

Chapter 6

Optimal Power Allocation for Downlink NOMA Systems



Yongming Huang, Jiaheng Wang and Jianyue Zhu

6.1 Introduction

With the popularity of smartphones and Internet of Things, there is an explosive demand for new services and data traffic for wireless communications. The capacity of the fourth-generation (4G) mobile communication system is insufficient to satisfy such a demand in the near future. The development of the fifth-generation (5G) mobile communication system has been placed on the agenda with higher requirements in data rates, latency, and connectivity [1]. In order to meet the new standards, some potential technologies, such as massive multiple-input–multiple-output (MIMO) [2], millimeter wave [3], and ultra densification [4, 5], will be introduced into 5G. Meanwhile, new multiple access technologies, which are flexible, reliable, and efficient in terms of energy and spectrum, are also considered for 5G communication.

Conventionally, cellular systems have adopted orthogonal multiple access (OMA) approaches, in which wireless resources are allocated orthogonally to multiple users. The common OMA techniques include frequency-division multiple access (FDMA), time division multiple access (TDMA), code-division multiple access (CDMA), and orthogonal frequency-division multiple access (OFDMA). Ideally, in OMA, the intra-cell interference does not exist as result of dedicated resource allocation. Also, for this reason, the information of multiple users can be retrieved at a low complexity. Nonetheless, the number of served users is limited by the number of orthogonal resources, which is generally small in practice. Consequently, it is difficult for OMA systems to support a massive connectivity.

Recently, non-orthogonal multiple access (NOMA) technologies are developed and proposed for 5G, which will contribute to disruptive design changes on radio access and alleviate the scarcity of suitable spectra. By using superposition coding

Y. Huang (✉) · J. Wang · J. Zhu
Southeast University, Nanjing 210096, China
e-mail: huangym@seu.edu.cn

J. Wang
e-mail: jhwang@seu.edu.cn

J. Zhu
e-mail: zhujy@seu.edu.cn

at the transmitter with successive interference cancellation (SIC) at the receiver, downlink NOMA allows one (frequency, time, code, or spatial) channel to be shared by multiple users simultaneously [6, 7], thus leading to better performance in terms of spectral efficiency, fairness, or energy efficiency [8]. Therefore, NOMA has received much attention recently. Its combinations with MIMO and multi-cell technologies were studied in [9, 10] and [11], respectively. NOMA was also considered to be used in, e.g., visible light communication [12] and millimeter wave communication [13].

The principle of NOMA is to implement multiple access in the power domain [14]. Hence, allocation is critical for NOMA systems. In the literature, there are a number of works on power allocation for NOMA. In particular [15, 16] focused on power allocation in a two-user NOMA system and [17–21] investigated power allocation for multiple users (more than two) sharing one channel, which is referred as multi-user NOMA (MU-NOMA). There were also some works, e.g., [12, 22–28], studying the resource allocation problems in multi-channel NOMA (MC-NOMA) systems, where multiple channels are available for NOMA. Different criteria, such as maximin fairness [18–20, 22], sum rate [15–17, 22, 28–30], and energy efficiency [21–23, 26], were considered.

This chapter focuses on power allocation for downlink NOMA. We first briefly review the basic concepts of downlink NOMA transmission and introduce the two-user NOMA, MU-NOMA, and MC-NOMA schemes. Then, we investigate the optimal power allocation strategies for these NOMA schemes under different performance criteria such as the maximin fairness, sum rate, and energy efficiency along with user weights and quality-of-service (QoS) constraints. We show that the optimal NOMA power allocation can be analytically characterized in most cases, otherwise it can be numerically computed via convex optimization methods.

This chapter is organized as follows. Section 6.2 introduces the fundamentals of downlink NOMA and the two-user NOMA, MU-NOMA, and MC-NOMA schemes. In Sects. 6.3–6.5, we investigate the optimal power allocation for two-user NOMA, MU-NOMA, and MC-NOMA schemes, respectively. The performance of the NOMA power allocation strategies is evaluated in Sect. 6.6 via simulations, and the conclusion is drawn in Sect. 6.7.

6.2 Fundamentals of Downlink NOMA

In this section, we review the basic concepts of downlink NOMA transmission in a single-cell network.¹ To begin with, we start from the simplest two-user case, where a base station (BS) serves two users, namely UE_1 and UE_2 , on the same frequency band with bandwidth B . The BS transmits a signal s_n for user n ($UE_n, n = 1, 2$) with transmission power p_n . The total power budget of the BS is P , i.e., $p_1 + p_2 \leq P$. Such a simple downlink NOMA system is displayed in Fig. 6.1 [31].

¹For multi-cell NOMA, the reader is referred to [11].

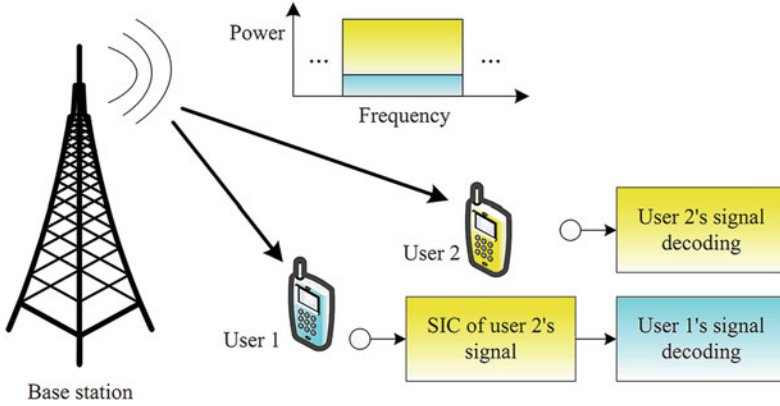


Fig. 6.1 A downlink NOMA system with one base station and two users

Superposition Coding: According to the NOMA principle, the BS exploits the superposition coding and broadcasts the signal

$$x = \sqrt{p_1}s_1 + \sqrt{p_2}s_2 \tag{6.1}$$

to both users. The received signal at UE_n is

$$y_n = h_n (\sqrt{p_1}s_1 + \sqrt{p_2}s_2) + z_n, \tag{6.2}$$

where $h_n = g_n d_n^{-\rho}$ is the channel coefficient from the BS to UE_n , g_n follows a Rayleigh distribution, d_n is the distance between the BS and UE_n , ρ is the path loss exponent, and z_n is the additive white Gaussian noise with zero mean and variance σ_n^2 , i.e., $z_n \sim \mathcal{C.N}(0, \sigma_n^2)$.

Successive Interference Cancellation (SIC): In NOMA systems, each user exploits SIC at its receiver. Let $\Gamma_n = |h_n|^2 / \sigma_n^2$ be the channel-to-noise ratio (CNR) of UE_n . Assume without loss of generality (w.l.o.g.) that the users are ordered by their normalized channel gains as $\Gamma_1 \geq \Gamma_2$, i.e., UE_1 and UE_2 are regarded as the strong and weak users, respectively. It is expected that more power is allocated to the weak user UE_2 and less power is allocated to the strong user UE_1 , i.e., $p_1 \leq p_2$ [14, 25]. Then, UE_1 first decodes the message of UE_2 and removes it from its received signal, while UE_2 treats the signal of UE_1 as interference and decodes its own message.

Achievable Rate: Suppose that the channel coding is ideal and UE_1 is able to decode the message of UE_2 successfully. Then, the achievable rates of UE_1 and UE_2 are given respectively by

$$R_1 = B \log (1 + p_1 \Gamma_1), \quad R_2 = B \log \left(1 + \frac{p_2 \Gamma_2}{1 + p_1 \Gamma_2} \right) \tag{6.3}$$

which are often used as the design objectives of NOMA systems.

Multi-User NOMA (MU-NOMA): Consider a more general case where a BS serves $N \geq 2$ users on the same spectrum, which are indexed by $n = 1, \dots, N$. The broadcast signal by the BS is then given by

$$x = \sum_{i=1}^N \sqrt{p_i} s_i \quad (6.4)$$

and then the received signal at each UE_n is given by

$$y_n = h_n \sum_{i=1}^N \sqrt{p_i} s_i + z_n. \quad (6.5)$$

Similarly, suppose that the users are ordered by their normalized channel gains as

$$\Gamma_1 \geq \Gamma_2 \geq \dots \geq \Gamma_N \quad (6.6)$$

and the NOMA protocol allocates higher powers to the users with lower CNRs, leading to $p_1 \leq p_2 \leq \dots \leq p_N$. Therefore, UE_n is able to decode the message of UE_l for $l > n$ and remove it from the received signal so that UE_n is only interfered by UE_j for $j < n$. Therefore, after SIC, the achievable rate of UE_n is

$$R_n = \log \left(1 + \frac{p_n \Gamma_n}{\sum_{j=1}^{n-1} p_j \Gamma_n + 1} \right) \quad (6.7)$$

for $n = 1, \dots, N$.

Multi-Channel NOMA (MC-NOMA): The frequency band shared by the users could be viewed as a channel, which may also be a time slot, spread code, or resource block. In cellular systems, there are often multiple channels available, which leads to a more general NOMA scheme called multi-channel NOMA (MC-NOMA), where multiple users share multiple channels. Specifically, in a downlink MC-NOMA network, the BS serves N users through M channels and the total bandwidth B is equally divided to M channels so the bandwidth of each channel is $B_c = B/M$. Let $N_m \in \{N_1, N_2, \dots, N_M\}$ be the number of users using channel m for $m = 1, 2, \dots, M$ and $UE_{n,m}$ denotes user n on channel m for $n = 1, 2, \dots, N_m$. The signal transmitted by the BS on each channel m can be expressed as

$$x_m = \sum_{n=1}^{N_m} \sqrt{p_{n,m}} s_n, \quad (6.8)$$

where s_n is the symbol of $UE_{n,m}$ and $p_{n,m}$ is the power allocated to $UE_{n,m}$. The received signal at $UE_{n,m}$ is

$$y_{n,m} = \sum_{i=1}^{N_m} \sqrt{p_{i,m}} h_{n,m} s_i + z_{n,m}. \quad (6.9)$$

It is easily seen that on each channel m is an MU-NOMA scheme. Similarly, assume w.l.o.g. that the CNRs of the users on channel m are ordered as

$$\Gamma_{1,m} \geq \cdots \geq \Gamma_{n,m} \geq \cdots \geq \Gamma_{N_m,m}, \quad (6.10)$$

which will lead to $p_{1,m} \leq \cdots \leq p_{n,m} \leq \cdots \leq p_{N_m,m}$. Then, the achievable rate of UE $_{n,m}$ using SIC is

$$R_{n,m} = B_c \log \left(1 + \frac{p_{n,m} \Gamma_{n,m}}{1 + \sum_{i=1}^{n-1} p_{i,m} \Gamma_{n,m}} \right). \quad (6.11)$$

The basic idea of NOMA is to implement multiple access in the power domain [14]. Hence, power allocation is the key to achieve the full benefit of NOMA transmission. In the following parts, we will investigate the optimal power allocation strategies for different NOMA schemes, including the simplest two-user case, the MU-NOMA scheme, and the MC-NOMA scheme, under different performance measures.

