Energy-Efficient Resource Allocation for Heterogeneous Services in OFDMA Downlink Networks: Systematic Perspective

Quansheng Xu, Xi Li, Hong Ji, *Senior Member, IEEE*, and Xiaojiang Du, *Senior Member, IEEE*

***Abstract*—In the area of energy-efficient (EE) resource allocation, only limited work has been done on consideration of both transmitter and receiver energy consumption. In this paper, we propose a novel EE resource-allocation scheme for orthogonal frequency-division multiple-access (OFDMA) networks, where both transmitter energy consumption and receiver energy consumption are considered. In addition, different quality-of-service (QoS) requirements, including minimum-rate guarantee service and best effort service, are taken into account. The time slot, subcarrier (frequency), and power-allocation policies are jointly considered to optimize system EE. With all these considerations, the EE resource-allocation problem is formulated as a mixed combinatorial and nonconvex optimization problem, which is extremely difficult to solve. To reduce the computational complexity, we tackle this problem in three steps. First, for a given power allocation, we obtain the time–frequency resource unit (RU) allocation policy via binary quantum-behaved particle swarm optimization (BQPSO) algorithm. Second, under the assumption of known RU allocation, we transform the original optimization problem into an equivalent concave optimization problem and obtain the optimal power-allocation policy through the Lagrange dual approach. Third, an iteration algorithm is developed to obtain the joint time–frequency power–resource-allocation strategy. We validate the convergence and effectiveness of the proposed scheme by extensive simulations.**

***Index Terms*—Energy efficiency (EE), heterogeneous service, mixed combinatorial and nonconvex optimization, orthogonal frequency-division multiple-access (OFDMA) network, resource allocation.**

I. Introduction

With the explosive growth of high-data-rate applications, incresingly more energy is consumed in wireless networks. Due to limited energy supply and the need for environment-friendly transmission behaviors [1]–[5], energy-efficient (EE) wireless communications is drawing increasing attention. Several international research projects dedicated to EE wireless communications are being carried out, such as Green Radio, EARTH, OPERA-Net, eWIN, etc. [1].

EE wireless communications includes many research areas, such as low-power circuit design, high-efficiency power amplifiers and digital signal processing technologies, EE resource management, EE network architecture and planning, adequate EE metric and energy consumption modeling, adaptive traffic pattern and load variation algorithms, and advanced cooling systems [1]–[3]. As an important aspect of EE resource management, EE resource allocation is very significant to enhance EE performance [6]–[8].

Several of EE resource-allocation algorithms have been proposed to maximize EE for different fading channels, such as frequency-selective fading channel, flat fading channel, etc. [9]–[11]. In addition, it has been shown that a unique global maximum EE exists and can be obtained by the proposed algorithms [8]–[11]. Moreover, some efficient resource-allocation algorithms have been proposed to optimize the tradeoff between spectrum efficiency (SE) and EE [12]–[14], bandwidth consumption and energy consumption [15], [16], and delay performance and energy consumption [17], [18]. However, almost all the aforementioned algorithms only optimize base-station (BS) energy consumption and do not consider user-equipment (UE) receiver circuit energy consumption that used to receive and process downlink traffic, which can significantly increase the UE receiver circuit energy consumption [19] and result in low EE. Therefore, the proposed algorithms are not EE from the system perspective.

Energy supply of UE is limited. Discontinuous reception (DRX) technology is always used to save the circuit energy of UE because circuit energy consumption increases with data transmission time [20], [21]. With the inspiration of DRX, some resource-allocation algorithms have been proposed [7], [22]. In orthogonal frequency-division multiple-access (OFDMA) systems, traffic to one UE can be scheduled into fewer time slots to reduce the energy consumption; thus, in [7], a green resource-allocation algorithm is proposed to minimize the total receiving energy consumption of UE. In [22], a DRX-aware scheduling method is proposed, where DRX parameters are used for scheduling, to reduce packet loss rate and UE energy consumption. However, these studies only optimize UE energy consumption, which imposes a strict restriction on resource allocation, causes services not able to use the most suitable resource, and may result in low EE.

Manuscript received August 19, 2013; revised January 10, 2014; accepted February 27, 2014. Date of publication March 18, 2014; date of current version June 12, 2014. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61271182, by the NSFC Young Scientists Research Fund under Grant 61302080, by the Beijing Higher Education Young Elite Teacher Project under Grant YETP0432, and by the U.S. National Science Foundation under Grant CNS-1065444. The review of this paper was coordinated by Q. Xu.

Q. Xu, X. Li, and H. Ji are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: quansheng@bupt.edu.cn; lixi@bupt.edu.cn; jihong@bupt.edu.cn).

X. Du is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: dxj@ieee.org).

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/TVT.2014.2312288

EE enhancement at the system level can be achieved only if energy consumption of the entire communication chains is considered [19]. However, if UE circuit energy consumption and BS transmission energy consumption are not comparable, it is unnecessary to consider jointly the transmitter and the receiver for designing EE resource-allocation algorithms. In traditional macrocellular scenarios, when transmission distance is large, circuit energy consumption of UE receivers is always much lower than the BS transmission energy consumption. However, in many short-range wireless communication systems [e.g., femtocells, wireless sensor networks, etc.], the circuit energy consumption of receiver becomes comparable to or even exceeds the transmission energy consumption [3], [20], [23]. Therefore, when designing EE resource-allocation algorithms for short-range communication situations, it is feasible and even necessary to consider jointly transmitter and receiver energy consumption.

So far, few works have jointly considered transmitter and receiver energy consumption when designing EE resource-allocation algorithms. In [19], a packet-scheduling algorithm is proposed, which can minimize both BS transmission and UE circuit energy consumption while meeting the service QoS requirement. However, this study has several weaknesses. First, with the objective of minimizing BS transmission energy and UE circuit energy consumption, it is not necessarily EE [8], [23]. Second, to decrease UE circuit energy consumption, at the beginning of each scheduling period, only a fraction of time is allowed to transmit data, which induces low resource utilization efficiency. Third, EE can be further improved, for it does not consider BS circuit energy consumption. Moreover, in [24] an end-to-end EE resource-allocation algorithm is proposed; however, only a heuristic method is used to solve the formulation problem.

In this paper, we investigate the performance optimization of EE for downlink communications in OFDMA networks from a systematic perspective, where the BS transmission, BS circuit, and UE circuit energy consumption are all taken into account. The resource-allocation problem is formulated as a mixed combinatorial and nonconvex optimization problem, where the time slot, subcarrier (frequency), and power-allocation policies are considered together to optimize EE. To reduce the computational complexity of the formulated problem, we tackle this problem in three steps. First, for a given power allocation, we obtain the time–frequency resource unit (RU) allocation policy via a binary quantum-behaved particle swarm optimization (BQPSO) algorithm [25]. Second, under the assumption of known RU allocation, we transform the original optimization problem to an equivalent concave optimization problem and obtain the optimal power-allocation policy through the Lagrange dual approach. Finally, based on the first and second steps, an iteration algorithm is developed to obtain the time–frequency power–resource-allocation strategy. We validate the convergence and effectiveness of the proposed scheme by extensive simulations. The distinct features of this paper are summarized as follows.

- Different from most existing works, we consider EE resource allocation from a systematic perspective. In the problem formulation, the BS transmission, BS circuit, and UE energy consumption are jointly considered, which can achieve better performance of EE.
- Heterogeneous services, including minimum-rate guarantee service and best effort service, are supported by our proposed resource-allocation scheme, which is realistic, for heterogeneous services may simultaneously request system resource.
- Since the time slot, subcarrier, and power–resource are jointly considered in our problem formulation, the proposed scheme can be regarded as a multidimensional resource-allocation scheme. The more resource dimensions we consider, the harder it is to solve the formulated problem. In fact, only few works have been done in multidimensional resource allocation.

TABLE I
SOME NOTATIONS USED IN THIS PAPER

N	number of sub-carriers
M	number of time slots in each scheduling period
K	number of users
K_1	number of users with minimum-rate guarantee service
W	bandwidth of each sub-carrier
T	duration of one time slot
Ω_A	set of users with minimum-rate guarantee service
Ω_B	set of users with best-effort service
R_k^{min}	minimum rate threshold for user k, $k \in \Omega_A$
η_k	proportional-fairness factor for user k, $k \in \Omega_B$
R_{tot}	total transmitted data during one scheduling period
P_{tot}	total energy consumption
P_{max}	transmission power budget of BS
R_k	transmitted data of user k
$R_{n,m,k}$	transmitted data of user k on the RU (n, m)
$\gamma_{n,m,k}$	SNR of unit transmission power, i.e., CNR
$a_{n,m,k}$	Boolean variable indicating the RU allocation
$p_{n,m,k}$	transmission power of user k on RU (n, m)
$\mathbf{A}$	RU allocation policy with the element $a_{m,n,k}$
$\mathbf{P}$	power allocation strategy with the element $p_{m,n,k}$
$\mathbb{A}_k$	set composed of RUs that are allocated to user k
P_c	circuit power of BS
P_{r_k}	circuit power of UE k at receiving mode
P_{nr_k}	circuit power of UE k at non-receiving mode
$\mathbf{D}$	particle position in BQPSO algorithm
$U(\cdot,\cdot)$	fitness function in BQPSO algorithm
$\mathbf{M}_{be}(\cdot)$	mean best position of all particles
$\mathbf{B}_i^{be}(\cdot)$	best position of the i-th particle
$\mathbf{G}_{be}(\cdot)$	global best position of all particles
$\mathbf{L}_i(\cdot)$	local attractor for particle i

The remainder of this paper is organized as follows. Section II gives the system model and problem formulation. In Section III, the *time–frequency RU allocation for a given power allocation* is discussed. In Section IV, the *power allocation for a given RU allocation* is presented. The *time–frequency power–resource allocation* is developed in Section V. The performance analysis and discussions are given in Section VI. Finally, we conclude this paper in Section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Here, we introduce the system model and formulate the problem of EE resource allocation. To make the rest of this paper easy to follow, we list some frequently used notations in Table I.

