# Joint Optimization of User Association, Subchannel Allocation, and Power Allocation in Multi-Cell Multi-Association OFDMA Heterogeneous Networks

Feng Wang, Wen Chen, *Senior Member, IEEE*, Hongying Tang, and Qingqing Wu, *Student Member, IEEE*

*Abstract*—Heterogeneous network is a novel network architecture proposed in long-term-evolution, which highly increases the capacity and coverage compared with the conventional networks. However, in order to provide the best services, appropriate resource management must be applied. In this paper, we consider the joint optimization problem of user association, subchannel allocation, and power allocation for downlink transmission in multi-cell multi-association orthogonal frequency division multiple access heterogeneous networks. To solve the optimization problem, we first divide it into two subproblems: 1) user association and subchannel allocation for fixed power allocation and 2) power allocation for fixed user association and subchannel allocation. Subsequently, we obtain a locally optimal solution for the joint optimization problem by solving these two subproblems alternately. For the first subproblem, we derive the globally optimal solution based on graph theory. For the second subproblem, we obtain a Karush–Kuhn–Tucker optimal solution by a low complexity algorithm based on the difference of two convex functions approximation method. In addition, the multi-antenna receiver case and the proportional fairness case are also discussed. Simulation results demonstrate that the proposed algorithms can significantly enhance the overall network throughput.

*Index Terms*—Heterogeneous networks, multi-association, user-association, subchannel allocation, power allocation.

## I. INTRODUCTION

A S A NOVEL candidate technology in 5th generation (5G) wireless networks, heterogeneous network is proposed to increase network throughput and coverage, and reduce energy consumption. In the homogeneous networks, the transmission power and coverage of each BS is similar. Nevertheless, macro and micro base stations (BSs) with different transmission power and processing capability are deployed in heterogeneous networks to meet various communication demands [1]. The micro BSs include picocell BSs, femtocell BSs [2] and relays. Picocell BSs and femtocell BSs are connected to the network by wired backhaul and relays are connected to the network by wireless backhaul. A heterogeneous architecture brings in a rich topology, but the deployment of different low power BSs over existing macro BSs coverage causes severe interference, which poses new challenges on interference management and resource allocation.

Different kinds of user association schemes have been discussed for heterogeneous networks. To balance the traffic load between the BSs, range-expansion based scheme is proposed in [3] and [4], where a bias factor is used to balance the load in macro BSs and micro BSs. In [5], WU *et al.* propose a novel user association model with dual connectivity and constrained backhaul. In [6], Siddique *et al.* propose a channel-access-aware user association scheme to enhance the spectral efficiency and achieve traffic load balancing. A Voronoi-based user association scheme is proposed in [7] to maximize the number of admitted users. Subchannel allocation is another hot issue in Orthogonal Frequency Division Multiple Access (OFDMA) heterogenous networks. In general cases, each subchannel will be allocated to the user equipments (UEs) that has the best channel condition. Various methods are investigated for subchannel allocation, such as the worst user first (WUF) Greedy algorithm [8] and the proportional fair method [9]. A few recent literatures investigate joint optimization of user association and subchannel allocation in multi-cell OFDMA networks, such as [10] and [11]. In [10], the user association and subchannel allocation problem is solved separately. Chen *et al.* [11] propose an iterative method which only achieves a suboptimal solution. To our best knowledge, the optimal solution for the joint user association and subchannel allocation problem in multi-cell OFDMA networks has not been achieved yet.

As the spectrum becomes rare and expensive, the co-channel deployment (CCD) scheme, where all BSs operate on the full set of subchannels, are highly desirable [12]. Some works focus on the resource allocation when the CCD is consid-

ered. In [13], a distributed power allocation method based on iterative water-filling (IW) is presented. In [14], Lopez-Perez *et al.* propose a dynamic algorithm to jointly allocate subchannel and power to mitigate inter-cell interference. In [15], Tabassum *et al.* investigate the subchannel and power allocation in high signal-to-interference-noise-ratio (SINR) regime. In [16], Kim *et al.* propose the joint subchannel allocation and power control based on polyblock outer approximation (JSPPA) algorithm to get the optimal solution of the joint subchannel and power allocation problem. However, its computational complexity increases exponentially with the number of UEs and subchannels. In addition, all the literatures above assume that one user can only be connected to one BS in each time slot. Recently, Ghimire *et al.* assume that one user can be connected to multiple BSs, which is referred to as multi-association [17]. Intuitively, this will yield higher throughput since it allows for higher flexibility. However, the authors adopt the exhaustive search for the user association and subchannel allocation problem in CCD networks.

In this paper, we also consider multi-association scenario. In particular, we assume that one user can be connected to different BSs on different subchannels. We maximize the weighted sum-rate for downlink transmission in multi-cell OFDMA heterogeneous networks. Our main contributions are summarized as follows:

- We develop a more general mathematical model considering both multi-association and CCD. To the best of our knowledge, this is the first work that jointly optimizes user association, subchannel allocation, and power allocation for our considered system.
- To solve this optimization problem, we first divide it into two subproblems. The first one is joint optimization of user association and subchannel allocation for a fixed power allocation. We transform it into an equivalent bipartite matching problem [18] and obtain the globally optimal solution by Hungarian algorithm [19]. As far as we know, this is the first optimal solution for the joint problem in multi-cell networks.
- The second subproblem is power allocation for fixed user association and subchannel allocation, which is transformed into a series of convex problems based on the difference of two convex functions approximation (DCA) method. Then a low complexity algorithm is proposed to solve these problems. Moreover, we prove that the DCA method converges to a Karush-Kuhn-Tucker (KKT) optimal point under some mild conditions.
- Based on that, we obtain a locally optimal solution for the joint optimization problem by an alternating optimization method. Simulation results show that our algorithm can achieve significant performance gain compared with the existing algorithms. We also find that the iteration number of the conventional Lagrange Dual method [20], [21] is nearly 9 times larger than that of our low complexity method.
- We also extend the problem into the multi-antenna receiver case and the proportional fairness case, respectively. Our algorithm performs well when the UEs apply multi-antenna receivers, and it helps achieve a good
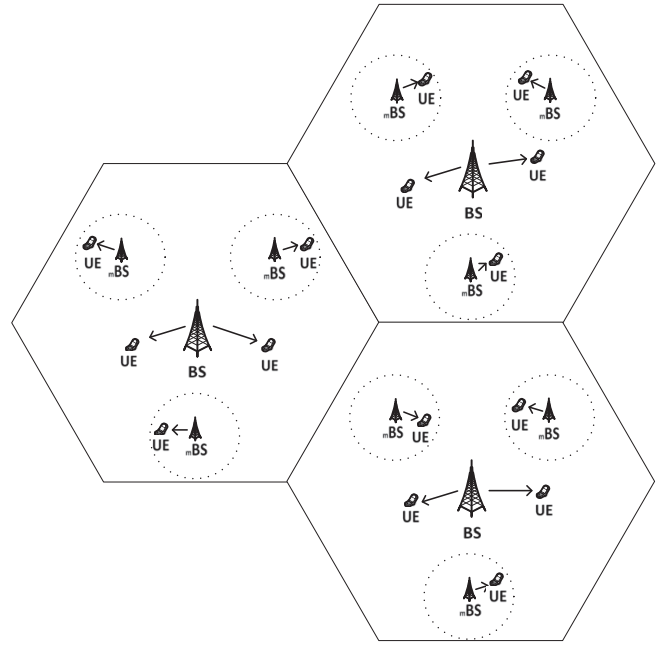


Fig. 1. The system model of a downlink heterogeneous network.

tradeoff between throughput and fairness when considering proportional fairness.

The remainder of this paper is organized as follows. In Section II, we establish the multi-cell OFDMA heterogeneous networks model and formulate the weighted sum-rate maximization problem. In Section III, a joint optimization method is proposed to obtain the locally optimal solution by dividing the maximization problem into two subproblems, where the multi-antenna receiver case and the proportional fairness case are also discussed. Simulation results are provided in Section IV. Finally, the paper is concluded in Section V.

## II. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

### A. System Description

We consider an OFDMA multi-cell heterogeneous network in the downlink transmission as shown in Fig. 1. The total number of cells in the network is $N_c$. Each cell is associated with one macro BS in the center and $N_m - 1$ uniformly deployed micro BSs. There are $N_u$ UEs uniformly distributed in the network. The total transmission time and frequency band are equally divided into multiple time slots and multiple subchannels, respectively. Each subchannel consists of several consecutive subcarriers. The channel is modeled to capture both the large-scale attenuation and the small-scale fading. Four basic assumptions are made in our system model as follows:

1) Back-haul connectivity: There is a centralized controller which is associated to all the BSs by optical fiber. Perfect channel state information (CSI) is available at the centralized controller. The CSI can be collected by the following way: each BS broadcasts the pilot signal to all the UEs. Next, each UE estimates the CSI and sends it to the related BS

via a feedback channel. Then all the BSs send the CSI to the centralized controller by optical fiber. Due to the high speed data exchange between the centralized controller and the BSs, the time cost of CSI overhead is negligible.

2) *Co-channel deployment (CCD):* All BSs (both macro BSs and micro BSs) operate on the full set of subchannels. Although CCD may lead to severe interference, its system performance can outperform the spectrum splitting method in our system model, as shown in the simulation results.

3) *Multi-association:* A UE can be served by multiple BSs in each time slot. Conventionally, a UE can only associate with one BS in each time slot [22], [23]. To further improve the network throughput, we assume that each UE can be served by different BSs on different subchannels.

4) *Channel fading:* The small-scale fading is assumed to be frequency selective and independent among different subchannels, while the channel in each subchannel is assumed to be flat fading. The channel coefficients remains unchanged within each time slot.

