2 Theoretical Framework toward Green Networks

Traditional designs in cellular networks focuses on spectrum efficiency, which is defined as the amount of bits transmitted by each unit of bandwidth. Since the first-generation cellular system, spectrum efficiency improvement, along with network coverage enhancement, has been the most important issue in network design. Many appealing technologies, such as orthogonal frequency-division multiplexing (OFDM), multiple-input multiple-output (MIMO), small-cell networking, and full-duplex communications, have been proposed in this regard.

With the explosion of wireless data applications in recent years, energy consumption of wireless networks has aroused much interest in the 5G era. The motivation of so-called energy-efficient communications or green networks is to save the energy consumption of the whole cellular network [1]. This chapter focuses on the theoretical framework toward green radio networks. In this chapter, we will first introduce the definition of energy efficiency and some important metrics for green network design. Following that, the study of energy efficiency from information theoretical aspects will be outlined. Then some fundamental trade-offs in green radio networks will be introduced, especially the energy efficiency (EE)–spectral efficiency (SE) trade-off in orthogonal frequency-division multiple access (OFDMA) networks. The EE design from the perspective of optimization theory will be also introduced in this chapter. We will finally present the EE-oriented radio resource allocation algorithms for both orthogonal and non-orthogonal systems.

2.1 Metrics for Green Radio

There are various definitions of green metrics [2], which can be roughly classified as two kinds: energy efficiency metrics and energy consumption metrics. Table 2.1 depicts some typical definitions from link level, access level, and network level.

Link-Level Metric for Green Radio

The energy consumption of a point-to-point communication link presents the required energy for transmitting one bit information (Joules/bit). Minimizing the energy consumption has been considered for a long time. Radio resource allocation aims to

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Table 2.1 Green metrics for different levels.

	Energy efficiency	Energy consumption
Link level	b/Joule; b/s/W; b/s/W; (b · m)/s/Hz/W	Joule/bit
Access level	GEE; WSEE; WMEE	
Network level	m^2/W; user/W	W/m^2 W/user

minimize the average required transmit power for a given average data rate requirement. The corresponding energy efficiency metric could be defined as the amount of transmitted bit for each Joule of energy (bits/Joule), as

$$\mathrm{EE} = \frac{R}{\zeta P_t + P_c}, \tag{2.1}$$

where ζ is the inverse of power amplifier efficiency, R is the transmit data rate, P_t relates to the transmit power, and P_c corresponds to the circuit power consumption, which can be modeled as a linear function of data rate

$$P_c = P_s + \xi R. \tag{2.2}$$

Here, P_s is the static circuit power in the transmit mode and ξ is a constant denoting dynamic power consumption per unit data rate. In some literatures, the dynamic power consumption is ignored, which corresponds to a special case that $\xi = 0$ of the mode just discussed. From (2.2), this energy efficiency metric can also be interpreted as the achievable data rate for a given transmit power (b/s/W).

Some other metrics have also been often used in certain circumstances. Corresponding to the spectral efficiency (b/s/Hz), power efficiency can be defined as the achievable spectral efficiency for a given supplied power resources (b/s/Hz/W). More generally, the radio efficiency ((b·m)/s/Hz/W) provides a more thorough definition, which takes into account the transmission distance.

These metrics mainly focus on a single point-to-point data link and can be applied in modulation and coding design, SE and EE trade-off, fundamental energy efficiency analysis, and other related topics. Link-level energy efficiency largely depends on the data rate of communication channels and the energy consumption in the transmitter and the receiver. Therefore, for a wireless channel, link-level energy efficiency is greatly impacted by the channel fading, such as path loss, shadowing, and fast fading. Also, given a point-to-point data link, energy consumption mainly consists of transmit power and circuit power. Link-level energy efficiency is important for the fundamental and theoretical study of green radios. For example, the fundamental limits of power consumption for each bit information is $\ln 2 \cdot N_0$ for an additive white Gaussian noise (AWGN) channel with a noise power density of N_0.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Access-Level Metric for Green Radio

The point-to-point link-level metric can be extended into a multiple-user link, e.g., multiple access network, multiuser interference, and device-to-device communications. In these scenarios, each communication node/user has its own energy efficiency, which should be considered for the EE metric. There are mainly three well-established metrics to aggregate the different EEs. The global energy efficiency (GEE) is defined as

$$\text{GEE} = \frac{\sum_{k=1}^{K} R_k}{\sum_{k=1}^{K} \zeta P_{t,k} + P_{c,k}}, \tag{2.3}$$

in which K is the number of total links, and $R_k, P_{t,k}$, and $P_{c,k}$ refer to the data rate, transmit power, and circuit power of user k, respectively. This GEE is actually the overall energy efficiency of the whole network consisting of K links.

The GEE represents the ratio between the overall amount of bit and the overall energy consumption, and thus can be interpreted as the benefit–cost ratio of the entire network. However, the GEE does not consider the individual energy efficiency, leading to potential unfairness among different users' EE. The maximization of GEE does not always provide the maximum EE for each user.

On the other hand, weighted sum energy efficiency (WSEE) is defined as the weighted summation of all EEs, as

$$\text{WSEE} = \sum_{k=1}^{K} \omega_k \frac{R_k}{\zeta P_{t,k} + P_{c,k}}, \tag{2.4}$$

where ω_k is the weight for link k. Moreover, weighted minimum energy efficiency (WMEE) is defined as

$$\text{WMEE} = \min_{k=1,\ldots,K} \omega_k \frac{R_k}{\zeta P_{t,k} + P_{c,k}}. \tag{2.5}$$

The WSEE and the WMEE are capable of characterizing the complete Pareto-optimal EE region of users, as it can prioritize individual EE by varying the weight value, ω_k. In regard, the maximization of WSEE or WMEE makes a good balance, or trade-off, for the EE of each individual link.

Network-Level Metric for Green Radio

While considering the network level, there are some other green metrics that characterize network energy efficiency performance, and that involve the overall consumed power of the whole network and the system-level performance. Again, there are two kinds of metrics: energy consumption metrics and energy efficiency metrics.

The area energy consumption is defined as the average network power consumption divided by the network coverage (W/m^2). This metric takes into account all network power consumptions, including radio transmission power, fixed circuit power related to the operation system, cooling system, etc. Thus, it is more related to carbon dioxide

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

emissions and the carbon footprint. A counterpart area energy efficiency (AEE) metric can be defined as the average network coverage per consumed power (m^2/W), as

$$\text{AEE} = \frac{A}{P} \quad m^2/W, \tag{2.6}$$

where A is the coverage size and P is the overall consumed power. The AEE is more suitable for the rural environment than for the urban environment, since the network in the rural environment is usually coverage-limited.

In the urban environment where the network is capacity-limited, a more useful metric defined as the number of supported users per power unit can be used, which has a unit of users/W or W/user. With the densification of 5G small-cell base stations where the number of access points could be comparable to that of the associated users, the average power consumption per user becomes an important metric related to the operation cost, particularly the electricity bill, of cellular operators.

2.2 EE Study from Information Theory

Energy efficiency has been investigated from information theory since the very beginning. The capacity of a band-limited AWGN channel is given as

$$R = B \log_2 \left(1 + \frac{P_t}{N_0 B}\right), \tag{2.7}$$

where B is the bandwidth, P_t is the transmit power, and N_0 is the noise power density. In this section, without loss of generality, we assume that $N_0 = 1$. Spectral efficiency (SE) is defined as $\eta = R/B$, which means the amount of transmitted bits for a given unit bandwidth.

On the other hand, EE can be defined as the transmitted bit for one Joule energy, as $\epsilon = R/P_t$. Thus, the unit of EE is bit/Joule. Then, without considering circuit power consumption, the capacity region in (2.7) can be rewritten as

$$\epsilon = \frac{\eta}{2^\eta - 1}. \tag{2.8}$$

The above equation illustrates the SE–EE trade-off for point-to-point AWGN channel [3]. According to (2.8), EE monotonically decreases with SE, indicating that high SE and high EE cannot be simultaneously achieved in general. Moreover, we can also analyze that the maximum value of EE can be achieved at $(\ln 2)^{-1}$ when η approaches 0.

The relation of SE–EE trade-off can also be extended into multiuser channel scenario [4]. Multiple access channel (MAC) is one of the important multiuser channel models, which corresponds to the multiuser uplink communication in cellular networks. MAC is also the only channel model whose capacity region is already known.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

The capacity region for Gaussian MAC with K users can be expressed as a convex region restricted by the following inequalities [3]

$$\sum_{i\in\mathcal{S}} R_i \leq B\log_2\left(1+\frac{\sum_{i\in\mathcal{S}} P_i}{N_0 B}\right), \forall \mathcal{S}\subset\{1,2,\ldots,K\}, \tag{2.9}$$

where P_i and R_i are the transmit power and data rate of user i, respectively.

We now investigate the EE of Gaussian MAC, as well as the EE–SE trade-off. The above inequalities enclose the SE region, which characterizes the achievable data rate trade-off among different users. That is, each user cannot increase its own SE without degrading the SE of other users. Similarly, there also exists a fundamental EE trade-off among different users, i.e., the EE of one particular user will be generally decreased as EEs of other users increase. Therefore, to analyze the EE–SE trade-off, we shall first investigate the EE region of MAC.

Similar to the point-to-point AWGN channel, the SE and EE of user i can be defined as $\eta_i = R_i/B$ and $\epsilon_i = R_i/P_i$, respectively. Then, by substituting these definitions into the MAC capacity region, we can obtain the following expression

$$\sum_{i\in\mathcal{S}} \frac{\eta_i}{\epsilon_i} \geq 2^{\sum_{i\in\mathcal{S}}\eta_i} - 1, \forall \mathcal{S}\subset\{1,2,\ldots,K\}. \tag{2.10}$$

Both SE and EE are involved in (2.10). Therefore, two fundamental and significant insights can be observed from it. The first is the EE trade-off among different users, which is similar to the SE trade-off in (2.9). The second insight is the SE–EE trade-off for each user in the MAC. Considering a particular user k, its SE–EE trade-off is a function of all other users' SE and EE, i.e., η_i and ϵ_i, $i \neq k$.

Assuming $K = 2$, the EE–SE relation of user 1 can be derived as

$$\epsilon_1(\epsilon_2,\eta_1,\eta_2) = \begin{cases} \dfrac{\eta_1}{2^{\eta_1}-1}, & 0 \leq \epsilon_2 \leq \dfrac{\eta_2}{2^{\eta_1}(2^{\eta_2}-1)} \\ \dfrac{\eta_1}{2^{\eta_1+\eta_2}-(1+\frac{\eta_2}{\epsilon_2})}, & \dfrac{\eta_2}{2^{\eta_1}(2^{\eta_2}-1)} < \epsilon_2 \leq \dfrac{\eta_2}{2^{\eta_2}-1}. \end{cases} \tag{2.11}$$

With some simple mathematical analysis, we can derive that

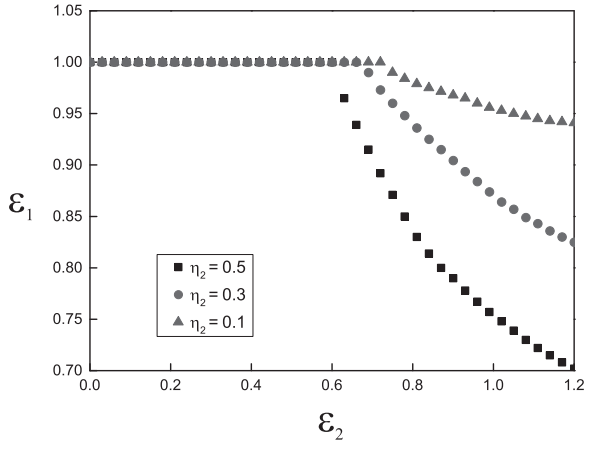
- ϵ_1 decreases with η_1, indicating the EE–SE trade-off of a particular user.
- ϵ_1 non-increases with ϵ_2 and η_2, which means the EE of one user will be generally degraded if the other user increases its EE or SE.

The SE–EE trade-off and the EE region of the two-user MAC are illustrated in Fig. 2.1.

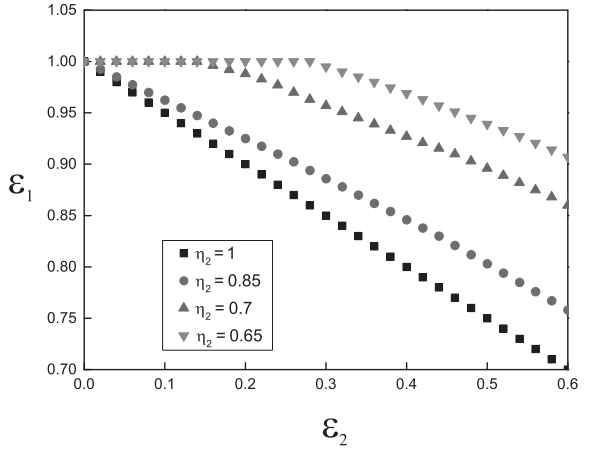
As we have discussed, the relation between EE and SE without considering circuit power can be expressed in closed form, for both point-to-point channel and MAC. In particular, the EE–SE curve is a cup shape curve for point-to-point transmission.

Whereas, considering circuit power, the relation between EE and SE is more complicated and cannot be expressed in closedform. It becomes a bell-shaped curve for point-to-point transmission, which will be discussed in detail later.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press



(a)



(b)

Figure 2.1 The EE trade-off and EE–SE trade-off of the two-user MAC, $\eta_1 = 1$. (a) The relationship between ϵ_1 and ϵ_2. (b) The relationship between ϵ_1 and η_2.

EE Region of TDMA and FDMA

EE regions in FDMA and TDMA systems are also attainable. In a FDMA system, we assume that the total bandwidth is B and the α_i proportion of the entire bandwidth is allocated for user i. Then the achievable rate region in FDMA system is given by [3]

$$\bigcup_{(\alpha_1,...,\alpha_M)} \left\{ (R_1,...,R_M) \,\middle|\, R_i \le \alpha_i B \log_2\left(1+\frac{P_i}{\alpha_i B N_0}\right), \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \ge 0, \forall i = 1,2,...M \right\}. \tag{2.12}$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

By substituting the definitions of SE and EE into (2.12), we can obtain the EE region for the FDMA system as

$$\bigcup_{(\alpha_1,...,\alpha_M)} \left\{ (\epsilon_1,...,\epsilon_M) \,\middle|\, \epsilon_i \le \frac{\eta_i}{\alpha_i \left(2^{\frac{\eta_i}{\alpha_i}} - 1\right)}, \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \ge 0, \forall i = 1,2,...M \right\}. \tag{2.13}$$

In a TDMA system, we assume that the total transmission time is T and $\alpha_i T$ is the time period allocated for each user. If we restrict the transmission power of each user in its own transmission period $\alpha_i T$ to be $P_{Ti} = P_i$, then the average achievable data rate in time period T can be expressed as

$$\bigcup_{(\alpha_1,...,\alpha_M)} \left\{ (R_1,...,R_M) \,\middle|\, R_i \le \alpha_i B \log_2\left(1 + \frac{P_i}{BN_0}\right), \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \ge 0, \forall i = 1,2,...M \right\}. \tag{2.14}$$

On the other hand, if we restrict the average transmission power of each user in the total transmission period T, that is $P_{Ti} = \frac{P_i}{\alpha_i}$, then the average achievable rate in T can be written as

$$\bigcup_{(\alpha_1,...,\alpha_M)} \left\{ (R_1,...,R_M) \,\middle|\, R_i \le \alpha_i B \log_2\left(1 + \frac{P_i}{\alpha_i BN_0}\right), \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \ge 0, \forall i = 1,2,...M \right\}. \tag{2.15}$$

For the first kind of power constraint, the definitions of SE and EE could be $\eta_i = \frac{R_i}{B}$ and $\epsilon_i = \frac{R_i}{\alpha_i P_i}$, respectively. For the second kind of power constraint, the definitions of SE and EE can be rewritten as $\eta_i = \frac{R_i}{B}$ and $\epsilon_i = \frac{R_i}{P_i}$, respectively.

