Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas

Derrick Wing Kwan Ng and Robert Schober
The University of British Columbia, Canada

***Abstract*— In this paper, resource allocation for energy efficient communication in orthogonal frequency division multiple access (OFDMA) downlink networks with large numbers of base station (BS) antennas is studied. Assuming perfect channel state information at the transmitter (CSIT), the resource allocation algorithm design is modeled as a non-convex optimization problem for maximizing the energy efficiency of data transmission (bit/Joule delivered to the users), where the circuit power consumption and a minimum required data rate are taken into consideration. Subsequently, by exploiting the properties of fractional programming, an efficient iterative resource allocation algorithm is proposed to solve the problem. In particular, the power allocation, subcarrier allocation, and antenna allocation policies for each iteration are derived. Simulation results illustrate that the proposed iterative resource allocation algorithm converges in a small number of iterations and unveil the trade-off between energy efficiency and the number of antennas.**

I. Introduction

The demand for high data rate multi-media applications with certain guaranteed quality of service (QoS) has been growing rapidly over the last decade. Multiple-input multiple-output (MIMO) technology is considered as a viable solution for addressing this issue, as it provides extra spatial degrees of freedom for resource allocation. Recently, the concept of large numbers of antennas has received growing research interest [1]-[3]. In [1], the authors investigated channel estimation and linear precoding techniques in time division duplex (TDD) system with a large number of base station (BS) antennas. In [2], high throughputs were shown in both uplink and downlink for a TDD multi-cell system which employs multiple BSs equipped with large numbers of antennas. In [3], a low complexity signal detection algorithm was proposed for large-scale MIMO systems. A substantial capacity gain (bit/s/Hz) and a better interference management were observed with MIMO compared to single antenna systems in all studies [1]-[3]. Yet, the advantages of MIMO do not come for free. The extra energy consumption in antenna circuitries have significant financial implications for service providers, which has been largely overlooked in the literature so far. As a result, energy efficient system designs, which adopt *energy efficiency* (bits-per-Joule) as the performance metric, have also recently drawn much attention in both industry and academia.

In [4], the authors studied the energy efficiency of MIMO sensors networks and demonstrated that MIMO systems may not always be more energy-efficient than single antenna systems for short range communications. In [5], a power loading algorithm was designed to minimize the energy-per-goodbit in MIMO systems. In [6], the optimal transmission mode selection for maximizing energy efficiency was studied for MIMO multi-carrier systems. In [7], the tradeoff between energy efficiency, bandwidth, and number of antennas was investigated in MIMO multi-hop networks. However, these works considered single user systems with a small number of antennas and the results in [4]-[7] may not be applicable in multi-user multi-carrier systems with a large number of transmitter antennas.

This work has been supported in part by the Natural Science and Engineering Council of Canada (NSERC) Strategic Project Grant STPGP 396545.

In this paper, we address the above issues. For this purpose, we formulate the resource allocation algorithm design as an optimization problem and we maximize the energy efficiency of communication in orthogonal frequency division multiple access (OFDMA) systems. By exploiting the properties of fractional programming, the considered non-convex optimization problem in fractional form is transformed into an equivalent optimization problem in subtractive form with a tractable solution, which can be found with an iterative algorithm.

II. OFDMA Downlink Network Model

We consider an OFDMA network which consists of a BS with multiple antennas, K mobile users equipped with a single antenna, and n_F subcarriers. The channel gains are assumed to be time-invariant (slow fading) and known at the BS. The downlink received symbol at user $k \in \{1, \ldots, K\}$ on subcarrier $i \in \{1, \ldots, n_F\}$ is given by

$$y_{i,k} = \sqrt{P_{i,k} l_k g_k} \mathbf{h}_{i,k}^T \mathbf{f}_{i,k} x_{i,k} + z_{i,k}, \tag{1}$$

where $x_{i,k}$, $P_{i,k}$, and $\mathbf{f}_{i,k} \in \mathbb{C}^{N_{T_{i,k}} \times 1}$ are the transmitted symbol, transmitted power, and precoding vector for the link from the BS to user k on subcarrier i, respectively. $\mathbb{C}^{N \times M}$ is the space of all $N \times M$ matrices with complex entries and $[\cdot]^T$ denotes the transpose operation. $N_{T_{i,k}}$ is the number of active antennas allocated to user k on subcarrier i for transmission. $\mathbf{h}_{i,k} \in \mathbb{C}^{N_{T_{i,k}} \times 1}$ is the vector of small scale fading coefficients between the BS and user k. The elements in $\mathbf{h}_{i,k}$ are assumed to independent and identically distributed (i.i.d.). l_k and g_k represent the path loss and shadowing between the BS and user k, respectively. $z_{i,k}$ is additive white Gaussian noise (AWGN) on subcarrier i at user k with zero mean and power spectral density N_0.

