Energy- and Spectral-Efficiency Tradeoff in Downlink OFDMA Networks

Cong Xiong†‡, Geoffrey Ye Li†, Shunqing Zhang‡, Yan Chen‡, and Shugong Xu‡
† School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
‡ GREAT Research Team, Huawei Technologies Co., Ltd., Shanghai, China

Abstract—Conventional design of wireless networks mainly focuses on system capacity and spectral efficiency (SE). As green radio (GR) becomes an inevitable trend, energy-efficient design in wireless networks is becoming more and more important. In this paper, the fundamental relation between energy efficiency (EE) and SE in downlink orthogonal frequency division multiple access (OFDMA) networks is addressed. We first set up a general EE-SE tradeoff framework, where the overall EE, SE and per-user quality-of-service (QoS) are all considered, and prove that EE is strictly quasiconcave in SE. We also find a tight upper bound and a tight lower bound on the EE-SE curve for general scenarios, which reflect the actual EE-SE relation. We then focus on a special case that priority and fairness are considered and develop a low-complexity but near-optimal resource allocation algorithm for practical application of the EE-SE tradeoff. Numerical results corroborate the theoretical findings and demonstrate the effectiveness of the proposed resource allocation scheme for achieving a flexible and desirable tradeoff between EE and SE.

Index Terms—Energy efficiency (EE), green radio (GR), orthogonal frequency division multiple access (OFDMA), spectral efficiency (SE)

I. INTRODUCTION

In recent years, the widespread application of high-data-rate wireless services and requirement of ubiquitous access have triggered rapidly booming energy consumption. Meanwhile, the escalation of energy consumption in wireless networks leads to large amount of greenhouse gas emission and high operation expenditure. Green radio (GR) [1], which emphasizes on energy efficiency (EE) besides spectral efficiency (SE), has been proposed as an effective solution and becomes an inevitable trend for future wireless network design. Unfortunately, EE and SE do not always coincides and sometimes may even conflict [1]. Hence, how to balance EE and SE is well worth studying.

Orthogonal frequency division multiple access (OFDMA) has been extensively studied from the SE perspective and proposed for next generation wireless communication systems, such as WiMAX and the 3GPP LTE. While OFDMA can provide high throughput and SE, its EE is previously not much concerned. To keep pace with GR, it is necessary for OFDMA to guarantee a certain level of EE at the same time. Recently, more attention has been paid to energy efficient design in OFDMA networks. For uplink OFDMA transmission with flat fading channels, it is shown that EE-oriented design always consumes less energy than the traditional fixed power schemes [2]. Meanwhile, we notice that there is only limited work on joint design of EE and SE for downlink OFDMA networks.

In this paper, we address the EE-SE relation in downlink OFDMA networks. We build a general EE-SE tradeoff framework, prove that EE is quasiconcave in SE, then bound the EE-SE curve for general scenarios by a double-side approximation process, which relies on a tight upper bound and a lower bound obtained by Lagrange dual decomposition [3], [4]. When priority and fairness are considered, we propose a computationally efficient algorithm for resource allocation to facilitate application of EE-SE tradeoff.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we introduce the system model of downlink OFDMA and formulate the problem of EE-SE tradeoff.

A. System Model
We consider the downlink of a single cell OFDMA network consisting of $K$ active users. The total bandwidth $B$ is equally divided into $N$ subcarriers, each with a bandwidth of $W = B/N$. Let $K = \{1, 2, \ldots, K\}$ and $N = \{1, 2, \ldots, N\}$ denote the sets of all users and all subcarriers, respectively.

Assume that each subcarrier is exclusively assigned to at most one user each time to avoid interference among different users. Denote the transmit power of user $k$ on subcarrier $n$ as $P_{k,n}$. Then, the achievable data rate of user $k$ on subcarrier $n$ is accordingly

$$r_{k,n} = W \log_2 \left(1 + \frac{P_{k,n}g_{k,n}}{\sigma^2}\right),$$

where $g_{k,n} = |h_{k,n}|^2$ is the channel power gain of user $k$ on subcarrier $n$, $h_{k,n}$ is the corresponding frequency response and is assumed to be accurately known at the transmitter, and $\sigma^2$ is the noise power, which is, without loss of generality, assumed to be the same for all users on all subcarriers.

