Energy-Efficient Power Allocation for Pilots in Training-Based Downlink OFDMA Systems

Zhikun Xu, *Student Member, IEEE*, Geoffrey Ye Li, *Fellow, IEEE*, Chenyang Yang, *Senior Member, IEEE*, Shunqing Zhang, Yan Chen, and Shugong Xu, *Senior Member, IEEE*

***Abstract*—In this paper, power allocation between pilots and data symbols is investigated to maximize *energy efficiency* (EE) for downlink *orthogonal frequency division multiple access* (OFDMA) networks. We first derive an EE function considering channel estimation error, which depends on large-scale channel gains of multiple users, allocated power to pilots and data symbols, and circuit power consumption. Then an optimization problem is formulated to maximize the EE under overall transmit power constraint. Exploiting the quasiconcavity property of the EE function, we propose an alternating optimization method in the low transmit power region and reformulate a joint quasiconcave problem in the high transmit power region. Analysis and simulation results show that the power ratio for pilots decreases with the circuit power. When the circuit power is small, the optimal overall transmit power increases with the circuit power. Otherwise, the optimal transmit power does not depend on it. Transmitting more data symbols to the users with higher channel gains improves the EE but at a cost of sacrificing the fairness among multiple users. Simulation results also demonstrate that compared with *spectral efficiency* (SE)-oriented design, the EE-oriented design can improve the EE performance significantly with a relatively small SE loss.**

***Index Terms*—Energy efficiency (EE), power allocation, pilot, orthogonal frequency division multiple access (OFDMA).**

I. INTRODUCTION

MOTIVATED by protection of natural environment and the scarcity of energy resources, efficient energy usage has been widely advocated. Although wireless communications only take up a small portion of today's total carbon footprint among the information and communication technologies [1], [2], explosive growth of high-quality wireless services indicates that it will play a more and more important role in the future. Consequently, design of energy-efficient wireless communication systems becomes an urgent task.

Paper approved by O. Oyman, the Editor for Cooperative and Heterogeneous Networks of the IEEE Communications Society. Manuscript received April 21, 2011; revised January 26 and March 29, 2012.

Z. Xu and C. Yang are with the School of Electronics and Information Engineering, Beihang University, Beijing, 100191, P. R. China (e-mail: xuzhikun@ee.buaa.edu.cn; cyyang@buaa.edu.cn).

G. Y. Li is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: liye@ece.gatech.edu).

S. Zhang, Y. Chen, and S. Xu are with the GREAT (Green Radio Excellence in Architectures and Technologies) Research Team of Huawei Technologies Co., Ltd., Shanghai 201206, P. R. China (e-mail: {sqzhang, bigbird.chenyan, shugong}@huawei.com).

The work is supported in part by a research gift from Huawei Technologies Co., Ltd., the National Basic Research Program of China, 973 Program 2012CB316000, the National Natural Science Foundation of China (NSFC) under Grant 61120106002, and the Innovation Foundation of BUAA for Ph.D. Graduates.

Digital Object Identifier 10.1109/TCOMM.2012.071812.110242

Energy efficiency (EE) is first studied from information-theoretic point of view. It is shown in [3] that when only transmission energy consumption is considered, the maximum EE is achieved when *spectral efficiency* (SE) approaches zero. When circuit energy consumption in practical systems is taken into account, the relationship between EE and SE completely changes, which depends on network architectures, transmission schemes, and resource allocation strategies, etc. [1]. Recently, various issues have been addressed to improve the EE of wireless systems. For example, a tradeoff among transmission energy, circuit energy, and transmission time is investigated for different modulation schemes in [4]. Energy-efficient power adaptation in frequency-selective channels is addressed in [5]. Adaptive switching between *multiple-input multiple-output* (MIMO) and *single-input multiple-output* (SIMO) is studied in [6] to save energy for mobile terminals.

In wireless communication systems, pilots are usually inserted into data streams to facilitate channel estimation for coherent detection. Pilot design has been studied from different aspects under various criteria. The number, positions, and power of pilot symbols are designed to maximize the capacity lower bound in [7] and to minimize the Cramer-Rao bound of channel estimation error in [8]. It is shown that the optimal placement of pilot symbols is periodical insertion in frequency domain. Power allocation for pilots is discussed for MIMO systems in [9] and for doubly selective fading channels in [10], respectively. A comprehensive overview of pilot-assisted transmission is provided in [11], from both information theory and signal processing point of view.

In contrast to the flourishing on pilot design from the perspective of channel capacity and channel estimation error, little attention is paid to the EE. Pilot design that maximizes capacity, i.e., maximizes the SE, does not necessarily maximize the EE. Moreover, in EE-oriented pilot resource allocation, the overall transmit power needs to be optimized besides the power ratio between pilots and data symbols. The energy-efficient pilot design is first studied in [12] for single user case. By assuming the interference incurred by channel estimation error as Gaussian, the overall transmit power is optimized to maximize the EE. It is shown that the maximum EE is achieved at a certain nonzero overall transmit power, which is totally different from the relationship between the EE and transmit power when perfect channel information is known at the receiver. Power allocation between pilots and data symbols in multi-user case is more complicated since different users suffer from different channel fadings,

TABLE I
NOTATION LIST

$(\mathbf{A})_{ij}$	the (i,j)th element of matrix $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of matrix $\mathbf{A}$
$\mathbf{A}^T$	transpose of matrix$\mathbf{A}$
$\mathbf{A}^H$	Hermitian of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of matrix $\mathbf{A}$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbb{E}_{\mathbf{x}}[\cdot]$	expectation operation over $\mathbf{x}$
$\partial f/\partial x$	partial derivative of function f with respect to x
∇f	gradient of function f
$\nabla^2 f$	Hessian matrix of function f
$\text{diag}\{x_1, \cdots, x_N\}$	$N \times N$ diagonal matrix with diagonal elements $x_1, \cdots, x_N$
$\text{diag}\{(x_i)_{i\in\mathcal{S}}\}$	diagonal matrix with diagonal elements $x_i,\ i \in \mathcal{S}$
$\max\{x_1, \cdots, x_N\}$	the maximum value of $\{x_n\}_{n=1}^N$

which results in different requirements on the power for pilot transmission. On the other hand, all users may share the common pilots to estimate channels, e.g., in downlink 3GPP LTE systems [13], which causes that the method in single user case cannot be used any more. How to allocate power between pilots and data symbols to maximize the EE in multi-user case is still unknown.

In this paper, we will investigate energy-efficient power allocation between pilots and data symbols in training-based downlink *orthogonal frequency division multiple access* (OFDMA) systems. We consider *minimum mean-square error* (MMSE) channel estimation. Based on the correlation matrix of the channel estimation error, we will first derive the ergodic capacity of each user and the EE function considering the circuit power consumption. An optimization problem is then formulated to maximize the EE with the overall transmit power constraint. A method based on the quasiconcavity of the EE function is developed to solve the problem. The impacts of the circuit power and multiple users are analyzed. Simulation results will show a substantial EE gain of the EE-oriented design over the SE-oriented design.

Notations used in this paper are summarized in Table I.

II. SYSTEM MODEL

Consider a downlink OFDMA cellular network with M users and N_{tot} subcarriers. K pilot symbols are periodically placed in the frequency domain and shared by different users for channel estimation [7]. The index set of subcarriers for pilots is denoted as $\mathcal{S}_p$. Except K subcarriers for pilot transmission, the remaining $N_{tot}-K$ subcarriers can be assigned to different users for data transmission without overlapping. We denote $\mathcal{S}_m$ as the index set of subcarriers occupied by user m with size N_m. For example, for the system in Fig. 1, $M=3$, $N_{tot}=8$, $\mathcal{S}_p=\{1,3,5,7\}$, $\mathcal{S}_1=\{0,4\}$, $\mathcal{S}_2=\{2\}$, and $\mathcal{S}_3=\{6\}$.

We optimize the power allocation based on statistical channel information of multiple users. Since the frequency-

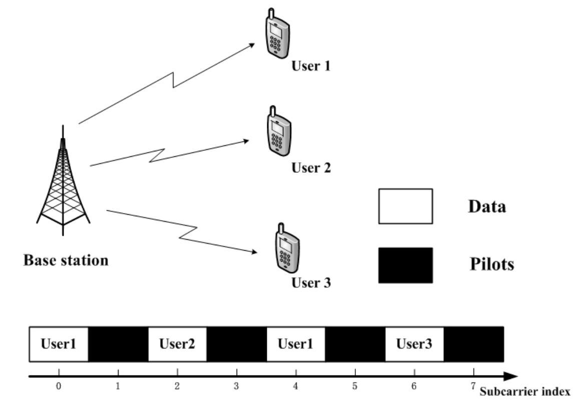


Fig. 1. Pilot structure of an OFDMA system.

domain channel coefficients have the same distribution over all subcarriers occupied by each user, the transmit power for pilots and data symbols of each user is uniformly allocated in frequency domain. Denote α and β_m as the overall transmit power for pilot and data symbols of user m, respectively. Then the powers per subcarrier for pilot symbols and data symbols are α/K and β_m/N_m, respectively. The noise at the receiver of each user is assumed to be additive white Gaussian with zero mean and variance σ^2.