Denote $h_{i,j,k}^n$ as the channel coefficients from the $j$th BS in the $i$th cell to the $k$th UE on the $n$th subchannel, where $j = 1$ for macro BSs, and $j > 1$ for micro BSs. Denote $N_r$ as the total number of subchannels. The total transmit power of the macro BSs and the micro BSs are given by $P_b$ and $P_m$, respectively. Assuming that the $n$th subchannel of the $j$th BS in the $i$th cell is allocated to the $k$th UE, then the received signal-to-interference-noise-ratio (SINR) on this subchannel can be expressed by

$$\text{SINR}_{i,j,k}^n = \frac{p_{i,j}^n h_{i,j,k}^n}{\sum\limits_{i',j' \neq i,j} p_{i',j'}^n h_{i',j',k}^n + N_0}, \quad (1)$$

where $p_{i,j}^n$ is the power allocated on the $n$th subchannel at the $j$th BS in the $i$th cell and $N_0$ is the variance of the Additive White Gaussian Noise (AWGN). Then the data rate of the $k$th UE received from the $j$th BS in the $i$th cell on the $n$th subchannel in terms of bit/s/Hz is given by

$$R_{i,j,k}^n = \log(1 + \text{SINR}_{i,j,k}^n). \quad (2)$$

Let binary variable $u_{i,j,k}$ represent the user association, where $u_{i,j,k} = 1$ if the $k$th UE is associated with the $j$th BS in the $i$th cell and $u_{i,j,k} = 0$ otherwise. Let binary variable $s_{i,j,k}^n$ represent the subchannel allocation, where it equals 1 if $n$th subchannel of the $j$th BS in the $i$th cell is allocated to the $k$th UE, and equals 0 otherwise. Then the sum-rate of the heterogeneous network is given as follows

$$\sum_{i=1}^{N_c}\sum_{j=1}^{N_m}\sum_{k=1}^{N_u}\sum_{n=1}^{N_s} u_{i,j,k} s_{i,j,k}^n R_{i,j,k}^n. \quad (3)$$

### B. Problem Formulation

In this paper, our goal is to jointly optimize user association, subchannel allocation, and power allocation with the objective of maximizing the weighted sum-rate. Mathematically, the considered problem is formulated as

$$\max_{u_{i,j,k}, s_{i,j,k}^n, p_{i,j}^n} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m}\sum_{k=1}^{N_u}\sum_{n=1}^{N_s} \omega_k u_{i,j,k} s_{i,j,k}^n R_{i,j,k}^n \quad (4a)$$

$$\text{s.t.} \quad \sum_{i=1}^{N_c}\sum_{j=1}^{N_m} s_{i,j,k}^n \leq 1, \quad \forall k, n, \quad (4b)$$

$$\sum_{k=1}^{N_u} s_{i,j,k}^n \leq 1, \quad \forall i, j, n, \quad (4c)$$

$$s_{i,j,k}^n \in \{0, 1\}, \quad \forall i, j, k, n, \quad (4d)$$

$$u_{i,j,k} \in \{0, 1\}, \quad \forall i, j, k, \quad (4e)$$

$$\sum_{n=1}^{N_s} p_{i,j}^n \leq P_{i,j}, \quad \forall i, j, \quad (4f)$$

$$0 \leq p_{i,j}^n \leq P_{i,j}^n, \quad \forall i, j, n. \quad (4g)$$

In problem (4), (5a) represents the weighted sum-rate of the heterogeneous network, where $\omega_k$ is the weight of the $k$th UE. Constraint (5b) means that on each subchannel, one UE can be connected to at most one BS. Constraint (5c) means that on each subchannel, one BS can serve at most one UE. Transmit powers are constrained by both total power limits given in (4f) and spectral masks in (4g). For the user association variable $u_{i,j,k}$, we do not impose the constraint $\sum_{i=1}^{N_c}\sum_{j=1}^{N_m} u_{i,j,k} \leq 1, \forall k$, which means that multi-association is permissible in our formulation. Note that our formulation in (4) does not include any instantaneous QoS constraints for individual UEs. When the channel condition of a certain UE is very week, supporting its instantaneous QoS will consume vast resources (bandwith and power) or even be infeasible. Furthermore, since we assume CCD in the network, supporting the QoS of the UEs with bad channels creates severe interference to other UEs. Note that different weights are introduced to represent different priorities and provide different QoS. By giving each UE a weight, we not only consider the different QoS of the UEs but also avoid wasting of resources compared to the instantaneous QoS constraints.

Sovling problem (4) is challenging due to the existence of the binary variables and the non-convex SINR structure. A direct method would involve an exhaustive search over all possible user association and subchannel allocation, followed by finding the optimal power allocation for each of them. However, the complexity is exponential which makes the exhaustive search infeasible in practice. Moreover, even for fixed user association and subchannel allocation, it is still difficult to optimize the power due to the non-convex structure. Therefore, the conventional convex and quasi-convex optimization methods are not applicable to obtaining the optimal solution of problem (4).

## III. JOINT OPTIMIZATION OF USER ASSOCIATION, SUBCHANNEL ALLOCATION AND POWER ALLOCATION

In this section, we propose an alternating optimization method to solve the joint optimization problem (4). First,

we divide the problem into two subproblems: 1) joint optimization of user association and subchannel allocation for a fixed power allocation, 2) power allocation for fixed user association and subchannel allocation. For the first subproblem, the globally optimal solution is derived. For the second subproblem, we will obtain a local optimal solution. Second, we obtain a locally optimal solution of the joint optimization problem by solving these subproblems alternately.

## A. Joint Optimization of User Association and Subchannel Allocation for Fixed Power Allocation

In this subsection, we show how to obtain the optimal user association and subchannel allocation for given power allocation. In the conventional schemes, each UE can only associate with one BS in each time slot. Most previous works only consider either user association [3]–[7] or subchannel allocation [8], [9]. A few recent work consider joint optimization of user association and subchannel allocation, such as [10] and [11]. However, only a suboptimal solution is achieved in the two articles. In the following, we will propose a new user association and subchannel allocation method by exploiting the graph theory, which yields the optimal solution in multi-association system.

The joint user association and subchannel allocation problem is formulated as follows

$$\max_{u_{i,j,k}, s^n_{i,j,k}} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m}\sum_{k=1}^{N_u}\sum_{n=1}^{N_s} \omega_k u_{i,j,k} s^n_{i,j,k} R^n_{i,j,k} \quad (5a)$$

$$\text{s.t.} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m} s^n_{i,j,k} \le 1, \quad \forall k, n, \quad (5b)$$

$$\sum_{k=1}^{N_u} s^n_{i,j,k} \le 1, \quad \forall i, j, n, \quad (5c)$$

$$s^n_{i,j,k} \in \{0, 1\}, \quad \forall i, j, k, n, \quad (5d)$$

$$u_{i,j,k} \in \{0, 1\}, \quad \forall i, j, k. \quad (5e)$$

Denote $\rho^n_{i,j,k} = u_{i,j,k} s^n_{i,j,k}$, problem (5) is simplified as

$$\max_{\rho^n_{i,j,k}} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m}\sum_{k=1}^{N_u}\sum_{n=1}^{N_s} \omega_k \rho^n_{i,j,k} R^n_{i,j,k} \quad (6a)$$

$$\text{s.t.} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m} \rho^n_{i,j,k} \le 1, \quad \forall k, n, \quad (6b)$$

$$\sum_{k=1}^{N_u} \rho^n_{i,j,k} \le 1, \quad \forall i, j, n, \quad (6c)$$

$$\rho^n_{i,j,k} \in \{0, 1\}, \quad \forall i, j, k, n. \quad (6d)$$

One can observe that the summations and the constraints in the above problem are independent with respect to $n$, which suggests that problem (6) can be decomposed into $N_s$ subproblems. Without loss of generality, we concentrate on the $m$th subchannel for analysis, where $m \in \{1, \cdots, N_s\}$, i.e.,

$$\max_{\rho^m_{i,j,k}} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m}\sum_{k=1}^{N_u} \omega_k \rho^m_{i,j,k} R^m_{i,j,k} \quad (7a)$$

$$\text{s.t.} \sum_{i=1}^{N_c}\sum_{j=1}^{N_m} \rho^m_{i,j,k} \le 1, \quad \forall k, \quad (7b)$$

$$\sum_{k=1}^{N_u} \rho^m_{i,j,k} \le 1, \quad \forall i, j. \quad (7c)$$

Note that, multi-association is not supported in problem (7a), since it only focuses on the $m$th subchannel. Namely, one UE can be connected to at most one BS and one BS can serve at most one UE, if there is only one subchannel. However, problem (5) supports multi-association due to that one user can be connected to different BSs on different subchannels. Problem (7a) is a combinatorial optimization problem [24] which can always be solved by exhaustive search for all the possible cases. Obviously, this leads to a prohibitive computational complexity especially when $N_c$, $N_m$, and $N_k$ are large. Therefore, it is necessary to explore a new way to solve problem (7a). In the following, we will transform problem (7a) into an equivalent bipartite matching problem [18].

