Resource Allocation Scheme for Energy Saving in Heterogeneous Networks

Xun Sun and Shaowei Wang, *Senior Member, IEEE*

***Abstract*—Energy efficiency in communication networks has received increasing attention in both industry and academia. In this paper, we investigate the energy saving issue in a heterogeneous network (HetNet), which is introduced to cellular radio networks to improve capacity and enhance coverage. A HetNet consists of base stations (BSs) with different transmission powers, resulting in systematic power control that is more complex than that in the conventional cellular networks. The main difficulty is addressing the mutual interference between different kinds of BSs. In this paper, we try to minimize the power consumption of an OFDM-based HetNet while satisfying all users' rate requirements, as well as considering the inter-cell interference. Our general problem formulation leads to a nonconvex optimization task that is generally hard to tackle. We derive a concave lower bound of user's achievable rate for a given power allocation, based on which an efficient iterative algorithm is developed to solve the formulated problem efficiently. Numerical results show that our proposed resource allocation scheme works well in different scenarios. The energy consumption of the cellular system is reduced dramatically compared to other schemes. Moreover, our proposed algorithm converges quickly and stably, showing great potential for applications.**

***Index Terms*—Energy efficiency, heterogeneous network, interference management, nonconvex optimization.**

I. Introduction

Wireless data traffic is dramatically growing and the monthly demand is forecasted to reach 6.3 EB on 2015, a 26-fold increase over 2010 [1]. However, as a scarce natural resource, radio spectrum is very crowded in the band for mobile communications, which is becoming a bottleneck to develop diverse wireless applications. Though some promising spectrum utilization schemes have been proposed to address spectrum crisis [2], [3], most of them are far from implementation because of technical limitations. A practical solution is Heterogeneous Network (HetNet), which consists of various radio access nodes, such as macro Base Stations (BSs), pico BSs, femto BSs and relays. These nodes have different capacities and operating functionalities, which can potentially improve spectrum efficiency significantly by enhancing area spectrum reuse. On the other hand, HetNet also entails a new paradigm shifting from traditional centralized macro BS only framework to a more autonomous, uncoordinated, and intelligent one, resulting in more complex interference management and power control issues that should be properly addressed.

Besides improving system throughput, HetNet also has a potential from the viewpoint of energy saving. Statistical data show that Information and Communication Technology (ICT) industry is an increasingly key contributor with an 8% of the worldwide energy consumption in 2008, and is expected to double by 2020. Particularly, mobile communication systems occupy 0.5% of global energy consumption [4]. The potential energy saving of the HetNet comes from the fact that low power access nodes, such as pico BSs, femto BSs and relays, are generally much closer to end users than macro BSs, thus the radio links between these low power access nodes and the end users suffer lower path losses as compared to the links between the macro BSs and the end users. As a result, the low power access nodes can not only offload the traffic of the macro BSs, but also reduce the total power consumption of the cellular system while satisfying the same users' rate requirements.

Energy consumption issue has been investigated extensively in recent years. To increase the energy efficiency (EE) of the cellular systems, traffic-aware transmission strategies have been proposed in [5]–[7], where underutilized BSs are recommended to switch to sleep mode or be shut off during off-peak time of traffic loads. It is practical because the deployments of existing cellular networks are always designed for peak load traffics, leading to very inefficient usage of BSs during off-peak time. The cell discontinuous transmission strategies proposed in these works are recognized as promising approaches to improve the EE of the cellular system. There are also energy saving schemes from the perspective of network deployment. In [8], *bit/Joule* is introduced to evaluate the EE of a cognitive radio system, based on which an efficient resource allocation algorithm is developed. In [9], the optimal BS density for both homogeneous and heterogeneous cellular networks is analyzed, where all BSs transmit data with fixed power and the objective is to minimize the total energy cost. With variable BS transmission powers, the BS density should be re-optimized and the network energy consumption can be significantly reduced, as show in [10], which focuses on the energy efficient deployment for both homogeneous and heterogeneous cellular networks under coverage performance constraints by using stochastic geometry tools.

As a widely investigated topic in Orthogonal Frequency Division Multiplexing (OFDM)-based cellular networks, resource allocation also plays an important role in interference

Manuscript received December 23, 2013; revised September 17, 2014, January 5, 2015, and March 19, 2015; accepted March 25, 2015. Date of publication April 6, 2015; date of current version August 10, 2015. This research was partially supported by the NSFC and JiangsuSF. Part of this work was presented at the IEEE WCNC 2013, Shanghai, China, April 7–10, 2013. The associate editor coordinating the review of this paper and approving it for publication was S. Cui. *(Corresponding author: Shaowei Wang).*

The authors are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: wangsw@nju.edu.cn; sxun@smail.nju.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2015.2420558

management and energy saving. In [11], fundamental tradeoff between EE and spectral efficiency (SE) in the OFDM system is studied. Energy-efficient spectrum sharing in heterogeneous cognitive radio network with femto BSs is investigated in [12], where Stackelberg game is employed to address the formulated problem. In [13], an energy-efficient power optimization scheme is presented for interference-limited communication. In [14], non-cooperative game is introduced to perform subchannel assignment, followed by joint multi-cell power allocation to minimize the power consumption of each user. Generally, full channel state information (CSI) is necessary for the mentioned resource allocation schemes. In a HetNet, femto BSs are typically consumer deployed (unplanned) access nodes for indoor application with a network backhaul facilitated by the consumers home digital subscriber line (DSL) or cable modem [15]. However, due to the feedback latency and limited signal overhead that can be exchanged among BSs, perfect CSI may not usually be known by all BSs. In [16], the authors designed a game-theoretical resource allocation scheme considering both EE and interference control in the HetNet with incomplete CSI. As for the femto BSs installed by users in an *ad-hoc* manner, the mutual interference between macro BSs and femto BSs makes the resource allocation task challenging [17], [18]. Interference management algorithms for femto BSs can be found in [19]–[21].

Most of the previous works focus on the improvement of EE rather than directly reducing the energy consumption. It is reasonable since maximizing the EE of the cellular network certainly reduces the energy consumption for a given system throughput target. As shown in [11], the optimal EE of an OFDM system can be reached in the case that the system throughput equals to a specific value. However, in most situations, the total rate requirement of users is not equal to a specific one, which means that it is impossible to get the maximum EE. Moreover, even if the sum rate of all users happens to reach the optimal point that yields the maximum EE, each user's rate requirement may not be satisfied at this point. It is of great concern to satisfy the QoS requirements of users for service providers. Therefore, it is more reasonable to minimize the total energy consumption of the cellular system while satisfying all users' QoS requirements, which differs from the objective of maximizing the system throughput and is also the motivation of this work. We try to save energy as much as possible while keeping all users' achievable rates above their required thresholds. By exploiting the properties of the lower bound of the user's rate, the formulated nonconvex optimization task can be relaxed to a convex optimization problem, which we develop an iterative algorithm to address efficiently. At each iteration, we minimize the total power consumption to produce power and subchannel allocation solutions with dual decomposition. It should be noted that in this work, perfect CSI exchanges between BSs are assumed, the study of imperfect CSI is out of the scope of this paper and the related research can be found in [16]. Our proposed schemes not only reduces the energy consumption of the system greatly but also converges quickly and stably, which are verified by numerical results.

The rest of this paper is organized as follows. In Section II, we give system model and formulate our optimization task.

