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

Do-Yup Kim, *Graduate 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.

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 exacerbate 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 high spectral efficiency compared to orthogonal multiple access (OMA)-based technologies [3]–[6]. Especially, in [7], the potential performance gains of NOMA over OMA have been extensively studied in various system setups. Unlike OMA, which essentially excludes inter-user interference (IUI), NOMA is based on the premise that IUI is acceptable, and two popular categories of NOMA include power-domain NOMA and code-domain NOMA [8]. This paper focuses on the power-domain NOMA that multiplexes multiple users on the same resource based on superposition coding in the power domain, and then, at the receiver, multi-user detection is realized by mitigating IUI based on successive interference cancellation (SIC) (see [9]–[11] and references therein). However, despite the 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., [12]–[14], cannot be easily applied to the NOMA systems, and provide limited performance even if they can. In this vein, resource scheduling with low computational complexity is one of the most paramount issues in NOMA systems, and thus, many studies have been conducted. 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 [15]–[19], and power allocation and user selection [20], [21]. 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 [22]–[24]. 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 [25]. 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., [25]–[27], 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 have not been addressed therein 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 [19]), and accordingly the Joint-SAPA problem to maximize the weighted sum rate is known to be a strongly NP-hard problem [28]. 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 [29]–[36]. In [29], the power allocation for each subchannel in a two-user MC-NOMA system has been investigated. In [30], the Joint-SAPA problem in a multi-user MC-NOMA system has been investigated without considering the essential SIC capacity constraint. In [31], 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 [32], [33], heuristic Joint-SAPA algorithms using the difference-

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
[29]	✓					✓	
[30]	✓				✓	✓	✓
[31]–[35]	✓		✓		✓	✓	✓
[36]	✓	✓	✓		✓	✓	
Our work	✓	✓	✓	✓	✓	✓	✓

of-convex programming (DCP) approach have been developed under the assumption that each subchannel is occupied by up to two users. In [34], the power allocation and the subchannel assignment are performed based on the geometric programming (GP) approach and the many-to-many matching game, respectively. In [35], the authors have proposed a Joint-SAPA algorithm utilizing the Lagrangian dual and dynamic programming (DP) approaches. Most recently, in [36], 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

Joint-SAPA algorithms to maximize the weighted sum rate have been extensively studied in the literature. However, all of them are still based on approaches that typically need high computational complexity (e.g., approaches based on FTPC and ES [31], DCP [32], [33], GP and matching game [34], and DP [35], [36]). Such high computational complexity will become increasingly burdensome for practical use in future cellular networks with very short time slots.[1] Hence, Joint-SAPA algorithms with much lower computational complexity are needed to make it possible to generate transmit signals at the base station (BS) in a very short time slot.

---

[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 [37]. 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).

In addition, not all weighted sum rate maximization studies have considered explicit QoS requirements, as shown in Table I. Instead, the authors in [30]–[35] have realized proportional 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 using the Joint-SAPA-LCC algorithm, and thus, it is very effective for practical use. Through simulation, we show that our opportunistic MC-NOMA scheduling 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.

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 includes path loss, shadowing, and multipath fading. 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,[2] so that the BS can jointly perform user scheduling, subchannel assignment (i.e., user pairing per subchannel), and power allocation based on them.

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 for 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 complex Gaussian noise with variance $\sigma_{k,i}^2$, and the first, second, and third terms are the desired, interference, and noise signals, respectively. For compact notation, 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 [42]

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

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

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

---

[2]This work focuses on resource scheduling from a system-level optimization perspective. Accordingly, channel estimation is beyond the scope of this work, as in [29]–[36]. For readers interested in channel estimation, we refer to [38]–[41] and references therein.

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 simplicity, we interchangeably use R_i^t and $R_i(\mathbf{p}^t, \mathbf{q}^t; \mathbf{h}^t)$ without confusion. We now 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, $\bar{R}_{\mathrm{WSR}}$, which is what we are trying to maximize, as

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

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, $\bar{R}_{\mathrm{min},i}$, which is represented as

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

Because of SIC, in each time slot, the BS can schedule multiple users on the same subchannel. However, due to the high computational complexity and the potential error propagation in 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

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

By introducing this SIC capacity constraint with an appropriate M, we assume a perfect SIC without taking into account the error propagation of the SIC.