6.3 Two-User NOMA

In this section, we investigate the optimal power allocation for the two-user NOMA scheme. Although the two-user scheme is the simplest case of NOMA, the results and insights obtained in this case will serve the more complicated MU-NOMA and MC-NOMA schemes.

6.3.1 Optimal Power Allocation for MMF

The NOMA scheme enables a flexible management of users' achievable rates and provides an efficient way to enhance user fairness. A widely used fairness metric is the maximin fairness (MMF), which is achieved by maximizing the worst (i.e., minimum) user rate. According to (6.3), the power allocation to achieve the MMF is given by the solution to the following optimization problem:

$$TU^{\text{MMF}} : \begin{array}{l} \max_{p_1, p_2} \min \{R_1(p_1, p_2), R_2(p_1, p_2)\} \\ \text{s.t. } 0 \leq p_1 \leq p_2, p_1 + p_2 \leq P \end{array}$$

This problem admits a closed-form solution as follows.

Proposition 1 Suppose that $\Gamma_1 \geq \Gamma_2$. Then, the optimal solution to TU^{MMF} is given by $p_1^* = \Lambda$ and $p_2^* = P - p_1^*$, where $\Gamma_1 \triangleq |h_1|^2 / \sigma_1^2$ and

$$\Lambda \triangleq \frac{-(\Gamma_1 + \Gamma_2) + \sqrt{(\Gamma_1 + \Gamma_2)^2 + 4\Gamma_1\Gamma_2^2 P}}{2\Gamma_1\Gamma_2}. \quad (6.12)$$

Proof Please refer to the proof of Proposition 1 in [22].

Remark 1 It can be verified that at the optimal point $R_1(p_1^*, p_2^*) = R_2(p_1^*, p_2^*)$, i.e., UE₁ and UE₂ achieve the same rate. This indicates that, under the MMF criterion, the NOMA system will provide absolute fairness for two users on one channel.

To elaborate another important insight, we introduce the following definition.

Definition 1 A NOMA system is called *SIC stable* if the optimal power allocation satisfies $p_1 < p_2$ on one channel.

Remark 2 In NOMA systems, SIC is performed according to the order of the CNRs of the users on one channel [14, 25], which is guaranteed by imposing an inverse order of the powers allocated to the users, i.e., $p_1 \leq p_2$. Specifically, UE₁ (the stronger user) first decodes the signal of UE₂ (the weaker user) and then subtracts it from the superposed signal. Therefore, from the SIC perspective, a large difference between the signal strengths of UE₂ and UE₁ is preferred [32]. However, even with the power order constraint $p_1 \leq p_2$, the power optimization may lead to $p_1 = p_2$; i.e., UE₁ and UE₂ have the same signal strength, which is the worst situation for SIC. In this case, SIC may fail or has a large error propagation and thus is unstable. Indeed, the authors in [33] pointed out that the power of the weak user must be strictly larger than that of the strong user, otherwise the users' outage probabilities will always be one. Definition 1 explicitly concretizes such a practical requirement in NOMA systems.

Lemma 1 The NOMA system is SIC stable for TU^{MMF} .

Proof Please refer to the proof of Lemma 1 in [22].

Indicated by Lemma 1, the two-user NOMA system is always SIC stable under the MMF criterion, as in this case the optimal power allocation always satisfies $p_1^* < p_2^*$. On the other hand, in the subsequent subsections, we will show that a NOMA system may not always be SIC stable under different criteria.

6.3.2 Optimal Power Allocation for SR Maximization

In this subsection, we seek the optimal power allocation for maximizing the sum rate (SR). In SR maximization, to take user priority or fairness into account, user weights or quality-of-service (QoS) constraints are often adopted.

6.3.2.1 Weighted SR Maximization (SR1)

According to (6.3), the problem of maximizing the weighted SR (WSR) is given by

$$TU^{SR1} : \begin{array}{l} \max_{p_1, p_2} W_1 R_1(p_1, p_2) + W_2 R_2(p_1, p_2) \\ \text{s.t. } 0 \leq p_1 \leq p_2, p_1 + p_2 \leq P \end{array}$$

where W_i denotes the weight of UE $_i$ for $i = 1, 2$. Note that TU^{SR1} is a nonconvex problem due to the interference between UE $_1$ and UE $_2$. Nevertheless, its optimal solution can be found as follows.

Proposition 2 Suppose that $\Gamma_1 \geq \Gamma_2$, $W_1 < W_2$ and $P > 2\Omega$, with

$$\Omega \triangleq \frac{W_2 \Gamma_2 - W_1 \Gamma_1}{\Gamma_1 \Gamma_2 (W_1 - W_2)}. \quad (6.13)$$

Then, the optimal solution to TU^{SR1} is given by $p_1^* = \Omega$ and $p_2^* = P - p_1^*$.

Proof Please refer to the proof of Proposition 2 in [22].

Remark 3 In Proposition 2, the conditions $W_1 < W_2$ and $P > 2\Omega$ are both to avoid a failure of SIC. Indeed, if $W_1 \geq W_2$, the solution to TU^{SR1} is $p_1^* = p_2^* = P/2$; i.e., the NOMA system is unstable according to Definition 1. SIC may also fail if $P \leq 2\Omega$, which will lead to $p_1^* = p_2^* = P/2$ as well. Therefore, the two-user NOMA system is SIC stable for the WSR maximization if and only if $W_1 < W_2$ and $P > 2\Omega$.

6.3.2.2 SR Maximization with QoS (SR2)

Now, we consider maximizing the SR with QoS constraints. In this case, the power allocation problem is given by

$$TU^{SR2} : \begin{array}{l} \max_{p_1, p_2} R_1(p_1, p_2) + R_2(p_1, p_2) \\ \text{s.t. } 0 \leq p_1 \leq p_2, p_1 + p_2 \leq P \\ R_i \geq R_i^{\min}, i = 1, 2. \end{array}$$

where R_i^{\min} is the QoS threshold of UE $_i$. The optimal solution to TU^{SR2} is provided in the following result.

Proposition 3 Suppose that $\Gamma_1 \geq \Gamma_2$, $A_2 \geq 2$, and $P \geq \Upsilon$, with

$$A_i = 2^{R_i^{\min}}, \quad \Upsilon \triangleq \frac{A_2(A_1 - 1)}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2}, \quad \Xi \triangleq \frac{\Gamma_2 P - A_2 + 1}{A_2 \Gamma_2}. \quad (6.14)$$

Then, the optimal solution to TU^{SR2} is given by $p_1^* = \Xi$ and $p_2^* = P - p_1^*$.

Proof Please refer to the proof of Proposition 3 in [22].

Remark 4 Similarly, in Proposition 3, the conditions $A_2 \geq 2$ and $P \geq \Upsilon$ are to guarantee the SIC stability. Indeed, if $A_2 < 2$, then $\Xi > P/2$ and the optimal solution will be $p_1^* = p_2^* = P/2$, which may lead a failure of SIC. At the same time, SIC may also fail if $P < \Upsilon$, which will lead to $p_1^* = p_2^* = P/2$ as well. Therefore, the NOMA system is SIC stable in this case if and only if $A_2 \geq 2$ and $P \geq \Upsilon$.

According to Proposition 3, if the NOMA system is SIC stable, the optimal solution will be $p_1^* = \Xi$ and $p_2^* = P - p_1^*$. Hence, we have $R_2(p_1^*, p_2^*) = R_2^{\min}$, implying that the user with a lower CNR (i.e., UE₂) receives the power to meet its QoS requirement exactly, while the remaining power is used to maximize the rate of the user with a higher CNR (i.e., UE₁).

6.3.3 Optimal Power Allocation for EE Maximization

In this subsection, we investigate the optimal power allocation for maximizing the energy efficiency (EE), which is defined as the ratio between the rate and the consumed power. Similarly, user weights and QoS constraints are considered.

6.3.3.1 Weighted EE Maximization (EE1)

The problem of maximizing the weighted EE is formulated as follows:

$$TU_a^{\text{EE1}} : \begin{aligned} \max_{p_1, p_2} \quad & \eta = \frac{W_1 R_1(p_1, p_2) + W_2 R_2(p_1, p_2)}{P_T + p_1 + p_2} \\ \text{s.t.} \quad & 0 \leq p_1 \leq p_2, \quad p_1 + p_2 \leq P \end{aligned}$$

where P_T is the power consumption of the circuits. Given the fraction form of the objective, TU_a^{EE1} is more complicated than TU^{SR1} . In the following, we show that this problem can also be optimally solved.

We introduce an auxiliary variable q with $p_1 + p_2 = q$. Then, TU_a^{EE1} can be equivalently written into

$$TU_b^{\text{EE1}} : \begin{aligned} \max_{p_1, q} \quad & \eta = \frac{W_1 R_1(p_1) + W_2 R_2(p_1, q)}{P_T + q} \\ \text{s.t.} \quad & q \geq 2p_1, \quad q \leq P \end{aligned}$$

where the rate of UE₂ can be expressed as

$$R_2(p_1, q) = B \log \left(\frac{1 + q\Gamma_2}{1 + p_1\Gamma_2} \right). \quad (6.15)$$

To deal with the fractional form, let us introduce the following objective function:

$$H(p_1, q, \alpha) \triangleq W_1 R_1(p_1) + W_2 R_2(p_1, q) - \alpha(P_T + q), \quad (6.16)$$

where α is a positive parameter. Then, we consider the following problem for given α :

$$TU_c^{EE1} : \begin{array}{l} \max_{p_1, q} H(p_1, q, \alpha) \\ \text{s.t. } q \geq 2p_1, q \leq P. \end{array}$$

The relation between TU_c^{EE1} and TU_b^{EE1} is stated in the following result.

Lemma 2 ([34, pp. 493–494]) *Let $H^*(\alpha)$ be the optimal objective value of TU_c^{EE1} and $\mathbf{p}^*(\alpha)$ be the optimal solution of TU_c^{EE1} . Then, $\mathbf{p}^*(\alpha)$ is the optimal solution to TU_b^{EE1} if and only if $H^*(\alpha) = 0$.*

According to Lemma 2, the optimal solution to TU_b^{EE1} can be found by solving TU_c^{EE1} parameterized by α such that $H^*(\alpha) = 0$. Since $H^*(\alpha)$ is monotonic in α , one can use any line search method, e.g., the bisection method, to find α such that $H^*(\alpha) = 0$. Then, the left question is how to solve TU_c^{EE1} with given α .

Theorem 1 *Suppose that $\Gamma_1 \geq \Gamma_2$. Then, TU_c^{EE1} is a convex problem if one of the following conditions hold*

$$\begin{array}{l} C1 : W_1 \geq W_2; \\ C2 : 1 < \frac{W_2}{W_1} \leq \frac{(\Gamma_1 + \Gamma_1 \Gamma_2 P)^2}{(\Gamma_2 + \Gamma_1 \Gamma_2 P)^2}. \end{array}$$

Proof See Appendix A.

Theorem 1 reveals that TU_c^{EE1} is in fact a convex problem if condition C1 or C2 holds. Consequently, TU_c^{EE1} can be efficiently solved via convex optimization methods. The optimal solution to TU_c^{EE1} can further be analytically characterized.

