is calculated by

$$R_j = \sum_i x_{ij} y_{ij} r_{ij},\ j = 1, 2 \cdots J. \tag{6}$$

On the other hand, the power consumption of the MBS can be expressed as [25]

$$P_i = \varepsilon_i p_i + M p_{\mathrm{a}} + p_{i,\mathrm{s}},\ i = 1, \tag{7}$$

where ε_i denotes the power amplifier efficiency of the i-th BS, p_{a} describes the circuit power per antenna, and $p_{i,\mathrm{s}}$ represents the static circuit power term independent of the antenna number. In contrast, the power consumption of the i-th PBS is given by

$$P_i = \varepsilon_i p_i + p_{\mathrm{a}} + p_{i,\mathrm{s}},\ i > 1. \tag{8}$$

Then, the total network power consumption can be obtained by

$$P_{\mathrm{tot}} = \sum_i P_i + P_{\mathrm{BH}}, \tag{9}$$

where P_{BH} denotes the power consumption for backhaul.

Specifically, the backhaul power consists of the power consumption of fiber-based and microwave-based backhaul links, which is expressed as

$$P_{\mathrm{BH}} = P_{\mathrm{bh}}^{\mathrm{Fiber}} + P_{\mathrm{bh}}^{\mathrm{Micro}}. \tag{10}$$

In (10), the power consumption for the fiber-based backhaul links is calculated based on the model in [26], [27], which is given by

$$P_{\mathrm{bh}}^{\mathrm{fiber}} = \left\lceil \frac{N_{\mathrm{dl}}}{\max_{\mathrm{dl}}} \right\rceil p_{\mathrm{switch}}^{\mathrm{fiber}} + N_{\mathrm{dl}} p_{\mathrm{dl}} + N_{\mathrm{ul}} p_{\mathrm{ul}}, \tag{11}$$

where $\max_{\mathrm{dl}}$ denotes the highest number of available downlink interfaces for one aggregation switch; $p_{\mathrm{switch}}^{\mathrm{fiber}}$ describes the maximum power consumption of one fiber-based aggregation switch. N_{dl} and p_{dl} represent the number of downlink interfaces and the power consumption of each downlink interface, respectively. N_{ul} denotes the number of uplink interfaces, which is a function of total aggregate traffic collected at the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

switch ($AT_{\rm tot}$) and the maximum transmission rate of an uplink interface ($U_{\max}$), i.e., $N_{\rm ul}=\left\lceil\frac{AT_{\rm tot}}{U_{\max}}\right\rceil$, and $p_{\rm ul}$ indicates the power consumption of each uplink interface. Similarly, The power consumed by microwave-based backhaul [26], [27] can be expressed as

$$P_{\rm bh}^{\rm micro}=\left\lceil\frac{AT_{\rm micro}}{C_{\rm switch}^{\rm micro}}\right\rceil P_{\rm switch}^{\rm micro}+N_{\rm PBS}P_{\rm low}+N_{\rm sink}P_{\rm high}, \quad (12)$$

where $AT_{\rm micro}$ is the aggregated microwave-based backhaul traffic, and $C_{\rm switch}^{\rm micro}$ and $P_{\rm switch}^{\rm micro}$ represent the capacity and the power consumption of each switch for microwave backhaul, respectively. $P_{\rm low}$ describes the power associated with microwave backhaul operations at each PBS, and $P_{\rm high}$ corresponds to the power consumption at the sink node. Correspondingly, $N_{\rm PBS}$ and $N_{\rm sink}$ represent the number of PBSs and the sink, respectively. In particular, we assume that the capacity of the backhaul link for the i-th BS is bounded by $C_{i,\rm bh}$.

B. SE, EE and Proportional Fairness

In this paper, the system SE (bits/s/Hz) is defined as the total system throughput per unit bandwidth, which is given by

$$\eta_{\rm SE}=\sum_j R_j=\sum_j\sum_i x_{ij}y_{ij}r_{ij}. \quad (13)$$

The system EE (bit/Joule/Hz) is further defined as the ratio of the system SE over the total power consumption, which can be expressed as

$$\eta_{\rm EE}=\frac{\eta_{\rm SE}}{P_{\rm tot}}=\frac{\sum_j\sum_i x_{ij}y_{ij}r_{ij}}{\sum_i P_i+P_{\rm BH}}. \quad (14)$$

It is known that SE and EE are conflicting objectives in a wireless communication system with limited radio resources [28]. This is because maximizing SE is equivalent to utilizing all the available resources, such as the maximum transmit power and all the available antennas to increase the throughout, while in such case the EE may become very low due to high power consumption. Therefore, only maximizing SE or EE may not satisfy the performance requirement of decision makers, and it is necessary to investigate the tradeoff between EE and SE. As explained in [29], the tradeoff between EE and SE can be obtained by solving a simplified MOO problem of minimizing the total power consumption and maximizing SE for specific ranges of the weighting parameter.

However, rate fairness among users can not be guaranteed in the above analysis, since only the system EE and SE are maximized. As proportional fairness can achieve a good balance between SE maximization and fairness [30], in this paper, the system SE is replaced in the MOO problem with the summation of the logarithmic rates of users, which is given by

$$\mathrm{U}\left(\mathbf{x},\mathbf{y},\mathbf{p},M\right)=\sum_j \ln\left(R_j\right). \quad (15)$$

III. Problem Formulation

As mentioned before, the tradeoff between EE and SE while ensuring proportional rate fairness among users can be equivalently formulated as a MOO problem maximizing the sum-log-rate utility and minimizing the total power consumption simultaneously. Thus, the joint optimization problem of user association, resource allocation, power coordination and number of activated antennas can be expressed as

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \ -\mathrm{U}\left(\mathbf{x},\mathbf{y},\mathbf{p},M\right),\\
&\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \ P_{\rm tot}\left(\mathbf{x},\mathbf{y},\mathbf{p},M\right),\\
&\text{s.t. } \mathrm{C1}: x_{ij}\in\{0,1\},\forall i,j,\\
&\quad \mathrm{C2}: \sum_i x_{ij}=1,\forall j,\\
&\quad \mathrm{C3}: y_{ij}\in[0,1],\forall i,j,\\
&\quad \mathrm{C4}: \sum_j x_{ij}y_{ij}\le 1,\forall i,\\
&\quad \mathrm{C5}: M\le M_{\max},\\
&\quad \mathrm{C6}: p_i\le p_{i,\max},\forall i,\\
&\quad \mathrm{C7}: \sum_j x_{ij}y_{ij}r_{ij}\le C_{i,\rm bh},\forall i,
\end{aligned} \quad (16)$$

where $M_{\max}$ and $p_{i,\max}$ denote the number of equipped antennas at the MBS and the maximum transmit power of BS i, respectively.

In (16), C1 and C2 indicates that each user can be only associated with one BS at any time. C3 and C4 are the constraints of spectrum allocation among users associated with the same BS. Then, C5 denotes that the number of activated antennas at the MBS is bounded by $M_{\max}$, and C6 ensures that the transmit power of the i-th BS is not more than $p_{i,\max}$, $\forall i$. Particularly, C7 represents that the capacity of the backhaul link for the i-th BS is bounded by $C_{i,\rm bh}$.

For the sake of consistent comparison, we first normalize the sum log-utility function and total power consumption in (16) as the following

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \ \mathrm{F}_1\left(\mathbf{x},\mathbf{y},\mathbf{p},M\right)=\frac{U_{\max}-\sum_j \ln(R_j)}{U_{\max}-U_{\min}},\\
&\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \ \mathrm{F}_2\left(\mathbf{x},\mathbf{y},\mathbf{p},M\right)=\frac{P_{\rm tot}}{P_{\max}},
\end{aligned} \quad (17)$$

where $P_{\max}$ denotes the maximum total power consumption, and $U_{\max}$ and $U_{\min}$ represent the maximum and minimum sum log-utility, respectively. Specifically, the maximum power consumption can be calculated by setting $\mathbf{p}=[p_{i,\max}]$ and $M=M_{\max}$, which is

$$P_{\max}=\sum_{i=1}^{I}\left(\varepsilon_i p_{i,\max}+p_{i,\rm s}\right)+p_{\rm a}\left(M_{\max}+I-1\right)+P_{\rm BH}. \quad (18)$$

Then, by omitting mutual interferences among BSs, we further obtain the upper bound of the sum log-utility, which must be maximized at $\mathbf{p}=[p_{i,\max}]$ and $M=M_{\max}$, i.e.,

$$U_{\max}=\max_{\mathbf{x},\mathbf{y}}\sum_j \ln\left(\sum_i x_{ij}y_{ij}\hat{r}_{ij}\right)\Bigg|_{p_i=p_{i,\max},M=M_{\max}}, \quad (19)$$

where $\hat{r}_{ij}$ is the approximate data rate without the consideration of mutual interferences. Thus, the Lagrange dual decomposition (LDD) method presented in Section IV can be used to find the user association and resource allocation solution of (19). On the other hand, the minimum utility can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

be approximated as

$$U_{\min} = J \ln(\delta), \tag{20}$$

where δ is a predefined and sufficiently small value, and we assume that $R_j \geq \delta, \forall j$.