B. Problem Formulation
For a downlink OFDMA network, EE and SE are, respectively, defined as

$$\eta_{EE} = \frac{R}{P + P_c} \quad \text{and} \quad \eta_{SE} = \frac{R}{B},$$

where $R$, $P$ and $P_c$ denote the system overall throughput, total transmit power, and circuit power, respectively. To obtain a
high EE as well as a desirable SE and guarantee quality-of-service (QoS) for each user, it is reasonable to maximize EE under a satisfying minimum overall throughput requirement, \( \hat{R} \), a series of (minimum) rate requirements, \( R_k \)'s, depending on the traffic of the corresponding user, and the peak transmit power, \( P_T \). Since the capability of providing differentiate services is an important feature for future wireless networks, a heterogeneous traffic model includes both real-time and non-real-time traffic is considered in our work. For users with real-time services [5], such as video conferencing and online gaming, fixed rates \( \hat{R}_k \)'s are required. For users with non-real-time services [5], such as file transfers and online video, only minimum rate requirements \( \hat{R}_k \)'s are demanded. Let \( K_1 = \{1, 2, \cdots, K_0 - 1\} \) and \( K_2 = \{K_0, K_0 + 1, \cdots, K\} \) denote the sets of real-time users and non-real-time users, respectively. Accordingly, the optimization problem can be formulated as

\[
\max_{\rho, \eta_{EE}} \eta_{EE}
\]

subject to

\[
\begin{align*}
\sum_{k=1}^{K} \rho_{k,n} &\leq 1, \forall n, \rho_{k,n} \in \{0, 1\}, \forall k \in K, n, \\
\sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} r_{k,n} &= R \geq \hat{R}, \\
\sum_{n=1}^{N} \rho_{k,n} r_{k,n} &= \hat{R}_k, \forall k \in K_1, \\
\sum_{n=1}^{N} \rho_{k,n} r_{k,n} &\geq \hat{R}_k, \forall k \in K_2, \\
\sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n} &= P \leq P_T, \; p_{k,n} \geq 0, \forall k \in K, n,
\end{align*}
\]

where \( \rho = [\rho_{k,n}]_{K \times N} \) and \( P = [p_{k,n}]_{K \times N} \) are the subcarrier allocation indicator matrix and transmit power matrix, respectively.

For convenience, some other notations used in this paper are listed in advanced as follows.

- \( S_k \): set of subcarriers assigned to user \( k \).
- \( m_k \): number of assigned subcarriers of user \( k \).
- \( K_E \): set of users without assigning any subcarrier (with empty \( S_k \)'s).
- \( P_k \): overall transmit power for user \( k \), and \( P_k = \sum_{n=1}^{N} \rho_{k,n} p_{k,n} \).
- \( R_k \): overall data rate for user \( k \), and \( R_k = \sum_{n=1}^{N} \rho_{k,n} r_{k,n} \).
- \( f(R_k, S_k) \): power needed by water-filling to fulfill rate \( R_k \) over subcarrier set \( S_k \).

### III. EE-SE Relation

In this section, we will study the EE-SE tradeoff.

#### A. Fundamentals for EE-SE Tradeoff Relation

The following theorem demonstrates the quasiconvexity of EE, \( \eta_{EE} \), in the SE, \( \eta_{SE} \), and is proved in the Appendix.