Proposition 4 *Suppose that $\Gamma_1 \geq \Gamma_2$, C2 holds, and $P \geq 2\Omega$ with*

$$\Omega = \frac{W_2 \Gamma_2 - W_1 \Gamma_1}{\Gamma_1 \Gamma_2 (W_1 - W_2)}. \quad (6.17)$$

Then, the optimal solution to TU_c^{EE1} is $p_1^ = \Omega$ and*

$$q^* = \left[\frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2} \right]_{2\Omega}^P = \max \left\{ 2\Omega, \min \left\{ \frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2}, P \right\} \right\}. \quad (6.18)$$

Proof See Appendix B.

Remark 5 Similarly, in Proposition 4, the conditions C2 and $P > 2\Omega$ are to avoid a failure of SIC. Although TU_c^{EE1} is convex under C1, condition C1 will lead to $p_1^* = p_2^*$, which, according to Definition 1, is SIC-unstable.

6.3.3.2 EE Maximization with QoS Constraints (EE2)

Then, we consider maximizing the EE with QoS constraints. In this case, the power allocation problem is given by

$$TU_a^{\text{EE2}} : \begin{aligned} & \max_{p_1, p_2} \frac{R_1(p_1, p_2) + R_2(p_1, p_2)}{P_T + p_1 + p_2} \\ & \text{s.t. } 0 \leq p_1 \leq p_2, \quad p_1 + p_2 \leq P \\ & \quad R_i \geq R_i^{\min}, \quad i = 1, 2. \end{aligned}$$

This problem can be optimally solved following the similar steps as in the previous subsection.

Specifically, using $p_1 + p_2 = q$, TU_a^{EE2} can be equivalently transformed into

$$TU_b^{\text{EE2}} : \begin{aligned} & \max_{p_1, q} \frac{R_1(p_1) + R_2(p_1, q)}{P_T + q} \\ & \text{s.t. } q \geq 2p_1, \quad q \leq P \\ & \quad R_i \geq R_i^{\min}, \quad i = 1, 2. \end{aligned}$$

Then, we consider the following problem with given α :

$$TU_c^{\text{EE2}} : \begin{aligned} & \max_{p_1, q} Q(p_1, q, \alpha) \\ & \text{s.t. } q \geq 2p_1, \quad q \leq P \\ & \quad R_i \geq R_i^{\min}, \quad i = 1, 2 \end{aligned}$$

where

$$Q(p_1, q, \alpha) \triangleq R_1(p_1) + R_2(p_1, q) - \alpha(P_T + q). \quad (6.19)$$

According to Lemma 2, the optimal solution to TU_b^{EE2} can be found by solving TU_c^{EE2} for a given α and updating α until the optimal objective value of TU_c^{EE2} , denoted by $Q^*(\alpha)$, satisfies $Q^*(\alpha) = 0$.

From Theorem 1, under condition C1, the objective of TU_c^{EE2} is concave and TU_c^{EE2} is a convex problem too. Therefore, TU_c^{EE2} can be efficiently solved. In fact, the optimal solution to TU_c^{EE2} can also be analytically characterized.

Proposition 5 Suppose that $\Gamma_1 \geq \Gamma_2$, $A_2 \geq 2$, and $P \geq \Upsilon$, with

$$A_l = 2 \frac{R_l^{\min}}{B}, \quad \Upsilon \triangleq \frac{A_2(A_1 - 1)}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2}. \quad (6.20)$$

Then, the optimal solution to TU_c^{EE2} is

$$p_1^* = \frac{1 + q^* \Gamma_2 - A_2}{A_2 \Gamma_2}, \quad (6.21)$$

$$q^* = \left[\frac{1}{\alpha} - \frac{A_2}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2} \right]_r^P = \max \left\{ \gamma, \min \left\{ \frac{1}{\alpha} - \frac{A_2}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2}, P \right\} \right\}. \quad (6.22)$$

Proof See Appendix C.

Similarly, it can be verified that the power allocation obtained from TU_c^{EE2} (or TU_a^{EE2}) is SIC stable if and only if $P \geq \gamma$ and $A_2 \geq 2$.

6.4 MU-NOMA

In this section, we consider the more general MU-NOMA scheme, where a BS serves $N \geq 2$ users on the same channel. Similarly, the optimal MU-NOMA power allocation is investigated under the MMF, SR, and EE criteria with user weights or QoS constraints.

6.4.1 Optimal Power Allocation for MMF

According to (6.7), the power allocation problem under the MMF criterion is formulated as

$$MU_a^{\text{MMF}} : \begin{array}{l} \max_{\mathbf{p}} \min_{i=1, \dots, N} \{R_i\} \\ \text{s.t. } 0 < p_1 < \dots < p_N, \sum_{i=1}^N p_i \leq P \end{array}$$

where $\mathbf{p} = \{p_i\}_{i=1}^N$ denotes the powers allocated to the users. It has been shown in [20] that, though nonconvex, MU_a^{MMF} is a quasi-convex problem. Thus, the optimal solution to MU_a^{MMF} can be found by solving a sequence of convex problems.

Specifically, MU_a^{MMF} is equivalent to

$$MU_b^{\text{MMF}} : \begin{array}{l} \max_{\mathbf{p}, t} t \\ \text{s.t. } 0 < p_1 < \dots < p_N, \sum_{i=1}^N p_i \leq P \\ R_i \geq t, i = 1, \dots, N. \end{array}$$

The constraint $R_i \geq t$ can be rewritten into

$$p_i \geq \frac{2^t - 1}{\Gamma_i} \left(\sum_{j=1}^{i-1} p_j \Gamma_j + 1 \right). \quad (6.23)$$

Hence, for fixed t , MU_b^{MMF} is a linear program (LP) and can be efficiently solved by a number of LP solvers. Then, one can exploit the bisection method to search the optimal t . Note that the optimal solution to MU_b^{MMF} for fixed t can be analytically characterized if there is no power order constraint.

Proposition 6 *In the absence of power order constraint, the solution to MU_b^{MMF} is given by*

$$p_i = \frac{2^t - 1}{\Gamma_i} \left(\sum_{j=1}^{i-1} p_j \Gamma_j + 1 \right), i = 1, \dots, N. \quad (6.24)$$

Proof Please refer to the proof of Theorem 1 in [20].

The solution in (6.24) implies that all users achieve the same data rate equal to t . Hence, in this case, the NOMA system will provide absolute fairness for all users. Note that, however, the solution in (6.24) is obtained without the power order constraint. One may wonder if this solution is still optimal if the power order constraint is not omitted. The following result provides a sufficient condition to characterize the optimality of (6.24).

Theorem 2 *The solution in (6.24) is optimal for MU_b^{MMF} if $P \geq \chi$, where $\chi = \sum_{i=1}^N \frac{2^{N-i}}{\Gamma_i}$.*

Proof See Appendix E.

Theorem 2 indicates that the power order constraints can be omitted under some conditions. In this case, the solution in (6.24) is indeed optimal for MU_b^{MMF} . On the other hand, it is unknown if the solution in (6.24) is optimal if the condition in Theorem 2 is not satisfied. Nevertheless, in this case, one can always numerically solve the linear problem MU_b^{MMF} for fixed t .

6.4.2 Optimal Power Allocation for SR Maximization

In this subsection, we investigate the SR maximization problems in MU-NOMA systems with user weights or QoS constraints.

6.4.2.1 Weighted SR Maximization (SR1)

The weighted SR maximization for MU-NOMA is formulated as

$$\begin{aligned}
& \max_p R_{\text{sum}} = \sum_{i=1}^N W_i R_i \\
MU_a^{\text{SR1}} : & \quad \text{s.t.} \quad \sum_{i=1}^N p_i \leq P \\
& \quad p_1 \leq p_2 \leq \dots \leq p_N.
\end{aligned}$$

Unlike MU_a^{MMF} for MMF, MU_a^{SR1} in its original form is neither a convex nor quasi-convex problem, making it difficult to solve it. Nevertheless, we show that MU_a^{SR1} can be transformed into a convex problem via a linear transformation of the optimization variables.

Introduce the following variable transformation: $q_i = \sum_{j=1}^i p_j$ for $i = 1, 2, \dots, N$; and conversely $p_i = q_i - q_{i-1}$ for $i = 2, \dots, N$ and $p_1 = q_1$. In this way, we have $R_1 = \log(q_1 \Gamma_1 + 1)$ and

$$R_i = \log \left(\frac{\sum_{j=1}^i p_j \Gamma_i + 1}{\sum_{j=1}^{i-1} p_j \Gamma_i + 1} \right) = \log \left(\frac{q_i \Gamma_i + 1}{q_{i-1} \Gamma_i + 1} \right) = \log(q_i \Gamma_i + 1) - \log(q_{i-1} \Gamma_i + 1) \quad (6.25)$$

for $i = 2, \dots, N$. Therefore, the weighted sum rate can be expressed as

$$\sum_{i=1}^N W_i R_i = W_1 \log(q_1 \Gamma_1 + 1) + \sum_{i=2}^N W_i (\log(q_i \Gamma_i + 1) - \log(q_{i-1} \Gamma_i + 1)) = \sum_{i=1}^N f_i(q_i), \quad (6.26)$$

where

$$f_i(q_i) = W_i \log(q_i \Gamma_i + 1) - W_{i+1} \log(q_i \Gamma_{i+1} + 1) \quad (6.27)$$

for $i = 1, \dots, N-1$ and $f_N(q_N) = W_N \log(q_N \Gamma_N + 1)$. The power constraint $\sum_{i=1}^N p_i \leq P$ is equal to $q_N \leq P$. The power order constraint $p_1 \leq p_2 \leq \dots \leq p_N$ is equal to $q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1}$. Consequently, problem MU_a^{SR1} can be equivalently transformed into the following problem:

$$\begin{aligned}
MU_b^{\text{SR1}} : & \quad \max_q \quad \sum_{i=1}^N f_i(q_i) \\
& \quad \text{s.t.} \quad q_N \leq P \\
& \quad 0 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1}
\end{aligned}$$

Then, the following result identifies the convexity of MU_a^{SR1} (or MU_b^{SR1}).

Theorem 3 MU_a^{SR1} (or MU_b^{SR1}) is a convex problem if one of the following conditions hold for $i = 1, \dots, N-1$:

$$\begin{aligned}
T1 : & \quad W_i \geq W_{i+1}; \\
T2 : & \quad 1 < \frac{W_{i+1}}{W_i} \leq \frac{(\Gamma_i + \Gamma_i \Gamma_{i+1} P)^2}{(\Gamma_{i+1} + \Gamma_i \Gamma_{i+1} P)^2}.
\end{aligned}$$

Proof Please refer to the proof of Theorem 1 in [17].

Remark 6 Theorem 3 indicates that MU_a^{SR1} (or MU_b^{SR1}) is a convex problem under some conditions of the user weights. From T1, if the user weights are in the same

order as the channel gains, i.e., $W_1 \geq W_2 \geq \dots \geq W_N$, then the objective function is concave and the problem is convex. Note that this situation includes the most common sum rate as a special case. On the other hand, the user weights can also be in the inverse order of the channel gains, i.e., $W_1 \leq W_2 \leq \dots \leq W_N$, but in this case the ratio between W_{k+1} and W_k cannot be too large according to T2. Consequently, one can find the optimal power allocation via standard convex optimization methods, e.g., the interior point method.