All users are assumed to undergo frequency-selective block fading channels with L resolvable paths whose values are subject to Gaussian distribution. The channel impulse response vector from the *base station* (BS) to user m, $\mathbf{h}_m^t$, consists of L elements which are independent random variables with zero mean and variance $\{\gamma_{m,l}\}_{l=1}^L$. Hence, its correlation matrix is

$$\mathbf{R}_{\mathbf{h}_m^t} = \mathbb{E}_{\mathbf{h}_m^t}\left[\mathbf{h}_m^t \mathbf{h}_m^{t,H}\right] = \text{diag}\{\gamma_{m,1}, \cdots, \gamma_{m,L}\}. \quad (1)$$

Denote the channel coefficient of user m at the ith subcarrier as $h_{m,i}^f$. Then from Parseval's theorem, its variance can be obtained as follows,

$$\mathbb{E}_{h_{m,i}^f}\left[|h_{m,i}^f|^2\right] = N_{tot}\mathbf{f}_i \mathbf{R}_{\mathbf{h}_m^t}\mathbf{f}_i^H = \sum_{l=1}^L \gamma_{m,l}, \quad (2)$$

where $\mathbf{f}_i$ denotes the ith row of a truncated FFT matrix $\mathbf{F}_L = \left[\frac{1}{\sqrt{N_{tot}}}e^{-j2\pi nl/N_{tot}}\right]_{n,l=0}^{N_{tot}-1,L-1}$.

The frequency-domain pilot vector received by user m is

$$\begin{aligned}\mathbf{y}_{p,m} &= \mathbf{T}_p\mathbf{h}_{p,m}^f + \mathbf{n}_{p,m} = \sqrt{N_{tot}}\mathbf{T}_p\mathbf{\Psi}_p\mathbf{F}_L\mathbf{h}_m^t + \mathbf{n}_{p,m} \\ &= \mathbf{\Phi}_m\mathbf{h}_m^t + \mathbf{n}_{p,m}, \end{aligned} \quad (3)$$

where $\mathbf{h}_{p,m}^f \triangleq \sqrt{N_{tot}}\mathbf{\Psi}_p\mathbf{F}_L\mathbf{h}_m^t$ is a vector consisting of the frequency-domain channel coefficients of user m at the subcarriers occupied by pilot symbols, $\mathbf{\Psi}_p$ is a $K \times N_{tot}$ selection matrix with one at the positions given by $\mathcal{S}_p$, $\mathbf{T}_p$ denotes a $K \times K$ diagonal matrix consisting of pilot symbols $\{s_k\}_{k=1}^K$, $\mathbf{\Phi}_m \triangleq \sqrt{N_{tot}}\mathbf{T}_p\mathbf{\Psi}_p\mathbf{F}_L$, and $\mathbf{n}_{p,m}$ is the noise vector.

The frequency-domain data vector received by user m can be expressed as

$$\mathbf{y}_{d,m} = \mathbf{H}_m^f\mathbf{d}_m + \mathbf{n}_{d,m} = \hat{\mathbf{H}}_m^f\mathbf{d}_m + \underbrace{\Delta\mathbf{H}_m^f\mathbf{d}_m + \mathbf{n}_{d,m}}_{\mathbf{v}_m}, \quad (4)$$

where $\mathbf{d}_m$ is the data vector, $\mathbf{n}_{d,m}$ is the noise vector, $\mathbf{H}_m^f \triangleq \text{diag}\{(h_{m,i}^f)_{i\in\mathcal{S}_m}\}$ is an $N_m \times N_m$ diagonal matrix, $\hat{\mathbf{H}}_m^f$ and $\Delta\mathbf{H}_m^f$ denote the estimation and estimation error of $\mathbf{H}_m^f$, respectively, and $\mathbf{v}_m \triangleq \Delta\mathbf{H}_m^f\mathbf{d}_m + \mathbf{n}_{d,m}$ denotes the total signal distortion.

III. Problem Formulation

In this section, we will first investigate the performance of MMSE channel estimator. Then ergodic channel capacity of each user is derived for downlink transmission and an optimization problem is formulated to maximize the EE under overall transmit power constraint.

A. Channel Estimation

Based on MMSE criterion [14, Ch.8], we can readily obtain the estimate of $\mathbf{h}_m^t$ from (3) as

$$\hat{\mathbf{h}}_m^t = \mathbf{R}_{\mathbf{h}_m^t}\mathbf{\Phi}_m\left(\mathbf{\Phi}_m\mathbf{R}_{\mathbf{h}_m^t}\mathbf{\Phi}_m^H + \sigma^2\mathbf{I}_K\right)^{-1}\mathbf{y}_{p,m},$$

where we have used the fact that noise is white and with zero mean and the same variance over all subcarriers. The correlation matrix of estimation errors can be expressed as

$$\begin{aligned}\mathbf{R}_{\Delta\mathbf{h}_m^t} &= \mathbb{E}_{\Delta\mathbf{h}_m^t}\left[(\mathbf{h}_m^t - \hat{\mathbf{h}}_m^t)(\mathbf{h}_m^t - \hat{\mathbf{h}}_m^t)^H\right] \\ &= \left(\mathbf{R}_{\mathbf{h}_m^t}^{-1} + \frac{1}{\sigma^2}\mathbf{\Phi}_m^H\mathbf{\Phi}_m\right)^{-1}. \end{aligned} \tag{5}$$

From Appendix A, we can obtain

$$\mathbf{\Phi}_m^H\mathbf{\Phi}_m = \alpha\mathbf{I}_L. \tag{6}$$

After substituting (1) and (6), (5) becomes

$$\mathbf{R}_{\Delta\mathbf{h}_m^t} = \text{diag}\left\{\frac{1}{\frac{1}{\gamma_{m,1}} + \frac{\alpha}{\sigma^2}}, \cdots, \frac{1}{\frac{1}{\gamma_{m,L}} + \frac{\alpha}{\sigma^2}}\right\}. \tag{7}$$

Denote $\Delta h_{m,i}^f$ and $\hat{h}_{m,i}^f$ as the estimation error and estimate of $h_{m,i}^f$, respectively. From (2) and (7), the variances of $\Delta h_{m,i}^f$, and $\hat{h}_{m,i}^f$ can be derived as

$$\mathbb{E}_{\Delta h_{m,i}^f}\left[|\Delta h_{m,i}^f|^2\right] = N_{tot}\mathbf{f}_i^H\mathbf{R}_{\Delta\mathbf{h}_m^t}\mathbf{f}_i = \sum_{l=1}^{L}\frac{\gamma_{m,l}\sigma^2}{\gamma_{m,l}\alpha + \sigma^2}, \tag{8}$$

and

$$\begin{aligned}\mathbb{E}_{\hat{h}_{m,i}^f}\left[|\hat{h}_{m,i}^f|^2\right] &= \mathbb{E}_{h_{m,i}^f}\left[|h_{m,i}^f|^2\right] - \mathbb{E}_{\Delta h_{m,i}^f}\left[|\Delta h_{m,i}^f|^2\right] \\ &= \sum_{l=1}^{L}\frac{\gamma_{m,l}^2\alpha}{\gamma_{m,l}\alpha + \sigma^2}, \end{aligned} \tag{9}$$

respectively.

B. Ergodic Capacity

If we assume that the elements of data vector $\mathbf{d}_m$ are independent and with Gaussian distribution and treat the term $\Delta\mathbf{H}_m^f\mathbf{d}_m$ in (4) as Gaussian noise as in [7], the ergodic channel capacity of user m can be written as

$$C_m = \Delta f\mathbb{E}_{\hat{\mathbf{H}}_m^f}\left[\log_2\det(\mathbf{I}_{N_m} + \mathbf{R}_{\mathbf{v}_m}^{-1}\hat{\mathbf{H}}_m^f\mathbf{R}_{\mathbf{d}_m}\hat{\mathbf{H}}_m^{f,H})\right], \tag{10}$$

where Δf is the subcarrier spacing, $\mathbf{R}_{\mathbf{d}_m}$ and $\mathbf{R}_{\mathbf{v}_m}$ are the correlation matrices of the data vector and the total signal distortion vector, respectively.

Since the transmit power for user m is equally allocated to its occupied subcarriers, the correlation matrix of the data vector can be expressed as

$$\mathbf{R}_{\mathbf{d}_m} = \frac{\beta_m}{N_m}\mathbf{I}_{N_m}. \tag{11}$$

Since data and noise are independent, the correlation matrix of the signal distortion vector in (4) becomes

$$\begin{aligned}\mathbf{R}_{\mathbf{v}_m} &\triangleq \mathbb{E}_{\mathbf{v}_m}\left[\mathbf{v}_m\mathbf{v}_m^H\right] \\ &= \mathbb{E}_{\Delta\mathbf{H}_m^f,\mathbf{d}_m,\mathbf{n}_{d,m}}\left[(\Delta\mathbf{H}_m^f\mathbf{d}_m + \mathbf{n}_{d,m})(\Delta\mathbf{H}_m^f\mathbf{d}_m + \mathbf{n}_{d,m})^H\right] \\ &= \mathbb{E}_{\Delta\mathbf{H}_m^f,\mathbf{d}_m}\left[\Delta\mathbf{H}_m^f\mathbf{d}_m\mathbf{d}_m^H\Delta\mathbf{H}_m^{f,H}\right] + \mathbb{E}_{\mathbf{n}_{d,m}}\left[\mathbf{n}_{d,m}\mathbf{n}_{d,m}^H\right] \\ &= \mathbb{E}_{\Delta\mathbf{H}_m^f}\left[\Delta\mathbf{H}_m^f\mathbf{R}_{\mathbf{d}_m}\Delta\mathbf{H}_m^{f,H}\right] + \sigma^2\mathbf{I}_{N_m}. \end{aligned} \tag{12}$$