**Theorem 1.** For any given SE, \( \eta_{SE} \geq \frac{\hat{R}}{P_T} \), achieved with subcarrier allocation matrix, \( \rho \), and power allocation matrix, \( P \), that satisfy all constraints but not necessarily including the peak transmit power one in (1), the maximum EE, \( \eta_{EE}^{*} (\eta_{SE}) = \max_{\rho, P} \eta_{EE} (\eta_{SE}) \), is strictly quasiconcave in \( \eta_{SE} \) if there is a sufficiently large number of subcarriers. Moreover, in the region between \( \eta_{SE} = \frac{\hat{R}}{P_T} \) and \( \eta_{SE} > \frac{\hat{R}}{P_T} \), \( \eta_{EE}^{*} (\eta_{SE}) \) is

(i) strictly decreasing with \( \eta_{SE} \) and maximized at \( \eta_{SE} = \frac{\hat{R}}{P_T} \) if

\[
\frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \bigg|_{\eta_{SE} = \frac{\hat{R}}{P_T}} \leq 0,
\]

(ii) strictly increasing with \( \eta_{SE} \) and maximized at \( \eta_{SE} = \frac{\hat{R}}{P_T} \) if

\[
\frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \bigg|_{\eta_{SE} = \frac{\hat{R}}{P_T}} > 0,
\]

and

(iii) first strictly increasing and then strictly decreasing with \( \eta_{SE} \) and maximized at \( \eta_{SE} = \frac{\hat{R}}{P_T} \) if

\[
\frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \bigg|_{\eta_{SE} = \frac{\hat{R}}{P_T}} > 0,
\]

and

\[
\frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \bigg|_{\eta_{SE} = \frac{\hat{R}}{P_T}} < 0,
\]

where \( \hat{R} \) is the maximum throughput under all constraints in (1), which is obviously achieved by transmitting at the peak transmit power, \( P_T \), and \( R_{EE,\text{max}} \) is the throughput which achieves the global maximum EE, \( \eta_{EE}^{\text{max}} \), under all constraints except the peak transmit power one in (1).

For any continuous and strictly quasiconcave function, there is always a unique global maximum over a finite domain [6, Ch. 8]. Thus, according to Theorem 1, a unique globally optimal EE of (1) always exists. More importantly, as a result of this quasiconvexity, (1) can be decomposed into two layers and solved iteratively,

(i) Inner layer: For a given SE, \( \eta_{SE} \), find the maximum EE, \( \eta_{EE}^{*} (\eta_{SE}) \), and its derivative, \( \frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \).

(ii) Outer layer: Find the optimal EE, \( \eta_{EE}^{\text{opt}} \), by bisection search like the GABS algorithm in [7].

The corresponding joint inner- and outer-layer optimization (JIOO) algorithm is listed in Table I. Then the key of the JIOO algorithm lies in the inner-layer algorithm that finds \( \eta_{EE}^{*} (\eta_{SE}) \) and \( \frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \) and will be studied in detail in the following sections.

#### B. Bounds on the EE-SE Tradeoff

As indicated before, the solution of (1) now relies on finding \( \eta_{EE}^{*} (\eta_{SE}) \) and \( \frac{d\eta_{EE}^{*} (\eta_{SE})}{d\eta_{SE}} \). Since the exact solution is too complicated to obtain in reality, we will use Lagrange dual decomposition to approximately solve it, which has been used in [3], [4] for similar problems and demonstrated to be quite
TABLE I
J I O O

Algorithm J I O O

Input: initial value of SE; \( \eta_{SE}^{(0)} = \frac{\tilde{R}}{P} \)

Output: optimal subcarrier and power allocation matrices, \( P_{opt} \) and \( \rho_{opt} \), which result in the optimal EE, \( \eta_{opt}^{SE} \)