6.4.2.2 SR Maximization with QoS (SR2)

Then, we consider the SR maximization problem with QoS constraints for MU-NOMA, which is given by

$$MU_a^{\text{SR2}} : \begin{aligned} & \max_p \sum_{i=1}^N R_i \\ & \text{s.t.} \quad \sum_{i=1}^N p_i \leq P, p_1 \leq p_2 \leq \dots \leq p_N \\ & \quad R_i \geq R_i^{\min}, i = 1, \dots, N \end{aligned}$$

Similarly, although MU_a^{SR2} is nonconvex in its original formulation, it can be transformed into a convex problem.

In particular, we exploit the same variable transformation: $q_i = \sum_{j=1}^i p_j$ for $i = 1, 2, \dots, N$. Then, MU_a^{SR2} is transformed into

$$MU_b^{\text{SR2}} : \begin{aligned} & \max_q \sum_{k=1}^N g_i(q_i) \\ & \text{s.t.} \quad q_N \leq P \\ & \quad (a_1 - 1) / \Gamma_1 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1} \\ & \quad q_{i-1} \leq a_i q_i - \varepsilon_i, i = 2, \dots, N \end{aligned}$$

where

$$g_i(q_i) = \log(q_i \Gamma_i + 1) - \log(q_i \Gamma_{i+1} + 1) \quad (6.28)$$

for $i = 1, \dots, N - 1$ and $g_N(q_N) = \log(q_N \Gamma_N + 1)$, $a_i = 2^{-R_i}$ and $\varepsilon_i = (1 - a_i) / \Gamma_i$. According to condition T1 in Theorem 3, the objective in MU_b^{SR2} is concave and thus MU_b^{SR2} is a convex problem. Therefore, MU_b^{SR2} can also be efficiently solved via convex optimization methods. Moreover, we show that if the power order constraint $p_1 \leq p_2 \leq \dots \leq p_N$ is absent, the optimal solution to MU_b^{SR2} can be analytically characterized.

Proposition 7 Suppose that $P \geq \sum_{i=1}^N \varphi_i$, where $\varsigma_i = 2^{R_i} - 1 / \Gamma_i$,

$$\varphi_i = \begin{cases} \varsigma_1, & i = 1 \\ \max \left\{ \varphi_{i-1}, \varsigma_i \left(1 + \Gamma_i \sum_{j=1}^{i-1} \varphi_j \right) \right\}, & i = 2, \dots, N \end{cases} \quad (6.29)$$

and the power order constraint is absent in MU_a^{SR2} and MU_b^{SR2} . Then, the solution to MU_b^{SR2} is

$$\tilde{q}_i = \begin{cases} a_{i+1}\tilde{q}_{i+1} - \varepsilon_{i+1}, & k = 1, \dots, N-1 \\ P, & k = N \end{cases} \quad (6.30)$$

and the solution to problem MU_a^{SR2} is

$$\tilde{p}_i = \begin{cases} \tilde{q}_1, & k = 1 \\ (1 - a_i)\tilde{q}_i + \varepsilon_i, & k = 2, \dots, N. \end{cases} \quad (6.31)$$

Proof Please refer to the proof of Proposition 3 and Lemma 2 in [17].

Then, a natural question is when the solution in Proposition 7 is indeed optimal with the power order constraint. The answer is given below.

Theorem 4 *The solution in (6.31) is optimal for problem MU_a^{SR2} with the power order constraint if T3: $R_2^{\min} \geq 1$ and*

$$R_i^{\min} \geq \log\left(2 - 2^{-R_{i+1}^{\min}}\right), \quad i = 3, \dots, N. \quad (6.32)$$

Proof Please refer to the proof of Theorem 2 in [17].

Corollary 1 *Condition T3 in Theorem 4 holds if $R_i^{\min} \geq 1$ for $i = 2, \dots, N$.*

Theorem 4 indicates that the power order constraint can be omitted without loss of optimality if the QoS thresholds of the last $N - 1$ users are not small. Corollary 1 specifies that the QoS thresholds are only required to be no less than 1bps/Hz, which is usually satisfied in practice. Therefore, for the SR maximization with QoS constraints, the optimal power allocation is given by Proposition 7 in practical MU-NOMA systems.

6.4.3 Optimal Power Allocation for EE Maximization

In this subsection, we investigate the EE maximization for MU-NOMA systems.

6.4.3.1 Weighted EE Maximization (EE1)

The EE maximization with user weights in an MU-NOMA system is formulated as

$$MU_a^{EE1} : \quad \begin{aligned} \max_{\mathbf{p}} \quad & \eta = \frac{\sum_{i=1}^N W_i R_i}{P_T + \sum_{i=1}^N p_i} \\ \text{s.t.} \quad & \sum_{i=1}^N p_i \leq P \\ & p_1 \leq p_2 \leq \dots \leq p_N. \end{aligned}$$

To address this problem, we follow the similar steps for MU_a^{SR1} to simplify MU_a^{EE1} . Specifically, using the variable transformation: $q_i = \sum_{j=1}^i p_j$ for $i = 1, 2, \dots, N$, MU_a^{EE1} can be reformulated as

$$MU_b^{\text{EE1}} : \begin{aligned} \max_{\mathbf{q}} \quad & \eta = \frac{\sum_{k=1}^N f_k(q_k)}{P_T + q_N} \\ \text{s.t.} \quad & q_N \leq P \\ & 0 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1} \end{aligned}$$

where $f_i(q_i)$ is defined in (6.27).

Then, we introduce the following objective function:

$$H(\mathbf{q}, \alpha) \triangleq \sum_{i=1}^N f_i(q_i) - \alpha \left(P_T + \sum_{i=1}^N p_i \right) \quad (6.33)$$

and consider the following problem parameterized by α :

$$MU_c^{\text{EE1}} : \begin{aligned} \max_{\mathbf{q}} \quad & H(\mathbf{q}, \alpha) \\ \text{s.t.} \quad & q_N \leq P \\ & 0 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1}. \end{aligned}$$

According to Lemma 2, the optimal solution to MU_b^{EE1} can be found by solving MU_c^{EE1} with α chosen such that $H^*(\alpha) = 0$, where $H^*(\alpha)$ is the optimal objective value of MU_c^{EE1} . The desirable α can be found via a line search method by exploring the monotonicity of $H^*(\alpha)$. To solve MU_c^{EE1} , we provide the following result.

Theorem 5 Given $\Gamma_1 \geq \Gamma_2 \geq \dots \geq \Gamma_N$, MU_c^{EE1} is a convex problem if T1 or T2 in Theorem 3 holds for $i = 1, \dots, N - 1$.

Proof Please refer to the proof of Theorem 1 in [17].

Theorem 5 indicates that, under the same condition in Theorem 3, MU_c^{EE1} is a convex problem. Therefore, one can efficiently compute its optimal solution via optimization tools, e.g., the interior method.

6.4.3.2 EE Maximization with QoS Constraints (EE2)

Then, we focus on maximizing EE with QoS constraints for MU-NOMA and the corresponding optimization problem is given by

$$MU_a^{\text{EE2}} : \begin{aligned} \max_{\mathbf{p}} \quad & \eta = \frac{\sum_{i=1}^N R_i}{P_T + \sum_{i=1}^N p_i} \\ \text{s.t.} \quad & \sum_{i=1}^N p_i \leq P \\ & p_1 \leq p_2 \leq \dots \leq p_N \\ & R_i \geq R_i^{\min}, i = 1, \dots, N. \end{aligned}$$

By using the same variable transformation: $q_i = \sum_{j=1}^i p_j$ for $i = 1, 2, \dots, N$, MU_a^{EE2} can be transformed into

$$MU_b^{\text{EE2}} : \begin{aligned} \max_{\mathbf{q}} \quad & \eta = \frac{\sum_{k=1}^N g_k(q_k)}{P_T + q_N} \\ \text{s.t.} \quad & q_N \leq P \\ & (a_1 - 1) / \Gamma_1 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1} \\ & q_{i-1} \leq a_i q_i - \varepsilon_i, i = 2, \dots, N \end{aligned}$$

where $g_i(q_i)$ is given in (6.28). Similarly, we introduce the following objective function

$$Q(\mathbf{q}, \alpha) \triangleq \sum_{k=1}^N g_k(q_k) - \alpha \left(P_T + \sum_{i=1}^N p_i \right), \quad (6.34)$$

and consider the problem parameterized by α :

$$MU_c^{\text{EE2}} : \begin{aligned} \max_{\mathbf{q}} \quad & Q(\mathbf{q}, \alpha) \\ \text{s.t.} \quad & q_N \leq P \\ & (a_1 - 1) / \Gamma_1 \leq q_1 \leq q_2 - q_1 \leq \dots \leq q_N - q_{N-1} \\ & q_{i-1} \leq a_i q_i - \varepsilon_i, i = 2, \dots, N. \end{aligned}$$

Similarly, to obtain the optimal solution to MU_b^{EE2} , one can solve MU_c^{EE2} for given α and search α such that the optimal objective value of MU_c^{EE2} satisfies $Q^*(\alpha) = 0$, for which we refer the reader to the previous subsection. To solve MU_c^{EE2} , we provide the following result.

Proposition 8 Suppose that $P \geq \sum_{i=1}^N \varphi_i$, where $\varsigma_i = 2^{R_i} - 1 / \Gamma_i$ and

$$\varphi_k = \begin{cases} \varsigma_1, & i = 1 \\ \max \left\{ \varphi_{i-1}, \varsigma_i \left(1 + \Gamma_i \sum_{j=1}^{i-1} \varphi_j \right) \right\}, & i = 2, \dots, N \end{cases}, \quad (6.35)$$

then MU_c^{EE2} is feasible and convex.

Proof Please refer to the proof of Theorem 1 and Proposition 3 in [17].

Proposition 8 indicates that if the power budget of BS is not too small, MU_c^{EE2} can be solved by convex optimization methods, e.g., the interior point method.

6.5 MC-NOMA

In this section, we consider the MC-NOMA scheme, where multiple users share multiple channels. In this case, the resource optimization includes power allocation and channel assignment. However, the joint optimization results in a mixed

integer problem and finding its solution requires exhaustive search [35], which causes prohibitive computational complexity. Therefore, in practice, power allocation and channel assignment are often separately and alternatively optimized [26, 28, 35]. In this section, we focus on seeking the optimal power allocation for given channel assignment.