We construct a bipartite graph [18] A $= (V_{BS} \times V_{UE}, E)$, where the two sets of vertices, $V_{BS}$ and $V_{UE}$, are sets of BSs and UEs, respectively. Denote $E$ as the set of edges that connect to the vertices in the different set. Vertice $v_{BS}(i, j)$ denotes the $j$th BS in the $i$th cell and vertice $v_{UE}(k)$ denotes the $k$th UE. Let $e(i, j, k)$ denote the edge connecting $v_{BS}(i, j)$ and $v_{UE}(k)$, and $w(i, j, k)$ is the weight of $e(i, j, k)$. We use $|\cdot|$ to represent the cardinality of a set. Then $|V_{BS}| = N_c N_m$, $|V_{UE}| = N_u$ and $|E| = N_c N_m N_u$. Given a graph G $= (V, E)$, a matching M in G is a set of pairwise non-adjacent edges. That is, no two edges share a common vertex. Let $S_M$ be the set consisting of all possible matchings. According to the description above, if we denote $w(i, j, k) = \omega_k R^m_{i,j,k}$, then we can solve problem (7a) by finding a set of edges $E^*(E^* \in S_M)$ in the bipartite graph, which maximizes the sum-weight of the edges in $E^*$. This can be explained as follows:

- The value of $\rho^m_{i,j,k}$ can be equally represented by the selection of the edge in $E^*$. $e(i, j, k) \in E^*$ represents $\rho^m_{i,j,k} = 1$ and $e(i, j, k) \notin E^*$ represents $\rho^m_{i,j,k} = 0$.
- The constraints in problem (7a) is equal to $E^* \in S_M$.
- The maximum weighted sum-rate in problem (7a) is equal to the maximum sum-weight of the edges in $E^*$.

This bipartite matching problem is called maximum weighted bipartite matching (MWBM) problem [18]. The Hungarian algorithm [19] is a classical algorithm to solve the MWBM problem. By converting problem (7a) to the MWBM problem, we can obtain the globally optimal solution in polynomial time. After obtaining the optimal $\rho^n_{i,j,k}$, one can derive the corresponding optimal user association and

subchannel allocation by the following equations:

$$
u_{i,j,k} = \begin{cases} 1, \text{if } \sum_{n=1}^{N_s} \rho_{i,j,k}^n > 0, \\ 0, \quad \text{otherwise}, \end{cases} \tag{8a}
$$

$$
s_{i,j,k}^n = \rho_{i,j,k}^n. \tag{8b}
$$

Note that, if a UE is required to associate to one BS only, the existing works can only obtain a suboptimal solution for the joint user association and subchannel allocation problem. Therefore, the multi-association assumption not only helps us improve the network throughput, but also makes the problem more tractable.

### B. Power Allocation for Fixed User Association and Subchannel Allocation

In this subsection, we discuss power allocation under fixed user association and subchannel allocation. Denote $\mathbf{p} \triangleq (p_{1,1}^1, , ..., p_{1,1}^{N_s}, p_{1,2}^1, ..., p_{1,2}^{N_s}, ..., p_{N_c,N_m}^1, ..., p_{N_c,N_m}^{N_s})^T \in \mathbb{R}^{N_c N_M N_S}$. For given $\rho_{i,j,k}^n$, problem (4) is simplified as

$$
\max_{p_{i,j}^n} \quad \sum_{i=1}^{N_c} \sum_{j=1}^{N_m} \sum_{n=1}^{N_s} w_{k_{i,j,n}^*} R_{i,j,k_{i,j,n}^*}^n(\mathbf{p}) \tag{9a}
$$

$$
\text{s.t.} \quad \sum_{n=1}^{N_s} p_{i,j}^n \le P_{i,j}, \quad \forall i, j, \tag{9b}
$$

$$
0 \le p_{i,j}^n \le P_{i,j}^n, \quad \forall i, j, n, \tag{9c}
$$

where $k_{i,j,n}^*$ represents the $k$th UE associated with the $j$th BS in the $i$th cell on the $n$th subchannel, and it has been determined by the given user association and subchannel allocation.

Our goal is to obtain the optimal $p_{i,j}^n$ that maximizes the weighted sum-rate while satisfying the power constraints in (9). Note that for the special case when $N_c = 1$ and $N_m = 1$, problem (9) can be optimally solved by the conventional water-filling algorithm [25]. However, for the multi-cell scenario, the problem becomes much more complicated due to the existence of inter-cell and intra-cell interference. In this case, any power allocation change will bring impact on the resulting interference as well as the SINR. Therefore, the conventional water-filling algorithm is not applicable any more. To deal with this problem, Yu *et al.* [13] propose the iterative water-filling (IW) algorithm: With a fixed total power constraint in each BS and uniform power allocation initially, the first BS updates its power allocation by the classical water-filling method, treating signals transmitted from all the other BSs as noise. Then the same process will be done for all the BSs one after another, and so forth until the process converges. However, since each BS never considers its interference to other BSs, the IW algorithm is doomed to be unable to achieve an ideal network throughput. In the next, we will propose a novel method by exploiting its implicit DC structure in problem (9).

Note that the objective function of problem (9) is differentiable and can be written as the difference of two concave functions (10) and (11), as shown at the bottom of this page. Such a problem is recognized as the difference of two concave functions (DC) programming problem, which can be efficiently solved via the DCA method [26], [27]. The main idea of the DCA method is replacing the minuend by its first order Taylor expansion around some point and then solving the resulting convex problem. For problem (9), it is approximated as the following problem at the $s$th iteration

$$
\max_{\mathbf{p}} \quad g(\mathbf{p}) - h(\mathbf{p}[s-1]) - \nabla h^T(\mathbf{p}[s-1])(\mathbf{p} - \mathbf{p}[s-1]) \tag{12a}
$$

$$
\text{s.t.} \quad \sum_{n=1}^{N_s} p_{i,j}^n \le P_{i,j}, \quad \forall i, j, \tag{12b}
$$

$$
0 \le p_{i,j}^n \le P_{i,j}^n, \quad \forall i, j, n. \tag{12c}
$$

Alg. 1 shows how the DCA method works for problem (9). Its convergence has been discussed in previous works. Kha *et al.* [26] prove the convergence of the DCA method. Khabbazibasmenj *et al.* [28] have shown that it can converge to a point which satisfies KKT optimality conditions. However, this proof depends heavily on the specific structure of their mathematical problem. In Lemma 1, we will prove that the DCA method converges to a KKT optimal point, as long as the non-convex problem satisfies some mild conditions. For the best of our knowledge, this is the first proof which shows that the DCA method converges to a KKT optimal point for general cases.

---

**Algorithm 1** The DCA Method for Solving Problem (9)

---

1: Choose an initial feasible point $\mathbf{p}[0]$ and set $s = 1$.
2: Solve problem (12) and obtain $\mathbf{p}[s]$ .
3: Increase $s$ and go to step 2 until $\mathbf{p}[s]$ converges.

---

*Lemma 1:* For any maximization problem with a convex feasible set, if the objective function is differentiable and can be written as the difference of two concave functions, then at least a KKT point can be obtained by the DCA method.

*Proof:* Without loss of generality, consider the following non-convex problem:

$$
\max \quad f_0(\mathbf{x})
$$
$$
\text{s.t.} \quad f_i(\mathbf{x}) \le 0, \quad i = 1, 2, ...m, \tag{13}
$$

---

$$
g(\mathbf{p}) \triangleq \sum_{i=1}^{N_c} \sum_{j=1}^{N_m} \sum_{n=1}^{N_s} w_{k_{i,j,n}^*} \log \left( \sum_{i'=1}^{N_c} \sum_{j'=1}^{N_m} p_{i',j'}^n h_{i',j',k_{i,j,n}^*}^n + N_0 \right), \tag{10}
$$

$$
h(\mathbf{p}) \triangleq \sum_{i=1}^{N_c} \sum_{j=1}^{N_m} \sum_{n=1}^{N_s} w_{k_{i,j,n}^*} \log \left( \sum_{i'=1}^{N_c} \sum_{j'=1}^{N_m} p_{i',j'}^n h_{i',j',k_{i,j,n}^*}^n - p_{i,j}^n h_{i,j,k_{i,j,n}^*}^n + N_0 \right). \tag{11}
$$

where $f_0$ can be written as the difference of two concave functions, i.e., $f_0(\mathbf{x}) = g(\mathbf{x}) - h(\mathbf{x})$, and both $g(\mathbf{x})$ and $h(\mathbf{x})$ are concave functions. The feasible set created by the constraints is convex. Then we can introduce a new additional variable $t$ and express problem (13) equivalently as

$$\max \quad t$$
$$\text{s.t.} \quad t - f_0(\mathbf{x}) \leq 0,$$
$$f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, ..., m. \tag{14}$$

Denote $\mathbf{x}^{(s)}$ as the optimal solution of the $s$th convex problem. Let $f(\mathbf{x}, t) \triangleq t - f_0(\mathbf{x}) = t - g(\mathbf{x}) + h(\mathbf{x})$. According to the DCA method, we use $\tilde{f}(\mathbf{x}, t)$ approximate $f(\mathbf{x}, t)$ by replacing $h(\mathbf{x})$ with its first order Taylor expansion around $\mathbf{x}^{(s-1)}$, i.e., $\tilde{f}(\mathbf{x}, t) = t - g(\mathbf{x}) + h(\mathbf{x}^{(s-1)}) + \nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x} - \mathbf{x}^{(s-1)})$, to form the $s$th convex problem. In the next, we provide three properties of DCA.