1. \( \eta_{SE}^{(0)} = \eta_{SE}^{(0)}, \ d_1 \leftarrow \frac{d_{EE}(\eta_{SE}^{(0)})}{d_{EE}(\eta_{SE}^{(1)})}, \ \eta_{SE}^{(1)} = \eta_{SE}^{(0)} \)
2. if \( d_1 \leq 0 \)
3. then \( \eta_{opt}^{SE} \leftarrow \eta_{SE}^{(0)} \)
4. else \( \eta_{opt}^{(0)} \leftarrow \eta_{SE}^{(1)}, \ \eta_{opt}^{(1)} \leftarrow \eta_{SE}^{(0)}, \ P_{opt} = P^{*}(\eta_{opt}^{SE}), \)
   \( d_1 \leftarrow \frac{d_{EE}(\eta_{opt}^{SE})}{d_{EE}(\eta_{opt}^{(0)})}, \ \eta_{opt}^{(0)} = \eta_{opt}^{SE} \)
5. while \( d_1 > 0 \) \&\& \( P_{opt} < P_T \)
6. do \( \eta_{SE}^{(2)} = \frac{\eta_{SE}^{(1)} + \eta_{opt}^{(1)}}{2}, \ P_{opt} = P^{*}(\eta_{SE}^{(2)}), \)
   \( d_{opt} \leftarrow \frac{d_{EE}(\eta_{opt}^{(2)})}{d_{EE}(\eta_{opt}^{(0)})}, \ \eta_{opt}^{(2)} = \eta_{opt}^{SE} \)
7. if \( d_{opt} \leq 0 \) \&\& \( P_{opt} > P_T \)
8. then \( \eta_{opt}^{SE} \leftarrow \eta_{SE}^{(2)} \)
9. else \( \eta_{opt}^{SE} \leftarrow \eta_{opt}^{(2)} \)
10. return \( \rho_{opt}, P_{opt} \)

In general, the dual decomposition solution for subcarrier allocation indicator matrix, transmit power matrix and total transmit power, \( (\rho_d^*, P_d^*, P_T^*) \), to the dual problem \( (2) \) yields a lower bound of total transmit power on the optimal solution, \( (\rho_{opt}, P_T^{opt}) \), to the primal problem as a result of the nonconvexity of the primal one \([3]\). To guarantee the solution feasible, we find an achievable upper bound on the minimum transmit power based on the subcarrier allocation strategy obtained from the dual decomposition. It is easy to verify that for the subcarrier allocation strategy, \( \rho = \rho_d^* \), the tightest and achievable upper bound, \( P_{ub}^* \), on total transmit power can be achieved in two stages: in the first stage, power is distributed individually among the subcarriers of each user, \( S_k \), by water-filling, to merely fulfill its own (minimum) rate requirement, \( \tilde{R}_k \); in the second stage, extra power is then distributed among all subcarriers of the non-real-time users (each has an initial water level due to the first phase processing) by water-filling till the throughput, \( R \), is achieved.

Using dual decomposition, we approximate the minimum transmit power, \( P_{ub}^* \), tightly from both sides, i.e., \( P_{ub} \leq P_T^* \leq P_{ub}^* \), which allows us quite an accurate EE-SE tradeoff relation.

On the other hand, using \( \eta_{SE}^{opt} = \frac{R}{P_T} \) and \( \eta_{SE} = \frac{R}{P} \), the above bounds on total transmit power, \( P \), correspond to inverse bounds on SE, \( \eta_{SE}^{opt}(\eta_{SE}) \). Since it was impossible to find out the closed-form expression for \( \eta_{SE}^{opt}(\eta_{SE}) \), \( \frac{d_{EE}(\eta_{SE}^{(2)})}{d_{EE}(\eta_{opt})} \) could only be calculated via the original definition of derivative, i.e., \( \frac{d_{EE}(\eta_{SE})}{d_{EE}(\eta_{opt})} = \lim_{\Delta \eta_{SE} \to 0} \eta_{EE}(\eta_{SE} + \Delta \eta_{SE}) - \eta_{EE}(\eta_{SE}) \Delta \eta_{SE} \). For practical implementation, we can choose a small positive \( \Delta \eta_{SE} = \frac{\tilde{R}}{R} \). To further reduce complexity, we do not need to repeat the whole process for the new SE, \( \eta_{SE} + \Delta \eta_{SE} \).