TABLE I
NOTATIONS

Symbol	Semantics
$\mathcal{N}_{macro}$	The sets of macro BSs
$\mathcal{N}_{pico}$	The sets of pico BSs
$\mathcal{N}$	The sets of all BSs
N	The number of all BSs
$P_{n,max}$	The maximum transmission power of BS n
B	The total bandwidth
$\mathcal{K}$	The sets of users
K	The number of users
$R_{k,min}$	The required rate of user k
$g_s^{k,n}$	The channel gain between BS n to user k on subchannel s
Γ	The SINR gap
σ_s^k	Noise power spectrum density
$A_{n,s,k}$	The subchannel assignment index
p_s^n	The proportion of power allocated to subchannel s by BS n

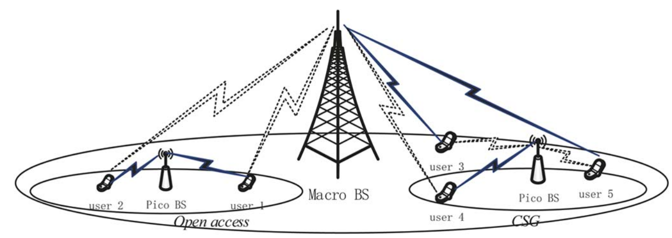

Fig. 1. An exemplary system setup, where the solid links and the dashed links represent communication links and cochannel interfering links for users in system.

In Section III, we propose resource allocation algorithms. Numerical results are given in Section IV, as well as discussions. Conclusion is given in Section V.

II. System Model and Problem Formulation

A. System Model

The frequently used terminologies and notations are given in Table I. Consider a heterogeneous cellular network shown in Fig. 1. The sets of macro BSs and low power BSs (pico BSs in Fig. 1) are denoted as $\mathcal{N}_{macro}$ and $\mathcal{N}_{low}$, respectively. The total transmission power of BS n is limited to $P_{n,max}$. For practical cellular systems, the maximum transmission power of a macro BS (e.g., 46 dBm) is much higher than that of a low power BS (e.g., 30 dBm). As a result, the coverage area of the low power BS is much smaller than that of the macro BS. For the low power BS, there are two kinds of access mode: *open access* and *closed subscriber groups (CSG)* [22]. For the *open access* mode, users are allowed to connect to either a macro BS or a low power BS. That is, if a user is within the coverage of a low power BS, it will be served by the low power BS (user 1, user 2 in Fig. 1); otherwise, it will be served by the macro BS (users 3 in Fig. 1). For the *CSG* mode, only subscribed users can connect to the low power BS. As shown in Fig. 1, user 4 is a subscribed user which connects to the pico BS. However, user 5 can only be served by the macro BS even though it is in the coverage of pico BS because it is not in the subscribed user group.

Denote $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{N} = \mathcal{N}_{macro} \cup \mathcal{N}_{low} = \{1, \ldots, N\}$ as the set of users and BSs, respectively. Each user can be served by only one BS. Denote $\mathcal{K}_n$ as the set of users associated with BS n, i.e., $\mathcal{K} = \mathcal{K}_1 \cup \cdots \cup \mathcal{K}_n$ and $\mathcal{K}_n \cap \mathcal{K}_m = \varnothing$, for $n \neq m$. The total bandwidth is B and is available for both macro BSs and low power BSs. The bandwidth is divided into S OFDM subchannels. The set of subchannels is denoted as $\mathcal{S} = \{1, \ldots, S\}$. Each user $k \in \mathcal{K}$ requires a minimal transmission rate $R_{k,min}$. We denote the channel power gain on subchannel s from BS n to user k by $g_s^{k,n}$. The number of users served by each BS is smaller than S, and each subchannel of any BS can be allocated to only one user. We ignore the feedback latency in the message exchange and assume the BSs know $\{g_s^{k,n}, \forall\, n, \forall\, s, \forall\, k\}$ precisely.

Each BS needs to determine: (i) which user should be scheduled on subchannel n; and (ii) how much power should be allocated for the scheduled user on this subchannel. We define a binary variable $A_{n,s,k}$ as subchannel assignment index,

$$A_{n,k,s} = \begin{cases} 1 & \text{subchannel } s \text{ of BS } n \text{ is allocated to user } k, \\ 0 & \text{otherwise.} \end{cases}$$

We stack $\{A_{n,k,s}\}_{k=1}^{K}$ into the vector $\mathbf{A}_{n,s} = [A_{n,1,s}, \ldots, A_{n,K,s}]$, and stack $\mathbf{A}_{n,s}$'s of all users in BS n to a matrix $\mathbf{A}_n$ column by column. Finally, we define a matrix $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_N]$, which indicates how subchannels are allocated to all users. Since each subchannel for a BS can only be used by at most one user, we have:

$$\sum_{k \in \mathcal{K}_n} A_{n,k,s} \leq 1,\ \forall\, n \in \mathcal{N}, s \in \mathcal{S}. \tag{1}$$

Define p_s^n as the proportion of $P_{n,max}$ allocated to subchannel s by BS n. We stack $\{p_s^n\}_{n=1}^{N}$ into an $N \times 1$ vector $\mathbf{p}_s = [p_s^1, \ldots, p_s^N]$. Each BS has a power budget:

$$\sum_{s=1}^{S} p_s^n \leq 1,\ \forall\, n \in \mathcal{N}. \tag{2}$$

In this paper, we do not consider advanced multiuser detection or interference cancellation, and the interference from other BSs is treated as noise. Obviously, if special interference management techniques were employed, the performance of our proposed scheme could be improved further. For a given power vector $\mathbf{p}_s$, the rate of user k served by BS n can be written as

$$R_k = \sum_{s=1}^{S} A_{n,k,s} \cdot R_k(\mathbf{p}_s), \tag{3}$$

where $R_k(\mathbf{p}_s)$ stands for the rate on subchannel s allocated to user k in BS n, and can be calculated as follows,

$$R_k(\mathbf{p}_s) = \frac{B}{S} \ln\left(1 + \frac{\gamma_{n,k,s}(\mathbf{p}_s)}{\Gamma}\right), \tag{4}$$

where $\gamma_{n,k,s}(\mathbf{p}_s)$ is the signal-to-interference-plus-noise ratio (SINR), and Γ represents the SINR gap to capacity which is typically a function of the desired bit error ratio (BER), the coding gain and noise margin, e.g., $\Gamma = \frac{-\ln(5BER)}{1.5}$ in MQAM [23]. We assume $\Gamma = 1$ here for the study of modulation is out of scope of this work. Particularly, $\gamma_{n,k,s}(\mathbf{p}_s)$ can be calculated as

$$\begin{aligned} \gamma_{n,k,s}(\mathbf{p}_s) &= \frac{g_s^{k,n} p_s^n P_{n,max}}{\sum_{m \neq n} g_s^{k,m} p_s^m P_{m,max} + \sigma_s^k} \\ &= \frac{p_s^n}{I_{n,k,s}(\mathbf{p}_s)}, \end{aligned} \tag{5}$$

where

$$I_{n,k,s}(\mathbf{p}_s) = x_{n,k,s} + \sum_{m \neq n} p_s^m y_{n,k,s},$$

$$x_{n,k,s} = \frac{\sigma_s^k}{g_s^{k,n} P_{n,max}},$$

and

$$y_{n,k,s} = \frac{g_s^{k,m} P_{m,max}}{g_s^{k,n} P_{n,max}}.$$

where σ_s^k is the noise power spectrum density.