III. Resource Allocation

A. Instantaneous Channel Capacity

In this subsection, we define the adopted system performance measure. Given perfect channel state information (CSI) at the receiver, the channel capacity between the BS and user k on subcarrier i with channel bandwidth W is given by

$$C_{i,k} = W \log_2 \Big(1 + \Gamma_{i,k}\Big) \text{ and } \Gamma_{i,k} = \frac{P_{i,k} l_k g_k |\mathbf{h}_{i,k}^T \mathbf{f}_{i,k}|^2}{N_0 W}, \tag{2}$$

where $\Gamma_{i,k}$ is the received signal-to-noise ratio (SNR) at user k on subcarrier i and $|\cdot|$ denotes the absolute value of a complex-valued scalar.

The *weighted sum system capacity* is defined as a weighted sum of the number of bits per second successfully delivered to the K mobile users (bits-per-second) and is given by

$$U(\mathcal{P}, \mathcal{A}, \mathcal{S}) = \sum_{k=1}^{K} \sum_{i=1}^{n_F} w_k s_{i,k} C_{i,k}, \tag{3}$$

where $\mathcal{P}$, $\mathcal{A}$, and $\mathcal{S}$ are the power allocation, antenna allocation, and subcarrier allocation policies, respectively. w_k are positive constants provided by the upper layers, which allow the resource allocator to prioritize certain users for the sake of fairness. $s_{i,k} \in \{0,1\}$ is the binary valued subcarrier allocation variable. Besides, we model the power dissipation in the system as the sum of three terms [5] which can be expressed as

$$\begin{aligned} & U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S}) \\ = \quad & \max_{i,k}\{s_{i,k}N_{T_{i,k}}\} \times P_C + \sum_{k=1}^{K}\sum_{i=1}^{n_F} \varepsilon P_{i,k}s_{i,k} + P_0, \end{aligned} \quad (4)$$

where P_C is the constant *circuit power consumption per antenna*, which includes the power dissipation of the transmit filter, mixer, frequency synthesizer, and digital-to-analog converter and is independent of the actual transmitted power. In the considered system, we assume that there are a maximum number of active antennas and a minimum number of active antennas, i.e., $N_{\max}$ and $N_{\min}$, at the BS. However, we do not necessarily activate the maximum number of antennas for the sake of energy efficient communication and the optimal number of active antennas will be found in the next section based on optimization. The physical meaning of the term $\max\limits_{i,k}\{s_{i,k} \times N_{T_{i,k}}\}$ in (4) is that if an antenna is activated, it consumes power even if it is used only by some of the users on some of the subcarriers. In order words, the first term in (4) represents the total power consumed by the activated antennas. The second term in (4) denotes the total power consumption in the radio frequency (RF) power amplifier of the BS. $\varepsilon \geq 1$ is a constant which accounts for the inefficiency in the power amplifier and the power efficiency is defined as $\frac{1}{\varepsilon}$. P_0 is the basic power consumed independent of the number of transmit antennas. Hence, the *weighted energy efficiency* of the considered system is defined as the total average number of bits delivered to the users/Joule

$$U_{eff}(\mathcal{P},\mathcal{A},\mathcal{S}) \quad = \quad \frac{U(\mathcal{P},\mathcal{A},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})}. \quad (5)$$

B. Optimization Problem Formulation

The optimal resource allocation policies $(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)$ can be obtained by solving

$$\begin{aligned} & \max_{\mathcal{P},\mathcal{A},\mathcal{S}} \; U_{eff}(\mathcal{P},\mathcal{A},\mathcal{S}) \qquad (6) \\ \text{s.t. } & \text{C1: } \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k}C_{i,k} \geq r, \quad \text{C2: } \sum_{k=1}^{K}\sum_{i=1}^{n_F} P_{i,k}s_{i,k} \leq P_T, \\ & \text{C3: } \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k} \leq 1, \qquad \text{C4: } P_{i,k} \geq 0, \; \forall i,k, \\ & \text{C5: } s_{i,k} \in \{0,1\}, \; \forall i,k, \\ & \text{C6: } N_{T_{i,k}} \in \{N_{\min}, N_{\min}+1, N_{\min}+2, \ldots, N_{\max}\}, \forall i,k. \end{aligned}$$

Here, C1 ensures that a minimum system data rate r is achieved. C2 is a transmit power constraint for the BS in the downlink. C3 and C5 are imposed to guarantee that each subcarrier can serve one user only. In other words, inter-user interference does not exist. C4 are the boundary constraints for the power allocation variables.