A. System Model

A single-cell OFDMA network with K users and N subcarriers is considered. Assume that these K users have heterogeneous service requirements and can be classified into two classes: users with minimum-rate guarantee services and users with best effort services [26]. The corresponding sets of these two user classes are denoted $\Omega_A = \{1, \ldots, K_1\}$ and $\Omega_B = \{K_1+1, \ldots, K\}$, respectively. Assume that each subcarrier has a bandwidth of W and can be modeled as Rayleigh block fading. We further assume that the channel state information can be estimated perfectly. A RU represents one subcarrier in one time slot with duration T, and one scheduling period contains M time slots. At each beginning of a scheduling period, the BS is responsible for allocating all the $N \times M$ RUs and power–resource among the K users.

B. Problem Formulation

The classical performance metric of EE "bits per joule" [4], i.e., the number of delivered bits per consuming unit energy, is adopted in this paper. This means that EE is defined as the amount of system-transmitted data R_{tot} divided by the total energy consumption P_{tot}.

The amount of system-transmitted data R_{tot} during one scheduling period is given as

$$R_{\mathrm{tot}} = \sum_{k=1}^{K} R_k \tag{1}$$

where R_k is the amount of transmitted data of user k during one scheduling period, which can be expressed as

$$R_k = \sum_{m=1}^{M} \sum_{n=1}^{N} TW \log_2(1 + a_{n,m,k}\gamma_{n,m,k}p_{n,m,k}) \tag{2}$$

where $a_{n,m,k}$ is the RU allocation indicator. $a_{n,m,k} = 1$ denotes that RU (n, m) is allocated to user k; otherwise, $a_{n,m,k} = 0$. $\gamma_{n,m,k} = |h_{n,m,k}|^2/N_0W$ is the signal-to-noise ratio of unit transmission power, i.e., the channel-gain-to-noise ratio (CNR). $h_{n,m,k}$ denotes the channel gain of user k on RU (n, m), and N_0 represents single-sided noise power spectral density. $p_{n,m,k} \geq 0$ denotes the transmission power of user k on RU (n, m).

The total energy consumption P_{tot} of transmitting R_{tot}-bit information can be calculated as follows. The total number of time slots, where there are data for user k, can be formulated as $M_{\mathrm{r}_k}(\mathbb{A}_k) = \sum_{m=1}^{M} f(\sum_{n=1}^{N} a_{n,m,k})$, where $\mathbb{A}_k$ is a set composed of RUs that are allocated to user k. If we know $\mathbf{A}$, we can obtain the set $\mathbb{A}_k$ easily. $f(x)$ is an integer step function, where $f(x) = 0$ when $x = 0$, and $f(x) = 1$ when $x \in \{1, \ldots, N\}$. To simplify the analysis, assume that there are only two work modes of UE in downlink transmission, i.e., receiving mode and nonreceiving mode. The circuit power of UE k at the receiving mode and that at the nonreceiving mode are P_{r_k} and P_{nr_k}, respectively. Assume that the circuit power of the BS is always P_c; thus, the total energy consumption P_{tot} can be given as

$$P_{\mathrm{tot}} = T\left[\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{n,m,k} + \sum_{k=1}^{K} P_{\mathrm{r}_k} M_{\mathrm{r}_k}(\mathbb{A}_k) + \sum_{k=1}^{K}(M - M_{\mathrm{r}_k}(\mathbb{A}_k))P_{\mathrm{nr}_k}) + P_c M\right]. \tag{3}$$

Then, the resource-allocation problem formulation from the systematic perspective can be given as

$$\begin{aligned}
&\max_{\mathbf{A},\,\mathbf{P}} \frac{R_{\mathrm{tot}}(\mathbf{A},\,\mathbf{P})}{P_{\mathrm{tot}}(\mathbf{A},\,\mathbf{P})} \\
\text{s.t.}\quad &\mathrm{C1}: \quad a_{n,m,k} \in \{0, 1\} \quad \forall n, m, k \\
&\mathrm{C2}: \quad \sum_{k=1}^{K} a_{n,m,k} \leq 1 \quad \forall n, m \\
&\mathrm{C3}: \quad \sum_{k=1}^{K}\sum_{n=1}^{N} p_{n,m,k} \leq P_{\max} \quad \forall m \\
&\mathrm{C4}: \quad R_k \geq R_k^{\min} \quad \forall k \in \Omega_A \\
&\mathrm{C5}: \quad \frac{R_k}{\sum_{k=K_1+1}^{K} R_k} = \eta_k \quad \forall k \in \Omega_B
\end{aligned} \tag{4}$$

where $\mathbf{A}$ with element $a_{m,n,k}$ and $\mathbf{P}$ with element $p_{m,n,k}$ are the RU allocation policy and the power allocation strategy, respectively. They are both $N \times M \times K$ matrices. Constraints C1 and C2 are RU allocation constraints. C2 means that one RU can be only assigned to one user at most. C3 is a power-allocation constraint, which gives the maximum transmission power of the BS, and $P_{\max}$ is the transmission power threshold. C4 is used to guarantee the minimum rate of user k in Ω_A, and $R_k^{\min}$ is the minimum rate threshold. C5 can ensure the fairness of user k in Ω_B, and η_k is the proportional fairness factor that is a predetermined value.

The optimal resource-allocation problem in (4) is a mixed combinatorial and nonconvex optimization problem. The combinatorial nature comes from the RU allocation constraints C1 and C2. The nonconvexity feature is caused by the proportional fairness constraint C5 and the fractional form of the objective function. Furthermore, the UE receiver energy consumption is considered and formulated as $T(\sum_{k=1}^{K} P_{\mathrm{r}_k} M_{\mathrm{r}_k}(\mathbb{A}_k) + \sum_{k=1}^{K}(M - M_{\mathrm{r}_k}(BBA_k))P_{\mathrm{nr}_k})$, which is nondifferential for arguments $a_{m,n,k}$. Therefore, the resource-allocation problem is very difficult to solve. In this paper, to solve the problem and obtain the resource-allocation policies, we develop the following three algorithms: *time–frequency RU allocation for a given power allocation*, *power allocation for a given RU allocation*, and *time–frequency power–resource allocation*.

III. Time–Frequency Resource Unit Allocation for a Given Power Allocation

Here, for a given power allocation, we present an RU allocation algorithm, which is based on BQPSO. BQPSO is a novel simulated evolution algorithm, which can effectively solve complicated combinatorial optimization problem with its desirable performance in finding a global optimal solution [25]. First, we introduce the BQPSO algorithm and then present the BQPSO-based RU allocation algorithm.

The BQPSO algorithm has three important parts, i.e., particle position, fitness function, and evolution equation. The position of each particle represents the possible solution of the optimization problem. In this paper, the position of each particle represents the possible RU allocation policy, i.e., decides how to assign $N \times M$ RUs to K users. Therefore, the $N \times M$ RUs are regarded as $N \times M$ decision variables, and each decision variable with $\lceil \log_2 K \rceil$ bits, where $\lceil \cdot \rceil$ means rounding up the value. The particle position is defined as

$$\mathbf{D} = \left(d_{1,1,1}, \ldots, d_{1,1,\lceil \log_2 K \rceil}, \ldots, d_{N,M,\lceil \log_2 K \rceil}\right) \tag{5}$$

which is a binary string with the length of $N \times M \times \lceil \log_2 K \rceil$. The bits $\mathbf{D}^1 = (d_{1,1,1}, \ldots, d_{1,1,\lceil \log_2 K \rceil})$ in $\mathbf{D}$ belong to the first decision variable, i.e., the RU (1, 1).

According to the position $\mathbf{D}$, we can get the RU allocation policy $\mathbf{A}$. For example, the $n \times M + m$th decision variable, i.e., the RU (n, M), should be allocated to user kk, i.e., $kk = d_{n,m,1} 2^{\lceil \log_2 K \rceil - 1} + d_{n,m,2} 2^{\lceil \log_2 K \rceil - 2} + \cdots + d_{n,m,\lceil \log_2 K \rceil} 2^0 + 1$. This means that $a_{n,m,k} = 0$ if $k \neq kk$ and $a_{n,m,k} = 1$ if $k = kk$.

The fitness function is used to evaluate the quality of the obtained solution, which is constructed by the original optimization problem. Assume that the power-allocation policy is given as $\mathbf{P}^t$. Then, using the method of penalty function, the fitness function is given as

$$U(\mathbf{A}, \mathbf{P}^t) = F(\mathbf{A}, \mathbf{P}^t) - \alpha F_p(\mathbf{A}, \mathbf{P}^t) \tag{6}$$

where $F(\mathbf{A}, \mathbf{P}^t) = R_{\text{tot}}(\mathbf{A}, \mathbf{P}^t)/P_{\text{tot}}(\mathbf{A}, \mathbf{P}^t)$ is the objective function, α is the penalty factor, and $F_p(\mathbf{A}, \mathbf{P}^t)$ represents the penalty function that consists of constraints related to $a_{m,n,k}$. The particle position in BQPSO is a binary string, and each RU is regarded as a decision variable allocated to one user at most; hence, C1 and C2 in (4) have been included. Then, the penalty function can be written as

$$F_p(\mathbf{A}, \mathbf{P}^t) = \sum_{k=1}^{K_1} \left[\max\left(0, R_k^{\min} - R_k\right)\right]^2 + \sum_{k=K_1+1}^{K} \left(\eta_k \sum_{k=K_1+1}^{K} R_k - R_k\right)^2 \tag{7}$$

where $\max(\cdot, \cdot)$ returns a larger value of the two variables.

