Title

Low-Complexity Dynamic Resource Scheduling for Downlink MC-NOMA Over Fading Channels

Authors

Kim, Do-Yup
Jafarkhani, Hamid
Lee, Jang-Won

Publication Date

2021-03-18

Peer reviewed

Low-Complexity Dynamic Resource Scheduling for Downlink MC-NOMA Over Fading Channels

Do-Yup Kim, *Student Member, IEEE,* Hamid Jafarkhani, *Fellow, IEEE,* and Jang-Won Lee, *Senior Member, IEEE*

Abstract

In this paper, we investigate dynamic resource scheduling (i.e., joint user, subchannel, and power scheduling) for downlink multi-channel non-orthogonal multiple access (MC-NOMA) systems over time-varying fading channels. Specifically, we address the weighted average sum rate maximization problem with quality-of-service (QoS) constraints. In particular, to facilitate fast resource scheduling, we focus on developing a very low-complexity algorithm. To this end, by leveraging Lagrangian duality and the stochastic optimization theory, we first develop an opportunistic MC-NOMA scheduling algorithm whereby the original problem is decomposed into a series of subproblems, one for each time slot. Accordingly, resource scheduling works in an online manner by solving one subproblem per time slot, making it more applicable to practical systems. Then, we further develop a heuristic joint subchannel assignment and power allocation (Joint-SAPA) algorithm with very low computational complexity, called Joint-SAPA-LCC, that solves each subproblem. Finally, through simulation, we show that our Joint-SAPA-LCC algorithm provides good performance comparable to the existing Joint-SAPA algorithms despite requiring much lower computational complexity. We also demonstrate that our opportunistic MC-NOMA scheduling algorithm in which the Joint-SAPA-LCC algorithm is embedded works well while satisfying given QoS requirements.

A preliminary version of this work [1] will be presented at IEEE VTC2021-Spring, in which only single-channel transmission has been taken into account, and neither subchannel scheduling nor inter-channel power scheduling has been dealt with.

D.-Y. Kim and J.-W. Lee are with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea (e-mail: danny.doyup.kim@yonsei.ac.kr; jangwon@yonsei.ac.kr).

H. Jafarkhani is with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, CA 92697, USA (e-mail: hamidj@uci.edu).

Index Terms

Low complexity, multi-channel non-orthogonal multiple access (MC-NOMA), quality of service (QoS), resource scheduling, time-varying fading channels, weighted sum rate.

I. Introduction

With the exponential proliferation of mobile devices, overall mobile data traffic is expected to grow to 77 exabytes per month by 2022 [2]. Such surge in mobile data traffic will intensify resource shortages, which in turn will necessitate high levels of connectivity and spectral efficiency. In these circumstances, non-orthogonal multiple access (NOMA) has been envisioned as a promising technology for future cellular networks thanks to its potential to achieve high connectivity and spectral efficiency compared to orthogonal multiple access (OMA)-based resource sharing technologies [3]–[6]. Unlike OMA techniques without inter-user interference (IUI), NOMA allows IUI by multiplexing multiple users on the same resource based on superposition coding in the power domain. Then, at the receiver, multi-user detection is realized by mitigating IUI based on successive interference cancellation (SIC) (see [7], [8] and references therein). However, despite of high network performance, the presence of IUI makes resource scheduling, which is very important in wireless networks, more difficult in NOMA systems. In addition, resource scheduling techniques developed for the OMA systems, e.g., [9]–[11], cannot be easily applied to the NOMA systems, and provide limited performance even if they can. In this context, resource scheduling with low computational complexity is one of the paramount issues in the NOMA systems, and thus, many studies have been conducted in this regard. Nevertheless, they still have practical limitations, especially in terms of computational complexity.

A. Related Work

Early studies in this area have focused on single-channel NOMA (SC-NOMA). Accordingly, various studies have been conducted in SC-NOMA systems, in terms of power allocation [12]–[16], and power allocation and user selection [17], [18]. More recently, the research focus in this area has been shifted from SC-NOMA to multi-channel NOMA (MC-NOMA). MC-NOMA systems take multi-channel transmission into account; however, compared to SC-NOMA systems, resource allocation becomes much more complicated because of the additional burden of subchannel assignment. As a result, the algorithms for SC-NOMA are usually inapplicable to MC-NOMA, and even if they are applicable, they provide limited performance. Thus, to take full advantage of MC-NOMA, a new joint subchannel assignment and power allocation (Joint-SAPA) algorithm tailored to MC-NOMA systems is needed.

Joint-SAPA for minimizing total power consumption in MC-NOMA systems has been investigated thanks to the corresponding simple linear objective function [19]–[21]. Later, it has shown that the sum rate becomes a concave function of power allocation even though each user's data rate is a nonconcave function [22]. Therefore, a basic power allocation problem to maximize the sum rate can be easily solved with well-known convex optimization solvers. However, the convexity in optimization gets lost in more general Joint-SAPA problems that take into account subchannel assignment and practical constraints, e.g., a so-called *SIC capacity constraint* that limits the number of users who can be served simultaneously through the same resource. Hence, many heuristic Joint-SAPA algorithms have been proposed, e.g., [22]–[24], but almost all of them are still based on the concavity of the sum rate function with respect to power allocation.

Despite many studies on the sum rate maximization, the relative importance and/or fairness among users has not been addressed due to the nature of the sum rate performance metric. Thereby, users with poor channel conditions may experience starvation because no resource might be allocated to them. On the other hand, different tradeoffs can be achieved between the sum rate performance and the user fairness by controlling user weights in the weighted sum rate maximization problem. However, unlike the (equally weighted) sum rate, the weighted sum rate is generally a nonconcave function of power allocation (even in SC-NOMA [16]), and accordingly the Joint-SAPA problem to maximize the weighted sum rate is known to be a very challenging NP hard problem. Hence, in most cases, the ideas and underlying theory exploited in the sum rate maximization cannot be fully leveraged in the weighted sum rate maximization. To address these, the Joint-SAPA problem to maximize the weighted sum rate has received much attention [25]–[32]. In [25], the power allocation for each subchannel in a two-user MC-NOMA system has been investigated. In [26], the Joint-SAPA problem in a multi-user MC-NOMA system has been investigated without taking into account the essential SIC capacity constraint. In [27], the authors have proposed a heuristic Joint-SAPA algorithm, considering the SIC capacity constraint, based on the fractional transmit power control (FTPC) and exhaustive search (ES) algorithms. In [28], [29], heuristic Joint-SAPA algorithms using the difference-of-convex programming (DCP) approach have been developed under the assumption that each subchannel is occupied by up to two users. In [30], the power allocation and the subchannel assignment are performed based on the geometric programming (GP) approach and

TABLE I

COMPARISON OF STUDIES ON THE WEIGHTED SUM RATE MAXIMIZATION IN THE MC-NOMA SYSTEM (✓: CONSIDERED)

Ref.	Constraint				Optimization		
	Total power limit	Subchannel power limit	SIC capacity	QoS requirement	Subchannel assignment	Power allocation	Scheduling over fading
[25]	✓					✓	
[26]	✓				✓	✓	✓
[27]–[31]	✓		✓		✓	✓	✓
[32]	✓	✓	✓		✓	✓	
Our work	✓	✓	✓	✓	✓	✓	✓

the many-to-many matching game, respectively. In [31], the authors have proposed a Joint-SAPA algorithm utilizing the Lagrangian dual and dynamic programming (DP) approaches. Most recently, in [32], the authors have studied a Joint-SAPA problem with further considering the individual subchannel power limits. They have developed a Joint-SAPA algorithm based on the DP approach and the projected gradient descent (PGD) method. The weighted sum rate maximization studies related to resource allocation for the MC-NOMA system are summarized in Table I.

B. Motivation and Contributions

Although Joint-SAPA algorithms to maximize the weighted sum rate have been extensively studied, thus far, all of them are still based on approaches with high computational complexity as discussed before. This is burdensome for practical use in future cellular networks with very short time slots.[1] Hence, Joint-SAPA algorithms with very low computational complexity are needed to make it possible to generate transmit signals at the base station (BS) in a very short time slot.

In addition, not all weighted sum rate maximization studies have considered explicit QoS requirements, as shown in Table I. Instead, the authors in [26]–[31] have realized proportional

[1]In recent standardization trends, the length of the slot, which is a unit to transmit 14 orthogonal frequency division multiplexing (OFDM) symbols, is reduced to achieve higher spectral efficiency and traffic capacity and lower user plane latency. For example, 5G New Radio (NR) supports flexible OFDM numerology with subcarrier spacing from 15 kHz to 240 kHz, resulting in a slot length as short as 62.5 μs [33]. Even more, 5G NR introduces a unit of *mini-slot*, which is even shorter than a slot, for the sake of fast data transmission for ultra-reliable low-latency communication (URLLC).

fair scheduling based on their own Joint-SAPA algorithms in simulation, using the fact that the proportional fair scheduling is a specific use case of the weighted sum rate maximization problem. Although the proportional fair scheduling provides high sum rate performance while closing the performance gap between users to some extent, it cannot explicitly guarantee given QoS requirements. Accordingly, in a practical QoS-aware system with individual user QoS requirements, a new scheduling technique that can meet the individual QoS requirements as well is needed. In particular, in wireless network systems that are subject to time-varying fading channels, the development of a scheduling technique that meets QoS requirements by exploiting the variability of the channels is necessary. Hence, in this paper, we aim to develop a novel low-complexity opportunistic resource scheduling algorithm for the downlink MC-NOMA system, which fully exploits the stochasticity of fading channels to maximize the weighted average sum rate while ensuring the individual QoS requirements of users.

The main contributions of this paper are summarized as follows:

- We address a dynamic resource scheduling problem for the downlink MC-NOMA system over time-varying fading channels. To the best of our knowledge, this is the first work to maximize the weighted average sum rate while ensuring explicitly given QoS requirements via joint optimization of user, subchannel, and power scheduling.
- We develop a Joint-SAPA algorithm with very low computational complexity, called Joint-SAPA-LCC, to maximize the instantaneous weighted sum rate. It has much lower computational complexity compared to the existing Joint-SAPA algorithms with the same objective.
  - We prove that it is optimal to select up to two users per subchannel, assuming that the noise power of users suffering from interference is neglected, and propose a very simple optimal user selection rule based on it.
  - In accordance with the proposed user selection rule, we derive closed-form optimal user power allocation formulas and a simple subchannel power allocation algorithm.
  - Through simulation, we verify that our Joint-SAPA-LCC algorithm provides good performance comparable to the existing Joint-SAPA algorithms despite requiring much lower computational complexity.
- By leveraging the Lagrangian duality and the stochastic optimization theory, we develop an opportunistic MC-NOMA scheduling algorithm that fully exploits time-varying fading channels. It operates in an online manner based on the Joint-SAPA-LCC algorithm, and thus,

it is very effective for practical use. Through simulation, we show that our opportunistic MC-NOMA scheduling with the Joint-SAPA-LCC algorithm works well and properly meets various QoS requirements.

C. Paper Structure and Notations

Paper Structure: The rest of the paper is organized as follows. In Section II, we formulate the system model and the dynamic resource scheduling problem. In Section III and Section IV, we develop the Joint-SAPA-LCC algorithm and the opportunistic MC-NOMA scheduling algorithm, respectively. We present simulation results in Section V and conclude in Section VI.

Notation: Scalars, vectors, and sets are denoted by italic, boldface, and calligraphic letters, respectively. A vector that consists of elements in the set $\{x_i : i \in \mathcal{X}\}$ is denoted by $(x_i)_{\forall i \in \mathcal{X}}$. The expectation operator is denoted by $\mathbb{E}[\cdot]$. For a complex number x, $|x|$ denotes its absolute value. For a real number x, y, and z, $[x]^+ = \max(0, x)$, and $[x]_y^z = \min(\max(x, y), z)$. For a real-valued vector $\mathbf{x}$, $[\mathbf{x}]^+$ is a vector whose ith element is $[x_i]^+$. We denote by $\mathbf{1}_{\{A\}}$ an indicator function taking the value of one if the statement A is true, and zero otherwise. The symbol $\Leftrightarrow$ denotes the logical connective *if and only if*. The floor function is denoted by $\lfloor \cdot \rfloor$, which gives the largest integer not exceeding its argument.

II. System Model and Problem Formulation

We consider the downlink of a single cell in the MC-NOMA system, in which one single-antenna BS transmits signals to N single-antenna users over K subchannels. The index sets of users and subchannels are denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$ and $\mathcal{K} = \{1, 2, \ldots, K\}$, respectively. We assume that the entire system bandwidth, B_{tot}, is divided into K orthogonal subchannels, so that there is no interference among them. The bandwidth of Subchannel k is denoted by B_k.

We consider a time-slotted system over doubly block fading channels, where the channel gain of each wireless link is time-varying and frequency-selective but remains constant during a time slot and flat within a subchannel. Let $\{h_{k,i}^t, t = 1, 2, \ldots\}$ be the fading process associated with User i on Subchannel k, where $h_{k,i}^t$ is a complex-valued continuous random variable representing the channel gain from the BS to User i on Subchannel k in time slot t. The fading process is assumed to be stationary and ergodic. Note that the channel gain is a comprehensive one including shadowing, multipath fading, and path loss. We assume that information on the underlying distributions of the fading process is unknown to the BS due to the practical difficulties in obtaining such information a priori. However, we assume that instantaneous channel gains are known to the BS at the beginning of each time slot.