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 maximum power constraint of $P_{\mathrm{max},k}$ 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

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

Note that since the sum of the maximum powers over all subchannels is usually greater than the total transmission power budget of the BS in practice, we assume that $\sum_{k \in \mathcal{K}} P_{\mathrm{max},k} \geq P_{\mathrm{max}}$.

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

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

$$
\text{(P1)} \quad \underset{\mathbf{p}^t, \mathbf{q}^t, \forall t}{\text{maximize}} \quad \bar{R}_{\text{WSR}}
$$

$$
\text{subject to} \quad \bar{R}_i \geq \bar{R}_{\text{min},i}, \ \forall i \in \mathcal{N},
$$

$$
\mathbf{p}^t \in \mathcal{P}, \ \mathbf{q}^t \in \mathcal{Q}, \ \forall t.
$$

We first note that dealing with Problem (P1) is not easy due to the nonconcave objective function, the QoS constraints composed of neither convex nor concave functions, integer decision variables, and the average operation over an infinite time horizon. To resolve these challenges, we first 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. More specifically, a subproblem is the Joint-SAPA problem to maximize the instantaneous weighted sum rate without the QoS constraints in that time slot, which will be defined as Problem (P_1^t) in the next section. As a consequence, we no longer need to solve Problem (P1) directly at once, but rather solve the Joint-SAPA problem at each time slot in an online manner without considering the average operation over an infinite time horizon and the QoS constraints. 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

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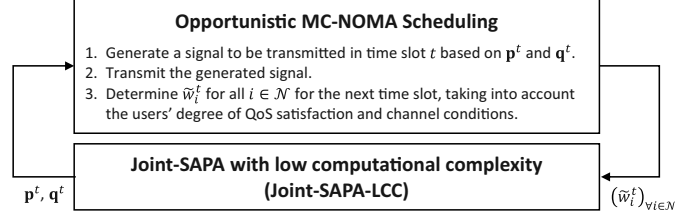$$

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

where $\tilde{w}_i^t$ is the effective weight of User i in time slot t. As mentioned before, this problem is not subject to the QoS constraints. Instead, the effective weights are systematically adjusted every time slot so that the QoS constraints in (7) are met. It is worth noting that the effective weight, $\tilde{w}_i^t$, differs from the weight, 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. This systematic update process will be rigorously elaborated later in Section IV. Although the QoS constraints are not considered in Problem (P_1^t), it is still an NP-hard problem and very difficult to solve using conventional methods since it contains not only a nonconcave objective function but also integer variables. Furthermore, we need a fast solution because 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 [43], [44]. 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

$$(\text{P}_2) \quad \underset{\mathbf{p},\, \mathbf{q},\, \bar{\mathbf{P}}}{\text{maximize}} \quad \sum_{i \in \mathcal{N}} \tilde{w}_i \sum_{k \in \mathcal{K}} R_{k,i}(\mathbf{p}_k, \mathbf{q}_k; \mathbf{h}_k)$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}} \bar{P}_k \leq P_{\max},$$

$$0 \leq \bar{P}_k \leq P_{\max,k}, \quad \forall k \in \mathcal{K},$$

$$\sum_{i \in \mathcal{N}} p_{k,i} \leq \bar{P}_k, \quad \forall k \in \mathcal{K},$$

$$p_{k,i} \geq 0, \quad \forall k \in \mathcal{K}, \ \forall i \in \mathcal{N},$$

$$\sum_{i \in \mathcal{N}} q_{k,i} \leq M, \quad \forall k \in \mathcal{K},$$

$$q_{k,i} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \ \forall i \in \mathcal{N}.$$

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

$$(S_k) \quad \underset{\mathbf{p}_k, \mathbf{q}_k}{\text{maximize}} \quad \sum_{i \in \mathcal{N}} \tilde{w}_i R_{k,i}(\mathbf{p}_k, \mathbf{q}_k; \mathbf{h}_k)$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}} p_{k,i} \leq \bar{P}_k,$$

$$p_{k,i} \geq 0, \ \forall i \in \mathcal{N},$$

$$\sum_{i \in \mathcal{N}} q_{k,i} \leq M,$$

$$q_{k,i} \in \{0, 1\}, \ \forall i \in \mathcal{N}.$$

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

$$(M) \quad \underset{\bar{\mathbf{P}}}{\text{maximize}} \quad \phi^*(\bar{\mathbf{P}}) = \sum_{k \in \mathcal{K}} \phi_k^*(\bar{P}_k)$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}} \bar{P}_k \leq P_{\max},$$

$$0 \leq \bar{P}_k \leq P_{\max,k}, \ \forall k \in \mathcal{K},$$

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. Due to the page limit, we leave the proof in Appendix A of the arXiv version [45]. □

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

---

Algorithm 1: Joint-SAPA-LCC

---

1 Initialize: $\bar{P}_k \leftarrow P_{\max}/K, \forall k \in \mathcal{K}$.

2 repeat

3   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.