Directly describing the evolution equation of BQPSO may be difficult to understand. Hence, first, the evolution equation of QPSO is introduced. Assume that there are I particles in search space. The evolution equation of particle $i (i = 1, \ldots, I)$ in the QPSO algorithm is given as follows [25], [27]:

$$\begin{cases} \mathbf{D}_i(l+1) = \mathbf{L}_i(l) + \upsilon \left|\mathbf{M}_{\text{be}}(l) - \mathbf{D}_i(l)\right| \ln\left(\frac{1}{u}\right), & \text{if } r \geq 0.5 \\ \mathbf{D}_i(l+1) = \mathbf{L}_i(l) - \upsilon \left|\mathbf{M}_{\text{be}}(l) - \mathbf{D}_i(l)\right| \ln\left(\frac{1}{u}\right), & \text{if } r < 0.5 \end{cases} \tag{8}$$

where l denotes the iteration time; υ is the contraction–expansion coefficient, which can be used to control the algorithm convergence rate; and u and r are both random variables between 0 and 1. $\mathbf{M}_{\text{be}}(l)$ is the mean best position of all particles in the lth iteration, which can be obtained by

$$\mathbf{M}_{\text{be}}(l) = \frac{1}{I} \sum_{i=1}^{I} \mathbf{B}_i^{\text{be}}(l) \tag{9}$$

where $\mathbf{B}_i^{\text{be}}(l)$ is the best position of the ith particle in the lth iteration. The $\mathbf{L}_i(l)$ in (8) is called the local attractor for particle i in the lth iteration, which can be given as

$$\mathbf{L}_i(l) = \theta \mathbf{B}_i^{\text{be}}(l) + (1-\theta) \mathbf{G}_{\text{be}}(l) \tag{10}$$

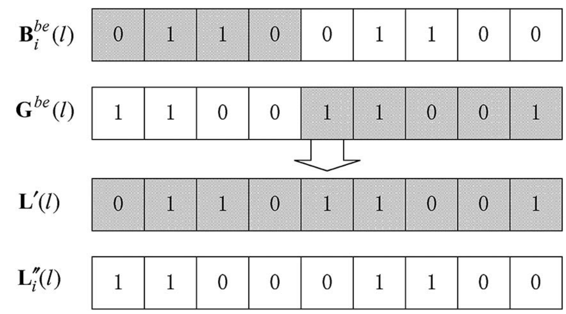


Fig. 1. $\mathbf{L}_i(l)$ producing process through single-point crossover.

where θ is a random variable between 0 and 1, and $\mathbf{G}_{\text{be}}(l)$ denotes the global best position of all particles in the lth iteration.

The particle location in BQPSO is a binary string; therefore, the evolution equation is different from that of QPSO. In BQPSO, the iterative equation (8) is replaced by the procedure of inversing the value of each bit in $\mathbf{L}_i(l)$ with a probability. All bits in the same decision variable have the same inverse probability. Specifically, the bits in $\mathbf{L}_i(l)$ that belong to the gth decision variable, i.e., $\mathbf{L}_i^g(l)$, are inversed with probability $p_i^g(l)$ to obtain $\mathbf{D}_i^g(l+1)$. $p_i^g(l)$ can be obtained as

$$b_i^g(l) = \upsilon \cdot d_H\left(\mathbf{M}_{\text{be}}^g(l), \mathbf{D}_i^g(l)\right) \cdot \ln(1/u) \tag{11}$$

$$p_i^g(l) = \begin{cases} b_i^g(l)/\lceil \log_2 K \rceil, & \text{if } b_i^g(l)/\lceil \log_2 K \rceil < 1 \\ 1, & \text{otherwise} \end{cases} \tag{12}$$

where $\mathbf{M}_{\text{be}}^g(l)$ and $\mathbf{D}_i^g(l)$ are the mean best position bits and the position bits belonging to the gth decision variable, respectively. $d_H(\cdot, \cdot)$ is a function that can obtain the Hamming distance of two input binary strings. In BQPSO, the jth bit of the $\mathbf{M}_{\text{be}}(l)$, i.e., $M_{\text{be}}^j(l)$, is determined by the states of the jth bit of all $\mathbf{B}_i^{\text{be}}(l)$. If more particles take on 1, the $M_{\text{be}}^j(l)$ will be 1; otherwise, it is 0.

The local attractor $\mathbf{L}_i(l)$ in BQPSO can be obtained after single-point crossover or multipoint crossover process. Fig. 1 shows how to obtain the local attractor from $\mathbf{B}_i^{\text{be}}(l)$ and $\mathbf{G}_{\text{be}}(l)$ through single-point crossover process. First, randomly select a number between 1 and $N \times M \times \lceil \log_2 K \rceil$ and regard it as the crossover point. Then, $\mathbf{L}_i'(l)$ and $\mathbf{L}_i''(l)$ are obtained from the offsprings of $\mathbf{B}_i^{\text{be}}(l)$ and $\mathbf{G}^{\text{be}}(l)$. Finally, $\mathbf{L}_i'(l)$ and $\mathbf{L}_i''(l)$ are selected randomly as the $\mathbf{L}_i(l)$.

Based on the BQPSO, the RU allocation algorithm is developed. The detailed steps are given in Algorithm 1.

Algorithm 1 RU Allocation for a Given Power Allocation

1: Initialization:

a) Set population size I, the maximum iteration times $L_{\text{iteration}}^{\text{BQPSO}}$, and iteration index $l = 1$.

b) Initialize the RU allocation policy $\mathbf{A}_i(1)$, and obtain $\mathbf{D}_i(1)$ according to the relationship between $\mathbf{A}_i(1)$ and $\mathbf{D}_i(1)$.

c) Set $\mathbf{B}_i^{\text{be}}(1) = \mathbf{D}_i(1)$, and according to the fitness function, choose a best position from $\mathbf{B}_i^{\text{be}}(1)$ as $\mathbf{G}_{\text{be}}(1)$;

for $l = 1, \ldots, L_{\text{iteration}}^{\text{BQPSO}}$ **do**
  Calculate $\mathbf{M}_{\text{be}}(l)$ and $\mathbf{L}_i(l)$ according to the related rules;
  for $i = 1, \ldots, I$ **do**
    Obtain $\mathbf{D}_i(l+1)$ according to the rules aforementioned;
    Get the updated RU allocation policy $\mathbf{A}_i(l+1)$ according to $\mathbf{D}_i(l+1)$;
    Get the individual best RU allocation policy $\mathbf{A}_i^{\text{ibe}}(l)$ according to $\mathbf{B}_i^{\text{be}}(l)$;
    if $U[\mathbf{A}_i(l+1), \mathbf{P}^t] > U[\mathbf{A}_i^{\text{ibe}}(l), \mathbf{P}^t]$, **then** the BS sets $\mathbf{B}_i^{\text{be}}(l+1) = \mathbf{D}_i(l+1)$; **else** it sets $\mathbf{B}_i^{\text{be}}(l+1) = \mathbf{B}_i^{\text{be}}(l)$; **endif**;
    Get the individual best RU allocation policy $\mathbf{A}_i^{\text{ibe}}(l+1)$ according to $\mathbf{B}_i^{\text{be}}(l+1)$;
    Get the global best RU allocation policy $\mathbf{A}^{\text{gbe}}(l)$ according to $\mathbf{G}_{\text{be}}(l)$;
    if $U[\mathbf{A}_i^{\text{ibe}}(l+1), \mathbf{P}^t] > U[\mathbf{A}^{\text{gbe}}(l), \mathbf{P}^t]$, **then** the BS sets $\mathbf{G}_{\text{be}}(l+1) = \mathbf{B}_i^{\text{be}}(l+1)$; **else** it sets $\mathbf{G}_{\text{be}}(l+1) = \mathbf{G}_{\text{be}}(l)$; **endif**;
    $i = i + 1$
  end for
  $l = l + 1$
end for
Obtain the RU allocation policy $\mathbf{A}^t$ according to $\mathbf{G}_{\text{be}}(l)$.

IV. Power Allocation for a Given Resource Unit Allocation

Here, under the assumption of known RU allocation, we transform the original nonconvex optimization problem to an equivalent concave optimization problem, and we obtain the optimal power-allocation policy by the Lagrange dual approach.

A. Problem Transformation

Assuming that the RU allocation policy $\mathbf{A}^t$ is known, then the BS only needs to do the power allocation for different users. Therefore, the resource-allocation problem in (4) can be reduced to

$$\max_{\mathbf{A}, \mathbf{P}} \frac{R_{\text{tot}}(\mathbf{A}, \mathbf{P})}{P_{\text{tot}}(\mathbf{A}, \mathbf{P})} \quad \text{s.t.} \quad \text{C3, C4, C5.} \tag{13}$$

Unfortunately, the optimization problem in (13) is still a nonconvex optimization problem due to C5 and the fractional form of the objective function. To develop an efficient resource-allocation algorithm, several transformations are needed to eliminate the nonconvexity and to make the problem more tractable. In the following, we first tackle C5 by changing the independent variable, and the original objective function is then transformed to an equivalent form, which is concave with respect to the new independent variable.