In MC-NOMA, a subchannel can be assigned to multiple users simultaneously by power-domain multiplexing. Let $x_{k,i}^t$, satisfying $\mathbb{E}[|x_{k,i}^t|^2] = 1$, be the information-bearing signal transmitted to User i on Subchannel k in time slot t, and $p_{k,i}^t$ be the power allocated to signal $x_{k,i}^t$. Also, let $q_{k,i}^t$ be the subchannel assignment indicator taking the value of one if Subchannel k is assigned to User i in time slot t, and zero otherwise. Then, the received signal at User i on Subchannel k in time slot t is given by

$$y_{k,i}^t = h_{k,i}^t q_{k,i}^t \sqrt{p_{k,i}^t}\, x_{k,i}^t + \sum_{j \in \mathcal{N}: j \neq i} h_{k,i}^t q_{k,j}^t \sqrt{p_{k,j}^t}\, x_{k,j}^t + n_{k,i}^t, \tag{1}$$

where $n_{k,i}^t$ is the additive zero-mean Gaussian noise with variance $\sigma_{k,i}^2$, i.e., $n_{k,i}^t \sim \mathcal{CN}(0, \sigma_{k,i}^2)$, and the first, second, and third terms represent the desired signal, interference, and noise, respectively. For convenience, we define the noise-to-channel ratio (NCR) of User i on Subchannel k in time slot t as

$$\eta_{k,i}^t = \frac{\sigma_{k,i}^2}{|h_{k,i}^t|^2}. \tag{2}$$

The NCR can be interpreted as the effective noise power when the channel gain is normalized to unity.

After receiving signal $y_{k,i}^t$, User i performs SIC to decode its own signal, $x_{k,i}^t$, from it. User i first decodes the signals for each User j whose NCR is not smaller than its NCR, i.e., $\eta_{k,j}^t \geq \eta_{k,i}^t$, and then subtracts the components associated with them from the received signal. Then, User i decodes its own signal by treating the signals for the other users whose NCRs are smaller than its NCR as noise. With a typical assumption that SIC has been successfully done, the maximum achievable data rate of User i on Subchannel k in time slot t is obtained as [34]

$$R_{k,i}(\mathbf{p}_k^t, \mathbf{q}_k^t; \mathbf{h}_k^t) = B_k \log_2\left(1 + \frac{q_{k,i}^t p_{k,i}^t}{\sum_{j \in \mathcal{N}: \eta_{k,j}^t < \eta_{k,i}^t} q_{k,j}^t p_{k,j}^t + \eta_{k,i}^t}\right), \tag{3}$$

where $\mathbf{p}_k^t = (p_{k,i}^t)_{\forall i \in \mathcal{N}}$, $\mathbf{q}_k^t = (q_{k,i}^t)_{\forall i \in \mathcal{N}}$, and $\mathbf{h}_k^t = (h_{k,i}^t)_{\forall i \in \mathcal{N}}$. From (3), the maximum achievable data rate of User i over all subchannels in time slot t is obtained as

$$R_i(\mathbf{p}^t, \mathbf{q}^t; \mathbf{h}^t) = \sum_{k \in \mathcal{K}} R_{k,i}(\mathbf{p}_k^t, \mathbf{q}_k^t; \mathbf{h}_k^t), \tag{4}$$

where $\mathbf{p}^t = (\mathbf{p}_k^t)_{\forall k \in \mathcal{K}}$, $\mathbf{q}^t = (\mathbf{q}_k^t)_{\forall k \in \mathcal{K}}$, and $\mathbf{h}^t = (\mathbf{h}_k^t)_{\forall k \in \mathcal{K}}$. For simple notation, we interchangeably use R_i^t and $R_i(\mathbf{p}^t, \mathbf{q}^t; \mathbf{h}^t)$ if there is no confusion. Using (4), we define the average data rate, $\bar{R}_i$, of User i as

$$\bar{R}_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R_i^t, \tag{5}$$

and the weighted average sum rate of the system, $\bar{R}_{\text{WSR}}$, which is what we are trying to maximize, as

$$\bar{R}_{\text{WSR}} = \sum_{i \in \mathcal{N}} w_i \bar{R}_i, \tag{6}$$

where w_i is the weight factor representing the relative importance of User i. Additionally, each User i has its own minimum average data rate requirement, $\bar{R}_{\min,i}$, which is represented as

$$\bar{R}_i \geq \bar{R}_{\min,i}, \ \forall i \in \mathcal{N}. \tag{7}$$

Because of SIC, in each time slot t, the BS can schedule multiple users on the same subchannel simultaneously. However, due to the high computational complexity and the potential error propagation of SIC as well as the limited processing capabilities of users, the number of users multiplexed simultaneously on the same subchannel is typically limited to a small number, M. In this regard, we define a feasible set for a subchannel assignment indicator vector, $\mathbf{q}^t$, in time slot t as

$$\mathcal{Q} = \left\{ \mathbf{q}^t \in \{0,1\}^{K \times N} \;\middle|\; \sum_{i \in \mathcal{N}} q_{k,i}^t \leq M, \ \forall k \in \mathcal{K} \right\}. \tag{8}$$

In addition, the BS should determine how much power to allocate to the scheduled users under given transmission power constraints. We assume that the BS has a limited total transmission power budget of $P_{\max}$ and an individual subchannel peak power constraint of $P_{\max,k}$ for each Subchannel k. In this regard, we define a feasible set for a power allocation vector, $\mathbf{p}^t$, in time slot t as

$$\mathcal{P} = \left\{ \mathbf{p}^t \in \mathbb{R}^{K \times N} \;\middle|\; \begin{array}{l} \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{K}} p_{k,i}^t \leq P_{\max}, \\ \sum_{i \in \mathcal{N}} p_{k,i}^t \leq P_{\max,k}, \ \forall k \in \mathcal{K}, \\ p_{k,i}^t \geq 0, \ \forall k \in \mathcal{K}, \ \forall i \in \mathcal{N} \end{array} \right\}. \tag{9}$$

With the performance metric function in (6), the QoS constraints in (7), and the feasible sets for decision variables in (8) and (9), we finally formulate the dynamic resource scheduling

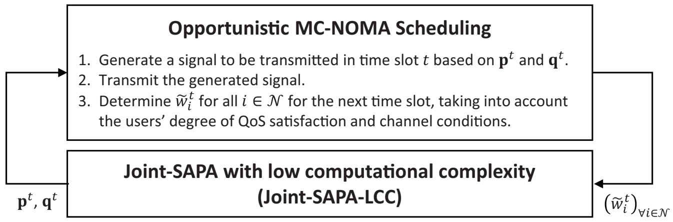


Fig. 1. The flow chart of the opportunistic MC-NOMA scheduling algorithm operating in an online manner to solve Problem (P1).

problem for joint user, subchannel, and power scheduling in the downlink MC-NOMA system over time-varying fading channels as

$$
\begin{aligned}
\text{(P1)} \quad & \underset{\mathbf{p}^t, \mathbf{q}^t, \forall t}{\text{maximize}} && \bar{R}_{\text{WSR}} \\
& \text{subject to} && \bar{R}_i \geq \bar{R}_{\min,i}, \;\; \forall i \in \mathcal{N}, \\
& && \mathbf{p}^t \in \mathcal{P}, \;\; \mathbf{q}^t \in \mathcal{Q}, \;\; \forall t.
\end{aligned}
$$

We first note that dealing with Problem (P1) is not easy due to the average operation over an infinite time horizon. To resolve this difficulty, we develop an opportunistic MC-NOMA scheduling algorithm by leveraging the Lagrangian duality and the stochastic optimization theory, whereby Problem (P1) is decomposed into a series of deterministic optimization subproblems, one for each time slot. Specifically, the subproblem for one time slot is identified as a Joint-SAPA problem to maximize the instantaneous weighted sum rate without QoS constraints in that time slot. As a consequence, we do not have to solve Problem (P1) directly at once, but rather solve the Joint-SAPA problem in every time slot in an online manner. Meanwhile, an important caveat is that the Joint-SAPA problem needs to be solved by a simple algorithm with very low computational complexity so that the BS can generate and transmit signals in every short time slot. Hence, we develop a heuristic algorithm to solve the Joint-SAPA problem with low computational complexity, called Joint-SAPA-LCC algorithm. The flow chart of the process for solving the dynamic resource scheduling problem, Problem (P1), is schematically illustrated in Fig. 1. Note that the Joint-SAPA-LCC algorithm is a built-in algorithm that runs every time slot within the opportunistic MC-NOMA scheduling algorithm. In the following, for ease of explanation, we first develop the Joint-SAPA-LCC algorithm in Section III, and then the opportunistic MC-NOMA scheduling algorithm in Section IV.

III. Joint-SAPA with low computational complexity (Joint-SAPA-LCC)

In this section, we develop our Joint-SAPA-LCC algorithm that solves the instantaneous weighted sum rate maximization problem for each time slot t, defined by

$$(\mathrm{P}_1^t) \quad \underset{\mathbf{p}^t \in \mathcal{P},\, \mathbf{q}^t \in \mathcal{Q}}{\text{maximize}} \quad \sum_{i \in \mathcal{N}} \tilde{w}_i^t R_i^t,$$

where $\tilde{w}_i^t$ is the effective weight of User i in time slot t.[2] Problem (P_1^t) is an NP-hard problem that is very difficult to solve using conventional methods since it contains not only a nonconcave objective function but also integer variables. In addition, we need a fast solution since the transmission signal should be generated and transmitted according to the solution in every short time slot. For these reasons, we develop a heuristic suboptimal algorithm that provides a near-optimal performance despite very low computational complexity. In the remainder of this section, since Problem (P_1^t) is focusing only on time slot t, we omit the superscript t for notational brevity.

To solve Problem (P_1^t), we exploit the primal decomposition method [35], [36]. Accordingly, by introducing a new coupling vector $\bar{\mathbf{P}} = (\bar{P}_k)_{\forall k \in \mathcal{K}}$, we reformulate Problem (P_1^t) equivalently as

$$\begin{aligned}
(\mathrm{P}_2) \quad \underset{\mathbf{p},\, \mathbf{q},\, \bar{\mathbf{P}}}{\text{maximize}} \quad & \sum_{i \in \mathcal{N}} \tilde{w}_i \sum_{k \in \mathcal{K}} R_{k,i}(\mathbf{p}_k, \mathbf{q}_k; \mathbf{h}_k) \\
\text{subject to} \quad & \sum_{k \in \mathcal{K}} \bar{P}_k \le P_{\max}, \\
& 0 \le \bar{P}_k \le P_{\max,k}, \ \forall k \in \mathcal{K}, \\
& \sum_{i \in \mathcal{N}} p_{k,i} \le \bar{P}_k, \ \forall k \in \mathcal{K}, \\
& p_{k,i} \ge 0, \ \forall k \in \mathcal{K}, \ \forall i \in \mathcal{N}, \\
& \sum_{i \in \mathcal{N}} q_{k,i} \le M, \ \forall k \in \mathcal{K}, \\
& q_{k,i} \in \{0, 1\}, \ \forall k \in \mathcal{K}, \ \forall i \in \mathcal{N}.
\end{aligned}$$

It is worth noting that Problem (P_2) could be decoupled for each Subchannel k if the coupling vector $\bar{\mathbf{P}}$ were fixed. Hence, we separate it into two levels of optimization. At the lower level,

[2]Note that $\tilde{w}_i^t$ differs from w_i in (6) in that it is systematically adjusted in every time slot based on the degree of QoS satisfaction of User i and its channel condition. It will be rigorously discussed later in Section IV.

we have K subproblems, one for each Subchannel k, defined by

$$
\begin{aligned}
(\mathrm{S}_k) \quad \underset{\mathbf{p}_k,\, \mathbf{q}_k}{\text{maximize}} \quad & \sum_{i\in\mathcal{N}} \tilde{w}_i R_{k,i}(\mathbf{p}_k, \mathbf{q}_k; \mathbf{h}_k) \\
\text{subject to} \quad & \sum_{i\in\mathcal{N}} p_{k,i} \le \bar{P}_k, \\
& p_{k,i} \ge 0, \ \forall i \in \mathcal{N}, \\
& \sum_{i\in\mathcal{N}} q_{k,i} \le M, \\
& q_{k,i} \in \{0, 1\}, \ \forall i \in \mathcal{N}.
\end{aligned}
$$

At the higher level, we have an optimization in charge of updating the coupling vector $\bar{\mathbf{P}}$, defined by

$$
\begin{aligned}
(\mathrm{M}) \quad \underset{\bar{\mathbf{P}}}{\text{maximize}} \quad & \phi^*(\bar{\mathbf{P}}) = \sum_{k\in\mathcal{K}} \phi_k^*(\bar{P}_k) \\
\text{subject to} \quad & \sum_{k\in\mathcal{K}} \bar{P}_k \le P_{\max}, \\
& 0 \le \bar{P}_k \le P_{\max,k}, \ \forall k \in \mathcal{K},
\end{aligned}
$$

where $\phi_k^*(\bar{P}_k)$ is the optimal value of Problem (S_k). Then, we can obtain a suboptimal solution, $\{\mathbf{p}^*, \mathbf{q}^*, \bar{\mathbf{P}}^*\}$, to Problem (P_2) by alternately solving Problems (S_k), $\forall k \in \mathcal{K}$, and Problem (M) until convergence. The pseudocode for this process is summarized in Algorithm 1. We can show that Algorithm 1 converges to a stationary point.

Theorem 1. *Algorithm 1 converges to a stationary point.*

Proof. See Appendix A. □

In the primal decomposition method, if the primal problem is a convex problem, not only the subproblems but also the master problem becomes a convex problem, resulting in the convergence to a global optimal solution. In our case, however, since Problem (P_1^t) is not a convex problem, the convergence to a global optimal solution is not guaranteed. Nevertheless, we show by simulation results that the algorithm provides a near-optimal performance with very low computational complexity. Also, it is worth noting that we will derive $\phi_k^*(\bar{P}_k)$ in a closed form in Section III-A, and effectively use it to solve Problem (M) in Section III-B.