Problem (16) is a mix-integer and non-concave problem due to the integer variables $\mathbf{x}$ and M and the existing inter-cell interference presented in r_{ij}, which is generally NP-hard. Weighted sum method is one of the most general methods for solving MOO problems, but it can not obtain points on non-convex portions of the Pareto optimal set [31]. Hence, in this paper, we employ weighted Tchebycheff method to transform problem (16) into a SOO problem, which is sufficient and necessary for achieving Pareto optimality and therefore provides the complete Pareto front of problem (16) [31]. According to the weighted Tchebycheff method, the MOO problem can be transformed into the following SOO problem as

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \max\left\{w\mathrm{F}_1(\mathbf{x},\mathbf{y},\mathbf{p},M), (1-w)\mathrm{F}_2(\mathbf{x},\mathbf{y},\mathbf{p},M)\right\}, \\ &\text{s.t. C1} - \text{C7}, \end{aligned} \tag{21}$$

where $w \in [0,1]$ is the weighting parameter representing the relative importance between two objectives. To illustrate the equivalence between the MOO problem (16) and the SOO problem (21), we have the following proposition.

Proposition 1: For any given $w \in [0,1]$, the unique optimal solution $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{p}^*, M^*)$ of problem (21) is Pareto optimal for the MOO problem (16).

Proof: See Appendix A.

IV. Proposed Algorithms

To further understand the tradeoff between SE and EE, it is necessary to find the optimal solution of (21). Specifically, we first introduce an additional variable φ to make problem (21) tractable, and the equivalent transformed problem is expressed as

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y},\mathbf{p},M,\varphi} \ \varphi, \\ &\text{s.t. C1} - \text{C7}, \\ &\quad \text{C8}: w\left(\frac{U_{\max} - \sum_j \ln(R_j)}{U_{\max} - U_{\min}}\right) \leq \varphi, \\ &\quad \text{C9}: (1-w)\left(\frac{P_{\text{tot}}}{P_{\max}}\right) \leq \varphi. \end{aligned} \tag{22}$$

However, it is still rather challenging to find the global optimum of problem (22) with affordable computational complexity, since problem (22) is mixed-integer and non-convex due to the integer variable x_{ij} and M and the inter-cell interference. Moreover, these optimization variables $(\mathbf{x},\mathbf{y},\mathbf{p},M)$ are closely coupling in terms of EE and SE. To solve this nontrivial problem, a series of decomposition algorithms will be proposed in the following. We first relax constraints C8-C9, and the corresponding Lagrange function can be presented as

$$\begin{aligned} \mathrm{L}(\mathbf{x},\mathbf{y},\mathbf{p},M,\varphi,\lambda,\mu) &= \varphi + \lambda\left(w\mathrm{F}_1(\mathbf{x},\mathbf{y},\mathbf{p},M) - \varphi\right) \\ &\quad + \mu\left((1-w)\mathrm{F}_2(\mathbf{x},\mathbf{y},\mathbf{p},M) - \varphi\right), \end{aligned} \tag{23}$$

where $\lambda \geq 0$ and $\mu \geq 0$ are Lagrangian multipliers. Then, the corresponding dual function is given by

$$\mathrm{H}(\lambda,\mu) = \begin{cases} \min\limits_{\mathbf{x},\mathbf{y},\mathbf{p},M,\varphi} \ \mathrm{L}(\mathbf{x},\mathbf{y},\mathbf{p},M,\varphi,\lambda,\mu), \\ \text{s.t. C1} - \text{C7}. \end{cases} \tag{24}$$

Rearranging the Lagrange function yields

$$\begin{aligned} &\mathrm{L}(\mathbf{x},\mathbf{y},\mathbf{p},M,\varphi,\lambda,\mu) \\ &= \underbrace{-\frac{\lambda w \mathrm{U}(\mathbf{x},\mathbf{y},\mathbf{p},M)}{U_{\max} - U_{\min}} + \frac{\mu(1-w)P_{\text{tot}}(\mathbf{x},\mathbf{y},\mathbf{p},M)}{P_{\max}}}_{\mathrm{L}_1(\mathbf{x},\mathbf{y},\mathbf{p},M,\lambda,\mu)} \\ &+ \underbrace{(1-\lambda-\mu)\varphi}_{\mathrm{L}_2(\varphi,\lambda,\mu)} + \frac{\lambda w U_{\max}}{U_{\max} - U_{\min}}. \end{aligned} \tag{25}$$

Thus, for the given (λ,μ), problem (24) can be decomposed into two subproblems, i.e., the joint user association and resource optimization problem

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y},\mathbf{p},M} \ \mathrm{L}_1(\mathbf{x},\mathbf{y},\mathbf{p},M,\lambda,\mu), \\ &\text{s.t. C1} - \text{C7}, \end{aligned} \tag{26}$$

and the adaptive φ selection problem

$$\min_{\varphi} \ \mathrm{L}_2(\varphi,\lambda,\mu). \tag{27}$$

A. Joint User Association and Resource Optimization

We first focus on finding the optimal solution of (26), i.e., the joint optimization of user association, spectrum allocation, power coordination and the number of activated antennas at the MBS. Since these optimization variables are coupled, the primal decomposition method [32] is appropriate to be applied to separate the original problem into the following two levels of optimization problem. Fixing user association matrix $\mathbf{x}$ and the spectrum allocation matrix $\mathbf{y}$ in problem (26) yields the *power coordination and antenna number optimization problem*, i.e.,

$$\begin{aligned} &\max_{\mathbf{p},M} \ \mathrm{f}_1(\mathbf{p},M) = \frac{\lambda w \sum_j \ln(R_j)}{U_{\max} - U_{\min}} - \frac{\mu(1-w)\left(\sum_i \varepsilon_i p_i + M p_{\text{a}}\right)}{P_{\max}}, \\ &\text{s.t. C5} - \text{C7}. \end{aligned} \tag{28}$$

On the contrary, by fixing the number of active antennas M and the transmit power vector $\mathbf{p}$ in problem (26), the *user association and spectrum allocation problem* can be obtained by

$$\begin{aligned} &\max_{\mathbf{x},\mathbf{y}} \ \mathrm{f}_2(\mathbf{x},\mathbf{y}) = \sum_j \ln\left(\sum_i x_{ij} y_{ij} r_{ij}\right), \\ &\text{s.t. C1} - \text{C4}, \text{C7}. \end{aligned} \tag{29}$$

Note that several constant terms are omitted in (28) and (29) for simplicity.

1. Subproblem 1: Power coordination and antenna number optimization

As mentioned before, because of the existing inter-cell interference and the integer variable M, the joint optimization of power coordination and the number of active antennas is a non-convex mixed-integer problem. To tackle this difficulty, we first relax the number of active antennas M to a real variable, and find the suboptimal solution $\tilde{M}^* = \lceil M^* \rceil$ with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

Algorithm 1 Sequential convex programming algorithm for Subproblem 1

1. Initialize $n = 0$, $flag_{\rm low} = 1$, $p_i^0 = p_{i,\max}$, and $M = M_{\max}$, $\forall i$.
2. Calculate the SINR $\tilde{\theta}_{ij}^0$.
3. **while** $flag_{\rm low} > 0.01$**, do**
4. $n = n + 1$;
5. Calculate a_{ij}^{n-1} and b_{ij}^{n-1} with $\tilde{\theta}_{ij}^{n-1}$, $\forall i, j$;
6. Solve problem (33), and obtain $(\mathbf{q}^n, M^n)$;
7. Update $p_i^n = 2^{q_i^n}$, $\forall i$;
8. Update SINR value $\tilde{\theta}_{ij}^n$ with $(\mathbf{p}^n, M^n)$, $\forall i, j$;
9. Calculate $\Delta\tilde{\theta}_{ij}^n = \left|\frac{\tilde{\theta}_{ij}^n - \tilde{\theta}_{ij}^{n-1}}{\tilde{\theta}_{ij}^{n-1}}\right|$, $\forall i, j$;
10. Calculate $flag_{\rm low} = \max\limits_{i,j}\left\{\Delta\tilde{\theta}_{ij}^n\right\}$.
11. **end while**

the top integer operation. Then, to deal with the non-convexity of user rate, the lower bound of the logarithmic function is employed [33], i.e.,

$$\log_2(1+\theta) \geq a\log_2\theta + b, \tag{30}$$

where $a = \frac{\theta'}{1+\theta'}$, and $b = \log_2(1+\theta') - \frac{\theta'}{1+\theta'}\log_2\theta'$. Note that when $\theta = \theta'$, the equality holds. Thus, with the transformation $q_i = \log_2 p_i$, $\mathrm{f}_1(\mathbf{p}, M)$ is lower-bounded by

$$\begin{aligned}\mathrm{f}_1(\mathbf{p}, M) \geq \tilde{\mathrm{f}}_1(\mathbf{q}, M) = &\frac{\lambda w \sum\limits_j \ln\left(\sum\limits_i x_{ij} y_{ij} \tilde{r}_{ij}(\mathbf{q}, M)\right)}{U_{\max} - U_{\min}} \\ &- \frac{\mu(1-w)\left(\sum\limits_i 2^{q_i}/\varepsilon_i + M p_a\right)}{P_{\max}},\end{aligned} \tag{31}$$

where the transmission rate from BS i to the j-th user is approximated to

$$\tilde{r}_{ij}(\mathbf{q}, M) = \begin{cases} N\left(a_{ij}\log_2\left(\mathrm{SINR}_{ij}(\mathbf{q}, M)\right) + b_{ij}\right), & i = 1, \\ a_{ij}\log_2\left(\mathrm{SINR}_{ij}(\mathbf{q}, M)\right) + b_{ij}, & i > 1, \end{cases} \tag{32}$$

and a_{ij} and b_{ij} are derived with a given SINR value $\tilde{\theta}_{ij}$. For further analysis, we have the following proposition about the approximate objective function $\tilde{\mathrm{f}}_1(\mathbf{q}, M)$.