B. Problem Formulation

Stack $\{\mathbf{p}_s\}_{s=1}^{S}$ into an $N \times S$ matrix $\mathbf{P}$, and we refer to $\mathbf{P}$ as the power allocation for the sake of convenience. Our target is to find the optimal $\mathbf{A}$ and $\mathbf{P}$ to minimize the total transmission power of BSs, while keeping the rate of each user above a predefined threshold. Specifically, the objective function is expressed as

$$f(\mathbf{A}, \mathbf{P}) = \sum_{n \in N} \sum_{s \in S} p_s^n P_{n,max}. \tag{6}$$

Mathematically, our optimization problem can be formulated as follows,

$$\begin{aligned} \min_{\mathbf{A},\mathbf{P}} \quad & f(\mathbf{A}, \mathbf{P}) \\ s.t. \quad & C_1 : \textstyle\sum_{s=1}^{S} p_s^n \leq 1,\ \forall\, n \in \mathcal{N}, \\ & C_2 : \sum_{k \in \mathcal{K}_n} A_{n,k,s} \leq 1,\ \forall\, n \in \mathcal{N},\ \forall\, s \in \mathcal{S}, \\ & C_3 : A_{n,k,s} \in \{0, 1\},\ \forall\, n \in \mathcal{N},\ \forall\, s \in \mathcal{S},\ \forall\, k \in \mathcal{K}, \\ & C_4 : R_k \geq R_{k,min},\ \forall\, k. \end{aligned} \tag{7}$$

C1 is the power constraint of each BS. C4 specifies user's minimum rate requirement $R_{k,min}$. It is notable that R_k is a nonconvex function associated with power allocation variables p_s^n's, which makes the formulated problem a nonconvex one. C2 and C3 are imposed to guarantee that each subchannel is used by only one user. Such optimization task of minimizing power with rate constraints is closely related to the problem of maximizing rate with power constraints in [24], but the

TABLE II
SUBCHANNEL ALLOCATION

1:	For $k = 1$ to K;
2:	$r_{k,n} = \max\{r_{k,1}, ...r_{k,N}\},\ n \in \mathcal{N}_{macro}$, $r_{k,m} = \max\{r_{k,1}, ...r_{k,M}\},\ m \in \mathcal{N}_{low}$;
3:	If $r_{k,n} \le r_{k,m}$, $k \in \mathcal{K}_m$, otherwise $k \in \mathcal{K}_n$;
4:	Endfor;
5:	**Initialization:**
6:	$\mathcal{S}_t$=$\mathcal{S}$, Ω_k=$\emptyset$, $\forall k$;
7:	Set the users' rate to zero: $R_k = 0$, for $k \in \mathcal{K}_n$;
8:	**Step 1**
9:	*While* $\mathcal{S}_t \ne \emptyset$ and $R_k < R_{k,min}$, $k \in \mathcal{K}_n$
10:	Find k^* satisfies $R_{k^*} - R_{k,min} \le R_k - R_{k,min}$ for all $k \in \mathcal{K}_n$;
11:	Find s^* satisfies $\gamma_{n,k^*,s^*} \ge \gamma_{n,k^*,s}$, $\forall s \in \mathcal{S}_t$;
12:	Update $R_{k^*} = R_{k^*} + \log(1 + \gamma_{n,k^*,s^*})$;
13:	Update $\Omega_{k^*} = \Omega_{k^*} \cup s^*$,$\mathcal{S}_t = \mathcal{S}_t \setminus s^*$, $A_{n,k^*,s^*} = 1$;
14:	*Endwhile*
15:	**Step 2**
16:	*For* $i = 1$ to length($\mathcal{S}_t$);
17:	For $s^* = \mathcal{S}_{t_i}$, find s^* satisfies $\gamma_{n,k^*,s^*} \ge \gamma_{n,k^*,s}$;
18:	Update $\Omega_{k^*} = \Omega_{k^*} \cup s^*$, $A_{n,k^*,s^*} = 1$.
19:	*Endfor*

interference introduced by other BSs makes this problem much more difficult.

III. Our Proposed Algorithm

Equation (7) is a nonconvex problem since the transmit power is coupled with integer variable $A_{n,k,s}$'s. The problem defined by Eq. (7) is a mixed-integer nonlinear programming (MINLP), which is generally computationally intractable. It is known that even if the simplified problem, in which subchannel scheduling issue is eliminated by relaxing the integer constraints to continuous ones, is hard to obtain the optimal solution because the remaining optimization task is nonconvex. To find global optimal solution, we need to fully search the space of the feasible power allocation for all BSs with a small granularity along with all possible combinations of subchannel allocations. Thus, it is infeasible for practical cellular systems to solve Eq. (7) directly. In this paper, we deal with it with a two-step procedure: subchannel assignment and power optimization.

A. Subchannel Assignment

For the case of *open access*, firstly we need to associate users with BSs, where the total achievable rate that users can get from the BSs are adopted as the measure. This process is shown in line1–4 of Table II, where $r_{k,n}$ and $r_{k,m}$ denote the maximum achievable rate that user k can get from macro BSs and low power BSs, respectively. Here the user rate from one BS is calculated as the sum rate of all subchannels under the condition of average power allocation. User k is served by the BS that can supply the highest rate. In this way, the user set of each BS is determined. Such a BS association scheme has incorporated path-loss fading, shadowing and frequency selective fading, which is more reasonable in realistic scenario. It is straightforward to extend our results to range extension, which increases the coverage of low power BS by adding a positive bias to their signal-strengths during BS association [25]. For the case of *CSG*, user set of each BS is fixed, and the subchannel assignment just starts from line 5 in Table II.

By initializing the power vector as $[\mathbf{P}^0]_{n,s} = (1/S),\ \forall n,\ \forall s$, we propose an efficient subchannel allocation method to figure out the binary variables $A_{n,k,s}$'s, specifying a subchannel assignment for each BS n. Ω_k is the set of subchannels occupied by the kth user, which cannot be reused by other users served by the same cell. The proposed subchannel allocation algorithm consists of two steps: Step 1, we allocate subchannels to users to meet their minimal rate requirements; Step 2, we allocate each subchannel in the remaining ones to the users with the highest SINR over it. The intuitiveness that lies in Step 1 is that the user whose current rate is the farthest away from its target has a priority to get a subchannel among all available ones. And the procedure continues until all users' rate requirements are satisfied. Preferably, the subchannel with the highest achievable rate associated with a user will be chosen at this step. At Step 2, each of the remaining subchannels is allocated to the user who has the highest achievable rate over it to potentially maximize the sum capacity of the system.

Notice that in [26], [27], subchannel assignment and power allocation are implemented in an integral algorithm with a compact form. However, these algorithms have a prohibitively high computational complexity due to inner and outer loops. Also, it requires multiple information exchanges per slot between BSs to reflect the updated interference level followed by the updated power. In this paper, we try to design a low complexity algorithm without information exchanges during the running of the algorithm.

B. Power Allocation Optimization

For a given subchannel assignment $\mathbf{A}$, the original problem degenerates to the following power allocation one:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \sum\nolimits_{n\in N} \sum\nolimits_{s\in S} p_s^n P_{n,max} \\ s.t. \quad & C_1 : \sum\nolimits_{s=1}^{S} p_s^n \le 1, \forall n \in \mathcal{N}, \\ & C_2 : R_k \ge R_{k,min}, \forall k. \end{aligned} \tag{8}$$

However, there is still no efficient algorithm to solve Eq. (8) because the user rate R_k is nonconvex, as a result of the existence of interference. To tackle this problem, we establish a concave lower bound of user rate to approximate the original problem. If the lower bound is tight, we can obtain promising solutions to the considered optimization problem.