Substituting (8) and (11) into (12), we have

$$\begin{aligned}\mathbf{R}_{\mathbf{v}_m} &= \text{diag}\left\{\left(\frac{\beta_m}{N_m}\mathbb{E}_{\Delta h_{m,i}^f}\left[|\Delta h_{m,i}^f|^2\right] + \sigma^2\right)_{i\in\mathcal{S}_m}\right\} \\ &= \left(\frac{\beta_m}{N_m}\sum_{l=1}^{L}\frac{\gamma_{m,l}\sigma^2}{\gamma_{m,l}\alpha + \sigma^2} + \sigma^2\right)\mathbf{I}_{N_m}. \end{aligned} \tag{13}$$

To facilitate the analysis of the impact of power allocation, we re-express the channel estimate as a product of its standard deviation and a normalized random variable as follows,

$$\hat{h}_{m,i}^f = \sqrt{\mathbb{E}_{\hat{h}_{m,i}^f}\left[|\hat{h}_{m,i}^f|^2\right]}\,g = \sqrt{\sum_{l=1}^{L}\frac{\gamma_{m,l}^2\alpha}{\gamma_{m,l}\alpha + \sigma^2}}\,g, \tag{14}$$

where (9) is used. Because $\hat{h}_{m,i}^f$ is subject to Gaussian distribution when linear MMSE channel estimation is applied [12], g is a complex Gaussian random variable with zero mean and unit variance.

Substituting (11) and (13) into (10) and considering (14), we have

$$\begin{aligned}C_m &= \Delta f\mathbb{E}_{\hat{h}_{m,i}^f}\left[\sum_{i=1}^{N_m}\log_2\left(1 + \frac{\frac{\beta_m}{N_m}|\hat{h}_{m,i}^f|^2}{\frac{\beta_m}{N_m}\sum_{l=1}^{L}\frac{\gamma_{m,l}\sigma^2}{\gamma_{m,l}\alpha+\sigma^2} + \sigma^2}\right)\right] \\ &= \Delta fN_m\mathbb{E}_g\left[\log_2\left(1 + \frac{\beta_m\sum_{l=1}^{L}\frac{\gamma_{m,l}^2\alpha}{\gamma_{m,l}\alpha+\sigma^2}|g|^2}{(N_m + \beta_m\sum_{l=1}^{L}\frac{\gamma_{m,l}}{\gamma_{m,l}\alpha+\sigma^2})\sigma^2}\right)\right] \\ &\triangleq \Delta fN_m\mathbb{E}_g\left[\log_2(1 + f(\alpha,\beta_m)|g|^2)\right], \end{aligned} \tag{15}$$

where $f(\alpha,\beta_m) \triangleq \frac{\beta_m\sum_{l=1}^{L}\frac{\gamma_{m,l}^2\alpha}{\gamma_{m,l}\alpha+\sigma^2}}{\left(N_m+\beta_m\sum_{l=1}^{L}\frac{\gamma_{m,l}}{\gamma_{m,l}\alpha+\sigma^2}\right)\sigma^2}$ is used.

C. Energy Efficiency Optimization

The overall transmit power in training-based downlink systems consists of the power for pilot transmission and that for data transmission and can be expressed as

$$P_t = \alpha + \sum_{m=1}^{M} \beta_m. \tag{16}$$

Except for transmit power, additional circuit power is consumed at the BS for transmitting and receiving, such as filtering, *digital-to-analog* (D/A) or *analog-to-digital* (A/D) converting, and signal processing. The overall consumed power at the BS is given by [4],

$$P_{tot} = \rho P_t + P_c = \rho \left(\alpha + \sum_{m=1}^{M} \beta_m \right) + P_c. \tag{17}$$

where P_c represents the circuit power consumption and $1/\rho$ denotes the power amplifier efficiency which is defined as the ratio of the output power to input power of a power amplifier.

The EE in dowlink transmission is defined as the overall number of bits transmitted from the BS per unit energy [5] and is equivalent to the sum capacity of multiple users per unit power. From (15) and (17), we can obtain the EE as

$$\begin{aligned} \eta &= \frac{\sum_{m=1}^{M} C_m}{P_{tot}} \\ &= \frac{\Delta f \sum_{m=1}^{M} N_m \mathbb{E}_{|g|^2} \left[\log_2(1 + f(\alpha, \beta_m)|g|^2) \right]}{\rho \left(\alpha + \sum_{m=1}^{M} \beta_m \right) + P_c}. \end{aligned} \tag{18}$$

Since the target application of our work is for best-effort services pursuing high EE, we do not consider minimum data rate requirements of multiple users. Then the optimization problem of power allocation to maximize the EE under transmit power constraint is finally formulated as follows,

$$\max_{\alpha, \{\beta_m\}_{m=1}^{M}} \frac{\Delta f \sum_{m=1}^{M} N_m \mathbb{E}_{|g|^2} \left[\log_2(1 + f(\alpha, \beta_m)|g|^2) \right]}{\rho(\alpha + \sum_{m=1}^{M} \beta_m) + P_c} \tag{19}$$

$$\text{s. t.} \quad \alpha + \sum_{m=1}^{M} \beta_m \leq P_{max}, \tag{19a}$$

$$\alpha \geq 0, \quad \beta_m \geq 0, \quad m = 1, \cdots, M, \tag{19b}$$

where P_{max} is the maximum transmit power.

IV. Solution of Power Allocation to Maximize EE

In this section, we solve problem (19) considering the quasiconcavity of the EE function over the powers for pilots and data symbols. To exploit this property of the EE function, we consider two cases that the optimal overall transmit power is less than or equal to the maximum value, P_{max}, respectively. When the optimal overall transmit power is less than P_{max}, we solve problem (19) by using the alternating optimization method. Otherwise, we first approximate the objective function of problem (19) and then apply existing quasiconcave optimization algorithm to find its solution.

A. Case 1: The Optimal Overall Transmit Power is Less than P_{max}

According to the optimization theory[15], when the optimal overall transmit power is less than P_{max}, constraint $(19a)$ will not affect the optimum of problem (19) and thus can be removed from the optimization problem. In the following, we will solve an optimization problem with the objective function of (19) and constraint $(19b)$ but without constraint $(19a)$, which is referred to as problem A to facilitate the description.

Before introducing the method to solve problem A, we first show some properties of function $f(\alpha, \beta_m)$. We prove in Appendix B that $f(\alpha, \beta_m)$ is a concave function of α or β_m, but it is not a jointly concave function of α and β_m. According to the composition rule of concave functions, we know from (15) that the channel capacity of user m is a concave function of α or β_m. Then the sum channel capacity over multiple users is concave with respect to α or $\{\beta_m\}_{m=1}^{M}$. Therefore, the EE function in (18) is a concave function over a linear function. Similar to the proof in [5], we can prove that the EE function is quasiconcave with respect to α or $\{\beta_m\}_{m=1}^{M}$. However, since $f(\alpha, \beta_m)$ is not jointly concave with respect to α and β_m, the EE function cannot be proved as a jointly quasiconcave function of α and $\{\beta_m\}_{m=1}^{M}$. In the following, we propose an efficient method based on the quasiconcavity of the EE function over α or $\{\beta_m\}_{m=1}^{M}$.

Since the objective function of problem (19) is quasiconcave and constraint $(19b)$ composes a convex set, problem A is a quasiconcave problem with respect to α or $\{\beta_m\}_{m=1}^{M}$ and the solution of problem A can be found by optimizing α and $\{\beta_m\}_{m=1}^{M}$ alternately. The iteration procedure will be stopped when the EE improvement is less than a predetermined small value. This alternating optimization method has been widely applied in MIMO and relay transmission systems[16], [17] and has been demonstrated to achieve a fairly good performance.

B. Case 2: The Optimal Overall Transmit Power is Equal to P_{max}

When the optimal overall transmit power is equal to the maximum value, i.e. $\alpha + \sum_{m=1}^{M} \beta_m = P_{max}$, we know from [18] that the alternating optimization method in previous subsection may get stuck at an unexpected point with poor performance and is no longer effective to find the solution for problem (19). Therefore, we need to develop a new method in this case.

In cellular networks, the maximum transmit power should guarantee that all users in the entire cell region can receive the pilots with high SNR. Therefore, when the BS transmits pilots and data symbols with the maximum transmit power, P_{max}, it is reasonable to assume that the power of received pilot symbols is much higher than the noise power, i.e., $\gamma_{m,i}\alpha \gg \sigma^2$[19], [20]. Then we can approximate $f(\alpha, \beta_m)$ as follows,

$$f(\alpha, \beta_m) \triangleq \frac{\beta_m \sum_{l=1}^{L} \frac{\gamma_{m,l}^2 \alpha}{\gamma_{m,l}\alpha + \sigma^2}}{\left(N_m + \beta_m \sum_{l=1}^{L} \frac{\gamma_{m,l}}{\gamma_{m,l}\alpha + \sigma^2} \right) \sigma^2}$$

TABLE II
PROCEDURE FOR SOLVING PROBLEM (19)

1.	Solve problem A by the alternating optimization method and obtain the overall transmit power, P_t^1.
2.	**if** $P_t^1 < P_{max}$,
3.	P_t^1 is the optimal overall transmit power for problem (19);
4.	**else**
5.	Substitute (20) into (18) and solve problem (19) when the equality in constraint (19a)
6.	holds by existing quasiconcave optimization algorithms;
7.	**end**

$$\approx \frac{\alpha\beta_m \sum_{l=1}^{L} \gamma_{m,l}}{(N_m\alpha + L\beta_m)\sigma^2} \triangleq \tilde{f}(\alpha, \beta_m), \tag{20}$$

where we define $\tilde{f}(\alpha, \beta_m)$ as the approximation of $f(\alpha, \beta_m)$. After substituting this approximation into (15) and (18), we can obtain the approximate ergodic capacity of user m and the approximate EE, respectively.