Proposition 2: $\tilde{\mathrm{f}}_1(\mathbf{q}, M)$ is a concave function on $(\mathbf{q}, M)$[1].

Proof: See Appendix B. ■

Thus, we can solve problem (28) by finding the optimum $(\mathbf{q}^*, M^*)$ of the convex optimization problem as the following

$$\begin{aligned}&\max_{\mathbf{q}, M} \ \tilde{\mathrm{f}}_1(\mathbf{q}, M), \\ &\text{s.t. } \mathrm{C5}: M \leq M_{\max}, \\ &\quad \mathrm{C6}: 2^{q_i} \leq p_{i,\max}, \forall i, \\ &\quad \mathrm{C7}: \sum_j x_{ij} y_{ij} \tilde{r}_{ij} \leq C_{i,\rm bh}, \forall i,\end{aligned} \tag{33}$$

which can be optimally solved by the interior-point method with affordable polynomial complexity [34]. However, the approximation in (31) can be loose, which may result in unexpected results. To tighten the lower bound in (31) and make the optimal solution of (33) closer to that of (28), it is indispensable to update a_{ij} and b_{ij} in iterative manner by setting new SINR value and solving problem (33) until convergence. Therefore, we borrow the idea of sequential convex programming (SCP) [32], i.e., finding a local optimum of problem (28) by solving a sequence of convex problems (33) with new obtained SINR value, which allows us to develop an effective algorithm with low computational complexity. The above procedure is summarized in **Algorithm 1**, and its convergence and optimality are analyzed in the following proposition.

Proposition 3: Algorithm 1 monotonically improves the value of $\mathrm{f}_1(\mathbf{p}, M)$ at each iteration, and eventually converges to one solution satisfying the KKT conditions of problem (28).

Proof: See Appendix C. ■

Particularly, the optimal number of activated antennas at the MBS can be characterized in the following proposition.

Proposition 4: When $w = 1$, the optimal number of activated antennas M^* always satisfies $M^* = M_{\max}$; when $w = w_{\rm EE}$[2], if $M_{\max}$ is sufficiently large, M^* diminishes with the increase of the circuit power consumption per antenna $p_{\rm a}$.

Proof: See Appendix D. ■

2. Subproblem 2: User association and spectrum allocation

To solve problem (29), the LDD method is adopted to find the optimal user association and spectrum allocation. For convenience, we first transform problem (29) into

$$\begin{aligned}&\max_{\mathbf{x}, \mathbf{y}} \ \hat{\mathrm{f}}_2(\mathbf{x}, \mathbf{y}) = \sum_i \sum_j x_{ij} \ln(y_{ij} r_{ij}), \\ &\text{s.t. } \mathrm{C1} - \mathrm{C4}, \mathrm{C7},\end{aligned} \tag{34}$$

and the following proposition illustrates the equivalence between two problems.

Proposition 5: Problem (34) and (29) are equivalent.

Proof: See Appendix E. ■

Thus, relaxing constraints C4 and C7, the Lagrange function of (34) is obtained by

$$\begin{aligned}\mathrm{T}(\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = &\sum_i \sum_j x_{ij} \ln(y_{ij} r_{ij}) \\ &+ \sum_i \alpha_i \left(1 - \sum_j x_{ij} y_{ij}\right) \\ &+ \sum_i \beta_i \left(C_{i,\rm bh} - \sum_j x_{ij} y_{ij} r_{ij}\right),\end{aligned} \tag{35}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_I]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_I]$ are Lagrange multipliers corresponding to constraints C4 and C7, respectively. The dual function is further expressed as

$$\mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} \max\limits_{\mathbf{x}, \mathbf{y}} \mathrm{T}(\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \\ \text{s.t. } \mathrm{C1} - \mathrm{C3}, \end{cases} \tag{36}$$

and the corresponding dual problem is given by

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}} \mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta}). \tag{37}$$

In order to solve the dual problem (37), we first investigate problem (36) for given $\boldsymbol{\alpha}, \boldsymbol{\beta}$. Specifically, the Lagrange

[1]Here M is the relaxed real number.

[2]$w_{\rm EE}$ corresponds to the point of the maximum EE.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

function (35) is rearranged as

$$\mathrm{T}\left(\mathbf{x},\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta}\right)=\sum_{j}\sum_{i}x_{ij}\mathrm{S}_{ij}\left(y_{ij}\right)+\sum_{i}\alpha_{i}+\sum_{i}\beta_{i}C_{i,\mathrm{bh}}, \tag{38}$$

where

$$\mathrm{S}_{ij}\left(y_{ij}\right)=\ln\left(y_{ij}r_{ij}\right)-\alpha_{i}y_{ij}-\beta_{i}y_{ij}r_{ij}. \tag{39}$$

It is easy to find that $\mathrm{S}_{ij}\left(y_{ij}\right)$ is a concave function over y_{ij}. Thus, when the j-th user is associated with the i-th BS, i.e., $x_{ij}=1$, the optimal spectrum allocation can be calculated by taking the derivative of $\mathrm{T}\left(\mathbf{x},\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta}\right)$ with respect to y_{ij} and setting it to zero, i.e.,

$$0=\frac{\partial\mathrm{T}\left(\mathbf{x},\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta}\right)}{\partial y_{ij}}=\frac{1}{y_{ij}}-\alpha_{i}-\beta_{i}r_{ij}, \tag{40}$$

and we can obtain

$$y_{ij}^{*}=\frac{1}{\alpha_{i}+\beta_{i}r_{ij}}. \tag{41}$$

Substituting (41) into (38), to maximize $\mathrm{T}\left(\mathbf{x},\mathbf{y}^{*},\boldsymbol{\alpha},\boldsymbol{\beta}\right)$, the optimal user association for user j must be

$$x_{ij}^{*}=\begin{cases}1, & i=\arg\max\limits_{l}\mathrm{S}_{lj}\left(y_{lj}^{*}\right),\\ 0, & \text{otherwise.}\end{cases} \tag{42}$$

Up to now, the user association and spectrum allocation policy for given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ has been obtained. Thus, problem (34) can be solved via the dual problem (37) by utilizing subgradient method. Specifically, at the k-th iteration, the dual function (36) is calculated with the given $\left(\boldsymbol{\alpha}^{k},\boldsymbol{\beta}^{k}\right)$. Then, $\left(\boldsymbol{\alpha}^{k+1},\boldsymbol{\beta}^{k+1}\right)$ is updated via the subgradient method, which can be expressed as

$$\alpha_{i}^{k+1}=\alpha_{i}^{k}+v^{k}\left(\sum_{j}x_{ij}^{k}y_{ij}^{k}-1\right), \tag{43a}$$

$$\beta_{i}^{k+1}=\beta_{i}^{k}+v^{k}\left(\sum_{j}x_{ij}^{k}y_{ij}^{k}r_{ij}^{k}-C_{i,\mathrm{bh}}\right), \tag{43b}$$

where v^{k} is the diminishing step size at the k-th iteration. Note that the subgradient updates of (43) are guaranteed to converge as long as v^{k} is chosen to be sufficiently small. In this paper, $v^{k}=\frac{0.1}{\sqrt{k}}$. Since the user association variable x_{ij} is naturally discrete, there may exist a non-zero duality gap between the primal problem (34) and the dual problem (37). Nevertheless, the Lagrangian dual method often provides good solutions of the primal optimization problem [14], and the following proposition further illustrates that the duality gap is bounded.

Proposition 6: For the user association and spectrum allocation problem (34), the duality gap between the objective $\hat{f}_{2}\left(\mathbf{x},\mathbf{y}\right)$ obtained via subgradient method and the global optimum of problem (34) is bounded by $\sum\limits_{i}\alpha_{i}+\sum\limits_{i}\beta_{i}C_{i,\mathrm{bh}}-J$.