Note that using SIC at each user's receiver causes additional complexity, which is proportional to the number of users on the same channel. Thus, in the multi-channel case, each channel is often restricted to be shared by two users [25, 26, 36], which is also beneficial to reduce the error propagation of SIC. In this section, we would also like to focus on this typical situation. In this case, suppose w.l.o.g. that the CNRs of UE_{1,m} and UE_{2,m} are ordered as $\Gamma_{1,m} \geq \Gamma_{2,m}$. Then, the rates of UE_{1,m} and UE_{2,m} on channel m are given, respectively, by

$$R_{1,m} = B_c \log(1 + p_{1,m} \Gamma_{1,m}), \quad R_{2,m} = B_c \log\left(1 + \frac{p_{2,m} \Gamma_{2,m}}{p_{1,m} \Gamma_{2,m} + 1}\right). \quad (6.36)$$

6.5.1 Optimal Power Allocation for MMF

In MC-NOMA systems, the power allocation problem under the MMF criterion is given by

$$MC_a^{\text{MMF}} : \begin{aligned} & \max_{\mathbf{p}_1, \mathbf{p}_2} \min_{m=1, \dots, M} \{R_{1,m}(p_{1,m}, p_{2,m}), R_{2,m}(p_{1,m}, p_{2,m})\} \\ & \text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, m = 1, \dots, M, \quad \sum_{m=1}^M p_{1,m} + p_{2,m} \leq P \end{aligned}$$

where $\mathbf{p}_1 = \{p_{1,m}\}_{m=1}^M$ and $\mathbf{p}_2 = \{p_{2,m}\}_{m=1}^M$. Note that MC_a^{MMF} is a nonconvex problem. To address it, we first introduce auxiliary variables $\mathbf{q} = \{q_m\}_{m=1}^M$, where q_m represents the power budget for channel m with $p_{1,m} + p_{2,m} = q_m$. Suppose that the channel power budgets $\{q_m\}_{m=1}^M$ are given. Then, MC_a^{MMF} is decomposed into a group of subproblems and each subproblem is same with TU^{MMF} in the two-user case with P replaced by q_m .

With given channel power budget q_m , the optimal power allocation for the two users on channel m has been provided in Proposition 1. We can use this result to further optimize the power budgets $\{q_m\}$. According to MC_a^{MMF} and TU^{MMF} , the corresponding power budget optimization problem is

$$MC_b^{\text{MMF}} : \begin{aligned} & \max_{\mathbf{q}} \min_{m=1, \dots, M} f_m^{\text{MMF}^*}(q_m) \\ & \text{s.t. } \sum_{m=1}^M q_m \leq P, \quad \mathbf{q} \geq \mathbf{0} \end{aligned}$$

where $f_m^{\text{MMF}^*}(q_m)$ is the optimal objective value of TU^{MMF} for each channel m . Using Proposition 1, we obtain

$$f_m^{\text{MMF}\star} \triangleq B_c \log \left(\frac{\Gamma_{2,m} - \Gamma_{1,m} + \sqrt{(\Gamma_{1,m} + \Gamma_{2,m})^2 + 4\Gamma_{1,m}\Gamma_{2,m}q_m}}{2\Gamma_{2,m}} \right). \quad (6.37)$$

Then, we show that MC_b^{MMF} has a closed-form solution.

Theorem 6 *The optimal solution to MC_b^{MMF} is given by*

$$q_m^\star = \frac{(Z(\lambda)\Gamma_{2,m} + \Gamma_{1,m})(Z(\lambda) - 1)}{\Gamma_{1,m}\Gamma_{2,m}}, \quad \forall m, \quad (6.38)$$

where

$$Z(\lambda) \triangleq X + \sqrt{X^2 + \frac{B_c}{2\lambda \sum_{m=1}^M 1/\Gamma_{1,m}}}, \quad X \triangleq \frac{\sum_{m=1}^M (\Gamma_{2,m} - \Gamma_{1,m}) / (\Gamma_{1,m}\Gamma_{2,m})}{4 \sum_{m=1}^M 1/\Gamma_{1,m}}, \quad (6.39)$$

and λ is chosen such that $\sum_{m=1}^M q_m^\star = P$.

Proof Please refer to the proof of Theorem 1 in [22].

Consequently, the optimal MC-NOMA power allocation under the MMF criterion is fully characterized by Theorem 6 and Proposition 1. It follows from (6.38) that q_m^\star is monotonically decreasing in λ , so the optimal λ satisfying $\sum_{m=1}^M q_m^\star = P$ can be efficiently found via a simple bisection method.

6.5.2 Optimal Power Allocation for SR Maximization

In this subsection, we investigate the SR maximization problem with weights or QoS constraints in MC-NOMA systems.

6.5.2.1 Weighted SR Maximization (SR1)

With given channel assignment, the problem of maximizing the weighted sum rate is formulated as the following power allocation problem:

$$MC_a^{\text{SR1}} : \begin{cases} \max_{p_1, p_2} \sum_{m=1}^M (W_{1,m} R_{1,m}(p_{1,m}, p_{2,m}) + W_{2,m} R_{2,m}(p_{1,m}, p_{2,m})) \\ \text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, m = 1, \dots, M, \sum_{m=1}^M (p_{1,m} + p_{2,m}) \leq P \end{cases}$$

To solve it, similarly we introduce auxiliary variables $\mathbf{q} = \{q_m\}_{m=1}^M$ that represent the power budgets on each channel m with $p_{1,m} + p_{2,m} = q_m$. Then, MC_a^{SR1} is decom-

posed into a group of subproblems, where each subproblem is the same with TU^{SR1} except P replaced by q_m and its solution has been provided in Proposition 2.

Next, we further optimize the power budget q_m for each channel m . According to Remark 3, to guarantee that the NOMA system is SIC stable, it is reasonable to assume that $q_m \geq \Theta_m > 2\Omega_m$ and $P \geq \sum_{m=1}^M \Theta_m$ for some positive Θ_m . Then, from MC_a^{SR1} and TU^{SR1} , the corresponding power budget optimization problem is given by

$$MC_b^{\text{SR1}} : \begin{aligned} & \max_q \sum_{m=1}^M f_m^{\text{SR1}\star}(q_m) \\ & \text{s.t.} \quad \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Theta_m, \quad \forall m \end{aligned}$$

where $f_m^{\text{SR1}\star}(q_m)$ is the optimal objective value of each subproblem. Using Proposition 2, we obtain

$$f_m^{\text{SR1}\star}(q_m) = W_{1,m} \log(1 + \Omega_m \Gamma_{1,m}) + W_{2,m} \log\left(\frac{q_m \Gamma_{2,m} + 1}{\Omega_m \Gamma_{2,m} + 1}\right). \quad (6.40)$$

It is easily seen that $f_m^{\text{SR1}\star}(q_m)$ is a concave function, so MC_b^{SR1} is a convex problem, whose solution is provided in the following result.

Theorem 7 *The optimal solution to MC_b^{SR1} is given by*

$$q_m^* = \left[\frac{W_{2,m} B_c}{\lambda} - \frac{1}{\Gamma_{2,m}} \right]_{\Theta_m}^{\infty}, \quad (6.41)$$

where λ is chosen such that $\sum_{m=1}^M q_m^* = P$.

Proof The solution to MC_b^{SR1} is given by the well-known waterfilling form.

Consequently, the optimal power allocation for the weighted sum rate maximization in MC-NOMA systems is jointly characterized by Theorem 7 and Proposition 2 under the SIC stability.

6.5.2.2 SR Maximization with QoS (SR2)

Now, we consider maximizing the SR with QoS constraints. In this case, the power allocation problem is given by

$$MC_a^{\text{SR2}} : \begin{aligned} & \max_{p_1, p_2} \sum_{m=1}^M (R_{1,m}(p_{1,m}, p_{2,m}) + R_{2,m}(p_{1,m}, p_{2,m})) \\ & \text{s.t.} \quad 0 \leq p_{1,m} \leq p_{2,m}, \quad m = 1, \dots, M, \quad \sum_{m=1}^M (p_{1,m} + p_{2,m}) \leq P \\ & \quad R_{n,m} \geq R_{n,m}^{\min}, \quad n = 1, 2, \quad m = 1, \dots, M. \end{aligned}$$

We use the similar method to address MC_a^{SR2} . By introducing the power budget q_m on each channel m , MC_a^{SR2} decomposes into several subproblems and each of them has the same structure as TU^{SR2} . Thus, the optimal solution to each subproblem is given in Proposition 3 with P replaced by q_m .

Then, we focus on optimizing the power budget q_m for each channel. Similarly, according to Remark 4, to guarantee the NOMA system is SIC stable, we assume that $q_m \geq \Upsilon_m$ and $P \geq \sum_{m=1}^M \Upsilon_m$. According to MC_a^{SR2} and TU^{SR2} , the corresponding power budget optimization problem is as follows

$$MC_b^{\text{SR2}} : \begin{aligned} & \max_q \sum_{m=1}^M f_m^{\text{SR2}^*}(q_m) \\ & \text{s.t.} \quad \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Upsilon_m, \quad \forall m \end{aligned}$$

where $f_m^{\text{SR2}^*}(q_m)$ is the optimal objective value of each subproblem and given by

$$f_m^{\text{SR2}^*}(q_m) = B_c \log \left(\frac{A_{2,m}\Gamma_{2,m} - A_{2,m}\Gamma_{1,m} + \Gamma_{1,m}\Gamma_{2,m}q_m + \Gamma_{1,m}}{A_{2,m}\Gamma_{2,m}} \right) + R_{2,m}^{\min}. \quad (6.42)$$

Since $f_m^{\text{SR2}^*}(q_m)$ is a concave function, MC_b^{SR2} is a convex problem, whose solution is also given in a waterfilling form.

Theorem 8 *The optimal solution to MC_b^{SR2} is given by*

$$q_m^* = \left[\frac{B_c}{\lambda} - \frac{A_{2,m}}{\Gamma_{1,m}} + \frac{A_{2,m}}{\Gamma_{2,m}} - \frac{1}{\Gamma_{2,m}} \right]_{\Upsilon_m}^{\infty}, \quad (6.43)$$

where λ is chosen such that $\sum_{m=1}^M q_m^* = P$.

Proof The proof is simple and thus omitted.

Therefore, the optimal power allocation for the SR maximization with QoS constraints in MC-NOMA systems is jointly characterized by Proposition 3 and Theorem 8.

6.5.3 Optimal Power Allocation for EE Maximization

In this subsection, we investigate the optimal power allocation for maximizing the EE with weights or QoS constraints in MC-NOMA systems.

6.5.3.1 EE Maximization with Weights (EE1)

With given channel assignment, the problem of maximizing the weighted EE is formulated as the following power allocation problem:

$$MC_a^{EE1} : \max_{p_1, p_2} \frac{\sum_{m=1}^M (W_{1,m} R_{1,m}(p_{1,m}, p_{2,m}) + W_{2,m} R_{2,m}(p_{1,m}, p_{2,m}))}{P_T + \sum_{m=1}^M (p_{1,m} + p_{2,m})}$$

$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, m = 1, \dots, M, \sum_{m=1}^M (p_{1,m} + p_{2,m}) \leq P.$$

The difficulties in solving MC_a^{EE1} lie in its nonconvex and fractional objective. In the following, we will show that this problem can also be optimally solved.

We use the similar trick to address this problem, i.e., introducing the auxiliary variables $\{q_m\}_{m=1}^M$ with $p_{1,m} + p_{2,m} = q_m$ for each channel m . Then, MC_a^{EE1} is decomposed into a group of subproblems. Each subproblem is the same with TU^{SR1} except P replaced by q_m , and thus, its solution is provided in Proposition 1.