With the help of this approximation, we prove in Appendix C that the approximate EE function is jointly quasiconcave with respect to α and $\{\beta_m\}_{m=1}^M$. Considering that constraints (19a) and (19b) are linear, we know that problem (19) is a quasiconcave optimization problem of α and $\{\beta_m\}_{m=1}^M$.

Consequently, problem (19) is converted into quasiconcave problems in both cases, which can be solved by existing efficient algorithms [15], [21]–[23]. Besides finding the solution of problem (19) in the two cases, we also need to judge whether the optimal overall transmit power for problem (19) equals the maximum value, P_{max}. This judgement can be implemented by solving problem A in Section IV-A. If the solution of problem A is less than P_{max}, the optimal overall transmit power for problem (19) is less than P_{max}. Otherwise, the optimal overall transmit power is equal to P_{max}. The entire procedure to find the solution of problem (19) is summarized in Table II.

V. IMPACT OF CIRCUIT POWER AND MULTIPLE USERS

In practical communication systems, besides the transmit power, the circuit power takes up a large proportion of the overall power consumption at the BS[24], [25]. For OFDMA transmission, different users suffer from different channel fading and occupy different numbers of subcarriers, which makes the EE analysis more complicated than that in the single user case. In this section, we study the impact of circuit power and multiple users on the EE-oriented pilot power optimization of OFDMA systems.

A. Impact of Circuit Power

In cellular networks, the transmit power only occupies a small fraction of the entire power consumption at the BS and a large proportion is consumed on the BS's hardware, such as RF chains, baseband processing modules, AC-DC units and cooling systems[24]–[26]. It is shown in [26] that the ratio of the circuit power decreases with the cell coverage. The ratio is up to 74% for pico-cells, while it is about 43% for macro-cells. Therefore, considering the circuit power in the energy-efficient design and investigating its impact on the EE of the BS are very meaningful.

It can be seen from problem (19) that the value of the objective function decreases with the circuit power, P_c, when α and $\{\beta_m\}_{m=1}^M$ are fixed. On the other hand, constraints (19a) and (19b) are independent of P_c. Therefore, the optimal EE obtained from problem (19) decreases with P_c.

To highlight the impact of circuit power on the optimal transmit power, we consider single user case and the user index m is omitted for simplicity. Denote the ratio of the power for pilots to the overall transmit power as θ. Then the power for pilots and data symbols can be respectively expressed as $\alpha = P_t\theta$ and $\beta = P_t(1-\theta)$, where β denotes the power for data symbols of the single user and P_t represents the overall transmit power. Substituting them into problem (19), we obtain an optimization problem with respect to P_t and θ as follows,

$$\max_{P_t,\theta} \frac{\Delta f N \mathbb{E}_g \left[\log_2\left(1 + \frac{P_t^2\theta(1-\theta)\sum_{l=1}^{L}\frac{\gamma_l^2}{\gamma_l P_t\theta+\sigma^2}|g|^2}{\left(N + P_t(1-\theta)\sum_{l=1}^{L}\frac{\gamma_l}{\gamma_l P_t\theta+\sigma^2}\right)\sigma^2}\right)\right]}{\rho P_t + P_c} \tag{21}$$

$$\text{s.t.} \quad 0 \le P_t \le P_{max}, \quad \theta \ge 0, \tag{21a}$$

where N denotes the number of subcarriers occupied by the user and $\{\gamma_l\}_{l=1}^L$ represents the channel gains of resolvable paths.

For general channels where $\{\gamma_l\}_{l=1}^L$ are arbitrary values, it is very hard to find the relationship among P_t, θ, and P_c. To show the impact of channel characteristics, we consider two extreme channel models. One is flat fading channel, i.e. $\gamma_1 = \zeta$ and $\gamma_2 = \cdots \gamma_L = 0$, and the other is that all the resolvable channel paths have the same variance as in [7], i.e. $\gamma_1 = \cdots = \gamma_L = \delta$. With these two channel models, we can find the optimal θ with a similar method in [9] as

$$\theta_1^* = \frac{1}{\sqrt{N\frac{\zeta P_t+\sigma^2}{\zeta P_t + N\sigma^2}} + 1}, \quad \text{and} \quad \theta_2^* = \frac{1}{\sqrt{N\frac{\delta P_t+\sigma^2}{L\delta P_t + N\sigma^2}} + 1},$$

where θ_1^* is the power ratio when the channel is flat fading and θ_2^* is for the other channel. After substituting them into the objective function of problem (21), the optimization problem only depends on P_t and can be rewritten as follows,

$$\max_{P_t} \quad \eta(P_t) \tag{22}$$

$$\text{s. t.} \quad 0 \le P_t \le P_{max}, \tag{22a}$$

where $\eta(P_t)$ is defined as $\frac{C(P_t)}{\rho P_t + P_c}$ and $C(P_t) \triangleq \Delta f N \mathbb{E}_g\left[\log_2\left(1 + W(P_t)|g|^2\right)\right]$ is the ergodic channel capacity of the user. $W(P_t)$ has different expressions in different channel models. For the flat-fading channel, we have

$$W_1(P_t) = \frac{(N+1)\zeta P_t + 2N\sigma^2 - 2\sqrt{N(N\sigma^2 + \zeta P_t)(\zeta P_t + \sigma^2)}}{(N-1)^2\sigma^2}. \tag{23}$$

When the channel gains of resolvable paths have the same variance, we have

$$W_2(P_t) \quad (24)$$
$$= \frac{L\left((N+L)\delta P_t + 2N\sigma^2 - 2\sqrt{N(N\sigma^2 + L\delta P_t)(\delta P_t + \sigma^2)}\right)}{(N-L)^2\sigma^2}.$$

To see the impact of P_c on the optimal solution of problem (22), we first study the impact of P_c on the optimal overall transmit power that maximizes its objective function, $\eta(P_t)$ without the transmit power constraint $(22a)$.

It is easy to show that $\frac{\partial \eta(P_t)}{\partial P_t}|_{P_t=0} > 0$ and $\lim_{P_t\to\infty} \frac{\partial \eta(P_t)}{\partial P_t} < 0$. Denote the optimal transmit power that maximizes $\eta(P_t)$ as P_t^*. Then P_t^* satisfies

$$\frac{\partial \eta(P_t)}{\partial P_t}|_{P_t=P_t^*} = 0. \quad (25)$$

Considering the expression of $\eta(P_t)$ and after some manipulations, (25) can be rewritten as

$$\frac{C^{'}(P_t^*)(\rho P_t^* + P_c) - \rho C(P_t^*)}{(\rho P_t^* + P_c)^2} = 0,$$

where $C^{'}(P_t^*)$ denotes the derivative of $C(P_t)$ at the point $P_t = P_t^*$. Then we can obtain the expression of P_c as a function of P_t^* as

$$P_c = \frac{\rho C(P_t^*)}{C'(P_t^*)} - \rho P_t^*.$$

The derivative of P_c over P_t^* can then be expressed as

$$\frac{\mathrm{d}P_c}{\mathrm{d}P_t^*} = -\frac{\rho C(P_t^*) C^{''}(P_t^*)}{\left(C'(P_t^*)\right)^2}, \quad (26)$$

where $C^{''}(P_t^*)$ denotes the second derivative of $C(P_t)$ at the point $P_t = P_t^*$.

Due the complicated expression of $W(P_t^*)$ and the expectation operation in the expression of ergodic channel capacity, it is very hard to judge the sign of $C^{''}(P_t^*)$. To gain useful insights, we further consider two extreme cases as follows.

When P_t^* is very small, we can approximate $W(P_t^*)$ in (23) and (24) as

$$W_1(P_t^*) \approx \frac{(N+1)\zeta P_t^*}{(N-1)^2\sigma^2}, \quad (27)$$

and

$$W_2(P_t^*) \approx \frac{(N+L)L\delta P_t^*}{(N-L)^2\sigma^2}, \quad (28)$$

respectively. When P_t^* is very large, we can omit the noise item in (23) and (24) and approximate $W(P_t^*)$ as

$$W_1(P_t^*) \approx \frac{(\sqrt{N}-1)^2\zeta P_t^*}{(N-1)^2\sigma^2}, \quad (29)$$

and

$$W_2(P_t^*) \approx \frac{(\sqrt{N}-\sqrt{L})^2 L\delta P_t^*}{(N-L)^2\sigma^2}, \quad (30)$$

respectively. After substituting $(27)-(30)$ into the expression of $C(P_t^*)$, we can find that $C(P_t^*)$ is a concave function of P_t^* and thus $C^{''}(P_t^*) < 0$. Then we can obtain from (26) that $\frac{\mathrm{d}P_c}{\mathrm{d}P_t^*} > 0$, which means that the optimal power, P_t^*, increases with P_c.

Based on this relationship between P_t^* and P_c, we can further study the impact of P_c on the optimal solution of problem (22). When the value of P_c is small such that $P_t^* \le P_{max}$, the optimal solution of problem (22) is P_t^* and increases with P_c. When P_c is so large that $P_t^* > P_{max}$, the optimal solution of problem (22) is P_{max}, which does not change with the value of P_c.