By substituting the definitions of SE and EE into (2.14) and (2.15), we can obtain the EE expression in a TDMA system. However, under the two different constraints, the same expressions of the EE-region can be achieved, as

$$\bigcup_{(\alpha_1,...,\alpha_M)} \left\{ (\epsilon_1,...,\epsilon_M) \,\middle|\, \epsilon_i \le \frac{\eta_i}{\alpha_i \left(2^{\frac{\eta_i}{\alpha_i}} - 1\right)}, \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \ge 0, \forall i = 1,2,...M \right\}. \tag{2.16}$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

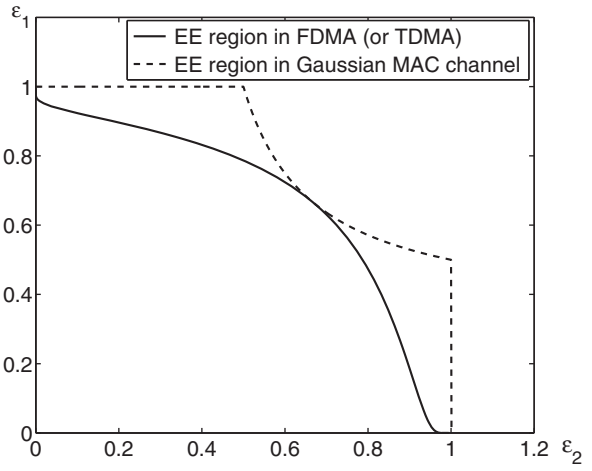


Figure 2.2 Achievable energy efficiency region in FDMA and TDMA systems, $\eta_1 = \eta_2 = 1$.

It is obvious that the EE-region of the TDMA system shown in (2.16) is the same as the EE-region of FDMA system shown in (2.13) assuming that the average power constraints in these systems are the same.

In Fig. 2.2, the achievable EE-regions in two-user FDMA and TDMA systems are shown and compared with the MAC, whose EE region is given in (2.10).

There is an intersection point of the two curves in Fig. 2.2. When $\alpha_i = \dfrac{\eta_i}{\eta_1 + \eta_2}, i = 1, 2$, the EE trade-off point in FDMA and TDMA is the same as that in MAC, that is, $\epsilon_1 = \epsilon_2 = \dfrac{\eta_1 + \eta_2}{2^{\eta_1+\eta_2} - 1}$. Recalling that there is also a point that the capacity region of FDMA or TDMA is the same to the MAC when $\alpha_i = \dfrac{P_i}{P_1 + P_2}$. Interestingly, this point is exactly the same as the intersection point in Fig. 2.2.

2.3 Fundamental EE–SE Trade-Off

The EE–SE trade-off can be expressed in closed form without considering the fixed circuit power. However, if the circuit power cannot be ignored in practical networks, it is difficult to analyze the EE–SE relation in a closed form. In this section, we use an example of single-cell downlink OFDMA network to build a general EE–SE trade-off framework. We also demonstrate that EE is quasiconcave in SE and discuss the potential impact of channel power gain and circuit power consumption on the EE–SE trade-off.

Considering a single-cell downlink OFDMA networks with K active users and a total bandwidth B_t, which is equally divided into N subcarriers. Denote $\mathcal{K} = \{1, 2, \ldots, K\}$ as

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

the set of K active users and $\mathcal{N} = \{1, 2, \ldots, N\}$ as the set of N subcarriers, each with a bandwidth of B. Let $p_{k,n}$ and $g_{k,n}$ be the transmit power and the channel power gain of user k ($k \in \mathcal{K}$) on subcarrier n ($n \in \mathcal{N}$), respectively. Then, for an AWGN channel with noise power density N_0, the channel capacity of user k on subcarrier n can be written as

$$r_{k,n} = B \log_2\left(1 + \frac{p_{k,n} g_{k,n}}{N_0 B}\right). \tag{2.17}$$

The overall system throughput and the total transmit power can be expressed as

$$\begin{aligned} R &= \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} r_{k,n}, \\ P_t &= \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n}, \end{aligned} \tag{2.18}$$

respectively. In the above, $\rho_{k,n} \in \{1, 0\}$ is the subcarrier allocation indicator. We let $\rho_{k,n} = 1$ if subcarrier n is allocated to user k; $\rho_{k,n} = 1$ otherwise. The overall transmit power at the base station is constrained as $P_{\max}$, that is $P_t \leq P_{\max}$. Moreover, as in many practical OFDMA systems, we assume that each subcarrier can only be used by at most one user to guarantee the orthogonality, i.e., $\sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall\, n \in \mathcal{N}$.

Let the set $\boldsymbol{\rho} = [\rho_{k,n}]_{K \times N}$ denote the feasible subcarrier assignment indicator matrix and the set $\boldsymbol{P} = [p_{k,n}]_{K \times N}$ denote the feasible power allocation matrix, which can be expressed as

$$\begin{aligned} \boldsymbol{\rho} \in \varrho \overset{\text{def}}{=} \Big\{ [\rho_{k,n}]_{K \times N} \big| & \sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall\, n \in \mathcal{N}; \\ & \rho_{k,n} = \{0, 1\}, \forall\, k \in \mathcal{K}, n \in \mathcal{N} \Big\}, \\ \boldsymbol{P} \in \mathcal{P} \overset{\text{def}}{=} \Big\{ [p_{k,n}]_{K \times N} \big| & p_{k,n} \geq 0, \forall\, k \in \mathcal{K}, \forall\, n \in \mathcal{N}; \\ & \sum_{k=1}^{K} \sum_{n=1}^{N} p_{k,n} \leq P_{\max} \Big\}, \end{aligned} \tag{2.19}$$

respectively.

To analyze the EE–SE trade-off, it is equivalent to maximize the EE for a given SE requirement, or data rate requirement. Let $\check{R}_k$ denote the data rate requirement of user k. As in practical LTE networks, we assume that there are two kinds of users: real-time users and non-real-time users. Let $\mathcal{K}_1 = \{1, 2, \ldots, K_0 - 1\}$ denote the set of $K_0 - 1$ ($K_0 \geq 1$) real-time users and $\mathcal{K}_2 = \{K_0, K_0+1, \ldots, K\}$ represent the set the remaining $K - K_0 + 1$ non-real-time users. The real-time users have a fixed data rate requirement, which is equal to $\check{R}_k, \forall k \in \mathcal{K}_1$, while the data rate requirement of the non-real-time users should be greater than $\check{R}_k, \forall k \in \mathcal{K}_2$.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

We can now mathematically formulate the EE maximization problem as

$$\max_{\boldsymbol{\rho}\in\mathcal{Q},\,\boldsymbol{P}\in\mathcal{P}} \ \eta_{\mathrm{EE}} \left(= \frac{\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_{k,n}r_{k,n}}{\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_{k,n}\left(\zeta p_{k,n}+\xi r_{k,n}\right)+P_s} \right), \tag{2.20}$$

subject to

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_{k,n}r_{k,n} \geq \breve{R}, \tag{2.21}$$

$$\sum_{n=1}^{N}\rho_{k,n}r_{k,n} = \breve{R}_k, \forall\, k \in \mathcal{K}_1, \tag{2.22}$$

$$\sum_{n=1}^{N}\rho_{k,n}r_{k,n} \geq \breve{R}_k, \forall\, k \in \mathcal{K}_2. \tag{2.23}$$

In the above, we can further assume that $\breve{R} \geq \sum_{k=1}^{K}\breve{R}_k$, indicating that the overall data rate requirement is no less than the summation of each user's data rate requirement.

2.3.1 EE–SE Relation

In the following discussion, we will study the fundamental EE–SE relation of the system in (2.21)–(2.23). By solving the problem in (2.20) for a given $\breve{R}$, we can obtain the optimal EE as a function of SE. However it is impossible to express the EE function as a closed-form function due to the complicated optimization problem. Nevertheless, we can reveal some insightful properties of the EE function. In what follows, we first demonstrate that the EE is a quasiconcave function in SE. In addition, we also discuss how channel power gain and fixed circuit power consumption impact the EE–SE trade-off.

Assuming the number of subcarriers, N, is sufficiently large, the quasiconcavity of the EE function $\eta_{\mathrm{EE}}(\boldsymbol{R})$ can be presented in the following theorem. Interested readers can refer to [5] for the detailed proof.

THEOREM 1 *For any achievable data rate vector, $\boldsymbol{R} = [R_k]_{K\times 1}$, achieved with a feasible subcarrier allocation and power allocation, the maximum EE, $\eta^*_{\mathrm{EE}}(\boldsymbol{R}) = \max_{\boldsymbol{\rho}\in\mathcal{Q},\, p_{k,n}\geq 0} \eta_{\mathrm{EE}}(\boldsymbol{R})$, is strictly quasiconcave in $\boldsymbol{R}$ given sufficiently large number of subcarriers. Moreover, in the SE region $\left[\frac{\breve{R}}{B_t}, \frac{\hat{R}}{B_t}\right]$, $\eta^*_{\mathrm{EE}}\left(\eta_{\mathrm{SE}}\right)$ has the following monotonic properties.*

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

(a) It strictly decreases with η_{SE} *and achieves its maximum at* $\eta_{\mathrm{SE}} = \frac{\check{R}}{B_t}$ *if*

$$\left.\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}\left(\eta_{\mathrm{SE}}\right)}{\mathrm{d}\eta_{\mathrm{SE}}}\right|_{\eta_{\mathrm{SE}}=\frac{\check{R}}{B_t}} \leq 0,$$

(b) It strictly increases with η_{SE} *and achieves its maximum at* $\eta_{\mathrm{SE}} = \frac{\hat{R}}{B_t}$ *if*

$$\left.\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}\left(\eta_{\mathrm{SE}}\right)}{\mathrm{d}\eta_{\mathrm{SE}}}\right|_{\eta_{\mathrm{SE}}=\frac{\check{R}}{B_t}} > 0 \text{ and } \left.\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}\left(\eta_{\mathrm{SE}}\right)}{\mathrm{d}\eta_{\mathrm{SE}}}\right|_{\eta_{\mathrm{SE}}=\frac{\hat{R}}{B_t}} \geq 0,$$

(c) It first strictly increases and then strictly decreases with η_{SE} *and achieves its maximum at* $\eta_{\mathrm{SE}} = \frac{R_{\mathrm{EE,\,max}}}{B_t}$ *if*

$$\left.\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}\left(\eta_{\mathrm{SE}}\right)}{\mathrm{d}\eta_{\mathrm{SE}}}\right|_{\eta_{\mathrm{SE}}=\frac{\check{R}}{B_t}} > 0 \text{ and } \left.\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}\left(\eta_{\mathrm{SE}}\right)}{\mathrm{d}\eta_{\mathrm{SE}}}\right|_{\eta_{\mathrm{SE}}=\frac{\hat{R}}{B_t}} < 0.$$

Here, $\hat{R}$ *is the maximum throughput and* $R_{\mathrm{EE,\,max}}$ *is the throughput that corresponds to the maximum EE,* $\eta_{\mathrm{EE}}^{\max}$*, under all constraints except the peak transmit power constraint in problem* (2.20)*.*

This theorem not only shows that the EE is a quasiconcave function of the overall SE but also presents an effective way to achieve the optimal EE–SE trade-off based on the quasiconcavity. According to [6, ch. 8], a unique global optimum always exists for any continuous and strictly quasiconcave function. Therefore, a unique and global optimal EE of the problem in (2.20) can always be achieved by this theorem.

The η_{EE}^{*}-versus-η_{SE} curve is referred to as the EE–SE trade-off curve in the sequel. In Fig. 2.3, we further illustrate the detailed EE–SE trade-off curves in the three possible cases in Theorem 1, as following.

- Case A (condition a): the EE decreases with SE for the entire feasible region where the optimal EE is achieved at the minimum achievable rate point.
- Case B (condition b): the EE is an increasing function of SE where the optimal EE is achieved at the maximum achievable rate point.
- Case C (condition c): the EE first increases and then decreases with SE in the feasible SE region where the maximum EE is achieved at the stationary point.

As discussed earlier, both EE and SE are related to the channel power gain and the fixed circuit power consumption. Therefore, it is important to investigate the impact of channel power gain and the circuit power consumption on the EE–SE trade-off. We present the following three properties.

Property 1. Given SE, the EE is a non-decreasing function of the channel power gain, $g_{k,n}$, and a strictly decreasing function of the circuit power consumption, P_c.

This property is rather intuitive since larger channel power gain leads to better channel capacity given a fixed transmit power, whereas larger circuit power consumption results in the degradation of EE. According to this property, it is better to schedule users with

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

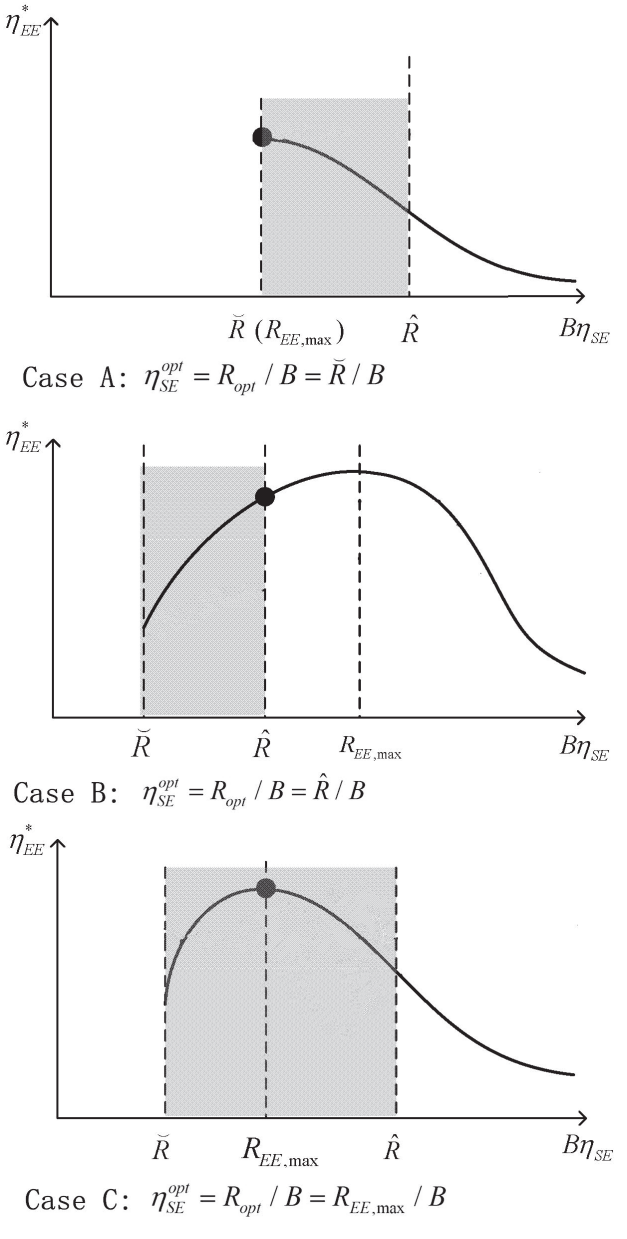


Figure 2.3 EE–SE relation in downlink OFDMA.

good channel quality, such as those users near the base station, for an improved EE–SE trade-off.