1) The following inequality holds for arbitrary $\mathbf{x}$ and $\mathbf{x}^{(s-1)}$ due to the concavity of $h(\mathbf{x})$

$$h(\mathbf{x}) \leq h(\mathbf{x}^{(s-1)}) + \nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x} - \mathbf{x}^{(s-1)}). \tag{15}$$

Thus, we have the following result for arbitrary $\mathbf{x}$

$$f(\mathbf{x}, t) = t - g(\mathbf{x}) + h(\mathbf{x})$$
$$\leq t - g(\mathbf{x}) + h(\mathbf{x}^{(s-1)}) + \nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x} - \mathbf{x}^{(s-1)})$$
$$= \tilde{f}(\mathbf{x}, t). \tag{16}$$

2) The optimal solution of the *(s-1)*th convex problem satisfies the following equations

$$\tilde{f}(\mathbf{x}^{(s-1)}, t)$$
$$= t - g(\mathbf{x}^{(s-1)}) + h(\mathbf{x}^{(s-1)})$$
$$\quad + \nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x}^{(s-1)} - \mathbf{x}^{(s-1)})$$
$$= t - g(\mathbf{x}^{(s-1)}) + h(\mathbf{x}^{(s-1)})$$
$$= f(\mathbf{x}^{(s-1)}, t). \tag{17}$$

3) The gradient of $\tilde{f}(\mathbf{x}, t)$ is

$$\nabla \tilde{f}(\mathbf{x}, t)$$
$$= \nabla(t - g(\mathbf{x}) + h(\mathbf{x}^{(s-1)}) + \nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x} - \mathbf{x}^{(s-1)}))$$
$$= \nabla t - \nabla g(\mathbf{x}) + \nabla(\nabla h^T(\mathbf{x}^{(s-1)})(\mathbf{x} - \mathbf{x}^{(s-1)}))$$
$$= \nabla t - \nabla g(\mathbf{x}) + \nabla h(\mathbf{x}^{(s-1)}). \tag{18}$$

Then we have the following equation

$$\nabla \tilde{f}(\mathbf{x}^{(s-1)}, t) = \nabla t - \nabla g(\mathbf{x}^{(s-1)}) + \nabla h(\mathbf{x}^{(s-1)})$$
$$= \nabla f(\mathbf{x}^{(s-1)}, t). \tag{19}$$

Note that, the equations (16), (17), and (19) are the sufficient conditions for the convergence to a KKT point [29]. Therefore, the DCA method will converge to a KKT point. ∎

We can easily know that Alg. 1 will converge to a KKT point, since problem (9) actually satisfies the above constraints in Lemma 1. Now the remaining task is to solve problem (12), where Lagrange dual technique [20], [21] can be employed. The Lagrangian function is

$$\mathcal{L}(\mathbf{p}, \boldsymbol{\lambda}) \triangleq g(\mathbf{p}) - h(\mathbf{p}[s-1]) - \nabla h^T(\mathbf{p}[s-1])(\mathbf{p} - \mathbf{p}[s-1])$$
$$+ \sum_{i=1}^{N_c} \sum_{j=1}^{N_m} \lambda_{i,j}\left(P_{i,j} - \sum_{n=1}^{N_s} p_{i,j}^n\right). \tag{20}$$

where $\boldsymbol{\lambda}$ is a Lagrange multiplier vector corresponding to the maximum transmit power constraint (10b). Then the dual optimization problem is given by

$$\min_{\boldsymbol{\lambda}} \quad \max_{\mathbf{p}} \mathcal{L}(\mathbf{p}, \boldsymbol{\lambda})$$
$$\text{s.t.} \quad \lambda_{i,j} \geq 0. \tag{21}$$

The above dual problem can be solved iteratively by decomposing it into two nested loops: the inner loop that maximizes $\mathbf{p}$ for given $\boldsymbol{\lambda}$, and the outer loop that determines the optimal $\boldsymbol{\lambda}$. In the following, we will discuss them in detail.

1) The inner loop: Denote $\boldsymbol{\lambda}^{(l)}$ as the $l$th iteration dual variable. For given $\boldsymbol{\lambda}^{(l)}$, we can derive the following power allocation under the constraint (10c), by setting the derivative of $\mathcal{L}(\mathbf{p}, \boldsymbol{\lambda})$ with respect to $p_{i,j}^n$ to zero.

$$p_{i,j}^n = \left[ \frac{1}{\lambda_{i,j}^{(l)} + d_{i,j}^n} - \frac{\sum\limits_{(u,v) \neq (i,j)} p_{u,v}^n h_{u,v,k_{i,j,n}^*}^n + N_0}{h_{i,j,k_{i,j,n}^*}^n} \right]_0^{P_{i,j}^n}, \tag{22}$$

where

$$d_{i,j}^n = \sum_{(i',j') \neq (i,j)} \left( \frac{h_{i,j,k_{i',j',n}^*}^n}{\sum\limits_{(u,v) \neq (i',j')} p_{u,v}^n[s-1]h_{u,v,k_{i',j',n}^*}^n + N_0} \right.$$
$$\left. - \frac{h_{i,j,k_{i',j',n}^*}^n}{\sum\limits_{u,v} p_{u,v}^n h_{u,v,k_{i',j',n}^*}^n + N_0} \right), \tag{23}$$

which is a taxation term related to the interference to other scheduled users. Since $p_{i,j}^n$ also appears on the right side in (22), a closed-form power allocation expression cannot be obtained directly. However, a unique optimal power allocation can be obtained by the fixed point method as shown in [30].

2) The outer loop: Once the optimal power allocation is achieved, the solution of the dual problem can be updated by the subgradient method as follows

$$\lambda_{i,j}^{(l+1)} = \left[ \lambda_{i,j}^{(l)} + \delta\left(\sum_{n=1}^N p_{i,j}^n - P_{i,j}\right) \right]^+, \tag{24}$$

where $\delta$ is a sufficiently small step size. Since the dual problem is always a convex optimization problem, the subgradient method will converge to the globally optimal solution. The detailed algorithm to solve problem (12) is summarized as follows.

The computational complexity of Alg. 2 is $O\left(K_\lambda K_P (N_C N_m)^2 N_S\right)$, where $K_\lambda$ and $K_P$ are the required number of iterations for updating $\boldsymbol{\lambda}$ and $\mathbf{p}$, respectively. From the simulation results in Section IV, we will see that $K_\lambda$ is large, which inherently increases the computational complexity of Alg. 2.

In the remainder of this subsection, we will propose a low complexity algorithm to solve problem (10) by avoiding the gradient descent search on the dual variable $\boldsymbol{\lambda}$. The basic idea behind this algorithm is that $\lambda_{i,j}^{(l)}$ and $p_{i,j}^{n,(l)}$ can be updated simultaneously in the $l$th iteration. More specifically, in the

**Algorithm 2** The Lagrange Dual Algorithm for Solving Problem (12)

1: Set $l = 1$. Initialize $\lambda_{i,j}^{(l)}$ and $\delta$.
2: For given $\lambda_{i,j}^{(l)}$, obtain the optimal $p_{i,j}^n$ from (22) by the fixed point method.
3: Update $\lambda_{i,j}^{(l+1)}$ by (24).
4: Increase $l$ and go to step 2 until $\frac{\left|\lambda_{i,j}^{(l+1)} - \lambda_{i,j}^{(l)}\right|}{\lambda_{i,j}^{(l+1)}} \leq \varepsilon$, for $\forall i, j$.

$l$th iteration, we first obtain the taxation term $d_{i,j}^{n,(l)}$ for all BSs according to (23) in parallel. Then we update $\lambda_{i,j}^{(l)}$ and $p_{i,j}^{n,(l)}$ according to the KKT conditions

$$p_{i,j}^{n,(l)} = \left[ \frac{1}{\lambda_{i,j}^{(l)} + d_{i,j}^{n,(l)}} - \frac{\sum\limits_{(u,v)\neq(i,j)} p_{u,v}^{n,(l-1)} h_{u,v,k_{i,j,n}^*}^n + N_0}{h_{i,j,k_{i,j,n}^*}^n} \right]_0^{P_i^n}, \tag{25a}$$

$$\lambda_{i,j}^{(l)} \left( P_{i,j} - \sum_{n=1}^{N_s} p_{i,j}^{n,(l)} \right) = 0, \quad \forall i, j, \tag{25b}$$

$$\sum_{n=1}^{N_s} p_{i,j}^{n,(l)} \leq P_{i,j}, \quad \forall i, j, \tag{25c}$$

$$\lambda_{i,j}^{(l)} \geq 0, \quad \forall i, j. \tag{25d}$$

In theorem 1, we will prove that there always exists a unique $\lambda_{i,j}^{(l)}$ which satisfies the conditions in (25). Furthermore, since $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)}$ is decreasing with respect to $\lambda_{i,j}^{(l)}$, we can efficiently obtain this $\lambda_{i,j}^{(l)}$ via the bisection method [31]. Our low complexity algorithm to solve problem (12) is summarized as Alg. 3. Its computational complexity is $O\left(K_T (N_C N_m)^2 N_S\right)$, where $K_T$ is the number of iterations.

*Theorem 1:* Given $p_{i,j}^{n,(l-1)}$, there exists a unique $\lambda_{i,j}^{(l)}$ which satisfies the conditions in (25).

*Proof:* From (25a), it is easy to know that $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)}$ is strictly decreasing with respect to $\lambda_{i,j}^{(l)}$, when $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)} > 0$. For given $p_{i,j}^{n,(l-1)}$, set $\tilde{\lambda}_{i,j}^{(l)} = 0$ and calculate $\tilde{p}_{i,j}^n$ by (25a). Then we have the following two cases.

1) For the case $\sum\limits_{n=1}^{N_s} \tilde{p}_{i,j}^n \leq P_{i,j}$, $\tilde{\lambda}_{i,j}^{(l)} = 0$ satisfies (25). Furthermore, for any $\lambda_{i,j}^{(l)} > 0$, we have $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)} < P_{i,j}$ since $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)}$ is strictly decreasing with respect to $\lambda_{i,j}^{(l)}$. Therefore, we have $\lambda_{i,j}^{(l)} \left( P_{i,j} - \sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)} \right) > 0$, which violates equation (25b). Hence we know that $\tilde{\lambda}_{i,j}^{(l)} = 0$ is the unique $\lambda_{i,j}^{(l)}$ which satisfies the conditions in (25).