Note that the above optimization problem formulation can be extended to the case of imperfect CSI and subcarrier reuse by different users, as shown in the journal version of this paper [8].

IV. Solution of the Optimization Problem

The objective function in (6) is a non-convex function. In general, a brute force approach is required for obtaining a global optimal solution which is computationally infeasible even for small size systems. In order to derive an efficient resource allocation algorithm, we introduce the following transformation.

TABLE I
ITERATIVE RESOURCE ALLOCATION ALGORITHM.

Algorithm 1 Iterative Resource Allocation Algorithm

Initialize the maximum number of iterations $L_{\max}$ and the maximum tolerance ϵ
Set maximum energy efficiency $q = 0$ and iteration index $n = 0$
repeat {Main Loop}
  Solve the inner loop problem in (9) for a given q and obtain resource allocation policies $\{\mathcal{P}',\mathcal{A}',\mathcal{S}'\}$
  if $U(\mathcal{P}',\mathcal{A}',\mathcal{S}') - qU_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}') < \epsilon$ **then**
    Convergence $=$ **true**
    return $\{\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*\} = \{\mathcal{P}',\mathcal{A}',\mathcal{S}'\}$ and $q^* = \frac{U(\mathcal{P}',\mathcal{A}',\mathcal{S}')}{U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')}$
  else
    Set $q = \frac{U(\mathcal{P}',\mathcal{A}',\mathcal{S}')}{U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')}$ and $n = n+1$
    Convergence $=$ **false**
  end if
until Convergence $=$ **true** or $n = L_{\max}$

A. Transformation of the Objective Function

The objective function in (5) can be classified as nonlinear fractional program [9]. For the sake of notational simplicity, we define Θ as the set of feasible solutions of the optimization problem in (6) and $\{\mathcal{P},\mathcal{A},\mathcal{S}\} \in \Theta$. Without loss of generality, we define the maximum energy efficiency q^* of the considered system as

$$q^* = \frac{U(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)}{U_{TP}(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)} = \max_{\mathcal{P},\mathcal{A},\mathcal{S}} \frac{U(\mathcal{P},\mathcal{A},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})}. \quad (7)$$

We are now ready to introduce the following Theorem.

Theorem 1: The maximum energy efficiency q^* is achieved *if and only if*

$$\begin{aligned} \max_{\mathcal{P},\mathcal{A},\mathcal{S}} \quad & U(\mathcal{P},\mathcal{A},\mathcal{S}) - q^*U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S}) \\ = \quad & U(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*) - q^*U_{TP}(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*) = 0, \end{aligned} \quad (8)$$

for $U(\mathcal{P},\mathcal{A},\mathcal{S}) \geq 0$ and $U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S}) > 0$.

Proof: Please refer to Appendix A for a proof of Theorem 1. Theorem 1 reveals that for an optimization problem with an objective function in fractional form, there exists an equivalent[1] objective function in subtractive form, e.g. $U(\mathcal{P},\mathcal{A},\mathcal{S}) - q^*U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})$ in the considered case. As a result, we can focus on the equivalent objective function in the rest of the paper.

B. Iterative Algorithm for Energy Efficiency Maximization

In this section, we propose an iterative algorithm (known as the Dinkelbach method [9]) for solving (6) with an equivalent objective function. The proposed algorithm is summarized in Table I and the convergence to optimal energy efficiency is guaranteed.

Proof: Please refer to Appendix B for the proof of convergence.

As shown in Table I, in each iteration in the main loop, we solve the following optimization problem for a given q:

$$\begin{aligned} \max_{\mathcal{P},\mathcal{A},\mathcal{S}} \quad & U(\mathcal{P},\mathcal{A},\mathcal{S}) - qU_{TP}(\mathcal{P},\mathcal{A},\mathcal{S}) \\ & \text{s.t. C1, C2, C3, C4, C5, C6.} \end{aligned} \quad (9)$$

[1]Here, "equivalent" means both problem formulations will lead to the same resource allocation policies.

The transformed problem in (9) is a mixed combinatorial and non-convex optimization problem. To obtain an optimal solution, an exhaustive search is needed with complexity $\sum_{t=N_{\min}}^{N_{\max}} t \times n_F^K$, which is computational infeasible for $N_{T_{i,k}}, K, n_F \gg 1$. In order to derive an efficient resource allocation algorithm, we solve the above problem in two steps.