Property 2. The optimal SE, defined as $\eta_{\mathrm{SE}}^{\mathrm{opt}} \overset{\mathrm{def}}{=} \frac{R_{\mathrm{opt}}}{B_t}$, is a non-decreasing function of the static circuit power consumption, P_s. The maximum SE, defined as

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

$\eta_{\mathrm{SE}}^{\max} \overset{\mathrm{def}}{=} \frac{R_{\mathrm{EE,\,max}}}{B_t}$, is a strictly increasing function of P_s. On the other hand, both $\eta_{\mathrm{SE}}^{\mathrm{opt}}$ and $\eta_{\mathrm{SE}}^{\max}$ are independent of the dynamic circuit power consumption rate, ξ.

The independence of ξ on the maximal or optimal SE is interesting and a little bit unexpected. On some occasions if ξ or the static circuit power consumption, P_s, is large, a high SE may be still achievable despite the low EE performance. This leads to a relatively low and flat EE–SE curve, i.e., EE is insensitive to the change of SE. However, in this case, the locations of $\eta_{\mathrm{SE}}^{\mathrm{opt}}$ and $\eta_{\mathrm{SE}}^{\max}$ are not impacted.

Property 3. Even for the case of very small circuit power consumption, i.e., $P_c \approx 0$, the optimal EE, $\eta_{\mathrm{EE}}^{\mathrm{opt}}$, is not necessarily achieved when the SE is minimized, i.e., $\eta_{\mathrm{SE}} = \frac{\check{R}}{B_t}$.

The property 3 is a little bit counter-intuitive as the optimal EE is always achieved when the data rate is minimized in the single user point-to-point link. This property also indicates that even if the transmit power dominates the power consumption, the most energy-efficient communication scheme is not necessarily the least spectral-efficient. In this regard, it is possible for us to design both spectral-efficient and energy-efficient communication simultaneously.

2.3.2 Bounds on the EE–SE Curve

In the previous section, we demonstrated that EE is a quasiconcave function of SE, and introduced some insightful properties related to the EE–SE trade-off. In this section, we further analyze the upper and the lower bounds of the EE–SE trade-off curve.

From Theorem 1, the optimal EE–SE trade-off can be obtained by solving $\eta_{\mathrm{EE}}^{*}(\eta_{\mathrm{SE}})$ and $\frac{\mathrm{d}\eta_{\mathrm{EE}}^{*}(\eta_{\mathrm{SE}})}{\mathrm{d}\eta_{\mathrm{SE}}}$. However, the exact and closed-form solution to (2.20) is rather difficult to achieve. In what follows, we apply the Lagrange dual decomposition (LDD) to approximately approach it. The LDD is an effective method to solve this kind of problems with a good accuracy and reasonable computational complexity.

The Lagrange dual problem of minimizing the total transmit power consumption for a given data rate requirement, $R \geq \check{R}$, can be expressed as

$$\max_{\substack{\boldsymbol{\lambda}_1 \succeq \mathbf{0}, \boldsymbol{\lambda}_2 \succeq \mathbf{0}, \\ \lambda_3 \geq 0}} \min_{\boldsymbol{\rho} \in \mathcal{Q},\, \boldsymbol{P} \in \mathcal{P}} \left\{ \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n} + \sum_{n=1}^{N} \lambda_{1,n} \left(\sum_{k=1}^{K} \rho_{k,n} - \right) \right.$$
$$\left. + \sum_{k=1}^{K} \lambda_{2,k} \left(\check{R}_k - \sum_{n=1}^{N} \rho_{k,n} r_{k,n} \right) + \lambda_3 \left(R - \sum_{k=1}^{K_0 - 1} \check{R}_k - \sum_{k=K_0}^{K} \sum_{n=1}^{N} \rho_{k,n} r_{k,n} \right) \right\}, \tag{2.24}$$

where $\boldsymbol{\lambda}_1 = [\lambda_{1,1}, \lambda_{1,2}, \ldots, \lambda_{1,N}]^T$, $\boldsymbol{\lambda}_2 = [\lambda_{2,1}, \lambda_{2,2}, \ldots, \lambda_{2,K}]^T$, and λ_3 are the Lagrange multipliers related to the corresponding constraints, and $\succeq$ denotes the component-wise inequality.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Consequently, the problem in (2.20) can be decomposed into two layers. The inner layer solves the subordinate problem for each subcarrier n, which can be formulated as

$$U_n = \min_{\boldsymbol{\rho}\in\mathcal{Q},\, \boldsymbol{P}\in\mathcal{P}} \sum_{k=1}^{K} \rho_{k,n} u_{k,n},\ n \in \mathcal{N}, \tag{2.25}$$

subject to

$$\begin{aligned} u_{k,n} &= p_{k,n} + \lambda_{1,n} - \lambda_{2,k} r_{k,n},\ \text{if } k \in \mathcal{K}_1, \\ u_{k,n} &= p_{k,n} + \lambda_{1,n} - (\lambda_{2,k} + \lambda_3) r_{k,n},\ \text{if } k \in \mathcal{K}_2. \end{aligned} \tag{2.26}$$

The outer layer is for the master problem and can be written as

$$\max_{\substack{\boldsymbol{\lambda}_1\succeq\mathbf{0},\boldsymbol{\lambda}_2\succeq\mathbf{0},\\ \lambda_3\geq 0}} \left\{ \sum_{n=1}^{N} U_n - \sum_{n=1}^{N} \lambda_{1,n} + \sum_{k=1}^{K} \lambda_{2,k} \check{R}_k + \lambda_3 \left(R - \sum_{k=1}^{K_0-1} \check{R}_k \right) \right\}. \tag{2.27}$$

The problems in (2.25) and (2.27) can be iteratively solved. In the inner layer, if $\rho_{k,n} = 1$, the optimal transmit power can be expressed similarly in a water-filling way, as

$$\begin{aligned} p_{k,n} &= B \left[\frac{\lambda_{2,k}}{\ln 2} - \frac{N_0}{g_{k,n}} \right]^+,\ \text{if } k \in \mathcal{K}_1, \\ p_{k,n} &= B \left[\frac{(\lambda_{2,k} + \lambda_3)}{\ln 2} - \frac{N_0}{g_{k,n}} \right]^+,\ \text{if } k \in \mathcal{K}_2, \end{aligned} \tag{2.28}$$

where $[x]^+ = \max(x, 0)$. The above equation should be solved for all $k \in \mathcal{K}$. After that, the optimal transmit power, $p_{k,n}$'s, should be substituted back into (2.26), and $\rho_{k,n}$'s should be set to 1 for the user with the minimum $u_{k,n}$ and 0 for all other users.

Moreover, the Lagrange multipliers can be updated by the subgradient method as

$$\lambda_{2,k}^{(i+1)} = \left[\lambda_{2,k}^{(i)} - s^{(i)} \left(\sum_{n=1}^{N} \rho_{k,n}^{(i)} r_{k,n}^{(i)} - \check{R}_k \right) \right]^+, \tag{2.29a}$$

$$\lambda_3^{(i+1)} = \left[\lambda_3^{(i)} - s^{(i)} \left(\sum_{k=K_0}^{K} \sum_{n=1}^{N} \rho_{k,n}^{(i)} r_{k,n}^{(i)} - \left(R - \sum_{k=1}^{K_0-1} \check{R}_k \right) \right) \right]^+, \tag{2.29b}$$

in which $s^{(i)}$ is a sufficiently small stepsize for the ith iteration. We shall note that the duality gap between the original problem and the LDD problem would not be zero due to the non-convex nature of the problem. However, the duality gap approaches zero when the number of subcarriers becomes large enough [7]. In practice, the duality gap is already very small for the case with $N = 64$ subcarriers.

Until now, the subcarrier allocation strategy has been derived based on the LDD method, as discussed earlier. In the next step, we shall consequently determine transmit power allocation, and derive upper and lower bounds, respectively.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Upper-Bound Power Allocation Strategy

For a given subcarrier allocation result, $\boldsymbol{\rho} = [\rho_{k,n}]_{K,N}$, from the LDD method, we design a two-stage power allocation strategy. In the first stage, transmit power is allocated by the water-filling method for each user to fulfill its own data rate requirement, $\check{R}_k$. In the second stage, the extra power can be allocated among the subcarriers of all non-real-time users also by the water-filling method until the data rate reaches R.

It is very straightforward that this power allocation strategy achieves an upper bound of the transmit power, which also corresponds to the upper bound of the EE–SE trade-off curve.

Lower-Bound Power Allocation Strategy

A lower bound on the minimum transmit power can be achieved by relaxing the binary variables, $\rho_{k,n}$'s, into continuous real variables within $[0, 1]$, which can be also interpreted as time sharing of the subcarrier. With this manipulation, the data rate expression can be rewritten as

$$r_{k,n} = B \log_2 \left(1 + \frac{p_{k,n} g_{k,n}}{\rho_{k,n} B N_0} \right). \tag{2.30}$$

Then, the original problem in (2.20) can be transformed into a convex optimization problem and then be solved by the standard methods [8]. Since the binary constraint on subcarrier allocation has been relaxed, the result from this method can serve as a lower bound of the total transmit power, which also corresponds to the lower bound of the EE–SE trade-off curve.

An example of EE–SE trade-off curve is depicted in Fig. 2.4. In the example, there are 72 OFDM subcarriers and each subcarrier has a bandwidth of 15 kHz. The frequency-selective Rayleigh fading is according to the ITU Pedestrian-B model and has an identical average channel-gain-to-noise ratio of 20 dB. The noise power density is -174 dBm/Hz. There are two real-time users and four non-real-time users. It is assumed that the data rate requirement for each real-time user is 200 kbps and the minimum data rate requirement for each non-real-time user is 50 kbps. The power amplifier efficiency is set to be 38%. From the figure, it is clear that the EE–SE trade-off has a bell-shaped curve, which is also quasiconcave. Figure 2.4 also shows that the maximum EE point decreases while the maximum SE point increases with the static circuit power consumption. However, the latter is almost independent to the dynamic circuit power factor. These results demonstrate the effectiveness of the analysis in this section.

We further plot the optimal EE with different SE in Fig. 2.5. The results therein indicate that the optimal EE design also leads to the same throughput when the minimum data rate requirement of each user is no greater than a given threshold. However, when the minimum data rate requirement is greater than the threshold, it is better to operate at the minimum data rate point to maximize the system EE.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

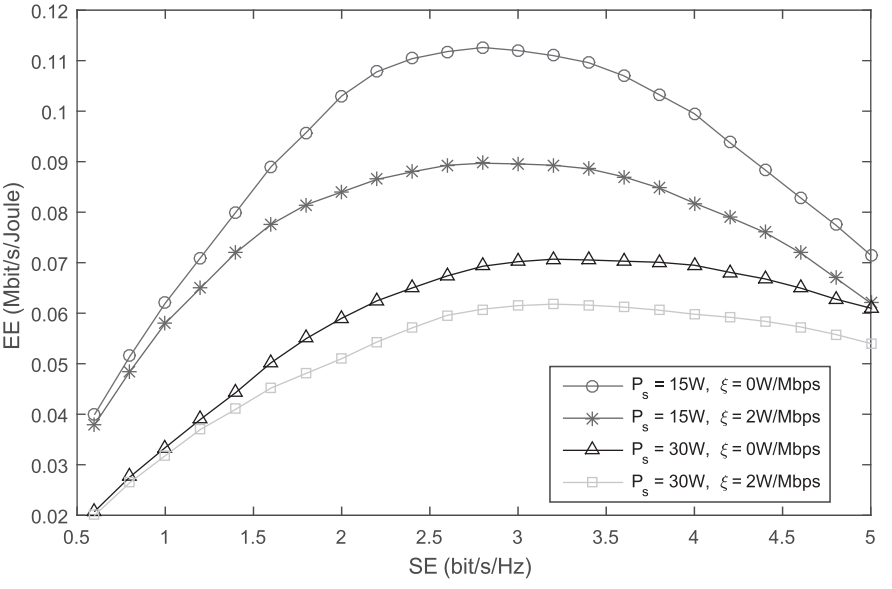


Figure 2.4 The EE–SE trade-off.

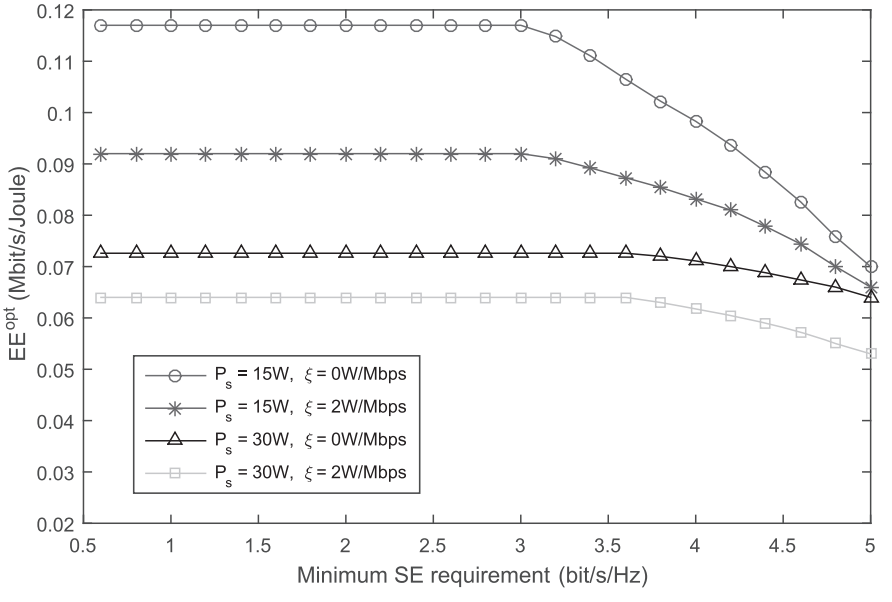


Figure 2.5 The optimal EE versus SE.

2.3.3 Further Discussion

In this section, we have discussed the EE–SE trade-off with the consideration of circuit power consumption. The fundamental results introduced here can be applied to many

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

other different network models and scenarios. In [9], the EE–SE trade-off in downlink OFDMA networks has been further investigated by taking into account the user fairness. It shows that a certain EE degradation should be compromised as a cost to ensure users' rate fairness. The EE–SE trade-off has also been extended into the amplify-and-forward-based relay network, and the result shows that the EE–SE trade-off curve is also quasiconcave in this scenario [10]. In [11, 12], the EE–SE trade-off in type-I ARQ system and cognitive radio has been studied, respectively. The EE–SE study in homogeneous cellular networks with random distributed base stations shows that, with respect to the outage constraint, the EE–SE trade-off only occurs under a given network situation [13].

In [14, 15], an alternative multi-objective optimization approach has been applied to analyze the EE–SE trade-off. The quasiconcave EE–SE relation is also revealed in these works. Moreover, a practical theoretical framework for analyzing the EE–SE trade-off in single-cell cellular networks to achieve tractable results has been developed in [16]. Leveraging the stochastic geometry method, the framework has also been extended to the multicell scenario with the presence of intercell interference. The EE–SE trade-off and the corresponding upper and lower bounds have also been investigated for video transmission over mobile ad hoc networks in [17].

In addition to the EE–SE trade-off in practical networks, there are also three other fundamental trade-offs regarding the green radio design [18].

- Deployment efficiency (DE)–EE trade-off: The EE can be increased by shrinking the cell radius. However, higher density of base stations would certainly lead to increased deployment complexity, or decreased DE. Moreover, more deployment of base stations also results in larger circuit power consumption. With this regard, the energy-efficient design should also be in coordination with DE.
- Bandwidth-power trade-off: In traditional SE design, the relation between bandwidth and power is monotonic. However, considering EE design when circuit power consumption scales with the transmission bandwidth, their relation is non-monotonic. To this end, the optimal bandwidth and power trade-off should be revisited in the EE oriented design.
- Delay–EE trade-off: Without considering fixed-circuit power, the relation between per-bit power and packet delay is monotonically decreasing. However, as the packet delay increases, more circuit power consumption would be introduced. With this regard, the delay–EE relation would be no longer monotonic, and there would exist an optimal delay–EE trade-off.

