Energy-Adaptive Downlink Resource Allocation in Wireless Cellular Systems

Ya-Ju Yu, *Member, IEEE,* Ai-Chun Pang, *Senior Member, IEEE,* Pi-Cheng Hsiu, *Member, IEEE,* and Yuguang Fang, *Fellow, IEEE*

Abstract—Mobile devices have increasingly been used to run multimedia applications which are extremely downlink-intensive. The conventional *rate adaptive* and/or *margin adaptive* approach for radio resource allocation may result in unnecessary energy consumption on mobile devices, which will not be energy efficient for mobile multimedia applications. In this paper, we develop an *energy adaptive* approach and design an energy-efficient downlink resource allocation scheme to support multimedia applications. The objective is to minimize the total energy consumption of mobile devices for data reception while meeting the data rate requirements at mobile devices and the transmit power constraint at the base station. We show that the optimization problem is $\mathcal{NP}$-hard and then propose an efficient algorithm that has a provable performance guarantee under a certain condition. We have conducted extensive simulations to evaluate the efficacy of the proposed algorithm and our results provide useful insights into the design of energy-efficient resource allocation for wireless systems.

Index Terms—Downlink resource allocation, energy efficiency, mobile devices, wireless systems.

1 Introduction

Advances in wireless communications have enabled ubiquitous personal computing and facilitated the development of various mobile services. Wireless systems have experienced three phases of growth. The first phase (e.g., the first generation cellular systems) focused on voice traffic and the second phase (e.g., the second generation cellular systems) added the support of data traffic [1]. In the third phase, video traffic became the main focus and has been anticipated to account for about 60 percent of mobile data traffic by 2017 [2]. One of the salient features of multimedia applications such as video is their asymmetry, i.e., the data delivery is extremely downlink-intensive in the sense that the amount of downlink data of mobile devices is considerably larger than that of uplink data for such applications [3]. As indicated in some studies, mobile devices require significantly higher power consumption for data reception [4, 5], and thus, energy-efficient system design has become an important issue for future wireless systems. This motivates us to study the energy-efficient downlink resource allocation for mobile devices in wireless cellular systems.

The *orthogonal frequency division multiple access* (OFDMA) technology is adopted by the fourth-generation (4G) wireless systems, e.g., LTE [6] and WiMAX [7], for downlink transmissions. The downlink resource allocation problem for power and channel in OFDMA-based networks has already been extensively studied, which can roughly be classified into *rate adaptive* (RA) and *margin adaptive* (MA) [8]. With respect to power and channel allocation, attention was first paid to maximizing the data rate for all devices with a constraint on the transmit power at the base station, e.g., [8–13], referred to as the RA scheme. However, data rate maximization is not appropriate for real-time multimedia data whose bit rate is generally constant and determined by the adopted compression algorithm [14]. Therefore, some papers aimed at minimizing the transmit power at the base station while meeting devices' data rate requirements, e.g., [14–19] known as the MA scheme. Nevertheless, the transmit power only occupies about 5 percent of the operating power consumption at a base station [20], which will not significantly help the design.

More importantly, we observe that the RA and MA schemes may increase the energy consumption on mobile devices for data reception, and, thus, put further burden on mobile devices with limited battery capacity. This is because the design objectives for both schemes tend to distribute the radio resource for the same device over different time slots, thereby causing a mobile device to stay in the receiving mode for a longer time and consume more energy (we will use a simple example to illustrate this later). Thus, downlink resource allocation in wireless cellular systems should minimize the energy consumption on mobile devices for data reception as the primary objective.

- *Y.-J. Yu is with the Research Center for Information Technology Innovation (CITI), Academia Sinica, No. 128, Sec. 2, Academia Rd., Nankang Dist., Taipei 115, Taiwan, R.O.C., and the Smart Network System Institute, Institute for Information Industry (III), 7F, No. 133, Sec. 4, Minsheng E. Rd., Taipei 105, Taiwan, R.O.C. E-mail: yajuyu@iii.org.tw*
- *A.-C. Pang is with the Graduate Institute of Networking and Multimedia, Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan, R.O.C. E-mail: acpang@csie.ntu.edu.tw.*
- *P.-C. Hsiu is with the Research Center for Information Technology Innovation (CITI), and the Institute of Information Science (IIS), Academia Sinica, No. 128, Sec. 2, Academia Rd., Nankang Dist., Taipei 115, Taiwan, R.O.C. E-mail: pchsiu@citi.sinica.edu.tw*
- *Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611. E-mail: fang@ece.ufl.edu*

A preliminary version of this paper was published in the Proceedings of the IEEE Global Communications Conference (GLOBECOM) 2013.

In this paper, we introduce an alternative design, called *energy adaptive* (EA), into downlink resource allocation for wireless cellular systems. The contributions of this paper are as follows. First, we study the conventional RA and MA, and discover that these two schemes consume unnecessary energy on mobile devices for data reception. This may hinder the development of mobile multimedia applications. Next, we propose EA as an alternative and formulate the energy-efficient downlink resource allocation as an optimization problem to minimize the total energy consumption on mobile devices for data reception while simultaneously meeting the data rate requirement at every device and the transmit power constraint at the base station. Then, we show that the target optimization problem is $\mathcal{NP}$-hard, and then develop a polynomial-time optimal algorithm based on dynamic programming for a special case where only one device and one time slot are considered. For the general case, based on the dynamic-programming algorithm, we develop an efficient algorithm to find approximate solutions and show that, its energy consumption based on the solution derived by the algorithm will be no more than twice of that of the optimal allocation under a certain condition.

Finally, we conduct extensive simulations, with real video sequences encoded by H.264 and the parameters set according to LTE [6], to evaluate the performance of the proposed algorithm. To provide further insights, we compare our algorithms with the two algorithms developed, respectively, based on RA [13] and MA [19], as well as a lower bound estimated for the optimal solution. The RA algorithm proposed in [13] is a greedy heuristic, while the MA algorithm is a dynamic-programming algorithm proposed in [19]. The simulation results agree with what we have observed that RA and MA tend to consume significantly more energy at mobile devices than EA. In addition, the simulation results show a tradeoff between the energy consumption at a base station and that at mobile devices. The tradeoff explains why EA is an alternative for downlink radio resource allocation.

The remainder of this paper is organized as follows. Section 2 reviews related work on radio resource allocation in OFDMA-based networks. In Section 3, we describe the system model and formulate the optimization problem. In Section 4, we show that the problem is $\mathcal{NP}$-hard and propose an efficient algorithm with performance guarantee. The simulation results and some useful insights are discussed in Section 5. Section 6 contains some concluding remarks.

2 Related Work

Radio resource allocation has been considered as one of the most important issues in OFDMA-based wireless networks [21]. In recent years, many researchers have developed effective algorithms to allocate limited radio resources (i.e., power and channel) with various optimizing objective functions (system performance metrics), especially the rate-adaptive scheme (RA) and the margin-adaptive scheme (MA).

RA attempts to maximize the total data rate of all devices under a given transmit power constraint of their base station, e.g., [8–13]. In particular, Jang et al. [9] showed that the data rate in *one single symbol* can be maximized when each subcarrier is assigned to only one device which has the best channel gain for that subcarrier and the transmit power is distributed over the subcarriers by *water-filling policy*. In order to reduce the computational complexity of the water-filling policy, Yu et al. [10] proposed a constant-power water-filling scheme, and showed that the maximum difference between the respective data rates achieved by the constant-power water-filling and the original water-filling policy is bounded by a constant of 1.44 b/s/Hz. Furthermore, Lee et al. [11] proposed a suboptimal subcarrier algorithm with equal power allocation, while Zaki et al. [12] developed optimal and suboptimal algorithms based on graph theory and Lagrangian relaxation. However, devices with poor channel gains (e.g., those far away from the base station) may suffer from starvation. To achieve the fairness of data rates among devices, some studies considered the proportional rate constraint [8] or individual rate constraint [12]. Additionally, Biagioni et al. [13] considered multiple symbols for each resource allocation and provided the "max-min fair" data rate for devices.

On the other hand, because the data rate requirements of some applications (e.g., video streaming) are fixed, Wong et al. [14] proposed a MA scheme to minimize the total transmit power of the base station while satisfying every device's data rate requirement. In [14], a subcarrier allocation algorithm with a lower bound on the total transmit power was proposed, leading to the launch of several studies on the resource allocation problem based on MA. Zhang [15] proposed a heuristic algorithm and demonstrated that the the heuristic algorithm outperforms the suboptimal algorithm proposed in [14] by simulation results. Seong et al. [16] employed the Lagrange dual decomposition method to efficiently solve both the weighted sum rate maximization and weighted sum power minimization problems. In order to reduce the computational complexity for devising an optimal solution, dynamic programming and branch-and-bound based algorithms were proposed in [17–19]. However, the resource allocation algorithms based on RA or MA do not consider the battery capacity of mobile devices and tend to increase the energy consumption on mobile devices for data reception.

Recently, some studies aimed at minimizing the downlink energy consumption for video streaming applications on mobile devices in cellular systems. Video streaming services send traffic at a fixed rate, resulting in mobile devices operating continuously in the active mode and depleting their energy quickly. Siekkinen et al. [22] evaluated the energy saving potential of shaping streaming traffic into bursts before transmitting the traffic over cellular networks to mobile devices. Their results confirmed that traffic shaping is an effective way to save the mobile device's energy, since, in between the bursts, the device has sufficient time to switch from the high-power active state to low-power states. Thus, a number of researchers, e.g., Hoque et al. [23] and Hefeeda et al. [24], explored how to determine an optimal burst interval or minimize the total number of bursts, while maintaining the video stream's smooth playback. An essential problem behind this research direction is to determine *when* a data burst should be transmitted for a mobile device.

In another direction, researches explored *how* a data burst should be transmitted by allocating OFDMA radio resources appropriately, because the energy consumed by a device is considered to be proportional to the number of symbols it receives. In [25], assuming that mobile devices have been allocated with specific subchannels, a heuristic algorithm was proposed to determine the allocation of symbols so as to minimize the total number of symbols received by all mobile devices. In [26], Chu et al. proposed an optimal algorithm only when the data rate requirements of mobile devices can be partitioned into several sets and each set can be satisfied by exactly one symbol. In our previous work [27], we proposed approximation algorithms for energy-efficient video multicast in OFDMA-based wireless networks. Nevertheless, the above algorithms have not considered the situation that a device may have different channel gains on different subchannels, and have not addressed how to allocate the available transmit power of the base station among the subchannels, which is an essential issue of EA in minimizing the downlink energy consumption while meeting every device's data rate requirement.