Step 1 (Asymptotic Channel Capacity for $N_{T_{i,k}} \to \infty$): In the first step, the beamforming vector $\mathbf{f}_{i,k}$ adopted at the BS is chosen to be the eigenvector corresponding to the maximum eigenvalue of $\mathbf{h}_{i,k}\mathbf{h}_{i,k}^{\dagger}$, i.e., $\mathbf{f}_{i,k} = \frac{\mathbf{h}_{i,k}}{\|\mathbf{h}_{i,k}\|}$ where $\|\cdot\|$ denotes the Euclidean norm of a vector. The adopted beamforming scheme is known as maximum ratio transmission (MRT). As a result, for sufficiently large $N_{T_{i,k}}$, the capacity equation in (2) can be expressed as

$$C_{i,k} \overset{(a)}{\approx} W \log_2\Big(\Gamma_{i,k}\Big) \text{ and } \Gamma_{i,k} = \frac{P_{i,k} l_k g_k N_{T_{i,k}}}{W N_0}, \qquad (10)$$

where (a) is due to the law of large numbers and the high SNR assumption, i.e., $\lim_{N_{T_{i,k}} \to \infty} \frac{\mathbf{h}_{i,k}\mathbf{h}_{i,k}^{\dagger}}{N_{T_{i,k}}} = 1$ and $\log_2(1+x) \approx \log_2(x)$ for $x \gg 1$, respectively.

Step 2 (Constraint Relaxations): In the second step, we handle the combinatorial constraints in C5 and C6 by relaxing the corresponding variables such that $0 \le s_{i,k} \le 1$ and $N_{T_{i,k}}$ is a positive real number. For facilitating the derivation of the resource allocation algorithm, we introduce two auxiliary variables and define them as $\widetilde{P}_{i,k} = P_{i,k} s_{i,k}$ and $\widetilde{N}_{T_{i,k}} = N_{T_{i,k}} s_{i,k}$. Then we substitute them into (9). By doing so, the optimization problem in (9) becomes jointly concave with respect to (w.r.t.) $\widetilde{P}_{i,k}$, $s_{i,k}$, and $\widetilde{N}_{T_{i,k}}$. As a result, solving the dual problem is equivalent to solving the primal. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as $\mathcal{L}(\mu, \gamma, \boldsymbol{\beta}, \mathcal{P}, \mathcal{A}, \mathcal{S})$

$$= \sum_{k=1}^{K}\sum_{i=1}^{n_F}(w_k+\gamma)s_{i,k}C_{i,k} - \gamma r - \mu\Big(\sum_{k=1}^{K}\sum_{i=1}^{n_F} P_{i,k}s_{i,k} - P_T\Big)$$
$$- \sum_{i=1}^{n_F}\beta_i\Big(\sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k} - 1\Big) - q\Big(U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\Big) \qquad (11)$$

where $\gamma \ge 0$ and $\mu \ge 0$ are the Lagrange multipliers corresponding to the required minimum capacity constraint C1 and maximum transmit power allowance C2, respectively. $\boldsymbol{\beta}$ is the Lagrange multiplier vector associated with the subcarrier assignment constraint C3 with elements $\beta_i \ge 0$, $i \in \{1, \ldots, n_F\}$. Boundary constraints C5 and C6 will be absorbed into the Karush-Kuhn-Tucker (KKT) conditions when deriving the optimal resource allocation policies in the following. Thus, the dual problem of (9) for a given parameter q is

$$\min_{\mu,\gamma,\boldsymbol{\beta} \ge 0} \max_{\mathcal{P},\mathcal{A},\mathcal{S}} \quad \mathcal{L}(\mu, \gamma, \boldsymbol{\beta}, \mathcal{P}, \mathcal{A}, \mathcal{S}). \qquad (12)$$

Since (9) is transformed into a concave optimization problem after step 1 and step 2, the KKT conditions are the necessary and sufficient conditions for the optimal solution. Thus, from (12), the closed-form resource allocation policies for the BS serving user k in subcarrier i for a given parameter q are obtained as:

$$P_{i,k}^{*} = \left[\frac{\mathcal{B}/n_F(w_k+\gamma)}{\ln(2)(\mu+q\varepsilon)}\right]^{+}, \qquad (13)$$

$$N_{T_{i,k}}^{*} = \left[\frac{\mathcal{B}/n_F(\max_{k\in\Psi_i} w_k + \gamma)}{\ln(2)P_C(q/\Phi_i)}\right]_{N_{\min}}^{N_{\max}}, \text{ and} \qquad (14)$$

$$s_{i,k}^{*} = \begin{cases} 1 & \text{if } k = \arg\max_b M_{i,b},\ M_{i,b} \ge \beta_i = 0, \forall i \\ 0 & \text{otherwise} \end{cases}, \qquad (15)$$