Proof: See Appendix F. ■

Observing (41)-(43), the Lagrange multipliers $(\boldsymbol{\alpha},\boldsymbol{\beta})$ can be treated as the message between users and BSs. Thus, we can solve the joint user association and spectrum allocation problem (34) distributively. The specific procedures of distributive algorithms for users and BSs are presented in **Algorithm 2** and **Algorithm 3**. Now, the solutions of problem (28) and (29) have been found, respectively. According to primal decomposition, the original problem (26) can be solved by solving problem (28) and (29) in iterative manner until convergence. As a consequence, the two-stage iterative algorithm for joint optimization of user association, spectrum allocation, power coordination, and number of activated MBS antennas is summarized in **Algorithm 4**.

Algorithm 2 Distributed User Association and Spectrum Allocation Algorithm at the j-th User

1. **if** $k=0$
2. Estimate r_{ij}, $\forall i$, via pilot signals from all BSs.
3. **end if**
4. $k=k+1$;
5. Receive $\boldsymbol{\alpha}^{k}$ and $\boldsymbol{\beta}^{k}$ transmitted from all BSs;
6. Calculate y_{ij}^{k}, $\forall i$, according to (41);
7. Calculate x_{ij}^{k}, $\forall i$, according to (42);
8. Calculate the auxiliary variable $z_{ij}^{k}=x_{ij}^{k}y_{ij}^{k}$, $\forall i$;
9. Feedback the message $z_{ij}^{k}>0$ to BS i, $\forall i$.

Algorithm 3 Distributed User Association and Spectrum Allocation Algorithm at the i-th BS

1. **if** $k=0$
2. Initialize α_{i}^{1} and β_{i}^{1}.
3. **else**
4. Receive the information $z_{ij}^{k}>0$, $\forall j$;
5. Update α_{i}^{k+1} and β_{i}^{k+1} with z_{ij}^{k} according to

$$\alpha_{i}^{k+1}=\alpha_{i}^{k}+v^{k}\left(\sum_{j}z_{ij}-1\right),$$

$$\beta_{i}^{l+k}=\beta_{i}^{k}+v^{k}\left(\sum_{i}z_{ij}r_{ij}-C_{i,\mathrm{bh}}\right);$$

6. **end if**
7. $k=k+1$;
8. Broadcast the updated multipliers α_{i}^{k} and β_{i}^{k}.

B. Adaptive Optimization of φ

Observing the linear optimization problem (27), its optimal solution can be readily obtained by

$$\varphi^{*}=\begin{cases}1, & \text{if } \lambda+\mu>1,\\ \max\left\{w\mathrm{F}_{1},(1-w)\,\mathrm{F}_{2}\right\}, & \text{otherwise.}\end{cases} \tag{44}$$

Note that $\varphi\leq 1$ always holds since the two objectives in (17) and the weight w range between 0 and 1. Thus, problem (22) can be finally solved by updating (λ,μ) like (43).

C. Summary and Complexity Analysis

As presented before, the transformed problem (24) for the tradeoff between EE and SE is solved through a series of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

Algorithm 4 Two-stage iterative algorithm

1. For any given weighting parameter w,
2. Initialize $m = 0$, $flag = 1$, $M^0 = M_{\max}$, and $p_i^0 = p_{i,\max}$.
3. **while** $flag > 0.01$**, do**
4. $m = m + 1$;
5. Calculate $(\mathbf{x}^m, \mathbf{y}^m)$ via Algorithm 2 and 3;
6. Calculate $(\mathbf{p}^m, M^m)$ via Algorithm 1;
7. Calculate $flag = \max\limits_{i,j} \left| \frac{x_{ij}^m y_{ij}^m - x_{ij}^{m-1} y_{ij}^{m-1}}{x_{ij}^{m-1} y_{ij}^{m-1}} \right|$.
8. **end while**

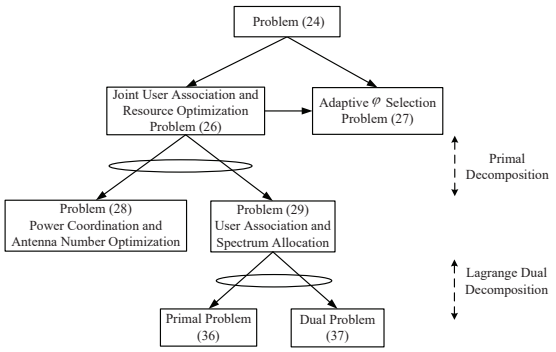

Fig. 2. The problem decomposition.

decomposition algorithms, which is summarized in Fig. 2 for clarity. Problem (24) is first separated into the joint optimization problem (26) and the adaptive φ selection problem (27). Then, with the help of primal decomposition, problem (26) is decomposed into the power coordination and the number of activated antennas optimization problem (28), and the user association and spectrum allocation problem (29). Next, with the help of SCP algorithm, Algorithm 1 is developed to solve problem (28), and the power coordination and the number of activated antennas at the MBS are obtained. Given solutions of problem (28), problem (29) is then solved via Lagrange dual decomposition, and the corresponding dual problem is problem (37). By distributively solving problem (37) via Algorithm 2 and 3, the user association and spectrum allocation matrices are determined. Finally, with the solution of problem (24) obtained by Algorithm 4, the adaptive φ selection problem (27) is easily solved.

For the user association and spectrum allocation problem, the computational complexity of our proposed distributive algorithms, i.e., Algorithm 2 and 3, is $\mathcal{O}(IJ)$ at each iteration, and the complexity of the outer Lagrangian multipliers update based on sub-gradient method is a polynomial function of the dual problem dimension, i.e., $2I$ for $\mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ [35]. Therefore, the complexity to update all multipliers is in the order of I^{ϕ}, where ϕ denotes a positive constant [36]. Thus, the overall computational complexity for the user association and spectrum allocation problem is $\mathcal{O}\left(I^{\phi+1}J\right)$.

On the other hand, the power coordination and the number of antenna optimization problem is a non-trivial problem due to the existence of inter-cell interference. Therefore, finding its global optimum with affordable complexity is challenging,

TABLE I
SIMULATION PARAMETERS

Parameters	Default value	Parameters	Default value
Cell radius	500 m	Bandwidth	20 MHz
J	50	$M_{\max}$	300
N	10	p_{a}	0.5 W
$p_{1,\max}$	43 dBm	$p_{i,\max}, i > 1$	30 dBm
Noise	-174 dBm/Hz	$\varepsilon_i, i = 1$	21.45
$\varepsilon_i, i > 1$	5.5	$p_{i,\mathrm{s}}, i = 1$	30W
$p_{i,\mathrm{s}}, i > 1$	10 W	$\max_{\mathrm{dl}}$	24
$p_{\mathrm{switch}}^{\mathrm{fiber}}$	200 W	$p_{\mathrm{switch}}^{\mathrm{micro}}$	23 W
p_{dl}	1 W	p_{ul}	2 W
P_{low}	17 W	P_{high}	42.5 W
$U_{\max}$	10 Gb/s	$C_{\mathrm{switch}}^{\mathrm{micro}}$	36 Gb/s
$C_{i,\mathrm{bh}}, i = 1$	500 bit/s/Hz	$C_{i,\mathrm{bh}}, i > 1$	20 bit/s/Hz

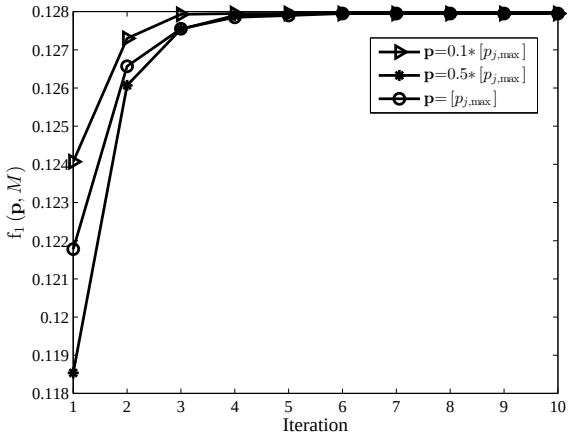

Fig. 3. The convergence procedure of Algorithm 1.

which actually requires exponential computational complexity. Consequently, a computationally-efficient algorithm is adopted to solve problem (28), i.e., Algorithm 1. At each iteration of Algorithm 1, we only need to solve a standard convex optimization problem with the requirement of polynomial complexity, which is defined as $\mathrm{c}(I)$. In addition, the required number of iteration n before convergence of Algorithm 1 is quite small, whose value is consistently below 10 in our numerical results (see Fig. 3). Furthermore, the iteration between problem (28) and (29) in Algorithm 4 also converges fast as demonstrated in the numerical study (see Fig. 4).

In summary, the overall computational complexity of Algorithm 4 for the joint user association and resource optimization problem (26) is $\mathcal{O}\left(I^{\phi+1}J\mathrm{c}(I)\right)$. Hence, the proposed algorithm only requires affordable polynomial complexity to find the solution of the original problem (22).