1) A Concave Lower Bound of User Rate: The concave lower bound of user rate is associated with a given power allocation $\mathbf{P}^*$ [27]. Denote $\mathbf{p}_s^*$ as the sth column of $\mathbf{P}^*$, and the corresponding SINR and user rate are $\gamma_{n,k,s}(\mathbf{p}_s^*)$ and $R_{n,k,s}(\mathbf{p}_s^*)$, respectively. Define

$$R_{n,k,s}^*(e^{\mathbf{q}}) = \alpha_{n,k,s}^* \cdot \ln\left(\gamma_{n,k,s}(e^{\mathbf{q}})\right) + \beta_{n,k,s}^*, \tag{9}$$

$$\alpha_{n,k,s}^* = \frac{\gamma_{n,k,s}\left(\mathbf{p}_s^*\right)}{\Gamma + \gamma_{n,k,s}\left(\mathbf{p}_s^*\right)}, \tag{10}$$

$$\beta_{n,k,s}^* = R_{n,k,s}\left(\mathbf{p}_s^*\right) - \alpha_{n,k,s}^* \cdot \ln\left(\gamma_{n,k,s}(\mathbf{p}_s^*)\right), \tag{11}$$

where $\mathbf{q}$ is an $N \times 1$ vector. Note that $\mathbf{P}^*$ belongs to the set $P_+ = \{P | \forall n, \forall s, p_{n,s} \in (0, 1]\}$, to guarantee that $\forall k, \forall n, \forall s, \gamma_{n,k,s}(\mathbf{P}_s^*) > 0, \alpha_{n,k,s}^* > 0$.

To exploit the properties of $R_{n,k,s}^*(e^{\mathbf{q}})$, we first show the following mathematical fact:

Fact 1: Given positive values z and z^*, we have

$$\ln(1+z) \geq \ln(1+z^*) + \frac{z^*}{1+z^*}\left(\ln(z) - \ln(z^*)\right), \quad (12)$$

where the equality holds when $z = z^*$ [28]. Then we can have the following theorem:

Theorem 1: $R_{n,k,s}(e^{\mathbf{q}}) \geq R_{n,k,s}^*(e^{\mathbf{q}})$ and the equality holds when $e^{\mathbf{q}} = \mathbf{p}_s^*$; $R_{n,k,s}^*(e^{\mathbf{q}})$ is a concave function of $\mathbf{q}$.

Proof: The first claim can be easily got by Fact.1. Particularizing z and z^* with $\gamma_{n,k,s}(e^{\mathbf{q}})/\Gamma$ and $\gamma_{n,k,s}(\mathbf{p}_s^*)/\Gamma$, respectively. $R_{n,k,s}(e^{\mathbf{q}}) \geq R_{n,k,s}^*(e^{\mathbf{q}})$ follows and the equality holds when $e^{\mathbf{q}} = \mathbf{p}_s^*$.

To prove the second claim, we expand $R_{n,k,s}^*(e^{\mathbf{q}})$ as

$$R_{n,k,s}^*(e^{\mathbf{q}}) = \alpha_{n,k,s}^* \cdot \left(q_n - \ln\left(I_{n,k,s}(e^{\mathbf{q}})\right)\right) + \beta_{n,k,s}^* \quad (13)$$

where q_n is the nth entry of $\mathbf{q}$. $\ln\left(I_{n,k,s}(e^{\mathbf{q}})\right) = \ln(x_{n,k,s} + \sum_{m \neq n} y_{n,k,s} e^{q_m})$ is a convex function of $\{q_m\}_{m \neq n}$, details of this proof can be found in Appendix. Since $\alpha_{n,k,s}^* > 0$, it can be easily shown that $R_{n,k,s}^*(e^{\mathbf{q}})$ is a concave function of $\mathbf{q}$. ■

Define

$$R_k^*\left(A_{n,k,s}, e^{\mathbf{Q}}\right) = \sum_{n=1}^{N}\sum_{s=1}^{S} A_{n,k,s} \cdot R_{n,k,s}^*(e^{\mathbf{q}}), \quad (14)$$

where $\mathbf{Q}$ is an $N \times S$ matrix whose the sth column is $\mathbf{q}_s$. Based on Theorem 1, we know that $R_k^*(A_{n,k,s}, e^{\mathbf{Q}})$ is a concave function of $\mathbf{Q}$, as well as a lower bound of R_k. Besides, the bound is tight when $e^{\mathbf{Q}} = \mathbf{P}^*$. In this way, we can transform the original nonconvex problem into a convex relaxation.

2) Power Allocation Algorithm: Define the sets $\boldsymbol{P}_\epsilon = \{\mathbf{P} | \forall n, \forall s, P_{n,s} \in [e^\epsilon, 1]\}$ and $\boldsymbol{Q}_\epsilon = \{\forall n, \forall s, q_{n,s} \in [\epsilon, 0]\}$, ϵ is a prescribed negative value. By replacing R_k with $R_k^*(A_{n,k,s}, e^{\mathbf{Q}})$, we can get a convex approximation to Eq. (8) as follows,

$$\begin{aligned} \min_{\mathbf{Q}} \quad & \sum_{n=1}^{N}\sum_{s=1}^{S} e^{q_{n,s}} \cdot P_{n,max} \\ s.t. \quad & C_1 : \mathbf{Q} \in \boldsymbol{Q}_\epsilon, \\ & C_2 : \sum\nolimits_{s=1}^{S} e^{q_{n,s}} \leq 1, \ \forall n \in \mathcal{N}, \\ & C_3 : R_k^*(A_{n,k,s}, e^{\mathbf{Q}}) \geq R_{k,min}, \ \forall k. \end{aligned} \quad (15)$$

Once we get a feasible power allocation for the original problem, define it as $\mathbf{P}^*$ and solve Eq. (15) for optimal result $\mathbf{Q}_0 \in \boldsymbol{Q}_\epsilon$. As $\mathbf{Q}_0$ is feasible for Eq. (15), and $R_k^*(A_{n,k,s}, e^{\mathbf{Q}_0}) \leq R_k(A_{n,k,s}, e^{\mathbf{Q}_0})$, then $\mathbf{P}_0 = e^{\mathbf{Q}_0}$ is guaranteed to be feasible for Eq. (8). Since $\mathbf{P}^* \in \mathbf{P}_\epsilon$, there must exist a unique element $\mathbf{Q}^* \in \boldsymbol{Q}_\epsilon$ which satisfies $\mathbf{P}^* = e^{\mathbf{Q}^*}$. We can also find $f(\mathbf{A}, e^{\mathbf{Q}_0}) \leq f(\mathbf{A}, e^{\mathbf{Q}^*})$ holds because $\mathbf{Q}_0$ and $\mathbf{Q}^*$ are optimal and feasible to Eq. (15), which means $\mathbf{P}_0$ is at least as good as $\mathbf{P}^*$.

TABLE III
POWER ALLOCATION OPTIMIZATION ALGORITHM

Initialize: $i = 0$, $\mathbf{P}^0 = \mathbf{P}$;
repeat
Solve (15) with $\mathbf{P}^* = \mathbf{P}^i$ for $\mathbf{Q}_0$ using Duality-based Algorithm;
$\mathbf{P}^{i+1} = e^{\mathbf{Q}_0}$;
$i = i + 1$;
until $\lVert vec(\mathbf{Q}_0 - \mathbf{Q}^*) \rVert < \Delta_1$ or $i = I$
Output: $\mathbf{P} = \mathbf{P}^i$.

Based on the analysis above, we proposed an iterative power allocation algorithm for a given $\mathbf{A}$, which is depicted in Table III, where i denotes the iteration number $(i > 0)$, and $\mathbf{P}^i$ denotes the tentative power allocation after the ith iteration. In each iteration, we solve Eq. (15) with $\mathbf{P}^* = \mathbf{P}^i$ by a duality-based algorithm to get the optimal solution $\mathbf{Q}_0$, and set $\mathbf{P}^{i+1} = e^{\mathbf{Q}_0}$.