In [12], only the case that $P_c = 0$ is considered while in [7], only the case that P_c is so large that $P_t = P_{max}$ is studied. Different from their work, we have analyzed the relationship among P_c, the EE, and the overall transmit power in this section. The impact of P_c on the power ratio is also shown by simulation in Section VI.

B. Impact of Multiple Users

In OFDMA systems, the number of subcarriers occupied by each user affects the optimal EE. To simplify the analysis, we consider that the channels of multiple users have the same normalized power delay profile but with different average channel gains. Denote $\{\phi_l\}_{l=1}^L$ as the normalized power delay profile and Γ_m as the average channel gain of user m, respectively. Then the channel gain of the lth path of user m, $\gamma_{m,l} = \Gamma_m \phi_l$, and we can rewrite $f(\alpha, \beta_m)$ as

$$f(\alpha,\beta_m) \triangleq \frac{\beta_m \sum_{l=1}^{L} \frac{\gamma_{m,l}^2 \alpha}{\gamma_{m,l}\alpha + \sigma^2}}{\left(N_m + \beta_m \sum_{l=1}^{L} \frac{\gamma_{m,l}}{\gamma_{m,l}\alpha+\sigma^2}\right)\sigma^2}$$
$$= \frac{\beta_m \sum_{l=1}^{L} \frac{\Gamma_m^2 \phi_l^2 \alpha}{\Gamma_m \phi_l \alpha + \sigma^2}}{\left(N_m + \beta_m \sum_{l=1}^{L} \frac{\Gamma_m \phi_l}{\Gamma_m \phi_l \alpha+\sigma^2}\right)\sigma^2}. \quad (31)$$

We can readily prove that $f(\alpha, \beta_m)$ is an increasing function of Γ_m (see Appendix D for details). Denote the largest average channel gain as $\Gamma_{\max}$. Then we can obtain

$$f(\alpha,\beta_m) \le \frac{\beta_m \sum_{l=1}^{L} \frac{\Gamma_{\max}^2 \phi_l^2 \alpha}{\Gamma_{\max} \phi_l \alpha + \sigma^2}}{\left(N_m + \beta_m \sum_{l=1}^{L} \frac{\Gamma_{\max} \phi_l}{\Gamma_{\max} \phi_l \alpha+\sigma^2}\right)\sigma^2}, \; m = 1, \cdots, M.$$

Correspondingly, the EE in (18) can be upper bounded as follows,

$$\eta \le \frac{\Delta f \sum_{m=1}^{M} N_m \mathbb{E}_g \left[\log_2(1 + f_{max}(\alpha,\beta_m)|g|^2)\right]}{\rho\left(\alpha + \sum_{m=1}^{M} \beta_m\right) + P_c}, \quad (32)$$

where $f_{max}(\alpha,\beta_m) \triangleq \frac{\beta_m \sum_{l=1}^{L} \frac{\Gamma_{\max}^2 \phi_l^2 \alpha}{\Gamma_{\max} \phi_l \alpha + \sigma^2}}{\left(N_m + \beta_m \sum_{l=1}^{L} \frac{\Gamma_{\max} \phi_l}{\Gamma_{\max} \phi_l \alpha+\sigma^2}\right)\sigma^2}$ is used.

It is readily to prove that $f(\alpha, \beta_m)$ increases with β_m and thus we can obtain

$$\sum_{m=1}^{M} N_m \mathbb{E}_g \left[\log_2(1 + f_{max}(\alpha,\beta_m)|g|^2)\right] \quad (33)$$

$$\leq \sum_{m=1}^{M} N_m \mathbb{E}_g \left[\log_2\left(1+f_{max}\left(\alpha, \sum_{m=1}^{M}\beta_m\right)|g|^2\right)\right],$$

where $\beta_m \leq \sum_{m=1}^{M}\beta_m$ is used. Substituting this inequality into (32), we can obtain

$$\eta \leq \frac{\Delta f \sum_{m=1}^{M} N_m \mathbb{E}_g \left[\log_2\left(1+f_{max}\left(\alpha, \sum_{m=1}^{M}\beta_m\right)|g|^2\right)\right]}{\rho\left(\alpha+\sum_{m=1}^{M}\beta_m\right)+P_c}, \tag{34}$$

where the equality is achieved when the BS only transmits data to the user with the highest average channel gain with the number of subcarriers, $\sum_{m=1}^{M} N_m$, and the total transmit power, $\sum_{m=1}^{M}\beta_m$. (34) implies that there exists a tradeoff between the EE performance and the fairness among the users, i.e., fairness needs to be taken into account in the EE-oriented design for multi-user systems. This cannot be observed from the single user case in [12].

Though the observation is easy to understand, the design considering the fairness is not straightforward herein. In fact, user fairness can be defined under various criteria. From the aspect of the fairness in terms of resource, each user can be assigned with equal time-frequency resource or equal transmit power in OFDMA systems[27]. From the aspect of the fairness in terms of the data rate, proportional fairness or max-min fairness rules can be applied [28]. These criteria are already well explored in the SE-oriented design, and can be extended to the EE-oriented design. From the aspect of the fairness in terms of the EE, the proportional fairness among multiple users has been considered for uplink OFDMA transmission in [29]. However, for the downlink OFDMA systems, because the circuit power at the BS is consumed by all users, the contribution of each user on the circuit power is not clear and thus the EE function of each user is hard to derive. In the current work, we assign equal number of subcarriers to each user to guarantee the resource fairness in the later simulations.

VI. Simulation Results

In this section, we evaluate the EE, the overall transmit power, and the ratio of the power for pilots to the overall transmit power of downlink OFDMA systems using the proposed method. The parameters are listed in Table III. The value of power amplifier efficiency is referred to [30] and the value of circuit power ranges from 0 dBm to 60 dBm, which covers the circuit power consumption in existing practical systems[26].

We first evaluate the performance of the proposed method in Section IV by comparing it with the optimal solution. Because the EE function in (18) does not have a good property in general cases, we employ exhaustive searching to find the optimal power allocation for problem (19). Due to high complexity of the exhaustive searching, only one user is considered. Fig. 2 shows the EE versus the circuit power under different maximum transmit power, P_{max}. The curves with the circle marks represent the optimal performance while the curves with the square marks denote the performance of the proposed method. We can see that the EEs obtained by the proposed method overlap with the maximal EEs obtained by the exhaustive searching. In the following, we will employ the proposed method to analyze the performance of the OFDMA systems.

TABLE III
List of Simulation Parameters

Subcarrier spacing, Δf	15 kHz
Number of subcarriers, N_{tot}	1024
Number of pilots, K	16
Number of users, M	5
Number of subcarriers occupied by each user, $\{N_m\}_{m=1}^{M}$	201
Cell radius, R	250 m
Distribution of M users	uniformly distributed in the cell region
Number of the channel resolvable paths of User m, L	8
Channel power delay profile	All channel paths have the same variance
Power spectral density of noise	-174 dBm/Hz
Standard deviation of Shadowing	8 dB
Noise amplifier gain	7 dBi
Minimum distance from the BS to users, d_{min}	35 m
Path loss (dB)	35+38$\log_{10} d$
Maximum transmit power, P_{max}	20, 30, 40 dBm
Power amplifier efficiency, $1/\rho$	38%

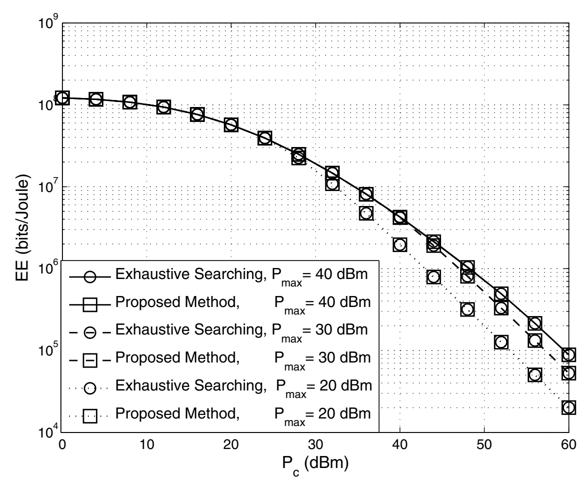
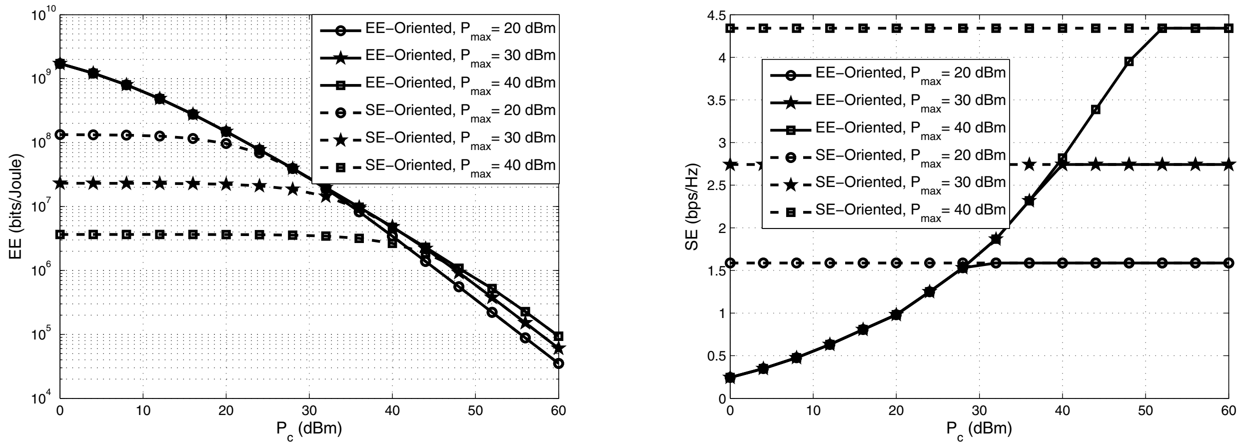


Fig. 2. Comparison of the performance of the proposed method with the maximal EE obtained by exhaustive searching.