2.4 EE Design in Orthogonal Systems

In the previous section, we introduced the EE–SE relation of an OFDMA network. In this section, we take an uplink OFDMA network as an example to show how to maximize the overall EE, summation of EE, and minimal EE of orthogonal systems. Through this section, we try to illustrate the problem formulation, mathematical analysis, and

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

algorithm development for practical energy-efficient wireless resource allocation. Note that the detailed mathematical optimization theory is illustrated in Appendix 2.1.

Similar to Section 2.3, we also consider a single-cell uplink OFDMA system with K users and N subcarriers, each of which has a bandwidth of B Hz. In OFDMA systems, each user can transmit on several subcarriers, whereas each subcarrier can only be occupied by one user to ensure orthogonality.

The overall transmit rate of user k can be written as

$$R_k = \sum_{n=1}^{N} R_{k,n} = \sum_{n=1}^{N} \rho_{k,n} B \log_2 \left(1 + \frac{p_{k,n} h_{k,n}}{\rho_{k,n} B N_0}\right), \rho_{k,n} \in \{0,1\}. \tag{2.31}$$

In the above, $p_{k,n}$ denotes the transmit power and $\rho_{k,n}$ stands for the binary channel allocation indicator.

In the OFDMA network, the overall power consumption of each user can be expressed as

$$P_k = \sum_{n=1}^{N} \left(\zeta p_{k,n} + \rho_{k,n} P_{\mathrm{e}}\right) + P_{\mathrm{fix}}. \tag{2.32}$$

Here, we use P_{e} to denote the radio frequency (RF) circuit power consumption per used subcarrier and P_{fix} to denote the fixed power consumption irrespective of the number of used subcarriers. Note that the overall circuit power $P_c \overset{\mathrm{def}}{=} \sum_{n=1}^{N} \rho_{k,n} P_{\mathrm{e}} + P_{\mathrm{fix}}$ is consistent with the definition in (2.2).

The EE for each individual user, say, user k, can be expressed as

$$\eta_k = \frac{R_k}{P_k}. \tag{2.33}$$

Instead of the GEE in Section 2.3, we aim at maximizing the WSEE or WMEE in this section. Since the WSEE or WMEE is related to the EE of each individual user, this objective function has the merit of providing better insight on the EE trade-off among users. Users in practical cellular networks may have different priorities as well as different status of battery levels. Through maximizing the WSEE or WMEE, one can allocate more wireless resources to those users with higher priorities or lower battery levels by setting different weight values to different users. In this way, higher EEs and better quality of experience (QoE) could be achieved by those users. Moreover, we can also provide fairness among users by maximizing individual EE rather than maximizing the overall system EE or GEE.

2.4.1 Weighted Summation EE Maximization

Based on the discussion in the previous section, the maximization of WSEE can be mathematically defined as

$$\eta_{\mathrm{ws}}^{\max} = \max_{(\boldsymbol{\rho}, \mathbf{p}) \in \boldsymbol{\Theta}} \left\{ \sum_{k \in \mathcal{K}} \epsilon_k \eta_k \right\}, \tag{2.34}$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

subject to

$$\rho_{k,n} \in \{0,1\}, \forall k,n, \tag{2.35}$$

$$\sum_{k=1}^{K} \rho_{k,n} \leqslant 1, \forall n, \tag{2.36}$$

$$\sum_{n=1}^{N} p_{k,n} \leqslant P_{\max}, \forall k, \tag{2.37}$$

$$p_{k,n} \geqslant 0, \forall k,n, \tag{2.38}$$

$$R_k \geqslant R_{\min}, \forall k. \tag{2.39}$$

In the above, ϵ_k is the weight for user k, $\boldsymbol{\Theta}$ denotes the feasible subcarrier and power allocation set, (2.35) is the constraint of binary subcarrier allocation, (2.36) limits that each subcarrier can only be allocated to at most one user, (2.37) and (2.38) are the transmit power constraints of each user, and (2.39) is the QoS requirement.

We shall note that the WSEE maximization problem might be infeasible due to the potential conflicts of user QoS requirement and the maximum transmit power limitation. In this way, some users should be dropped through the admission control strategy. Here, admission control is not considered, and we assume that the problem is always feasible and try to find some effective algorithms to approach it.

We notice that the WSEE optimization problem aims to maximize the summation of several fractional functions, and thus is a sum-of-ratios optimization problem, which can be solved by the method introduced in Appendix 2.1. However, due to the binary subcarrier allocation indicator, $\rho_{k,n} \in \{0,1\}, \forall k,n$, the problem cannot be directly solved. To make it more tractable, we first relax $\rho_{k,n}$ into continuous variables within [0,1]. After that, the problem can be rewritten as

$$\tilde{\eta}_{\text{ws}}^{\max} = \max_{(\tilde{\boldsymbol{\rho}},\mathbf{p})\in\tilde{\Theta}} \left\{ \sum_{k=1}^{K} \epsilon_k \tilde{\eta}_k \right\} = \max_{(\tilde{\boldsymbol{\rho}},\mathbf{p})\in\tilde{\Theta}} \left\{ \sum_{k=1}^{K} \epsilon_k \frac{\tilde{R}_k}{\tilde{P}_k} \right\}, \tag{2.40}$$

where $\tilde{R}_k$ and $\tilde{P}_k$ are the data rate and power consumption after the binary subcarrier allocation $\rho_{k,n}$ is replaced with $\tilde{\rho}_{k,n}$, respectively, and $\tilde{\boldsymbol{\Theta}}$ is the feasible subcarrier allocation and power control region with constraint (2.35) being replaced by

$$0 \leqslant \tilde{\rho}_{k,n} \leqslant 1, \forall k,n. \tag{2.41}$$

We shall note that since the original constraint (2.35) has been relaxed, the solution to (2.40) serves as an upper bound of the original problem in (2.34), that is $\tilde{\eta}_{\text{ws}}^{\max} \geq \eta_{\text{ws}}^{\max}$, since $\boldsymbol{\Theta} \subset \tilde{\boldsymbol{\Theta}}$.

To tackle the problem, we first present the following important property.

PROPOSITION 1 *The generalized EE,* $\tilde{\eta}_k(\tilde{\boldsymbol{\rho}},\mathbf{p}) = \dfrac{\tilde{R}_k}{\tilde{P}_k}, \forall k$, *is jointly quasiconcave in variables* $\tilde{\rho}_{k,n}$ *and* $p_{k,n}, \forall k,n$.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Proof: We will first show that $\tilde{R}_{k,n}$ is a concave function over $\tilde{\rho}_{k,n}$ and $p_{k,n}$. Define $x \triangleq B\tilde{\rho}_{k,n}, y \triangleq \frac{p_{k,n}h_{k,n}}{N_0}$, and $f(x,y) = -x\log_2\left(1+\frac{y}{x}\right)$, then $\tilde{R}_{k,n} = -f(x,y)$. The Hessian of $f(x,y)$ is

$$\mathbf{H} = \begin{bmatrix} \frac{y^2/x}{(x+y)^2} & -\frac{y}{(x+y)^2} \\ -\frac{y}{(x+y)^2} & \frac{x}{(x+y)^2} \end{bmatrix}, \tag{2.42}$$

which is positive semi-defined, since the eigenvalues of $\mathbf{H}$ are

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_2 &= \frac{x^2+y^2}{x^3+2x^2y+xy^2} \geq 0. \end{aligned}$$

Therefore $\tilde{R}_{k,n}$ is concave. Clearly, $\tilde{R}_k$ is concave since it is a linear combination of $\tilde{R}_{k,n}$. We further define $\bar{\eta}_k(\tilde{\boldsymbol{\rho}}, \mathbf{p}) = -\tilde{\eta}_k(\tilde{\boldsymbol{\rho}}, \mathbf{p}) = -\frac{\tilde{R}_k}{\tilde{P}_k}$ and its sublevel sets as

$$\tau_\alpha = \left\{\tilde{\rho}_{k,n} \geqslant 0, p_{k,n} \geqslant 0, \forall n | \bar{\eta}_k(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \leq \alpha\right\}, \tag{2.43}$$

which equals

$$\tau_\alpha = \left\{\tilde{\rho}_{k,n} \geqslant 0, p_{k,n} \geqslant 0, \forall n | -\alpha\tilde{P}_k - \tilde{R}_k \leqslant 0\right\}. \tag{2.44}$$

We see that τ_α is convex due to the convexity of $-\alpha\tilde{P}_k - \tilde{R}_k$, which leads to the quasiconcavity of $\tilde{\eta}_k(\tilde{\boldsymbol{\rho}}, \mathbf{p})$. This ends the proof. □

According to this property, we can equivalently transform the problem into

$$\max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}, \psi} \left\{\sum_{k=1}^{K} \psi_k\right\}, \tag{2.45}$$

subject to

$$\psi_k \leq \epsilon_k \frac{\tilde{R}_k}{\tilde{P}_k}, \forall k. \tag{2.46}$$

Then, based on the sum-of-ratios optimization method (please see the details in Appendix 2.1), we have the following theorem.

THEOREM 2 *If* $\left(\tilde{\boldsymbol{\rho}}^*, \mathbf{p}^*, \boldsymbol{\Psi}^*\right)$ *is the optimal solution to* (2.45)*, then there exists* $\boldsymbol{\kappa}^* = \left(\kappa_1^*, \kappa_2^*, \ldots, \kappa_K^*\right)$*, such that* $\left(\tilde{\boldsymbol{\rho}}^*, \mathbf{p}^*\right)$ *is the optimal solution to the following problem for* $\boldsymbol{\kappa} = \boldsymbol{\kappa}^*$ *and* $\boldsymbol{\Psi} = \boldsymbol{\Psi}^*$

$$\max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \left\{\sum_{k=1}^{K} \kappa_k \left(\epsilon_k \tilde{R}_k - \psi_k \tilde{P}_k\right)\right\}. \tag{2.47}$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

And $(\widetilde{\boldsymbol{\rho}}^*, \mathbf{p}^*)$ *also satisfies the following system of equations for* $\boldsymbol{\kappa} = \boldsymbol{\kappa}^*$ *and* $\boldsymbol{\Psi} = \boldsymbol{\Psi}^*$*:*

$$\kappa_k = \frac{1}{\tilde{P}_k}, \forall k, \tag{2.48}$$

$$\epsilon_k \tilde{R}_k - \psi_k \tilde{P}_k = 0, \forall k. \tag{2.49}$$

According to this theorem, to solve the problem in (2.40) is equivalent to solving the problem in (2.47). Moreover, the latter can be iteratively solved by the following two nested steps: the inner step solves $\left(\tilde{\boldsymbol{\rho}}^*, \mathbf{p}^*\right)$ for given $(\boldsymbol{\kappa}, \Psi)$, and the outer step updates the parameter $(\boldsymbol{\kappa}, \Psi)$ satisfying (2.48) and (2.49).

We now discuss the optimality of the algorithm. As shown in [24], the sum-of-ratios algorithm achieves at least a KKT point of the original problem, which means at least a local optimal solution can be obtained. In addition, as we have proved that the generalized EE is quasiconcave, the local optimal solution is also the global optimal one.

Although the global optimum of the problem in (2.40) can be obtained, the proposed algorithm is not exactly the optimal one due to the continuous relaxation of the binary channel allocation indicators. That is, the resulting optimal solution does not necessarily guarantee that $\rho_{k,n}$'s are binary variables. In what follows, we shall develop a sub optimal algorithm to approach the original problem based on the LDD method and take into account the binary variables as well.

The suboptimal solution is based on the resulting Lagrange parameters, κ_k^* and ψ_k^*, $\forall k$, in the upper-bound algorithm. Once κ_k^* and ψ_k^*, $\forall k$ are obtained, the objective function for the suboptimal problem can be redefined as

$$\max_{(\boldsymbol{\rho}, \mathbf{p}) \in \boldsymbol{\Theta}} \sum_{k=1}^{K} \kappa_k^* \left(\epsilon_k R_k - \psi_k^* P_k\right). \tag{2.50}$$

In the above, $\boldsymbol{\rho}$ is now a binary variable.

We can rewrite the Lagrangian function of (2.50) as

$$\begin{aligned} L\left(\boldsymbol{\rho}, \boldsymbol{p}, \boldsymbol{\lambda}\right) = & \sum_{k=1}^{K} \kappa_k^* \left(\epsilon_k R_k - \psi_k^* P_k\right) + \sum_{n=1}^{N} \lambda_n^{(1)} \left(1 - \sum_{k=1}^{K} \rho_{k,n}\right) \\ & + \sum_{k=1}^{K} \lambda_k^{(2)} \left(P_{\max} - \sum_{n=1}^{N} p_{k,n}\right) + \sum_{k=1}^{K} \lambda_k^{(3)} \left(R_k - R_{\min}\right), \end{aligned} \tag{2.51}$$

where $\lambda_n^{(1)}$, $\lambda_k^{(2)}$, and $\lambda_k^{(3)}$ are the Lagrange multipliers related to the corresponding constraints. Furthermore, the Lagrange dual problem can be expressed as

$$\min_{\boldsymbol{\lambda}} \max_{\boldsymbol{\rho}, \mathbf{p}} L(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\lambda}), \tag{2.52}$$

subject to

$$\lambda_n^{(1)} \geqslant 0, \lambda_k^{(2)} \geqslant 0, \lambda_k^{(3)} \geqslant 0, \forall k, n. \tag{2.53}$$

Since the problem is now non-convex because of the binary $\rho_{k,n}$, the LDD method and the KKT condition cannot be applied to find the optimal solution. In the following,

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

we decouple the problem into two sub-problems: the channel assignment sub-problem and the power allocation sub-problem.

In the channel assignment sub-problem, we again relax the variables, ρ. In this way, we can utilize the KKT condition to find the optimal Lagrange multipliers, $\lambda_n^{(1)*}$, $\lambda_k^{(2)*}, \lambda_k^{(3)*}, \forall k, n$, and the optimal results, $\tilde{\rho}_{k,n}^*, p_{k,n}^*, \forall k, n$. By substituting these results into (2.52), the channel assignment problem can be formulated as

$$\max_{\boldsymbol{\rho}} \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} u_{k,n} + v, \tag{2.54}$$

subject to

$$\rho_{k,n} \in \{0, 1\}, \forall k, n, \tag{2.55}$$

$$\sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall n, \tag{2.56}$$

where

$$u_{k,n} = \left(\kappa_k^* \epsilon_k + \lambda_k^{(3)*}\right) R_{k,n}^* - \lambda_n^{(1)*} - \psi_k^* P_{\mathrm{e}}, \forall k, n,$$

and

$$v = -\sum_{k=1}^{K} \sum_{n=1}^{N} \psi_k \left(\zeta p_{k,n}^* + P_{\mathrm{fix}}\right) + \sum_{n=1}^{N} \lambda_n^{(1)*} + \sum_{k=1}^{K} \lambda_k^{(2)*} \left(P_{\max} - \sum_{n=1}^{N} p_{k,n}^*\right) - \sum_{k=1}^{K} \lambda_k^{(3)*} R_{\min},$$

is a constant irrespective to $\rho_{k,n}$. Apparently, this problem is an assignment problem whose optimal solution can be obtained by the classical Hungarian algorithm with a computational complexity of $O(N^3)$ [35].