Then we develop a duality-based algorithm to solve Eq. (15). Denote the dual variables related to the power constraint of BS n and the rate constraint of user k by λ_n and ν_k, respectively. We stack all λ_n's and ν_k's into a vector $\omega = [\lambda_1, \ldots, \lambda_N, \nu_1, \ldots, \nu_K]$.

Reformulate the dual function of Eq. (15) as

$$\begin{aligned} D(\omega) &= \min_{\mathbf{Q} \in \boldsymbol{Q}_\epsilon} \sum_{n=1}^{N}\sum_{s=1}^{S} e^{q_{n,s}} P_{n,max} + \sum_{n=1}^{N} \lambda_n \left(\sum_{s=1}^{S} e^{q_{n,s}} - 1\right) \\ &\quad + \sum_{k=1}^{K} \nu_k \left(R_{k,min} - R_k^*(A_{n,k,s}, e^{\mathbf{Q}})\right) \\ &= \min_{\mathbf{Q} \in \boldsymbol{Q}} L(\omega, Q), \end{aligned} \quad (16)$$

and the dual problem is

$$\begin{aligned} \max_{\omega} \min_{\mathbf{Q}} \quad & \sum_{n=1}^{N}\sum_{s=1}^{S} e^{q_{n,s}} P_{n,max} + \sum_{n=1}^{N} \lambda_n \left(\sum_{s=1}^{S} e^{q_{n,s}} - 1\right) \\ & + \sum_{k=1}^{K} \nu_k \left(R_{k,min} - R_k^*(A_{n,k,s}, e^{\mathbf{Q}})\right) \\ & = \max_{\omega} D(\omega). \end{aligned} \quad (17)$$

Define $Q_\omega = \arg\min_{\mathbf{Q} \in \boldsymbol{Q}_\epsilon} L(\omega, Q)$. To solve Eq. (15), we should first find out the optimal dual variable $\omega^* = \arg\max_\omega D(\omega)$, and then compute Q_{ω^*} as the optimal solution to Eq. (15). We can work out ω^* and Q_{ω^*} by using the method proposed in [29]. To this end, we develop a duality-based algorithm as shown in Table IV. Here, i denotes the iteration number $(i \geq 0)$, ω^i represents the dual variables produced after the ith iteration, Δ_2 is a prescribed small positive values.

At the beginning, we initialize the dual variable ω^0. In each iteration, $\mathbf{Q}_{\omega^i}$ can be computed by the gradient-projection algorithm proposed in the next section. When $\mathbf{Q}_{\omega^i}$ is worked out, we can update the dual variable ω^i by solving Eq. (17). Equation (17) can be solved by ellipsoid method [30] or

TABLE IV
THE DUALITY-BASED ALGORITHM TO SOLVE EQ. (15)

```
Initialize: i = 0, ω^0;
    repeat
        Compute Q_{ω^i} with the gradient-projection algorithm;
        Update ω^{i+1} by Eq.(18), Eq.(19);
        i = i + 1;
    until ||ω^i − ω^{i−1}|| < Δ_2
Output: Q_{ω^i} is the optimal solution to Eq.(15).
```

TABLE V
THE GRADIENT-PROJECTION ALGORITHM

```
Initialize;
    If i = 0, t = 0
        Q^t = Q^*;
    else
        Q^t = Q_{ω^{i−1}};
    repeat
        Update B by Eq.(20);
        Q^t = [Q^{t−1} − τ · B]_{Q_ε};
    until ||vec(Q^t − Q^{t−1})|| < Δ_3;
Output: Q_{ω^i} = Q^t .
```

subgradient method [31]–[33]. In this paper, we use subgradient method to update the dual variables as follows,

$$\lambda_n^{i+1} = \left[\lambda_n^i + \varphi^i \left(\sum_{s=1}^{S} e^{q_{n,s}} - 1\right)\right]^+, \tag{18}$$

$$\nu_k^{i+1} = \left[\nu_k^i + \varphi^i \left(R_{k,min} - R_k^*(A_{n,k,s}, e^{\mathbf{Q}})\right)\right]^+, \tag{19}$$

where the superscript i indicates the associated dual variable is an entry in ω^i, $[x]^+ = \max\{0, x\}$, and φ^i is a sufficiently small positive step size for the ith iteration. For example, φ^i can be determined in a diminishing way, such as $\varphi^i = \frac{0.1}{\sqrt{i}}$. The convergence of this duality-based algorithm is guaranteed if φ^i is sufficiently small. The algorithm is terminated when $\|\omega^i - \omega^{i-1}\|$ is smaller than a prescribed small positive value Δ_2, then $\mathbf{Q}_{\omega^i}$ obtained in the last iteration can be taken as the optimal solution to Eq. (15).

Other power updating methods can also be applied by smoothing the objective function or optimizing step sizes, and some of them may lead to even faster convergence. These advanced updating techniques is beyond the scope of this paper and the readers interested in this aspect can refer to [34] and references therein.

3) Gradient-Projection Algorithm to Compute $\boldsymbol{Q}_{\omega^i}$: To find $\mathbf{Q}_{\omega^i}$, we need to solve a constrained optimization problem over the convex set $\boldsymbol{Q}_\epsilon$, namely $Q_{\omega^i} = \arg\min_{\mathbf{Q}\in\boldsymbol{Q}_\epsilon} L(\omega^i, Q)$. Because of the convexity of $L(\omega^i, Q)$, we can adopt a gradient-projection based iterative algorithm, which features simplicity as well as guaranteed convergence to find $\mathbf{Q}_{\omega^i}$ [29].

The algorithm is shown in Table V, where t denotes the iteration times. The algorithm starts with initializing $\mathbf{Q}^t$ by $\mathbf{Q}^*$ if $i = 0$; otherwise by $\mathbf{Q}_{\omega^{i-1}}$. Then, $\mathbf{Q}^t$ is iteratively updated by $\mathbf{Q}^t = [\mathbf{Q}^{t-1} - \tau \cdot B]_{\boldsymbol{Q}_\epsilon}$, where $\mathbf{B}$ is a matrix containing the gradients of $L(\omega^i, Q)$ with respect to every entry of $\mathbf{Q}$, $[\cdot]_{\boldsymbol{Q}_\epsilon}$ is the operator of projection into $\boldsymbol{Q}_\epsilon$, and τ is a prescribed small positive value that guarantees convergence of the gradient-projection based algorithm. By simple mathematical arrangement, every entry of $\mathbf{Q}$ is updated by

$$\forall n, \forall s, q_{n,s} = \begin{cases} \epsilon & \text{if } q_{n,s} - \tau \cdot [\mathbf{B}]_{n,s} \le \epsilon \\ 0 & \text{if } q_{n,s} - \tau \cdot [\mathbf{B}]_{n,s} \ge 0 \\ q_{n,s} - \tau \cdot [\mathbf{B}]_{n,s} & \text{otherwise} \end{cases}$$

where $[\mathbf{B}]_{n,s} = (\partial L(\omega^i, Q))/\partial q_{n,s}$ is derived as

$$[\mathbf{B}]_{n,s} = e^{q_{n,s}} P_{n,max} + \lambda_n^i e^{q_{n,s}} - \sum_{k=1}^{K} \nu_k^i \Bigg(A_{n,k,s}\alpha_{n,k,s}^* - \sum_{m=1,m\ne n}^{N} A_{m,k,s}\alpha_{m,k,s}^* \frac{e^{q_{n,s}} P_{n,max} g_s^{k,n}}{\sum\limits_{j=1,j\ne m}^{N} e^{q_{j,s}} P_{j,max} g_s^{k,j} + \sigma_s^k} \Bigg). \tag{20}$$

The iteration is terminated when $\|vec(\mathbf{Q}^t - \mathbf{Q}^{t-1})\|$ is smaller than a prescribed value Δ_3. Then $\mathbf{Q}^t$ produced in the last iteration is taken as $\mathbf{Q}_{\omega^i}$.