C5 makes the feasible set nonconvex. In general, to solve the problem efficiently, one needs to linearize C5. We introduce a new independent variable $R_{n,m,k} = WT \log 2(1 + \gamma_{n,m,k} p_{n,m,k})$ to the problem (13), which can decouple the proportional rate constraints. After introducing $R_{n,m,k}$, C5 can be rewritten as C5′, i.e.,

$$\sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} = \frac{\eta_k}{\eta_{K_1+1}} \sum_{(n,m)\in\mathbb{A}_{K_1+1}^t} R_{n,m,K_1+1}, \forall k \in \Omega_B. \tag{14}$$

Similarly, C3 and C4 can be rewritten as

$$\begin{aligned} \text{C3}': &\quad \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}} \le P_{\max} \quad \forall m \\ \text{C4}': &\quad \sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} \ge R_k^{\min} \quad \forall k \in \Omega_A. \end{aligned} \tag{15}$$

Since $R_{n,m,k}$ is a nonnegative variable, it is necessary to add a new constraint, i.e., C6: $R_{n,m,k} \ge 0$. Furthermore, it easy to verify that C3′ and C4′ are also concave functions or affine functions with respect to $R_{n,m,k}$. Therefore, the feasible set is a convex set after the transformation.

With a convex feasible set, the work in [28] and [29] show that the fractional program problem in (13) can be transformed to a easily solvable form. We define the maximum EE q^t of the considered system as

$$q^t = \frac{R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t)}{P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t)} = \max_{\mathbf{P}} \frac{R_{\text{tot}}(\mathbf{A}^t, \mathbf{P})}{P_{\text{tot}}(\mathbf{A}^t, \mathbf{P})}. \tag{16}$$

Then, we can use the following theorem, which had been proven in [28] and [29].

Theorem 1: The maximum EE q^t is achieved if and only if

$$\begin{aligned} &\max_{\mathbf{P}} R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) - q^t P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) \\ &\quad = R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t) - q^t P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t) = 0 \end{aligned} \tag{17}$$

for $R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) \le 0$ and $P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) > 0$.

Theorem 1: For an optimization problem with a fractional form objective function, there exists an equivalent objective function in subtractive form, e.g., $R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t) - q^t P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}^t)$. When the RU allocation policy is known, $p^{\text{const}} = \sum_{k=1}^{K} P_{\text{r}_k} M_{\text{r}_k}(\mathbb{A}_k) + \sum_{k=1}^{K}(M - M_{\text{r}_k}(BBA_k))P_{\text{nr}_k}) + P_c M$ is a constant. Then, the objective function can be transformed with the independent variable $R_{n,m,k}$ as

$$U^{\text{eff}}(\mathbf{R}) = \sum_{k=1}^{K} \sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} - qT \times \left[P^{\text{const}} + \sum_{k=1}^{K} \sum_{(n,m)\in\mathbb{A}_k^t} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}} \right]. \tag{18}$$

It is easy to verify that $U^{\text{eff}}(\mathbf{R})$ is a concave function with respect to $R_{n,m,k}$. As a result, the transformed problem, i.e.,

$$\max_{\mathbf{R}} \quad U(\mathbf{R})^{\text{eff}} \quad \text{s.t.} \quad \text{C3}', \text{C4}', \text{C5}', \text{C6} \tag{19}$$

is a concave optimization problem. Hence, we can first solve (19), and we then can use iterative algorithms, such as the Dinkelbach method [28], to solve (13).

In this paper, the Dinkelbach method is adopted to design the power-allocation algorithm, which is described in Algorithm 2. The proposed algorithm can converges to the optimal EE, which is proved in the Appendix. Furthermore, since $q_{l+1} = R_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l, R_{\rm be}^l)/P_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l, R_{\rm be}^l)$, the algorithm converges to the optimal EE with a superlinear convergence rate [9].

Algorithm 2 Power Allocation for Given RU Allocation

1: Initialization:
   a) Set the maximum iteration times $L_{\rm iteration}^{\rm Dinkelbach}$ and the maximum tolerance ϵ;
   b) Initialize the optimal EE $q_1 = 0$ and iteration index $l = 1$.
2: **for** $l = 1, \ldots, L_{\rm iteration}^{\rm Dinkelbach}$ **do**
3: For a given q_l, the BS solves the problem in (19) and obtains the resource-allocation policy $\{\mathbf{A}^t, \mathbf{P}_l\}$;
4: **if** $|R_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l) - q_l P_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l)| < \epsilon$, **then** the BS obtains the power-allocation policy $\mathbf{P}^t = \mathbf{P}_l$;
5: **else** BS sets $q_{l+1} = (R_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l)/P_{\rm tot}(\mathbf{A}^t, \mathbf{P}_l))$ and $l = l + 1$; **endif**;
6: **end for**
7: Output the optimal power-allocation policy $\mathbf{P}^t = \mathbf{P}_l$.

B. Power Allocation for Transformed Problem

The optimization problem in (19) is a concave optimization problem; thus, under some mild conditions, it can be shown that strong duality holds, and the duality gap is equal to zero [30]. In other words, solving the optimization problem in (19) is equivalent to solving the Lagrange dual problem. The Lagrange function of the transformed problem is given as

$$\begin{aligned} L(R_{m,n,k},\, \lambda_m,\, \beta_k,\, \xi_k) &= \sum_{k=1}^{K} \sum_{(n,m)\in \mathbb{A}_k^t} R_{n,m,k} - qT \\ &\quad \times \left[P^{\rm const} + \sum_{k=1}^{K} \sum_{(n,m)\in \mathbb{A}_k^t} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}} \right] \\ &\quad + \sum_{k=1}^{K_1} \beta_k \left(\sum_{(n,m)\in \mathbb{A}_k^t} R_{n,m,k} - R_k^{\min} \right) \\ &\quad + \sum_{m=1}^{M} \lambda_m \left(P_{\max} - \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}} \right) \\ &\quad + \sum_{k=K_1+2}^{K} \xi_k \left(\sum_{(n,m)\in \mathbb{A}_k^t} R_{n,m,k} \right. \\ &\quad \left. - \frac{\eta_k}{\eta_{K_1+1}} \sum_{(n,m)\in \mathbb{A}_{K_1+1}^t} R_{n,m,K_1+1} \right). \end{aligned} \tag{20}$$

where $\lambda_m \geq 0$ $(m = 1, \ldots, M)$, $\beta_k \geq 0$ $(k = 1, \ldots, K_1)$, and $\xi_k \geq 0$ $(k = K_1 + 2, \ldots, K)$ are the Lagrangian multipliers. When deriving the power-allocation policy, the boundary constraints $p_{n,m,k} \geq 0$ and $R_{n,m,k} \geq 0$ will be absorbed into the Karush–Kuhn–Tucker (KKT) conditions. Thus, the dual problem of (19) is as follows:

$$\min_{\lambda_m,\, \beta_k,\, \xi_k} \max_{R_{m,n,k}} L(R_{m,n,k},\, \lambda_m,\, \beta_k,\, \xi_k). \tag{21}$$

In the following, we solve the dual problem iteratively by decomposing it into two layers: layer 1 subproblem, i.e., allocating power for a fixed set of Lagrange multipliers, and layer 2 master problem, i.e., obtaining the Lagrange multipliers with the gradient method.

1) Solution for Layer 1: By dual decomposition, the BS first solves the following Layer 1 subproblem:

$$\max_{R_{m,n,k}} L(R_{m,n,k}, \lambda_m, \beta_k, \xi_k) \tag{22}$$

with a given parameter q and a fixed set of Lagrange multipliers $\{\lambda_m,\, \beta_k,\, \xi_k\}$. Using standard optimization techniques and the KKT conditions, the power-allocation policy $p_{n,m,k}$ is obtained as

$$\begin{cases} \left[\frac{(1+\beta_k)TW}{(qT+\lambda_m)\ln 2} - \frac{1}{\gamma_{n,m,k}} \right]^+ & \forall k \in \Omega_A \\ \left[\frac{(1-\sum_{K_1+2}^{K} \frac{\xi_k \eta_k}{\eta_{K_1+1}})TW}{(qT+\lambda_m)\ln 2} - \frac{1}{\gamma_{n,m,k}} \right]^+, & k = K_1 + 1 \\ \left[\frac{(1+\xi_k)TW}{(qT+\lambda_m)\ln 2} - \frac{1}{\gamma_{n,m,k}} \right]^+ & \forall k \in \Omega_B;\, k \neq K_1 + 1 \end{cases} \tag{23}$$

where $[x]^+ = \max\{0,\, x\}$. The power allocation has the form of multilevel water-filling. It can be observed that the EE variable $q \geq 0$ prevents energy inefficient transmission by truncating the water levels.