Algorithm 1: Joint-SAPA-LCC

```
Initialize: $\bar{P}_k \leftarrow P_{\max}/K, \forall k \in \mathcal{K}$.
repeat
    Obtain $\phi_k^*(\bar{P}_k)$ by solving Problem (S_k) for the given $\bar{P}_k$, $\forall k \in \mathcal{K}$, using Algorithm 2.
    Obtain the solution, $\bar{\mathbf{P}}^*$, to Problem (M) for the given $\phi^*(\bar{\mathbf{P}})$ using Algorithm 3.
    Update $\bar{\mathbf{P}} \leftarrow \bar{\mathbf{P}}^*$
until convergence
Obtain $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$ by solving Problem (S_k) for the given $\bar{P}_k$, $\forall k \in \mathcal{K}$, using Algorithm 2.
return $\{\mathbf{p}^*, \mathbf{q}^*\}$.
```

A. Solution to Problem (S_k)

In this subsection, we discuss how to solve Problem (S_k) for a given $\bar{P}_k$. To be specific, to maximize the weighted sum rate over Subchannel k, we find out which users to be assigned to Subchannel k and how much power to be allocated to them under the limited power of $\bar{P}_k$. In addition, we derive the corresponding objective value, $\phi_k^*(\bar{P}_k)$, as a closed-form function of $\bar{P}_k$.

Problem (S_k) is difficult to solve in general mainly due to the integer variables, i.e., the subchannel assignment indicators, $\mathbf{q}_k$. To address this difficulty, we first consider a problem defined by

$$(\mathrm{Q}_k) \quad \underset{\mathbf{p}_k}{\text{maximize}} \quad \sum_{i\in\mathcal{N}} \tilde{w}_i R_{k,i}(\mathbf{p}_k;\mathbf{h}_k)$$
$$\text{subject to} \quad \sum_{i\in\mathcal{N}} p_{k,i} \le \bar{P}_k,$$
$$p_{k,i} \ge 0, \ \forall i \in \mathcal{N},$$

where

$$R_{k,i}(\mathbf{p}_k;\mathbf{h}_k) = B_k \log_2\left(1 + \frac{p_{k,i}}{\sum_{j\in\mathcal{N}:\eta_{k,j}<\eta_{k,i}} p_{k,j} + \eta_{k,i}}\right). \tag{10}$$

Although Problem (Q_k) is different from Problem (S_k), the optimal solution to Problem (S_k) can be easily derived from that to Problem (Q_k) if a certain condition is met, as stated in the following theorem.

Theorem 2. *Let* $\mathbf{p}_k^\dagger = (p_{k,i}^\dagger)_{\forall i\in\mathcal{N}}$ *be an optimal solution to Problem* (Q_k)*, and suppose that it satisfies*

$$\sum_{i\in\mathcal{N}} \mathbf{1}_{\left\{p_{k,i}^\dagger>0\right\}} \le M. \tag{11}$$

Then, an optimal solution, $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$*, to Problem* (S_k) *can be obtained as*

$$\mathbf{p}_k^* = \mathbf{p}_k^\dagger \text{ and } \mathbf{q}_k^* = (q_{k,i}^*)_{\forall i \in \mathcal{N}}, \tag{12}$$

where $q_{k,i}^* = \mathbf{1}_{\left\{p_{k,i}^\dagger > 0\right\}}$ *for all* $i \in \mathcal{N}$*.*

Proof. See Appendix B. □

Note that the solution to Problem (Q_k) that will be obtained by our proposed algorithm always satisfies (11) for any $M > 1$. Therefore, a solution to Problem (S_k) can be easily obtained according to Theorem 2.

Now, we discuss how to solve Problem (Q_k). Even though Problem (Q_k) does not have any integer variables, it is still NP-hard due to its nonconcave objective function. Hence, we focus on developing a heuristic algorithm that provides a near-optimal solution to Problem (Q_k) with very low computational complexity. To this end, we first find *candidate* users who might be allocated positive power on Subchannel k. Then, we derive the optimal power allocation for them in closed forms. For compact notation, we assume, without loss of generality, that users are ordered such that $\eta_{k,i} > \eta_{k,j}$ if $i < j$, and define a *last SIC user* as follows.

Definition 1. *A last SIC user* refers to a user who does not experience any interference signals after the SIC process.

We start with the assumption that User φ_k is the last SIC user on Subchannel k and has been allocated a certain amount of power. Accordingly, we assume that p_{k,φ_k} is given as a fixed positive value, and $p_{k,i}$ for $i > \varphi_k$ is given as zero so that User φ_k does not experience any interference signals after the SIC process. Under this assumption, $p_{k,i}$'s for $i \geq \varphi_k$ are no longer decision variables. Note that how to select the last SIC user and how much power to allocate to it will be discussed later. Now, Problem (Q_k) can be reformulated as

$$\begin{aligned}
(\mathrm{Q}_{k,1}^{\varphi_k}) \quad \underset{p_{k,i},\, i<\varphi_k}{\text{maximize}} \quad & \sum_{i=1}^{\varphi_k - 1} \tilde{w}_i B_k \log_2\left(1 + \frac{p_{k,i}}{\sum_{j>i} p_{k,j} + \eta_{k,i}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{p_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right) \\
\text{subject to} \quad & \sum_{i=1}^{\varphi_k - 1} p_{k,i} + p_{k,\varphi_k} \leq \bar{P}_k, \\
& p_{k,i} \geq 0, \;\; i < \varphi_k.
\end{aligned}$$

Note that the purpose of this problem is not to find power allocation to users but to find *candidate* users when User φ_k is selected as the last SIC user on Subchannel k. Based on the fact that interference power is usually much greater than noise power, the noise power of users suffering from the interference signals can be assumed to be negligible. Hence, by assuming $\sigma_{k,i}^2 = 0$ for $i < \varphi_k$ and letting $\rho_{k,i} = \sum_{j=i}^{\varphi_k} p_{k,j}$ for $i \leq \varphi_k$, we can approximate Problem $(\mathrm{Q}_{k,1}^{\varphi_k})$ as

$$(\mathrm{Q}_{k,2}^{\varphi_k}) \quad \underset{\rho_{k,i},\, i<\varphi_k}{\text{maximize}} \quad \sum_{i=1}^{\varphi_k-1} \tilde{w}_i B_k \log_2\left(\frac{\rho_{k,i}}{\rho_{k,i+1}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{\rho_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right) \tag{13}$$

$$\text{subject to} \quad \prod_{i=1}^{\varphi_k-1} \frac{\rho_{k,i}}{\rho_{k,i+1}} \leq \frac{\bar{P}_k}{\rho_{k,\varphi_k}}, \tag{14}$$

$$\frac{\rho_{k,i}}{\rho_{k,i+1}} \geq 1, \;\; i < \varphi_k, \tag{15}$$

where the inequality constraints are equivalent to those in Problem $(\mathrm{Q}_{k,1}^{\varphi_k})$, which can be derived by simple arithmetic operations. In succession, by letting $r_{k,i} = \log_2(\rho_{k,i}/\rho_{k,i+1})$ for $i < \varphi_k$ and $r_{k,\varphi_k} = \log_2(\rho_{k,\varphi_k})$, and taking the logarithm of the both sides of the constraints, we can reformulate Problem $(\mathrm{Q}_{k,2}^{\varphi_k})$ equivalently as

$$\begin{aligned} (\mathrm{Q}_{k,3}^{\varphi_k}) \quad & \underset{r_{k,i},\, i<\varphi_k}{\text{maximize}} \quad \sum_{i=1}^{\varphi_k-1} \tilde{w}_i B_k r_{k,i} + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{2^{r_{k,\varphi_k}}}{\eta_{k,\varphi_k}}\right) \\ & \text{subject to} \quad \sum_{i=1}^{\varphi_k-1} r_{k,i} \leq \log_2(\bar{P}_k) - r_{k,\varphi_k}, \\ & \qquad\qquad\quad r_{k,i} \geq 0, \;\; i < \varphi_k. \end{aligned}$$

In Problem $(\mathrm{Q}_{k,3}^{\varphi_k})$, since r_{k,φ_k} is not a decision variable, the second term, $\tilde{w}_{\varphi_k} B_k \log_2(1 + 2^{r_{k,\varphi_k}}/\eta_{k,\varphi_k})$, of the objective function and the right-hand side, $\log_2(\bar{P}_k) - r_{k,\varphi_k}$, of the first constraint are constants. Also, the decision variables are linearly combined in the objective function, and the feasible set is a unit simplex. Hence, it is obvious that the objective function is maximized when all the decision variables, except the one with the largest weight, are zero. Also, by the definition of $r_{k,i}$, we can easily see that, for any $i < \varphi_k$, $p_{k,i}$ is zero if and only if $r_{k,i}$ is zero. Thus, only one user with the largest weight is selected as the other *candidate* user on Subchannel k together with the last SIC user, i.e., User φ_k. We state this result in the following theorem.

Theorem 3. *Under the assumption that the noise power of users suffering from the interference signals are neglected, on each Subchannel k, at most two users are allocated power according to the optimal solution to Problem $(\mathrm{Q}_{k,1}^{\varphi_k})$. To be specific, when User φ_k has been selected as* *the last SIC user on Subchannel k, User $\psi_k(\varphi_k)$ is accordingly selected as the other candidate user, where*

$$\psi_k(\varphi_k) = \underset{i<\varphi_k}{\operatorname{argmax}}\{\tilde{w}_i\}. \tag{16}$$

By Theorem 3, we can reduce Problem $(\mathrm{Q}_{k,1}^{\varphi_k})$ to the power allocation problem for the two-user case as

$$\begin{aligned}
(\mathrm{Q}_{k,4}^{\varphi_k}) \quad \underset{p_{k,\psi_k},\, p_{k,\varphi_k}}{\text{maximize}} \quad & \tilde{w}_{\psi_k} B_k \log_2\!\left(1 + \frac{p_{k,\psi_k}}{p_{k,\varphi_k} + \eta_{k,\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\!\left(1 + \frac{p_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right) \\
\text{subject to} \quad & p_{k,\psi_k} + p_{k,\varphi_k} \leq \bar{P}_k, \\
& p_{k,\psi_k} \geq 0,\ \ p_{k,\varphi_k} \geq 0,
\end{aligned}$$

where $\psi_k(\varphi_k)$ is replaced with ψ_k for notational simplicity. This two-user power allocation problem can be optimally solved in closed forms.

Theorem 4. *Let $\{p^\star_{k,\psi_k}, p^\star_{k,\varphi_k}\}$ be the optimal solution to Problem $(\mathrm{Q}_{k,4}^{\varphi_k})$. Then, it can be obtained as*

$$p^\star_{k,\varphi_k} = \begin{cases} 0, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \leq C_k^1, \\ \bar{P}_k, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > C_k^2(\bar{P}_k), \\ \dfrac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}, & \text{otherwise}, \end{cases} \tag{17}$$

$$p^\star_{k,\psi_k} = \bar{P}_k - p^\star_{k,\varphi_k}, \tag{18}$$

where

$$C_k^1 = \frac{\eta_{k,\varphi_k}}{\eta_{k,\psi_k}} \quad \text{and} \quad C_k^2(\bar{P}_k) = \frac{\bar{P}_k + \eta_{k,\varphi_k}}{\bar{P}_k + \eta_{k,\psi_k}}. \tag{19}$$

Proof. See Appendix C. □

All derivations so far have been made on the assumption that User φ_k is allocated a certain amount of power as the last SIC user. However, in the case where $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k(\varphi_k)} \leq C_k^1$, the optimal power allocation to the last SIC user becomes zero. Thus, we consider this case to be contradictory, and set the objective value (i.e., the weighted sum rate on Subchannel k) to a

Algorithm 2: Solution to Problem (S_k)

for *each User* $\varphi_k \in \mathcal{N}$ **do**
    Suppose that User φ_k is the last SIC user.
    Select the other *candidate* user using (16).
    Obtain their power allocation using Theorem 4.
    Obtain $\phi_k(\bar{P}_k;\varphi_k)$ using (20).
Obtain φ_k^* and $\phi_k^*(\bar{P}_k)$ using (21) and (22), respectively.
Let Users φ_k^* and $\psi_k(\varphi_k^*)$ be the optimal *candidate* users.
Obtain $\mathbf{p}_k^\dagger$ using Theorem 4.
Derive $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$ from $\mathbf{p}_k^\dagger$ using Theorem 2.
return $\{\mathbf{p}_k^*, \mathbf{q}_k^*, \phi_k^*(\bar{P}_k)\}$.

negative infinity in this case. Then, the weighted sum rate on Subchannel k under the assumption that User φ_k is the last SIC user on this subchannel can be obtained as

$$\phi_k(\bar{P}_k;\varphi_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{p^\star_{k,\psi_k}}{p^\star_{k,\varphi_k} + \eta_{k,\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{p^\star_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > C_k^1, \\ -\infty, & \text{otherwise,} \end{cases} \tag{20}$$

where $p^\star_{k,\varphi_k}$ and $p^\star_{k,\psi_k}$ are given by (17) and (18), respectively. Then, the optimal last SIC user on Subchannel k, indexed by φ_k^*, can be determined as

$$\varphi_k^* = \operatorname*{argmax}_{\varphi_k \in \mathcal{N}} \phi_k(\bar{P}_k;\varphi_k), \tag{21}$$

and the corresponding optimal value is given by

$$\phi_k^*(\bar{P}_k) = \phi_k(\bar{P}_k;\varphi_k^*). \tag{22}$$

Consequently, Users φ_k^* and $\psi_k(\varphi_k^*)$ are selected as the optimal *candidate* users on Subchannel k. Then, by Theorem 4, the power allocation solution to Problem (Q_k) can be obtained as $\mathbf{p}_k^\dagger = (p^\dagger_{k,i})_{\forall i \in \mathcal{N}}$, where $p^\dagger_{k,\varphi_k^*} = p^\star_{k,\varphi_k^*}$, $p^\dagger_{k,\psi_k(\varphi_k^*)} = p^\star_{k,\psi_k(\varphi_k^*)}$, and $p^\dagger_{k,i} = 0$ for all $i \in \mathcal{N} \setminus \{\varphi_k^*, \psi_k(\varphi_k^*)\}$. Finally, the solution, $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$, to Problem ($\mathrm{S}_k$) can be derived from $\mathbf{p}_k^\dagger$ using Theorem 2. We summarize this process in Algorithm 2.