TABLE VI
SIMULATION PARAMETERS

Bandwidth	20*MHz*
Coverage of macro BS	0.5 km
Coverage of pico BS	0.1 km
S	64
Path loss (macro BS)	$131.1 + 42.8log10(R)$ dB, R in km
Path loss (pico BS)	$145.4 + 37.5log10(R)$ dB, R in km
Shadowing standard deviation	10 *dB*
Maximum macro BS transmit power	$40W$
Maximum pico BS transmit power	$1W$

IV. NUMERICAL EXPERIMENTS

Consider the downlink of an OFDM-based HetNet, where pico BSs are placed with uniform intervals around a circle of radius $R_p = 400$ m, and a macro BS is located at the center of the circle. Simulation parameters are listed in Table VI and all results are obtained by averaging 1000 Monte Carlo experiments (unless otherwise specified).

We compare the energy consumption of the HetNet with different modes *(open access and CSG)* and homogeneous network that is only equipped with one macro BS. For the *CSG* mode, equal number of users are uniformly distributed in the macrocell and picocells, which can simulate the scenario of dense users distribution. Each user has the same rate requirement in the HetNet and the homogeneous network. From Fig. 2 we can see that for both modes in HetNet, the energy consumption of the HetNet is less than the homogeneous one. Moreover, the gap between them becomes larger as the number of users grows. The main reason is that more users will be served by the pico BSs as the number of users increases. These users consume much less energy because of the shorter

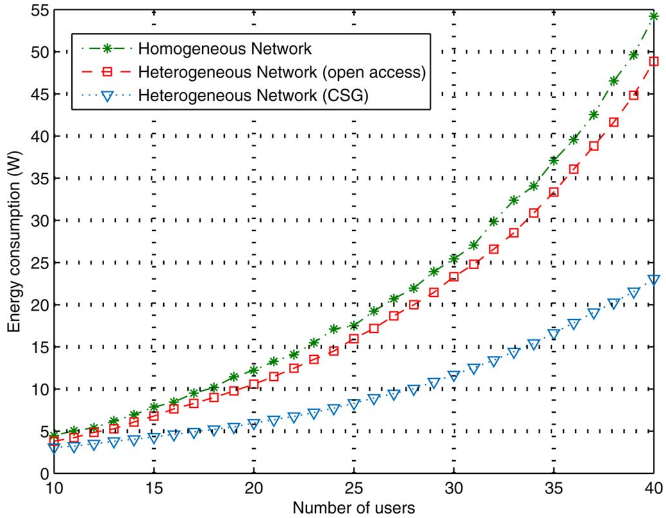

Fig. 2. Energy consumption of homogeneous network and heterogeneous network with different numbers of users.

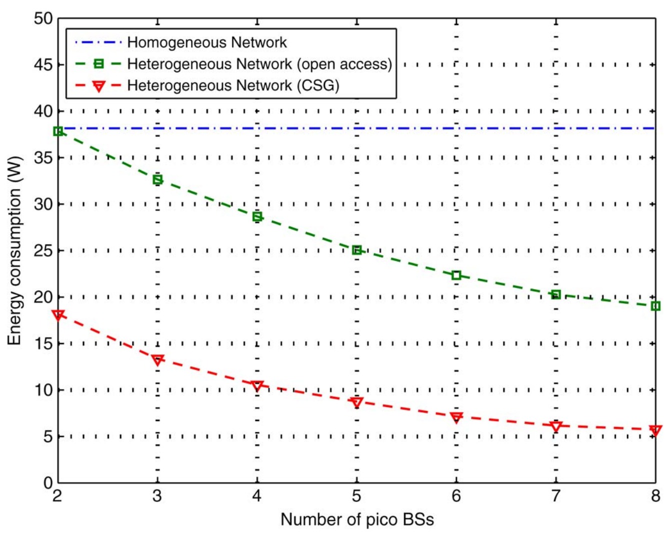

Fig. 3. Energy consumption of heterogeneous network with different numbers of pico BSs.

distance between users and the pico BSs which serve them. It is worthwhile to note that the homogeneous network cannot supply services to all users when the number of users reaches a threshold because the energy consumption exceeds the maximum transmit power (40 W for the macro BS); in contrast, for the *CSG* mode, the HetNet has sufficient power margin to serve more users. We can conclude that deploying pico BSs in the areas with dense users can greatly save the energy.

Notice that the power consumption between the two access modes in the HetNet differs greatly as can be seen from Fig. 2. The reason is that we associate some specific users with the pico BSs for the *CSG* mode while maximum achievable rate association is adopted for the *open access* mode, which results that the number of users served by the pico BSs in the *open access* mode is less than that of the *CSG* mode. In Fig. 3, we can see that the energy consumptions of the HetNet with different number of pico BSs is also less than that of the homogeneous network. However, since the pico BSs for the *CSG* mode is not accessible for all users, the power consumption gap between the two access modes becomes smaller when the number of pico BSs grows because more users can be served by the pico BSs even for the *open access* mode.

TABLE VII
ENERGY CONSUMPTION OF OUR ALGORITHM, UPPER BOUND AND REFIM

User rate $1.5Mbps$	Energy consumption (W)			
	Macro BS	Pico BS1	Pico BS2	Pico BS3
Our algorithm	32.96	0.4142	0.4344	0.4209
Upper bound	30.80	0.3184	0.3412	0.3235
REFIM [26]	36.12	0.8897	0.8887	0.8889

To show the effectiveness of our proposed algorithm, it is more convincing to compare our results to the optimum one or other existing schemes. However, the problem Eq. (7) is NP hard, which is difficult or impossible to get the optimal solution even for a middle scale case. On the other hand, as far as the authors known, previous works in the literature generally focus on maximizing the system throughput or the energy efficiency, which are different from the formulated problem of this work. Instead, we compare our results with an upper bound that is obtained in the following way.

Obtain an upper bound: It is easy to know that an upper bound can be worked out if all mutual interferences are reduced to the least. In this case, the SINR of user k in subchannel s of BS n can be calculated as follows:

$$\gamma_{n,k,s}(\mathbf{p}_s) = \frac{g_s^{k,n} p_s^n P_{n,max}}{x_s^k + \sigma_s^k}. \tag{21}$$

From Eq. (21) we can see that only the lowest interference in the corresponding subchannel from other BSs is involved. It is treated as a constant and denoted by x_s^k. Take Eq. (21) back into Eq. (3) and Eq. (4), then the users' rates can be calculated easily. Moreover, the objective function in Eq. (7) is convex under this condition. Then the original power allocation problem Eq. (7) becomes a convex one and can be solved by using Water-Filling (WF) method [35]. Note that the result obtained in this way is a relaxed solution and can only serve as an upper bound.

Table VII shows the energy consumption of our proposed algorithm with the *CSG* mode, the upper bound and the subchannel allocation scheme proposed in [26] (REFIM). We can see that the energy consumption of our algorithm differs slightly from the upper bound for the macro BS while the gap between our proposal and the REFIM is large. The pico BSs employing the REFIM need about twice energy as much as our proposed algorithm. Fig. 4 shows the energy consumptions of pico BSs (index 1), the macro BS (index 2) and the cellular system (index 3). We can find that the gap between our proposal and the upper bound is less than 7% for the macro BS. Furthermore, our proposed algorithm can reduce more than 12% energy consumption as compared to that proposed in [26]. We can conservatively conclude that our proposed algorithm can produce solutions close to the upper bound and reduce the energy consumption of the cellular system efficiently.