2) Solution for Layer 2: The dual function is differentiable with respect to optimization variables $R_{n,m,k}(p_{n,m,k})$. Therefore, using the solutions of the Layer 1 subproblems, the gradient method [29] can be used to solve the Layer 2 master problem, which leads to

$$\lambda_m^{l+1} = \left[\lambda_m^l - \nu_m^l \times \left(P_{\max} - \sum_{k=1}^{K} \sum_{n=1}^{N} p_{n,m,k} \right) \right]^+ \quad \forall m \tag{24}$$

$$\begin{aligned} \beta_k^{l+1} = \Bigg[\beta_k^l - \vartheta_k^l \times \Bigg(&\sum_{(n,m)\in \mathbb{A}_k^t} TW \log_2(1 + \gamma_{n,m,k} p_{n,m,k}) \\ &- R_k^{\min} \Bigg) \Bigg]^+ \quad \forall k \in \Omega_A \end{aligned} \tag{25}$$

$$\begin{aligned} \xi_k^{l+1} = \Bigg[\xi_k^l + \upsilon_k^l \times \Bigg(&R_{\rm be}\eta_k - \sum_{(n,m)\in \mathbb{A}_k^t} TW \log_2 \\ &\times (1 + \gamma_{n,m,k} p_{n,m,k}) \Bigg) \Bigg]^+ \\ &\forall k \in \Omega_B; \quad k \neq K_1 + 1 \end{aligned} \tag{26}$$

where index $l \geq 0$ is the iteration index, and ν_m^l, ϑ_k^l, and υ_k^l are the positive step sizes. $R_{\rm be}$ is the total transmission data of all the best effort service, i.e.,

$$R_{\rm be} = \sum_{k \in \Omega_B} \sum_{(n,m) \in \mathbb{A}_k} TW \log_2(1 + \gamma_{n,m,k} p_{n,m,k}). \quad (27)$$

Therefore, for each set of Lagrange multipliers $\{\lambda_m, \beta_k, \xi_k\}$, we can obtain the optimized power allocation $p_{n,m,k}$ and $R_{\rm be}$ from (23) and (27), respectively. After obtaining $p_{n,m,k}$ and $R_{\rm be}$, we can use (24)–(26) to update the Lagrange multipliers. The process is repeated until convergence is achieved. Since the transformed problem in (19) is concave in nature, if the chosen step sizes satisfy the general conditions stated in [30], then the iteration between Layers 1 and 2 will converge to the optimal solution of (19).

V. Time–Frequency Power–Resource Allocation

Based on the aforementioned works, initially, a joint time–frequency power–resource-allocation scheme is developed, and the complexity of the proposed scheme is then analyzed.

A. Time–Frequency Power–Resource Allocation

As discussed earlier, first, when the transmission power in each RU is known, the RU allocation policy can be obtained by using Algorithm 1. Second, based on the achieved RU allocation results, the optimal power allocation can be obtained by using Algorithm 2. In the third step, we substitute the power allocation results obtained in the second step into the first step and calculate the RU allocation again. The third step is shown in Algorithm 3. This iteration runs repeatedly until the results converge. The initial power allocation $\mathbf{P}^1$ is uniform among all RUs.

B. Analysis of Complexity and Feasibility

Here, we analyze the time complexity of the proposed time–frequency-power–resource-allocation scheme. First, we analyze the time complexity of the RU allocation, i.e., Algorithm 1. The complexity of the RU allocation is $\mathcal{O}(L_{\rm iteration}^{\rm BQPSO} \times I \times M \times N \times \lceil \log_2 K \rceil)$ [25]. Second, we analyze the time complexity of the power allocation, i.e., Algorithm 2. Since the original optimization problem has been transformed to a concave problem with respect to $R_{m,n,k}$, and dual decomposition is used to obtain the power-allocation policy. Therefore, similar to the analysis in [31], the complexity of the power allocation is $\mathcal{O}(L_{\rm iteration}^{\rm Dinkelbach} \times L_{\rm iteration}^{\rm Power} \times M \times N \times K)$, where $L_{\rm iteration}^{\rm Power}$ is the iteration time of the gradient method used to solve the Lagrange dual problem. Hence, the total complexity of the proposed resource-allocation scheme is $\mathcal{O}([L_{\rm iteration}^{\rm JTFPR} \times M \times N \times (L_{\rm iteration}^{\rm BQPSO} \times I \times \lceil \log_2 K \rceil + L_{\rm iteration}^{\rm Dinkelbach} \times L_{\rm iteration}^{\rm Power} \times K)])$. We find that the complexity of the proposed scheme is linear with respect to the number of time slots, users, subcarriers, and iteration times. Therefore, if the proposed scheme has a good convergence property, the time complexity of the proposed scheme is acceptable. In Section VI, the convergence performance of the proposed scheme is evaluated through the method of simulation, and we can find that it is acceptable.

Algorithm 3 Time–Frequency Power–Resource Allocation

1: Initialization:
  a) Each UE estimates the $h_{n,m,k}$ and sends $h_{n,m,k}$, energy consumption $P_{{\rm r}_k}$, and $P_{{\rm nr}_k}$ to the BS;
  b) Set the maximum iteration times $T_{\rm iteration}^{\rm JTFPR}$ and the maximum tolerance ε;
  c) The BS initializes the time–frequency RU allocation policy $\mathbf{A}^1$, the power-allocation policy $\mathbf{P}^1$, and iteration index $t = 1$.
2: **for** $t = 1, \ldots, T_{\rm iteration}^{\rm JTFPR}$ **do**
3:   For a given power-allocation policy $\mathbf{P}^t$, the BS obtains time–frequency RU allocation policy $\mathbf{A}^{t+1}$ via **Algorithm 1**;
4:   After getting the RU allocation policy $\mathbf{A}^{t+1}$, the BS calculates the power-allocation policy $\mathbf{P}^{t+1}$ by **Algorithm 2**;
5:   **if** $|p_{n,m,k}^{t+1} - p_{n,m,k}^t| \leq \varepsilon, \forall n, m, k,$ **then** the BS obtains the resource-allocation policy $\{\mathbf{A}^*, \mathbf{P}^*\} = \{\mathbf{A}^t, \mathbf{Pq}^t\}$;
6:   **else** $t = t + 1$; **endif**
7: **end for**
8: Output the resource-allocation policy $\{\mathbf{A}^*, \mathbf{P}^*\} = \{\mathbf{A}^t, \mathbf{P}^t\}$.

VI. Performance Evaluation and Discussions

The simulation parameters are set as follows. The total bandwidth, i.e., 1.08 MHz, is equally divided into $N = 72$ orthogonal subcarriers. The scheduling period includes $M = 10$ time slots and each time slot with a duration of $T = 0.5$ ms. Assume that there are $K = 5$ UE devices unless noted otherwise. UE 1 and UE 2 are users with the minimum-rate requirement of 500 and 750 kb/s, respectively. UE 3, UE 4, and UE 5 are users with best effort service. The channel of the kth UE is modeled as Rayleigh fading with an average CNR of $\overline{\gamma}_k$. In our simulation results, the average CNR in the horizontal axis represents the CNR of the lowest CNR UE. Unless specifically noted, $I = 3000$, $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, $\eta_3 = \eta_4 = \eta_5 = 1/3$, $P_{\max} = 40$ dBm, $P_c = 36.99$ dBm, $P_{{\rm r}_k} = [31.14, 31.46, 30.79, 31.14, 31.46]$ dBm, and $P_{{\rm nr}_k} = [20, 23.01, 20, 23.01, 23.01]$.

A. Convergence of the Proposed Resource-Allocation Scheme

Fig. 2 shows the convergence of the proposed RU allocation algorithm for given power allocation. The given power-allocation policy are $\mathbf{P}^1$ and $\mathbf{P}^2$. The results in Fig. 2 were averaged over 500 adaptation processes. It can be seen that no matter which power-allocation policy and channel condition are given, the proposed RU allocation algorithm always converges to 90% of the upper bound performance within $L_{\rm iteration}^{\rm BQPSO} = 700$.

Fig. 3 shows the convergence of the gradient method used to solve the Lagrange dual problem in power allocation. If

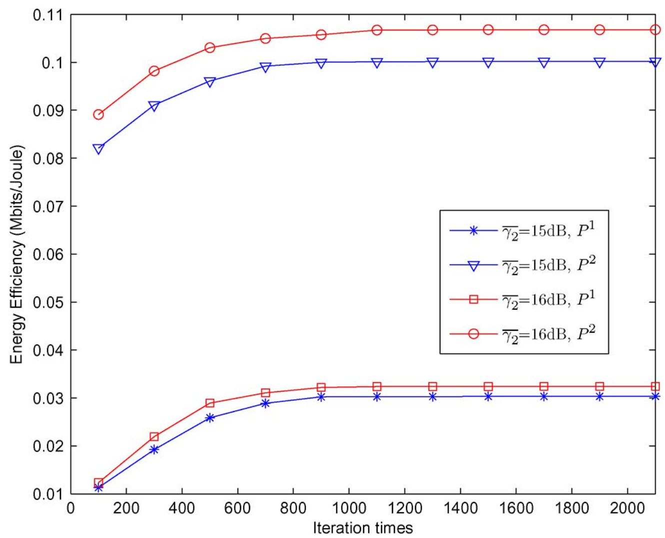


Fig. 2. Convergence of Algorithm 1.

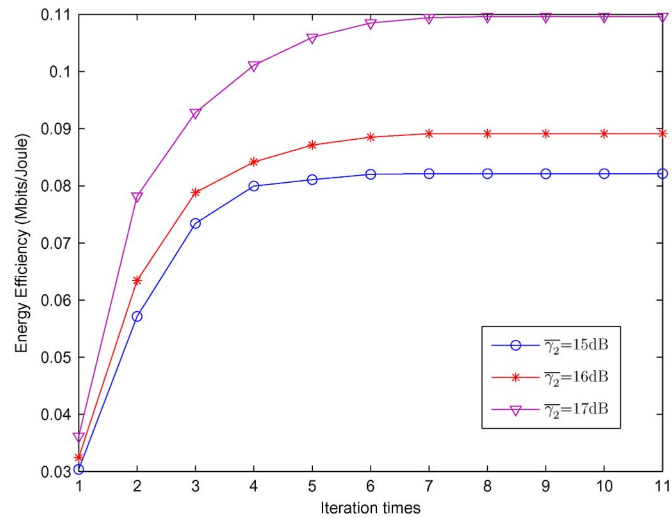


Fig. 4. Convergence of Algorithm 2.