Our work considers how to maximize the EE while traditional design concerns about how to maximize the SE. To show the difference of the two design criteria, we compare their performance in terms of the EE and SE. We call them as EE-oriented and SE-oriented criteria for short. For the EE-oriented criterion, we can solve problem (19) by using the method in Section IV. For the SE-oriented criterion, the optimization problem is to maximize the downlink ergodic channel capacity in (15) with overall transmit power constraints in $(19a)$ and $(19b)$. Since the downlink ergodic channel capacity increases with the transmit power, it is maximized when the pilots



(a) The EE vs. circuit power

(b) The SE vs. circuit power

Fig. 3. The EE and SE of EE-oriented and SE-oriented design.

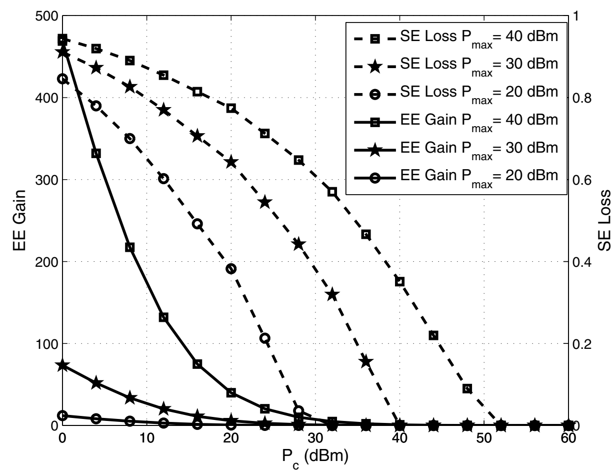


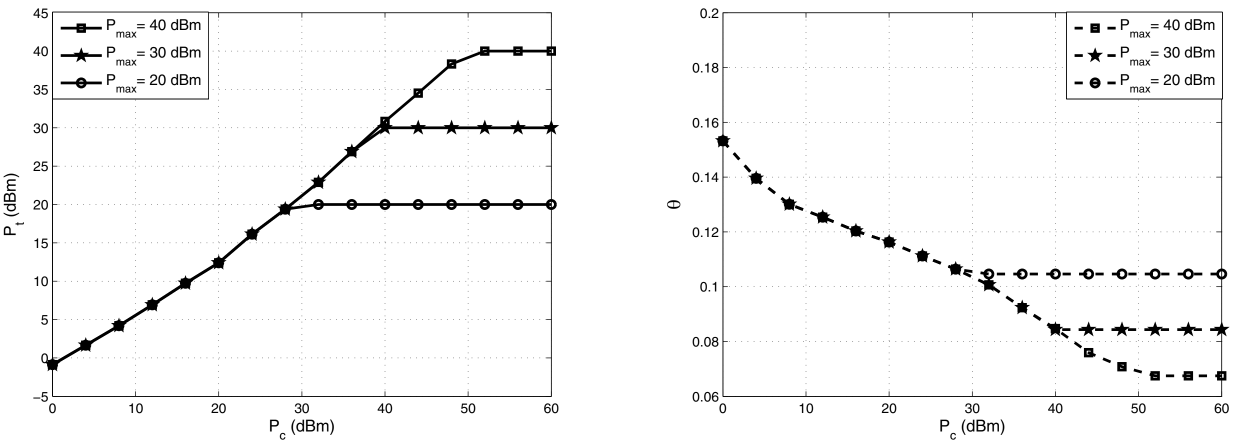
Fig. 4. The EE gain and the SE loss of the EE-oriented design over the SE-oriented design.

and data symbols are transmitted with the maximum transmit power, P_{max} and thus we can solve this problem by using the method in Section IV-B.

Fig. 3 shows the performance of the EE-oriented and SE-oriented design in terms of the EE and the SE. Considering the large dynamic range of the EE, we use the log-scale axis for the EE value. We can see that the EE decreases with the circuit power for both design criteria. The SE of EE-oriented design increases with the circuit power until the circuit power becomes dominant such that the maximum transmit power is used, while the SE of SE-oriented design keep a constant. When the circuit power is low, the EE-oriented design outperforms SE-oriented design in terms of the EE significantly but its SE is lower than that of SE-oriented design. With the increase of the circuit power, these two criteria have the same performance on both the EE and the SE. This observation can be explained as follows. When the circuit power is low, the optimal overall transmit power for EE-oriented design is much lower than the maximum value, P_{max}, which results in the gap on the EE and SE performance. As the circuit power increases, the optimal overall transmit power for EE-oriented design approaches P_{max} and the circuit power is dominant in the overall power consumption. The EE-oriented design reduces to the SE-oriented design. To observe the EE and SE differences of both design criteria more clearly, we show the EE gain and the SE loss of the EE-oriented design over the SE-oriented design in Fig. 4. The EE gain is defined as the ratio of EE difference between these two design criteria to the EE value of SE-oriented design and the SE loss is defined as the ratio of SE difference between two criteria to the SE of the SE-oriented design. It can be seen that both the EE gain and the SE loss decrease with the circuit power. When $P_{max} = 40$ dBm and $P_c = 20$ dBm, the EE gain is about 40 times over the EE of SE-oriented design and the SE loss is about 80% compared with the SE of SE-oriented design. This implies that the EE-oriented design can improve the EE significantly with a relatively small SE loss from the SE-oriented design.

The required overall transmit power and the ratio of the power for pilots to the overall transmit power for the EE-oriented design is shown in Fig. 5. When the circuit power is low, the overall transmit power increases with the circuit power, which is consistent with the analysis in Section V-A. The power ratio decreases with the circuit power, which implies that more power is used for data transmission rather than channel estimation. As the circuit power becomes higher, the overall transmit power is constrained to the maximum value and the power ratio is a constant.

To observe the impact of multiple users of OFDMA system on the EE-oriented optimization, we show the EE of a two user system, where the number of subcarriers allocated to them varies. We assume that these two users are 50 m and 100 m away from the BS, and we call the user closer to the BS as user 1. As shown in Fig. 6, we can see that the EE increases with the subcarrier ratio for user 1 with different value of the circuit power, i.e., transmitting more data symbols to the user with higher channel gain can improve the EE. This implies that maximizing the EE of a multi-user system will be at a cost of sacrificing the fairness among the users, which is analogous to the SE-oriented design.



(a) Overall transmit power vs. circuit power

(b) Pilot power ratio vs. circuit power

Fig. 5. Overall transmit power and pilot power ratio for pilots for the EE-oriented design.

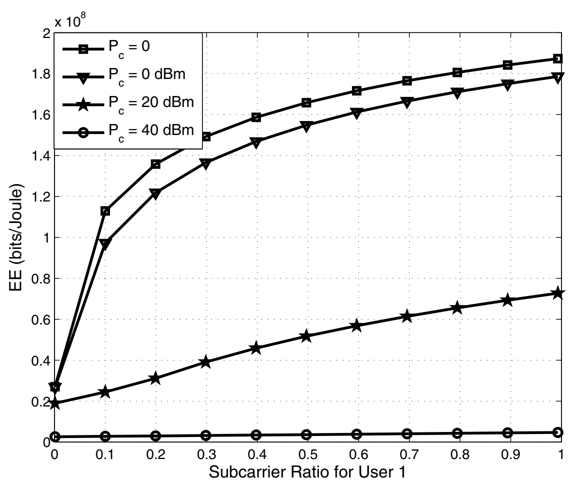


Fig. 6. Impact of multiple users on the EE of the EE-oriented design.

VII. Conclusion

In this paper, we have studied power allocation between pilots and data symbols to maximize the EE of training-based downlink OFDMA systems when circuit power consumption is considered. We have derived the EE function when channel estimation error exists, formulated the optimization problem, and analyzed the impact of the circuit power and multiple users. To provide an efficient solution to the optimization problem, an alternating optimization method is developed in the low transmit power region and an approximation on the EE is used to reformulate a joint quasiconcave problem in the high transmit power region. Analysis and simulation results show that the EE-oriented design can improve the EE performance significantly with a relatively small SE loss compared with SE-oriented design. When $P_{max} = 40$ dBm and $P_c = 20$ dBm, the EE gain is about 40 times over the EE of SE-oriented design while the SE of the EE-oriented design achieves about 80% of that of the SE-oriented design. When the circuit power becomes dominant, the EE-oriented design reduces to the SE-oriented design. The overall transmit power increases and the power ratio for pilots decreases with the circuit power. Transmitting more data symbols to the user with higher channel gain can improve the EE.