V. SIMULATION RESULTS

In the simulations, we consider a two-tier HetNet, where one massive MIMO macro BS is in the center, and three pico-BSs are symmetrically placed along the circle with radius of 200 m. Users are randomly located in the cell. The pathloss between BSs and users is modelled as $128.1 + 37.6\log_{10}d$ (km), and the standard derivation of shadow fading is 8 dB. The other simulation parameters are shown in Table I.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

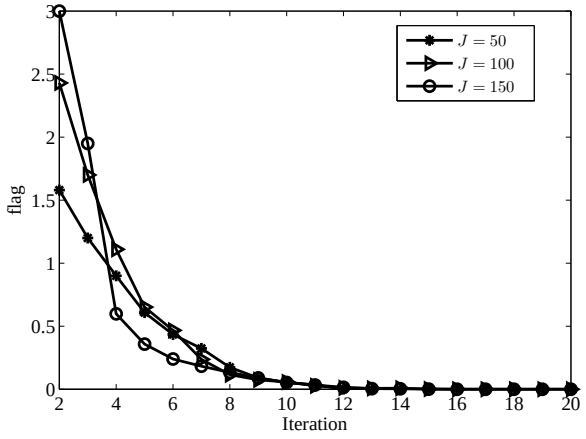

Fig. 4. The convergence procedure of Algorithm 4.

A. Performance of Convergence

We first provide insight on the convergence property of Algorithm 1, and Fig. 3 presents the objective function $\mathrm{f}_1(\mathbf{p}, M)$ of problem (28) versus iterations. Specifically, three curves with different initial transmit power are plotted in Fig. 3, and it can be observed that $\mathrm{f}_1(\mathbf{p}, M)$ increases consistently and converges to the maximum value within about 6 iterations. In addition, we can find that the choice of initial points has negligible effect on the convergence procedure of Algorithm 1. These numerical results also keep in line with Proposition 3 proved in Section IV.

Since our proposed algorithm for joint optimization of user association and resource optimization (i.e., Algorithm 4) is a two-stage iterative algorithm, it is significant to demonstrate its convergence property. Fig. 4 plots the flag value versus iterations under different number of users, where

$$flag^m = \left\| \mathbf{z}^m - \mathbf{z}^{m-1} \right\|_1, \tag{45}$$

and m represents the iteration number. When the flag value turns into zero, the obtained user association and spectrum allocation solution stays the same between two iterations, and thus the optimized transmit power and the number of antenna are also stable, which indicates the convergence of Algorithm 4. From Fig. 4, it can be observed that Algorithm 4 converges quickly, and the increase of users has slight impact on the speed of convergence.

B. Performance Comparisons

We first compare the performance of the system considered in this paper, i.e., massive MIMO enabled HetNets, with that of two special cases where only one massive MIMO enabled marco BS or only single-antenna pico cells are deployed. Note that the parameter settings for the two cases also keep in line with those in Table I. It can be observed from Fig. 5 that the combination of HetNets and massive MIMO contributes a much greater level of performance improvement in EE and SE as compared to single form of the network.

Specifically, compared to the massive MIMO enabled two-tier HetNet, the system with only one marco BS requires higher power consumption to transmit data to users located in the cell edge, which may result in a lower EE. Accordingly, as the transmit power of the macro BS is upper-bounded, the maximum achievable SE for the case of only one marco BS is also reduced compared to that in two-tier HetNets. On the other hand, the case of only single-antenna pico BSs also achieves a worse EE-SE tradeoff in contrast to the system considered in this paper. Since the transmit powers of pico BSs are much lower than the macro BS, the coverage area of each pico BS is limited. In this case, there may exist users which can not be well served by pico BSs. Moreover, as a single antenna is deployed in each pico BS, no performance gain from spatial multiplexing can be obtained.

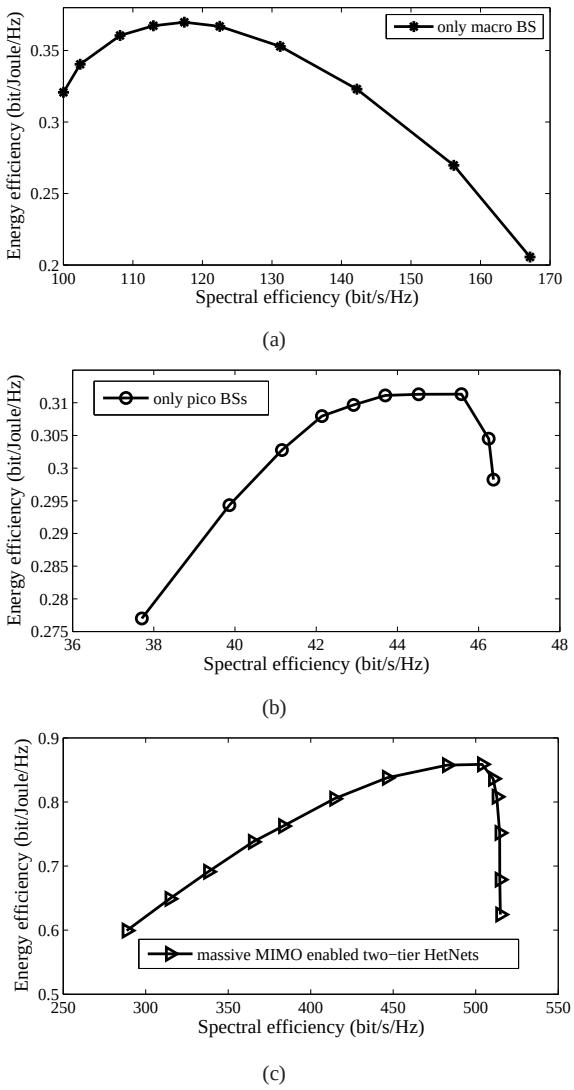

Fig. 5. (a) Energy efficiency *v.s.* spectral efficiency for the system with only one macro BS. (b) Energy efficiency *v.s.* spectral efficiency for the system with only pico BSs. (c) Energy efficiency *v.s.* spectral efficiency for the massive MIMO enabled HetNets.

Furthermore, to demonstrate the optimality of our 'proposed algorithm', we first compare it with the exhaustive search method which serves as the benchmark. Due to the exponential computational complexity requirement of exhaustive search, we consider a small-scale problem of $K = 8$, $N = 4$ as an example. It can be observed in Fig. 6 that the performance of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

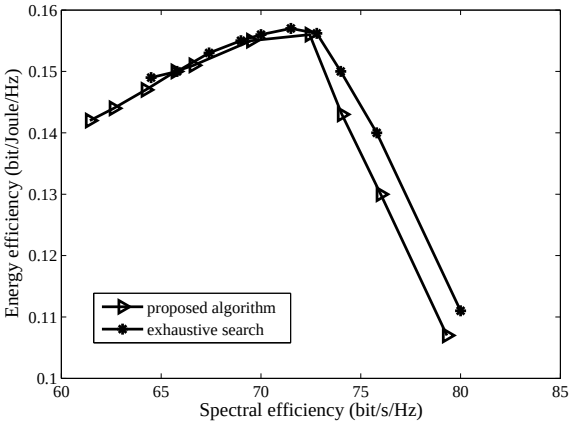

Fig. 6. Energy efficiency *vs.* spectral efficiency for the proposed algorithm and the exhaustive search method.

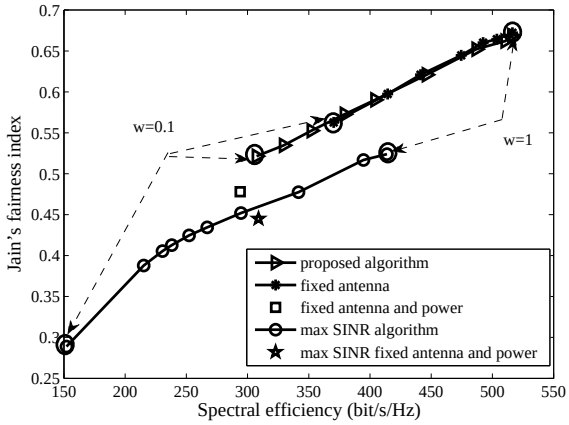

Fig. 8. Jain's fairness *vs.* spectral efficiency for different algorithms.

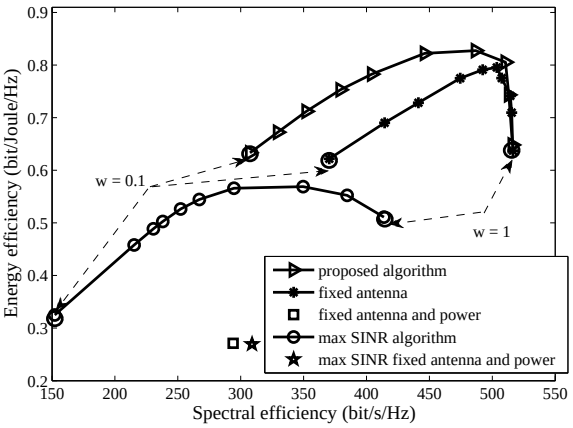

Fig. 7. Energy efficiency *vs.* spectral efficiency for different algorithms.

proposed algorithm is close to that of the exhaustive search method in terms of the EE-SE tradeoff.