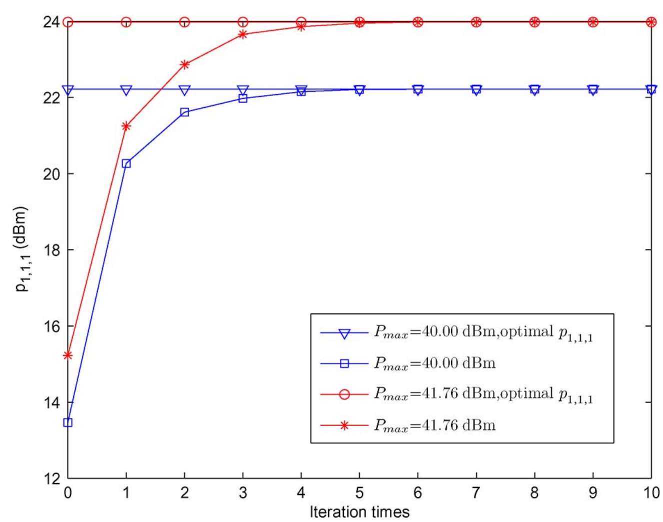


Fig. 3. Convergence of the gradient method.

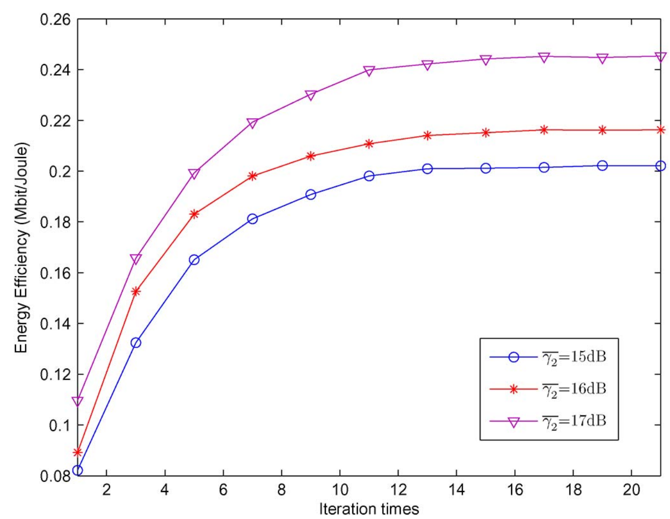


Fig. 5. Convergence of Algorithm 3.

$\overline{\gamma}_2 = 15$ dB, the RU allocation policy is $\mathbf{A}^1$ with $L_{\text{iteration}}^{\text{BQPSO}} = 1000$ and $q = 0.1$ Mbit/J. $L_{\text{iteration}}^{\text{BQPSO}} = 1000$ can ensure the RU allocation algorithm convergence. It can be seen that the gradient method has fast convergence rate; it converges to 90% of the upper bound performance within five iterations.

Fig. 4 shows the convergence of the proposed power-allocation algorithm for given RU allocation, i.e., the Dinkelbach method. The RU allocation policy is $\mathbf{A}^1$ with $L_{\text{iteration}}^{\text{BQPSO}} = 1000$ and $L_{\text{iteration}}^{\text{Power}} = 10$. It can be seen that the Dinkelbach method has a fast convergence rate; it converges to 90% of the upper bound performance within six iterations.

Fig. 5 shows the convergence of the proposed joint time–frequency power–resource-allocation scheme. $L_{\text{iteration}}^{\text{BQPSO}} = 1000$, $L_{\text{iteration}}^{\text{Power}} = 10$, and $L_{\text{iteration}}^{\text{Dinkelbach}} = 10$. Similarly, it can be seen that the proposed scheme has a satisfactory convergence rate; it converges to 90% of the upper bound performance within 11 iterations.

In Figs. 2–5, we can find that the proposed resource-allocation scheme has good convergence performance.

B. Performance Comparison of Different Resource-Allocation Schemes

To show that the proposed scheme is necessary for some scenarios, the simulation is performed in different scenarios. The communication scenarios in practical can be roughly classified into the following three scenarios. In Scenario 1, the UE circuit power is little compared with BS transmission power, i.e., $P_{\max} = 43.01$ dBm, $P_c = 40$ dBm, $P_{\mathrm{r}_k} = [28.45, 29.03, 28.75, 29.03, 28.45]$ dBm, and $P_{\mathrm{nr}_k} = [20, 23.01, 20, 23.01, 20]$ dBm. In Scenario 2, the UE circuit power is comparable to the BS transmission power, i.e., $P_{\max} = 40$ dBm, $P_c = 36.99$ dBm, $P_{\mathrm{r}_k} = [31.14, 31.46, 30.79, 31.14, 31.46]$ dBm, and $P_{\mathrm{nr}_k} = [20, 23.01, 20, 23.01, 23.01]$ dBm. In Scenario 3, the UE circuit power plays an important role in the total energy consumption, i.e., $P_{\max} = 36.99$ dBm, $P_c = 33.01$ dBm, $P_{\mathrm{r}_k} = [31.14, 31.76, 30, 30.79, 31.46]$ dBm, and $P_{\mathrm{nr}_k} = [20, 23.01, 20, 24.77, 20]$ dBm. Furthermore, to evaluate the proposed resource-allocation scheme, we compare it with three traditional resource-allocation schemes. Comparison Scheme 1 only considers the BS energy consumption, as in [9] and [11]. Comparison

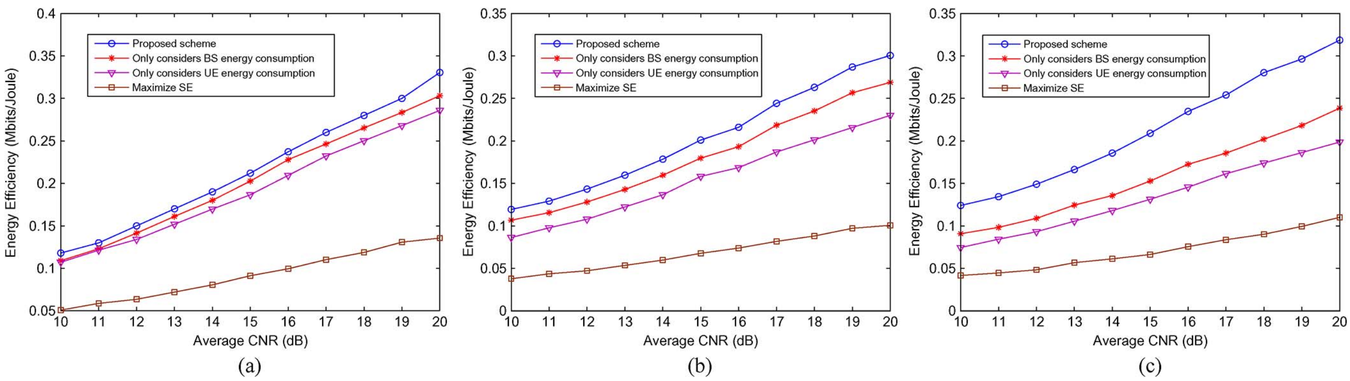


Fig. 6. EE of different resource-allocation schemes. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3.

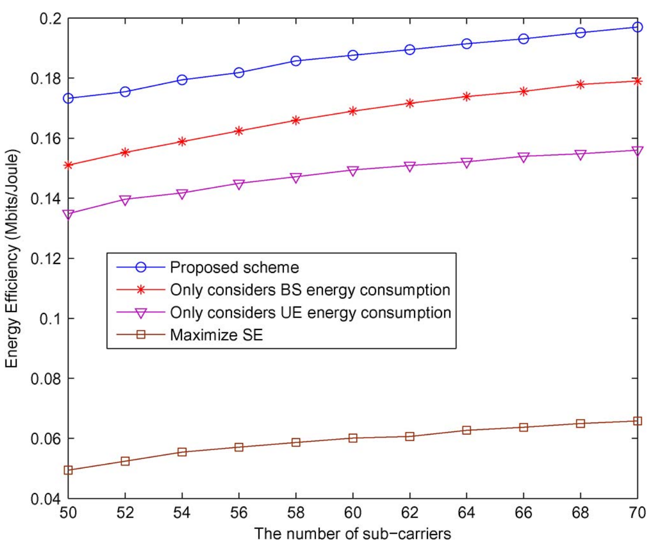


Fig. 7. EE of different resource-allocation schemes versus the number of subcarriers ($\overline{\gamma}_2 = 15$ dB).

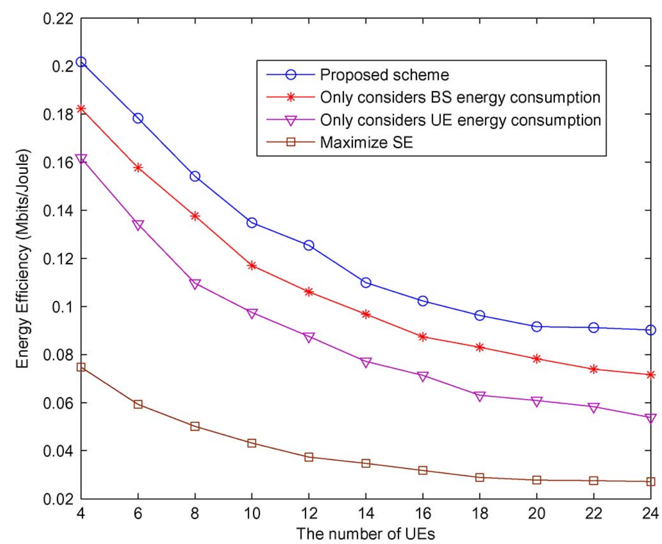


Fig. 8. EE of different resource-allocation schemes versus the number of UE devices.