Then, we concentrate on searching the optimal power budget q_m for each channel. Similarly, to guarantee the NOMA system is SIC stable, it is assumed that $q_m \geq \Theta_m > 2\Omega_m$ and $P \geq \sum_{m=1}^M \Theta_m$ for some positive Θ_m . According to Proposition 1 and MC_a^{EE1} , the power budget optimization problem is formulated as

$$MC_b^{EE1} : \max_q \eta(\mathbf{q}) \triangleq \frac{\sum_{m=1}^M f_m^{SR1*}(q_m)}{P_T + \sum_{m=1}^M q_m} \quad (6.44)$$

$$\text{s.t. } \sum_{m=1}^M q_m \leq P, q_m \geq \Theta_m, \forall m$$

where $f_m^{SR1*}(q_m)$ is given in (6.40). Although $f_m^{SR1*}(q_m)$ is a concave function, MC_b^{EE1} is nonconvex due to the fraction form. To solve it, we introduce the following objective function:

$$H(\mathbf{q}, \alpha) \triangleq \sum_{m=1}^M f_m^{SR1*}(q_m) - \alpha \left(P_T + \sum_{m=1}^M q_m \right)$$

$$= \sum_{m=1}^M \left(\tilde{R}_{1,m} + W_{2,m} \log \left(\frac{q_m \Gamma_{2,m} + 1}{\Omega_m \Gamma_{2,m} + 1} \right) \right) - \alpha \left(P_T + \sum_{m=1}^M q_m \right), \quad (6.45)$$

where $\tilde{R}_{1,m} \triangleq W_{1,m} \log(1 + \Omega_m \Gamma_{1,m})$ and α is a positive parameter. Then, we consider the following convex problem with given α :

$$MC_c^{EE1} : \begin{aligned} & \max_{\mathbf{q}} H(\mathbf{q}, \alpha) \\ & \text{s.t. } \sum_{m=1}^M q_m \leq P, q_m \geq \Theta_m, \forall m. \end{aligned}$$

According to Lemma 2, the optimal solution to MC_b^{EE1} can be found by solving MC_c^{EE1} with given α and then updating α until $H^*(\alpha) = 0$. Hence, we first solve MC_c^{EE1} with given α , whose solution is provided in the following result.

Theorem 9 *The optimal solution to MC_c^{EE1} is*

$$q_m^* = \left[\frac{W_{2,m} B_c}{\alpha + \lambda} - \frac{1}{\Gamma_{2,m}} \right]_{\Theta_m}^{\infty}, \quad (6.46)$$

where λ is chosen such that $\sum_{m=1}^M q_m^* = P$.

Proof The solution is obtained by exploiting the KKT conditions of MC_c^{EE1} .

After the optimal solution to MC_c^{EE1} is obtained, we shall find an α such that $H^*(\alpha) = 0$. Since $H^*(\alpha)$ is monotonic in α , one can use the bisection method to find α . Thereby, the optimal power allocation for the weighted EE maximization in MC-NOMA systems is provided Proposition 2 and Theorem 9.

6.5.3.2 EE Maximization with QoS (EE2)

In this part, we consider maximizing the EE with QoS constraints. The corresponding power allocation problem is given by

$$MC_a^{EE2} : \begin{aligned} & \max_{p_1, p_2} \frac{\sum_{m=1}^M (R_{1,m}(p_{1,m}, p_{2,m}) + R_{2,m}(p_{1,m}, p_{2,m}))}{P_T + \sum_{m=1}^M (p_{1,m} + p_{2,m})} \\ & \text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, m = 1, \dots, M, \sum_{m=1}^M (p_{1,m} + p_{2,m}) \leq P \\ & R_{l,m} \geq R_{l,m}^{\min}, l = 1, 2, m = 1, \dots, M. \end{aligned}$$

We can use the similar method to solve MC_a^{EE2} . Briefly, we also adopt $\{q_m\}_{m=1}^M$ with $p_{1,m} + p_{2,m} = q_m$ and decompose MC_a^{EE2} into a group of subproblems, whose solution is coincided with TU^{SR2} and provided in Proposition 3.

Next, we optimize the channel power budget q_m for each channel. First, we assume that $q_m \geq \Upsilon_m$ and $P \geq \sum_{m=1}^M \Upsilon_m$ to guarantee the SIC stability. Then, according to Proposition 3 and MC_a^{EE2} , the power budget optimization problem is given by

$$MC_b^{EE2} : \max_{\mathbf{q}} \eta(\mathbf{q}) \triangleq \frac{\sum_{m=1}^M f_m^{SR2^*}(q_m)}{P_T + \sum_{m=1}^M q_m}$$

$$\text{s.t. } \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Upsilon_m, \quad \forall m$$

where $f_m^{SR2^*}(q_m)$ is given in (6.42). To solve MC_b^{EE2} , we introduce the objective function parameterized by α :

$$Q(\mathbf{q}, \alpha) \triangleq \sum_{m=1}^M f_m^{SR2^*}(q_m) - \alpha \left(P_T + \sum_{m=1}^M q_m \right)$$

$$= \sum_{m=1}^M \left(W_{1,m} \log \left(\frac{A_{2,m} \Gamma_{2,m} - A_{2,m} \Gamma_{1,m} + \Gamma_{1,m} \Gamma_{2,m} q_m + \Gamma_{1,m}}{A_{2,m} \Gamma_{2,m}} \right) + R_{2,m}^{\min} \right)$$

$$- \alpha \left(P_T + \sum_{m=1}^M q_m \right), \quad (6.47)$$

and formulate the following problem with given α :

$$MC_c^{EE2} : \max_{\mathbf{q}} Q(\mathbf{q}, \alpha)$$

$$\text{s.t. } \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Upsilon_m, \quad \forall m.$$

Then, from Lemma 2, we shall solve MC_c^{EE2} , which is a convex problem since $Q(\mathbf{q}, \alpha)$ is concave in \mathbf{q} . The optimal solution to MC_c^{EE2} is provided below.

Theorem 10 *The optimal solution to MC_c^{EE2} is*

$$q_m^* = \left[\frac{W_{1,m} B_c}{\lambda + \alpha} - \frac{A_{2,m}}{\Gamma_{1,m}} + \frac{A_{2,m}}{\Gamma_{2,m}} - \frac{1}{\Gamma_{2,m}} \right]_{\Upsilon_m}^{\infty}, \quad (6.48)$$

where λ is chosen such that $\sum_{m=1}^M q_m^* = P$.

Proof The solution is obtained by exploiting the KKT conditions of MC_c^{EE2} .

Then, we can exploit the bisection method to find an α such that the optimal objective value of MC_c^{EE2} satisfies $Q^*(\alpha) = 0$. Consequently, the optimal power allocation for the EE maximization with QoS constraints in MC-NOMA systems is obtained by using Theorem 10 and Proposition 3.

6.6 Numerical Results

This section evaluates the performance of the optimal power allocation investigated in this chapter. In simulations, the BS is located in the cell center and the users are randomly distributed in a circular range with a radius of 500 m. The minimum distance between users is set to be 40 m, and the minimum distance between the users and the BS is 50 m. Each channel coefficient follows an i.i.d. Gaussian distribution as $g \sim \mathcal{CN}(0, 1)$ and the path loss exponent is $\rho = 2$. The total power budget of the BS is $P = 41$ dBm and the circuit power consumption is $P_T = 30$ dBm. The noise power is $\sigma^2 = BN_0/M$, where the bandwidth is $B = 5$ MHz and the noise power spectral density is $N_0 = -174$ dBm.

First, we evaluate the performance of the proposed optimal power solutions for two-user NOMA ($N = 2$) and MU-NOMA ($N = 6$) systems. The user weights satisfy $W_{i+1}/W_i = 0.5$ for $i = 1, \dots, N - 1$ and the QoS thresholds to be $R_i^{\min} = 2$ bps/Hz for $i = 1, \dots, N$. In addition, we compare the NOMA schemes with OFDMA and the DC (difference of two convex functions) approach in [26], where the power allocation is optimized via waterfilling and via DC programming, respectively.

Figure 6.2 shows the minimum user rates of the two-user NOMA and MU-NOMA schemes using the optimal power allocation under the MMF criterion and the minimum user rate of the OFDMA scheme for different total power budgets and user numbers. The minimum user rate in the NOMA system is higher than that in the OFDMA system especially in the two-user case, implying that NOMA provides better fairness than OFDMA.

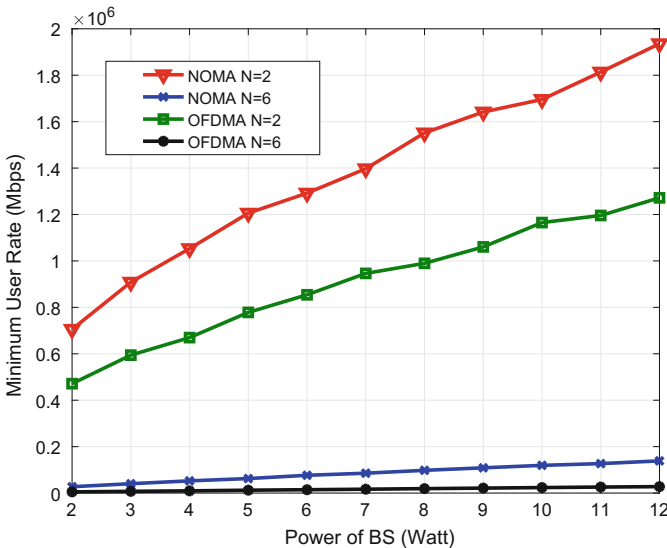


Fig. 6.2 Minimum user rate for different number of users versus BS power

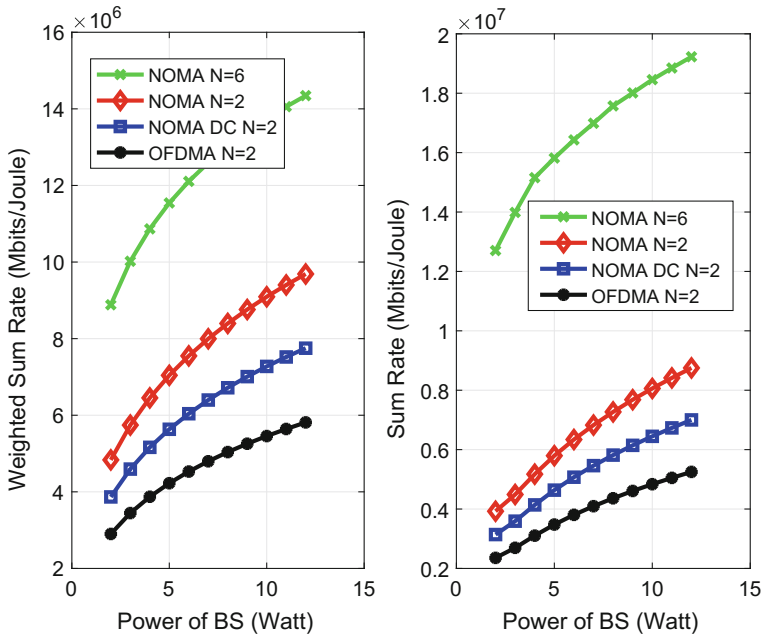


Fig. 6.3 Sum rate versus BS power

In Fig. 6.3 the left subfigure shows the weighted sum rate and the right subfigure shows the sum rate with QoS constraints. Here, in each subfigure, we compare the proposed methods with the OFDMA scheme and the NOMA scheme using DC programming in the two-user case. While NOMA outperforms OFDMA, NOMA with the optimal power allocation also achieves a higher sum rate than the DC approach, as the DC approach generally leads to a suboptimal power allocation. Meanwhile, as expected, the (weighted) sum rate increases with the user number, implying the potential of NOMA. In Fig. 6.4, the similar phenomenon can be observed, i.e., NOMA using the optimal power allocation outperforms OFDMA as well as the (suboptimal) DC approach in terms of energy efficiency.