Based on the default parameter settings in Table I, we then compare the proposed algorithm with other four algorithms: 1) 'fixed antenna' is almost in line with 'proposed algorithm', while the antenna number of the MBS is predefined as the number of maximum available antennas; 2) 'fixed antenna and power' denotes that both the the number of activated antennas and transmit power are fixed as maximum values; 3) 'max SINR fixed antenna and power' involves a different user association method that each user chooses the BS with the highest SINR, where spectrum allocation is also obtained by Lagrange dual decomposition; 4) 'max SINR algorithm' further includes the optimization of power coordination and antenna number by adopting our proposed Algorithm 1.

By adjusting the weighting parameter w from 0.1 to 1, Fig. 7 first presents the tradeoff performance between EE and SE for different algorithms. For 'proposed algorithm', 'fixed antenna', and 'max SINR algorithm', the system EE first increases to its peak and then decreases to a low level with the increase of SE. When SE is low, the increase of SE is much faster than that of the total power consumption. Thus, EE grows as SE increases. However, there is a maximum point for EE. After this point, the fixed circuit power does not dominate any longer and the increase of transmit power and antenna power consumption has a significant effect on the total power consumption. Under this circumstance, the growing of SE becomes slower than the total power consumption, which is due to the diminishing gradient of the logarithmic rate function. Furthermore, for a given SE, the optimization of the number of MBS antennas can significantly improve the system EE according to the comparison between 'proposed algorithm' and 'fixed antenna'. Besides, it can be observed in Fig. 7 that our proposed user association algorithm achieves significant improvement on both EE and SE in comparison with the basic 'max SINR' user association algorithm as our algorithm maximizes EE and SE simultaneously. By contrast, since the transmit power of each BS and the number of activated MBS antennas are fixed for 'fixed antenna and power' and 'max SINR fixed antenna and power', their performances are presented via single points, whose corresponding EE is much lower than the other three algorithms for the same SE.

Then, the performance of rate fairness for different algorithms is presented in Fig. 8, and the rate fairness is measured by Jain's fairness index, i.e.,

$$Index = \left(\sum_j R_j\right)^2 \Big/ \left(J \sum_j R_j^2\right). \tag{46}$$

which ranges from $\frac{1}{J}$ to 1. If the index equals 1, this indicates that all users have the same data rate and the system achieves absolute fairness. However, the index will gradually drop to $\frac{1}{J}$ with the increase of rate disparity among users. As presented in Fig. 8, for a given SE, the proposed user association algorithm can achieve higher level of fairness compared to 'max SINR' algorithm, since the logarithmic utility function is maximized in our algorithm instead of the system SE. In addition, it is reasonable to find that the optimization of antenna number has no effect on the rate fairness. To conclude, our proposed user association algorithm is superior to 'max SINR' algorithms in terms of rate fairness.

C. Characteristic of the Number of Activated Antennas

The performance of EE versus the optimal number of activated antennas at the MBS is illustrated in Fig. 9. We

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

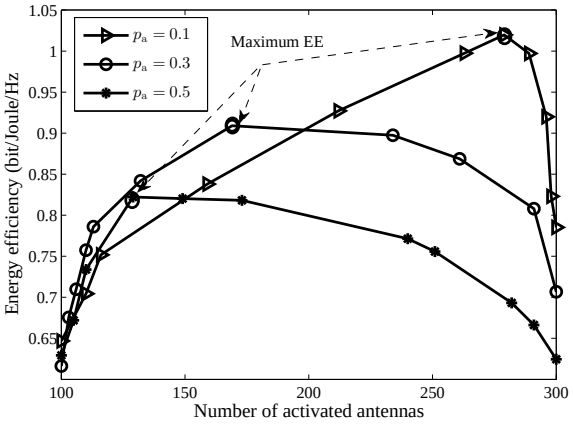

Fig. 9. Energy efficiency *vs.* the number of activated antennas.

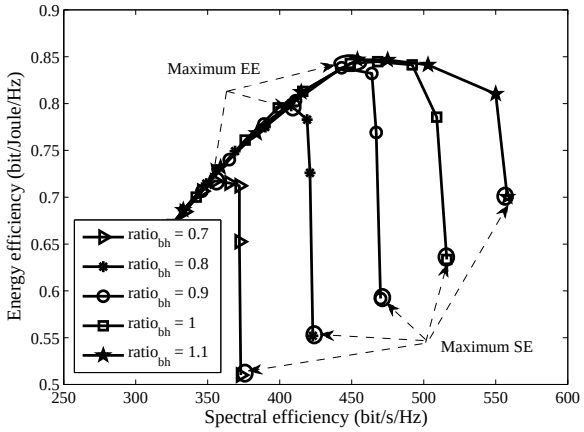

Fig. 10. Energy efficiency *vs.* spectral efficiency for different levels of backhaul capacity.

can find that all the three curves follow the same pattern, where the system EE first increases and then decreases as the optimal number of activated antennas goes up. This is because all the three selective values of $p_{\rm a}$ are sufficiently small so that the optimal number of activated antennas corresponding to the maximum EE is smaller than the number of equipped antennas $M_{\max}$. In particular, the maximum EE declines with the increase of $p_{\rm a}$, which can be easily understood by the fact that the more power each antenna consumes, the less energy-efficient the system is. Besides, as presented in Fig. 9, lowering the power consumption per antenna makes the optimal number of activated number corresponding to the maximum EE closer to the maximum value.

D. Performance Bottleneck: Backhaul Capacity Constraint

The tradeoff between EE and SE for different backhaul capacities is plotted in Fig. 10, and the parameter $ratio_{\rm bh}$ represents the ratio between the available backhaul capactiy and the default value in Table I. Observing the second part of curves where EE decreases with SE, we can find that more SE is achieved with larger backhaul capacity for the same EE, which indicates that the backhaul capacity constraint is indeed a bottleneck for the network performance. Particularly, with the increase of the backhaul capacity, the maximum SE increases, and its corresponding EE also rises up. This is because the maximum SE is bounded by the sum of the backhaul capacity for each BS. When the backhaul capacity rises up, more data can be transmitted between BSs and users. By contrast, the maximum EE first increases and then remains stable with the increase of the backhaul capacity. When the system EE reaches its peak, the data rates of BSs can be lower than its corresponding backhaul capacities with the relaxation of the backhaul constraint. In this case, the maximum EE can not be improved by raising the backhaul capactiy.

E. Impact of the Number of Users

Furthermore, we investigate the performance of rate fairness for different number of users, where the proposed algorithm is compared with the basic 'max SINR algorithm'. As observed in Fig. 11, when the number of users increases, the fairness

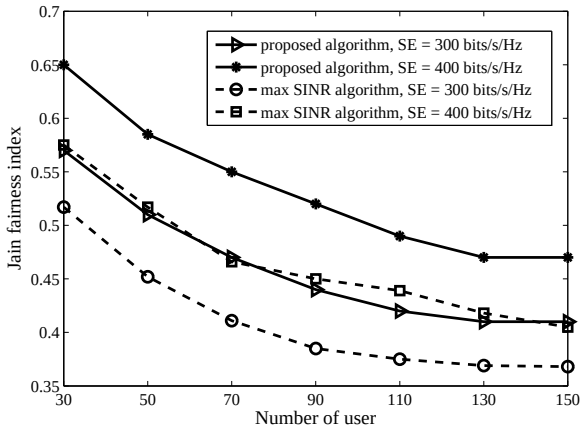

Fig. 11. Jain's fairness *vs.* the number of users for different algorithms.

index gradually decreases, and all curves follow the same pattern. This can be explained by the fact that the same rate fairness among more users is more difficult to achieve. Besides, for a given SE, our proposed algorithm consistently achieves higher rate fairness among users compared to 'max SINR algorithm'. Finally, the impact of the number of users on the tradeoff between SE and EE for the proposed algorithm is captured in Fig. 12. It can be observed that for a given SE, the system EE increases with the number of users at a diminishing speed, which indicates that it consumes less power to achieve the same level of SE when the number of users rises up. This is mainly because of the effect of multiuser diversity, and the gain decreases gradually as the number of users increases.

VI. Conclusion

In this paper, we studied the tradeoff between EE and SE while ensuring proportional rate fairness in massive MIMO enabled HetNets, and the MOO problem was formulated to maximize EE and SE simultaneously while ensuring proportional rate fairness, where the backhaul capacity constraint and power consumption were both taken into account. A computational-efficient algorithm was further proposed to comprehensively optimize user association, spectrum allocation, power coordination and number of activated antennas,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

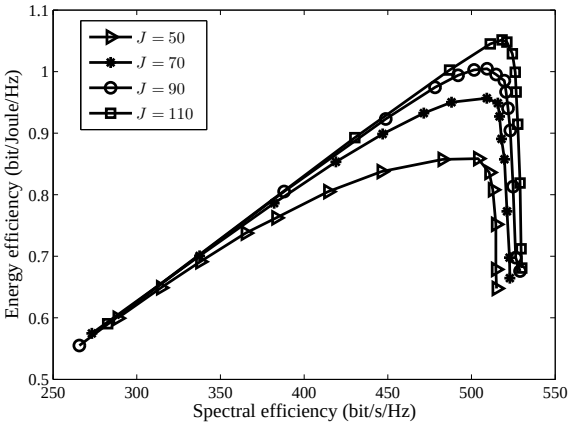

Fig. 12. Energy efficiency *vs.* spectral efficiency for different numbers of users.

which only required affordable polynomial complexity. The convergence and optimality of the proposed algorithm were verified via both theoretical analysis and numerical results. As presented in simulation results, our proposed algorithm converged fast within several iterations. In comparison with 'max SINR' algorithm, the proposed algorithm significantly improved the performance of the EE-SE tradeoff and rate fairness among users, which demonstrated its effectiveness. We further investigated the characteristics of the optimal number of activated antennas via theoretical analysis and numerical results, and also demonstrated numerically that the backhaul capacity constraint was a bottleneck for the network performance.