Scheme 2 only considers the UE energy consumption, as in [7]. Comparison Scheme 3 maximizes the system transmission data rate, i.e., maximize the SE.

Fig. 6 shows the EE of different resource-allocation schemes under the aforementioned three scenarios. Fig. 6 shows that the proposed scheme can achieve the best EE performance in all the scenarios. Furthermore, the results show that, when the proportion of UE circuit power to the total power becomes larger, i.e., from Scenario 1 to Scenario 3, the advantage of the proposed scheme increases. This is because all energy consumption during the communication process is considered when designing the proposed scheme, which is different from the comparing schemes. Therefore, better EE performance is achieved. In the current and future communication systems, increasingly more communications will happen in short-range situations, and circuit energy consumption in the receiver will play an important role in the total energy consumption. Therefore, this paper is meaningful.

Fig. 7 shows the EE of different resource-allocation schemes versus the number of subcarriers. We find that the EE rises up progressively as the number of subcarriers increases gradually. This is because as the number of subcarriers increases, more bandwidth resources are available, and better EE can be obtained, which is a classical conclusion obtained in [15] and [16]. In addition, similar to that in Fig. 6, we can also find that the proposed scheme can achieve the best performance of EE among all the resource-allocation schemes.

Fig. 8 shows the EE of different resource-allocation schemes versus the number of UE devices. To simplify the presentation, we assume $K_1 = \lfloor 0.4 * K \rfloor$, where $\lfloor \cdot \rfloor$ means rounding down the value. $R_k^{\min} = 1.25/K_1$ Mb/s, $\forall\, k \in \Omega_A$, and $\eta_k = 1/K_2$, $\forall\, k \in \Omega_B$. The channel conditions of different UE devices are independent, and all UE devices have the same average CNR of 15 dB. The circuit power of different UE devices is also set as the same, i.e., $P_{\mathrm{r}_k} = 31.14$ dBm and $P_{\mathrm{nr}_k} = 20$ dBm ($\forall\, k \in [1, \ldots, K]$). Furthermore, the BS maximum transmission power and circuit power are set as $P_{\max} = 40$ dBm and $P_c = 36.99$ dBm, respectively. In Fig. 8, we can obtain the conclusion that the EE decreases gradually as the number of UEs increases. Since the UE circuit power consumption is considered, the greater the number of UE devices, the more circuit power will be consumed, which results in low EE. The conclusion is different from existing results. When the receiver circuit power is not considered, the nominal EE will rise gradually as the number of UE devices increases, for the multiusers diversity gain. Furthermore, Fig. 8 also proves that the proposed scheme can achieve the best performance of EE.

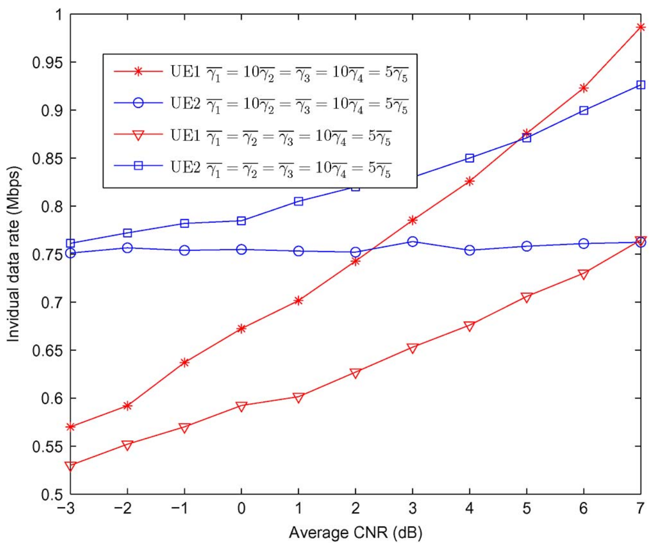


Fig. 9. Satisfying the transmission rate requirement of minimum-rate guarantee services.

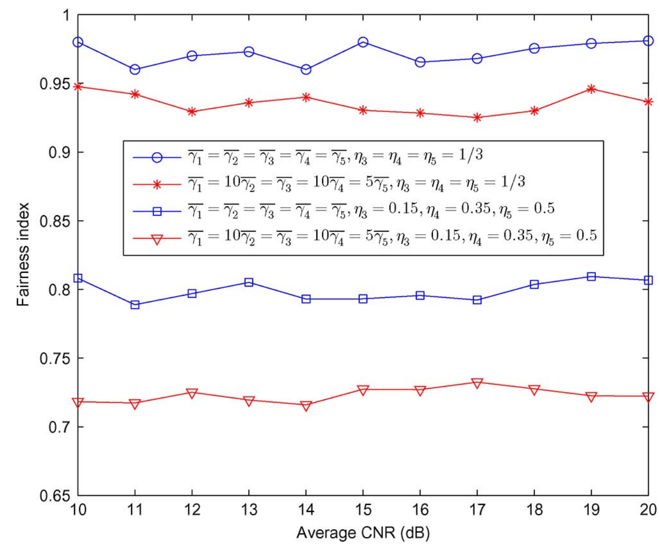


Fig. 10. Guaranteeing the fairness of best effort services.

C. Capability of the Proposed Resource-Allocation Scheme for Guaranteeing Heterogeneous Service QoS

Here, we discuss the performance of the proposed resource-allocation scheme for guaranteeing heterogeneous QoS requirements. Fig. 9 shows the capability of the proposed scheme for satisfying the requirement of minimum-rate guarantee service under different channel conditions. In Fig. 9, we can see that the proposed scheme can guarantee users' minimum-rate requirements under all channel conditions. Furthermore, under the condition of $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, we find that the rate of low-CNR UE (UE2) is almost fixed at the minimum-rate requirement, i.e., 750 kb/s, whereas the rate of high-CNR UE (UE1) increases with the CNR. This is because, in the proposed scheme, more resource is allocated to the UE with a good channel condition to achieve better EE.

For the best effort services, the fairness can be evaluated in term of fairness index [32], which is defined as

$$\phi = \frac{\left(\sum_{k=K_1+1}^{K} R_k\right)^2}{(K-K_1)\sum_{k=K_1+1}^{K} R_k^2}. \tag{28}$$

where ϕ is in the range of [0, 1], and if the value of ϕ is closer to 1, then a better fairness performance will be achieved.

Fig. 10 shows the capability of the proposed scheme for achieving the fairness of best effort services under different situations. We see that, no matter what the channel conditions are, the fairness performance of a situation with $\eta_3 = \eta_4 = \eta_5 = 1/3$ is always better than a situation with $\eta_3 = 0.15$, $\eta_4 = 0.35$, and $\eta_5 = 0.5$. Furthermore, we also find that, although the channel conditions of different UE devices are very different in the case of $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, if we set reasonable $\eta_k(\eta_3 = \eta_4 = \eta_5 = 1/3)$, the proposed scheme can still obtain satisfactory fairness performance. Therefore, we have the following conclusion: The proportional fairness factor η_k setting has a big impact on guaranteeing service fairness, and we can adjust the proportional-fairness factor to achieve the desirable performance of guaranteeing service fairness.

VII. Conclusion

In this paper, we have studied the problem of EE resource allocation for downlink communications in OFDMA networks that support heterogeneous services. Both the transmitter energy consumption and receiver energy consumption are considered. We formulated the problem of EE resource allocation as a mixed combinatorial and nonconvex optimization problem. To reduce the computational complexity, we solved the problem in three steps, where techniques such as BQPSO and some mathematical processes have been used. We run simulations to evaluate the performance of the proposed scheme. Our simulation results show the effectiveness of the proposed scheme. In the future work, we will consider the issues of how to design EE resource-allocation scheme with much lower computational complexity while maintaining the desirable system performance and how to evaluate the performance of the proposed scheme with realistic energy consumption models.

Appendix
Proof of Algorithm 2 Convergence

A similar approach in [28] is adopted to prove the convergence of the iterative algorithm, i.e., Algorithm 2. First, two propositions are introduced and then the convergence is demonstrated. To simplify the notations, the equivalent objective function in (13) is defined as $F_e(q') = \max_{\mathbf{P}}\{R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) - q'P_{\text{tot}}(\mathbf{A}^t, \mathbf{P})\}$.

Proposition 1: $F_e(q')$ is a nonnegative function in the domain of definition.

Proof: Assuming that $\{\mathbf{A}^t, \mathbf{P}'\}$ is an arbitrary solution for the problem and $q' = R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}')/P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}')$, then

$$\begin{aligned} F_e(q') &= \max_{\mathbf{P}}\left\{R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) - q'P_{\text{tot}}(\mathbf{A}^t, \mathbf{P})\right\} \\ &\geq R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') - q'P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') = 0. \end{aligned} \tag{29}$$

Proposition 2: $F_e(q')$ is a strictly monotonic decreasing function with respect to q', i.e., $F_e(q'') > F_e(q')$ as long as $q' > q''$.

Proof: Assuming that $\{\mathbf{A}^t, \mathbf{P}'\}$ and $\{\mathbf{A}^t, \mathbf{P}''\}$ are two different optimal policies for $F_e(q')$ and $F_e(q'')$, respectively, then

$$\begin{aligned} F_e(q'') &= \max_{\mathbf{P}} \left\{ R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) - q'' P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}) \right\} \\ &= R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}'') - q'' P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}'') \\ &> R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') - q'' P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') \\ &\geq R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') - q' P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}') = F(q'). \end{aligned} \quad (30)$$

Therefore, the convergence of Algorithm 2 can be proven as follows. First, we can prove that q increases in each iteration. Second, we demonstrate that, if iteration time is large enough, q will converge to the optimal solution q^t such that it meets the optimality condition of Theorem 1.