Note that we only need the sign of \( \frac{d_{EE}(\eta_{opt})}{d_{EE}(\eta_{SE})} \) but not necessarily for its value. Since \( \frac{d_{EE}(\eta_{SE} + \Delta \eta_{SE})}{d_{EE}(\eta_{opt})} = \frac{B \Delta R - \Delta P_{ub}^*}{P_{ub} + \Delta P} \), we have that \( \frac{\eta_{EE}(\eta_{SE} + \Delta \eta_{SE}) - \eta_{EE}(\eta_{SE})}{\Delta \eta_{SE}} = \text{sgn}(\frac{\Delta R}{R} - \eta_{EE}(\eta_{SE})) \), where \( \text{sgn}(x) \) denotes the sign of \( x \).

We can simply distribute \( \Delta R \) among all subcarriers of the non-real-time users by water-filling to approximately find out \( \Delta P \) based on the subcarrier allocation and power allocation matrix corresponding to \( \eta_{EE}^*(\eta_{SE}) \), and the sign of \( \frac{d_{EE}(\eta_{opt})}{d_{EE}(\eta_{SE})} \) is obtained accordingly.

C. Priority, Fairness and Low-Complexity Algorithm

For many practical scenarios, capability to provide different service priority and fairness among users is important. Let \( \tilde{R} = R - \sum_{k=1}^{K} \tilde{R}_k \), then the two rate constraints in \( (1) \) can be rewritten as

\[
\sum_{n=1}^{N} p_{k,n} r_{k,n} = \tilde{R}_k + \omega_k \tilde{R},
\]

where \( \omega_k \) is the weight factor for user \( k \). We can determine the weights according to user traffic types, fairness and priority requirements. For users with real-time services, \( \omega_k \) can be simply set to zero. For users with non-real time services, which have only minimum rate requirements \( \tilde{R}_k \)'s, we can prioritize them and enforce certain notions of fairness by
adjusting $\omega$. For example, we can determine $\omega_j$ by making user rates proportional to a set of predetermined values to ensure proportional fairness among non-real-time users [8].

With certain predetermined weight vector, $\omega$, the equivalent problem of (1) given the throughput $R$ can be regarded as the conventional margin adaptation (MA) optimization problem [3], [9]–[11] as follows.

$$
\begin{align*}
\min_{P} & \quad \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} P_{k,n} \\
\text{subject to} & \quad \sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall n, \rho_{k,n} \in \{0, 1\}, \forall k, n, \\
& \quad r_{k,n} \geq 0, \forall k, n, \\
& \quad \sum_{n=1}^{N} \rho_{k,n} r_{k,n} = \bar{R}_k + \omega_k \bar{R}, \forall k,
\end{align*}
$$

(3)

Although dual decomposition can be still employed similarly as before, it typically needs $O(NK^3)$ times of water-filling to converge [3] and its complexity is still very high. Motivated by the BABS algorithm [10] for finding distribution of subcarriers in flat-fading channels, we suggest the following maximum-power-decrease-first (MPDF) algorithm. In initialization of this approach, each user is only virtually assigned its worst subcarrier and its transmit power needed in this situation will be used as a benchmark to measure how much power can be saved if the user is actually assigned a subcarrier when he has none. Then, in each iteration, each user finds its optimal subcarrier among the unassigned ones and calculates its power decrease with the additional subcarrier. Only the user with the maximum power decrease (power saving) will be assigned its favorite subcarrier in this iteration. The above iteration process will proceed until all the subcarriers have been assigned or no user needs additional subcarriers. Note that to guarantee each user is assigned at least one subcarrier at last, when the number of unassigned subcarriers equals the number of users without any subcarriers, the subcarrier assignment should be implemented among such empty users.