2) For the case $\sum\limits_{n=1}^{N_s} \tilde{p}_{i,j}^n > P_{i,j}$, due to the monotonicity of $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)}$, there exists a unique $\lambda_{i,j}^{(l)} > 0$ which makes $\sum\limits_{n=1}^{N_s} p_{i,j}^{n,(l)} = P_{i,j}$ hold. Furthermore, it is clear that this $\lambda_{i,j}^{(l)}$ satisfies the other conditions in (25). Theorem 1 is proved. ∎

**Algorithm 3** The Low Complexity Algorithm for Solving Problem (12)

1: Choose an initial feasible point $p_{i,j}^{n,(0)}$ and set $l = 1$.
2: Update $d_{i,j}^{n,(l)}$ according to (23) for all BSs.
3: Update $\lambda_{i,j}^{(l)}$ via the bisection method, then update $p_{i,j}^{n,(l)}$ according to (25a).
4: Increase $l$ and go to step 2 until $\frac{\left|\lambda_{i,j}^{(l+1)} - \lambda_{i,j}^{(l)}\right|}{\lambda_{i,j}^{(l+1)}} \leq \varepsilon$, for $\forall i, j$.

Note that once Alg. 3 converges, one will obtain a KKT point for problem (12). Moreover, it is also a globally optimal solution, according to the convexity of problem (12). Deriving conditions under which Alg. 3 converges is intractable, although convergence has always been observed in our simulation results in Section IV.

### C. Joint Optimization of User Association, Subchannel Allocation, and Power Allotcation

In the previous subsections, we have obtained the globally optimal user association and subchannel allocation for fixed power allocation, and a KKT optimal power allocation for fixed user association and subchannel allocation. Now, we propose an alternating optimization method for the joint design of user association, subchannel allocation, and power allotcation, involving iterations between the Hungarian algorithm and Alg. 3 until convergence. The overall algorithm is summarized as follows, whose convergence is guaranteed by Theorem 2.

**Algorithm 4** The Joint Optimization Algorithm for Solving problem (4)

1: Initialize $\mathbf{p}[0]$ uniformly and set $i = 1$.
2: Update $\boldsymbol{\rho}[i]$ by the Hungarian algorithm for $\mathbf{p}[i-1]$.
3: Form the $i$-th approximated convex problem around $\mathbf{p}[i-1]$, and solve this convex problem by Alg. 3 to update $\mathbf{p}[i]$.
4: Increase $i$ and go to step 2 until the weighted sum-rate converges.

*Theorem 2:* The convergence of Alg. 4 can always be guaranteed.

*Proof:* From the above subsections, one can see the weighted sum-rate can only change in step 2 and step 3. In step 2, the Hungarian algorithm finds the globally optimal user association and subchannel allocation. In step 3, Alg. 3 finds the globally optimal solution for the $i$-th approximated convex problem. Therefore, the sequence of iterations produces a monotonically increasing weighted sum-rate. Meanwhile, it is obvious that the weighted sum-rate has an upper

bound for finite power constraints. Thus, the convergence of Alg. 4 is guaranteed. ∎

*Remark 1:* Note that even if we can obtain the globally optimal power for given user association and subchannel allocation, the alternating optimization method still converges to a locally optimal solution. The globally optimal solution of problem (4) needs exhaustive search for all the possible user association and subchannel allocation and get the optimal power allocation for each possible case, which is impossible in practice.

In the following, we will analyze the complexity of Alg. 4. In each iteration, the Hungarian algorithm is adopted in step 2 with complexity given by $O\left(N_u{}^3 N_s\right)$, and Alg. 3 is adopted in step 3 with complexity given by $O\left(K_T (N_C N_m)^2 N_S\right)$ as shown in the above subsection. Denote $K_J$ as the average iteration number, then the computational complexity of Alg. 4 is given by $O\left(K_J N_u{}^3 N_s + K_J K_T (N_c N_m)^2 N_s\right)$. From the simulation results, we can observe that both $K_J$ and $K_T$ are very small. Therefore, Alg. 4 has a relatively low computational complexity.

### D. Multi-Antenna Receivers

In this subsection, we consider the case that all UE have the same number of antennas, denoted as $N_a$ ($N_a > 1$). Note that by adopting multiple antennas, the receiver will obtain different SINR by different strategies. We focus on two different strategies: the maximum ratio combining (MRC) strategy and the interference rejection combining (IRC) strategy. According to what strategies the UEs apply, we partition them into two categories: MRC receivers and IRC receivers. In practice, MRC receivers and IRC receivers coexist in the network, which are indistinguishable to the BSs.

Denote $\boldsymbol{h}_{i,j,k}^n \in \mathcal{C}^{N_a}$ as the channel coefficients vector from the $j$th BS in the $i$th cell to the $k$th UE on the $n$th subchannel. Assuming that the $k$th UE is associated with the $j$th BS in the $i$th cell on the $n$th subchannel, then the received signal in $k$th UE is given by:

$$\mathbf{y}_{i,j,k}^n = \mathbf{h}_{i,j,k}^n \sqrt{p_{i,j}^n} x_{i,j}^n + \sum_{i',j'\neq i,j} \mathbf{h}_{i',j',k}^n \sqrt{p_{i',j'}^n} x_{i',j'}^n + \mathbf{n_0},$$
(26)

where $x_{i,j}^n$ is the data from the $j$th BS in the $i$th cell on the $n$th subchannel, and $\mathbf{n_0} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{N_a})$ is the complex AWGN vector.

For MRC receivers, the received signals in (26) is multiplied by the channel coefficient vector, i.e.,

$$(\mathbf{h}_{i,j,k}^n)^H \mathbf{y}_{i,j,k}^n = (\mathbf{h}_{i,j,k}^n)^H \mathbf{h}_{i,j,k}^n \sqrt{p_{i,j}^n} x_{i,j}^n$$
$$+ \sum_{i',j'\neq i,j} (\mathbf{h}_{i,j,k}^n)^H \mathbf{h}_{i',j',k}^n \sqrt{p_{i',j'}^n} x_{i',j'}^n$$
$$+ (\mathbf{h}_{i,j,k}^n)^H \mathbf{n_0},$$
(27)

then the corresponding SINR is given by

$$\text{SINR}_{i,j,k}^n = \frac{p_{i,j}^n \left|(\mathbf{h}_{i,j,k}^n)^H \mathbf{h}_{i,j,k}^n\right|^2}{\sum\limits_{i',j'\neq i,j} p_{i',j'}^n \left|(\mathbf{h}_{i,j,k}^n)^H \mathbf{h}_{i',j',k}^n\right|^2 + \left|(\mathbf{h}_{i,j,k}^n)^H \mathbf{n_0}\right|^2}.$$
(28)

Comparing formula (28) with formula (1), we know that the SINR of the multi-antenna system has the same structure with that of the single-antenna system. Therefore, Alg. 4 can also be directly applied to the multi-antenna system for the MRC receivers.

For IRC receivers, the received signals in (26) is multiplied by some predefined vector, i.e.,

$$\mathbf{w}^H \mathbf{y}_{i,j,k}^n = \mathbf{w}^H \mathbf{h}_{i,j,k}^n \sqrt{p_{i,j}^n} x_{i,j}^n$$
$$+ \sum_{i',j'\neq i,j} \mathbf{w}^H \mathbf{h}_{i',j',k}^n \sqrt{p_{i',j'}^n} x_{i',j'}^n + \mathbf{w}^H \mathbf{n_0}, \quad (29)$$

where $\mathbf{w}^H = (\mathbf{h}_{i,j,k}^n)^H (\sum\limits_{i',j'\neq i,j} p_{i',j'}^n \mathbf{h}_{i',j',k}^n (\mathbf{h}_{i,j,k}^n)^H + N_0 \mathbf{I})^{-1}$. Thus, the received SINR is given by

$$\text{SINR}_{i,j,k}^n = \frac{p_{i,j}^n \left|\mathbf{w}^H \mathbf{h}_{i,j,k}^n\right|^2}{\sum\limits_{i',j'\neq i,j} p_{i',j'}^n \left|\mathbf{w}^H \mathbf{h}_{i',j',k}^n\right|^2 + \left|\mathbf{w}^H \mathbf{n_0}\right|^2}.$$
(30)

From (30), we can see that neither the numerator nor the denominator is convex. Therefore, the network throughput in this case does not have a DC structure. However, since the BSs do not know which kind of receiver that the UEs apply in advance, they will continue to allocate power as if the UEs applied MRC receivers. In the following simulation results, an interesting observation is that Alg. 4 behaves even better when applying to those IRC receivers.