Once the channel assignment is obtained, each user's subcarrier set can be denoted as $\mathcal{S}_k$. Next, the power allocation problem can be solved for each user with the following problem

$$\max_{\mathbf{p}} \frac{\sum_{n \in \mathcal{S}_k} R_{k,n}}{\sum_{n \in \mathcal{S}_k} \left(\zeta p_{k,n} + P_{\mathrm{e}}\right) + P_{\mathrm{fix}}}, \forall k, \tag{2.57}$$

subject to

$$\sum_{n \in \mathcal{S}_k} p_{k,n} \leq P_{\max}, \tag{2.58}$$

$$p_{k,n} \geq 0, n \in \mathcal{S}_k, \tag{2.59}$$

$$\sum_{n \in \mathcal{S}_k} R_{k,n} \geq R_{\min}. \tag{2.60}$$

Note that this problem is a standard convex–concave fractional programming and therefore the Dinkelbach algorithm can be used to solve it optimally. The detailed theory and approaches are discussed in Appendix 2.1.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Although both the Hungarian algorithm and the Dinkelbach algorithm optimally solve the channel assignment and power control problem, respectively, the proposed solution to the problem in (2.50) is not necessarily optimal due to the problem decoupling as well as its non-convexity. However, at least a suboptimal solution can be obtained since the result corresponds to a KKT point of the problem. Moreover, the main merit of the proposed algorithm is its low computational complexity, which can be implemented in practical networks.

2.4.2 Maximum-Minimal EE Maximization

We now consider the WMEE maximization problem, which aims at maximizing the minimum EE. The mathematical optimization problem can be formulated as

$$\eta_{\mathrm{mm}}^{\max} = \max_{(\boldsymbol{\rho},\mathbf{p})\in\boldsymbol{\Theta}} \min_{k\in\mathcal{K}}\{\eta_k\}. \tag{2.61}$$

Again, binary subcarrier allocation indicator is involved. Thus, we need to relax the above problem into a continuous one, as

$$\tilde{\eta}_{\mathrm{mm}}^{\max} = \max_{(\tilde{\boldsymbol{\rho}},\mathbf{p})\in\tilde{\boldsymbol{\Theta}}} \min_{k\in\mathcal{K}} \left\{\frac{\tilde{R}_k}{\tilde{P}_k}\right\}. \tag{2.62}$$

This problem is a generalized fraction problem (GFP) as described in Appendix 2.1. We can use the generalized Dinkelbach algorithm to solve it. However, the "max-min" operation renders it difficult to solve. In the following, we shall introduce an alternative algorithm to solve the problem.

First, according to the fact that a quasiconvex function attains its maximum on the vertex of a convex polyhedron [32], we have the following proposition.

PROPOSITION 2 *The WMEE problem is equivalent to*

$$\min_{k\in\mathcal{K}} \left\{\frac{\tilde{R}_k}{\tilde{P}_k}\right\} = \min_{\mathbf{y}\in\mathcal{Y}} \left\{\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\mathbf{y},\tilde{\boldsymbol{\rho}},\mathbf{p}) = \frac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}}\right\}, \tag{2.63}$$

where $\mathcal{Y} \triangleq \left\{(y_1,\ldots,y_K)\,|y_k \geqslant 0, \forall k, \sum_{k=1}^{K} y_k = 1\right\}$, $\mathbf{P} = \left\{\tilde{P}_1, \tilde{P}_2, \cdots \tilde{P}_K\right\}$, *and* $\mathbf{R} = \left\{\tilde{R}_1, \tilde{R}_2, \cdots \tilde{R}_K\right\}$.

Furthermore, we can prove the quasiconcavity of $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}$, as presented in the following proposition.

PROPOSITION 3 $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\mathbf{y},\tilde{\boldsymbol{\rho}},\mathbf{p})$ *is quasiconcave over* $(\tilde{\boldsymbol{\rho}},p)$ *for a given* $\mathbf{y}\in\mathcal{Y}$, *and quasilinear over* $\mathbf{y}$, *for a given* $(\tilde{\boldsymbol{\rho}},p)\in\tilde{\boldsymbol{\Theta}}$.

Proof: We can first prove that $\mathbf{y}\mathbf{R}^{\mathrm{T}}$ is concave. Define $\bar{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\tilde{\boldsymbol{\rho}},\mathbf{p}) = -\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\tilde{\boldsymbol{\rho}},\mathbf{p}) = -\dfrac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}}$, then its sublevel set can be expressed as

$$\tau_\alpha = \left\{\tilde{\rho}_{k,n} \geqslant 0, p_{k,n} \geqslant 0, \forall k,n | \bar{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\tilde{\boldsymbol{\rho}},\mathbf{p}) \leq \alpha\right\}, \tag{2.64}$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

which is equivalent to

$$\tau_\alpha = \left\{ \tilde{\rho}_{k,n} \geqslant 0, p_{k,n} \geqslant 0, \forall k, n | -\alpha \mathbf{y}\mathbf{P}^{\mathrm{T}} - \mathbf{y}\mathbf{R}^{\mathrm{T}} \leq 0 \right\}. \tag{2.65}$$

Since $-\alpha \mathbf{y}\mathbf{P}^{\mathrm{T}} - \mathbf{y}\mathbf{R}^{\mathrm{T}}$ is convex over $(\tilde{\boldsymbol{\rho}}, p)$, τ_α is convex, which leads to the quasiconcavity of $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\tilde{\boldsymbol{\rho}}, \mathbf{p})$. Similarly, define the sublevel sets of $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}(\mathbf{y}) = \dfrac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}}$ for given $(\tilde{\boldsymbol{\rho}}, p)$, as

$$S_\alpha = \left\{ y_k \geqslant 0, \forall k | \tilde{\eta}_y(\mathbf{y}) \leq \alpha \right\}, \tag{2.66}$$

which is equivalent to

$$S_\alpha = \left\{ y_k \geqslant 0, \forall k | \mathbf{y}\mathbf{R}^{\mathrm{T}} - \alpha \mathbf{y}\mathbf{P}^{\mathrm{T}} \leq 0 \right\}. \tag{2.67}$$

Since $\mathbf{y}\mathbf{R}^{\mathrm{T}} - \alpha \mathbf{y}\mathbf{P}^{\mathrm{T}}$ is both convex and concave over $y_k, \forall k$, it is quasilinear. □

Since $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{sum}}$ is quasiconcave, we can further apply the Sion's min-max theorem [33] to finally convert the problem into a better tractable one, as

$$\max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \min_{k \in \mathcal{K}} \left\{ \frac{\tilde{R}_k}{\tilde{P}_k} \right\} = \max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \min_{\mathbf{y} \in \mathcal{Y}} \left\{ \frac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}} \right\} = \min_{\mathbf{y} \in \mathcal{Y}} \max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \left\{ \frac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}} \right\}. \tag{2.68}$$

Now, we can solve the problem in (2.68) in two steps: the inner layer finds the optimal $\tilde{\boldsymbol{\rho}}, \mathbf{p}$ for a given $\mathbf{y}$, and the outer layer solves the optimal $\mathbf{y}^*$. The detailed algorithm is similar to Algorithm 6 in Appendix 2.1.

Interestingly, the GEE, defined as the ratio of the overall achievable data rate and the overall power consumption, serves as an upper bound for the WMEE. This can be simply proved by the setting $y_k = \dfrac{1}{K}, \forall k$ in (2.63), as

$$\tilde{\eta}_{\mathrm{mm}}^{\mathrm{spup}} = \max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \left\{ \frac{\mathbf{y}\mathbf{R}^{\mathrm{T}}}{\mathbf{y}\mathbf{P}^{\mathrm{T}}} \right\} = \max_{(\tilde{\boldsymbol{\rho}}, \mathbf{p}) \in \tilde{\boldsymbol{\Theta}}} \left\{ \frac{\sum_{k=1}^{K} \tilde{R}_k}{\sum_{k=1}^{K} \tilde{P}_k} \right\} \geq \tilde{\eta}_{\mathrm{mm}}^{\max}. \tag{2.69}$$

The GEE can be achieved by the Dinkelbach algorithm (please see the details in Appendix 2.1). We denote

$$B(\theta) = \sum_{k=1}^{K} \tilde{R}_k - \theta \sum_{k=1}^{K} \tilde{P}_k. \tag{2.70}$$

Then $\tilde{\eta}_{\mathrm{mm}}^{\mathrm{spup}}$ is achieved if and only if $\max\limits_{(\tilde{\boldsymbol{\rho}}, p) \in \tilde{\boldsymbol{\Theta}}} B\left(\tilde{\eta}_{\mathrm{mm}}^{\mathrm{spup}}\right) = 0$.

Again, the above algorithm only achieves an upper bound to the WMEE maximization problem due to the relaxation of binary variables. Therefore, in the following, we shall develop a suboptimal algorithm to the original problem in (2.61) while considering

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

the binary subcarrier allocation constraint. Based on the optimal EE achieved by the upper-bound algorithm, a suboptimal heuristic problem can be formulated, as

$$\max_{(\boldsymbol{\rho},\mathbf{p})\in\boldsymbol{\Theta}} \min_{k\in\mathcal{K}} \left\{ R_k - \tilde{\eta}_{\mathrm{mm}}^{\max}(k)\, P_k \right\},$$

where $\tilde{\eta}_{\mathrm{mm}}^{\max}(k)$ is the k-th element in the optimal EE achieved by the upper-bound algorithm. The idea of the suboptimal heuristic algorithm is to achieve the EE as close as to the optimal EE for each individual user. We can rewrite the problem in a parametric optimization format, as

$$\max_{(\boldsymbol{\rho},\mathbf{p})\in\boldsymbol{\Theta}} \tau, \tag{2.71}$$

subject to

$$R_k - \tilde{\eta}_{\mathrm{mm}}^{\max}(k)\, P_k \geq \tau, \forall k. \tag{2.72}$$

Then, the Lagrange multiplier method can be applied to solve it. The Lagrange function can be written as

$$\begin{aligned} L(\boldsymbol{\rho}, \boldsymbol{p}, \boldsymbol{\lambda}) =& \tau + \sum_{n=1}^{N} \lambda_n^{(1)} \left(1 - \sum_{k=1}^{K} \rho_{k,n}\right) \\ &+ \sum_{k=1}^{K} \lambda_k^{(2)} \left(P_{\max} - \sum_{n=1}^{N} p_{k,n}\right) \\ &+ \sum_{k=1}^{K} \lambda_k^{(3)} \left(R_k - R_{\min}\right) \\ &+ \sum_{k=1}^{K} \lambda_k^{(4)} \left[\left(R_k - \tilde{\eta}_{\mathrm{mm}}^{\max}(k) P_k\right) - \tau\right], \end{aligned} \tag{2.73}$$

where $\lambda_n^{(1)}$, $\lambda_k^{(2)}$, $\lambda_k^{(3)}$ and $\lambda_k^{(4)}$ are Lagrange multipliers related to the corresponding constraints in (2.35)–(2.39). Furthermore, the Lagrange dual decomposition is

$$\min_{\boldsymbol{\lambda}} \max_{\boldsymbol{\rho},\mathbf{p}} L(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\lambda}), \tag{2.74}$$

subject to

$$\lambda_n^{(1)} \geqslant 0, \lambda_k^{(2)} \geqslant 0, \lambda_k^{(3)} \geqslant 0, \lambda_k^{(4)} \geqslant 0, \forall k, n. \tag{2.75}$$

This problem can be easily solved by a similar method to the WSEE problem.

2.4.3 Numerical Results

We now present simulation results to evaluate the performance of the algorithms developed in this section. A simple two-user scenario is considered. The base station has a radius of 500 m. There are 2–8 subcarriers in the system. The channel of each user follows independent and identically distributed (i.i.d.) Rayleigh fading model. Other parameters are summarized in Table 2.2.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Table 2.2 Simulation parameters

Parameter	Value
Cell radius	500 m
subcarriers bandwidth, B	$\{1,2,3,4,5\} * 0.25$ MHz
Noise spectral density, N_0	-174 dBm/Hz
Path loss model	$128.1+37.6\log_{10}(d[\text{km}])$ dB
Shadowing standard deviation	10 dB
$R_{\min}$	100 kbps
Number of users, K	2
Number of subcarriers, N	2–8
P_{fix}	1 W–3 W
P_{e}	0.42 W
Power efficiency	50%

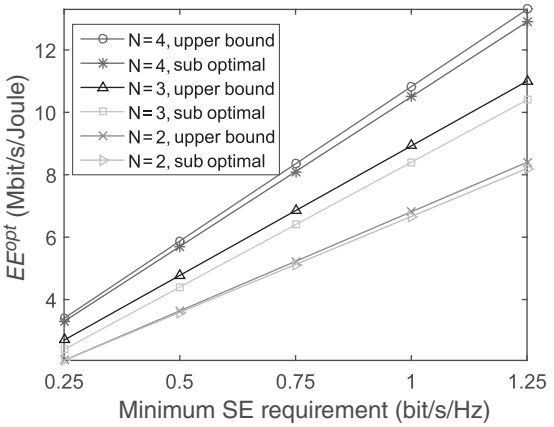


Figure 2.6 Weighted-sum EE versus subcarrier bandwidth. $P_{\max} = 26.7$ dBm, $P_{\text{fix}} = 2.13$ W.

In Fig. 2.6, we show the performance of WSEE algorithms where the weights are set equal for all users. Particularly, the upper-bound algorithm and the suboptimal algorithm are compared. As we can observe from the figure, the performance gap between the upper-bound algorithm and the suboptimal algorithm is very small, which verifies the effectiveness of our proposed algorithm. In addition, the WSEE increases as the number of subcarriers increases. This is because that more subcarriers will bring about more communication freedom for EE improvement.

In Fig. 2.7, the algorithms for maximizing the WMEE are compared, i.e., the upper-bound algorithm, the algorithm to maximize the GEE, and the suboptimal algorithm. From the figure, the GEE achieves the highest EE performance, as has been discussed before. However, this upper bound is very loose. On the other hand, the upper-bound algorithm can achieve a tight upper bound as we can observe that the WMEE gap between the upper-bound algorithm and suboptimal algorithm is very small. This result

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

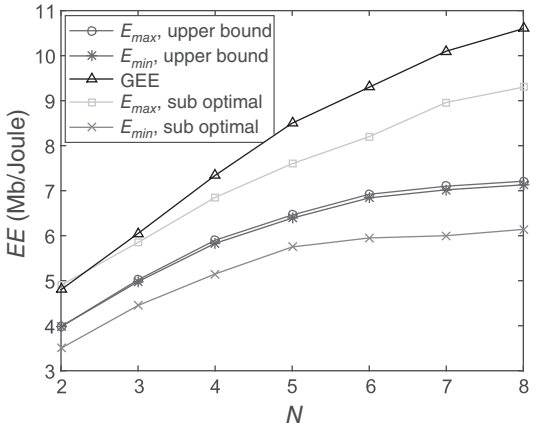


Figure 2.7 The minimum EE versus number of subcarriers. $P_{\max} = 26.7$ dBm, $P_{\text{fix}} = 2.13$ W, $B = 1.25$ MHz.

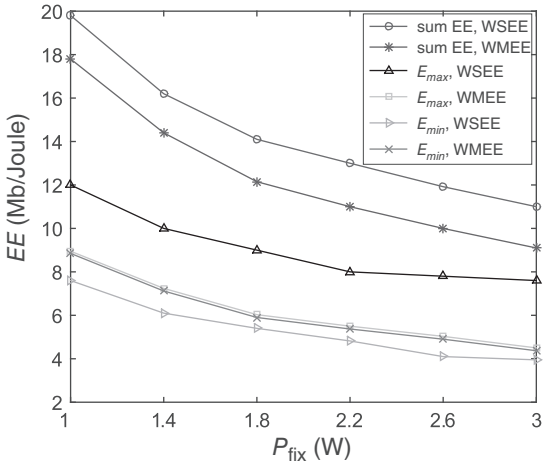


Figure 2.8 EE comparison of different algorithms. $N = 4$, $P_{\max} = 26.7$ dBm, $B = 1.25$ MHz.

verifies the effectiveness of our proposed algorithms for WMEE maximization. Again, from the figure, as the subcarrier number, N, increases, the EE performance increases.