Assume that $\{\mathbf{A}^t, \mathbf{P}_l\}$ is the optimal policy in the lth iteration and that $q_l (\neq q^t)$ and $q_{l+1} (\neq q^t)$ represent the EE in iterations l and $l+1$, respectively. According to Theorem 1 and Proposition 1, $F_e(q_l) > 0$ and $F_e(q_{l+1}) > 0$ must be true. Moreover, Since we calculate q_{l+1} as $q_{l+1} = R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}_l) / P_{\text{tot}}(\mathbf{A}^t, \mathbf{P}_l)$, $F_e(q_l)$ can be expressed as

$$\begin{aligned} F_e(q_l) &= R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}_l) - q_l R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}_l) \\ &= R_{\text{tot}}(\mathbf{A}^t . \mathbf{P}_l)(q_{l+1} - q_l) > 0. \end{aligned} \quad (31)$$

Since $R_{\text{tot}}(\mathbf{A}^t, \mathbf{P}_l) > 0$, then $q_{l+1} > q_l$.

Therefore, according to Propositions 1 and 2, and $q_{l+1} > q_l$, as long as the iteration time is large enough, $F_e(q_l)$ will eventually approach to zero and satisfy the optimality condition of Theorem 1, i.e., $F_e(q^t) = 0$.

References

[1] D. Feng, C. Jiang, G. Lim, J. Cimini, L. J., G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Survey Tuts.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.

[2] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Survey Tuts.*, vol. 13, no. 4, pp. 524–540, Nov. 2011.

[3] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communication tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[4] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.

[5] S. Chen, Y. Wang, W. Ma, and J. Chen, "Technical innovations promoting standard evolution: From TD-SCDMA to TD-LTE and beyond," *IEEE Wireless Commun.*, vol. 19, no. 1, pp. 60–66, Feb. 2012.

[6] S. Pollin, R. Mangharam, B. Bougard, L. V. der Perre, R. R. I. Moerman, and F. Catthoor, "MEERA: Cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 98–109, Jan. 2008.

[7] F.-S. Chu, K.-C. Chen, and G. Fettweis, "Green resource allocation to minimize receiving energy in OFDMA cellular systems," *IEEE Commun. Lett.*, vol. 16, no. 3, pp. 372–374, Mar. 2012.

[8] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–5.

[9] C. Isheden, Z. Chong, E. A. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[10] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–6.

[11] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[12] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 3874–3886, Nov. 2011.

[13] W. Cheng, X. Zhang, and H. Zhang, "Joint spectrum and power efficiencies optimization for statistical QoS provisionings over SISO/MIMO wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 903–915, May 2013.

[14] L. Zhou, R. Q. Hu, Y. Qian, and H.-H. Chen, "Energy-spectrum efficiency tradeoff for video streaming over mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 981–991, May 2013.

[15] B. Changhun and W. E. Stark, "End-to-end energy-bandwidth tradeoff in multihop wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4051–4066, Sep. 2009.

[16] C. L. Chen, W. E. Stark, and S. G. Chen, "Energy-bandwidth efficiency tradeoff in MIMO multi-hop wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1537–1546, Sep. 2011.

[17] C. Xiong, G. Y. Li, Y. Liu, and S. Xu, "Qos driven energy-efficient design for downlink OFDMA networks," in *Proc. IEEE GLOBECOM*, Dec. 2012, pp. 4320–4325.

[18] F. Meshkati, H. V. Poor, S. C. Schwartz, and R. V. Balan, "Energy-efficient resource allocation in wireless networks with quality-of-service constraints," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3406–3414, Nov. 2009.

[19] R. Gupta and E. C. Strinati, "Green scheduling to minimize base station transmit power and UE circuit power consumption," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2011, pp. 2424–2429.

[20] S. G. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[21] I. Stanojev, O. Simeone, Y. Bar-Ness, and H. K. Dong, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326–335, Jan. 2009.

[22] B. Huang, H. Tian, L. Chen, and J. Zhu, "DRX-aware scheduling method for delay-sensitive traffic," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1113–1115, Dec. 2010.

[23] G. Miao, N. Himayatyy, and G. Y. Li, "Cross-layer optimization for energy-efficient wireless communications: a survey," *Wireless Commun. Mobile Comput.*, vol. 9, no. 4, pp. 529–542, Apr. 2009.

[24] Q. Xu, X. Li, and H. Ji, "Multiple resource allocation in OFDMA downlink networks: End-to-end energy-efficient approach," in *Proc. IEEE ICC 2014*, Sydney, Australia, Jun. 2014.

[25] M. L. Xi, J. Sun, and Y. Wu, "Quantum-behaved particle swarm optimization with binary encoding," *Control Decision*, vol. 25, no. 1, pp. 99–104, Jan. 2010.

[26] R. Xie, F. R. Yu, and H. Ji, "Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 770–780, Feb. 2012.

[27] Y. S. Zhao, X. Li, Y. Li, and H. Ji, "Resource allocation for high-speed railway downlink MIMO-OFDM system using quantum-behaved particle swarm optimization," in *Proc. IEEE ICC*, Jun. 2013, pp. 2343–2347.

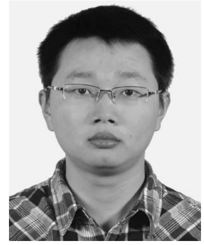
[28] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

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

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[31] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.

[32] E. L. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, pp. 1024–1039, Sep. 1991.



Quansheng Xu received the B.S. degree from Hangzhou Dianzi University, Hangzhou, China, in 2010. He is currently working toward the Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include green communications, high-speed railway communications, and heterogeneous networks.

Mr. Xu has served as a Reviewer for several journals and conference proceedings, including *Wiley Wireless Communications and Mobile Computing*, the *Wiley International Journal of Communication Systems and Networking*, *China Communications*, and the 2014 IEEE International Conference on Communications.

Xi Li received the B.E. and Ph.D. degrees in communication and information systems from Beijing University of Posts and Telecommunications, Beijing, China, in 2005 and 2010, respectively.

She is currently an Associate Professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. Her current research interests include broadband wireless communications and heterogeneous networks.

Dr. Li served as the Chair of the Special Track on Cognitive Test Beds for the 2011 International Conference on Communications and Networking in China and as a member of the Technical Program Committee Track on Medium Access Control (MAC) Layer of the 2012 and 2013 IEEE Wireless Communications and Networking Conference and of the TPC Track on MAC and Cross Layer Design at the 2012 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.

Hong Ji (SM'09) received the B.S. degree in communications engineering and the M.S. and Ph.D. degrees in information and communications engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1989, 1992, and 2002, respectively.

From June to December 2006, she was a Visiting Scholar with The University of British Columbia, Vancouver, BC, Canada. She is currently a Professor with the School of Information and Communication Engineering, BUPT. She is also engaged with national science research projects, including the High-Technology Research and Development Program of China (863 program) and the National Natural Science Foundation of China. Her research interests include heterogeneous networks, peer-to-peer protocols, cognitive radio networks, relay networks, Long-Term Evolution/fifth generation, and cooperative communications.

Dr. Ji has served on the technical program committees (TPCs) of numerous conferences, such as the 2010–2013 IEEE International Conference on Communications, the 2010–2013 IEEE Global Communications Conference, the 2012 and 2014 IEEE Wireless Communications and Networking Conference, the 2010 and 2011 IEEE International Conference on Computer Communications workshops, and the 2012 IEEE International Conference on Computer, Information, and Telecommunication Systems. She has also served as a TPC Cochair for the 2011 International Conference on Communications and Networking in China workshops and as a Program Cochair for the 2012–2014 International Conference on Wireless Information Networks and Systems. She serves on the editorial boards of several journals, including the *Wiley International Journal of Wireless Communications and Mobile Computing* and the *Wiley International Journal of Communication Systems*.

Xiaojiang (James) Du (SM'03) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 2002 and 2003, respectively, all in electrical engineering.

From August 2004 and July 2009, he was an Assistant Professor with the Department of Computer Science, North Dakota State University, Fargo, ND, USA. He is currently an Associate Professor with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA. He is the author of over 120 journal and conference papers in these areas. His research interests include security, cloud computing, wireless networks, and computer networks and systems.

Dr. Du is a Life Member of the Association for Computing Machinery (ACM). He was the Chair of the Computer and Network Security Symposium of the IEEE/ACM International Wireless Communication and Mobile Computing Conference from 2006 to 2010. He was a member of the technical program committees of several premier ACM and IEEE conferences, such as the IEEE International Conference on Computer Communications (2007–2015), the IEEE/International Federation for Information Processing (IFIP) International Symposium on Integrated Network Management, the IEEE/IFIP Network Operations and Management Symposium, the IEEE International Conference on Communications (IEEE ICC), the IEEE Global Communications Conference, the IEEE Wireless Communications and Networking Conference (IEEE WCNC), BroadNet, and the IEEE International Performance Computing and Communications Conference. He will serve as the Lead Chair for the Communication and Information Security Symposium of the 2015 IEEE ICC and as the Cochair of the Mobile and Wireless Networks Track for the 2015 IEEE WCNC. He serves on the editorial boards of four international journals. He received the Excellence in Research Award from North Dakota State University in May 2009. He has also received research grants from the U.S. National Science Foundation, the U.S. Army Research Office, the U.S. Air Force Research Laboratory, the National Aeronautics and Space Administration, the Commonwealth of Pennsylvania, and Amazon.