Then, we show the performance of the optimal power allocation in MC-NOMA systems. The user weights are set to be $W_{1,m} = 0.9$ and $W_{2,m} = 1.1$ for $\forall m$ and the QoS thresholds are set to be $R_{l,m}^{\min} = 2$ bps/Hz for $l = 1, 2, \forall m$. In Fig. 6.5, we compare the joint resource allocation (JRA) method, which uses the optimal power allocation and the matching algorithm [22, 26] for channel assignment, with the exhaustive search (ES), which provides the jointly optimal solution but has high complexity. We set the number of users $N = 6$ and the power budget of the BS ranges from 2 to 12 W. From Fig. 6.5, the performance of JRA is very close to the

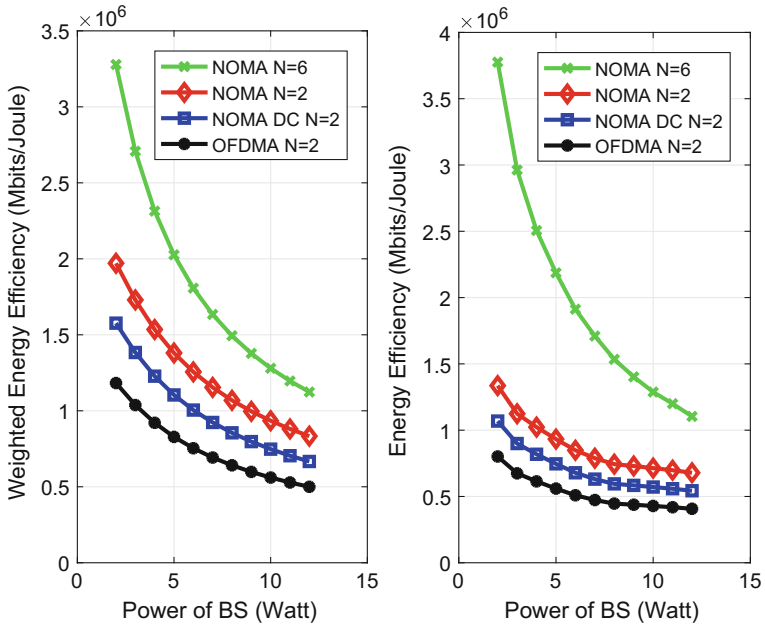


Fig. 6.4 Energy efficiency versus BS power

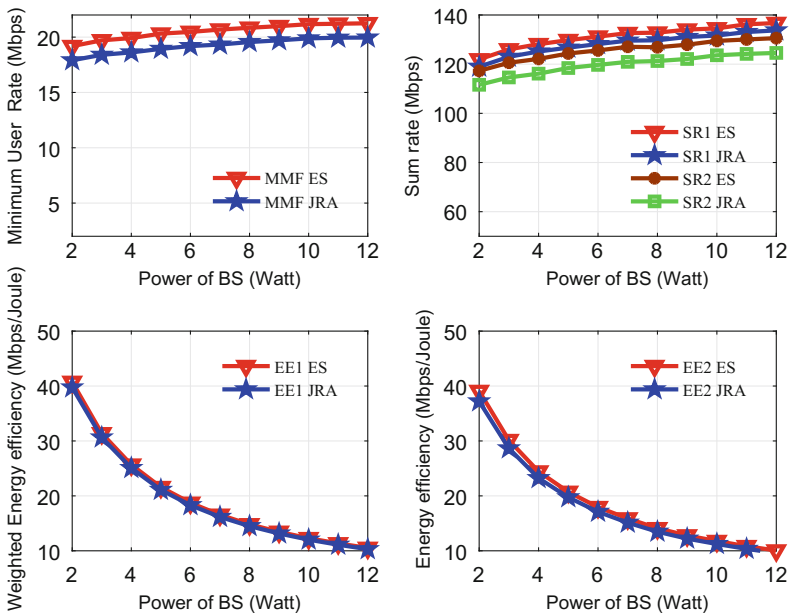


Fig. 6.5 Comparison with the exhaustive search (ES)

globally optimal value and the maximum gap is less than 5%. Therefore, the optimal power allocation method along with efficient (suboptimal) matching algorithm is able to achieve near-optimal performance with low complexity.

6.7 Conclusion

In this chapter, we discussed a promising multiple access technology, i.e., NOMA, for 5G networks and focused on the key problem of power allocation in NOMA systems. We have investigated the optimal power allocation for different NOMA schemes, including the two-user MU-NOMA, and MC-NOMA schemes. The optimal power allocation was derived under different performance measures, including the maximin fairness, weighted sum rate, and energy efficiency, wherein user weights or QoS constraints were also considered. We showed that in most cases the optimal NOMA power allocation admits an analytical solution, while in other cases it can be numerically computed via convex optimization methods.

Appendix

A. Proof of Theorem 1

Since the constraints in TU_c^{EE1} are all linear, it suffices to investigate the concavity of $H(p_1, q, \alpha)$. The second-order derivative of $H(p_1, q, \alpha)$ with respect to p_1 is

$$\frac{\partial^2 H}{\partial p_1^2} = \frac{1}{\ln 2} \left(\frac{\Upsilon \Theta}{(p_1 \Gamma_1 + 1)^2 (p_1 \Gamma_2 + 1)^2} \right), \quad (6.49)$$

where $\Upsilon = \sqrt{W_2 B} \Gamma_2 (p_1 \Gamma_1 + 1) + \sqrt{W_1 B} \Gamma_1 (p_1 \Gamma_2 + 1)$ and $\Theta = \sqrt{W_2 B} \Gamma_2 - \sqrt{W_1 B} \Gamma_1 + \sqrt{B} \Gamma_1 \Gamma_2 p_1 (\sqrt{W_2} - \sqrt{W_1})$. Given $\Gamma_1 \geq \Gamma_2$, if $W_1 \geq W_2$, then $\frac{\partial^2 H}{\partial p_1^2} \leq 0$. On the other hand, with $q \leq P$ and if C2 holds, we have

$$\sqrt{W_2 B} \Gamma_2 - \sqrt{W_1 B} \Gamma_1 + \sqrt{B} \Gamma_1 \Gamma_2 p_1 (\sqrt{W_2} - \sqrt{W_1}) \sqrt{W_2 B} \Gamma_2 - \sqrt{W_1 B} \Gamma_1 + (\sqrt{W_2} - \sqrt{W_1}) \sqrt{B} \Gamma_1 \Gamma_2 P \leq 0, \quad (6.50)$$

also implying $\frac{\partial^2 H}{\partial p_1^2} \leq 0$. Following the similar manner, it can be verified that $\frac{\partial^2 H}{\partial q^2} \leq 0$, $\frac{\partial^2 H}{\partial q \partial p_1} = 0$ and $\frac{\partial^2 H}{\partial p_1 \partial q} = 0$. Therefore, the Hessian matrix

$$\begin{pmatrix} \frac{\partial^2 H}{\partial p_1^2} & \frac{\partial^2 H}{\partial q \partial p_1} \\ \frac{\partial^2 H}{\partial p_1 \partial q} & \frac{\partial^2 H}{\partial q^2} \end{pmatrix}$$

is a negative semidefinite matrix, indicating that $H(p_1, q, \alpha)$ is a concave in (p_1, q) .

B. Proof of Proposition 4

The Lagrange of TU_c^{EE1} is given by

$$L = W_1 R_1(p_1) + W_2 R_2(p_1, q) - \alpha (P_T + q) + \mu (q - 2p_1) - \lambda (q - P) \quad (6.51)$$

with Lagrange multipliers μ and $\lambda \geq 0$. According to Theorem 1, TU_c^{EE1} is a convex problem under condition C1 or C2. Therefore, its optimal solution is characterized by the following Karush–Kuhn–Tucker (KKT) conditions:

$$\frac{\partial L}{\partial p_1} = \frac{W_1 B \Gamma_1}{\ln 2 (1 + p_1 \Gamma_1)} - \frac{W_2 B \Gamma_2}{\ln 2 (1 + p_1 \Gamma_2)} - 2\mu = 0, \quad (6.52)$$

$$\frac{\partial L}{\partial q} = \frac{W_2 B \Gamma_2}{\ln 2 (1 + q \Gamma_2)} - \alpha + \mu - \lambda = 0, \quad (6.53)$$

$$\mu (q - 2p_1) = 0, \quad (6.54)$$

$$\lambda (q - P) = 0. \quad (6.55)$$

According to Definition 1, if $p_1 = q/2$, then the NOMA system is SIC-unstable. Therefore, from (6.54), considering the SIC stability, we have $\mu = 0$. Hence, from (6.52) we obtain the optimal $p_1^* = \Omega$. It follows from (6.55) that if $q < P$, then $\lambda = 0$. Then, from (6.53) we obtain

$$2\Omega \leq q = \frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2} < P. \quad (6.56)$$

On the other hand, if $q = P$, then from (6.53) we have

$$\lambda = \frac{W_2 B \Gamma_2}{\ln 2 (1 + P \Gamma_2)} - \alpha \geq 0, \quad (6.57)$$

which leads to

$$\frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2} \geq P. \quad (6.58)$$

Therefore, the optimal q is given by $q^* = \left[\frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2} \right]_{2\Omega}^P$.

C. Proof of Proposition 5

The Lagrange of TU_c^{EE2} is given by

$$L = R_1(p_1) + R_2(p_1, q) - \alpha(P_T + q) + \mu(q - 2p_1) - \lambda(q - P) \quad (6.59)$$

$$+ \sigma_1 \left(p_1 - \frac{A_1 - 1}{\Gamma_1} \right) + \sigma_2(1 + q\Gamma_2 - A_2 - A_2p_1\Gamma_2),$$

where μ, λ, σ_1 , and σ_2 are the Lagrange multipliers. The optimal solution is characterized by the following KKT conditions:

$$\frac{\partial L}{\partial p_1} = \frac{B\Gamma_1}{\ln 2(1 + p_1\Gamma_1)} - \frac{B\Gamma_2}{\ln 2(1 + p_1\Gamma_2)} - 2\mu + \sigma_1 - \sigma_2 A_2 \Gamma_2 = 0, \quad (6.60)$$

$$\frac{\partial L}{\partial q} = \frac{B\Gamma_2}{\ln 2(1 + q\Gamma_2)} - \alpha + \mu - \lambda + \sigma_2 \Gamma_2 = 0, \quad (6.61)$$

$$\mu(q - 2p_1) = 0, \quad (6.62)$$

$$\lambda(q - P) = 0, \quad (6.63)$$

$$\sigma_1 \left(p_1 - \frac{A_1 - 1}{\Gamma_1} \right) = 0, \quad (6.64)$$

$$\sigma_2(1 + q\Gamma_2 - A_2 - A_2p_1\Gamma_2) = 0. \quad (6.65)$$

In (6.62), considering the SIC stability, we have $q > 2p_1$ and hence $\mu = 0$. Note that $\sigma_2 \neq 0$. To see this, if $\sigma_2 = 0$, according to (6.60), we have

$$\frac{B\Gamma_1}{\ln 2(1 + p_1\Gamma_1)} - \frac{B\Gamma_2}{\ln 2(1 + p_1\Gamma_2)} + \sigma_1 = 0 \quad (6.66)$$

which, however, does not hold since $\frac{B\Gamma_1}{\ln 2(1 + p_1\Gamma_1)} - \frac{B\Gamma_2}{\ln 2(1 + p_1\Gamma_2)} + \sigma_1 > 0$ with $\Gamma_1 \geq \Gamma_2$.