3 System Model and Problem Formulation

3.1 System Model

The OFDMA-based multi-carrier technology has been widely adopted by 4G wireless cellular networks for downlink transmissions. In 4G wireless systems, data is transmitted with OFDMA frames. A frame consists of *slots* in the time domain and *subchannels* in the frequency domain, as shown in Fig. 1. A slot comprises one or multiple *symbols*, and a subchannel consists of multiple *subcarriers*. Because the fast fading effect may not be the same for different subchannels, a device may have different levels of channel gains on different subchannels [12]. A base station can collect the information about the channel gain either by the measurement reports of devices or by the estimation of the base station itself [12]. In LTE, for example, the base station can set the *channel state information* (CSI) request field [28] to trigger CSI reporting, in which each device will report a *channel quality indicator* (CQI) for each subchannel. For WiMAX, a similar configuration can be set for mobile devices to feedback each sub-band CQI [7].

In the OFDMA frame structure, a combination of a slot and a subchannel, referred to as a *resource block* (RB), is the basic unit for resource allocation. For resource allocation, the base station has to determine which RB should be allocated to which device, referred to as *RB allocation*. Moreover, because *modulation-coding schemes* are adopted in wireless communications, the base station needs to determine the power level allocated to each

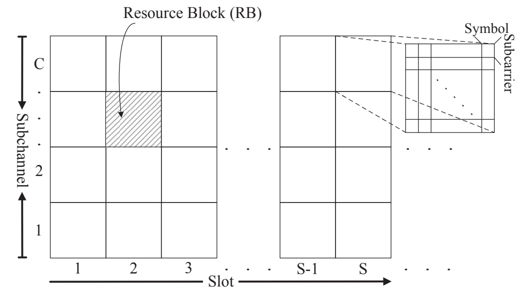


Fig. 1. The OFDMA frame structure.

RB so as to adopt the corresponding modulation-coding scheme, referred to as *power allocation*. For instance, when a higher-rate modulation-coding scheme (e.g., QAM64-3/4) is selected for a device, the base station must allocate a higher power level to the RB in order to meet the required signal-to-noise ratio (lowering the bit error rate). Notice that the available transmit power of a base station is limited. Once the resource allocation is determined, the power consumed by the transceivers at devices during data reception may be different although the transceiver's power level during data reception can be deemed constant [29, 30]. Consequently, the energy consumption for data reception depends on the reception time and is proportional to the number of slots the device needs to receive [25–27][1]. If a mobile device can receive its data in a shorter time, it can save power [7, 26]. As a result, an appropriate resource allocation, i.e., power and RB allocation, can reduce the energy consumption at mobile devices while satisfying their data rate requirements.

3GPP TS23.203 defines a set of *QoS class identifier* (QCI) values with different priorities for different kinds of services [32]. The base station needs to fulfill the demand of higher-priority services, such as VoIP calls and online gaming, before allocating radio resources for video streaming. Then, the remaining radio resources can be used to serve video streaming based on our proposal, as well as other lower-priority applications. Regarding the information about the data rate requirement of the video stream requested by each mobile device, the base station can acquire the information from the streaming service provider, based on a QoS rule, called *policy and charging control* (PCC), specified in [32].

3.2 Problem Formulation

In this paper, we study joint power and RB allocation for downlink transmissions in 4G wireless systems. The objective is to minimize the total energy consumption at mobile devices when receiving downlink data, provided that the data rate requirement of every device is satisfied and the available power of the base station is limited. For the sake of brevity, we omit "$\forall$" when the meaning is clear from the context.

1. In addition to the energy consumption for data reception, there is an extra energy cost, known as *tail energy*, spent by each mobile device in active state after each data transfer [23, 31]. The tail energy is not considered in our system because it cannot be reduced by allocating radio resources appropriately. However, our algorithms remain applicable and the proofs still hold when the tail energy is considered.

For each resource allocation, we consider a set of frames with $S \cdot C$ available RBs to be allocated to N devices, where S is the number of slots and C is the number of subchannels. The channel gains of a device on different subchannels may be different. Let the channel gain of device n on subchannel c be denoted as $g_{(n,c)}$. Every device can have a preferred data rate requirement. Let the data rate requirement of device n be denoted as R_n. The base station has L transmit power levels, denoted by $\ell = 1, 2, \ldots, L$, where a larger index indicates a higher transmit power level. The power of each level is increased by a constant δ. In other words, the base station consumes power $\ell\delta$ when the power level ℓ is used to transmit an RB. Moreover, the base station has M possible modulation-coding schemes, denoted by $m = 1, 2, \ldots, M$, where a larger index indicates a modulation-coding scheme with a higher data rate. A single RB applied with modulation-coding scheme m can provide a data rate of d_m, but the base station has to use a power level that can meet the corresponding signal-to-noise ratio, denoted as SNR_m, where $SNR_1 < SNR_2 < \ldots < SNR_M$. Consequently, if modulation-coding scheme m is used for the RB comprising slot s and subchannel c allocated to device n, the minimum transmit power required for the RB can be calculated by

$$p^n_m(s,c) = \min_{1 \le \ell \le L} \left\{ \ell\delta | \ell\delta \ge \frac{(2^{(SNR_m/10)} - 1) \times \sigma}{g_{(n,c)}} \right\}, \quad (1)$$

where σ is the noise power. The transmit power of the base station available for each slot is limited, i.e., $L\delta$. Moreover, we define $\chi^n_m(s,c)$ as an indicator function, which is 1 if the RB comprises slot s and subchannel c is allocated to device n with modulation-coding scheme m, and 0 otherwise. If any RB in slot s is allocated to device n, the device has to consume an amount of energy, denoted as J_n, for reception during the time slot. A resource allocation is *feasible* if all the following constraints are satisfied.

Resource constraint: Equation (2) ensures that each RB can be modulated with at most one modulation-coding scheme and only be allocated to a device at a time.

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \chi^n_m(s,c) \le 1, \forall\ s, c \quad (2)$$

Requirement constraint: Equation (3) states that every device's requirement must be satisfied with the desired data rate.

$$\sum_{m=1}^{M} \sum_{s=1}^{S} \sum_{c=1}^{C} d_m \cdot \chi^n_m(s,c) \ge R_n, \forall\ n \quad (3)$$

Available transmit power constraint: The total transmit power cannot exceed the available power at the base station.

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{c=1}^{C} \chi^n_m(s,c) \cdot p^n_m(s,c) \le L\delta,\ \forall\ s \quad (4)$$

We now define the target problem formally as follows.

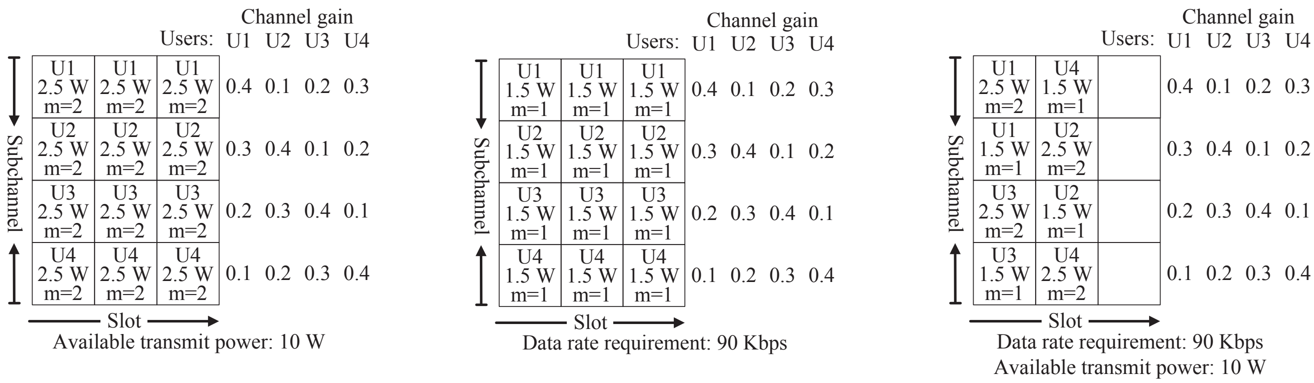
TABLE 1
Summary of Notations

N	The number of devices
S	The number of available slots
C	The number of available subchannels
L	The number of transmit power levels
δ	The power of each level is increased by a constant power δ
M	The number of modulation-coding schemes
SNR_m	The corresponding signal-to-noise ratio for using modulation-coding scheme m
d_m	The data rate provided by a single RB with modulation-coding scheme m
$L\delta$	The available transmit power of the base station
σ	The noise power
R_n	The data rate requirement of device n
J_n	The energy consumption of device n for data reception during a slot
$g_{(n,c)}$	The channel gain of device n on subchannel c
$p^n_m(s,c)$	The power required to transmit the RB, comprised of slot s and subchannel c, to device n with modulation-coding scheme m
$\chi^n_m(s,c)$	An indicator function, which is 1 if the RB comprised of slot s and subchannel c is applied with modulation-coding scheme m and allocated to device n, and 0 otherwise

The Energy-Efficient Downlink Resource Allocation Problem
Input instance: Consider $S \cdot C$ available RBs to be allocated to N devices. Device n has a channel gain $g_{(n,c)}$ on subchannel c and a data rate requirement R_n. The base station has L transmit power levels and M modulation-coding schemes. Each power level ℓ consumes power $\ell\delta$, while each modulation-coding scheme has a corresponding signal-to-noise ratio SNR_m. The RB, comprised slot s and subchannel c, applied with modulation-coding scheme m, can support a data rate of d_m for device n, but requires at least transmit power $p^n_m(s,c)$ calculated according to Equation (1). The available power of the base station for each slot is $L\delta$. The noise power is σ.
Objective: Let J_n be the energy consumption of device n for data reception during a slot. Our objective is to find a feasible resource allocation, $\chi^n_m(s,c)$, such that the total energy consumption of all the devices is minimized. The objective function is expressed as

$$\texttt{Minimize}\ \sum_{n=1}^{N} J_n \times \sum_{s=1}^{S} \left(\bigvee_{m=1}^{M} \bigvee_{c=1}^{C} \chi^n_m(s,c) \right),$$

subject to constraints (2)-(4), where $\bigvee_{m=1}^{M} \bigvee_{c=1}^{C}\ \chi^n_m(s,c)$ indicates whether device n needs to receive data during slot s, and $\sum_{s=1}^{S}(\bigvee_{m=1}^{M} \bigvee_{c=1}^{C}\ \chi^n_m(s,c))$ represents the number of slots during which device n needs to receive data. Table 1 summarizes the notations used in the



(a) Rate adaptive (b) Margin adaptive (c) Energy adaptive

Fig. 2. An illustrative example for the resource allocation problem with different objectives.

problem formulation.