B. Solution to Problem (M)

In this subsection, we discuss how to solve Problem (M) for the given $\phi^*(\bar{\mathbf{P}}) = \sum_{k \in \mathcal{K}} \phi_k^*(\bar{P}_k)$. To be specific, we find an optimal coupling vector, $\bar{\mathbf{P}}^*$, that maximizes $\phi^*(\bar{\mathbf{P}})$ under the limited total transmission power budget of $P_{\max}$. For compact notation, we write the optimal last SIC user, φ_k^*, and the other *candidate* user, $\psi_k(\varphi_k^*)$, on Subchannel k as φ_k and ψ_k, respectively, if there is no confusion.

By plugging the optimal power allocation solution derived in Theorem 4 into (22), we can obtain $\phi_k^*(\bar{P}_k)$ as

$$\phi_k^*(\bar{P}_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \le C_k^2(\bar{P}_k), \\ \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\varphi_k}}\right), & \text{otherwise,} \end{cases} \tag{23}$$

where $C_k^2(\bar{P}_k)$ is defined in (19), and

$$C_k^3 = \tilde{w}_{\psi_k} B_k \log_2\left(\frac{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}{\eta_{k,\varphi_k} - \eta_{k,\psi_k}} \cdot \frac{\eta_{k,\psi_k}}{\tilde{w}_{\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(\frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} \cdot \frac{\tilde{w}_{\varphi_k}}{\eta_{k,\varphi_k}}\right). \tag{24}$$

Note that, in (23), the case where $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \le C_k^1$ is excluded since $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k}$ is always greater than C_k^1 as long as the optimal *candidate* users are selected by Algorithm 2. In succession, since dealing with (23) is difficult due to $C_k^2(\bar{P}_k)$ that varies with $\bar{P}_k$, we reformulate it equivalently as

$$\phi_k^*(\bar{P}_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \bar{P}_k \ge C_k^4, \\ \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\varphi_k}}\right), & \text{otherwise,} \end{cases} \tag{25}$$

where

$$C_k^4 = \frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}. \tag{26}$$

The proof of the equivalence of (23) and (25) is provided in Appendix D. From (25), we can see that if $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \ge 1$, $\phi_k^*(\bar{P}_k)$ is a continuous logarithmic function. However, if $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1$, it is a piecewise nonlinear function with a breakpoint at $\bar{P}_k = C_k^4$. Nevertheless, we can show that it is a continuously differentiable concave function.

Proposition 1. *The function ϕ_k^* in* (25) *is a continuously differentiable concave function of $\bar{P}_k$ on* $[0, \infty)$.

Proof. See Appendix E. □

By Proposition 1, we can conclude that the objective function, ϕ^*, of Problem (M) is a concave function of $\bar{\mathbf{P}}$, and accordingly, Problem (M) is a convex optimization problem. Hence, we can obtain its optimal solution using the Karush–Kuhn–Tucker (KKT) conditions [37].

Algorithm 3: Solution to Problem (M)

Let $f(\mu) = \sum_{k \in \mathcal{K}} \bar{P}_k^* - P_{\max}$.
Initialize μ_L to zero, and μ_U to a sufficiently large value.
repeat
    $\mu_{\text{new}} \leftarrow (\mu_L + \mu_U)/2$.
    if $f(\mu_L) \cdot f(\mu_{\text{new}}) < 0$ **then** $\mu_U \leftarrow \mu_{\text{new}}$.
    if $f(\mu_U) \cdot f(\mu_{\text{new}}) < 0$ **then** $\mu_L \leftarrow \mu_{\text{new}}$.
until *convergence*
Obtain $\bar{\mathbf{P}}^*$ by plugging $\mu^* = (\mu_L + \mu_U)/2$ into (27).
return $\bar{\mathbf{P}}^*$.

Theorem 5. *The optimal solution, $\bar{\mathbf{P}}^* = (\bar{P}_k^*)_{\forall k \in \mathcal{K}}$, to Problem* (M) *is provided as follows. For each Subchannel $k \in \mathcal{K}$,*

$$\bar{P}_k^* = \begin{cases} \left[\mu^* \tilde{w}_{\psi_k} B_k - \eta_{k,\psi_k}\right]_0^{P_{\max,k}}, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \mu^* > C_k^5, \\ \left[\mu^* \tilde{w}_{\varphi_k} B_k - \eta_{k,\varphi_k}\right]_0^{P_{\max,k}}, & \text{otherwise,} \end{cases} \tag{27}$$

where

$$C_k^5 = \frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{B_k(\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k})}, \tag{28}$$

and μ^* is chosen to satisfy $\sum_{k \in \mathcal{K}} \bar{P}_k^* = P_{\max}$.

Proof. See Appendix F. □

Note that $\bar{P}_k^*$ is continuous, piecewise-linear, and increasing with respect to μ^*. Hence, there exists a unique solution, μ^*, which can be easily found by any simple root-finding method such as a bisection method. Once μ^* is determined, the optimal solution, $\bar{\mathbf{P}}^*$, to Problem (M) can be obtained using (27). The pseudocode based on the bisection method is provided in Algorithm 3. Finding the root of a single variable μ^*, as needed in our algorithm, in general, is much faster than standard convex programming tools required to solve Problem (M) consisting of K variables and $2K + 1$ constraints.

Remark. We can achieve the proportional fair scheduling by solving Problem (P_1^t) in every time slot using the Joint-SAPA-LCC algorithm, where the effective weight of User i in time slot t is given by

$$\tilde{w}_i^t = \frac{1}{R_{\text{EMA},i}^t}, \tag{29}$$

where $R^t_{\text{EMA},i}$ is the exponential moving average data rate of User i in time slot t, and it can be recursively updated by

$$R^{t+1}_{\text{EMA},i} = \left(1 - \frac{1}{\tau}\right) R^t_{\text{EMA},i} + \frac{1}{\tau} R^t_i, \tag{30}$$

where τ is the time-averaging window coefficient.

However, this proportional fair scheduling may violate the explicitly given QoS requirements as in (7).

IV. Opportunistic MC-NOMA Scheduling

In this section, we finally develop an opportunistic MC-NOMA scheduling algorithm that works in an online manner by decomposing Problem (P1) into a series of Joint-SAPA problems over time slots, i.e., Problem (P^t_1) for each time slot. To this end, we first take advantage of the fact that due to the ergodicity of the fading process, the long-term time average converges almost surely to the expectation for almost all realizations of the fading process. Thereby, by denoting a channel vector in a generic time slot as $\mathbf{h}$ without the superscript t and replacing the superscript t of the decision variables with $\mathbf{h}$, we reformulate Problem (P1) as

$$\begin{aligned} \text{(P2)} \quad \underset{\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}, \forall \mathbf{h}}{\text{maximize}} \quad & \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h})\right] \\ \text{subject to} \quad & \mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h})\right] \geq \bar{R}_{\min,i}, \ \forall i \in \mathcal{N}, \\ & \mathbf{p}^{\mathbf{h}} \in \mathcal{P}, \ \mathbf{q}^{\mathbf{h}} \in \mathcal{Q}, \ \forall \mathbf{h}. \end{aligned}$$

At each time slot t where the channel vector is realized as $\mathbf{h}^t$, subchannel assignment and power allocation can be done according to the solution for $\mathbf{p}^{\mathbf{h}}$ and $\mathbf{q}^{\mathbf{h}}$ obtained by solving Problem (P2) with $\mathbf{h} = \mathbf{h}^t$.

There is still a big challenge in solving Problem (P2). Since no information on the underlying distributions of the fading process is provided, we have to solve the stochastic optimization problem without such information. To resolve it, we leverage the Lagrangian duality and the stochastic optimization theory to develop the opportunistic MC-NOMA scheduling algorithm. Its core mechanism is to take advantage of the time-varying channel conditions opportunistically to maximize the weighted average sum rate. Also, the effective weights are systemically adjusted so that the QoS requirements (i.e., the individual minimum average data rate requirements) are fulfilled. To develop the algorithm, by introducing a Lagrange multiplier, λ_i, for the minimum average data rate constraint of User i, we first define a Lagrangian function, L, associated with Problem (P2) as

$$L(\bar{\mathbf{p}}, \bar{\mathbf{q}}, \boldsymbol{\lambda}) = \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h})\right] + \sum_{i \in \mathcal{N}} \lambda_i \Big(\mathbb{E}_{\mathbf{h}}\big[R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h})\big] - \bar{R}_{\min,i}\Big)$$
$$= \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} (w_i + \lambda_i) R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h})\right] - \sum_{i \in \mathcal{N}} \lambda_i \bar{R}_{\min,i}, \tag{31}$$

where $\bar{\mathbf{p}} = (\mathbf{p}^{\mathbf{h}})_{\forall \mathbf{h}}$, $\bar{\mathbf{q}} = (\mathbf{q}^{\mathbf{h}})_{\forall \mathbf{h}}$, and $\boldsymbol{\lambda} = (\lambda_i)_{\forall i \in \mathcal{N}}$. Then, the dual problem associated with Problem (P2) can be defined by

$$\text{(D)} \quad \underset{\boldsymbol{\lambda}}{\text{minimize}} \quad F(\boldsymbol{\lambda})$$
$$\text{subject to} \quad \lambda_i \geq 0, \ \forall i \in \mathcal{N},$$

where

$$F(\boldsymbol{\lambda}) = \underset{\mathbf{p}^{\mathbf{h}} \in \mathcal{P},\, \mathbf{q}^{\mathbf{h}} \in \mathcal{Q},\, \forall \mathbf{h}}{\text{maximize}} \quad L(\bar{\mathbf{p}}, \bar{\mathbf{q}}, \boldsymbol{\lambda}). \tag{32}$$

Since Problem (P2) is nonconvex, even though its dual problem, Problem (D), is optimally solved, there may be a duality gap. However, the duality gap vanishes in our problem, resulting in no loss of optimality.

Theorem 6. *The strong duality (i.e., zero duality gap) holds between Problem* (P2) *and its dual problem, Problem* (D).

Proof. See Appendix G. □

We thus develop an algorithm that solves Problem (D). To this end, we first focus on obtaining its objective function, $F(\boldsymbol{\lambda})$. The first (expectation) term in (31) is separable for each channel realization, and the second term is independent of the decision variables, $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$. Hence, for a given Lagrange multiplier vector, $\boldsymbol{\lambda}$, the maximization in (32) can be solved by separately solving the subproblem for each channel realization, defined by

$$(\text{D}^{\mathbf{h}}) \quad \underset{\mathbf{p}^{\mathbf{h}} \in \mathcal{P},\, \mathbf{q}^{\mathbf{h}} \in \mathcal{Q}}{\text{maximize}} \quad \sum_{i \in \mathcal{N}} (w_i + \lambda_i) R_i(\mathbf{p}^{\mathbf{h}}, \mathbf{q}^{\mathbf{h}}; \mathbf{h}).$$

Since the expectation has disappeared in Problem ($\text{D}^{\mathbf{h}}$), it can be solved without knowledge of the underlying distributions of the fading process once the channel realization is provided. Thus, for any given $\boldsymbol{\lambda}$ and $\mathbf{h}$, Problem ($\text{D}^{\mathbf{h}}$) turns into a deterministic optimization problem for Joint-SAPA that aims to maximize the instantaneous weighted sum rate with weight $w_i + \lambda_i$

Algorithm 4: Opportunistic MC-NOMA Scheduling

Initialize: $t = 1$ and $\boldsymbol{\lambda}^t = \mathbf{0}$.
for *each time slot* t **do**
    Solve Problem ($\mathrm{D}^{\mathbf{h}}$) with $\mathbf{h} = \mathbf{h}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$ using the Joint-SAPA-LCC algorithm (Algorithm 1).
    Transmit the signal generated by the obtained solution.
    Calculate $\boldsymbol{\lambda}^{t+1}$ according to (33) and (34).
    $t \leftarrow t + 1$.

for User i. This problem can be solved using the Joint-SAPA-LCC algorithm developed in the previous section with letting $\tilde{w}_i = w_i + \lambda_i$.

We now focus back on solving Problem (D). Even though Problem ($\mathrm{D}^{\mathbf{h}}$) can be solved for given $\mathbf{h}$ and $\boldsymbol{\lambda}$, the underlying distributions of the fading process are still required to solve Problem (D). Nevertheless, thanks to the fact that Problem (D) is a convex stochastic optimization problem, we can solve the problem without resorting it using the stochastic subgradient method [38], where the Lagrange multiplier vector, $\boldsymbol{\lambda}$, is iteratively updated by

$$\boldsymbol{\lambda}^{t+1} = \left[\boldsymbol{\lambda}^t - \zeta^t \mathbf{v}^t\right]^+, \tag{33}$$

where $\boldsymbol{\lambda}^t$ and ζ^t are the Lagrange multiplier vector and the positive step size in time slot t, respectively, and $\mathbf{v}^t = (v_i^t)_{\forall i \in \mathcal{N}}$ is the stochastic subgradient of $F(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$. By Danskin's min-max theorem [39], the stochastic subgradient, $\mathbf{v}^t$, can be obtained by

$$v_i^t = R_i^t - \bar{R}_{\min,i}, \ \forall i \in \mathcal{N}, \tag{34}$$

where R_i^t is the instantaneous data rate of User i in time slot t defined as in (4), which is achieved when the subchannel assignment and power allocation are performed according to the solution to Problem ($\mathrm{D}^{\mathbf{h}}$) with $\mathbf{h} = \mathbf{h}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$. With the update process of (33), the Lagrange multiplier vector converges almost surely to the optimal solution, $\boldsymbol{\lambda}^*$, of Problem (D) if the step size ζ^t is square-summable, but not summable [40], i.e.,

$$\zeta^t \geq 0, \ \sum_{t=1}^{\infty} \zeta^t = \infty, \ \text{and} \ \sum_{t=1}^{\infty} (\zeta^t)^2 < \infty. \tag{35}$$

The proposed algorithm for the opportunistic MC-NOMA scheduling is outlined in Algorithm 4.