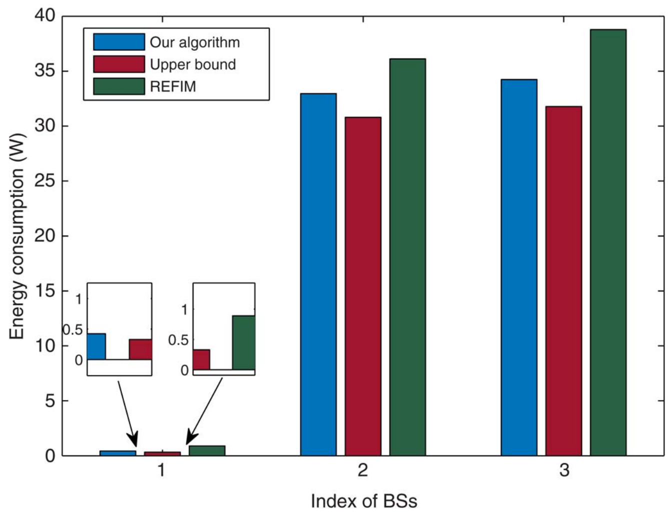

Fig. 4. Energy consumption of pico BSs, macro BS and the cellular system.

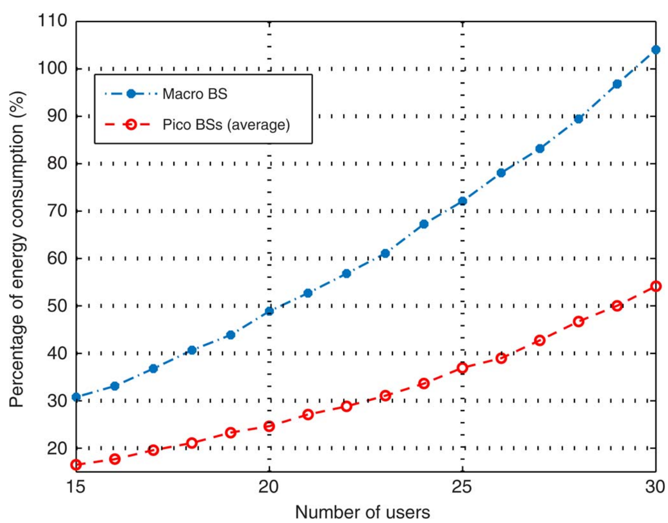

Fig. 5. Percentage of energy consumption as a function of the number of users in each BS.

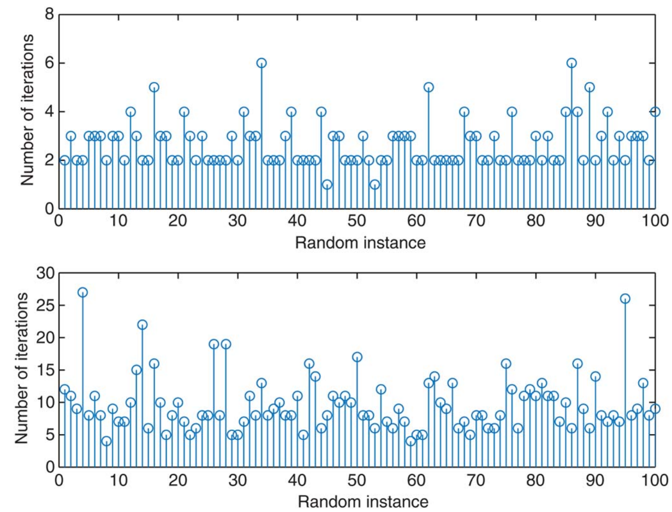

Fig. 6. Number of iterations with different convergence precision. Upper: $\Delta_1 = 10^{-1}$; Lower: $\Delta_1 = 10^{-2}$.

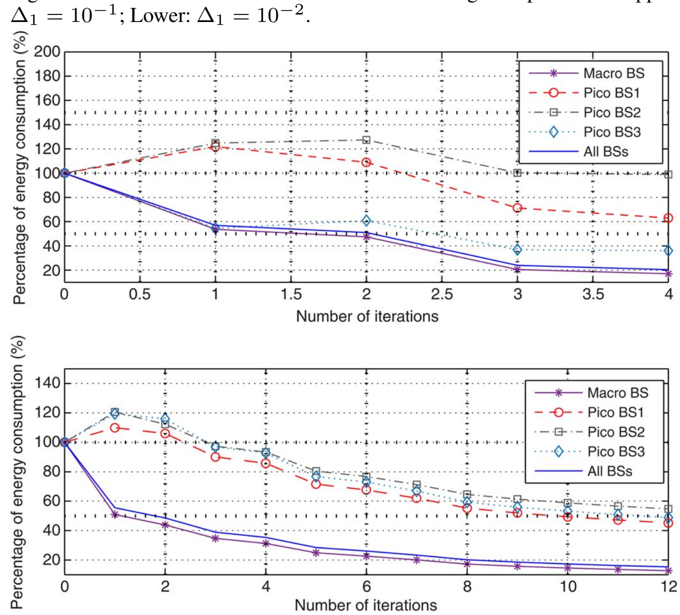

Fig. 7. Power consumption with different convergence precision. Upper: $\Delta_1 = 10^{-1}$; Lower: $\Delta_1 = 10^{-2}$.

Fig. 5 shows the percentage of energy consumption for the macro BS and the pico BSs as the number of users served by each BS increases with the *CSG* mode. The macro BS needs 31% of the maximum power budge for the case of 15 users, which corresponds to 17% for the pico BSs. As the number of users served by each BS increases, the percentage of energy consumption for the macro BS increases more quickly than that for the pico BSs. Even if the the macro BS uses up its power budget, the pico BSs still have power margin to serve more users. It can be explained that the pico BSs usually consume much less power compared to the macro BS to serve a user due to the proximity to the user.

Fig. 6 illustrates the convergence performance of our proposed algorithm. When the convergence precision is 10^{-1}, the number of iterations varies in a narrow range with an average of 3. It is still stable when the convergence precision decreases to 10^{-2}, for which the average number is about 10. Fig. 7 depicts the energy consumption of each BS during iterations with different convergence precisions. We can observe that the energy consumption of the macro BS decreases as the number of iteration increases. Though the transmit power of the pico BSs may increase in some certain iterations (even higher than 100%), the sum power consumption of all BSs keeps decreasing as the number of iterations increases, and the final energy consumptions of the pico BSs are all lower than 100% when the iteration ends as shown in Fig. 7.

V. Conclusion

In this paper, we studied the energy saving problem in heterogeneous networks, which extended our preliminary research [36]. Our optimization objective is to minimize the total energy consumption while guaranteeing the demands of users' rates and considering the mutual interference between different kinds of BSs. We give the lower bound of the user's rate and prove its convexity, based on which we develop efficient algorithms for subchannel assignment and power distribution. The effectiveness and the efficiency of our algorithms are validated by numerical experiments. Since the problem formulation is based on practical scenarios of cellular networks, our proposed resource allocation scheme is competitive for applications.