3.3 An Illustrative Example

We use an example, as shown in Fig. 2, to explain our observation on the impact of different resource allocation strategies (i.e., with different objectives) on the energy consumption of mobile devices for data reception. Consider four devices (i.e., $N = 4$), each of which has a data rate requirement of 90 Kbps (i.e., $R_n = 90\text{K}, \forall\, n$). The numbers of slots and subchannels are set, respectively, at 3 and 4 (i.e., $S = 3$ and $C = 4$). The channel gains of each device on the four subchannels are set to be some permutation of 0.4, 0.3, 0.2, and 0.1, as shown in the figure. The base station has 20 transmit power levels (i.e., $L = 20$), and the power of each level is increased by 0.5 W (i.e., $\delta = 0.5$). The power available for each slot is 10 W (i.e., $L\delta = 10$). There are two modulation-coding schemes (i.e., $M = 2$) with corresponding signal-to-noise ratios 6 dB and 10 dB (i.e., $SNR_1 = 6$ and $SNR_2 = 10$), respectively. According to Equation (1), the respective power levels required to use the two modulation-coding schemes are 1.5 W and 2.5 W. A single RB modulated with the two schemes can support a data rate of 30 Kbps and 60 Kbps (i.e., $d_1 = 30\text{K}$ and $d_2 = 60\text{K}$), respectively. The noise power is set at 1 W (i.e., $\sigma = 1$). We assume that the energy consumption required for data reception during any slot is set at 320 μJ (i.e., $J_n = 320\mu$). Then, the base station determines the power and RB allocation under RA, MA, and EA, respectively.

Fig. 2(a) shows an optimal solution for the resource allocation problem with RA. In order to maximize the data rate, the base station always allocates each RB to the device with the highest channel gain. As a result, each RB is allocated to a device with channel gain 0.4. Moreover, modulation-coding scheme 2 is used and each RB is allocated equally with 2.5 W since the available transmission power is 10 W. Consequently, each device needs to receive data during all the three slots, and the total energy consumption for data reception is $3 \times 4 \times 320$ μJ = 3840 μJ. Fig. 2(b) shows an optimal solution for MA. To minimize the transmit power, the base station also tries to allocate each RB to a device with the highest channel gain 0.4. However, it uses modulation-coding scheme 1 and allocates only 1.5 W to each RB so that all devices' requirements can be satisfied at least. In this example, the four devices consume 3840 μJ as well.

With the EA objective, as shown in Fig. 2(c), the base station tends to allocate RBs in the same slot to each device. As a result, some RBs are allocated to devices with channel gain 0.4, while some RBs allocated to devices with channel gain 0.3. Moreover, 2.5 W and 1.5 W are, respectively, allocated to the RBs using modulation-coding schemes 2 and 1 associated with channel gain 0.4 and 0.3. Based on the allocation, each device can be satisfied with a data rate of 60 Kbps + 30 Kbps = 90 Kbps, and the total transmit power in a slot is $2\times$(2.5 W+1.5 W) = 8 W, which is less than the available power 10 W. More importantly, the total energy consumption for data reception is only 4×320 μJ = 1280 μJ. Comparing Fig. 2(b) and Fig. 2(c), we observe an interesting phenomenon that, in order to save the transmit power of the base station, each mobile device will stay in receive mode for a longer time and consume more energy to achieve the required data rate.

4 Energy-Efficient Downlink Resource Allocation

In this section, we consider the energy-efficient downlink resource allocation problem. In Section 4.1, we show that the problem is $\mathcal{NP}$-hard. To simplify the problem, in Section 4.2, we consider a special case with one device in one slot and propose a polynomial-time optimal algorithm based on dynamic programming to maximize the data rate received by the device during a single slot. Then, in Section 4.3, we present an efficient algorithm, which relies on the algorithm presented in Section 4.2, for the general case and prove that it is a 2-approximation algorithm under a certain condition.

4.1 Problem Hardness

In this subsection, we show that the target problem is $\mathcal{NP}$-hard by a reduction from the *partition problem*, which is known to be $\mathcal{NP}$-complete [33].

***Theorem** 1: The energy-efficient downlink resource allocation problem is $\mathcal{NP}$-hard.*

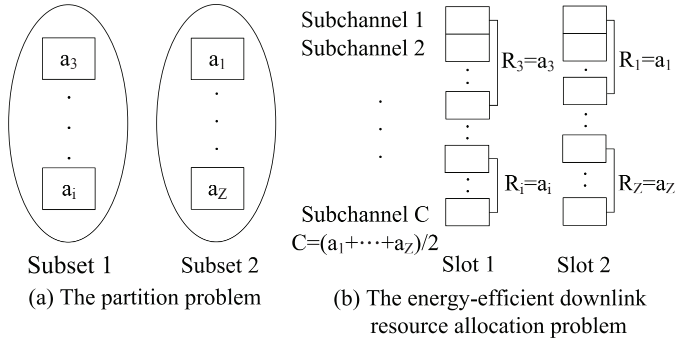


Fig. 3. An illustration of the NP-hard proof

Proof*:* The input instance of the partition problem is a set of Z integers, $A = \{a_1, a_2, \ldots, a_Z\}$. The output is YES if and only if A can be partitioned into two subsets A' and $A \backslash A'$ with the same sum, i.e., $\sum_{a_i \in A'} a_i = \sum_{a_i \notin A'} a_i = \frac{1}{2}\sum_{a_i \in A} a_i$, as illustrated in Fig. 3(a).

Given an instance $\langle A \rangle$ of the partition problem, we explain how to construct an instance $\langle N, S, C, L, \delta, M, SNR_m, d_m, \sigma, R_n, g_{(n,c)} \rangle$ of our problem in polynomial time such that A can be evenly partitioned if and only if there exists a resource allocation whose total energy consumption is not more than Z. The construction is as follows. As illustrated in Fig. 3(b), consider Z devices (i.e., $N = Z$), and each device i requires a data rate of a_i bps (i.e., $R_i = a_i$, $\forall 1 \le i \le Z$). The numbers of available slots and subchannels are, respectively, set as $S = 2$ and $C = \frac{1}{2}\sum_{a_i \in A} a_i$. There are $\frac{1}{2}\sum_{a_i \in A} a_i$ transmit power levels (i.e., $L = \frac{1}{2}\sum_{a_i \in A} a_i$) and the power of each level is increased by 1 W (i.e., $\delta = 1$). There is only one modulation-coding scheme (i.e., $M = 1$) whose corresponding signal-to-noise ratio is 10 dB (i.e., $SNR_1 = 10$). Each RB modulated with the modulation-coding scheme can provide a data rate of 1 bps (i.e., $d_1 = 1$). The channel gain of all devices on any subchannel is set at 1 (i.e., $g_{(n,c)} = 1$, $\forall$ $1 \le c \le C$, $1 \le n \le 2$). Let the noise power be $\sigma = 1$. According to Equation (1), the power required for a single RB allocated to any device is 1 W (i.e., $p^n_m(s, c) = 1, \forall 1 \le n \le Z, 1 \le s \le 2, 1 \le c \le C$). Moreover, we set the energy consumption of any device for date reception during one slot at 1 (i.e., $J_n = 1, \forall 1 \le n \le Z$).

To complete the proof, we show that two partitioned subsets can be used to derive a resource allocation whose total energy consumption is not more than Z, and vice versa. In one direction, if A can be evenly partitioned into two subsets, we explain how the two subsets can be used to derive a resource allocation whose total energy consumption is not more than Z. Because each subset corresponds to a slot and each integer corresponds to a device's data rate requirement, the assignment of integer a_i to a subset implies that a_i RBs in the corresponding slot are allocated to device i such that the device's requirement can be satisfied. For example, as shown in Fig. 3, the assignment of integer a_1 to subset 2 implies the allocation of a_1 RBs in slot 2 to device 1. In this resource allocation, every device needs to receive data during exactly one slot and its energy consumption is 1. Because there are Z devices, the total energy consumption is Z.

In the other direction, if there exists a resource allocation whose energy consumption is not more than Z, we describe how the resource allocation can be used to partition A evenly into two subsets. Because the total energy consumption is not more than Z and every device's requirement must be satisfied, no device should receive data during more than one slot. By a similar argument, the allocation of a_i RBs in a slot to satisfy the requirement of device i implies the assignment of integer a_i to the corresponding subset. Therefore, the set can be evenly partitioned by assigning the integers into two subsets in accordance with the resource allocation. The existence of a polynomial-time algorithm for the partition problem implies the same for ours, which completes the proof. □

4.2 A Special Case

4.2.1 A Polynomial Time Optimal Algorithm

Next, we consider a special case of the target problem when there is only one device and one slot (i.e., $N = 1$ and $S = 1$). Since there is only one slot, the total energy consumed by the only device is J_n, and the problem is thus to determine whether the data rate requirement R_n can be satisfied by the single slot. In other words, an algorithm is optimal if it can derive the maximum data rate provided by the single slot for the device.

To solve this special case, we propose a polynomial-time optimal algorithm based on dynamic programming. It determines which subchannel should be allocated to the device and with which modulation-coding scheme so that the data rate provided by the slot is maximized. The proposed algorithm is based on the recursive formula given in Equation (5). Let $f(c, m, \ell)$ be the maximum data rate achieved by the first c subchannels, where subchannel c can be applied with one of the first m modulation-coding schemes and any other subchannel can be applied with one of the M modulation-coding schemes, when the transmit power available for the slot is $\ell\delta$. We delineate three possible cases in Equation (5):

(1) If $c = 0$ or $m = 0$, $f(c, m, \ell)$ is set at 0. That is, no data rate can be provided because there is no subchannel or modulation-coding scheme.

(2) If $\ell\delta - p^n_m(s, c) < 0$, then $f(c, m, \ell)$ is set as $\max(f(c, m-1, \ell), f(c-1, M, \ell))$. That is, it is impossible to apply subchannel c with modulation-coding scheme m when the available power is only $\ell\delta$. Therefore, the maximum data rate is achieved by one of the two possible selections: 1) subchannel c is applied with one of the first $m-1$ modulation-coding schemes; or 2) subchannel c is not used, and the maximum data rate is provided by the first $c-1$ subchannels, each of which can be applied with any of the M modulation-coding schemes.