where $M_{i,b} = \frac{\mathcal{B}}{n_F}(w_b+\gamma)\Big[\log_2\Big(\frac{N_{T_{i,b}}^{*}P_{i,b}^{*}l_b g_b}{N_0 W}\Big) - 2/\ln(2)\Big]$. $\big[x\big]_b^a = a$, if $x > a$; $\big[x\big]_b^a = x$, if $b \le x \le a$; $\big[x\big]_b^a = b$, if $b > x$ and $\big[x\big]^{+} = \max\{0, x\}$. The optimal power allocation solution in (13) is in the form of *multi-level water-filling*. Note that if a user has a higher value of w_k (higher priority), a higher power will be allocated to the user since she has a higher water level $\frac{\mathcal{B}/n_F(w_k+\gamma)}{\ln(2)(\mu+q\varepsilon)}$ compared to other users. In (14), Ψ_i denotes a selected user set for using subcarrier i and $\Phi_i = \sum_{b\in\Psi_i} 1(\max_{k\in\Psi_i} w_k = w_b)$ counts the number of w_k which have a value equal to $\max_{k\in\Psi_i} w_k$ for all selected users on subcarrier i. $1(\cdot)$ denotes an indicator function which is 1 when the event is true and 0 otherwise. On the other hand, since the dual function in (12) is differentiable, the optimal values of μ and γ can be found by using numerical methods such as the gradient method [8] or the bisection method. Besides, updating β_i is not necessary as it has the same value for each user. Therefore, setting $\beta_i = 0$ does not affect the subcarrier allocation in (15).

TABLE II
COORDINATE ASCENT METHOD.

Algorithm 2 Coordinate Ascent Method for Solving (9)

1: Set the iteration indices $t = 0, m = 0$, maximum number of iterations $t_{\max}, m_{\max}$
2: Initialize the Lagrange multipliers μ, γ and resource allocation policies $\{\mathcal{P}_t, \mathcal{A}_t, \mathcal{S}_t\}$ for $t = 0$
3: **repeat** {Outer loop}
4: **repeat** {Inner loop}
5: Solve the power allocation and antenna allocation by using (13) and (14) for all subcarriers with subcarrier allocation policy $\mathcal{S}_t$. Assign the solutions to $\mathcal{P}_{t+1}$ and $\mathcal{A}_{t+1}$
6: Solve the subcarrier allocation for all subcarriers by using (15) together with $\mathcal{P}_{t+1}$ and $\mathcal{A}_{t+1}$. Assign the solution to $\mathcal{S}_{t+1}$; $t = t+1$
7: **until** Convergence = **true** or $t = t_{\max}$
8: Update μ and γ by gradient method or bisection method; $m = m+1$
9: **until** Convergence= **true** or $m = m_{\max}$
10: **return** $\{\mathcal{P}_t, \mathcal{A}_t, \mathcal{S}_t\}$ as $\{\mathcal{P}', \mathcal{A}', \mathcal{S}'\}$ to line 4 in Algorithm 1

Coordinate Ascent Method for Implementing (13)-(15): Theoretically, equations (13)-(15) provide a complete solution for the considered resource allocation problem. However, the dependency between (14) and (15) is a hurdle for practical implementation. In this section, we present another iterative algorithm to bridge the gap between theory and practice. The algorithm is outlined in Table II. The overall iterative algorithm in solving (9) is implemented by two nested loops in Algorithm 2. In Table II, $t_{\max}$ and $m_{\max}$ are the maximum number of iterations for the two nested loops. $\mathcal{P}_t$, $\mathcal{S}_t$, $\mathcal{A}_t$ are the power allocation, subcarrier allocation, and antenna allocation policies in the t-th iteration, respectively. The inner loop, i.e., line 4 to line 7, is solving the maximization in (12) by using the coordinate ascent method for a given set of Lagrange multipliers. In particular, in line 5, we first keep the subcarrier allocation fixed and optimize the power allocation policy and antenna allocation policy. Then, in line 6, we use the optimized policies $\mathcal{P}_{t+1}$ and

$\mathcal{A}_{t+1}$ from line 5 to optimize the subcarrier allocation policies. Convergence of the inner loop to the optimum point for a given set of Lagrange multipliers is ensured for convex optimization problems [10]. On the other hand, the outer loop, i.e., line 3 to line 9, solves the minimization in the (12) by updating the Lagrange multipliers.

Note that the resource allocation policies obtained in Algorithm 1 and Algorithm 2 are optimal w.r.t. the relaxed problem in high SNR. For implementing the final resource allocation policies, we need to apply a ceiling function $\lceil\cdot\rceil$ to the antenna allocation solution from the output of Algorithm 1, i.e., $N_{T_{i,k}} = \lceil N^*_{T_{i,k}} \rceil$.