**TABLE II**

<table>
<thead>
<tr>
<th>Maximum-Power-Decrease-First (MPDF) Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Initialize: $K_E \leftarrow K, S_k \leftarrow \emptyset, m_k \leftarrow 0, \forall k \in K$.</td>
</tr>
<tr>
<td>2) Calculate benchmarks: $\hat{n}<em>k \leftarrow \arg \max</em>{n \in N} g_{k,n}, P_k = f(R_k, \hat{n}_k), \forall k \in K$.</td>
</tr>
<tr>
<td>3) $\hat{n}<em>k \leftarrow \arg \max</em>{n \in N} g_{k,n}, \Delta P_k \leftarrow P_k - f(R_k, S_k \cup {\hat{n}_k}), \forall k \in K$.</td>
</tr>
<tr>
<td>$\hat{k} \leftarrow \arg \max_{k \in K} \left(1 + \delta \left(N - \sum_{k=1}^{K} m_k -</td>
</tr>
<tr>
<td>Assign and update: $S_k \leftarrow S_k \cup {\hat{n}_k}, N \leftarrow N \setminus {\hat{n}_k}, m_k \leftarrow m_k + 1, P_k \leftarrow P_k - \Delta P_k, K_E \leftarrow K_E \setminus {\hat{k}}$.</td>
</tr>
<tr>
<td>4) Repeat Step 3) until $\sum_{k=1}^{K} m_k = N$ or $\max_{k \in K} \Delta P_k = 0$.</td>
</tr>
</tbody>
</table>

**Theorem 2.** The power required by water-filling with a fixed target rate is convex and non-increasing with the number of subcarriers assigned if the subcarriers are added in descending order of the power gains of subcarriers. (The proof is omitted.)

From Theorem 2, the MPDF algorithm naturally prevents one user getting too many subcarriers since benefit of acquiring subcarriers is decreasing. Meanwhile, the MPDF algorithm needs $O(NK)$ times of water-filling in total.

On the other hand, to obtain the sign of $\frac{d\eta_{SE}(R_E)}{dR_E}$ to implement the JIOO algorithm, we can follow the same way as for the general case indicated before.

**IV. Numerical Results and Conclusion**

In this section, we present some simulation results to verify the theoretical analysis and the effectiveness of the proposed approaches. In our simulation, the channel is modeled as a frequency-selective multipath fading channel consisting of six independent Rayleigh multipaths with power delay profile, $e^{-2\lambda\ell}$, $l = 0, 1, \cdots, 5$. The total bandwidth, $1.28$ MHz, is equally divided into 64 non-overlapping subcarriers. The circuit power is $2.5$ W. There are two real-time users each has a fixed rate requirement of $125$ kbps, and two non-real-time users each has a minimum rate requirement of $25$ kbps. The fairness notion employed here for non-real-time users is the partial proportional constraint (PPC) modified from the proportional constraint in [8], where $\omega_1$ : $\omega_2$ : $\cdots$ : $\omega_K$ : $\alpha_1$ : $\alpha_2$ : $\cdots$ : $\alpha_K$. For simplicity, we let $\alpha_k = \bar{R}_k \log_2 g_{k,n}$. Consequently $\omega_k = \bar{R}_k \log_2 g_{k,n}/\sum_{k=1}^{K} \bar{R}_k \log_2 g_{k,n}$. From Theorem 2, the EE-SE relation in the case that all the non-real-time users has an average CNR of $20$ dB and no specific overall throughput requirement is imposed here, i.e., $\bar{R} = 0$. From the figure, the EE-SE relation has a bell shape curve and is also quasiconcave, since the upper bound and the lower bound derived from Lagrange dual decomposition almost perfectly match together and they are in a bell shape.

Figure 2 demonstrates the relation between EE, $\eta_{SE}$, and the minimum prescribed SE, $\bar{R}_E$, and the proposed SE, $\eta_{SE} = \bar{R}_E$. In this case, the two real-time users have the same average CNR of $20$ dB. One of the two non-real-time users has an average CNR of $20$ dB, while the other has an average CNR of $17$ dB. Here no peak power constraint is imposed to investigate performance limit. From the figure, the performance of the MPDF-based method is quite close to that of the dual method, where the performance loss is within $3\%$; and it is slightly better than the DPRA [11]-based method in the high power regime.