(3) Otherwise, subchannel c is either applied or not applied with modulation-coding scheme m. If it is applied with modulation-coding scheme m, the base station has to consume power $p^n_m(s, c)$ and the device can obtain a data rate of d_m. Then, the

$$f(c,m,\ell)=\begin{cases}0, & \text{if } c=0 \text{ or } m=0;\\ \max\Big(f(c,m-1,\ell),\ f(c-1,M,\ell)\Big), & \text{else if } \ell\delta-p^n_m(s,c)<0;\\ \max\Big(f(c-1,M,\ell-\frac{p^n_m(s,c)}{\delta})+d_m,\ f(c,m-1,\ell),\ f(c-1,M,\ell)\Big), & \text{otherwise.}\end{cases} \tag{5}$$

remaining transmit power, i.e., $\ell\delta-p^n_m(s,c)$, can be used for the first $c-1$ subchannels, each of which can be applied with one of the M modulation-coding schemes. Thus, the maximum data rate achieved by this selection is $f(c-1,M,\ell-\frac{p^n_m(s,c)}{\delta})+d_m$. In contrast, if subchannel c is not applied with modulation-coding scheme m, the maximum data rate can only be achieved by the two possible selections as discussed in Case (2). Thus, $f(c,m,\ell)$ is set as the maximum of the data rates achieved by the three selections.

Algorithm 1

Require: $N=1$, $S=1$, C, L, δ, M, SNR_m, d_m, σ, R_1, $g_{(1,c)}$
Ensure: $\chi^n_m(s,c)$
1: $\chi^n_m(s,c)\leftarrow 0, \forall m,c$
2: FILL-TABLE(C,M,L)
3: BACK-TRACE(C,M,L)
4: **return** $\chi^n_m(s,c)$, $\forall\ m,c$

Algorithm 1 implements the dynamic-programming formula in Equation (5). First, an indicator function $\chi^n_m(s,c)$, indicating whether subchannel c is applied with modulation-coding scheme m and allocated to the device, is initialized as 0, $\forall m,c$. Then, the algorithm maintains a 3-dimensional table $f[]$, each entry of which stores the solution derived by $f(c,m,\ell)$. When Procedure FILL-TABLE() is invoked, it simply fills in the corresponding table $f[]$ according to Equation (5). After the table is completed, Procedure BACK-TRACE() is invoked to select the subchannels and the corresponding modulation-coding schemes such that the data rate provided in the slot is maximized by back tracing the table, and set the corresponding $\chi^n_m(s,c)$, $\forall\ m,c$, as 1. Finally, the algorithm return the solution $\chi^n_m(s,c)$.

Procedure FILL-TABLE(C,M,L)
1: **for** $c\leftarrow 0$ to C **do**
2:   **for** $m\leftarrow 0$ to M **do**
3:     **for** $\ell\leftarrow 0$ to L **do**
4:       **if** $c=0$ or $m=0$ **then**
5:         $f[c,m,\ell]\leftarrow 0$
6:       **else if** $\ell\delta-p^n_m(s,c)<0$ **then**
7:         $f[c,m,\ell]\leftarrow\max(f[c,m-1,\ell],f[c-1,M,\ell])$
8:       **else**
9:         $f[c,m,\ell]\leftarrow\max(f[c-1,M,\ell-\frac{p^n_m(s,c)}{\delta}]+d_m,$ $f[c,m-1,\ell],f[c-1,M,\ell])$

Procedure FILL-TABLE() takes C, M, and L as inputs. It fills in each table entry $f[c,m,\ell]$ with the value derived based on Equation (5) (Lines 4-9). Because filling in a table energy may refer to some other entries, the table entries are computed in sequence, i.e., the dimension from $\ell=0$ to L first, then the dimension from $m=0$ to M and, finally, the dimension from $c=0$ to C.

Procedure BACK-TRACE(C,M,L)
1: $c\leftarrow C$
2: $m\leftarrow M$
3: $\ell\leftarrow L$
4: **while** $c>0$ **do**
5:   **if** $f[c,m,\ell]=f[c-1,M,\ell]$ **then**
6:     $c\leftarrow c-1$
7:     $m\leftarrow M$
8:   **else if** $f[c,m,\ell]=f[c,m-1,\ell]$ **then**
9:     $m\leftarrow m-1$
10:   **else**
11:     $\chi^n_m(s,c)\leftarrow 1$
12:     $c\leftarrow c-1$
13:     $m\leftarrow M$
14:     $\ell\leftarrow\ell-\frac{p^n_m(s,c)}{\delta}$

Procedure BACK-TRACE() also takes C, M, and L as inputs. It selects subchannels and modulation-coding schemes to provide the maximum data rate for the device by back tracing table $f[]$. We begin with the last entry (i.e., $f[C,M,L]$) by setting three indexes c, m, and ℓ as C, M, and L respectively. During Procedure FILL-TABLE(), $f[c,m,\ell]$ is set as the maximum among $f[c-1,M,\ell]$, $f[c,m-1,\ell]$, and $f[c-1,M,\ell-\frac{p^n_m(s,c)}{\delta}]+d_m$. We discuss the three cases. If the maximum is the first term, then subchannel c is not used for transmission, so c is updated to $c-1$ and m is updated to M (Lines 5-7). If the maximum is the second term, then subchannel c is not applied with modulation-coding scheme m but one of the first $m-1$ modulation-coding schemes; thus m is updated to $m-1$ (Lines 8-9). Otherwise, the maximum is the third term, and c is applied with modulation m. Therefore, $\chi^n_m(s,c)$ is set as 1, and the three indexes are updated to their corresponding values (Lines 10-14). Then, we start with the entry indexed by the updated c, m, and ℓ, and repeat the above process until all the subchannels have been examined.

4.2.2 *The Properties of Algorithm 1*

In this section, we analyze the time complexity of Algorithm 1, and prove that it is a polynomial-time optimal algorithm.

***Lemma** 1: The time complexity of Algorithm 1 is $O(CML)$.*

Proof: The time complexity of the algorithm depends on the number of table entries and the time required to derive the value of each entry. The 3-dimensional table comprises CML table entries. Moreover, deriving the value of each entry $f[]$ by referring to at most three other entries takes $O(1)$ time. Since a derived value will never be changed, the table can be completed in $O(CML)$. On the other hand, constructing the corresponding allocation by back tracing the table has to examine at most $O(CM)$ entries, and each examination takes constant time $O(1)$. Thus, the time complexity of Algorithm 1 is $O(CML)$. □

***Theorem** 2:* Algorithm 1 yields the maximum data rate for a device during a slot.

Proof*:* We prove this theorem by two-dimensional mathematical induction on the indexes c and m. As the induction basis, $m = 0$, there is no modulation-coding scheme that can be used to provide any data rate. Thus, the maximum data rate is 0, i.e., $f(c, 0, \ell) = 0, \forall c, \ell$. For the induction hypothesis, suppose that $f(0, m-1, \ell), \forall \ell$, is correct for some positive integer m. We show that the formula $f(0, m, \ell), \forall \ell$, is also correct. When $c = 0$, there is no subchannel and it is impossible to provide any data rate. Thus, the maximum data rate is 0, i.e., $f(0, m, \ell) = 0, \forall m, \ell$. The theorem is correct when $c = 0$.

Next, let us consider the case when $c = 1$. For the induction hypothesis, we also suppose that the formula $f(0, m-1, \ell), \forall \ell$, is correct for some positive integer m. We show that the formula $f(1, m, \ell), \forall \ell$, is also correct. If the available transmit power $\ell\delta$ is less than $p^n_m(s, 1)$, it is impossible to apply subchannel 1 with modulation-coding scheme m. Then, the maximum data rate must be achieved by subchannel 1 applied with one of the first $m-1$ modulation-coding scheme or the first 0 subchannels. By the induction hypothesis and the correctness proved when $c = 0$, the maximum data rate $f(1, m, \ell) = \max(f(1, m-1, \ell), f(0, M, \ell))$ if $\ell\delta - p^n_m(s, 1) < 0$. Otherwise, if subchannel 1 is applied with modulation-coding scheme m, the maximum data rate $f(1, m, \ell) = f(0, M, \ell - \frac{p^n_m(s,1)}{\delta}) + d_m$. In contrast, if it is not applied with modulation-coding scheme m, the maximum data rate $f(1, m, \ell) = \max(f(1, m-1, \ell), f(0, M, \ell))$. Because subchannel 1 is either applied or not applied with modulation-coding scheme m, the maximum data rate is the larger one of the two values. Thus, the theorem is correct when $c = 1$. The validity of the theorem when $c \geq 2$ can be proved similarly. Thus, we conclude that the formula $f(c, m, \ell), \forall c, m, \ell$, is correct. □

4.3 The General Case

4.3.1 Algorithm Description

In this section, we propose an efficient algorithm to solve the general case with multiple devices in multiple slots. It relies on the dynamic-programming algorithm presented in Section 4.2 to derive an indicator function that maximizes the data rate for each device during a single slot. Then, based on the indicator functions, it allocates RBs in column-major order to all the devices one by one to reduce the energy consumption for data reception.