V. Results

In this section, we evaluate the system performance through simulations. A single cell with a radius of 1 km is considered. There are $n_F = 128$ subcarriers with carrier center frequency 2.5 GHz and a total system bandwidth of $W = 5$ MHz. We assume a noise power of $N_0W = -128$ dBm in each subcarrier and $w_k = 1\,\forall k$. The K desired users are uniformly distributed between the reference distance and the cell boundary at 1 km. The 3GPP path loss model is used with a reference distance of $d_0 = 35$ m and log-normal shadowing with a standard deviation of 8 dB. The small scale fading coefficients of the BS-to-user link are modeled as i.i.d. Rayleigh random variables with zero means and unit variances. The average system energy efficiency is obtained by counting the amount of data which is successfully decoded by the users and dividing it by the total power consumption averaged over both macroscopic and microscopic fading. We assume a static circuit power consumption of $P_0 = 40$ dBm [11] and a data rate requirement of $r = 80$ Megabit/s. $P_C = 41$ dBm denotes the additional power dissipation incurred by each extra antenna for transmission [12]. In practice, the value of P_C depends on the application-specific integrated circuit (ASIC) and the specific implementation. On the other hand, we assume a power efficiency of 20% in the RF power amplifier. i.e., $\varepsilon = \frac{1}{0.2} = 5$. The maximum and minimum numbers of active antennas are set to $N_{\max} = 100$ and $N_{\min} = 10$, respectively. Note that if the resource allocator is unable to guarantee the minimum data rate in a time slot, we set the energy efficiency in that particular time slot to zero to account for the corresponding failure. On the other hand, in the following results, the "number of iterations" is referring to the number of iterations of Algorithm 1 in Table I.

Figure 1 illustrates the energy efficiency versus the total transmit power for $K = 30$ users. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 10. It can be seen that the performance difference between 5 iterations and 10 iterations is negligible which confirms the practicality of our proposed iterative resource allocation algorithm. It can be observed that when the maximum transmit power at the power amplifier is large enough, e.g., $P_T \geq 34$ dBm, the energy efficiency of the proposed algorithm approaches a constant value since the resource allocator is not willing to consume more power or activate more antennas, when the maximum energy efficiency is achieved. For comparison, Figure 1 also contains the energy efficiency of a baseline resource allocation scheme in which resource allocation is performed for maximizing the spectral efficiency (bit/s/Hz) in (9) except that the number of transmit antennas is fixed to $N_{T_{i,k}} = 10, 15, 20, \forall i, k$, respectively. It can be observed that the energy efficiency of the baseline scheme is far from optimal in the high transmit power regime and a fixed number of transmit antennas $N_{T_{i,k}}$ always degrades the system performance in terms of energy efficiency. This is because in the baseline scheme, either more power is consumed by the circuitries for operating the antennas or the number of antennas is not large enough for satisfying the minimum data rate requirement. Note that the energy efficiency will drop dramatically if the number of antennas further increases from $N_{T_{i,k}} = 20$ to $N_{T_{i,k}} > 20$ in the baseline scheme. More simulation results and detailed explanations can be found in the journal version of this paper [8].

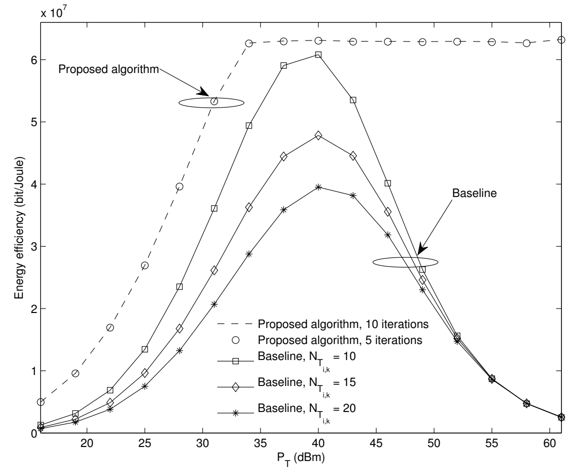

Fig. 1. Energy efficiency versus maximum transmit power, P_T, for $K = 30$ users.

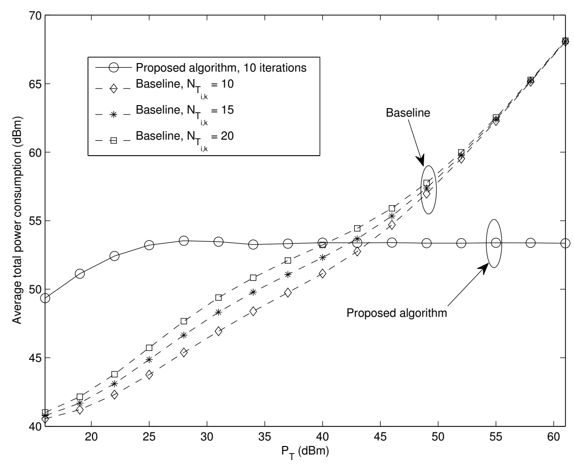

Fig. 2. Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{S})\}$, versus maximum transmit power, P_T, for different resource allocation algorithms, 10 iterations, and $K = 30$ users.