We further compare the sum EE, the maximum EE, and the minimum EE of the two users with WSEE and WMEE maximization algorithms in Fig. 2.8. Both upper bounds for WSEE and WMEE maximization algorithms are plotted. From the figure, the WSEE algorithm has a better sum EE than the WMEE algorithm. However, from this result, WSEE algorithm doesn't necessarily guarantee user fairness. On the other hand, although the sum EE in the WMEE algorithm is smaller than that in the WSEE algorithm, the two users achieve almost the same EE as can be observed from the figure. In this regard, the WMEE maximization can achieve fair EE performance among users.

Our proposal in this section indeed provides an insightful reference for network operator to choose which algorithm to be used in practice.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

2.5 EE Design in Non-Orthogonal Systems

In the previous section, we illustrated how to use the optimization theory to maximize various EEs for orthogonal systems, such as OFDMA networks. In this section, we will further analyze the EE trade-off problem in non-orthogonal systems, where non-convex optimization theory will be applied.

As shown in Fig. 2.9, we consider a network with K pairs of users communicating with each other, denoted as $\mathcal{K} = \{1, 2, \ldots, K\}$. Let $\{\mathrm{Tx}_1, \mathrm{Tx}_2, \ldots, \mathrm{Tx}_K\}$ and $\{\mathrm{Rx}_1, \mathrm{Rx}_2, \ldots, \mathrm{Rx}_K\}$ denote the transmitter and the receiver, respectively, and B denote the overall bandwidth of the considered system. Moreover, let $h_{m,k}$ denote the channel power gain between Tx_m and Rx_k.

The capacity region of interference channel is still an open challenge and the best known achievable region is the Han–Kobayashi region [36], which is very difficult to achieve. Here, we simply treat interference as noise, which serves as a lower bound of the data rate for interference networks.

The data rate of user pair k in the interference network can be expressed as

$$R_k = B\log_2\left(1 + \frac{p_k h_{k,k}}{\sum\limits_{m=1, m\neq k}^{K} p_m h_{m,k} + BN_0}\right), \forall k, \tag{2.76}$$

where p_k is the transmit power of the transmitter k and N_0 represents the variance of the noise spectral density. Denote $\mathbf{p} = \{p_k\}$. The EE of each user pair can be defined as

$$\eta_k = \frac{R_k}{p_k + P_\mathrm{c}}, \forall k. \tag{2.77}$$

where p_k and P_c are the RF power consumption and the fixed circuit power consumption respectively. Here, for notational simplicity, we have omitted the power amplifier efficiency, which would not lose the generality of the following analysis.

Instead of maximizing a single-objective EE function, the emphasis of this section is to maximize the EE of each user from the perspective of multi-objective optimization

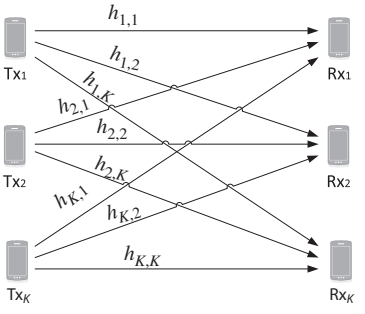


Figure 2.9 The model of non-orthogonal systems.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

theory, as detailed in Appendix 2.1. Therefore, the optimization problem can be mathematically formulated as

$$\boldsymbol{\eta}_{\max} = \max_{\mathbf{p}\in\mathcal{P}} \{\eta_1, \eta_2, \ldots, \eta_K\}, \tag{2.78}$$

where $\mathcal{P}$ is the feasible power allocation strategy set satisfying

$$0 \le p_k \le P_{\max}, \forall k, \tag{2.79}$$

$$R_k \ge R_{\min}, \forall k. \tag{2.80}$$

In the above, (2.79) limits the maximum transmit power for each user and (2.80) guarantees the minimum data rate for each user.

2.5.1 Utopia EE

We use the weighted Tchebycheff method to solve (2.78) (please see the details in Appendix 2.1). We shall first introduce the utopia EE for each individual user, as defined in the following.

Definition 1. The utopia EE for user k, η_k^{u}, is defined as the maximal EE this user could achieve, i.e., $\eta_k^{\mathrm{u}} = \max_{\eta\in\mathcal{F}}\{\eta_k\}$, where $\mathcal{F}$ denotes the set of all Pareto-optimal EEs.

According to the definition, the utopia EE can be solved by the following problem

$$\eta_k^{\mathrm{u}} = \max \eta_k, \tag{2.81}$$

subject to (2.79) and (2.80).

The above problem is a single-ratio fractional program, therefore Dinkelbach-like algorithm can be utilized to solve it. Let us define $\mathcal{U}(\varphi) = \max_{\mathbf{p}\in\mathcal{P}}\{R_k - \varphi P_k\}$, then η_k^{u} can be achieved if and only if $\mathcal{U}\left(\eta_k^{\mathrm{u}}\right) = 0$.

Unfortunately, it is still difficult to obtain the optimum value since the objective function is non-concave due to the interference in the denominator of R_k. However, this problem has a difference-of-convex (d.c.) structure, which can be solved by the method elaborated as follows.

The objective function can be rewritten as

$$f(\mathbf{p}) = R_k - \varphi P_k \triangleq f_{\mathrm{cave1}}(\mathbf{p}) - f_{\mathrm{cave2}}(\mathbf{p}), \tag{2.82}$$

where

$$f_{\mathrm{cave1}}(\mathbf{p}) = B\log_2\left(\sum_{m=1}^{K} p_m h_{m,k} + BN_0\right),$$

and

$$f_{\mathrm{cave2}}(\mathbf{p}) = B\log_2\left(\sum_{m=1,m\neq k}^{K} p_m h_{m,k} + BN_0\right) + \varphi(p_k + P_{\mathrm{c}}).$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

We can observe that both $f_{\text{cave1}}(\mathbf{p})$ and $f_{\text{cave2}}(\mathbf{p})$ are strictly concave on $\mathbf{p}$. Therefore, the original problem can be expressed as the following d.c. problem

$$\max_{\mathbf{p}\in\mathcal{P}} \{f_{\text{cave1}}(\mathbf{p}) - f_{\text{cave2}}(\mathbf{p})\}. \tag{2.83}$$

There are many algorithms that can effectively solve the d.c. problem, such as the d.c. algorithm (DCA) [37]. In this problem, the function $f_{\text{cave2}}(\mathbf{p})$ is differentiable. Therefore, the concave-convex procedure (CCCP) method can be utilized to solve the problem in (2.83) [38]. Therefore, we can develop the following theorem to solve the above problem by following the majorization-maximization approximation.

THEOREM 3 *The problem in* (2.83) *can be iteratively solved by the following concave programming*

$$\mathbf{p}^{(j+1)} = \arg\max_{\mathbf{p}\in\mathcal{P}} \left\{ f_{\text{cave1}}(\mathbf{p}) - \mathbf{p}^{\mathrm{T}} * \nabla f_{\text{cave2}}\left(\mathbf{p}^{(j)}\right)\right\}, \tag{2.84}$$

where $\mathbf{p}^{\mathrm{T}}$ *denotes the transpose of* $\mathbf{p}$ *and* $\nabla f_{\text{cave2}}\left(\mathbf{p}^{(j)}\right) \triangleq \left[\nabla_1^{(j)}, \nabla_2^{(j)}, \ldots, \nabla_K^{(j)}\right]$ *denotes the gradient of* $f_{\text{cave2}}(\mathbf{p})$ *at* $\mathbf{p}^{(j)}$*, where*

$$\nabla_i^{(j)} = \begin{cases} \dfrac{Bh_{i,k}}{\left(\sum\limits_{m=1,m\neq k}^{K} p_m^{(j)} h_{m,k} + BN_0\right)\ln 2}, & \forall i \neq k, \\ \varphi, & i = k. \end{cases}$$

In the above theorem, we can further prove that the first part in (2.84), $f_{\text{cave1}}(\mathbf{p})$, is concave, whereas the second part, $-\mathbf{p}^{\mathrm{T}} * \nabla f_{\text{cave2}}\left(\mathbf{p}^{(j)}\right)$, is linear. Therefore, it is a standard concave optimization problem and classic convex optimization methods can be used to solve it. The detailed procedures of applying the CCCP approach to solve our problem can be summarized in Algorithm 1.

Algorithm 1 CCCP for the d.c. problem

1: **Initialize**
2: Set $\epsilon > 0$, $j = 0$, and $\forall \mathbf{p}^{(0)} \in \mathcal{P}$.
3: **Do**
4: $\mathbf{p}^{(j+1)} = \arg\max_{\mathbf{p}\in\mathcal{P}} \left\{ f_{\text{cave1}}(\mathbf{p}) - \mathbf{p}^{\mathrm{T}} * \nabla f_{\text{cave2}}\left(\mathbf{p}^{(j)}\right)\right\}$.
5: $j = j + 1$.
6: **Until** $\left\|\mathbf{p}^{(j)} - \mathbf{p}^{(j-1)}\right\| < \epsilon$.

Furthermore, the detailed procedures of our algorithm to find the Utopia EE for user k are summarized in Algorithm 2.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Algorithm 2 The algorithm to find the Utopia EE for user k.

1: **Initialize**
2: Set $\epsilon > 0$, $j = 0$, and $\forall \mathbf{p}^{(0)} \in \mathcal{P}$.
3: **Do**
4: $\varphi_j = \eta_k^{(j)}$.
5: Calculate $\mathbf{p}^{(j+1)} = \arg\max_{\mathbf{p}\in\mathcal{P}} \left\{ R_k - \varphi_j P_k \right\}$ by CCCP.
6: $j = j + 1$.
7: **Until** $\left| \max_{\mathbf{p}\in\mathcal{P}} \left\{ R_k - \varphi_{j-1} P_k \right\} \right| < \epsilon$.

We now show that the above algorithm can converge. First, according to the CCCP, the objective function in (2.84) is non-decreasing on the generated sequence $\left\{\mathbf{p}^{(j)}\right\}$, which will eventually converge to the stationary point $\mathbf{p}^{(\infty)}$ when $\mathbf{p}^{(j+1)} = \mathbf{p}^{(j)}$. From [38], the stationary point also satisfies the Karush–Kuhn–Tucker (KKT) conditions of the problem in (2.83). Therefore, by jointly utilizing the Dinkelbach method and the CCCP, at least a local optimal solution can be obtained. In fact, it is quite challenging to prove the global optimality. However, from [37, 39], the global optimum can be always achieved if the starting point is appropriately chosen.

2.5.2 Pareto-Optimal EE

After the utopia EE for each user is attained, we can apply the weighted Tchebycheff method to convert the multi-objective optimization problem into a single-objective function one, as

$$\begin{aligned} \eta^{\mathrm{wt}} &= \min_{\mathbf{p}\in\mathcal{P}} \max_{k\in\mathcal{K}} \left\{ \phi_k \left(\eta_k^{\mathrm{u}} - \frac{R_k}{P_k} \right) \right\} \\ &\triangleq \min_{\mathbf{p}\in\mathcal{P}} \max_{k\in\mathcal{K}} \left\{ \frac{\phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k \right)}{P_k} \right\}, \end{aligned} \tag{2.85}$$

where ϕ_k is an arbitrary positive weight for user k. The details of the weighted-Tchebycheff method are discussed in Appendix 2.1.

The utopia EE and the Pareto-optimal EE in a two-user case is illustrated in Fig. 2.10. In the figure, the shadowed area depicts all achievable EE for the two users in the problem in (2.78) and its boundary is the Pareto-optimal EE, as illustrated by the solid curve. Furthermore, η_1^{u} and η_2^{u} are the utopia EE for each user, which is exactly the maximum EE that the user can achieve. This figure also clearly illustrates how the weighted Tchebycheff method achieves the Pareto-optimal EE by minimizing the distance between the utopia EE point and achieved EE point, i.e., η_1^* and η_2^*. Moreover, by varying the weight $\phi_k, k = 1, 2$, all Pareto-optimal EEs on the boundary can be achieved.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

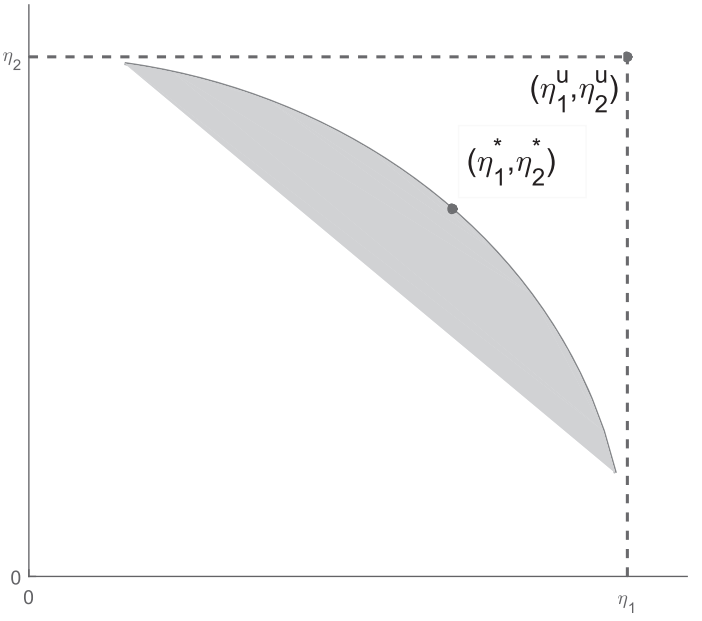


Figure 2.10 The weighted Tchebycheff method for a two-user case.

We now introduce an algorithm to solve the problem in (2.85) for given ϕ_k and η_k^{u}. The objective function in the weighted Tchebycheff method can be rewritten as

$$\eta^{\mathrm{wt}} = \min_{\mathbf{p}\in\mathcal{P}} \max_{k\in\mathcal{K}} \left\{ \frac{\phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right)}{P_k} \right\}. \tag{2.86}$$

The above problem is also a fractional optimization and can be solved by the Dinkelbach method [19], which is discussed in Appendix 2.1 in detail. We have the following theorem, whose proof is similar to that in [26].

THEOREM 4 *Define*

$$V(\alpha) = \min_{\mathbf{p}\in\mathcal{P}} \max_{k\in\mathcal{K}} \left\{ \phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right) - \alpha P_k \right\}, \tag{2.87}$$

then η^{wt} *is achieved if and only if* $V\left(\eta^{\mathrm{wt}}\right) = 0$.

Unfortunately, the problem in (2.87) is also non-convex due to the involved interference in R_k. In the following, we develop a suboptimal algorithm to achieve the local optimum point by jointly utilizing the Lagrange method and the CCCP method. First, define $l = \max_{k\in\mathcal{K}} \left\{ \phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right) - \alpha P_k \right\}$. Then, based on the parametric method, the problem in (2.87) can be equivalently transformed into

$$\min_{\mathbf{p}\in\mathcal{P}} l, \tag{2.88}$$

subject to

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

$$l \geq \phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right) - \alpha P_k, \forall k. \tag{2.89}$$

According to the LDD method, we can express the Lagrangian function of the above problem as

$$\begin{aligned} L\left(\mathbf{p}, l, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}\right) = & l + \sum_{k=1}^{K} \lambda_k \left\{\phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right) - \alpha P_k - l\right\} \\ & + \sum_{k=1}^{K} \mu_k \left(p_k - P_{\max}\right) \\ & + \sum_{k=1}^{K} \nu_k \left\{\left(2^{\frac{R_{\min}}{B}} - 1\right)\left(\sum_{m=1, m\neq k}^{K} p_m h_{m,k} + BN_0\right) - p_k h_{k,k}\right\}, \end{aligned}$$

where $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$ are Lagrange multiplier vectors corresponding to the respective constraints. Furthermore, the Lagrangian dual problem can be written as

$$\max_{\boldsymbol{\lambda}\geq 0, \boldsymbol{\mu}\geq 0, \boldsymbol{\nu}\geq 0} \min_{\mathbf{p}\geq 0, l} L\left(\mathbf{p}, l, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}\right). \tag{2.90}$$

Then, the Algorithm 3 is presented to solve the above problem, which contains the inner loop and the outer loop.