The algorithm, as shown in 2, starts with initialization. An indicator function $\chi^n_m(s, c)$ registers if an RB, comprised of slot s and subchannel c applied with modulation-coding scheme m, is allocated to device n,

Algorithm 2

Require: N, S, C, L, δ, M, SNR_m, d_m, σ, R_n, $g_{(n,c)}$
Ensure: $\chi^n_m(s, c)$
1: $\chi^n_m(s, c) \leftarrow 0, \forall n, m, s, c$
2: $\hat{s} \leftarrow 1$
3: $P_{BS} \leftarrow L\delta$
4: **for** $\hat{n} \leftarrow 1$ to N **do**
5: $\quad \hat{\chi}^{\hat{n}}_m(\hat{s}, c), \forall m, c \leftarrow$ Algorithm 1 $(\hat{n}, \hat{s})$
6: $\quad$ **if** $\hat{\chi}^{\hat{n}}_m(\hat{s}, c) = 0, \forall m$ **then**
7: $\quad\quad \hat{\chi}^{\hat{n}}_1(\hat{s}, c) \leftarrow 1, \forall c$
8: $\quad$ **while** $R_{\hat{n}} > 0$ and $\hat{s} \leq S$ **do**
9: $\quad\quad \hat{c} \leftarrow \arg\max_c \{\hat{m} | \hat{\chi}^{\hat{n}}_{\hat{m}}(\hat{s}, c) = 1$ and $\chi^n_m(\hat{s}, c) = 0, \forall n, m\}$
10: $\quad\quad P_{BS} \leftarrow P_{BS} - p^{\hat{n}}_{\hat{m}}(\hat{s}, \hat{c})$
11: $\quad\quad$ **if** $P_{BS} \geq 0$ and $\hat{c} \neq \emptyset$ **then**
12: $\quad\quad\quad R_{\hat{n}} \leftarrow R_{\hat{n}} - d_{\hat{m}}$
13: $\quad\quad\quad \chi^{\hat{n}}_{\hat{m}}(\hat{s}, \hat{c}) \leftarrow 1$
14: $\quad\quad$ **else**
15: $\quad\quad\quad \hat{s} \leftarrow \hat{s} + 1$
16: $\quad\quad\quad P_{BS} \leftarrow L\delta$
17: **return** $\chi^n_m(s, c)$, $\forall$ n, m, s, c

and is initialized at 0, $\forall$ n, m, s, c (Line 1). A variable $\hat{s}$ initialized at 1 is employed to index the currently used slot (Line 2); and P_{BS} maintains the remaining transmit power for the currently used slot and initialized at $L\delta$ (Line 3). Then, the algorithm attempts to minimize the total energy consumption, while satisfying each device's requirement sequentially (Line 4). For each device $\hat{n}$, the algorithm uses Algorithm 1 to derive $\hat{\chi}^{\hat{n}}_m(\hat{s}, c), \forall m, c$, which indicates the modulation-coding scheme for each subchannel such that the data rate at slot $\hat{s}$ is maximized for the device $\hat{n}$ (Line 5). In the indicator function derived by Algorithm 1, some subchannels may not be applied with any modulation-coding scheme due to the limited transmit power. Those subchannels are assigned with the lowest-rate modulation-coding scheme by default (Lines 6-7). This design is to further improve the radio resource utilization of a partially used slot because some subchannels that can be allocated to device $\hat{n}$ may have already been allocated to other devices.

Next, the algorithm starts to allocate RBs to device $\hat{n}$ until its data rate requirement is satisfied or the radio resources are exhausted (Line 8). In order to minimize the energy consumption for data reception, RBs are allocated one by one in column-major order. Based on the indicator function derived by Algorithm 1 in Line 5, Algorithm 2 finds an unused subchannel $\hat{c}$ applied with the highest modulation-coding scheme $\hat{m}$ so as to allocate as few RBs in the current slot $\hat{s}$ as possible to device $\hat{n}$ (Line 9). Note that the tie is broken by selecting the subchannel that requires the least transmit power. To apply subchannel $\hat{c}$ with modulation-coding scheme $\hat{m}$, the transmit power at the base station has to be decreased by $p^{\hat{n}}_{\hat{m}}(\hat{s}, \hat{c})$ (Line 10). For the current slot $\hat{s}$, if there are remaining transmit power (i.e., $P_{BS} \geq 0$) and any unused subchannel (i.e., $\hat{c} \neq \emptyset$), the subchannel $\hat{c}$ can be allocated to device $\hat{n}$ and provides a data rate of $d_{\hat{m}}$; moreover, the counterpart of the indicator function $\chi^{\hat{n}}_{\hat{m}}(\hat{s}, \hat{c})$ is set as 1 accordingly (Lines 11-13). Otherwise, the variable $\hat{s}$ moves to the next slot and the transmit power available for this new slot is reset at $L\delta$ (Lines 14-16). Finally, after all the devices have been considered, the algorithm returns the derived resource allocation $\chi_m^n(s,c)$ (Line 17). Based on the returned $\chi_m^n(s,c)$, the base station can examine whether all the devices' requirements are satisfied by the available radio resources[2].

4.3.2 The Properties of Algorithm 2

***Lemma** 2: The time complexity of Algorithm 2 is $O(NCML + SC^2)$.*

Proof*:* For each device, Algorithm 1 is invoked once and takes $O(CML)$ time to derive an indicator function for the device, as analyzed in Lemma 1. Thus, it takes $O(NCML)$ time for power allocation. Moreover, when allocating an RB to the device, Algorithm 2 finds an unused subchannel $\hat{c}$ and an appropriate modulation-coding scheme $\hat{m}$ for the slot (see Line 9). In other words, we have to ensure subchannel $\hat{c}$ is unused (i.e., $\chi_m^n(\hat{s},\hat{c}) = 0, \forall n, m$), and search for the indicator function derived by Algorithm 1 to determine the modulation-coding scheme $\hat{m}$ that can be applied to $\hat{c}$ (i.e., $\hat{\chi}_{\hat{m}}^{\hat{n}}(\hat{s},\hat{c}) = 1$). To this end, we can employ two additional variables for each subchannel to indicate whether it is unused or not and which modulation-coding scheme is applied, so that finding an unused subchannel with the highest modulation-coding scheme can be done in $O(C)$ time. Because there are $S \times C$ RBs to be allocated one by one, it takes $O(SC^2)$ time for RB allocation. □

***Theorem** 3: Algorithm 2 is a 2-approximation algorithm when $\sum_{n=1}^{N} \left\lceil \frac{R_n}{r_n^*} \right\rceil \leq S$, where r_n^* is the maximum data rate provided by a single slot for device n.*

Proof*:* We prove this theorem by showing that, based on the allocation $\chi_m^n(s,c)$ returned by Algorithm 2, each device needs to receive at most one slot more than that it receives in any optimal solution. For any device n in the network system, following the column-major order for resource allocation, Algorithm 2 starts to allocate RBs to device n in the same (current) slot as the previous device if there is any vacant subchannel in that slot. Let $\hat{C}_n$ denote the set of vacant subchannels in the current slot, where $1 \leq |\hat{C}_n| \leq C$. If all the subchannels in $\hat{C}_n$ are allocated to device n, the device can obtain a data rate of at most ρ_n. We delineate two possible cases: 1) If the data rate requirement of device n is less than or equal to ρ_n, it only needs to receive data during the current slot. 2) Otherwise, the remaining requirement must be satisfied by other slots. If all the subchannels in a new slot are allocated to device n, it can obtain a data rate of at most r_n^*, where r_n^* denotes the maximum data rate provided by a slot for device n. Note that r_n^* can be achieved by Algorithm 1, as proved in Theorem 2. Let β_n denote the number of slots during which device n needs to receive data. Then, we have

$$\beta_n = \begin{cases} 1, \text{ if } 1 \leq R_n \leq \rho_n, \\ 1 + \left\lceil \frac{R_n - \rho_n}{r_n^*} \right\rceil, \text{ otherwise.} \end{cases} \tag{6}$$

On the other hand, let β_n^* denote the number of slots during which device n needs to receive data in an optimal solution. Since r_n^* is the maximum data rate provided by a single slot for device n, the requirement R_n must be associated with at least $\lceil \frac{R_n}{r_n^*} \rceil$ slots, in the sense that

$$\beta_n^* \geq \left\lceil \frac{R_n}{r_n^*} \right\rceil \geq 1. \tag{7}$$

Based on (6) and (7), we have

$$\beta_n \leq 1 + \left\lceil \frac{R_n}{r_n^*} \right\rceil \leq 1 + \beta_n^*.$$

Now, we compute the total energy consumption of all devices for data reception in the allocation $\chi_m^n(s,c)$ derived by Algorithm 2.

$$\sum_{n=1}^{N} J_n \times \sum_{s=1}^{S} \left(\bigvee_{m=1}^{M} \bigvee_{c=1}^{C} \chi_m^n(s,c) \right) = \sum_{n=1}^{N} J_n \times \beta_n$$
$$\leq \sum_{n=1}^{N} J_n \times (1 + \beta_n^*) \leq 2 \times \sum_{n=1}^{N} J_n \times \beta_n^*$$

Notice that when $\sum_{n=1}^{N} \left\lceil \frac{R_n}{r_n^*} \right\rceil \leq S$, all the devices can be satisfied by the available radio resources. Thus we can conclude that under the condition, Algorithm 2 can derive a feasible solution whose energy consumption is no more than twice that of any optimal solution. □

2. We investigate the probability of returning a feasible solution via extensive simulations, and discuss how to handle the circumstance that some devices's requirements cannot be satisfied in Section 5.

TABLE 2
Modulation-Coding Schemes with SNR Ranges

m	Modulation	Coding rate	d_m (kbps)	SNR range (dB)
1	QPSK	0.438	3.95	3
2	QPSK	0.588	5.29	5
3	QAM16	0.476	8.61	8.5
4	QAM16	0.602	10.82	10
5	QAM64	0.455	12.29	11.5
6	QAM64	0.554	14.95	13.5

5 Performance Evaluation

In this section, we carry out the evaluation of our proposed allocation scheme by conducting simulation study.

5.1 Simulation Setup

We develop a simulation model to evaluate the performance of the proposed algorithm, i.e., Algorithm 2. The parameter settings are based on the LTE specification for a 10 MHz spectrum [6]. The number of available slots for the resource allocation was set at 40, with each

TABLE 3
Data Rates of Video Sequences Encoded by H.264

Video sequence	Data rate requirement
Akiyo	23.69Kbps
Grandma	29.51Kbps
Container	32.65Kbps
Hall monitor	46.62Kbps
News	71.1Kbps

slot time 0.5 ms. The number of available subchannels is set to 50. The network consists of one base station and multiple mobile devices (MDs), each of which is placed randomly in the coverage area of the base station with a distance between 100 to 300 meters from the base station. The base station supports a number of discrete transmit power levels, where the difference in power between two adjacent levels is 1 W. Moreover, the base station has six modulation-coding schemes. The corresponding data rate and the signal-to-noise (SNR) required to overcome, listed in Table 2, are based on the LTE specification [28]. The power consumption of each device in receive mode is assigned randomly at 620mW [4] or 1400mW [5] to consider heterogeneous mobile devices. Accordingly, the energy consumed by a device to receive data during one slot is 310μJ or 700μJ. Path loss, shadowing, and multipath fading are taken into consideration in the simulation model. We adopt the *log-distance path loss model* with propagation loss exponent of 3 for cellular networks [34, 35] and a reference distance of 100 meters. The shadowing follows a *log-normal distribution* with zero mean and standard deviation of 8 dB. Moreover, the subchannels are assumed to suffer from *multipath Rayleigh fading*.

We use five standard video test sequences with QCIF resolution, namely Akiyo, Grandma, Container, Hall Monitor, and News, all of which can be downloaded from [36]. Each device randomly selects a test sequence. The video sequences are encoded by H.264 with JM, version 18.4, and the IBBP encoding structure is used. The other encoding parameters are set at the default values of H.264 with JM. The data rate of each video sequence encoded is listed in Table 3.