4   Obtain the solution, $\bar{\mathbf{P}}^*$, to Problem (M) for the given $\phi^*(\bar{\mathbf{P}})$ using Algorithm 3.

5   Update $\bar{\mathbf{P}} \leftarrow \bar{\mathbf{P}}^*$

6 until *convergence*

7 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.

8 return $\{\mathbf{p}^*, \mathbf{q}^*\}$.

---

complexity.[3] 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.

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 mainly due to the integer variables, i.e., the subchannel assignment indicators, $\mathbf{q}_k$. To address this difficulty, we first consider the following problem:

$$(Q_k) \quad \underset{\mathbf{p}_k}{\text{maximize}} \quad \sum_{i \in \mathcal{N}} \tilde{w}_i R_{k,i}(\mathbf{p}_k; \mathbf{h}_k)$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}} p_{k,i} \leq \bar{P}_k,$$

$$p_{k,i} \geq 0, \ \forall i \in \mathcal{N},$$

where

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

---

[3]In [36], it is demonstrated that the proposed Joint-SAPA algorithm therein (called Joint-SAPA-DP later in simulation results) achieves near-optimal performance. Accordingly, by comparing our Joint-SAPA-LCC algorithm with the Joint-SAPA-DP algorithm via simulation, we verify that our algorithm provides near-optimal performance with very low computational complexity.

Although Problem (Q_k) is different from Problem (S_k), we can easily derive the optimal solution to Problem (S_k) from that to Problem (Q_k) if a certain condition is met, by 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*

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

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. Due to the page limit, we leave the proof in Appendix B of the arXiv version [45]. □

Note that in the remaining of this subsection, we focus on developing a low-complexity heuristic algorithm that solves Problem (Q_k) without considering (11). However, it will be revealed later that the solution to Problem (Q_k) obtained by our proposed algorithm always satisfies (11) as long as $M > 1$. Hence, a solution to Problem (S_k) can be easily obtained from the solution to Problem (Q_k) based on Theorem 2.

We now discuss how to solve Problem (Q_k). Even though Problem (Q_k) does not have any integer variables, it is still known to be strongly NP-hard [28]. 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

$$
(Q_{k,1}^{\varphi_k}) \quad \underset{p_{k,i},\, i<\varphi_k}{\text{maximize}} \quad \sum_{i=1}^{\varphi_k-1} \tilde{w}_i B_k \log_2\!\left(1 + \frac{p_{k,i}}{\sum_{j>i} p_{k,j} + \eta_{k,i}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\!\left(1 + \frac{p_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right)
$$

$$
\text{subject to} \quad \sum_{i=1}^{\varphi_k-1} p_{k,i} + p_{k,\varphi_k} \le \bar{P}_k,
$$

$$
p_{k,i} \ge 0, \ i < \varphi_k.
$$

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. Note that, in general, enough power is usually allocated to the last SIC user's signal for its successful decoding. Accordingly, the amount of interference power experienced by users other than the last SIC user will usually be non-negligible and exceed the thermal noise power. In other words, we assume that the mode of operation is interference limited, so that the noise power of the users suffering from the interference signals is neglected, i.e., $\sigma_{k,i}^2 = 0$ (accordingly, $\eta_{k,i} = 0$) for $i < \varphi_k$. Then, letting $\rho_{k,i} = \sum_{j=i}^{\varphi_k} p_{k,j}$ for $i \le \varphi_k$, we can approximate Problem $(Q_{k,1}^{\varphi_k})$ as