V. Simulation Results

In this section, we present simulation results to evaluate the performance of our proposed algorithms. We first investigate the Joint-SAPA-LCC algorithm and then the opportunistic MC-NOMA scheduling algorithm. Throughout the simulations, we consider a circular cell with a radius of 300 m, in which one BS located at the center of the cell serves N users over K subchannels. The system bandwidth, B_{tot}, is set to 5 MHz, and the bandwidth of each subchannel is set equally to B_{tot}/K. Unless otherwise specified, the numbers of users and subchannels are set to 10, i.e., $N = 10$ and $K = 10$. The total transmission power budget, $P_{\max}$, of the BS is set to 43 dBm, and the transmission power budget for each subchannel is set to $\gamma P_{\max}/K$ with $\gamma = 1.15$. The large-scale path loss is modeled by the HATA model for urban environments [41], where effective antenna heights of the BS and the users are set to 30 m and 2 m, respectively, the transmission frequency is set to 900 MHz, and the antenna gains at the BS and the users are set to 15 dBi and 0 dBi, respectively [42]. Additionally, the shadow fading with a standard deviation of 8 dB and the Rayleigh small-scale fading with unit variance are considered. The noise power spectral density, N_0, is set to -174 dBm/Hz. For a straightforward analysis of simulation results, we express data rates in units of bps/Hz. In Algorithm 4, the step size in (33) is set to $\zeta^t = 1/t$, which satisfies the conditions in (35) so that the convergence of the algorithm is guaranteed.

A. Joint Subchannel Assignment and Power Allocation

In this subsection, we provide the performance of our Joint-SAPA-LCC algorithm that aims to maximize the weighted sum rate by solving Problem (P_1^t). For comparison, we additionally provide the performance of three other algorithms. The first one is the Joint-SAPA-FTPC algorithm [27], where Joint-SAPA is performed based on the FTPC and ES algorithms. The second one is the Joint-SAPA-DCP algorithm [28], where Joint-SAPA is performed based on the DCP approach. The last one is the Joint-SAPA-DP algorithm [32], where Joint-SAPA is performed based on the DP approach and the PGD method. In the following simulations, we assume that N users are uniformly distributed within the circular cell with at least 30 m away from the BS, and their weights are randomly set between 0 and 1. Also, taking into account the high computational complexity of the above baseline algorithms, we assume that each subchannel can be assigned to up to 5 users, i.e., $M = 5$. All the simulation results are averaged over 3000 independent trials. For each trial, locations, weights, and channel gains of all users are independently generated.

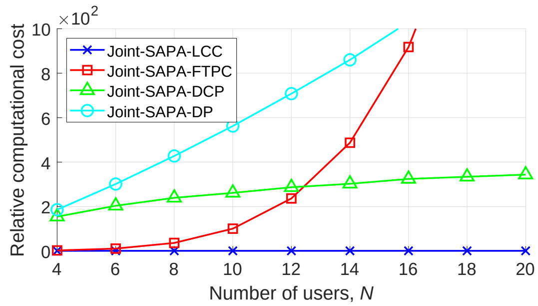


Fig. 2. The relative computational cost versus the number of users.

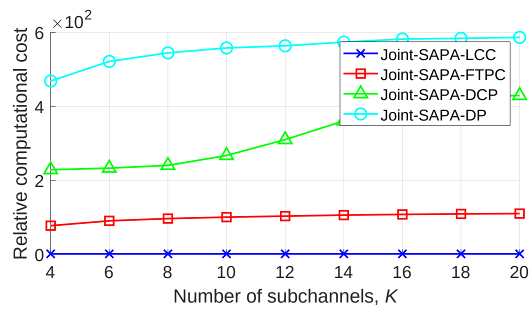


Fig. 3. The relative computational cost versus the number of subchannels.

We first compare the computational complexity of the Joint-SAPA algorithms. To this end, we define the relative computational cost of an algorithm as its execution time normalized to that of our Joint-SAPA-LCC algorithm, and then show the corresponding results for different numbers of users and subchannels in Figs. 2 and 3, respectively. The execution times were measured by MATLAB R2020a software on a computer with Intel Core i7-9700K CPU (3.60 GHz) and 32.0 GB RAM. From the figures, we can see that our Joint-SAPA-LCC algorithm is much faster than the other algorithms. For example, when $N = 10$ and $K = 10$, the computational cost of Joint-SAPA-LCC is about 100, 250, and 550 times lower than those of Joint-SAPA-FTPC, Joint-SAPA-DCP, and Joint-SAPA-DP, respectively. The main reason why our Joint-SAPA-LCC algorithm is fast is that the *candidate* users who might be allocated positive power are determined simply based on the closed-form power allocation formulas. Furthermore, although not proven theoretically, we were able to observe experimentally that Algorithm 1 converges in only a single iteration in most cases. That is, in most cases, the Joint-SAPA-LCC algorithm is performed in a 3-step procedure: i) to obtain ϕ^* by selecting *candidate* users based on equal subchannel power allocation, ii) to refine the subchannel power allocation $\bar{\mathbf{P}}$ based on ϕ^* obtained in the first

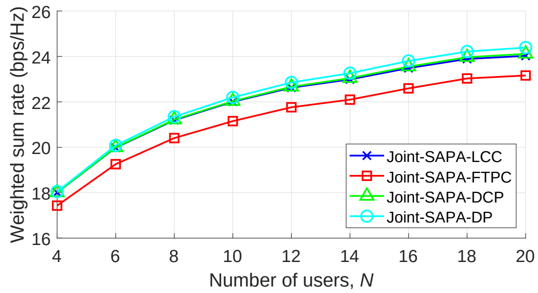


Fig. 4. The weighted sum rate versus the number of users.

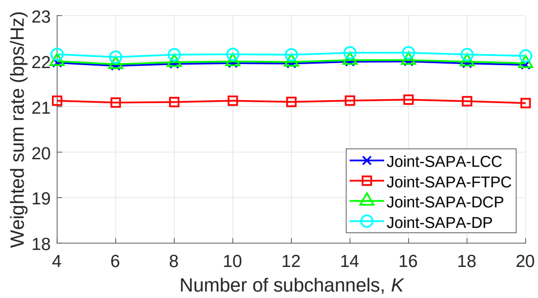
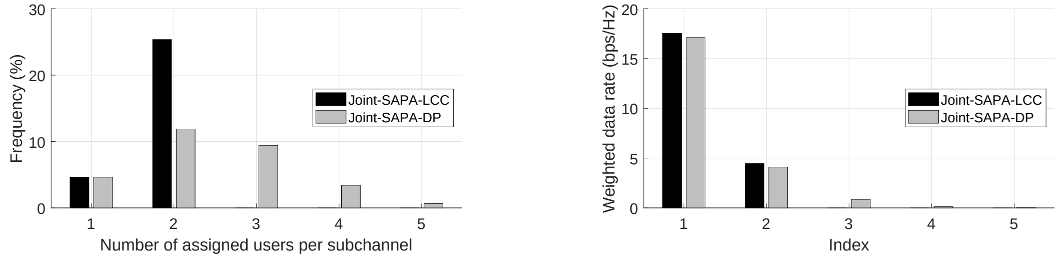


Fig. 5. The weighted sum rate versus the number of subchannels.

step, and iii) to obtain the final Joint-SAPA solution based on $\bar{\mathbf{P}}$ obtained in the second step. This computational complexity comparison confirms that our Joint-SAPA-LCC algorithm is very effective and well suited to be implemented in practical systems where Joint-SAPA should be performed in every very short time slot.

In Figs. 4 and 5, we compare the weighted sum rate performance with varying numbers of users and subchannels, respectively. First, as shown in Fig. 4, as the number of users increases, the weighted sum rates increase in all the Joint-SAPA algorithms thanks to the increase in the multi-user diversity gain. Also, we can see that our Joint-SAPA-LCC algorithm, despite its very low computational complexity, has only a little performance drop compared to the Joint-SAPA-DP and Joint-SAPA-DCP algorithms, and provides higher performance compared to the Joint-SAPA-FTPC algorithm. On the other hand, Fig. 5 shows that the weighted sum rates hardly change despite the increase in the number of subchannels. This is because the total bandwidth of the system is fixed at 5 MHz. However, the order between the Joint-SAPA algorithms in terms of the weighted sum rate performance remains the same as in Fig. 4. For example, when $N = 10$ and $K = 10$, the weighted sum rate of our Joint-SAPA-LCC algorithm is only 0.86 % and 0.12 %



(a) The frequency histogram of the number of assigned users per subchannel.

(b) The weighted data rates in descending order.

Fig. 6. The comparison between the Joint-SAPA-LCC and Joint-SAPA-DP algorithms for the case where $N = 10$ and $K = 10$.

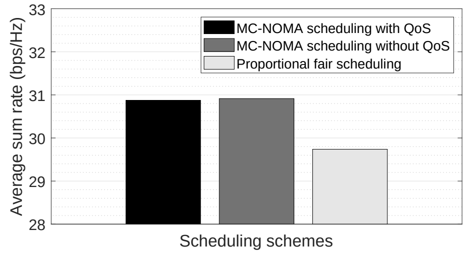
lower than those of the Joint-SAPA-DP and Joint-SAPA-DCP algorithms, respectively, whereas it is more than 4 % higher than that of the Joint-SAPA-FTPC algorithm.

Next, in Fig. 6, we compare our Joint-SAPA-LCC algorithm with the Joint-SAPA-DP algorithm, which provides the highest weighted sum rate, in more depth for the case where $N = 10$ and $K = 10$. Fig. 6a shows the frequency histogram of the number of assigned users per subchannel, and Fig. 6b shows the weighted data rates arranged in descending order. Specifically, the value of the bar for Index i represents the average value of the ith highest weighted data rates over all subchannels and all channel realizations. According to Fig. 6a, unlike our Joint-SAPA-LCC algorithm which is supposed to assign up to two users per subchannel, the probability of selecting three or more users per subchannel reaches about 45 % in the Joint-SAPA-DP algorithm, which is not small. However, Fig. 6b demonstrates that the sum of the top two weighted data rates in the Joint-SAPA-DP algorithm occupies more than 95 % of the weighted sum rate, which implies that the remaining bottom three weighted data rates have a very little impact on the weighted sum rate performance. This is why the weighted sum rate performance of our Joint-SAPA-LCC algorithm is very close to that of the Joint-SAPA-DP algorithm, even though ours allows only up to two users per subchannel. To sum up, the simulation results thus far confirm that not only does our Joint-SAPA-LCC algorithm provide good performance close to that of the Joint-SAPA-DP algorithm, but it also has much lower computational complexity compared to the other Joint-SAPA algorithms, which is critical for implementation in practical systems.

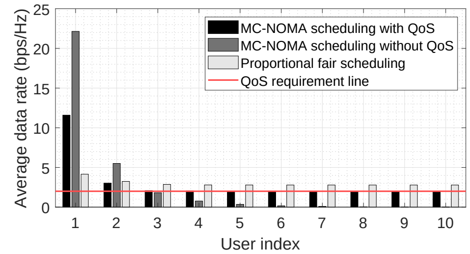
B. Opportunistic MC-NOMA scheduling

In this subsection, we provide the performance of our opportunistic MC-NOMA scheduling algorithm, taking into account the time-varying and frequency-selective channel conditions and various QoS requirements, i.e., the required minimum average data rates of users. To show the effectiveness of our MC-NOMA scheduling algorithm, we compare its simulation results with those of two other scheduling algorithms: MC-NOMA scheduling without QoS requirements and proportional fair scheduling. To be specific, the MC-NOMA scheduling without QoS requirements is achieved by solving Problem (P1) using our opportunistic MC-NOMA scheduling algorithm, where $\bar{R}_{\min,i}$ is set to 0 for all $i \in \mathcal{N}$. Meanwhile, the proportional fair scheduling is achieved by solving Problem (P_1^t) using our Joint-SAPA-LCC algorithm at each time slot, where the weight of each user is given as the reciprocal of its time-averaged data rate up until to that time slot as in (29). The time-averaging window coefficient, τ, in (30) is set to 1000. Throughout the following simulations, we consider a system where there are 10 users with equal weights in the cell, and the ith user is $30\times i$ m away from the BS. Thus, the lower the user index, the closer it is to the BS, resulting in a higher channel gain on average. The performance results of the scheduling algorithms are investigated in the following two scenarios. In one scenario, we assume that all users have the same QoS requirements. Specifically, the minimum average data rates of all the users are set to 2 bps/Hz. In the other scenario, we assume that the users have individually different QoS requirements. Specifically, the minimum average data rates of Users 1, 2, 5, 6, 9, and 10 are set to 3.5 bps/Hz, while those of Users 3, 4, 7, and 8 are set to 1 bps/Hz.

Figs. 7 and 8 show the performance results for the first and second scenarios, respectively. In Figs. 7a and 8a, we can see that the average sum rates of the MC-NOMA scheduling without QoS requirements are slightly higher than those of the MC-NOMA scheduling with QoS requirements in both cases. As can be expected, this result is obvious because the feasible space of Problem (P1) with positive $\bar{R}_{\min,i}$'s is a subspace of that of Problem (P1) with zero $\bar{R}_{\min,i}$'s. However, the lack of the QoS requirements makes the performance of users with poor channel conditions compromised to maximize the sum rate. Consequently, as can be seen in Figs. 7b and 8b, only a few users close to the BS exploit the resources exclusively, and thereby users far from the BS do not meet their QoS requirements. On the contrary, the QoS requirements of all users are satisfied well in the MC-NOMA scheduling with QoS requirements. Meanwhile,

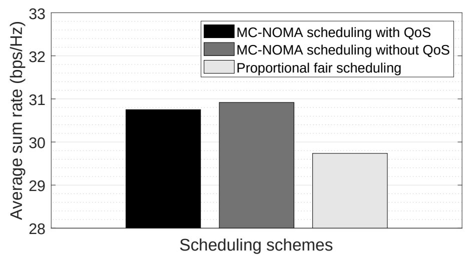


(a) The average sum rate.