We compared the proposed algorithm, denoted as *EA*, with two algorithms. The first algorithm was developed for the RA objective [13], denoted as *RA*, which attempts to maximize the total data rate of all devices under a given transmit power constraint. Initially, the transmit power available for each slot is allocated equally to all the RBs in the slot; meanwhile, such power allocation ensures that the transmit power constraint will not be violated. Then, the device with the lowest data rate is selected and an RB is allocated to it. Among those available RBs, the RB to be allocated is selected to provide the best signal-to-noise ratio (which implies the highest data rate) to that device. Ties are broken at random. The above process is repeated until all the RBs are allocated.

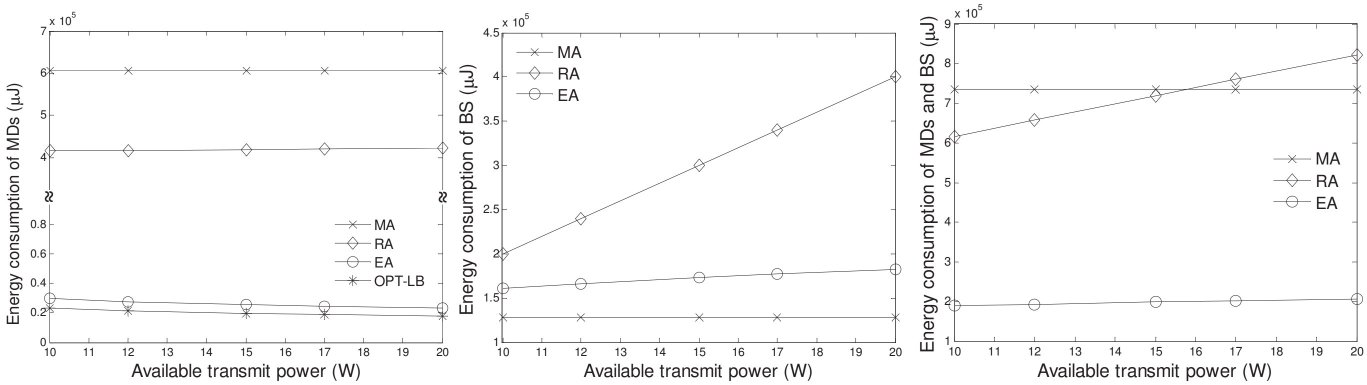
The second algorithm, denoted as *MA*, was proposed in [19] and intended to minimize the total transmit power of the base station while satisfying every device's data rate requirement. Because the MA algorithm considers only one single slot, we divide each device's requirement by S, and apply the resource allocation determined for one slot to all the S slots. Firstly, the algorithm sorts the C RBs in the slot in a descending order according to their maximum channel gains (recall that an RB is associated with one of N channel gains). Then, it allocates all the RBs temporarily to every device, and makes a series of C decisions. For the jth decision, the jth RB is given to one of the N devices and simultaneously removed from all the others. Based on the specific RB allocation, a modified water-filling power allocation algorithm is employed to calculate the minimum transmit power required to satisfy each device's data rate requirement. By considering all the N possible allocations, the jth RB is actually allocated to the device such that the total transmit power is minimized. After the C decisions made in sequence, each RB in the slot will be allocated to only one device.

In addition to the three algorithms, a lower bound estimated based on Equation (7) for the optimal solution, denoted as *OPT-LB*, is adopted as a baseline for comparison. The performance metrics are the respective amounts of energy consumed by mobile devices and by the base station, as well as the sum of the two amounts. The metrics including the percentage of unallocated RBs and the total data rate of mobile devices are further considered. We investigate the impacts of the available transmit power of the base station in the range 10 to 20 W when the number of devices is set at 40, as well as the impacts of the number of devices in the range 5 to 50 when the available transmit power is set at 20 W [20].

All of the simulations were run on a platform with an Intel i5 Dual-Core 1.6 GHz CPU and a 6GB DDR3 RAM. The simulation results reported are the average values of 1000 independent runs. In addition to the above studies, we evaluate the average running time required by each of the three algorithms per run. We also investigate the percentage of the feasible solutions derived over the 1000 runs to demonstrate the feasibility of the proposed algorithm. Note that *MA* and *RA* always return feasible solutions (in terms of their objectives) because *RA* does not ensure that every device's requirement is satisfied while *MA* does not limit the base station's available transmit power.

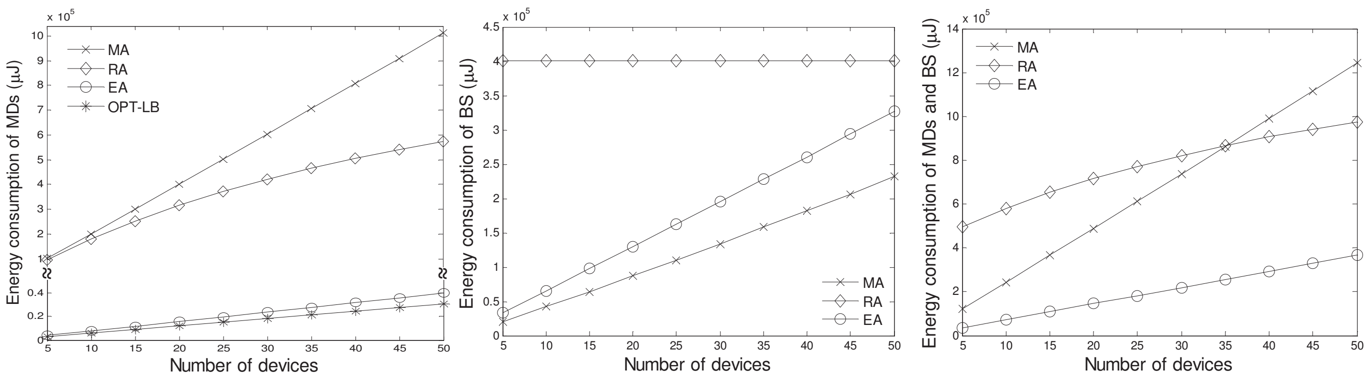
5.2 Energy Consumption

Fig. 4(a) shows the impact of the available transmit power of the base station on the energy consumption of mobile devices. The energy consumption decreases slightly as the available transmit power increases under *EA* and *OPT-LB*. The result is as expected because an RB allocated with larger power can be applied with a higher rate modulation-coding scheme. As a result, the requirement of each device can be satisfied with fewer RBs and, consequently, the device needs to receive data during fewer slots. In contrast, the available transmit power does not have a significant impact on the energy consumption of mobile devices under *RA* and *MA*. It



(a) Mobile Devices (MDs) (b) Base Station (c) Mobile Devices and Base Station

Fig. 4. The impacts of the available transmit power on the energy consumption under 40 devices.



(a) Mobile Devices (b) Base Station (c) Mobile Devices and Base Station

Fig. 5. The impacts of the number of devices on the energy consumption under the available transmit power of 20 W.

implies that both *RA* and *MA* do not leverage the available transmit power to reduce the energy consumed by mobile devices. The simulation results show that the mobile devices consume 14 to 26 times more energy under *RA* and *MA* than under *EA*. Moreover, the energy consumption of mobile devices under *EA* is close to that under *OPT-LB*.

We are also interested in the energy consumption of the base station under different objective functions. As shown in Fig. 4(b), the energy consumption of the base station increases as the available transmit power increases under *EA* and *RA*, and the increase is more obvious under *RA* than under *EA*. This is because *RA* tends to exhaust the available power over all the slots to improve the data rate, while *EA* attempts to use as few slots as possible to satisfy the data rate requirements. On the other hand, the base station consumes 1.2 to 1.4 times more energy under *EA* than under *MA*. Comparing Fig. 4(a) with Fig. 4(b), we observe an interesting tradeoff between the energy consumption of the base station and that of mobile devices. If the base station attempts to save the transmit power, then mobile devices may consume more energy for data reception, and vice versa. This tradeoff also explains why we use *energy adaptive* to represent the new concept raised in this paper for resource allocation. More importantly, when the total energy consumption of the base station and mobile devices is considered, *EA* is more energy efficient than *MA* by 3.7 times, as shown in Fig. 4(c).

Fig. 5(a) shows the impact of the number of devices on the energy consumption of the mobile devices. The energy consumption increases as the number of devices increases under all the algorithms. The reason for the increase under *EA* is that, as the number of devices increases, the amount of the data rate requirements increases. The more the RBs allocated to satisfy the requirements, the larger the energy the mobile devices consume. The increase is much more evident under *RA* and *MA*. This is because they usually scatter the RBs allocated to the same device over different slots; consequently, the total energy consumption increases, in general, proportionally to the number of devices. The simulation results show that the mobile devices consume 14 to 26 times more energy under *RA* and *MA* than under *EA*; moreover, the energy consumption under *EA* is close to that under *OPT-LB*. The result is similar to that in Fig. 4(a) and also demonstrates that both *RA* and *MA* cause mobile devices to consume more energy for data reception.

As shown in Fig. 5(b), the energy consumption of the base station increases as the number of devices increases under *MA* and *EA*. This is because there are more data rate requirements, so the base station has to consume more energy to transmit adequate RBs to satisfy the

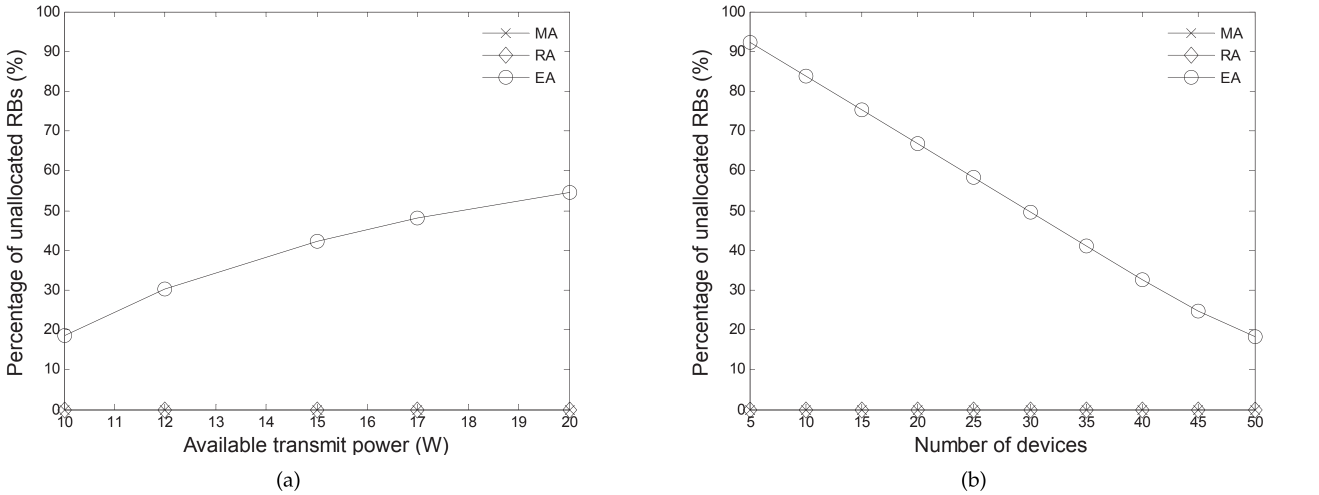


Fig. 6. The respective impacts of the available transmit power and the number of devices on the percentage of unallocated RBs.