Appendix A
Derivation of $\mathbf{\Phi}_m^H \mathbf{\Phi}_m = \alpha \mathbf{I}_L$

From the definition $\mathbf{\Phi}_m \triangleq \sqrt{N_{tot}}\mathbf{T}_p\mathbf{\Psi}_p\mathbf{F}_L$, we have

$$\mathbf{\Phi}_m^H \mathbf{\Phi}_m = N_{tot}\mathbf{F}_L^H\mathbf{\Psi}_p^H\mathbf{T}_p^H\mathbf{T}_p\mathbf{\Psi}_p\mathbf{F}_L. \tag{35}$$

Since the power for pilots, α, is equally allocated to the subcarriers, we have

$$\mathbf{T}_p^H\mathbf{T}_p = \frac{\alpha}{K}\mathbf{I}_K.$$

Upon substituting the above equation, (35) becomes

$$\mathbf{\Phi}_m^H \mathbf{\Phi}_m = \frac{\alpha N_{tot}}{K}\mathbf{F}_L^H\mathbf{\Psi}_p^H\mathbf{\Psi}_p\mathbf{F}_L = \frac{\alpha N_{tot}}{K}\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}, \tag{36}$$

where $\tilde{\mathbf{F}} \triangleq \mathbf{\Psi}_p\mathbf{F}_L$.

Since pilots are periodically inserted in frequency domain, the indexes of subcarriers occupied by pilots can be expressed as $\{kT+c\}_{k=0}^{K-1}$, where $T \triangleq \frac{N_{tot}}{K}$ denotes the period and c represents the position of the first pilot symbol. For example, in Fig. 1, $T = 2$ and $c = 1$. Hence, the subcarrier index set can be expressed as $\mathcal{S}_p = \{c, T+c, \cdots, (K-1)T+c\}$. Correspondingly, the (k,n)th element of $\mathbf{\Psi}_p$ is

$$(\mathbf{\Psi}_p)_{k,n} = \begin{cases} 1, & \text{when}\, n = kT+c \\ 0, & \text{when}\, n \neq kT+c \end{cases} \quad k = 0, \cdots, K-1.$$

When $\mathbf{F}_L$ is multiplied by $\mathbf{\Psi}_p$, the (k,r)th element of $\tilde{\mathbf{F}}$ can be written as

$$\left(\tilde{\mathbf{F}}\right)_{k,r} = \frac{1}{\sqrt{N_{tot}}}e^{-j\frac{2\pi(kT+c)r}{N_{tot}}},\ k = 0, \cdots, K-1, \quad \text{and}$$
$$r = 0, \cdots, L-1. \tag{37}$$

Upon substituting (37) into (36), the (n_1, n_2)th element of $\mathbf{\Phi}_m^H\mathbf{\Phi}_m$ can be derived as follows,

$$\left(\mathbf{\Phi}_m^H\mathbf{\Phi}_m\right)_{n_1,n_2} = \frac{\alpha N_{tot}}{K}\sum_{k=0}^{K-1}\left(\tilde{\mathbf{F}}\right)_{k,n_1}^*\left(\tilde{\mathbf{F}}\right)_{k,n_2}$$

$$\frac{\partial^2 f(\alpha,\beta_m)}{\partial\alpha^2} = \frac{2\beta_m(\beta_m \sum\limits_{l=1}^{L} \gamma_{m,l} + N_m\sigma^2)\left(-N_m \sum\limits_{l=1}^{L} Z_{m,l}^3 + \beta_m\left(\left(\sum\limits_{l=1}^{L} Z_{m,l}^2\right)^2 - \sum\limits_{l=1}^{L} Z_{m,l}\sum\limits_{l=1}^{L} Z_{m,l}^3\right)\right)}{\sigma^2\left(N_m + \beta_m \sum\limits_{l=1}^{L} Z_{m,l}\right)^3} \tag{38}$$

$$=\frac{\alpha}{K}\sum_{k=0}^{K-1} e^{j(\frac{2\pi k n_1}{K}+\frac{2\pi c n_1}{N_{tot}}-\frac{2\pi k n_2}{K}-\frac{2\pi c n_2}{N_{tot}})}$$
$$=\frac{\alpha}{K}e^{j\frac{2\pi c(n_1-n_2)}{N_{tot}}}\sum_{k=0}^{K-1} e^{j\frac{2\pi k(n_1-n_2)}{K}}.$$

We can readily derive that

$$\sum_{k=0}^{K-1} e^{j\frac{2\pi k(n_1-n_2)}{K}} = \begin{cases} K, & \text{when } n_1 = n_2 \\ 0, & \text{when } n_1 \neq n_2 \end{cases}.$$

Therefore, we can obtain $\mathbf{\Phi}_m^H\mathbf{\Phi}_m = \frac{\alpha}{K}K\mathbf{I}_L = \alpha\mathbf{I}_L$.

Appendix B
Proof of Concavity of $f(\alpha, \beta_m)$ over α or β_m

To facilitate the proof description, we define $Z_{m,l} \triangleq \frac{\gamma_{m,l}}{\gamma_{m,l}\alpha+\sigma^2}$. Then $f(\alpha,\beta_m)$ can be rewritten as

$$f(\alpha,\beta_m) = \frac{\beta_m\left(\sum\limits_{l=1}^{L}\gamma_{m,l} - \sum\limits_{l=1}^{L}\frac{\gamma_{m,l}\sigma^2}{\gamma_{m,l}\alpha+\sigma^2}\right)}{\left(N_m + \beta_m\sum\limits_{l=1}^{L}\frac{\gamma_{m,l}}{\gamma_{m,l}\alpha+\sigma^2}\right)\sigma^2}$$
$$= \frac{\beta_m\sum\limits_{l=1}^{L}\gamma_{m,l} + N_m\sigma^2}{\left(N_m + \beta_m\sum\limits_{l=1}^{L} Z_{m,l}\right)\sigma^2} - 1.$$

We can derive the second partial derivative of $f(\alpha,\beta_m)$ with respect to α as shown in (38) on the top of this page, where $\frac{\partial Z_{m,l}}{\partial\alpha} = -Z_{m,l}^2$ is used.

From Hölder's inequality, we have

$$\sum_{l=1}^{L} Z_{m,l}^2 \leq \left(\sum_{l=1}^{L} Z_{m,l}\right)^{\frac{1}{2}}\left(\sum_{l=1}^{L} Z_{m,l}^3\right)^{\frac{1}{2}}.$$

Then we can obtain

$$\frac{\partial^2 f(\alpha,\beta_m)}{\partial\alpha^2} \leq \frac{-2N_m\beta_m\left(\beta_m\sum\limits_{l=1}^{L}\gamma_{m,l} + N_m\sigma^2\right)\sum\limits_{l=1}^{L} Z_{m,l}^3}{\sigma^2\left(N_m + \beta_m\sum\limits_{l=1}^{L} Z_{m,l}\right)^3} < 0.$$

Therefore, $f(\alpha,\beta_m)$ is concave with respect to α.

Similarly, we can derive the second partial derivative of $f(\alpha,\beta_m)$ with respect to β_m and prove that $\frac{\partial^2 f(\alpha,\beta_m)}{\partial\beta_m^2} < 0$, which means that $f(\alpha,\beta_m)$ is a concave function of β_m.

To show that $f(\alpha,\beta_m)$ is not a jointly concave function with respect to α and β_m, we consider a special case that all the resolvable channel paths of user m have the same variance as in [7], i.e. $\gamma_{m,1} = \cdots = \gamma_{m,L} = \delta_m$. In this case,

$$f(\alpha,\beta_m) = \frac{L\delta_m^2\alpha\beta_m}{(N_m\delta_m\alpha + L\delta_m\beta_m + N_m\sigma^2)\sigma^2}. \tag{39}$$

The Hessian matrix of $f(\alpha,\beta_m)$ is given by

$$\nabla^2 f = \frac{LN_m\delta_m^2}{\sigma^2(\chi(\alpha,\beta_m))^3}\cdot \tag{40}$$
$$\begin{pmatrix} -2\delta_m\beta_m(L\delta_m\beta_m + N_m\sigma^2) & \sigma^2\chi(\alpha,\beta_m) + 2L\delta_m^2\alpha\beta_m \\ \sigma^2\chi(\alpha,\beta_m) + 2L\delta_m^2\alpha\beta_m & -2L\delta_m\alpha(\delta_m\alpha+\sigma^2) \end{pmatrix},$$

where $\chi(\alpha,\beta_m) \triangleq N_m\delta_m\alpha + L\delta_m\beta_m + N_m\sigma^2$.

The determinant of $\nabla^2 f$ is equal to $-\frac{L^2N_m^2\delta_m^4}{(N_m\delta_m\alpha+L\delta_m\beta_m+N_m\sigma^2)^4}$ and is less than zero, which means that the eigenvalues of $\nabla^2 f$ are not all negative. Therefore, $\nabla^2 f$ is not a negative semidefinite matrix and $f(\alpha,\beta_m)$ is not jointly concave with respect to α and β_m.

Appendix C
Proof of Quasiconcavity of the EE function

The Hessian matrix of $\tilde{f}(\alpha,\beta_m)$ is given by

$$\nabla^2\tilde{f} = \frac{-2\psi_m N_m L}{(N_m\alpha + L\beta_m)^3}\begin{pmatrix} \beta_m^2 & -\alpha\beta_m \\ -\alpha\beta_m & \alpha^2 \end{pmatrix}. \tag{41}$$

It is easy to prove that $\nabla^2\tilde{f}$ is a negative semidefinite matrix. Therefore, $\tilde{f}(\alpha,\beta_m)$ is a concave function with respect to α and β_m. The concavity of $\tilde{f}(\alpha,\beta_m)$ can be extended to the variable set $\{\alpha,\{\beta_m\}_{m=1}^M\}$. According to the composition rule that preserves concavity [15], we can conclude that both the capacity of user m, C_m, and the sum capacity, $\sum\limits_{m=1}^{M} C_m$, are concave. Then EE function (18) is expressed as a concave function over a linear function. Similar to the proof in [5], we can prove that the EE function is quasiconcave.