The complexity (average computational time) is evaluated using the profile in Matlab, which is shown in Fig. 3. Obviously, our MPDF-based method has relatively lower complexity and offers an attractive performance to complexity good tradeoff.

**APPENDIX**

**Proof:** Let $R_1^*, R_2^*$ and $R_3^*$ denote the optimal rate vectors corresponding to the overall throughput $R_1$, $R_2$, and $R_3$, respectively, and they also satisfy all constraints but not necessarily including the peak power constraint in (1). Without loss of generality, assume that $R_1 < R_2 < R_3$. Let $R_2$ denote the rate vector as follows.

$$
R_2 = \frac{R_3 - R_2}{R_3 - R_1} R_1^* + \frac{R_2 - R_1}{R_3 - R_1} R_3^* = \lambda R_1^* + (1 - \lambda) R_3^*.
$$
where $\lambda = (R_3 - R_2) / (R_3 - R_1)$ and $0 < \lambda < 1$. Obviously, $R_2$ is also in the feasible region of (1) and its sum rate is $R_2$. From [3], it is known that $P(R)$ is strictly convex in $R$, given a sufficiently large number of subcarriers. Thus, $P(R_2) < \lambda P(R_1) + (1 - \lambda) P(R_3)$. Since $P(R_2)$ is the optimal rate vector among all the rate vectors with a summation of $R_2$, we have $P(R_2) \leq P(R_2)$. Consequently, we have that $P(R_2) < \lambda P(R_1) + (1 - \lambda) P(R_3)$. Thus, the minimum transmit power needed given the throughput $R = B \eta_{SE}$, $P(R) = P(R^*)$, is strictly convex in $R$ (and $\eta_{SE}$).

Denote the superlevel set of $\eta_{EE}(\eta_{SE})$ as

$$S_{\beta} = \{ R \geq \bar{R} | \eta_{EE}(\eta_{SE}) \geq \beta, \beta \in \mathbb{R} \}.$$ When $\beta \geq 0$, $S_{\beta}$ is equivalent to $S_{\beta} = \{ R \geq \bar{R} | \beta P(\eta_{SE}) + \beta P, - B \eta_{SE} \leq 0 \}$, where $P(\eta_{SE})$ is the minimum total transmit power needed for any SE $\eta_{SE} \geq \bar{R}/B$. As a result of the convexity of $P(\eta_{SE})$ (i.e., $P(R)$) proved above, $S_{\beta}$ is strictly convex in $\eta_{SE}$. Hence, $\eta_{EE}(\eta_{SE})$ is strictly quasiconcave and has a unique global maximum.

One the other hand, $\lim_{\eta_{SE} \to \infty} \eta_{EE}(\eta_{SE}) = \lim_{\eta_{SE} \to \infty} B \eta_{SE} / P = \lim_{P \to \infty} B \eta_{SE} / P = \lim_{P \to \infty} \eta_{EE}(\eta_{SE}) = 0$. Thus, starting from $\eta_{SE} = \bar{R}/B$, $\eta_{EE}(\eta_{SE})$ is either strictly decreasing with $\eta_{SE}$ if $\frac{d\eta_{EE}(\eta_{SE})}{d\eta_{SE}} |_{\eta_{SE}=\bar{R}/B} < 0$, or first strictly increasing and then strictly decreasing with $\eta_{SE}$ if $\frac{d\eta_{EE}(\eta_{SE})}{d\eta_{SE}} |_{\eta_{SE}=\bar{R}/B} > 0$. And the maximum EE within the SE region, $\left[ \frac{\bar{R}}{B}, \frac{R_{PP}}{B} \right]$, is straightforward as indicated in Theorem 1.

**REFERENCES**