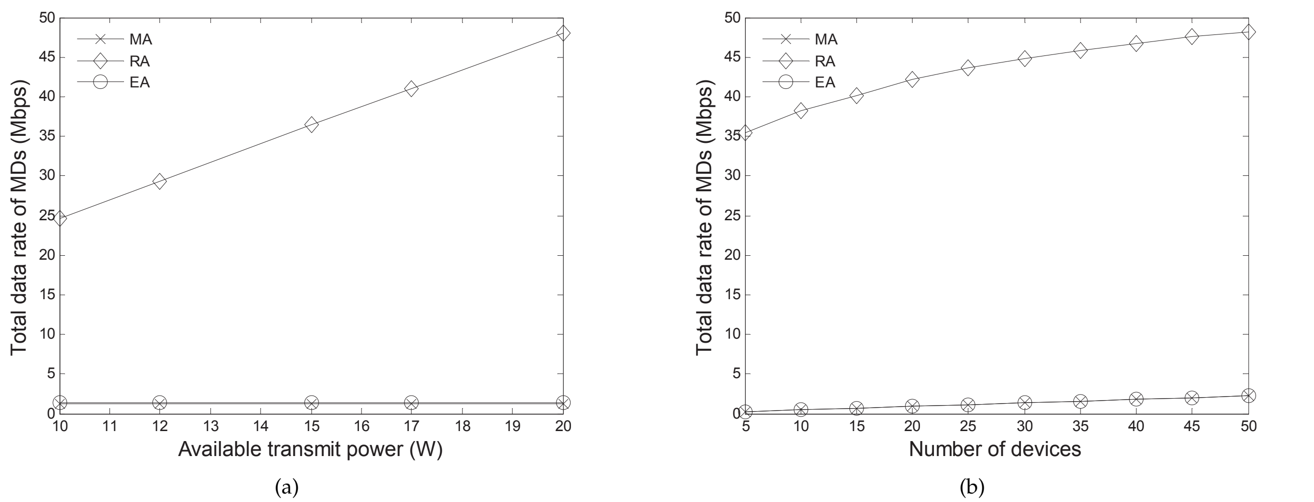


Fig. 7. The respective impacts of the available transmit power and the number of devices on the total data rate of mobile devices.

requirements. In contrast, the number of devices does not have a significant impact on the energy consumption of the base station under *RA*, because *RA* always tries to maximize the data rate regardless of the devices' requirements. The simulation results show that the base station consumes 1.4 to 1.6 times more energy under *EA* than under *MA*. As shown in Fig. 5(c), however, *EA* is still more energy efficient than *MA* by 3.4 times in terms of the total energy consumption of the base station and mobile devices.

5.3 Data Rate and Unallocated RBs

Fig. 6 shows the impacts of the available transmit power and the number of devices on the percentage of unallocated RBs to the total RBs. As shown in Fig. 6(a), the percentage of unallocated RBs under *EA* increases as the available transmit power increases. This is because an RB can be applied with a higher-rate modulation-coding scheme to provide a higher data rate when larger transmit power is allocated to the RB. Consequently, the data rate requirement of each device can be satisfied with fewer RBs. In contrast, the percentage under *EA* decreases as the number of devices increases, as shown in Fig. 6(b). The result can be expected because the base station has to use more RBs to satisfy more devices. When *EA* is adopted, the unallocated RBs can be used to serve other application services. On the other hand, both the number of devices and the available transmit power have no impact on the percentage of unallocated RBs under *RA* and *MA*. This is because *RA* always exhausts the available RBs to maximize the total data rate of mobile devices, while *MA* always exhausts the available RBs to minimize the transmit power of the base station.

Fig. 7 shows the impacts of the available transmit power and the number of devices on the total data rate of mobile devices. Under *RA*, as the available transmit power increases, the total data rate of mobile devices

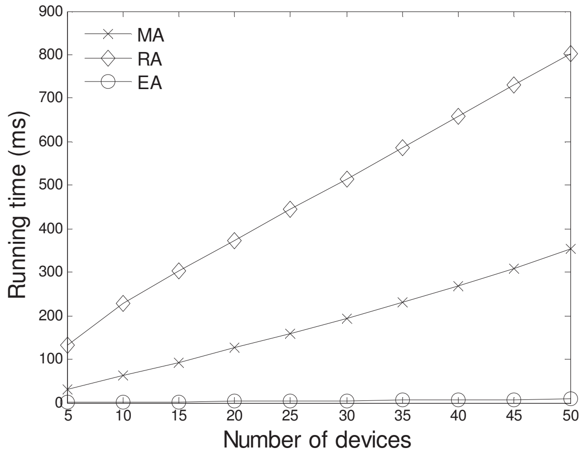


Fig. 8. The average running time required by each algorithm per run.

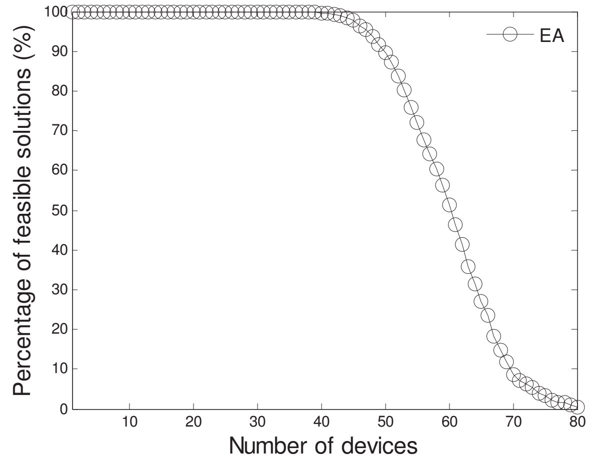


Fig. 9. The percentage of feasible solutions derived by *EA* under the available transmit power of 20 W.

increases, as shown in Fig. 7(a). This is because *RA* attempts to exhaust all the available RBs and transmit power to maximize the total data rate of mobile devices. However, data rate maximization is not beneficial for real-time multimedia data whose bit rate is generally constant. By contrast, the available transmit power does not have any impact on the total data rate under *MA* and *EA*, because they just utilize radio resources, i.e., transmit power and RBs respectively, exactly to satisfy the devices' data rate requirements. As shown in Fig. 7(b), under all the three algorithms, the total data rate increases with the number of devices. The increase is more evident under *RA* because it always utilizes all the available radio resources; moreover, when there are more devices, more diverse channel gains lead to a higher chance that an RB can be allocated to a device with a better channel gain, and, thus, a higher data rate provided by the RB can be achieved. On the other hand, *MA* and *EA* just provide the data rates exactly required by mobile devices. This explains why the total data rate under *MA* and *EA* is similar and increases slightly with the number of devices.

5.4 Running Time and Feasibility

Fig. 8 shows the impact of the number of devices on the average running time required by each of the three algorithms per run. The running times required by *MA* and *RA* increase significantly with the number of devices. In contrast, the number of devices does not have an obvious impact on the average running time required by *EA*. The reason is that *EA* allocates RBs to each device individually, while *MA* and *RA* involve a complex search on the other devices when allocating RBs to each device. The simulation results show that the average running time required by *EA* on our experiment platform for each resource allocation is shorter than 8.2 ms when the number of devices is smaller than 50. Thus, *EA* outperforms *MA* and *RA* in terms of the time complexity and is more applicable to resource allocation that requires timely computation.

Fig. 9 shows the impact of the number of devices on the percentage of feasible solutions derived by *EA*. The simulation results show that the percentage remains over 99.4% when the number of devices is smaller than 42; however, the percentage drops dramatically when the base station has to serve more than 47 devices. The reason for the dramatic drop is that radio resources are fixed and become insufficient for the devices' data rate requirements when the number of devices reaches the critical point. A recent study on unicast video services over LTE networks indicated that the average number of mobile devices that a base station can serve is approximately 10, and up to 42, according to the available bandwidth [1]. The simulation results, in conjunction with the study, could justify that *EA* is applicable to the application scenarios considered in this work. When *EA* returns a solution in which some devices' data rate requirements cannot be satisfied, the base station's admission control mechanism should not admit the devices into the network. The devices could be associated to other base stations nearby (if any) with lighter workloads instead.

6 Concluding Remarks

In this paper, we study energy efficient downlink resource allocation in 4G wireless networks and develop energy adaptive (EA) mechanisms with the objective of minimizing the total energy consumption of mobile devices for multimedia services, such as video streaming, with fixed data rates. With energy adaption in mind, we model the energy-efficient downlink resource allocation problem with two respective constraints to meet mobile devices' data rate requirements while limiting the base station's transmit power. We have shown that the problem is $\mathcal{NP}$-hard and then develop an heuristic algorithm to search for energy-efficient downlink resource allocation with theoretically provable performance guarantee. Through extensive simulations based on real video sequences, we have demonstrated that the algorithm proposed based on EA objective is very effective in reducing the energy consumption of mobile devices, compared with two previously known two algorithms. We have also observed that the efficacy is more evident when a base station serves either a larger number of mobile devices or the devices have higher data rate requirements.

ACKNOWLEDGEMENT

The authors would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions. The work of Ai-Chun Pang was partially supported by Excellent Research Projects of National Taiwan University under Grant 103R890821, by National Science Council under Grant NSC102-2221-E-002-075-MY2 and Grant NSC103-2221-E-002-142-MY3, Research Center for Information Technology Innovation (CITI)/Academia Sinica, Information and Communications Research Laboratories/Industrial Technology Research Institute (ICL/ITRI), and Institute for Information Industry (III). The work of Pi-Cheng Hsiu was partially supported by National Science Council under Grant NSC102-2221-E-001-007-MY2. The work of Yuguang Fang was partially supported by the US National Science Foundation under grant CNS-1343356.