Figure 2 depicts the average total power consumption, i.e., $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{S})\}$, versus maximum transmit power P_T for the proposed algorithm with 10 iterations and the baseline scheme, where $\mathcal{E}\{\cdot\}$ denotes statistical expectation. In the low transmit power regime, the proposed algorithm consumes more power than the baseline scheme. This is because more antennas have to be activated for satisfying the data rate requirement. However, as the maximum transmit power allowance P_T increases, the proposed algorithm gradually decreases the number of activated antennas and clips the transmit power to save energy. On the other hand, in the high transmit power regime, the proposed algorithm consumes much lesser power than the baseline scheme since the latter one always consumes all the power in order to maximize the spectral efficiency.

VI. Conclusion

In this paper, we formulated the resource allocation algorithm design for OFDMA networks with large numbers of BS antennas as a non-convex optimization problem, where the circuit power dissipation and the system data rate requirement were taken into consideration. An efficient iterative resource allocation algorithm with optimized power allocation, subcarrier allocation, and antenna allocation policies was proposed by using fractional programming and the law of large numbers. Our simulation results did not only show that the proposed algorithm converges within a small number of iterations, but also unveiled the trade-off between energy efficiency and the total transmit power.

Appendix

A. Proof of Theorem 1

We now prove the forward implication of Theorem 1 by following a similar approach as in [9]. Without loss of generality, we define q^* and $\{\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*\}\in\Theta$ as the optimal energy efficiency and the optimal resource allocation policy of the *original objective function* in (6), respectively. Then, the optimal energy efficiency can be expressed as

$$q^*=\frac{U(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)}{U_{TP}(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)}\geq\frac{U(\mathcal{P},\mathcal{A},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})},\ \forall\{\mathcal{P},\mathcal{A},\mathcal{S}\}\in\Theta,$$
$$\Longrightarrow U(\mathcal{P},\mathcal{A},\mathcal{S})-q^*U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\leq 0 \text{ and}$$
$$U(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)-q^*U_{TP}(\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*)=0. \quad (16)$$

Therefore, we conclude that $\max\limits_{\mathcal{P},\mathcal{A},\mathcal{S}} U(\mathcal{P},\mathcal{A},\mathcal{S}) - q^*U(\mathcal{P},\mathcal{A},\mathcal{S})=0$, which is achievable by resource allocation policy $\{\mathcal{P}^*,\mathcal{A}^*,\mathcal{S}^*\}$. This completes the forward implication.

Next, we prove the converse implication of Theorem 1. Suppose $\{\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*\}$ is the optimal resource allocation policy of the *equivalent objective function* such that $U(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)-q^*U_{TP}(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)=0$. Then, for any feasible resource allocation policy $\{\mathcal{P},\mathcal{A},\mathcal{S}\}\in\Theta$, we can obtain the following inequality

$$\begin{aligned}&U(\mathcal{P},\mathcal{A},\mathcal{S})-q^*U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\\ \leq\ &U(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)-q^*U_{TP}(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)=0.\end{aligned} \quad (17)$$

The above inequality implies

$$\begin{aligned}\frac{U(\mathcal{P},\mathcal{A},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})}&\leq q^* \quad \forall\{\mathcal{P},\mathcal{A},\mathcal{S}\}\in\Theta \quad \text{and}\\ \frac{U(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)}{U_{TP}(\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*)}&= q^*.\end{aligned} \quad (18)$$

In other words, the optimal resource allocation policy $\{\mathcal{P}_e^*,\mathcal{A}_e^*,\mathcal{S}_e^*\}$ for the equivalent objective function is also the optimal resource allocation policy for the original objective function.

This completes the proof of the converse implication of Theorem 1. In summary, the optimization of the original objective function and the optimization of the equivalent objective function result in the same resource allocation policy. □

B. Proof of Algorithm Convergence

We follow a similar approach as in [9] for proving the convergence of Algorithm 1. We first introduce the following two propositions. For the sake of notational simplicity, we define the equivalent objective function in (9) as $F(q')=\max\limits_{\mathcal{P},\mathcal{A},\mathcal{S}}\{U(\mathcal{P},\mathcal{A},\mathcal{S})-q'U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\}$.

Proposition 1: $F(q')$ is a strictly monotonic decreasing function in q', i.e., $F(q'')>F(q')$ if $q'>q''$.