APPENDIX A

Proof of Proposition 1: Define that

$$\mathrm{g}_1(\mathbf{x},\mathbf{y},\mathbf{p},M)=\frac{U_{\max}-\mathrm{U}(\mathbf{x},\mathbf{y},\mathbf{p},M)}{U_{\max}-U_{\min}}, \tag{47a}$$

$$\mathrm{g}_2(\mathbf{x},\mathbf{y},\mathbf{p},M)=\frac{P_{\mathrm{tot}}(\mathbf{x},\mathbf{y},\mathbf{p},M)}{P_{\max}}, \tag{47b}$$

$$\mathrm{G}(\mathbf{x},\mathbf{y},\mathbf{p},M)=\max\{w\mathrm{g}_1,(1-w)\,\mathrm{g}_2\}. \tag{47c}$$

For any given $w\in[0,1]$, since $(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*)$ is the unique optimal solution of problem (21), we have

$$\mathrm{G}(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*)\le\mathrm{G}(\mathbf{x},\mathbf{y},\mathbf{p},M), \tag{48}$$

for all $(\mathbf{x},\mathbf{y},\mathbf{p},M)$ satisfying constraints C1-C7.

Now we suppose that $(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*)$ is not Pareto optimal for the original MOO problem (16). Thus, there must exist another solution $(\mathbf{x}',\mathbf{y}',\mathbf{p}',M')$ of problem (21) that satisfies

$$\mathrm{g}_i(\mathbf{x}',\mathbf{y}',\mathbf{p}',M')\le\mathrm{g}_i(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*),\ \forall\, i=1,2, \tag{49}$$

and there is at least one $j\in\{1,2\}$ such that $\mathrm{g}_j(\mathbf{x}',\mathbf{y}',\mathbf{p}',M')<\mathrm{g}_j(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*)$. Under this circumstance, we have

$$\mathrm{G}(\mathbf{x}',\mathbf{y}',\mathbf{p}',M')\le\mathrm{G}(\mathbf{x}^*,\mathbf{y}^*,\mathbf{p}^*,M^*), \tag{50}$$

which contradicts with the uniqueness assumption. Therefore, the proposition is proved.

APPENDIX B

Proof of Proposition 2: For simplicity of the following analysis, we first introduce two coefficients:

$$d_i=\begin{cases}N, & i=1,\\ 1, & i>1,\end{cases}\qquad e_i=\begin{cases}\frac{M-N+1}{N}, & i=1,\\ 1, & i>1,\end{cases}$$

and the approximate data rate for user j associated with BS i can be further expressed as

$$\tilde{r}_{ij}(\mathbf{q},M)=d_i\cdot c_{ij}(\mathbf{q},M), \tag{51}$$

where

$$\begin{aligned}c_{ij}(\mathbf{q},M)=a_{ij}\left(q_i-\log_2\left(\sum_{l\neq i}2^{q_l}g_{lj}+\sigma_j^2\right)\right)\\ +a_{ij}\log_2(e_ig_{ij})+b_{ij}.\end{aligned} \tag{52}$$

$c_{ij}(\mathbf{q},M)$ is a concave function over $(\mathbf{q},M)$, which is due to the concavity of logarithm function and the convexity of the log-sum-exp function [34]. Furthermore, the first term of $\tilde{\mathrm{f}}_1(\mathbf{q},M)$ in (31) is in the form of composite functions $\log(c_{ij}(\mathbf{q},M))$. As the logarithmic function $\log(x)$ is an increasing concave function on x, the first term of $\tilde{\mathrm{f}}_1(\mathbf{q},M)$ is a concave function over $(\mathbf{q},M)$. On the other hand, the second term of $\tilde{\mathrm{f}}_1(\mathbf{q},M)$ is also a concave function because of the convexity of exponential function. Since $\tilde{\mathrm{f}}_1(\mathbf{q},M)$ is actually a sum of concave terms, it is straightforward to conclude that $\tilde{\mathrm{f}}_1(\mathbf{q},M)$ is concave over $(\mathbf{q},M)$.

APPENDIX C

Proof of Proposition 3: Let $\mathbf{q}^n=\log_2(\mathbf{p}^n)$ and M^n denote the optimal solution of problem (33) after the n-th iteration in Algorithm 1, and we have the following inequality

$$\begin{aligned}\mathrm{f}_1(\mathbf{p}^n,M^n)&\overset{(a)}{=}\tilde{\mathrm{f}}_1(\mathbf{q}^n,M^n)\overset{(b)}{\le}\tilde{\mathrm{f}}_1\left(\mathbf{q}^{n+1},M^{n+1}\right)\\ &\overset{(c)}{\le}\mathrm{f}_1\left(\mathbf{p}^{n+1},M^{n+1}\right),\end{aligned} \tag{53}$$

where the equality (a) is due to the fact that a_{ij} and b_{ij} are calculated with the given $\theta_{ij}^n=\mathrm{SINR}_{ij}(\mathbf{p}^n,M^n)$ and thus the relaxation (31) is tight; the inequality (b) holds since problem (33) is convex and $\left(\mathbf{q}^{n+1},M^{n+1}\right)$ is the global optimal solution; the inequality (c) is valid because $\tilde{\mathrm{f}}_1\left(\mathbf{q}^{n+1},M^{n+1}\right)$ is the lower bound of $\mathrm{f}_1\left(\mathbf{p}^{n+1},M^{n+1}\right)$ as illustrated in (31). Consequently, the value of $\mathrm{f}_1(\mathbf{p},M)$ increases at each iteration. Since $\mathrm{f}_1(\mathbf{p},M)$ is upper-bounded by the constraints of the maximum transmit power and the maximum number of antennas, Algorithm 1 must converge.

Assume that $(\mathbf{q}^*,M^*)$ is the obtained solution when Algorithm 1 converges, which must satisfy the KKT conditions of problem (33). When Algorithm 1 converges, the objectives in problem (28) and (33) have the same value, i.e., $\mathrm{f}_1(\mathbf{p}^*,M^*)=\tilde{\mathrm{f}}_1(\mathbf{q}^*,M^*)$. Besides, problem (28) and (33) actually have the same constraints. Hence, $(\mathbf{p}^*,M^*)$ must satisfy the KKT conditions of (28).

APPENDIX D

Proof of Proposition 4: When the weighting parameter w equals to 1, problem (28) turns into the SE maximization

problem. Because $\mathrm{f}_1(\mathbf{p}, M)$ increases monotonously with M when $w = 1$, the optimal number of activated antennas satisfies $M^* = M_{\max}$.

On the other hand, when $w = w_{\mathrm{EE}}$, problem (28) becomes the EE maximization problem. Thus, taking the derivative of $\mathrm{f}_1(\mathbf{p}, M)$ with respect to M yields

$$\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M} = A \sum_j \frac{B_j}{(1+\mathrm{SINR}_{1k}) \sum\limits_i x_{ij} y_{ij} r_{ij}} - C p_{\mathrm{a}}, \quad (54)$$

where $A = \frac{\lambda w_{\mathrm{EE}}}{U_{\max} - U_{\min}}$, $B_j = \frac{x_{1j} y_{1j} N p_1 g_{1j}}{\sum\limits_{i \neq 1} p_i g_{ij}}$, and $C = \frac{\mu(1 - w_{\mathrm{EE}})}{P_{\max}}$ are non-negative coefficients independent of M. As observed in (54), $\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M}$ decreases with M. Thus, when p_{a} is relatively small, there must exist $M_0 < M_{\max}$ that satisfies $\left.\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M}\right|_{M=M_0} > 0$. Since $M_{\max}$ is sufficiently large, there exists $M_1 \leq M_{\max}$ which satisfies $\left.\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M}\right|_{M=M_1} < 0$. Hence, $\mathrm{f}_1(\mathbf{p}, M)$ is maximized when $\left.\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M}\right|_{M=M^*} = 0$, and we have $M_0 < M^* < M_1 \leq M_{\max}$, where M^* denotes the optimal number of activated antennas at the MBS. Furthermore, as $\frac{\partial \mathrm{f}_1(\mathbf{p}, M)}{\partial M}$ decreases with M, it is easily concluded that M^* diminishes with the increase of p_{a}.