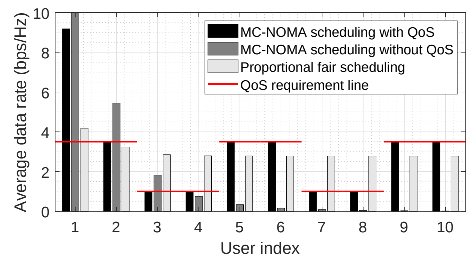


(b) The average data rate of each users.

Fig. 7. Performance comparison results between scheduling algorithms when all users have the same QoS requirements.



(a) The average sum rate.



(b) The average data rate of each users.

Fig. 8. Performance comparison results between scheduling algorithms when users have individually different QoS requirements.

the proportional fair scheduling follows the principle of giving high effective weights to users with low time-averaged data rates. Accordingly, it not only prevents users from starvation, but also provides similar average data rate performance among users to gratify its purpose of maximizing the fairness utility function. However, since it concentrates on the fairness between users and does not take into account the QoS requirements explicitly, a situation where the given QoS requirements are not satisfied may occur. Fig. 8b shows the case where the proportional fair scheduling cannot meet the QoS requirements, whereas our MC-NOMA scheduling with QoS requirements well satisfies them. In summary, the simulation results demonstrate that our proposed scheduling algorithm not only provides good performance, but also guarantees given QoS requirements.

VI. Conclusion

In this paper, we have studied the dynamic resource scheduling problem for joint user, subchannel, and power scheduling in the downlink MC-NOMA system over time-varying fading channels, which has the goal of maximizing the weighted average sum rate while ensuring given QoS requirements. To this end, we have first developed the Joint-SAPA-LCC algorithm to maximize the instantaneous weighted sum rate. By its characteristic that it leverages very simple user selection and power allocation based on closed-form equations, we could achieve much lower computational complexity compared to the existing Joint-SAPA algorithms. In succession, along with the proposal of the proportional fair scheduling based on our Joint-SAPA-LCC algorithm, we have developed the opportunistic MC-NOMA scheduling algorithm that systematically adjusts the effective weights so that the weighted average sum rate is maximized while the QoS requirements are met. Then, through the extensive simulation results, we have demonstrated that our Joint-SAPA-LCC algorithm provides good performance comparable to the Joint-SAPA-DP algorithm despite its much lower computational complexity, and that our opportunistic MC-NOMA scheduling algorithm satisfies given QoS requirements. This study will be the cornerstone of our future work on the development of scheduling for multi-cell MC-NOMA systems, massive multiple-input multiple-output (MIMO) MC-NOMA systems, and other scenarios.

Appendix A
Proof of Theorem 1

First, the objective value of Problem (P_2), i.e., the weighted sum rate over all subchannels, is bounded above since the total transmission power budget of the BS is limited. Hence, by the monotone convergence theorem, we can guarantee that Algorithm 1 converges to a stationary point as long as the weighted sum rate over all subchannels monotonically increases as the iteration progresses.

Let $\{\mathbf{p}^{(l)}, \mathbf{q}^{(l)}, \bar{\mathbf{P}}^{(l)}\}$ be the solution to Problem (P_2) obtained at the beginning of the lth iteration of Algorithm 1, and $y(\mathbf{p}^{(l)}, \mathbf{q}^{(l)}, \bar{\mathbf{P}}^{(l)})$ be the corresponding weighted sum rate over all subchannels. In Algorithm 1, two updates are performed in each iteration: one for both the power allocation and subchannel assignment indicator vectors (i.e., $\mathbf{p}$ and $\mathbf{q}$), and the other for the coupling vector (i.e., $\bar{\mathbf{P}}$). Accordingly, in the lth iteration, the weighted sum rate over all subchannels evolves as

$$y(\mathbf{p}^{(l)}, \mathbf{q}^{(l)}, \bar{\mathbf{P}}^{(l)}) \overset{(a)}{\rightarrow} y(\mathbf{p}^{(l+1)}, \mathbf{q}^{(l+1)}, \bar{\mathbf{P}}^{(l)}) \overset{(b)}{\rightarrow} y(\mathbf{p}^{(l+1)}, \mathbf{q}^{(l+1)}, \bar{\mathbf{P}}^{(l+1)}), \tag{A.1}$$

where (a) and (b) are done by solving Problem (S_k) for all $k \in \mathcal{K}$ and Problem (M), respectively. Since Problem (S_k) is to optimize $\{\mathbf{p}_k, \mathbf{q}_k\}$ so that the weighted sum rate on Subchannel k under the limited power of $\bar{P}_k^{(l)}$ is maximized, it is obvious that

$$y(\mathbf{p}^{(l)}, \mathbf{q}^{(l)}, \bar{\mathbf{P}}^{(l)}) \leq y(\mathbf{p}^{(l+1)}, \mathbf{q}^{(l+1)}, \bar{\mathbf{P}}^{(l)}). \tag{A.2}$$

Similarly, since Problem (M) is to optimize $\bar{\mathbf{P}}$ so that the weighted sum rate over all subchannels is maximized, it is obvious that

$$y(\mathbf{p}^{(l+1)}, \mathbf{q}^{(l+1)}, \bar{\mathbf{P}}^{(l)}) \leq y(\mathbf{p}^{(l+1)}, \mathbf{q}^{(l+1)}, \bar{\mathbf{P}}^{(l+1)}). \tag{A.3}$$

By combining (A.2) and (A.3), we can easily see that the weighted sum rate over all subchannels, i.e., the objective value of Problem (P_2), is monotonically increasing as the iteration progresses.

□

Appendix B
Proof of Theorem 2

Let us denote the objective function of Problem (S_k), its optimal solution, and its optimal value by $f_{(\mathrm{S}_k)}(\mathbf{p}_k, \mathbf{q}_k)$, $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$, and $f_{(\mathrm{S}_k)}^* = f_{(\mathrm{S}_k)}(\mathbf{p}_k^*, \mathbf{q}_k^*)$, respectively. We also denote the objective function of Problem (Q_k), its optimal solution, and its optimal value by $f_{(\mathrm{Q}_k)}(\mathbf{p}_k)$, $\mathbf{p}_k^\dagger$, and $f_{(\mathrm{Q}_k)}^\dagger = f_{(\mathrm{Q}_k)}(\mathbf{p}_k^\dagger)$, respectively. In addition, we define a feasible set for the subchannel assignment indicator vector for Subchannel k by $\mathcal{Q}_k = \{\mathbf{q}_k \in \{0,1\}^N : \sum_{i\in\mathcal{N}} q_{k,i} \leq M\}$, and denote a set of users accessing Subchannel k as $\mathcal{N}_k = \{i \in \mathcal{N} : q_{k,i} = 1\}$.

Now, suppose that $\mathbf{q}_k$ is arbitrarily given such that $\mathbf{q}_k \in \mathcal{Q}_k$. Then, Problem ($\mathrm{S}_k$) turns into

$$\begin{aligned} \underset{\mathbf{p}_k}{\text{maximize}} \quad & \hat{f}_{(\mathrm{S}_k)}(\mathbf{p}_k; \mathbf{q}_k) \\ \text{subject to} \quad & \sum_{i\in\mathcal{N}} p_{k,i} \leq \bar{P}_k, \\ & p_{k,i} \geq 0, \ \forall i \in \mathcal{N}, \end{aligned} \tag{B.1}$$

where

$$\hat{f}_{(\mathrm{S}_k)}(\mathbf{p}_k; \mathbf{q}_k) = \sum_{i\in\mathcal{N}_k} \tilde{w}_i B_k \log_2\left(1 + \frac{p_{k,i}}{\sum_{j\in\mathcal{N}_k:\eta_{k,j}<\eta_{k,i}} p_{k,j} + \eta_{k,i}}\right). \tag{B.2}$$

Here, it is worth noting that the problem in (B.1) is identical with Problem (Q_k) except that the user set is given as $\mathcal{N}_k$ instead of $\mathcal{N}$ in the objective function. Since $\mathcal{N}_k$ is a subset of $\mathcal{N}$, it is obvious that $\hat{f}_{(\mathrm{S}_k)}^*(\mathbf{q}_k) \leq f_{(\mathrm{Q}_k)}^\dagger$ for any $\mathbf{q}_k \in \mathcal{Q}_k$, where $\hat{f}_{(\mathrm{S}_k)}^*(\mathbf{q}_k)$ is the optimal value of the problem in (B.1) for a given $\mathbf{q}_k$. Thus, we can deduce that

$$f_{(\mathrm{S}_k)}^* \leq f_{(\mathrm{Q}_k)}^\dagger. \tag{B.3}$$

Next, suppose that the optimal solution, $\mathbf{p}_k^\dagger$, to Problem (Q_k) is given such that $\sum_{i\in\mathcal{N}} \mathbf{1}_{\{p_{k,i}^\dagger>0\}} \leq M$, and let us define

$$\hat{\mathbf{p}}_k = \mathbf{p}_k^\dagger \text{ and } \hat{\mathbf{q}}_k = (\hat{q}_{k,i})_{\forall i\in\mathcal{N}}, \tag{B.4}$$

where $\hat{q}_{k,i} = \mathbf{1}_{\{p_{k,i}^\dagger>0\}}$ for all $i \in \mathcal{N}$. Then, $\hat{\mathbf{p}}_k$ and $\hat{\mathbf{q}}_k$ satisfy the constraints in Problem (S_k). Also, by simple arithmetic operations, we can easily see that

$$f_{(\mathrm{S}_k)}(\hat{\mathbf{p}}_k, \hat{\mathbf{q}}_k) = f_{(\mathrm{Q}_k)}^\dagger. \tag{B.5}$$

By (B.3) and (B.5), we can deduce that $\{\hat{\mathbf{p}}_k, \hat{\mathbf{q}}_k\}$ is an optimal solution to Problem (S_k), i.e., $\mathbf{p}_k^* = \hat{\mathbf{p}}_k$ and $\mathbf{q}_k^* = \hat{\mathbf{q}}_k$. □

Appendix C
Proof of Theorem 4

Since the larger the transmission power, the higher the weighted sum rate, we can replace the first constraint in Problem ($\mathrm{Q}_{k,4}^{\varphi_k}$) with $p_{k,\psi_k} + p_{k,\varphi_k} = \bar{P}_k$. Then, by substituting p_{k,ψ_k} for $\bar{P}_k - p_{k,\varphi_k}$, we can reformulate Problem ($\mathrm{Q}_{k,4}^{\varphi_k}$) as

$$\underset{0\leq p_{k,\varphi_k}\leq \bar{P}^k}{\text{maximize}} \quad g(p_{k,\varphi_k}), \tag{C.1}$$

where

$$g(p_{k,\varphi_k}) = \tilde{w}_{\psi_k} B_k \log_2\left(\frac{\bar{P}_k + \eta_{k,\psi_k}}{p_{k,\varphi_k} + \eta_{k,\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{p_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right). \tag{C.2}$$

The derivative of $g(p_{k,\varphi_k})$ with respect to p_{k,φ_k} is given as

$$\begin{aligned} g'(p_{k,\varphi_k}) &= \frac{B_k}{\ln 2} \times \left[\frac{-\tilde{w}_{\psi_k}}{p_{k,\varphi_k} + \eta_{k,\psi_k}} + \frac{\tilde{w}_{\varphi_k}}{p_{k,\varphi_k} + \eta_{k,\varphi_k}}\right] \\ &= \frac{B_k}{\ln 2} \times \frac{(\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}) p_{k,\varphi_k} + \tilde{w}_{\varphi_k}\eta_{k,\psi_k} - \tilde{w}_{\psi_k}\eta_{k,\varphi_k}}{(p_{k,\varphi_k} + \eta_{k,\psi_k})(p_{k,\varphi_k} + \eta_{k,\varphi_k})}. \end{aligned} \tag{C.3}$$

From (C.3), we have $g'(\hat{p}_{k,\varphi_k}) = 0$ if and only if

$$\hat{p}_{k,\varphi_k} = \frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}. \tag{C.4}$$

Using (C.3) and (C.4), we can derive an optimal solution, $p_{k,\varphi_k}^\star$, to the problem in (C.1) by considering the following three mutually exclusive cases.

1) Suppose that $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > 1$. Then, according to (C.3), $g'(p_{k,\varphi_k}) < 0$ if $p_{k,\varphi_k} < \hat{p}_{k,\varphi_k}$ and $g'(p_{k,\varphi_k}) \geq 0$ otherwise. Also, according to (C.4), $\hat{p}_{k,\varphi_k} < 0$ since $\eta_{k,\varphi_k} < \eta_{k,\psi_k}$. Thus, $g(p_{k,\varphi_k})$ is an increasing function on $[0, \bar{P}_k]$, resulting in $p_{k,\varphi_k}^\star = \bar{P}_k$.