Appendix D
Proof of the Increasing of $f(\alpha,\beta_m)$ with Γ_m

By dividing the numerator and the denominator of (31) by Γ_m, we can obtain

$$f(\alpha,\beta_m) = \frac{\beta_m\sum\limits_{l=1}^{L}\frac{\Gamma_m\phi_l^2\alpha}{\Gamma_m\phi_l\alpha+\sigma^2}}{\left(\frac{N_m}{\Gamma_m} + \beta_m\sum\limits_{l=1}^{L}\frac{\phi_l}{\Gamma_m\phi_l\alpha+\sigma^2}\right)\sigma^2}. \tag{42}$$

It can be observed that the numerator of the right-hand side of (42) is an increasing function of Γ_m and the denominator of the right-hand side of (42) is a decreasing function of Γ_m. Therefore, we can conclude that $f(\alpha,\beta_m)$ is an increasing function of Γ_m.

References

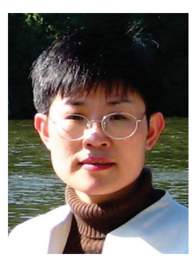
[1] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Commun. Mag.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[2] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental tradeoffs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, June 2011.
[3] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
[4] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.
[5] 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.
[6] H. Kim, C.-B. Chae, G. de Veciana, and R. W. Heath, "A cross-layer approach to energy efficiency for adaptive MIMO systems exploiting spare capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009.
[7] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.
[8] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.
[9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[10] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmission of doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
[11] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmission: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
[12] M. C. Gursoy, "On the capacity and energy efficiency of training-based transmissions over fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4543–4567, Oct. 2009.
[13] 3GPP, "Evolved universal terrestrial radio access (E-UTRA): Physical channels and modulation (release 9)," *TS 36.211 v9.0.0*, Dec. 2009. Available: www.3gpp.org
[14] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.
[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[16] Y. Liu, Y. Dai, and Z. Luo, "Max-min fairness linear transceiver design for a multi-user MIMO interference channel," in *Proc. 2011 IEEE ICC*.
[17] I. Hammerström and A. Wittneben, "On the optimal power allocation for nonregenerative OFDM relay links," in *Proc. 2006 IEEE ICC*.
[18] M. V. Solodov, "On the convergence of constrained parallel variable distritution algorithms," *SIAM J. Optim.*, vol. 8, no. 1, pp. 187–196, Feb. 1998.
[19] Y. Li, B. Vucetic, Z. Zhou, and M. Dohler, "Distributed adaptive power allocation for wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 948–958, Mar. 2007.
[20] I. Hammerström and A. Wittneben, "Joint power allocation for nonregenerative MIMO-OFDM relay links," in *Proc. 2006 IEEE ICASSP*.
[21] J. Gromicho, *Quasiconvex Optimization and Location Theory*. Kluwer Academic Publishers, 1998.
[22] S. Schaible and W. T. Ziemba, *Generalized Concavity in Optimization and Economics*. Academic Press, 1981.
[23] J. P. Penot, "Lagrangian approach to quasiconvex programing," *J. Optim. Theory Appl.*, vol. 117, no. 3, pp. 637–647, June 2003.
[24] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. V. der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.
[25] G. Auer, V. Giannini, C. Desset, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, D. Sebella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
[26] G. Auer, O. Blume, V. Giannini, I. Gódor, *et al.*, "D 2.3: energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *EARTH*, Nov. 2010. Available: https://www.ict-earth.eu/publications/deliverables/deliverables.html
[27] 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.
[28] G. Song and G. Y. Li, "Cross-layer optimization for OFDM wireless networks—part I: theorectical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–624, Mar. 2005.
[29] G. Miao, N. Himayat, G. Y. Li, and D. Bormann, "Energy-efficient design in wireless OFDMA," in *Proc. 2008 IEEE ICC*.
[30] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. 2010 ACM Future Netw. MobileSummit*.

Zhikun Xu received his B.S.E. degree in electronics engineering in 2007 and is now pursuing his Ph.D. degree in signal and information processing, both in the School of Electronics and Information Engineering, Beihang University, Beijing, China. From September 2009 to September 2010, he worked as a visiting student in the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests include cognitive radio, cross-layer resource allocation, and green radio.

Geoffrey Ye Li received his B.S.E. and M.S.E. degrees in 1983 and 1986, respectively, from the Department of Wireless Engineering, Nanjing Institute of Technology, Nanjing, China, and his Ph.D. degree in 1994 from the Department of Electrical Engineering, Auburn University, Alabama. He was a Teaching Assistant and then a Lecturer with Southeast University, Nanjing, China, from 1986 to 1991, a Research and Teaching Assistant with Auburn University, Alabama, from 1991 to 1994, and a Post-Doctoral Research Associate with the University of Maryland at College Park, Maryland, from 1994 to 1996. He was with AT&T Labs - Research, Red Bank, New Jersey, as a Senior and then a Principal Technical Staff Member from 1996 to 2000. Since 2000, he has been with the School of Electrical and Computer Engineering at the Georgia Institute of Technology as an Associate and then a Full Professor. He has also held the Cheung Kong Scholar title at the University of Electronic Science and Technology of China since March 2006.

His general research interests include statistical signal processing and telecommunications, with emphasis on OFDM and MIMO techniques, cross-layer optimization, and signal processing issues in cognitive radios. In these areas, he has published over 200 papers in refereed journals or conferences and two books, 20 of which with over 100 Google citations. He has over 20 granted patents. He once served or is currently serving as an editor, a member of the editorial board, and a guest editor for over 10 technical journals. He organized and chaired many international conferences, including technical program vice-chair of IEEE ICC'03 and co-chair of IEEE SPARC'11. He has been awarded an IEEE Fellow for his contributions to signal processing for wireless communications since 2006, selected as a Distinguished Lecturer for 2009–2010 by the IEEE Communications Society, and won the 2010 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communications theory.



Chenyang Yang received her M.S.E. and Ph.D. degrees in 1989 and 1997 in electrical engineering, from Beijing University of Aeronautics and Astronautics (BUAA, now renamed as Beihang University). She is now a full professor in the School of Electronics and Information Engineering, BUAA. She has published various papers and filed many patents in the fields of signal processing and wireless communications. She was nominated as an Outstanding Young Professor of Beijing in 1995 and was supported by the 1st Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by the Ministry of Education (P.R.C. "TRAPOYT") from 1999–2004. Currently, she serves as an associate editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an associate editor-in-chief of the *Chinese Journal of Communications*, and an associate editor-in-chief of the *Chinese Journal of Signal Processing*. She is chair of the Beijing Chapter of IEEE Communications Society. She has served as a TPC member for many IEEE conferences such as ICC and GLOBECOM. Her recent research interests include network MIMO, energy efficient transmission, and interference management in multi-cell systems.

Shunqing Zhang received his B.E. degree from Fudan University, Shanghai, China, in 2005 and his Ph.D. degree from HKUST in 2009. He joined the Green Radio Excellence in Architecture and Technology (GREAT) project at Huawei Technologies Co. Ltd. after graduation. His current research interests include energy-efficient resource allocation and optimization in cellular networks, joint baseband and radio frequency optimization, fundamental tradeoffs on green wireless network design, and other green radio technologies for energy saving and emission reduction.

Yan Chen received her B.Sc. and the Ph.D. degrees in information and communication engineering from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. She had been a visiting researcher at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, from 2008 to 2009. After graduation, she joined the Central Research Institute of Huawei Technologies Co., Ltd. She is now the project manager and technical leader of GREAT (Green Radio Excellence in Architectures & Technologies), an internal green research project of Huawei, focusing on the fundamental design of energy efficient solutions for radio access networks. Her current research interests include fundamental tradeoffs on green wireless network design, green network information theory, energy-efficient network architecture and management, as well as radio technologies, advanced processing, and resource management algorithms.

Shugong Xu received a B.Sc. degree from Wuhan University, China, and his M.E. and Ph.D. from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1990, 1993, and 1996, respectively. He is currently director of the Access Network Technology Research Department, principal scientist, and vice director of the Communication Laboratory, Huawei Corporate Research, and chief scientist of Wireless Access Technologies in the Nation al Key Laboratory on Wireless technologies at Huawei. He leads green research in Huawei including the GREAT (Green Radio Excellence in Architecture and Technologies) project, which focuses on power-efficient solutions for wireless radio access networks. Prior to joining Huawei Technologies in 2008, he was with Sharp Laboratories of America as a senior research scientist for seven years. Before starting his career in industrial research, he worked at universities including Tsinghua University, China, Michigan State University, and the City College of New York (CCNY).

In his over 18 years research on cutting edge topics of wireless/mobile networking and communication, home networking, and multimedia communications, he published more than 30 peer-reviewed research papers as lead author in top international conferences and journals, in which the most referenced one has over 900 Google Scholar citations. He holds more than 30 granted US patents or patent applications, of which technologies have been adopted in 802.11 and LTE standards, including the DRX protocol for power saving in the LTE standard. Dr. Xu is a senior member of IEEE, a Concurrent Professor at HUST, and the Technical Committee co-chair of the Green Touch consortium.