APPENDIX

We prove the second claim of Theorem 1, where the key is to show that $g(\mathbf{q}) = \ln(I_{n,k,s}(e^{\mathbf{q}})) = \ln(x_{n,k,s} + \sum_{m\neq n} y_{n,k,s}e^{q_m})$ is a convex function of $\mathbf{q} = [q_1, \ldots, q_{N-1}]$. Define $\Theta = [\theta_1, \ldots, \theta_{N-1}]^T = [y_{n,k,s}e^{q_1}, \ldots, y_{n,k,s}e^{q_{N-1}}]^T$, and $\eta = x_{n,k,s} + \sum_i \theta_i$. With intuitive mathematical arrangements, we can derive the Hessian of $g(\mathbf{q})$ with respect to $\mathbf{q}$ as

$$\bigtriangledown^2_{\mathbf{q}}\, g(\mathbf{q}) = \eta^{-2}\left(\eta \cdot diag(\Theta) - \Theta\Theta^T\right), \tag{22}$$

where $diag(\Theta)$ is an $N-1 \times N-1$ diagonal matrix with the ith diagonal entry equal to θ_i. For any vector $\mathbf{v} = [\upsilon_1, \ldots, \upsilon_{N-1}]^T$, we can show that

$$\begin{aligned}\mathbf{v}^T\bigtriangledown^2_{\mathbf{q}}\, g(\mathbf{q})\mathbf{v} &= \eta^{-2}\left(\eta\sum\nolimits_i \theta_i\upsilon_i^2 - \left(\sum\nolimits_i \theta_i\upsilon_i\right)^2\right)\\ &> \eta^{-2}\left(\sum\nolimits_i \theta_i \cdot \sum\nolimits_i \theta_i\upsilon_i^2 - \left(\sum\nolimits_i \theta_i\upsilon_i\right)^2\right)\\ &= \eta^{-2}\left(\sum\nolimits_i (\sqrt{\theta_i})^2 \cdot \sum\nolimits_i (\sqrt{\theta_i}\upsilon_i) - \left(\sum\nolimits_i \theta_i\upsilon_i\right)^2\right)\\ &\geq 0, \end{aligned} \tag{23}$$

where the last term follows from the Cauchy-Schwarz inequality. This means that $\bigtriangledown^2_{\mathbf{q}} g(\mathbf{q}) > 0$, therefore, $g(\mathbf{q})$ is convex with respect to $\mathbf{q}$.

ACKNOWLEDGMENT

The authors want to thank the associate editor and anonymous reviewers for helpful comments and suggestions.

REFERENCES

[1] " Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015," Cisco, San Jose, CA, USA, Feb. 2011. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/whitepaperc11-520862.html

[2] M. Dohler, R. Heath, A. Lozano, C. Papadias, and R. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.

[3] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[4] M. Pickavet *et al.*, "Worldwide energy needs for ICT: The rise of power-aware networking," in *Proc. IEEE ANTS*, 2008 pp. 1–3.

[5] Z. Niu, "Tango: Traffic-aware network planning and green operation," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 25–29, Oct. 2011.

[6] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.

[7] L. Correia *et al.*, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.

[8] S. Wang, M. Ge, and W. Zhao, "Energy-efficient resource allocation for OFDM-based cognitive radio networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3181–3191, Aug. 2013.

[9] D. Cao, S. Zhou, and Z. Niu, "Optimal combination of base station densities for energy-efficient two-tier heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4350–4362, Sep. 2013.

[10] J. Peng, P. Hong, and K. Xue, "Energy-aware cellular deployment strategy under coverage performance constraints," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 69–80, Jan. 2015.

[11] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy—And spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[12] R. Xie, F. Yu, and H. Ji, "Energy-efficient spectrum sharing and power allocation in cognitive radio femtocell networks," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1665–1673.

[13] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, Apr. 2011.

[14] Z. Han, Z. Ji, and K. Liu, "Non-cooperative resource competition game by virtual referee in multi-cell OFDMA networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1079–1090, Aug. 2007.

[15] A. Damnjanovic *et al.*, "A survey on 3gpp heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[16] S. Bu, F. Yu, and H. Yanikomeroglu, "Interference-aware energy-efficient resource allocation for heterogeneous networks with incomplete channel state information," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1036–1050, Mar. 2014.

[17] "3G home nodeb study item technical report," 3rd Generation Partnership Project, Sophia Antipolis Cedex, TR25.820, v8.2.0, Aug. 2008.

[18] R. Giuliano, C. Monti, and P. Loreti, "WiMAX fractional frequency reuse for rural environments," *IEEE Wireless Commun.*, vol. 15, no. 3, pp. 60–65, Jun. 2008.

[19] A. Gjendemsj, D. Gesbert, G. Oien, and S. Kiani, "Binary power control for sum rate maximization over multiple interfering links," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3164–3173, Aug. 2008.

[20] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2835–2848, Jul. 2009.

[21] K. Zheng, Y. Wang, C. Lin, X. Shen, and J. Wang, "Graph-based interference coordination scheme in orthogonal frequency-division multiplexing access femtocell networks," *IET Commun.*, vol. 5, no. 17, pp. 2533–2541, Nov. 2011.

[22] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 84–91, Feb. 2011.

[23] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[24] J. Bazerque and G. Giannakis, "Distributed scheduling and resource allocation for cognitive OFDMA radios," in *Proc. IEEE CROWNCOM*, 2007, pp. 343–350.

[25] I. Guvenc, "Capacity and fairness analysis of heterogeneous networks with range expansion and interference coordination," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1084–1087, Oct. 2011.

[26] K. Son, S. Lee, Y. Yi, and S. Chong, "REFIM: A practical interference management in heterogeneous wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1260–1272, Jun. 2011.

[27] T. Wang and L. Vandendorpe, "Iterative resource allocation for maximizing weighted sum min-rate in downlink cellular OFDMA systems," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 223–234, Jan. 2011.

[28] J. Papandriopoulos and J. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *Proc. IEEE ICC*, 2006, pp. 3270–3275.

[29] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Singapore: Athena Scientific, 2003.

[30] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[31] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE ISIT*, 2006, pp. 1394–1398.

[32] N. Mokari, M. Javan, and K. Navaie, "Cross-layer resource allocation in OFDMA systems for heterogeneous traffic with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 1011–1017, Feb. 2010.

[33] X. Zhou, G. Li, D. Li, D. Wang, and A. Soong, "Probabilistic resource allocation for opportunistic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2870–2879, Sep. 2010.

[34] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2230–2245, Apr. 2010.

[35] D. Palomar and J. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, 2005.

[36] X. Sun and S. Wang, "Interference management for energy saving in heterogeneous networks," in *Proc. IEEE WCNC*, 2013, pp. 1204–1208.

Xun Sun received the B.S. degree in communication engineering from Nanjing University, China, in 2012. He is currently pursuing the M.S. degree at the School of Electronic Science and Engineering, Nanjing University, China. His research interests include wireless communications and convex optimization. Currently, his research focuses on resource allocation in wireless networks.

Shaowei Wang (S'06–M'07–SM'13) received the B.S., M.S. and Ph.D. degrees in electronic engineering from Wuhan University, China, in 1997, 2003, and 2006, respectively. From 1997 to 2001, he was with China Telecom as an R & D Scientist. He has been with the School of Electronic Science and Engineering at Nanjing University from 2006, China, where he is currently a full Professor. From 2012 to 2013, he was also with Stanford University and The University of British Columbia as a Visiting Scholar/Professor. His research focuses on wireless communications and networking. In these areas he has published more than 60 papers in leading journals and conference proceedings. He organized the Special Issue on Enhancing Spectral Efficiency for LTE-Advanced and Beyond Cellular Networks for *IEEE Wireless Communications*, and the Feature Topic on Energy-Efficient Cognitive Radio Networks for *IEEE Communications Magazine*. He is on the editorial board of *IEEE Communications Magazine*, and serves/served on the technical or executive committee of reputable conferences including *IEEE INFOCOM, IEEE ICC, IEEE GLOBECOM, IEEE WCNC etc*.