2) Suppose that $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} = 1$. Then, according to (C.3), $g'(p_{k,\varphi_k}) > 0$ for any $p_{k,\varphi_k} \in [0, \bar{P}_k]$. Thus, $g(p_{k,\varphi_k})$ is an increasing function on $[0, \bar{P}_k]$, resulting in $p^{\star}_{k,\varphi_k} = \bar{P}_k$.
3) Suppose that $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1$. Then, according to (C.3), $g'(p_{k,\varphi_k}) > 0$ if $p_{k,\varphi_k} < \hat{p}_{k,\varphi_k}$ and $g'(p_{k,\varphi_k}) \leq 0$ otherwise. Also, according to (C.4), we have the following two inequalities.

$$\hat{p}_{k,\varphi_k} \leq 0 \;\Leftrightarrow\; \frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} \leq \frac{\eta_{k,\varphi_k}}{\eta_{k,\psi_k}}, \tag{C.5}$$

$$\hat{p}_{k,\varphi_k} > \bar{P}_k \;\Leftrightarrow\; \frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} > \frac{\bar{P}_k + \eta_{k,\varphi_k}}{\bar{P}_k + \eta_{k,\psi_k}}. \tag{C.6}$$

Consequently, if (C.5) is met, $g(p_{k,\varphi_k})$ is a decreasing function on $[0, \bar{P}_k]$, resulting in $p^{\star}_{k,\varphi_k} = 0$; if (C.6) is met, $g(p_{k,\varphi_k})$ is an increasing function on $[0, \bar{P}_k]$, resulting in $p^{\star}_{k,\varphi_k} = \bar{P}_k$; and if neither (C.5) nor (C.6) is met, $g(p_{k,\varphi_k})$ is an increasing function on $[0, \hat{p}_{k,\varphi_k}]$ but a decreasing function on $[\hat{p}_{k,\varphi_k}, \bar{P}_k]$, resulting in $p^{\star}_{k,\varphi_k} = \hat{p}_{k,\varphi_k}$.

Note that since $\eta_{k,\varphi_k} < \eta_{k,\psi_k}$, the right-hand side of (C.6) is always less than one, i.e., $(\bar{P}_k + \eta_{k,\varphi_k})/(\bar{P}_k + \eta_{k,\psi^k}) < 1$. Hence, the results for the three cases can be combined as

$$p^{\star}_{k,\varphi_k} = \begin{cases} 0, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \leq \eta_{k,\varphi_k}/\eta_{k,\psi_k}, \\ \bar{P}_k, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > (\bar{P}_k + \eta_{k,\varphi_k})/(\bar{P}_k + \eta_{k,\psi_k}), \\ \hat{p}_{k,\varphi_k} & \text{otherwise.} \end{cases} \tag{C.7}$$

Lastly, $p^{\star}_{k,\psi_k} = \bar{P}_k - p^{\star}_{k,\varphi_k}$ since $p_{k,\psi_k} + p_{k,\varphi_k} = \bar{P}_k$. □

Appendix D
Proof of the equivalence of (23) and (25)

To show the equivalence of (23) and (25), it is sufficient to show that

$$\frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} \leq C^2_k(\bar{P}_k) \;\Leftrightarrow\; \frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} < 1 \text{ and } \bar{P}_k \geq C^4_k. \tag{D.1}$$

First of all, the fact that the left-hand side holds if the right-hand side holds can be easily proved by simple arithmetic operations. Hence, we only prove that if the left-hand side holds, then its right-hand side holds. To this end, we first have

$$\frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} \leq C^2_k(\bar{P}_k) \overset{(a)}{\Leftrightarrow} \frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} \leq \frac{\bar{P}_k + \eta_{k,\varphi_k}}{\bar{P}_k + \eta_{k,\psi_k}} \tag{D.2}$$

$$\overset{(b)}{\Leftrightarrow} \bar{P}_k(\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}) \leq \tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}. \tag{D.3}$$

where (a) is done by the definition of $C_k^2(\bar{P}_k)$ in (19), and (b) is done by simple arithmetic operations. Then, we can deduce that if (D.2) holds, then

$$\frac{\tilde{w}_{\varphi_k}}{\tilde{w}_{\psi_k}} < 1 \text{ and } \bar{P}_k \geq C_k^4, \tag{D.4}$$

where the first inequality is derived from the fact that $C_k^2(\bar{P}_k)$ is always less than 1 since $\eta_{k,\psi_k} > \eta_{k,\varphi_k}$, and the second inequality is obtained by dividing by $\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}$ on both sides of (D.3). As a result, we can conclude that (D.1) holds.

Appendix E
Proof of Proposition 1

First, consider the case where

$$\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \geq 1. \tag{E.1}$$

Then, according to (25), we have $\phi_k^*(\bar{P}_k) = \tilde{w}_{\varphi_k} B_k \log_2(1 + \bar{P}_k/\eta_{k,\varphi_k})$, which is a logarithmic function that can be defined on $[0, \infty)$. Hence, ϕ_k^* is a continuously differentiable concave function of $\bar{P}_k$ on $[0, \infty)$ if (E.1) holds.

Next, consider the other case where

$$\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1. \tag{E.2}$$

Then, according to (25), we have

$$\phi_k^*(\bar{P}_k) = \begin{cases} \phi_{k,1}(\bar{P}_k), & \text{if } \bar{P}_k \geq C_k^4, \\ \phi_{k,2}(\bar{P}_k), & \text{otherwise,} \end{cases} \tag{E.3}$$

where

$$\phi_{k,1}(\bar{P}_k) = \tilde{w}_{\psi_k} B_k \log_2\left(1 + \frac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3, \tag{E.4}$$

$$\phi_{k,2}(\bar{P}_k) = \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{\bar{P}_k}{\eta_{k,\varphi_k}}\right). \tag{E.5}$$

Both $\phi_{k,1}$ and $\phi_{k,2}$ are logarithmic functions that can be defined on $[0, \infty)$. Hence, each of them is a continuously differentiable concave function of $\bar{P}_k$ on $[0, \infty)$. However, as shown in (E.3), ϕ_k^* is a piecewise function of $\bar{P}_k$ with a breakpoint at $\bar{P}_k = C_k^4$. Hence, to prove Proposition 1, we only need to show that $\phi_{k,1}(\bar{P}_k) = \phi_{k,2}(\bar{P}_k)$ and $\phi'_{k,1}(\bar{P}_k) = \phi'_{k,2}(\bar{P}_k)$ at the breakpoint, where $\phi'_{k,1}$ and $\phi'_{k,2}$ denote the derivative functions of $\phi_{k,1}$ and $\phi_{k,2}$, respectively, with respect to $\bar{P}_k$.

We first show that $\phi_{k,1}(\bar{P}_k) = \phi_{k,2}(\bar{P}_k)$ at $\bar{P}_k = C_k^4$. From the definitions of C_k^3 and C_k^4 given in (24) and (26), respectively, we have

$$\begin{aligned}
\phi_{k,1}(C_k^4) &= \tilde{w}_{\psi_k} B_k \log_2\left(1 + \frac{\frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}}{\eta_{k,\psi_k}}\right) \\
&\quad + \tilde{w}_{\psi_k} B_k \log_2\left(\frac{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}{\eta_{k,\varphi_k} - \eta_{k,\psi_k}} \cdot \frac{\eta_{k,\psi_k}}{\tilde{w}_{\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(\frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} \cdot \frac{\tilde{w}_{\varphi_k}}{\eta_{k,\varphi_k}}\right) \\
&= \tilde{w}_{\varphi_k} B_k \log_2\left(\frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} \cdot \frac{\tilde{w}_{\varphi_k}}{\eta_{k,\varphi_k}}\right) \\
&= \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{\frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}}{\eta_{k,\varphi_k}}\right) \\
&= \phi_{k,2}(C_k^4).
\end{aligned} \tag{E.6}$$

We now show that $\phi'_{k,1}(\bar{P}_k) = \phi'_{k,2}(\bar{P}_k)$ at $\bar{P}_k = C_k^4$. The derivative of $\phi_{k,1}$ with respect to $\bar{P}_k$ is given by

$$\phi'_{k,1}(\bar{P}_k) = \frac{\tilde{w}_{\psi_k} B_k}{\ln 2} \cdot \frac{1}{\bar{P}_k + \eta_{k,\psi_k}}. \tag{E.7}$$

Hence, we have

$$\begin{aligned}
\phi'_{k,1}(C_k^4) &= \frac{\tilde{w}_{\psi_k} B_k}{\ln 2} \cdot \frac{1}{\frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} + \eta_{k,\psi_k}} \\
&= \frac{B_k}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}
\end{aligned} \tag{E.8}$$

Next, the derivative of $\phi_{k,2}$ with respect to $\bar{P}_k$ is given by

$$\phi'_{k,2}(\bar{P}_k) = \frac{\tilde{w}_{\varphi_k} B_k}{\ln 2} \cdot \frac{1}{\bar{P}_k + \eta_{k,\varphi_k}}, \tag{E.9}$$

and thus,

$$\begin{aligned}
\phi'_{k,2}(C_k^4) &= \frac{\tilde{w}_{\varphi_k} B_k}{\ln 2} \cdot \frac{1}{\frac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} + \eta_{k,\varphi_k}} \\
&= \frac{B_k}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}{\eta_{k,\varphi_k} - \eta_{k,\psi_k}} \\
&= \phi'_{k,1}(C_k^4).
\end{aligned} \tag{E.10}$$

Hence, by (E.6) and (E.10), ϕ_k^* is a continuously differentiable concave function of $\bar{P}_k$ on $[0, \infty)$ if (E.2) holds. □

APPENDIX F
PROOF OF THEOREM 5

We first introduce two indicator variables e_k^1 and e_k^2, where e_k^1 is equal to one if $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1$ and $\bar{P}_k \geq C_k^4$ and zero otherwise, and e_k^2 is equal to one if $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \geq 1$ or $\bar{P}_k < C_k^4$ and zero otherwise. Note that since the indicators are mutually exclusive, $e_k^1 + e_k^2$ is always one. Then, $\phi_k^*(\bar{P}_k)$ in (25) can be expressed as

$$\phi_k^*(\bar{P}_k) = e_k^1 \cdot \left(\tilde{w}_{\psi_k} B_k \log_2\left(1 + \frac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3\right) + e_k^2 \cdot \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{\bar{P}_k}{\eta_{k,\varphi_k}}\right). \tag{F.1}$$

In addition, since the larger the total transmission power budget, the higher the objective value can be achieved, we can replace the first inequality constraint in Problem (M) with $\sum_{k\in\mathcal{K}} \bar{P}_k = P_{\max}$.

Now, let $\bar{\mathbf{P}}^* \in \mathbb{R}^K$ be an optimal solution to problem (M), $\underline{\boldsymbol{\mu}}^* \in \mathbb{R}^K$ and $\bar{\boldsymbol{\mu}}^* \in \mathbb{R}^K$ be Lagrange multiplier vectors for the inequality constraints $\bar{\mathbf{P}} \geq \mathbf{0}$ and $\bar{\mathbf{P}} \leq \bar{\mathbf{P}}_{\max}$, respectively, where $\mathbf{0}$ is the K-dimensional zero vector and $\bar{\mathbf{P}}_{\max} = (P_{\max,k})_{\forall k\in\mathcal{K}}$, and $\nu^* \in \mathbb{R}$ be a multiplier for the equality constraint $\sum_{k\in\mathcal{K}} \bar{P}_k = P_{\max}$. Then, the KKT conditions are obtained as

$$\frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\bar{P}_k^* + \eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\bar{P}_k^* + \eta_{k,\varphi_k}} + \underline{\mu}_k^* - \bar{\mu}_k^* - \nu^* = 0, \ \forall k \in \mathcal{K}, \tag{F.2}$$

$$\sum_{k\in\mathcal{K}} \bar{P}_k^* = P_{\max}, \ 0 \leq \bar{P}_k^* \leq P_{\max,k}, \ \forall k \in \mathcal{K}, \tag{F.3}$$

$$\underline{\mu}_k^* \geq 0, \ \underline{\mu}_k^* \bar{P}_k^* = 0, \ \forall k \in \mathcal{K}, \tag{F.4}$$

$$\bar{\mu}_k^* \geq 0, \ \bar{\mu}_k^* (\bar{P}_k^* - P_{\max,k}) = 0, \ \forall k \in \mathcal{K}. \tag{F.5}$$

We directly solve these equations to find $\bar{\mathbf{P}}^*$, $\underline{\boldsymbol{\mu}}^*$, $\bar{\boldsymbol{\mu}}^*$, and ν^*. To this end, we first eliminate $\underline{\boldsymbol{\mu}}^*$ by rearranging (F.2) for $\underline{\mu}_k^*$ and then plugging it into (F.4). Thereby, we have

$$\bar{\mu}_k^* + \nu^* - \left(\frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\bar{P}_k^* + \eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\bar{P}_k^* + \eta_{k,\varphi_k}}\right) \geq 0, \tag{F.6}$$

$$\left(\bar{\mu}_k^* + \nu^* - \left(\frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\bar{P}_k^* + \eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\bar{P}_k^* + \eta_{k,\varphi_k}}\right)\right) \times \bar{P}_k^* = 0. \tag{F.7}$$

We now find $\bar{\mathbf{P}}^*$, $\bar{\boldsymbol{\mu}}^*$, and ν^* with (F.3), (F.5), (F.6), and (F.7). First, suppose the case where

$$\bar{\mu}_k^* + \nu^* < \frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\eta_{k,\varphi_k}}. \tag{F.8}$$

Then, (F.6) can hold only if $\bar{P}_k^* > 0$. Thus, by (F.7), we have

$$\bar{\mu}_k^* + \nu^* = \frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\bar{P}_k^* + \eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\bar{P}_k^* + \eta_{k,\varphi_k}}. \tag{F.9}$$

Rearranging (F.9) for $\bar{P}_k^*$, we can conclude that if (F.8) holds,

$$\bar{P}_k^* = e_k^1 \cdot \left(\frac{\tilde{w}_{\psi_k} B_k}{(\bar{\mu}_k^* + \nu^*)\ln 2} - \eta_{k,\psi_k} \right) + e_k^2 \cdot \left(\frac{\tilde{w}_{\varphi_k} B_k}{(\bar{\mu}_k^* + \nu^*)\ln 2} - \eta_{k,\varphi_k} \right). \tag{F.10}$$

Now, suppose the opposite case where

$$\bar{\mu}_k^* + \nu^* \geq \frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\eta_{k,\varphi_k}}. \tag{F.11}$$