Inner Loop

The inner loop aims to solve $\min_{\mathbf{p}\geq 0, l} L\left(\mathbf{p}, l\right)$ for a given $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$. It is also non-convex but has a d.c. structure. Therefore, we can decompose $L\left(\mathbf{p}, l\right)$ into a subtraction of a concave function and a linear function, as

$$L\left(\mathbf{p}, l\right) = L_{\mathrm{vex1}}\left(\mathbf{p}, l\right) - L_{\mathrm{vex2}}\left(\mathbf{p}\right), \tag{2.91}$$

where

$$\begin{aligned} L_{\mathrm{vex1}}\left(\mathbf{p}, l\right) = & l - \sum_{k=1}^{K} \lambda_k \left\{\phi_k B \log_2 \left(\sum_{m=1}^{K} p_m h_{m,k} + BN_0\right)\right\} \\ & + \sum_{k=1}^{K} \lambda_k \left\{\phi_k \eta_k^{\mathrm{u}} P_k - \alpha P_k - l\right\} + \sum_{k=1}^{K} \mu_k \left(p_k - P_{\max}\right) \\ & + \sum_{k=1}^{K} \nu_k \left\{\left(2^{\frac{R_{\min}}{B}} - 1\right)\left(\sum_{m=1, m\neq k}^{K} p_m h_{m,k} + BN_0\right) - p_k h_{k,k}\right\}, \end{aligned}$$

and

$$L_{\mathrm{vex2}}\left(\mathbf{p}\right) = -\sum_{k=1}^{K} \lambda_k \left\{\phi_k B \log_2 \left(\sum_{m=1, m\neq k}^{K} p_m h_{m,k} + BN_0\right)\right\}.$$

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Now, $\min_{\mathbf{p}\geq 0, l} L(\mathbf{p}, l)$ has been expressed as the minimization of a d.c. function in a convex set, as

$$\min_{\mathbf{p}\in\mathcal{P}, l} \{L_{\text{vex1}}(\mathbf{p}, l) - L_{\text{vex2}}(\mathbf{p})\}. \tag{2.92}$$

The CCCP method can be used to effectively solve it. Similar to Theorem 3, we have the following theorem.

THEOREM 5 *The problem in* (2.92) *can be iteratively solved by the following sequential convex programming*

$$\mathbf{p}^{(j+1)} = \arg\min_{\mathbf{p}\in\mathcal{P}} \left\{ L_{\text{vex1}}(\mathbf{p}, l) - \mathbf{p}^{\mathrm{T}} * \nabla L_{\text{vex2}}\left(\mathbf{p}^{(j)}\right) \right\}, \tag{2.93}$$

where $L_{\text{vex2}}\left(\mathbf{p}^{(j)}\right) \triangleq \left[\nabla_1^{(j)}, \nabla_2^{(j)}, \ldots, \nabla_K^{(j)}\right]$ *denotes the gradient of* $L_{\text{vex2}}(\mathbf{p})$ *at* $\mathbf{p}^{(j)}$, *where*

$$\nabla_i^{(j)} = -\sum_{k=1, k\neq i}^{K} \lambda_k \phi_k B \left\{ \frac{h_{i,k}}{\left(\sum_{m=1, m\neq k}^{K} p_m^{(j)} h_{m,k} + BN_0\right) \ln 2} \right\}.$$

We can also prove that the first part of (2.93), $L_{\text{vex1}}(\mathbf{p}, l)$, is convex, and the second part of (2.93), $-\mathbf{p}^{\mathrm{T}} * \nabla L_{\text{vex2}}\left(\mathbf{p}^{(j)}\right)$, is linear. Therefore, (2.93) is a standard convex optimization problem and can be easily solved.

Outer Loop

Since the Lagrangian dual function is differentiable, we utilize the subgradient method to solve the outer loop, i.e., finding the optimal $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ for given $(\mathbf{p}, l)$. The subgradient update equations can be written as

$$\begin{aligned} \lambda_k &= \left[\lambda_k + \varsigma_1 \left\{\phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k\right) - \alpha P_k - l\right\}\right]^+, \forall k, \\ \mu_k &= \left[\mu_k + \varsigma_2 \left(p_k - P_{\max}\right)\right]^+, \forall k, \\ \nu_k &= \left[\nu_k + \varsigma_3 \left(\vartheta - p_k h_{k,k}\right)\right]^+, \forall k, \end{aligned} \tag{2.94}$$

where $\vartheta = \left(2^{\frac{R_{\min}}{B}} - 1\right)\left(\sum_{m=1, m\neq k}^{K} p_m h_{m,k} + BN_0\right)$, ς_1, ς_2, and ς_3 are all positive stepsizes.

We now summarize the proposed algorithm to achieve the Pareto-optimal EE, which contains three nested steps. The first step leverages the Dinkelbach method to update α in (2.87) until $V(\alpha) = 0$. The second step is a standard subgradient method to update the Lagrangian multipliers in (2.94). The third step utilizes the CCCP method to solve (2.93). The detailed procedures of the three-step algorithm can be found in Algorithm 3.

Again, since all three steps can converge, the convergence of the proposed algorithm can be guaranteed. However, due to the non-convexity of the optimization problem, the global optimality cannot be ensured although it converges. However, our algorithm can at least achieve a local optimal point that satisfies the KKT condition. Nevertheless,

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

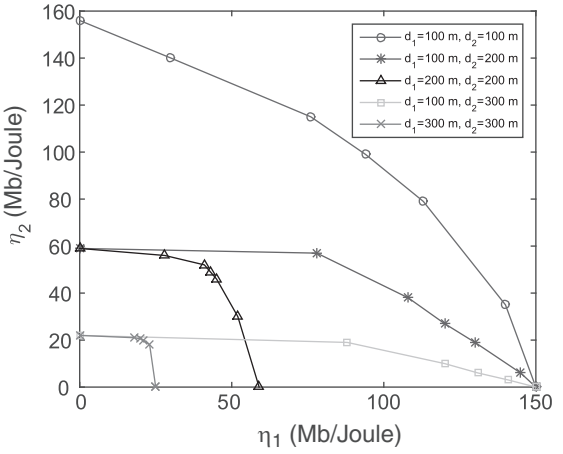


Figure 2.11 The EE trade-off with different communication distances. © 2016 IEEE. Reprinted, with permission, from Yu, G., 2016, "Energy Efficiency Tradeoff in Interference Channels," IEEE Access, Vol. 14, No. 6, pp. 3207–3218.

if carefully choosing the start point in the CCCP method, global optimum is often attainable.

Algorithm 3 The algorithm to find the Pareto-optimal EE.

Initialize

Set $\epsilon_1, \epsilon_2 > 0$, $j = 0$, and $\forall \mathbf{p}^{(0)} \in \mathcal{P}$.

Calculate $\eta_k^{\mathrm{u}}, \forall k$.

Do

Calculate $\alpha_j = \max\limits_{k \in \mathcal{K}} \left\{ \frac{\phi_k \left(\eta_k^{\mathrm{u}} P_k^{(j)} - R_k^{(j)} \right)}{P_k^{(j)}} \right\}$.

Initialize $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$, $i = 0$, $l^{(0)} = 0$.

Do

Using the CCCP to find $(\mathbf{p}^*, l^*) = \arg \min\limits_{\mathbf{p} \in \mathcal{P}, l} \{L_{\mathrm{vex1}}(\mathbf{p}, l) - L_{\mathrm{vex2}}(\mathbf{p})\}$.

Update $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$, $i = i + 1$, $\mathbf{p}^{(i)} = \mathbf{p}^*$, $l^{(i)} = l^*$.

Until $\left\| \left(\mathbf{p}^{(i)}, l^{(i)} \right) - \left(\mathbf{p}^{(i-1)}, l^{(i-1)} \right) \right\| < \epsilon_1$.

$j = j + 1$, $\mathbf{p}^{(j)} = \mathbf{p}^*$.

Until $\left| \min\limits_{\mathbf{p} \in \mathcal{P}} \max\limits_{k \in \mathcal{K}} \left\{ \phi_k \left(\eta_k^{\mathrm{u}} P_k - R_k \right) - \alpha_{j-1} P_k \right\} \right| < \epsilon_2$.

2.5.3 Numerical Results

A system with two pairs of users is considered. Let d_i denote the distance between the i-th user pair, which varies from 100–300 m. We assume that the channel models

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

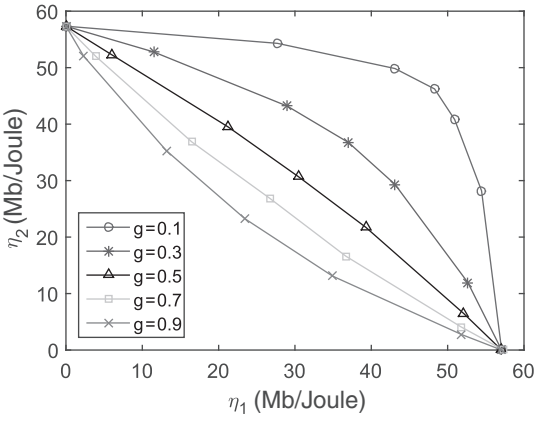


Figure 2.12 The EE trade-off in the two-user case with different g. © 2016 IEEE. Reprinted, with permission, from Yu, G., 2016, "Energy Efficiency Tradeoff in Interference Channels," IEEE Access, Vol. 14, No. 6, pp. 3207–3218.

between users are according to i.i.d Rayleigh fading and the path loss exponent is 4. Furthermore, we assume that each user has the same maximum transmit power, $P_{\max}$, and the fixed circuit power consumption is assumed to be 24 dBm. The minimum data rate requirement, $R_{\min}$, is not considered in the simulation. Moreover, for simplicity, we further assume that the average interference power gain to all other users are identical for each user, that is, $\overline{h}_{k,m} = g\overline{h}_{k,k}, \forall m \neq k$, where g is the interference-to-signal ratio.

Figure 2.11 plots the Pareto-optimal EE for the two-user case with different communication distances, where d_1 and d_2 denote the communication distances between user pairs. The EE trade-off can be easily observed from the figure. Moreover, the EE decreases with the communication distance. We further compare the EE trade-off curves with different interference-to-signal ratios, g, in Fig. 2.12. Here, both users have the same communication distance of 200 m. Also, the EE trade-off between the two users can be easily observed from the figure. Both users' EE decreases with g since large interference will certainly degrade both the spectral and energy efficiency.

Appendix 2.1 Optimization Theory for EE Design

This section discusses the mathematical optimization theory for EE design. Different from SE optimization problems, EE optimization problems generally involve fractional objective functions, which are known as fractional programming and are challenging to solve in general. New mathematical optimization theory and method are required to deal with it. The fractional programming can be effectively solved by the Dinkelbach approach, by converting the fractional objective function into a subtractive form. In this section, we will introduce the basic methodology of fractional programming and its extensions, such as sum-of-ratios optimization and generalized fractional programming

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

(GFP). We will also introduce the multi-objective optimization theory and attempt to deal with the EE trade-off problem.

A2.1.1 Fractional Programming and the Dinkelbach Algorithm

Fractional programming solves the mathematical optimization problems with fractional objective functions, which are in general nonlinear and non-convex. The ratio function to be optimized often reflects some kind of efficiency, defined as the ratio of utility and cost. In the EE optimization problem, the utility corresponds to the achievable data rate while the cost corresponds to the consumed power.

Definition 2 (Fractional programming problem). Let us define $f, g, h_j, j = 1, \ldots, M$ as some real-valued functions on a set $\mathcal{S}_0 \in \mathbb{R}^n$. The nonlinear programming

$$\max_x \frac{f(x)}{g(x)}, \tag{A2.1}$$

subject to

$$\begin{aligned} h_j(x) &\leq 0, j = 1, \ldots, M, \\ g(x) &> 0, \end{aligned}$$

is a fractional programming. Specifically, when f is nonnegative and concave, g is positive and convex, and $\mathcal{S}_0$ is a convex set, the problem is a concave fractional programming.

The Dinkelbach algorithm is effective to solve fractional programming problems. Define an auxiliary function

$$F(\lambda) = \max_x \{f(x) - \lambda g(x)\}. \tag{A2.2}$$

Then, the following theorem presents the solution to the problem in (A2.1).

THEOREM 6 *x^* is the optimal solution to the fractional programming in* (A2.1) *if and only if x^* also solves the problem in* (A2.2) *with the same constraints and when $F(\lambda) = 0$.*

According to Theorem 6, the fractional objective function now can be transformed into a subtractive function and can be solved by standard methods. Particularly, if f is concave and g is convex, the problem in (A2.2) can be solved by classical convex optimization methods, such as interior method, subgradient method, etc. To solve (A2.1), an iterative updating is further required to find the root of $F(\lambda) = 0$. The Dinkelbach algorithm [19] is an effective way to accomplish this, as presented in Algorithm 4.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

Algorithm 4 Dinkelbach algorithm to solve single-ratio fractional programming

Initialize $x_0 \in \mathcal{S}_0$. Set $\varepsilon > 0$;
Compute $\lambda_1 = f(x_0)/g(x_0)$. Set $k = 1$;
while $|F(\lambda_k)| \geq \varepsilon$ **do**
    $x^* = \arg\max_{x \in \mathcal{S}_0} \{f(x) - \lambda_k g(x)\}$
    $k = k + 1$;
    $\lambda_k = f(x^*)/g(x^*)$;
end **while**

The Dinkelbach algorithm can deal with all fractional programming problems, no matter whether the problem is concave. However, for concave fractional programming, a simple approach can be used to solve it.

Let us denote $t = \dfrac{1}{g(x)}$, then the concave fractional program can be equivalently transformed into the following parameter-free concave problem

$$\max_{\{x,t\}} tf(x), \tag{A2.3}$$

subject to

$$\begin{aligned} h_j(x) &\leq 0, j = 1, \ldots, M, \\ tg(x) &\leq 1, \\ t &\geq 0. \end{aligned}$$

Now the transformed problem is a concave one and can be easily solved by some classical methods. The fractional programming has been applied to design the energy-efficient resource allocation for green radio networks in [20–23].

A2.1.2 Sum-of-Ratios Optimization

The fractional programming aims to solve a single ratio problem, e.g., the GEE metric in (2.3). Whereas for the other EE metrics, such as WSEE and WMEE, the objective function involves multiple ratios, which can not be directly solved by the single-ratio Dinkelbach algorithm. In the following exploration, we will introduce two alternative methods to solve multi-ratio fractional programming, namely, sum-of-ratios optimization [24] and generalized fractional programming [25, 26].

Regarding the WSEE metric defined in (2.4), the objective function is the summation of several fractional functions, which is known as the sum-of-ratios programming. The sum-of-ratios optimization has been employed to solve various WSEE maximization problems in [27–30].

Similar to the definition of fractional programming, the sum-of-ratios programming is defined as follows.