### IV. SIMULATION RESULTS

In this section, we provide simulation results to demonstrate the effectiveness of the proposed methods in Multi-cell OFDMA Heterogeneous Networks. We consider a heterogeneous network consisting of 7 cells, where one macro BS is deployed at the center and three micro BSs are randomly distributed in each cell. The power of each macro BS and micro BS are 46dBm and 30dBm, respectively. The bandwidth of each subcarrier is 15kHZ and each subchannel consists of 12 subcarriers. The whole frequency band is divided into 50 subchannels. Unless otherwise specified, the inter-site distance (ISD) between macro BSs is set to be 500 meters so that the network is interference limited. For the large-scale fading, the distance-dependent path loss in dB is modeled as $PL_{NLOS} = 128.1 + 37.6\log_{10}(d)$, where $d$ is the distance from the user to the BS in kilometers. The log-normal shadowing is considered with $\sigma_{shadow} = 10$dB and the penetration loss is assumed to be 20dB. The small-scale fading is modeled as the normalized Rayleigh fading. The noise power spectral density is set to be $-174$dBm/Hz.
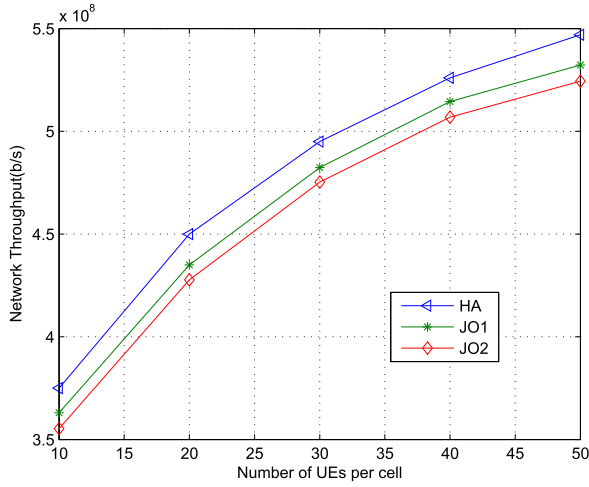
Fig. 2. The network throughput versus the number of UEs per cell of HA, JO1 and JO2.
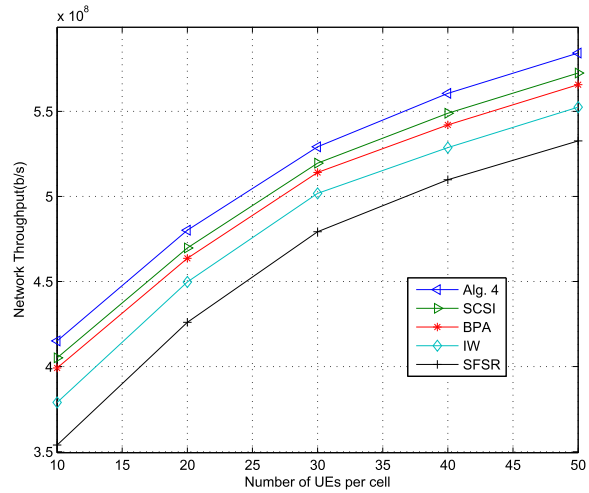


Fig. 3. The network throughput versus the number of UEs per cell of Alg. 4, SCSI, BPA, IW, and SFSR.

### A. Joint User Association and Subchannel Allocation

In this subsection, we compare the throughput between different joint user association and subchannel allocation schemes. The number of UEs in each cell changes from 10 to 50. In Fig. 2, we compare the network throughput of the following three methods: 1) the Hungarian algorithm (HA): our algorithm which can obtain the globally optimal solution; 2) JO1: the joint optimization method proposed in [11]; 3) JO2: the joint optimization method proposed in [10]. Uniform power allocation is performed in all the BSs. From the comparison, one can observe that our method outperforms the existing two joint optimization methods in multi-cell networks.

### B. Joint User Association, Subchannel Allocation, and Power Optimization

In this subsection, we show the performance of the network throughput versus the number of UEs in each cell. The number of UEs in each cell changes from 10 to 50. In Fig. 3, we compare the network throughput of the following five methods: 1) Alg. 4; 2) statistical CSI (SCSI): resource allocation by Alg.4 when the the intercell CSI is statistical, and the

mean value is used instead of the instantaneous intercell CSI; 3) BPA: the belief propagation algorithm proposed in [11]; 4) IW: power allocation by the IW algorithm for fixed user association based on cell range expansion; 5) Static full spectral reuse (SFSR): uniform power allocation for fixed user association by cell range expansion. From the comparison, we can observe that Alg. 4 performs best among all five schemes. It should be mentioned that, when the intercell CSI is statistical, Alg. 4 still outperforms the other three schemes. This result can be explained by the fact that the interference from other cells is weaker compared with the intra-cell interference. Hence, the SCSI method can be regarded as an alternative option in order to decrease the CSI overhead. Compared with BPA, we can observe the throughput gain by Alg.4 and SCSI due to the assumption of multiple-association and continuous power allocation.

In the following, we compare the computational complexity of the five methods. As mentioned above, the complexity of Alg. 4 and SCSI are both $O\left(K_J N_u{}^3 N_s + K_J K_T (N_c N_m)^2 N_s\right)$. The complexity of the BPA method is $O\left(K W H_v{}^2 + K W H_f{}^2\right)$, where $K$ is the average iteration number, $W$ is the avearage number of the scheduling options in variable nodes, $H_v$ is the average number of neighboring factor nodes of a variable node and $H_f$ is the average number of a factor node's neighboring variable nodes. The complexity of the IW algorithm and the SFSR method is $O\left(N_c N_m N_u + K (N_c N_m)^2 N_s\right)$ and $O\left(N_c N_m (N_u + N_s)\right)$, respectively. In conclusion, the complexity of Alg. 4 and SCSI is higher than that of the IW algorithm and the SFSR method. Note that, we can't compare the complexity between our algorithm with the BPA method due to the fact that $W$, $H_v$ and $H_f$ are not estimated in [11].

In additon, if multi-association is not considered, the globally optimal solution of the joint optimization problem can be obtained by the following two steps: 1) exhaustive search for all the possible user association; 2) get the optimal subchannel and power allocation for each possible case by the JSPPA algorithm [16]. The complexity in exhaustive search is $O\left((N_c N_m)^{N_u}\right)$. The complexity of the JSPPA algorithm is $O\left(R_1 (N_e)^{N_c N_s}\right)$, where $R_1$ is the number of iterations and $N_e$ is number of UEs in each cell. The globally optimal algorithm would take an unrealistically long time to return the globally optimal solution for practical multi-cell networks. Therefore, we compare Alg. 4 with the globally optimal algorithm in a single-cell scenario, where six UEs are distributed uniform in the cell. The total number of subchannels is four. The transmit power of the micro BS is 30dBm and the transmit power of the macro BS changes from 40dBm to 46dBm. Fig. 4 shows the performance of the network throughput versus the transmit power of the macro BS. From the simulation, we can see that the throughput gap between Alg. 4 and the above globally optimal (GO) algorithm is less than 3%, which is negligible considering the huge computational complexity reduction.

In the following, we compare Alg. 4 with the interference avoidance strategy where the BSs in each cell orthogonally utilize the resource. In Fig. 5, we compare the following three methods: 1) Alg. 4; 2) SFSR; 3) spectrum splitting (SS): subchannel allocation based on cell range expansion and power
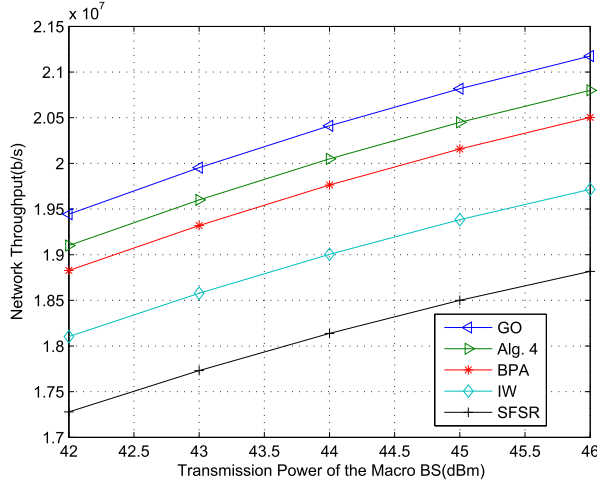
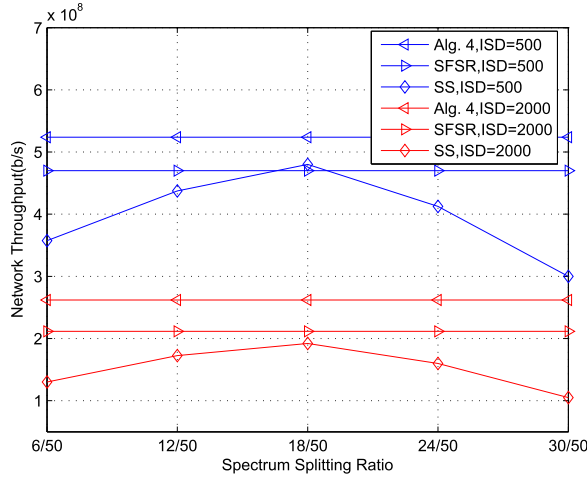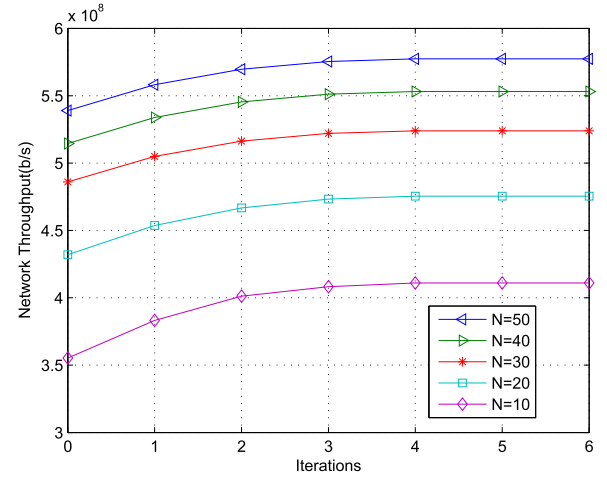Fig. 4. The network throughput of GO, Alg. 4, BPA, IW, and SFSR.



Fig. 6. The average network throughput after each iteration by Alg. 4.