Appendix E

Proof of Proposition 5: According to the constraints C1 and C2, each user can be associated with at most one BS. As a consequence, for any $\mathbf{x}, \mathbf{y}$ satisfying the constraints of problem (29), it is straightforward to obtain

$$\sum_i x_{ij} \ln(y_{ij} r_{ij}) = \ln\left(\sum_i x_{ij} y_{ij} r_{ij}\right). \quad (55)$$

Furthermore, we have

$$\sum_j \sum_i x_{ij} \ln(y_{ij} r_{ij}) = \sum_j \ln\left(\sum_i x_{ij} y_{ij} r_{ij}\right), \quad (56)$$

which means that problem (29) and (34) have equal objectives. Since their constraints are also identical, we can conclude that the two problems are equivalent.

Appendix F

Proof of Proposition 6: Assume that $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are the optimized Lagrange multipliers at the convergence of subgradient method, and $(\mathbf{x}, \mathbf{y})$ is the corresponding solution obtained from (41) and (42). Substituting (41) and (42) into (36), the dual function can be expressed as

$$\mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_j \max_i \ (\ln(y_{ij} r_{ij}) - 1) + \sum_i \alpha_i + \sum_i \beta_i C_{i,\mathrm{bh}}. \quad (57)$$

Thus, we have

$$\begin{aligned} \hat{f}_2(\mathbf{x}, \mathbf{y}) &= \sum_i \sum_j x_{ij} \ln(y_{ij} r_{ij}) = \sum_j \max_i \ (\ln(y_{ij} r_{ij})) \\ &= \mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \left(\sum_i \alpha_i + \sum_i \beta_i C_{i,\mathrm{bh}} - J\right). \end{aligned} \quad (58)$$

Then, we suppose that $(\mathbf{x}^*, \mathbf{y}^*)$ is the global optimal solution for problem (34). According to the weak duality, it always holds that

$$\mathrm{J}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq \hat{f}_2(\mathbf{x}^*, \mathbf{y}^*), \quad (59)$$

and from (58) we prove that

$$\hat{f}_2(\mathbf{x}, \mathbf{y}) \geq \hat{f}_2(\mathbf{x}^*, \mathbf{y}^*) - \left(\sum_i \alpha_i + \sum_i \beta_i C_{i,\mathrm{bh}} - J\right). \quad (60)$$

References

[1] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 94–101, Apr. 2016.

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

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

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] J. B. Rao and A. O. Fapojuwo, "An analytical framework for evaluating spectrum/energy efficiency of heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3568–3584, May 2016.

[6] D. Lopez-Perez, I. Guvenc, G. D. la Roche *et al.*, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, Jun. 2011.

[7] B. Yang, G. Mao, X. Ge, and T. Han, "A new cell association scheme in heterogeneous networks," in *2015 IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 5627 – 5632.

[8] H. Zhu, S. Wang, and D. Chen, "Energy-efficient user association for heterogenous cloud cellular networks," in *Proc. 2012 IEEE Globecom Workshops*, Dec. 2012, pp. 273–278.

[9] Y. Kwon, T. Hwang, and X. Wang, "Energy-efficient transmit power control for multi-tier MIMO HetNets," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2070–2086, Oct. 2015.

[10] C. Yang, J. Li, Q. Ni, A. Anpalagan, and M. Guizani, "Interference-aware energy efficiency maximization in 5G ultra-dense networks," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 728–739, Feb. 2017.

[11] C. Yang, J. Li *et al.*, "Joint power coordination for spectral-and-energy efficiency in heterogeneous small cell networks: A bargaining game-theoretic perspective," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 2, pp. 1364–1376, Feb. 2016.

[12] B. Zhuang, D. Guo, and M. L. Honig, "Energy-efficient cell activation, user association, and spectrum allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 823–831, Apr. 2016.

[13] A. He, L. Wang, M. Elkashlan, Y. Chen, and K.-K. Wong, "Spectrum and energy efficiency in massive MIMO enabled HetNets: A stochastic geometry approach," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2294–2297, Dec. 2015.

[14] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. Elkashlan, "Distributed energy efficient fair user association in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1770–1773, Oct. 2015.

[15] Y. Hao, Q. Ni, H. Li, and S. Hou, "Energy and spectral efficiency tradeoff with user association and power coordination in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2091–2094, Oct. 2016.

[16] H. H. Yang, G. Geraci, and T. Q. S. Quek, "Energy-efficient design of MIMO heterogeneous networks with wireless backhaul," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4914–4927, Jul. 2016.

[17] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: challgences and research advances," *IEEE Network*, vol. 28, no. 6, pp. 6–11, Nov. 2014.

[18] J. Tang, W. P. Tay, T. Q. S. Quek, and B. Liang, "System cost minimization in cloud RAN with limited fronthaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3371–3384, May 2017.

[19] K. Guo, M. Sheng, J. Tang, T. Q. S. Quek, and Z. Qiu, "Exploiting hybrid cooperative transmission and computation provisioning for green C-RAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 52, pp. 4063–4076, Dec. 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2730867, IEEE Transactions on Communications

[20] H. Beyranvand, W. Lim, M. Maier, C. Verikoukis, and J. A. Salehi, "Backhaul-aware user association in FiWi enhanced LTE-A heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 2992–3003, Jun. 2015.
[21] N. Wang, E. Hossain, and 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.*, vol. 15, no. 5, pp. 3251–3268, May 2016.
[22] Q. Han, B. Yang, G. Miao, C. Chen, X. Wang, and X. Guan, "Backhaul-aware user association and resource allocation for energy-constrained HetNets," *IEEE Trans. Veh. Technol.*, pp. 1–13, 2016.
[23] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1436–1449, Apr. 2013.
[24] D. Bethanabhotla, O. Y. Bursalioglu, and H. C. Papadopoulos, "User association and load balancing for cellular massive MIMO," in *Proc. IEEE Info. Theory and Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2014, pp. 1–10.
[25] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
[26] L. Suarez, M. A. Bouraoui, M. A. Mertah, M. Morvan, and L. Nuaymi, "Energy efficiency and cost issues in backhaul architectures for high data-rate green mobile heterogeneous networks," in *2015 IEEE PIMRC*, Hongkong, China, Aug. 2015, pp. 1563 – 1568.
[27] P. Monti, S. Tombaz, L. Wosinska, and J. Zander, "Mobile backhaul in heterogeneous network deployments: Technology options and power consumption," in *2012 14th ICTON*, 2012, pp. 1–7.
[28] Z. Song, Q. Ni, K. Navaie, S. Hou, S. Wu, and X. Sun, "On the spectral-energy efficiency and rate fairness tradeoff in relay-aided cooperative OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6342–6355, Sep. 2016.
[29] O. Amin, E. Bedeer, M. H. Ahmed, and O. A. Dobre, "Energy efficiency-spectral efficiency tradeoff: A multiobjective optimization approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1975–1981, Apr. 2016.
[30] H. Shi, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 1, pp. 5–24, Jan. 2014.
[31] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisc. Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.
[32] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Jul. 1978.
[33] J. Papandriopoulos and J. S. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, Jun. 2006, pp. 3270–3275.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[35] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.
[36] W. Dang, M. Tao, H. Mu, and J. Huang, "Subcarrier-pair based resource allocation cooperative multi-relay OFDM systems," *IEEE Trans. Wirel. Commun.*, vol. 9, no. 5, pp. 1640–1649, May 2010.

Yuanyuan Hao received her B.Sc. degree in Information Engineering, from Beijing Institute of Technology, Beijing, China, in 2014. She is currently working toward the Ph.D. degree with School of Information and Electronics, Beijing Institute of Technology, Beijing, China. From November 2016, she has been a visiting student in School of Computing and Communications, InfoLab21, Lancaster University, Lancaster, UK. Her research interests include radio resource management, green communications, multi-tier heterogeneous networks, massive MIMO and D2D communications.

Qiang Ni (M'04-SM'08) received the B.Sc., M.Sc., and Ph.D. degrees from Huazhong University of Science and Technology, China, all in engineering. He is a Professor and the Head of Communication Systems Group, School of Computing and Communications, Lancaster University, InfoLab21, Lancaster, U.K. His main research interests lie in the area of future generation communications and networking, including green communications and networking, cognitive radio network systems, heterogeneous networks, small cell and ultra dense networks, 5G, SDN, cloud networks, energy harvesting, wireless information and power transfer, IoTs and vehicular networks. He has published over 180 papers in these areas. He was an IEEE 802.11 Wireless Standard Working Group Voting Member and a contributor to the IEEE Wireless Standards.

Hai Li (M'05) received the B.Sc. and Ph.D. degrees from Beijing Institute of Technology, Beijing, China, in 1997 and 2002. He is currently an Associate Professor with School of Information and Electronics, Beijing Institute of Technology. His research interests include signal processing, protocol engineering, and wireless communications.

Shujuan Hou received the B.Sc., M.Sc., and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, all in signal and information processing. She is currently an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology. Her main research interests are digital signal processing and radio resource management in wireless communications.