Definition 3 (Sum-of-ratios programming). Define $f_i, g_i, i = 1, \ldots, N, h_j, j = 1, \ldots, M$ as some real-valued functions on a set $\mathcal{S}_0 \in \mathbb{R}^n$. The sum-of-ratios programming has the following format

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

$$\max_{x} \sum_{i=1}^{N} \frac{f_i(x)}{g_i(x)}, \tag{A2.4}$$

subject to

$$h_j(x) \leq 0, j = 1, \ldots, M,$$
$$g_i(x) > 0, i = 1, \ldots, N.$$

The sum-of-ratios programming can be solved by utilizing parametric algorithm and by converting the fractional format into subtractive format as well. It is easy to equivalently convert the above problem into the following one

$$\max_{x,\beta} \sum_{i=1}^{N} \beta_i, \tag{A2.5}$$

subject to

$$f_i(x) \geq \beta_i g_i(x), i = 1, \ldots, N,$$
$$h_j(x) \leq 0, j = 1, \ldots, M,$$
$$g_i(x) > 0, i = 1, \ldots, N.$$

Then, we introduce the following theorem to further transform the problem into a better tractable one [24].

THEOREM 7 *If* (x^*, β^*) *is the solution to the problem* (A2.5)*, then there exist* $u_i, i = 1, \ldots, N$, *such that* x^* *is a solution to the following problem for* $u = u^*$ *and* $\beta = \beta^*$

$$\max_{x} \sum_{i=1}^{N} u_i(f_i(x) - \beta_i g_i(x)), \tag{A2.6}$$

subject to

$$h_j(x) \leq 0, j = 1, \ldots, M.$$

And x^* *also satisfies the following system of equations for* $u = u^*$ *and* $\beta = \beta^*$

$$u_i = \frac{1}{g_i(x)}, i = 1, \ldots, N, \tag{A2.7}$$
$$f_i(x) - \beta_i g_i(x) = 0, i = 1, \ldots, N. \tag{A2.8}$$

According to the above theorem, the sum-of-ratios optimization problem can be iteratively solved by two nested loops. The inner loop solves the problem in (A2.6) and the outer loop finds the optimal u^* and β^* satisfying (A2.7) and (A2.8). If $f_i, \forall i$ are concave and $g_i, \forall i$ are convex, the inner problem is concave and its optimum can be obtained. In this case, the duality gap between the equivalent problem (A2.6) and the original problem (A2.5) is zero, which means that the problem can be optimally solved. Moreover, the outer loop can be solved by some numerical analysis approaches like Newton's method.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

A2.1.3 Generalized Fractional Programming

Generalized fractional programming (GFP) is another extension of the single-ratio fractional programming into the multi-ratio scenario, which aims to maximize the minimum of several ratio functions.

Definition 4 (Generalized fractional programming). Define $f_i, g_i, i = 1, \ldots, N, h_j$, $j = 1, \ldots, M$, as some real-valued functions on a set $\mathcal{S}_0 \in \mathbb{R}^n$. The GFP has the following format

$$\max_x \min_i \frac{f_i(x)}{g_i(x)}, \tag{A2.9}$$

subject to

$$h_j(x) \leq 0, j = 1, \ldots, M,$$
$$g_i(x) > 0, i = 1, \ldots, N.$$

Also, we can use the parametric algorithm to convert the fractional problem into a subtractive form, as

$$F(\mu) = \max_x \min_i \{f_i(x) - \mu g_i(x)\}. \tag{A2.10}$$

Then, similar to the single-ratio fractional programming, problem (A2.9) and problem (A2.10) have the same optimal solutions if $F(\mu) = 0$.

Accordingly, a Dinkelbach-type algorithm can be introduced to solve the above problem, as summarized in the following.

Algorithm 5 Dinkelbach-type algorithm to solve GFP

1: Initialize $x_0 \in \mathcal{S}_0$. Set $\varepsilon > 0$;
2: Compute $\mu_1 = \min_i f_i(x_0)/g_i(x_0)$. Set $k = 1$;
3: **while** $|F(\mu_k)| \geq \varepsilon$ **do**
4: $x^* = \arg\max_{x \in \mathcal{S}_0} \min_i \{f_i(x) - \mu_k g_i(x)\}$
5: $k = k + 1$;
6: $\mu_k = \min_i f_i(\mathrm{x}^*)/g_i(x^*)$;
7: end **while**

This algorithm is very similar to the single-ratio Dinkelbach algorithm except that a *min* operation is involved. It is useful to apply if the step (4) can be easily solved. However, the main drawback of the above algorithm is its slow convergence rate, which is superlinear as indicated in [26].

A fast Dinkelbach-type algorithm was developed in [31], which applies the Newton-like algorithm to update the ratio μ. Assuming x^* to be the optimal solution, the parametric subtractive problem can be reformulated as

$$F(\mu) = \max_x \min_i \left\{ \frac{f_i(x) - \mu g_i(x)}{g_i(x^*)} \right\}. \tag{A2.11}$$

In practice, the optimal x^* is impossible to be known a priori. Therefore, the previous iteration point, x_{k-1}, could be used as an approximation of x^*. Using (A2.11) instead of step (4) in the Dinkelbach-type algorithm, the convergence rate becomes quadratic.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

The preceding algorithms can effectively solve the GFP. However, the *min* operation in step (4) renders it difficult to solve in some cases. In the following, we introduce a new dual problem to solve the GFP.

We assume that $f_i, \forall i$ are concave on $\mathcal{S}_0$ and $g_i, \forall i$, are positive and convex on $\mathcal{S}_0$. Moreover, either $f_i, \forall i$ are nonnegative or $g_i, \forall i$ are affine on $\mathcal{S}_0$.

Let $f(x) = [f_1(x), f_2(x), \ldots, f_N(x)]$ and $g(x) = [g_1(x), g_2(x), \cdots, g_N(x)]$. Then, according to the fact that a quasiconvex function attains its maximum on the vertex of a convex polyhedron [32], we have

$$\min_i \frac{f_i(x)}{g_i(x)} = \min_{y \in Y} \frac{yf(x)}{yg(x)},$$

where $\mathcal{Y} \triangleq \{(y_1, ..., y_N) | y_n \geq 0, \forall n, \sum_{n=1}^{N} y_n = 1\}$.

Moreover, by applying Sion's mini-max theorem [33], the problem can be eventually transformed into

$$\max_x \min_i \frac{f_i(x)}{g_i(x)} = \max_x \min_{y \in Y} \frac{yf(x)}{yg(x)} = \min_{y \in Y} \max_x \frac{yf(x)}{yg(x)}.$$

Now the problem can be solved in two steps: the inner layer solves the optimal x^* for a given y and the outer layer finds the optimal y^*. Let us define $c(y) = \max_x \frac{yf(x)}{yg(x)}$. Obviously, the inner problem is a single-ratio fractional problem and can be easily solved by the Dinkelbach algorithm.

To solve the outer layer, we define $F(y, c(y)) = \max_x y(f(x) - c(y)g(x))$. Then, the optimal y^* can be achieved if $\min_y F(y, c(y)) = 0$. Based on the above analysis, we can introduce an iterative algorithm to solve the dual problem, as detailed in Algorithm 6.

Algorithm 6 Dual algorithm to solve the GFP

1: Initialize $y_0 \in Y, \varepsilon, \varphi$, and $k = 0$.
2: $x^* = \arg\max_{x \in \mathcal{S}_0} \{y_k f(x) - \varphi y_k g(x)\}$
3: If $\left| \max_{x \in \mathcal{S}_0} \{y_k f(x) - \varphi y_k g(x)\} \right| \leq \varepsilon$, then
4: Goto step (9)
5: Else
6: $\varphi = \frac{y_k f(x)}{y_k g(x)}$
7: Goto step (2)
8: End if
9: $y^* = \arg\min_{y \in Y} \max_{x \in \mathcal{S}_0} \{yf(x) - \varphi yg(x)\}$
10: If $\left| \max_{x \in \mathcal{S}_0} \{y^* f(x) - \varphi y^* g(x)\} \right| \leq \varepsilon$, then
11: Exit
12: Else
13: $k = k + 1, y_k = y^*$, go to step (2)
14: End if

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

A2.1.4 Multi-Objective Optimization and Weighted Tchebycheff Method

In the previous subsection, we have introduced the fractional programming and its extensions to solve EE-oriented resource optimization problems. The optimization solutions for both single-ratio and multi-ratio fractional programs have been discussed. The aforementioned problems are basically single-objective optimizations, no matter whether the objective is single-ratio or multi-ratio. However, in some occasions, we need to maximize many different EEs simultaneously and in this case, single-objective optimizations are not effective enough. For example, if jointly maximizing both the EE of users (uplink) and the EE of base stations (downlink), neither WSEE nor WMEE could provide a good metric for the optimization.

In light of the above, new optimization theory and approaches involving multi-objective optimization are demanded. Therefore, in this part, we will introduce the basic formulation of multi-objective optimization problems and their solutions.

Different from the single-objective optimization, multi-objective optimization aims to maximize many objectives simultaneously, or equivalently, an objective vector. The multi-objective optimization problem (MOOP) can be mathematically formulated as [34]

$$\max_{x \in \mathcal{S}_0} \left(f_1(x), f_2(x), \ldots, f_N(x)\right). \tag{A2.12}$$

In general, the solution to a MOOP is not unique since all objective functions cannot always be maximized simultaneously. The solutions to a MOOP are usually known as the Pareto-optimal solutions, which can be defined as follows.

Definition 5 (Pareto-optimal solution to the MOOP). A solution x^* is Pareto optimal to the MOOP defined in (A2.12) if and only if there does not exist any another x, such that $f_i(x) \geq f_i(x^*), \forall i$ and $f_j(x) > f_j(x^*)$ for any index j.

The Pareto-optimal solution for the problem in (A2.12) means that a single objective function cannot be increased without decreasing any other objective functions.

To directly solve a MOOP is very challenging due to the existence of multiple Pareto-optimal solutions. Instead, a MOOP can be effectively solved by converting it into some single-objective optimization problems, e.g., scalarizing the multiple objectives into a single one. There are many scalarization methods, including the weighted-summation of the objective functions, minimizing the maximal of the objective functions, etc. In the following discussion, we will introduce another scalarization method, namely the weighted Tchebycheff method, to solve MOOPs. In fact, the weighted Tchebycheff method can provide the complete Pareto-optimal solutions no matter whether the objective functions are concave. Also, in many occasions, the weighted Tchebycheff method also has a merit of low computational complexity.

To solve the MOOP in (A2.12) using the weighted Tchebycheff method, it is essential to introduce the concept of the utopia point, f_i^u, for each objection function f_i. The utopia point is also known as the idea point, which is defined as the maximal attainable point of each objective function, as

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

$$f_i^u = \max_{x \in \mathcal{S}_0} f_i(x).$$

By solving the above single-objective optimization problem, the utopia point is often attainable. Having the utopia point for each objective function, a scalarized single-objective problem can be formulated as

$$\min_{x \in \mathcal{S}_0} \left\{ \sum_{i=1}^{N} w_i [f_i^u - f_i(x)]^p \right\}^{\frac{1}{p}}, \tag{A2.13}$$

where w_i is the weight for the i-th objective function with $\sum_{i=1}^{N} w_i = 1$ and $w_i > 0, \forall i$, and $p > 0$. The above optimization problem aims to minimize the norm or distance between the solution point and the utopia point, while the weight vector $\mathbf{w}$ stands for the relative importance of each objective function. According to [34], the single-objective problem in (A2.13) provides sufficient and necessary conditions for the Pareto optimality of the original problem in case that its solution is unique.

The weighted Tchebycheff method is expressed as

$$\min_{x \in \mathcal{S}_0} \max_i \{f_i^u - f_i(x)\}, \tag{A2.14}$$

which is a special case of the problem in (A2.13) if $p \to \infty$.

References

[1] D. Feng, C. Jiang, G. Lim et al. "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, Jan. 2013.

[2] L. Suarez, L. Nuaymi, and J. M. Bonnin, "An overview and classification of research approaches in green wireless networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–18, Jan. 2012.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[4] G. Yu and Y. Jiang, "Energy-efficiency region for multiple access channels," *Electron. Lett.*, vol. 50, no. 13, pp. 959–961, Jun. 2014.

[5] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[6] R. K. Sundaram, *A First Course in Optimization*, Cambridge University Press, 1996.

[7] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[9] Z. Song, Q. Ni, K. Navaie et. al., "Energy- and spectral-efficiency tradeoff with α-fairness in downlink OFDMA systems," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1265–1268, Jul. 2015.

[10] S. Huang, H. Chen, J. Cai, and F. Zhao, "Energy efficiency and spectral-efficiency tradeoff in amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4366–4378, Nov. 2013.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press

[11] J. Wu, G. Wang, and Y. R. Zheng, "Energy efficiency and spectral efficiency tradeoff in type-I ARQ systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 356–366, Feb. 2014.

[12] W. Zhang, C. X. Wang, D. Chen, and H. Xiong, "Energy–spectral efficiency tradeoff in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2208–2218, Apr. 2016.

[13] J. B. Rao and A. O. Fapojuwo, "On the tradeoff between spectral efficiency and energy efficiency of homogeneous cellular networks with outage constraint," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1801–1814, May 2013.

[14] O. Amin, E. Bedeer, M. H. Ahmed, and O. A. Dobre, "Energy efficiency–spectral efficiency tradeoff: A multiobjective optimization approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1975–1981, Apr. 2016.

[15] L. Deng, Y. Rui, P. Cheng et al., "A unified energy efficiency and spectral efficiency tradeoff metric in wireless networks," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 55–58, Jan. 2013.

[16] D. Tsilimantos, J. M. Gorce, K. Jaffres-Runser, and H. V. Poor, "Spectral and energy efficiency trade-offs in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 54–66, Jan. 2016.

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

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

[19] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, pp. 492–498, Mar. 1967.

[20] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

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

[22] K. T. S. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: A fractional programming approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2746–2757, Jul. 2013.

[23] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf.*, vol. 11, no. 3–4, pp. 185–396, Jun. 2015.

[24] Y. Jong, "An efficient global optimization algorithm for nonlinear sum-of-ratios problem," May 2012. [Online]. Available: www.optimization-online.org/DB_FILE/2012/08/3586.pdf.

[25] J. P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Math. Program.*, vol. 52, no. 2, pp. 191–207, Oct. 1991.

[26] I. Barros, J. B. G. Frenk, S. Schaible, and S. Zhang, "A new algorithm for generalized fractional programs," *Math. Program.*, vol. 72, no. 2, pp. 147–175, Feb. 1996.

[27] G. Yu, Q. Chen, R. Yin, H. Zhang, and G. Y. Li, "Joint downlink and uplink resource allocation for energy-efficient carrier aggregation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3207–3218, Jun. 2015.

[28] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Sig. Proc.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.

[29] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical Non-Linear Energy Harvesting Model and Resource Allocation for SWIPT Systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.

[30] L. Xu, G. Yu, and Y. Jiang, "Energy-efficient resource allocation in single-cell OFDMA systems: Multi-objective approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5848–5858, Oct. 2015.

[31] J. P. Crouzeix, J. A. Ferland, and S. Scbaible, "A note on an algorithm for generalized tractional programs," *J. Optimiz. Theory App.*, vol. 50, pp. 183–187, 1986.

[32] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang, "Generalized concavity," in *Mathematical Concepts and Methods in Science and Engineering*, vol. 36, Plenum Press, 1988.

[33] M. Sion, "On general minimax theorems," *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, Mar. 1958.

[34] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscip. O.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.

[35] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logist. Q.*, vol. 2, no. 1–2, pp. 83–97, Mar. 1955.

[36] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[37] L. T. H. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, no. 1–4, pp. 23–46, 2005.

[38] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," *Neur. Inf. Process. Syst.*, pp. 1759–1767, 2009.

[39] M. Hast, K. J. Astrom, B. Bernhardsson, and S. Boyd, "PID design by convex-concave optimization," in *Proc. European Control Conference,* Jul. 2013.

https://doi.org/10.1017/9781108277389.003 Published online by Cambridge University Press