TABLE I
THE NUMBER OF ITERATIONS TO SOLVE THE CONVEX
PROBLEM IN STEP 3 OF ALG. 4

| parameters \ number of UEs | 30 | 60 | 90 | 120 | 150 |
|---|---|---|---|---|---|
| $K_\lambda$ | 11.314 | 12.015 | 11.941 | 11.736 | 10.928 |
| $K_P$ | 3.829 | 3.906 | 3.514 | 3.716 | 4.037 |
| $K_T$ | 4.235 | 4.419 | 4.137 | 3.819 | 4.386 |



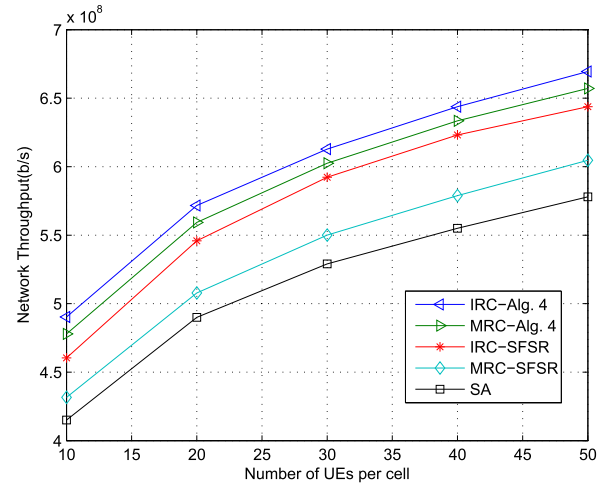Fig. 5. The network throughput of Alg. 4, SFSR, and the spectrum splitting policy.



Fig. 7. The network throughput versus the number of UEs per cell for the following five methods: IRC-Alg. 4, MRC-Alg. 4, IRC-SFSR, MRC-SFSR, and SA.

allocation based on classical water-filling algorithm [25] in each cell. Two scenarios are considered: the urban scenario where the ISD between macro BSs is 500m and the rural scenario where the ISD between macro BSs is 2000m. The number of UEs in each cell is 30. The x-axis represents the ratio of subchannels used by micro BSs among all 50 subchannels for SS method, which are divided equally to 3 micro BSs. As expected, the network throughputs of all the methods decrease with the increasing of ISD due to the increasing path loss. We can see that Alg. 4 is always better than the spectrum splitting policy. Another observation is that the best performance of SS is better than SFSR in the urban scenario. However, SFSR always outperforms SS in the rural scenario. The best performance of SS can be achieved at the ratio 18/50, which means that in each cell, the macro BS occupy 32 subchannels and every micro BS occupy 6 subchannels.

## C. Convergence of Algorithm 4

Fig. 6 shows the convergence behavior for Alg. 4, and the number of UEs in each cell changes from 10 to 50. In each iteration, the Hungarian algorithm and the DC approximation are implemented once, respectively. We set the precision to

be 0.01 and simulate 1000 times. As expected, the average network throughput increases after each iteration, and Alg. 4 converges within 6 iterations.

Tab. 1 shows the number of iterations to solve the convex problem in step 3 of Alg. 4, when the number of UEs changes from 30 to 150. Here, one iteration means the power update on all the BSs. The precision is set to be 0.01. $K_\lambda$ and $K_P$ are average numbers of required iterations for updating $\lambda$ and $\mathbf{p}$ respectively. $K_T$ is the average number of required iterations for Alg. 3. From Tab. 1, we can find that the iteration number of the Lagrange Dual method (given by $K_\lambda K_P$) is nearly 9 times larger than that of our method (given by $K_T$). Another observation is that the iteration number is almost independent of the number of UEs.
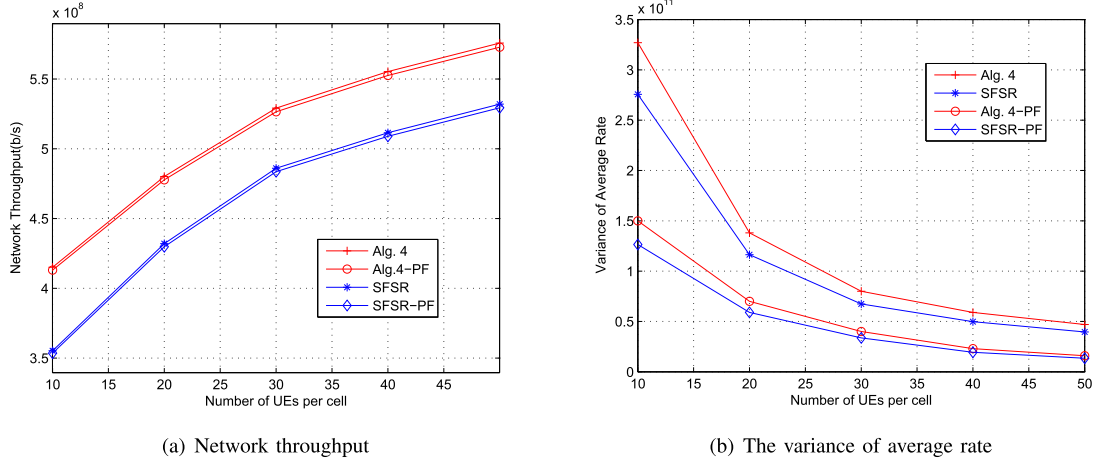
(a) Network throughput

(b) The variance of average rate

Fig. 8. The comparison of network throughput and the variance of average rate in dynamic scenario.



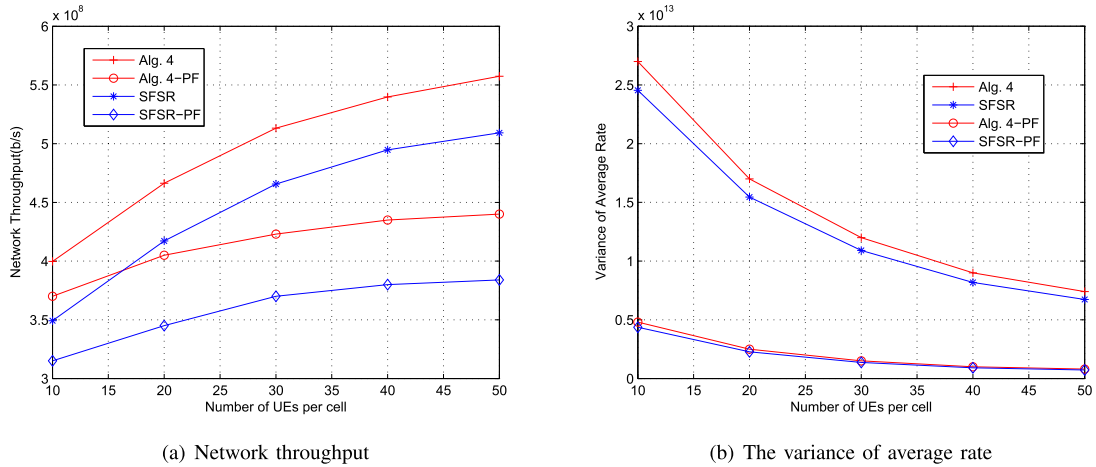(a) Network throughput

(b) The variance of average rate

Fig. 9. The comparison of network throughput and the variance of average rate in static scenario.

### D. Multi-Antenna Receivers

In Fig. 7, we compare the network throughput of the following five methods: 1) IRC-Alg. 4: the subchannel and power allocation is achieved by Alg.4 and the UEs apply IRC receivers; 2) MRC-Alg. 4: the subchannel and power allocation is achieved by Alg.4 and the UEs apply MRC receivers; 3)IRC-SFSR: the subchannel and power allocation is achieved by SFSR and the UEs apply IRC receivers; 4) MRC-SFSR: the subchannel and power allocation is achieved by SFSR and the UEs apply MRC receivers; 5) single antenna(SA). Without loss of generality, we set the number of antennas to be 2. We observe that IRC receivers always performs better than MRC receivers under the same resource allocation method. However, if the resource allocation is implemented by Alg. 4, the gap between the two kinds of receivers will diminish due to the fact that Alg. 4 can decrease the interference from other BSs.

### E. Proportional Fairness

The proportional fairness has been widely discussed since it strikes a good balance between network throughput and fairness by exploiting multiuser diversity and game-theoretic equilibrium [32]. In this subsection, we will show the performance of our algorithm when proportional fairness is considered. It has been proven that proportional fairness can be achieved by setting the weight of each user to the reciprocal of the average rate in each time slot, when the total time slots is large enough [33]. Denote the weight in time slot $i$ as

$$\omega_k(i) = \begin{cases} 1, & i = 1, \\ \frac{1}{C_k(i-1)}, & otherwise, \end{cases} \quad (31)$$

where $C_k(i - 1)$ is the average rate of the $k$th UE from the first time slot to the $i - 1$th time slot. Then we can achieve proportional fairness by adopting the weight in (31) in time slot $i$.

We focus on two scenarios: the dynamic scenario and the static scenario. In the dynamic scenario, we assume the location of the UEs are generated randomly and independently in each time slot for simplicity. While in the static scenario, the location of UEs remain static in all time slots. For each scenario, we compare the network throughput and the variance of average rate between the following four methods: 1) Alg. 4:

resource allocation by Alg.4 without considering the fairness; 2) Alg. 4-PF: resource allocation by Alg.4 considering proportional fairness; 3) SFSR: resource allocation by SFSR without considering the fairness; 4) SFSR-PF: resource allocation by SFSR considering proportional fairness.

Fig. 8 shows the performance in the dynamic scenario, and the total time slot is set as 1000. From Fig. 8(a), we find that the network throughput without considering the fairness is similar to that when considering proportional fairness. The reason is that the channel coefficients of all UEs are independent identically distributed in each time slot. When the number of the time slot tends to infinity, the network throughput when considering proportional fairness will tend to that without considering the fairness. From Fig. 8(b), we observe that the variance of average rate when considering proportional fairness is nearly half of that without considering the fairness, which demonstrate the promotion of fairness. Another observation is that Alg. 4-PF performs better than SFSR in both network throughput and fairness.