REFERENCES

[1] O. Oyman, J. Foerster, Y.-J. Tcha, and S.-C. Lee, "Toward Enhanced Mobile Video Services over WiMAX and LTE," *IEEE Commun. Magazine*, vol. 48, no. 8, pp. 68–76, Aug. 2010.

[2] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017," 2013. [Online]. Available: http://www.cisco.com/

[3] R. Trestian, A.-N. Moldovan, O. Ormond, and G.-M. Muntean, "Energy Consumption Analysis of Video Streaming to Android Mobile Devices," *IEEE NOMS*, pp. 444–452, 2012.

[4] A. Carroll and C. Heiser, "An Analysis of Power Consumption in a Smartphone," *USENIX Annual Technical Conference*, pp. 1–14, 2010.

[5] G. Perrucci, F. Fitzek, and J. Widmer, "Survey on Energy Consumption Entities on the Smartphone Platform," *IEEE VTC*, pp. 1–6, 2011.

[6] *3GPP TS 36.211 V10.0.0 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation(Release 10)*, Dec. 2010.

[7] *IEEE Std 802.16m, IEEE Standard for Local and metropolitan area networks-Part 16: Air Interface for Broadband Wireless Access Systems*, May 2011.

[8] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems with Proportional Rate Constraints," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[9] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE Journal on Selected Area in Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[10] W. Yu and J. M. Cioffi, "Constant-Power Waterfilling: Performance Bound and Low-Complexity Implementation," *IEEE Trans. on Commun.*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[11] H.-W. Lee and S. Chong, "Downlink Resource Allocation in Multi-Carrier Systems: Frequency-Selective vs. Equal Power Allocation," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 10, pp. 3738–3747, Oct. 2008.

[12] A. N. Zaki and A. O. Fapojuwo, "Optimal and Efficient Graph-Based Resource Allocation Algorithms for Multiservice Frame-Based OFDMA Networks," *IEEE Trans. on Mobile Computing*, vol. 10, no. 8, pp. 1175–1186, Aug. 2011.

[13] A. Biagioni, R. Fantacci, D. Marabissi, and D. Tarchi, "Adaptive Subcarrier Allocation Schemes for Wireless OFDMA Systems in WiMAX Networks," *IEEE Journal on Selected Area in Commun.*, vol. 27, no. 2, pp. 217–225, Feb. 2009.

[14] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE Journal on Selected Area in Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[15] G. Zhang, "Subcarrier and Bit Allocation for Real-Time Services in Multiuser OFDM Systems," *IEEE ICC*, pp. 2985–2989, June 2004.

[16] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal Resource Allocation for OFDMA Downlink Systems," *IEEE ISIT*, pp. 1394–1398, 2006.

[17] A. Feiten, R. Mathar, and M. Reyer, "Rate and Power Allocation for Multiuser OFDM: An Effective Heuristic Verified by Branch-and-Bound," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 1, pp. 60–64, Jan. 2008.

[18] Z. Mao and X. Wang, "Efficient Optimal and Suboptimal Radio Resource Allocation in OFDMA System," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 2, pp. 440–445, Feb. 2008.

[19] Y.-B. Lin, T.-H. Chiu, and Y. T. Su, "Optimal and Near-Optimal Resource Allocation Algorithms for OFDMA Networks," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 8, pp. 4066–4077, Aug. 2009.

[20] G. Auer et al., "How Much Energy is Needed to Run a Wireless Network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[21] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems," *IEEE Commun. Surveys & Tutorials*, vol. 11, no. 3, pp. 92–106, 3rd Quarter 2009.

[22] M. Siekkinen, M. A. Hoque, J. K. Nurminen, and M. Aalto, "Streaming over 3G and LTE: How to Save Smartphone Energy in Radio AccessNetwork-Friendly Way," *ACM MoVid*, pp. 13–18, 2013.

[23] M. A. Hoque, M. Siekkinen, J. K. Nurminen, S. Tarkoma, and M. Aalto, "Saving Energy in Mobile Devices for On-Demand Multimedia Streaming - A Cross-Layer Approach," *ACM Trans. on Multimedia Computing, Communications and Applications*, vol. 10, no. 3, p. Article 25, 2014.

[24] M. Hefeeda and C.-H. Hsu, "On Burst Transmission Scheduling in Mobile TV Broadcast Networks," *IEEE/ACM Trans. on Networking*, vol. 18, no. 2, pp. 610–623, 2010.

[25] J. Y. Kim and D. H. Cho, "Resource Allocation Scheme for Minimizing Power Consumption in OFDMA Cellular Systems," *IEEE VTC*, pp. 1862–1866, 2007.

[26] F.-S. Chu, K.-C. Chen, and G. Fettweis, "Green Resource Allocation to Minimize Receiving Energy in OFDMA Cellular Systems," *IEEE Commun. Letters*, vol. 16, no. 3, pp. 372–374, March 2012.

[27] Y.-J. Yu, P.-C. Hsiu, and A.-C. Pang, "Energy-Efficient Video Multicast in 4G Wireless Systems," *IEEE Trans. on Mobile Computing*, vol. 11, no. 10, pp. 1508–1522, Oct. 2012.

[28] *3GPP TS 36.213 V11.4.0 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 11)*, Sep. 2013.

[29] C.-C. Cheng and P.-C. Hsiu, "Extend Your Journey: Introducing Signal Strength into Location-based Applications," *IEEE INFOCOM*, pp. 2742–2750, 2013.

[30] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Constrained Modulation Optimization," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[31] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," *ACM MobiSys*, pp. 225–238, 2012.

[32] *3GPP TS 23.203 V12.4.0 Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 12)*, March 2014.

[33] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness*, 1st ed. New York: W. H. Freeman Co., Jan. 1979.

[34] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, Oct. 2002.

[35] N. Ksairi, P. Bianchi, and P. Ciblat, "Resource Allocation for Downlink Cellular OFDMA Systems-Part I: Optimal Allocation," *IEEE Trans. on Signal Processing*, vol. 58, no. 2, pp. 720–734, Feb. 2010.

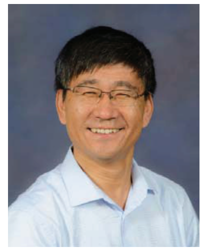
[36] Video sequences. [Online]. Available: http://trace.eas.asu.edu/yuv/.

Ya-Ju Yu received the BS degree in computer and communication engineering from National Kaohsiung First University of Science and Technology in 2005, the MS degree in communication engineering from National Central University in 2007, and Ph.D. degree in the Graduate Institute of Networking and Multimedia from National Taiwan University in 2012. He is currently a senior engineer in the Smart Network System Institute, Institute for Information Industry (III), Taiwan. His research interests include wireless mobile networks, multimedia communications, and cloud datacenter networking. He is a Member of IEEE.

Ai-Chun Pang received the B.S., M.S. and Ph.D. degrees in Computer Science and Information Engineering from National Chiao Tung University, Taiwan, in 1996, 1998 and 2002, respectively. She joined the Department of Computer Science and Information Engineering (CSIE), National Taiwan University (NTU), Taipei, Taiwan, as an Assistant Professor in 2002. She is now the Director of Graduate Institute of Networking and Multimedia (INM), and a Professor in CSIE and INM. She is also an Adjunct Research Fellow of Research Center for Information Technology Innovation, Academia Sinica, Taiwan. Her research interests include the design and analysis of wireless and multimedia networking, mobile communications, and cloud datacenter networking.

Dr. Pang was a Guest Editor of IEEE Wireless Communications, and is currently an Associate Editor of Wireless Networks, International Journal of Sensor Networks, and Security and Communication Networks. She served on the technical program committee of many international conferences, including IEEE INFOCOM, IEEE GLOBECOM, IEEE ICC, and IEEE VTC. She is a co-author of the book Wireless and Mobile All-IP Networks published by John Wiley & Sons Inc. Dr. Pang was a recipient of Outstanding Teaching Award at NTU in 2010, the Investigative Research Award of Pan Wen Yuan Foundation in 2006, Wu Ta You Memorial Award of National Science Council (NSC) in 2007, Excellent Young Engineer Award of Chinese Institute of Electrical Engineering in 2007, and K. T. Li Award for Young Researchers of ACM Taipei/Taiwan Chapter in 2007. Dr. Pang also receives the Republic of China Distinguished Women Medal in 2009. She is a Senior Member of IEEE.

Pi-Cheng Hsiu received the BS degree in computer information science from National Chiao Tung University in 2002 and the MS and PhD degrees in computer science and information engineering from National Taiwan University in 2004 and 2009, respectively. He is currently an associate research fellow in the Research Center for Information Technology Innovation (CITI), where he leads the Embedded and Mobile Computing Laboratory (EMCLab), and also a joint associate research fellow in the Institute of Information Science (IIS), Academia Sinica, Taiwan. He was a visiting scholar in the Department of Computer Science at the University of Illinois at Urbana-Champaign in 2007 and joined CITI as an assistant research fellow in 2009. His research interests include embedded systems, real-time systems, and mobile networks. He is a member of the IEEE and the ACM.



Yuguang "Michael" Fang (F'08) received a BS/MS degree from Qufu Normal University, Shandong, China in 1987, a Ph.D degree from Case Western Reserve University in 1994 and a Ph.D. degree from Boston University in 1997. He joined the Department of Electrical and Computer Engineering at University of Florida in 2000 and has been a full professor since 2005. He holds a University of Florida Research Foundation (UFRF) Professorship from 2006 to 2009, a Changjiang Scholar Chair Professorship with Xidian University, China, from 2008 to 2011, and a Guest Chair Professorship with Tsinghua University, China, from 2009 to 2012.

Dr. Fang received the US National Science Foundation Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002, and is the recipient of the Best Paper Award from IEEE ICNP (2006). He has also received a 2010-2011 UF Doctoral Dissertation Advisor/Mentoring Award, 2011 Florida Blue Key/UF Homecoming Distinguished Faculty Award, and the 2009 UF College of Engineering Faculty Mentoring Award. He is the Editor-in-Chief of IEEE Transactions on Vehicular Technology, was the Editor-in-Chief of IEEE Wireless Communications (2009-2012) and serves/served on several editorial boards of journals including IEEE Transactions on Mobile Computing (2003-2008, 2011-present), IEEE Transactions on Communications (2000-2011), and IEEE Transactions on Wireless Communications (2002-2009). He has been actively participating in conference organizations such as serving as the Technical Program Co-Chair for IEEE INOFOCOM'2014 and the Technical Program Vice-Chair for IEEE INFOCOM'2005. He is a fellow of the IEEE.