Proof: Let $\{\mathcal{P}',\mathcal{A}',\mathcal{S}'\}\in\Theta$ and $\{\mathcal{P}'',\mathcal{A}'',\mathcal{S}''\}\in\Theta$ be the two distinct optimal resource allocation policies for $F(q')$ and $F(q'')$, respectively.

$$\begin{aligned}F(q'')&=\max_{\mathcal{P},\mathcal{A},\mathcal{S}}\{U(\mathcal{P},\mathcal{A},\mathcal{S})-q''U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\}\\ &> U(\mathcal{P}',\mathcal{A}',\mathcal{S}')-q''U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')\\ &\geq U(\mathcal{P}',\mathcal{A}',\mathcal{S}')-q'U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')\\ &= F(q').\end{aligned} \quad (19)$$

□

Proposition 2: Let $\{\mathcal{P}',\mathcal{A}',\mathcal{S}'\}\in\Theta$ be an arbitrary feasible solution and $q'=\frac{U(\mathcal{P}',\mathcal{A}',\mathcal{S}')}{U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')}$, then $F(q')\geq 0$.

Proof:
$$\begin{aligned}F(q')&=\max_{\mathcal{P},\mathcal{A},\mathcal{S}}\{U(\mathcal{P},\mathcal{A},\mathcal{S})-q'U_{TP}(\mathcal{P},\mathcal{A},\mathcal{S})\}\\ &\geq U(\mathcal{P}',\mathcal{A}',\mathcal{S}')-q'U_{TP}(\mathcal{P}',\mathcal{A}',\mathcal{S}')=0.\end{aligned}$$

□

We are now ready to prove the convergence of Algorithm 1.

Proof of Convergence: We first prove that the energy efficiency q increases in each iteration. Then, we prove that if the number of iterations is large enough, the energy efficiency q converges to the optimal q^* such that it satisfies the optimality condition in Theorem 1, i.e., $F(q^*)=0$.

Let $\{\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n\}$ be the optimal resource allocation policy in the n-th iteration. Suppose $q_n\neq q^*$ and $q_{n+1}\neq q^*$ represent the energy efficiency of the considered system in iterations n and $n+1$, respectively. By Theorem 1 and Proposition 2, $F(q_n)>0$ and $F(q_{n+1})>0$ must be true. On the other hand, in the proposed algorithm, we calculate q_{n+1} as $q_{n+1}=\frac{U(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)}{U_{TP}(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)}$. Thus, we can express $F(q_n)$ as

$$\begin{aligned}F(q_n)&=U(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)-q_nU_{TP}(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)\\ &=U_{TP}(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)(q_{n+1}-q_n)>0 \quad (20)\\ \Longrightarrow q_{n+1}&>q_n, \qquad \because U_{TP}(\mathcal{P}_n,\mathcal{A}_n,\mathcal{S}_n)>0. \quad (21)\end{aligned}$$

By combining $q_{n+1}>q_n$, Proposition 1, and Proposition 2, we can show that as long as the number of iterations is large enough, $F(q_n)$ will eventually approach zero and satisfy the optimality condition as stated in Theorem 1. □

References

[1] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel Estimation and Linear Precoding in Multiuser Multiple-Antenna TDD Systems," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 2102–2116, Jun. 2011.

[2] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[3] P. Li and R. Murch, "Multiple Output Selection-LAS Algorithm in Large MIMO Systems," *IEEE Commun. Lett.*, vol. 14, pp. 399–401, May 2010.

[4] S. Cui, A. Goldsmith, and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 1089–1098, Aug. 2004.

[5] R. Prabhu and B. Daneshrad, "Energy-Efficient Power Loading for a MIMO-SVD System and its Performance in Flat Fading," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[6] H. S. Kim and B. Daneshrad, "Energy-Constrained Link Adaptation for MIMO OFDM Wireless Communication Systems," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2820–2832, Sep. 2010.

[7] C.-L. Chen, W. Stark, and S.-G. Chen, "Energy-Bandwidth Efficiency Tradeoff in MIMO Multi-Hop Wireless Networks," *IEEE J. Select. Areas Commun.*, vol. 29, pp. 1537–1546, Sep. 2011.

[8] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas," http://www.ece.ubc.ca/∼wingn/large_antennas_TWC.pdf, Tech. Rep.

[9] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, pp. 492–498, Mar. 1967. [Online]. Available: http://www.jstor.org/stable/2627691

[10] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.

[11] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks," in *Proc. Future Network and Mobile Summit*, 2010, pp. 1–8.

[12] R. Kumar and J. Gurugubelli, "How Green the LTE Technology Can be?" in *Intern. Conf. on Wireless Commun., Veh. Techn., Inform. Theory and Aerosp. Electron. Syst. Techn.*, Mar. 2011.