Fig. 9 shows the performance in the static scenario and the total time slot is set as 1000. The results are obtained by averaging over 50 independent large-scale channel realizations. From Fig. 9(a), we know Alg. 4 yields a better network throughput than Alg. 4-PF. The throughput gain comes at the cost of fairness among the UEs, as shown in Fig. 9(b). Due to the fixed large-scale fading, the channel conditions in most time tend to be better for some UEs, while worse for others. Hence more resource is required to sustain the rate for UEs with bad channels, resulting in the decrease of the network throughput. So the throughput gap between Alg. 4 and Alg. 4-PF in Fig. 9(a) is obviously larger than that in Fig. 8(a). Correspondingly, the variance gap in Fig. 9(b) is obviously larger than that in Fig. 8(b). Moreover, we notice that the network throughput of Alg. 4-PF outperforms SFSR when the number of UEs is small, while SFSR has a better performance when the number of UEs is large. This is because we only serve the UEs with superior channels for SFSR. However, for Alg. 4-PF, the number of UEs with weak channels increases with the total number of UEs, which costs more physical resource and thus inhibits the increase of the network throughput.

## V. Conclusion

In this paper, we maximized the weighted sum-rate for the downlink transmission in Multi-cell Multi-association OFDMA heterogeneous networks. A joint user association, subchannel allocation, and power allocation optimization problem was formulated. To solve the optimization problem, we divided it into two subproblems. The first subproblem is joint user association and subchannel allocation for a fixed power allocation, whose globally optimal solution can be obtained by the Hungarian algorithm. The second subproblem is power allocation for fixed user association and subchannel allocation, which can be transformed to a series of convex problems by the DCA method. To further reduce its complexity, we proposed a simplified but efficient algorithm to solve these problems, decreasing the number of iterations up to almost 90 percent off than the conventional Lagrange Dual method. Simulation results showed that our joint optimization algorithm achieves a better performance compared with the existing algorithms. We also extend the problem into the multi-antenna receiver case and the proportional fairness case, respectively. Our algorithm performs well when the UEs apply multi-antenna receivers, and it helps achieve a good tradeoff between throughput and fairness when considering proportional fairness.

## References

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

[2] T. Nakamura *et al.*, "Trends in small cell enhancements in LTE advanced," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 98–105, Feb. 2013.

[3] K. Okino, T. Nakayama, C. Yamazaki, H. Sato, and Y. Kusano, "Pico cell range expansion with interference mitigation toward LTE-advanced heterogeneous networks," in *Proc. ICC Workshops*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[4] *System Simulations for Downlink Co-Channel Interference Scenario*, document R1-130507, 3GPP TSG RAN WGC Meetig-72, Huawei, Jan. 2013.

[5] W. Zhouyun, X. Weiliang, Y. Fengyi, and B. Qi, "User association in heterogeneous network with dual connectivity and constrained backhaul," *China Commun.*, vol. 13, no. 2, pp. 11–20, Feb. 2016.

[6] U. Siddique, H. Tabassum, E. Hossain, and D. Kim, "Channel-access-aware user association with interference coordination in two-tier downlink cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5579–5594, Jul. 2016.

[7] A. Benmimoune, F. A. Khasawneh, and M. Kadoch, "User association for HetNet small cell networks," in *Proc. IEEE Conf. FiCloud*, Aug. 2015, pp. 113–117.

[8] N. Y. Ermolova and B. Makarevitch, "Performance of practical subcarrier allocation schemes for OFDMA," in *Proc. IEEE PIMRC*, Sep. 2007, pp. 1–4.

[9] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, "A proportional fair radio resource allocation for heterogeneous cellular networks with relays," in *Proc. IEEE GLOBECOM*, Anaheim, CA, USA, Dec. 2012, pp. 1–5.

[10] Y. Li, M. Sheng, Y. Sun, and Y. Shi, "Joint optimization of BS operation, user association, subcarrier assignment, and power allocation for energy-efficient HetNets," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3339–3353, Dec. 2016, doi: 10.1109/JSAC.2016.2600581.

[11] Y. Chen, J. Li, W. Chen, Z. Lin, and B. Vucetic, "Joint user association and resource allocation in the downlink of heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5701–5706, Jul. 2016.

[12] P. Bhat *et al.*, "LTE-advanced: An operator perspective," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 104–114, Feb. 2012.

[13] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[14] D. Lopez-Perez, X. Chu, and J. Zhang, "Dynamic downlink frequency and power allocation in OFDMA cellular networks," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2904–2914, Oct. 2012.

[15] H. Tabassum, Z. Dawy, and M. Alouini, "Sum rate maximization in the uplink of multi-cell OFDMA networks," in *Proc. IEEE Conf. IWCMC*, Jul. 2011, pp. 1152–1157.

[16] S. Y. Kim, J. A. Kwon, and J. W. Lee, "Sum-rate maximization for multicell OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4158–4169, Sep. 2015.

[17] J. Ghimire and C. Rosenberg, "Resource allocation, transmission coordination and user association in heterogeneous networks: A flow-based unified approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1340–1351, Mar. 2013.

[18] D. West, *Introduction to Graph Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2001.

[19] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics*, vol. 52, no. 1, pp. 7–21, Feb. 2005.

[20] Y. Liu and W. Chen, "Limited-feedback-based adaptive power allocation and subcarrier pairing for OFDM DF relay networks with diversity," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2559–2571, Jul. 2012.

[21] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation for secure OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2572–2585, Jul. 2012.

[22] P. Coucheney, C. Touati, and B. Gaujal, "Fair and efficient user-network association algorithm for multi-technology wireless networks," in *Proc. IEEE Conf. INFOCOM*, Apr. 2009, pp. 2811–2815.

[23] L. Wang, W. Chen, and J. Li, "Congestion aware dynamic user association in heterogeneous cellular network: A stochastic decision approach," in *Proc. IEEE Conf. ICC*, Jun. 2014, pp. 2636–2640.

[24] K. L. Hoffman and T. K. Ralphs, "Integer and combinatorial optimization," in *Encyclopedia of Operations Research and Management Science*, Springer US, pp. 771–783, 2013.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

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

[27] H. Tang, W. Chen, and J. Li, "Robust joint source-relay-destination design under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2639–2649, May 2015.

[28] A. Khabbazibasmenj and S. A. Vorobyov, "Robust adaptive beamforming for general-rank signal model with positive semi-definite constraint via POTDC," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 6103–6117, Dec. 2013.

[29] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul. 2007.

[30] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[31] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.

[32] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, Jan./Feb. 1997.

[33] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Network.*, vol. 13, no. 3, pp. 636–647, Jun. 2005.

**Wen Chen** (M'03–SM'11) received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999. From 1999 to 2001, he was a Researcher with the Japan Society for the Promotion of Sciences. In 2001, he joined the University of Alberta, Edmonton, AB, Canada, as a PIMS Post-Doctoral Fellow with the Information Research Laboratory, and then was as a Research Associate with the Department of Electrical and Computer Engineering. Since 2006, he has been a Full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, where he is also the Director of the Institute for Signal Processing and Systems. From 2014 to 2015, he was the Dean of the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, China. Since 2016, he has been the Chairman of Intellectual Property Management Corporation, Shanghai Jiao Tong University. He has authored 79 papers in the IEEE journals and transactions and over 100 papers in the IEEE conference publications. His research interests include multiple access, network coding, cooperative communications, and multiple-input-multiple output orthogonal-frequency-division multiplexing systems.

He received the Ariyama Memorial Research Prize in 1997 and the PIMS Post-Doctoral Fellowship in 2001. He received the honors of a New Century Excellent Scholar in China in 2006 and a Pujiang Excellent Scholar in Shanghai in 2007. He is elected to the Vice General Secretary of the Shanghai Institute of Electronics in 2008. He received the Best Service Award of the China Institute of Electronics in 2013, the Best Paper Award of Chinese Information Theory Society in 2014, and the Innovate 5G Competition Award in 2015, and was invited to deliver a keynote speech in APCC2016.

Dr. Chen was the TPC Chair of the IEEE ICCSC2008, the IEEE WiMob2012, and the General Chair of the IEEE ICIS2009, the IEEE WCNIS2010, and the IEEE WiMob2011. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of the IEEE ACCESS, and served as a Lead Guest Editor for the *EURASIP Journal on Wireless Communications and Networking* and the IEEE/CIC CHINA COMMUNICATIONS.



**Hongying Tang** received the Ph.D. degree from Shanghai Jiao Tong University, China, in 2015. She is currently an Assistant Professor with the Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences. Her research interests include cognitive radio and digital predistortion.



**Feng Wang** received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2010, where he is currently pursuing the Ph.D. degree with the Network Coding and Transmission Laboratory. His research interests include resource allocation in heterogeneous networks and interference management.



**Qingqing Wu** (S'13) received the B.S. degree in electronic engineering from the South China University of Technology, Guangzhou, China, in 2012. Since 2015, he has been a Visiting Scholar with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. He is currently pursuing the Ph.D. degree with the Network Coding and Transmission Laboratory, Shanghai Jiao Tong University. His current research interests include energy efficient optimization and wireless information and power transfer. He served as a TPC Member of the Chinacom 2015, the IEEE ICCC 2016, the IEEE Globecom 2016, and the IEEE VTC 2017. He was a recipient of the International Conference on Wireless Communications and Signal Processing Best Paper Award in 2015.