In this case, $\bar{P}_k^* > 0$ is impossible since it implies that

$$\begin{aligned}\bar{\mu}_k^* + \nu^* &\geq \frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\eta_{k,\varphi_k}} \\ &> \frac{e_k^1}{\ln 2} \cdot \frac{\tilde{w}_{\psi_k} B_k}{\bar{P}_k^* + \eta_{k,\psi_k}} + \frac{e_k^2}{\ln 2} \cdot \frac{\tilde{w}_{\varphi_k} B_k}{\bar{P}_k^* + \eta_{k,\varphi_k}},\end{aligned} \tag{F.12}$$

which violates (F.7). Hence, we can conclude that $\bar{P}_k^* = 0$ if (F.11) holds. Combining the results for the above two cases, we can simply express $\bar{P}_k^*$ as

$$\bar{P}_k^* = \max\left\{0,\ e_k^1 \cdot \left(\frac{\tilde{w}_{\psi_k} B_k}{(\bar{\mu}_k^* + \nu^*)\ln 2} - \eta_{k,\psi_k} \right) + e_k^2 \cdot \left(\frac{\tilde{w}_{\varphi_k} B_k}{(\bar{\mu}_k^* + \nu^*)\ln 2} - \eta_{k,\varphi_k} \right) \right\}. \tag{F.13}$$

In addition, by the complementary slackness conditions for $\bar{\boldsymbol{\mu}}^*$ in (F.5), we can see that if $\bar{P}_k^* < P_{\max,k}$, $\bar{\mu}_k^*$ should be zero; otherwise, $\bar{\mu}_k^*$ should be a certain value such that $\bar{P}_k^* = P_{\max,k}$. According to these facts, the Lagrange multiplier $\bar{\mu}_k^*$ can be eliminated in (F.13), leaving

$$\bar{P}_k^* = \left[e_k^1 \cdot \left(\mu^* \tilde{w}_{\psi_k} B_k - \eta_{k,\psi_k} \right) + e_k^2 \cdot \left(\mu^* \tilde{w}_{\varphi_k} B_k - \eta_{k,\varphi_k} \right) \right]_0^{P_{\max,k}}, \tag{F.14}$$

where $\mu^* = 1/(\nu^* \ln 2)$. For simple notation, let $f_{k,\psi_k}(\mu^*) = \mu^* \tilde{w}_{\psi_k} B_k - \eta_{k,\psi_k}$ and $f_{k,\varphi_k}(\mu^*) = \mu^* \tilde{w}_{\varphi_k} B_k - \eta_{k,\varphi_k}$. Then, by the definitions of e_k^1 and e_k^2, (F.14) can be expressed as

$$\bar{P}_k^* = [f_k(\mu^*)]_0^{P_{\max,k}}, \tag{F.15}$$

where

$$f_k(\mu^*) = \begin{cases} f_{k,\psi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } f_k(\mu^*) \geq C_k^4, \\ f_{k,\varphi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } f_k(\mu^*) < C_k^4, \\ f_{k,\varphi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \geq 1. \end{cases} \tag{F.16}$$

In the case where $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1$, $f_k(\mu^*)$ is piecewise linear with a breakpoint at $f_k(\mu^*) = C_k^4$, but continuous since both $f_{k,\psi_k}(\mu^*)$ and $f_{k,\varphi_k}(\mu^*)$ have the same value C_k^4 at $\mu^* = C_k^5$. Since both $f_{k,\psi_k}(\mu^*)$ and $f_{k,\varphi_k}(\mu^*)$ are monotonically increasing, (F.16) can be rewritten as

$$f_k(\mu^*) = \begin{cases} f_{k,\psi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \mu^* \geq C_k^5, \\ f_{k,\varphi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \mu^* < C_k^5, \\ f_{k,\varphi_k}(\mu^*), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \geq 1. \end{cases} \tag{F.17}$$

By plugging (F.17) into (F.15), we finally have (27). □

APPENDIX G
PROOF OF THEOREM 6

In order to prove the strong duality between Problem (P2) and its dual problem, Problem (D), we utilize the *time-sharing* condition proposed in [43], which is defined as follows.

Definition G.1. Let $\{\bar{\mathbf{p}}_x, \bar{\mathbf{q}}_x\}$ and $\{\bar{\mathbf{p}}_y, \bar{\mathbf{q}}_y\}$ be optimal solutions to Problem (P2) with $\bar{\mathbf{R}}_{\min} = \bar{\mathbf{R}}_x$ and $\bar{\mathbf{R}}_{\min} = \bar{\mathbf{R}}_y$, respectively, where $\bar{\mathbf{R}}_{\min} = (\bar{R}_{\min,i})_{\forall i \in \mathcal{N}}$, $\bar{\mathbf{R}}_x = (\bar{R}_{x,i})_{\forall i \in \mathcal{N}}$, and $\bar{\mathbf{R}}_y = (\bar{R}_{y,i})_{\forall i \in \mathcal{N}}$. Then, Problem (P2) is said to satisfy the time-sharing condition if for any $\bar{\mathbf{R}}_x$ and $\bar{\mathbf{R}}_y$, and for any $\theta \in [0,1]$, there always exists a feasible solution $\{\bar{\mathbf{p}}_z, \bar{\mathbf{q}}_z\}$ such that

$$\mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p}_z^{\mathbf{h}}, \mathbf{q}_z^{\mathbf{h}}; \mathbf{h})\right] \geq \theta \bar{R}_{x,i} + (1-\theta)\bar{R}_{y,i}, \ \forall i \in \mathcal{N}, \tag{G.1}$$

and

$$\mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p}_z^{\mathbf{h}}, \mathbf{q}_z^{\mathbf{h}}; \mathbf{h})\right] \geq \theta \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p}_x^{\mathbf{h}}, \mathbf{q}_x^{\mathbf{h}}; \mathbf{h})\right] + (1-\theta)\mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p}_y^{\mathbf{h}}, \mathbf{q}_y^{\mathbf{h}}; \mathbf{h})\right]. \tag{G.2}$$

It has been proven in [43] that if an optimization problem satisfies the time-sharing condition, the strong duality holds regardless of the convexity of the problem. Hence, we prove Theorem 6 by showing that Problem (P2) satisfies the time-sharing condition. First, for any $\{\bar{\mathbf{p}}_x, \bar{\mathbf{q}}_x\}$ and $\{\bar{\mathbf{p}}_y, \bar{\mathbf{q}}_y\}$, and for any $\theta \in [0,1]$, let us set $\{\mathbf{p}_z^t, \mathbf{q}_z^t\}$ to

$$\{\mathbf{p}_z^t, \mathbf{q}_z^t\} = \begin{cases} \{\mathbf{p}_x^t, \mathbf{q}_x^t\} & t \leq \lfloor \theta T \rfloor, \\ \{\mathbf{p}_y^t, \mathbf{q}_y^t\}, & t \geq \lfloor \theta T + 1 \rfloor. \end{cases} \tag{G.3}$$

Then, the first condition (G.1) holds as follows. For all $i \in \mathcal{N}$,

$$\mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p}_z^{\mathbf{h}}, \mathbf{q}_z^{\mathbf{h}}; \mathbf{h})\right] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} R_i(\mathbf{p}_z^t, \mathbf{q}_z^t; \mathbf{h}^t) \tag{G.4}$$

$$= \lim_{T\to\infty} \frac{1}{T}\left(\sum_{t=1}^{\lfloor \theta T \rfloor} R_i(\mathbf{p}_x^t, \mathbf{q}_x^t; \mathbf{h}^t) + \sum_{t=\lfloor \theta T+1 \rfloor}^{T} R_i(\mathbf{p}_y^t, \mathbf{q}_y^t; \mathbf{h}^t)\right) \tag{G.5}$$

$$= \theta \mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p}_x^{\mathbf{h}}, \mathbf{q}_x^{\mathbf{h}}; \mathbf{h})\right] + (1-\theta)\mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p}_y^{\mathbf{h}}, \mathbf{q}_y^{\mathbf{h}}; \mathbf{h})\right] \tag{G.6}$$

$$\geq \theta \bar{R}_{i,x} + (1-\theta)\bar{R}_{i,y}. \tag{G.7}$$

Similarly, the second condition (G.2) also holds as follows.

$$\mathbb{E}_{\mathbf{h}}\left[\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_z^{\mathbf{h}}, \mathbf{q}_z^{\mathbf{h}}; \mathbf{h})\right] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_z^t, \mathbf{q}_z^t; \mathbf{h}^t) \tag{G.8}$$

$$= \lim_{T\to\infty} \frac{1}{T}\left(\sum_{t=1}^{\lfloor \theta T \rfloor}\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_x^t . \mathbf{q}_x^t; \mathbf{h}^t) + \sum_{t=\lfloor \theta T+1 \rfloor}^{T}\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_y^t, \mathbf{q}_y^t; \mathbf{h}^t)\right) \tag{G.9}$$

$$= \theta \mathbb{E}_{\mathbf{h}}\left[\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_x^{\mathbf{h}}, \mathbf{q}_x^{\mathbf{h}}; \mathbf{h})\right] + (1-\theta)\mathbb{E}_{\mathbf{h}}\left[\sum_{i\in\mathcal{N}} w_i R_i(\mathbf{p}_y^{\mathbf{h}}, \mathbf{q}_y^{\mathbf{h}}; \mathbf{h})\right]. \tag{G.10}$$

Consequently, we can conclude that the time-sharing condition holds in Problem (P2), resulting in the strong duality. □

References

[1] D.-Y. Kim, H. Jafarkhani, and J.-W. Lee, "Low-complexity joint user and power scheduling for downlink NOMA over fading channels," in *Proc. IEEE VTC2021-Spring*, Virtual, Apr. 2021, pp. 1–5.

[2] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017—2022*, Cisco, San Jose, CA, USA, Feb. 2019.

[3] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart. 2018.

[4] J. Seo, Y. Sung, and H. Jafarkhani, "A high-diversity transceiver design for MISO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2591–2606, May 2019.

[5] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, no. 4, pp. 2192–2235, rth Quart. 2020.

[6] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 179–189, Jan. 2020.

[7] M. Ganji and H. Jafarkhani, "Interference mitigation using asynchronous transmission and sampling diversity," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[8] M. Ganji, X. Zou, and H. Jafarkhani, "Asynchronous transmission for multiple access channels: Rate-region analysis and system design for uplink NOMA," *IEEE Trans. Wireless Commun.*, to appear.

[9] B.-G. Kim and J.-W. Lee, "Opportunistic resource scheduling for OFDMA networks with network coding at relay stations," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 210–221, Jan. 2012.

[10] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288–296, Jan. 2009.

[11] S.-Y. Kim, J.-A. Kwon, and J.-W. Lee, "Sum-rate maximization for multicell OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4158–4169, Sep. 2014.

[12] H. Xing, Y. Liu, A. Nallanathan, Z. Ding, and H. V. Poor, "Optimal throughput fairness tradeoffs for downlink non-orthogonal multiple access over fading channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3556–3571, Jun. 2018.

[13] C.-L. Wang, J.-Y. Chen, and Y.-J. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 532–535, Oct. 2016.

[14] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[15] Z. Yang, W. Xu, C. Pan, Y. Pan, and M. Chen, "On the optimality of power allocation for NOMA downlinks with individual QoS constraints," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1649–1652, Jul. 2017.

[16] J. Wang, Q. Peng, Y. Huang, H.-M. Wang, and X. You, "Convexity of weighted sum rate maximization in NOMA systems," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1323–1327, Sep. 2017.

[17] F. Liu and M. Petrova, "Performance of proportional fair scheduling for downlink PD-NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 7027–7039, Oct. 2018.

[18] S. N. Datta and S. Kalyanasundaram, "Optimal power allocation and user selection in non-orthogonal multiple access systems," in *Proc. IEEE WCNC*, Doha, Qatar, Apr. 2016, pp. 1–6.

[19] X. Li, C. Li, and Y. Jin, "Dynamic resource allocation for transmit power minimization in OFDM-based NOMA systems," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2558–2561, Dec. 2016.

[20] F. Guo, H. Lu, D. Zhu, and Z. Gu, "Joint user association, grouping and power allocation in uplink NOMA systems with QoS constraints," in *Proc. IEEE ICC*, Shanghai, China, May 2019, pp. 1–6.

[21] Z. Yang, C. Pan, W. Xu, Y. Pan, M. Chen, and M. Elkashlan, "Power control for multi-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 927–942, Feb. 2017.

[22] Y. Fu, L. Salaün, C. W. Sung, and C. S. Chen, "Subcarrier and power allocation for the downlink of multicarrier NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11833–11847, Dec. 2018.

[23] E. C. Cejudo, H. Zhu, J. Wang, and O. Alluhaibi, "A fast algorithm for resource allocation in downlink multicarrier NOMA," in *Proc. IEEE WCNC*, Marrakesh, Morocco, Apr. 2019, pp. 1–5.

[24] S. Fu, F. Fang, L. Zhao, Z. Ding, and X. Jian, "Joint transmission scheduling and power allocation in non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8137–8150, Nov. 2019.

[25] F. Liu and M. Petrova, "Dynamic power allocation for downlink multi-carrier NOMA systems," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1930–1933, Sep. 2018.

[26] ——, "Performance of dynamic power and channel allocation for downlink MC-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1650–1662, Mar. 2020.

[27] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE PIMRC*, London, U.K., Sep. 2013, pp. 611–615.

[28] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *Proc. IEEE Globecom Workshops*, Austin, TX, USA, Dec. 2014, pp. 1026–1031.

[29] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for MC-NOMA systems," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[30] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[31] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[32] L. Salaün, M. Coupechoux, and C. S. Chen, "Joint subcarrier and power allocation in NOMA: Optimal and approximate algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 2215–2230, Mar. 2020.

[33] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[35] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods," Notes for EE364B, Stanford Univ., Stanford, CA, USA, pp. 1–36, 2007.

[36] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2014.

[39] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[40] S. Boyd and A. Mutapcic, *Stochastic subgradient methods*. Stanford, CA, USA: Stanford Univ., 2008.

[41] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, Aug. 1980.

[42] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) System Scenarios*, 3GPP TR 36.942 version 16.0.0 Release 16, Jul. 2020.

[43] 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.