We consider two cases: (1) $\sigma_1 \neq 0, \sigma_2 \neq 0$; and (2) $\sigma_1 = 0, \sigma_2 \neq 0$. First, if $\sigma_1 \neq 0, \sigma_2 \neq 0$, the optimal solution can be easily obtained as

$$p_1^* = \frac{1 + q^*\Gamma_2 - A_2}{A_2\Gamma_2}, \quad q^* = \Upsilon \quad (6.67)$$

from (6.64) and (6.65). Then, if $\sigma_1 = 0, \sigma_2 \neq 0$, according to (6.60) and (6.61), we have

$$\frac{A_2\Gamma_2}{(1 + q\Gamma_2)} + \left(\frac{1}{1/\Gamma_1 + p_1} - \frac{1}{1/\Gamma_2 + p_1} \right) = (\alpha + \lambda) A_2. \quad (6.68)$$

From (6.65), we obtain $p_1 = \frac{1+q\Gamma_2-A_2}{A_2\Gamma_2}$, which along with (6.68) leads to

$$q^* = \frac{1}{\alpha + \lambda} - \frac{A_2}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2}. \quad (6.69)$$

It follows from (6.63) that if $q < P$, then $\lambda = 0$. From (6.69), we obtain

$$\Upsilon \leq q = \frac{1}{\alpha} - \frac{A_2}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2} < P. \quad (6.70)$$

On the other hand, if $q = P$, then from (6.53) we have

$$\lambda = \frac{\Gamma_1\Gamma_2}{A_2\Gamma_2 - (A_2 - 1)\Gamma_1 + P\Gamma_2\Gamma_1} - \alpha \geq 0, \quad (6.71)$$

which leads to

$$\frac{1}{\alpha} - \frac{A_2}{\Gamma_1} + \frac{A_2 - 1}{\Gamma_2} \geq P. \quad (6.72)$$

Therefore, optimal q is given by $q^* = \left[\frac{W_2 B}{\alpha \ln 2} - \frac{1}{\Gamma_2} \right]^P$.

D. Proof of Theorem 2

Let $q_i = \sum_{j=1}^i p_j$, then $q_N = P$ and $p_i = \frac{2^t - 1}{\Gamma_i} \left(\sum_{j=1}^{i-1} p_j \Gamma_j + 1 \right)$ can be transformed into $q_i = q_{i-1} 2^t + \frac{2^t - 1}{\Gamma_i}$. Thus, we obtain $P = q_N = \sum_{i=1}^N \frac{(2^t - 1) 2^{(N-i)t}}{\Gamma_i} \geq \chi$, implying $t \geq 1$ and $p_i \geq p_{i-1}$ for $i = 2, \dots, N$. Therefore, this solution satisfies the power order constraint.

References

1. J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C. Soong, J.C. Zhang, What will 5G be? *IEEE J. Sel. Areas Commun.* **32**(6), 1065–1082 (2014)
2. V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apfelrojd, T. Svensson, The role of small cells, coordinated multipoint, and massive MIMO in 5G. *IEEE Commun. Mag.* **52**(5), 44–51 (2014)
3. M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G.K. Karagiannidis, E. Bjrnson, K. Yang, I. Chih-Lin, A. Ghosh, Millimeter wave communications for future mobile networks. *IEEE J. Sel. Areas Commun.* **35**(9), 1909–1935 (2017)
4. J.G. Andrews, H. Claussen, M. Dohler, S. Rangan, M.C. Reed, Femtocells: past, present, and future. *IEEE J. Sel. Areas Commun.* **30**(3), 497–508 (2012)
5. J. Wang, W. Guan, Y. Huang, R. Schober, X. You, Distributed optimization of hierarchical small cell networks: a GNEP framework. *IEEE J. Sel. Areas Commun.* **35**(2), 249–264 (2017)

6. Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, K. Higuchi, Non-orthogonal multiple access (NOMA) for cellular future radio access, in *Proceeding of IEEE Vehicular Technology Conference (VTC Spring)* Dresden, Germany, June 2013, pp. 1–5
7. Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, H.V. Poor, Application of non-orthogonal multiple access in LTE and 5G networks, *IEEE Commun. Mag.* **55**(2), 185–191 (2017)
8. L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, Z. Wang, Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. *IEEE Commun. Mag.* **53**(9), 74–81 (2015)
9. Z. Ding, R. Schober, H.V. Poor, A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment. *IEEE Trans. Wirel. Commun.* **15**(6), 4438–4454 (2016)
10. Z. Ding, F. Adachi, H.V. Poor, The application of MIMO to non-orthogonal multiple access. *IEEE Trans. Wirel. Commun.* **15**(1), 537–552 (2016)
11. W. Shin, M. Vaezi, B. Lee, D.J. Love, J. Lee, H.V. Poor, Non-orthogonal multiple access in multi-cell networks: theory, performance, and practical challenges. *IEEE Commun. Mag.* **55**(10), 176–183 (2017)
12. X. Zhang, Q. Gao, C. Gong, Z. Xu, User grouping and power allocation for NOMA visible light communication multi-cell networks. *IEEE Commun. Lett.* **21**(4), 777–780 (2017)
13. Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, X. You, Signal processing for MIMO-NOMA: present and future challenges. *IEEE Wirel. Commun.* **25**(2), 32–38 (2018)
14. L. Zhang, M. Xiao, G. Wu, M. Alam, Y.C. Liang, S. Li, A survey of advanced techniques for spectrum sharing in 5G networks. *IEEE Wirel. Commun.* **24**(5), 44–51 (2017)
15. C.-L. Wang, J.-Y. Chen, Y.-J. Chen, Power allocation for a downlink non-orthogonal multiple access system. *IEEE Wirel. Commun. Lett.* **5**(5), 532–535 (2016)
16. Z. Yang, Z. Ding, P. Fan, N. Al-Dhahir, A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems. *IEEE Trans. Wirel. Commun.* **15**(11), 7244–7257 (2016)
17. J. Wang, Q. Peng, Y. Huang, H.M. Wang, X. You, Convexity of weighted sum rate maximization in NOMA systems. *IEEE Signal Process. Lett.* **24**(9), 1323–1327 (2017)
18. J. Choi, Power allocation for max-sum rate and max-min rate proportional fairness in NOMA. *IEEE Commun. Lett.* **20**(10), 2055–2058 (2016)
19. J. Cui, Z. Ding, P. Fan, A novel power allocation scheme under outage constraints in NOMA systems. *IEEE Signal Process. Lett.* **23**(9), 1226–1230 (2016)
20. S. Timotheou, I. Krikidis, Fairness for non-orthogonal multiple access in 5G systems. *IEEE Signal Process. Lett.* **22**(10), 1647–1651 (2015)
21. Y. Zhang, H.M. Wang, T.X. Zheng, Q. Yang, Energy-efficient transmission design in non-orthogonal multiple access. *IEEE Trans. Veh. Technol.* **66**(3), 2852–2857 (2017)
22. J. Zhu, J. Wang, Y. Huang, S. He, X. You, L. Yang, On optimal power allocation for downlink non-orthogonal multiple access systems. *IEEE J. Sel. Areas Commun.* **35**(12), 2744–2757 (2017)
23. F. Fang, H. Zhang, J. Cheng, S. Roy, V.C.M. Leung, Joint user scheduling and power allocation optimization for energy efficient NOMA systems with imperfect CSI. *IEEE J. Sel. Areas Commun.* **35**(12), 2874–2885 (2017)
24. B. Di, S. Bayat, L. Song, Y. Li, Radio resource allocation for downlink non-orthogonal multiple access (NOMA) networks using matching theory, in *Proc. of IEEE Global Communication Conference (GLOBECOM)*, pp. 1–6
25. Z. Ding, M. Peng, H.V. Poor, Cooperative non-orthogonal multiple access in 5G systems. *IEEE Commun. Lett.* **19**(8), 1462–1465 (2015)
26. F. Fang, H. Zhang, J. Cheng, V.C. Leung, Energy-efficient resource allocation for downlink non-orthogonal multiple access network. *IEEE Trans. Commun.* **64**(9), 3722–3732 (2016)
27. L. Lei, D. Yuan, C.K. Ho, S. Sun, Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems. in *Proceeding of IEEE Global Communication Conference (GLOBECOM)*, San Diego, CA, Dec 2015, pp. 1–6

28. P. Parida, S.S. Das, Power allocation in OFDM based NOMA systems: a DC programming approach, in *Proceeding of IEEE Globecom Workshops*, Dec 2014, pp. 1026–1031
29. Y. Sun, D.W.K. Ng, Z. Ding, R. Schober, Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems. *IEEE Trans. Commun.* **65**(3), 1077–1091 (2017)
30. M.R. Hojeij, J.Farah, C.A. Nour, C. Douillard, Resource allocation in downlink non-orthogonal multiple access (NOMA) for future radio access, in *Proceeding of IEEE Vehicular Technology Conference (VTC Spring)*, Dresden, Germany, May 2015, pp. 1–6
31. Z. Wei, J. Yuan, D.W.K. Ng, M. ElKashlan, Z. Ding, A Survey of Downlink Non-orthogonal Multiple Access for 5G Wireless Communication Networks. CoRR. [arXiv: 1609.01856](https://arxiv.org/abs/1609.01856), <https://dblp.org/rec/bib/journals/corr/WeiYNED16> (2016)
32. M.S. Ali, H. Tabassum, E. Hossain, Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems. *IEEE Access* **4**, 6325–6343 (2016)
33. Z. Ding, Z. Yang, P. Fan, H.V. Poor, On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Lett.* **21**(12), 1501–1505 (2014)
34. W. Dinkelbach, On nonlinear fractional programming. *Manag. Sci.* **13**(7), 492–498 (1967)
35. S. Zhang, B. Di, L. Song, Y. Li, Radio resource allocation for non-orthogonal multiple access (NOMA) relay network using matching game,” in *Proceeding of IEEE International Conference Communication (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6
36. Z. Ding, P. Fan, H.V. Poor, Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. *IEEE Trans. Veh. Technol.* **65**(8), 6010–6023 (2016)