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

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

$$
\frac{\rho_{k,i}}{\rho_{k,i+1}} \ge 1, \ i < \varphi_k,
$$

where the inequality constraints are equivalent to those in Problem $(Q_{k,1}^{\varphi_k})$, 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 $(Q_{k,2}^{\varphi_k})$ equivalently as

$$
(Q_{k,3}^{\varphi_k}) \quad \underset{r_{k,i},\, i<\varphi_k}{\text{maximize}} \quad \sum_{i=1}^{\varphi_k-1} \tilde{w}_i B_k r_{k,i} + \tilde{w}_{\varphi_k} B_k \log_2\!\left(1 + \frac{2^{r_{k,\varphi_k}}}{\eta_{k,\varphi_k}}\right)
$$

$$
\text{subject to} \quad \sum_{i=1}^{\varphi_k-1} r_{k,i} \le \log_2(\bar{P}_k) - r_{k,\varphi_k},
$$

$$
r_{k,i} \ge 0, \ i < \varphi_k.
$$

Note that 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 is neglected, on each Subchannel k, at most two users are allocated power according to the optimal solution to Problem ($Q_{k,1}^{\varphi_k}$). 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*

$$\psi_k(\varphi_k) = \underset{i < \varphi_k}{\operatorname{argmax}}\{\tilde{w}_i\}. \tag{13}$$

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

$$(Q_{k,4}^{\varphi_k}) \quad \underset{p_{k,\psi_k}, p_{k,\varphi_k}}{\operatorname{maximize}} \quad \tilde{w}_{\psi_k} B_k \log_2\left(1 + \frac{p_{k,\psi_k}}{p_{k,\varphi_k} + \eta_{k,\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \frac{p_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right)$$

$$\text{subject to} \quad p_{k,\psi_k} + p_{k,\varphi_k} \le \bar{P}_k,$$

$$p_{k,\psi_k} \ge 0, \ p_{k,\varphi_k} \ge 0,$$

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 $\{p_{k,\psi_k}^\star, p_{k,\varphi_k}^\star\}$ be the optimal solution to Problem ($Q_{k,4}^{\varphi_k}$). It can be obtained as*

$$p_{k,\varphi_k}^\star = \begin{cases} 0, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \le C_k^1, \\ \bar{P}_k, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > C_k^2(\bar{P}_k), \\ \dfrac{\tilde{w}_{\psi_k}\eta_{k,\varphi_k} - \tilde{w}_{\varphi_k}\eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}, & \text{otherwise,} \end{cases} \tag{14}$$

$$p_{k,\psi_k}^\star = \bar{P}_k - p_{k,\varphi_k}^\star, \tag{15}$$

where

$$C_k^1 = \frac{\eta_{k,\varphi_k}}{\eta_{k,\psi_k}} \quad \text{and} \quad C_k^2(\bar{P}_k) = \frac{\bar{P}_k + \eta_{k,\varphi_k}}{\bar{P}_k + \eta_{k,\psi_k}}. \tag{16}$$

Proof. Due to the page limit, we leave the proof in Appendix C of the arXiv version [45]. □

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 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

$$
\phi_k(\bar{P}_k;\varphi_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{p^{\star}_{k,\psi_k}}{p^{\star}_{k,\varphi_k} + \eta_{k,\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{p^{\star}_{k,\varphi_k}}{\eta_{k,\varphi_k}}\right), & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} > C_k^1, \\[6pt] -\infty, & \text{otherwise,} \end{cases}
$$
(17)

where $p^{\star}_{k,\varphi_k}$ and $p^{\star}_{k,\psi_k}$ are given by (14) and (15), respectively. Then, the optimal last SIC user on Subchannel k, indexed by φ_k^*, can be determined as

$$
\varphi_k^* = \underset{\varphi_k \in \mathcal{N}}{\operatorname{argmax}}\, \phi_k(\bar{P}_k;\varphi_k),
$$
(18)

and the corresponding optimal value is given by

$$
\phi_k^*(\bar{P}_k) = \phi_k(\bar{P}_k;\varphi_k^*).
$$
(19)

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 $\mathbf{p}_k^{\dagger} = (p_{k,i}^{\dagger})_{\forall i \in \mathcal{N}}$, where $p_{k,\varphi_k^*}^{\dagger} = p^{\star}_{k,\varphi_k^*}$, $p_{k,\psi_k(\varphi_k^*)}^{\dagger} = p^{\star}_{k,\psi_k(\varphi_k^*)}$, and $p_{k,i}^{\dagger} = 0$ for all $i \in \mathcal{N} \setminus \{\varphi_k^*, \psi_k(\varphi_k^*)\}$. Finally, the solution, $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$, to Problem (S_k) can be derived from $\mathbf{p}_k^{\dagger}$ using Theorem 2. We summarize this process in Algorithm 2.

Before moving on to the next subsection, we analyze the computational complexity of Algorithm 2. First, once any one user is selected as the last SIC user, the computational complexity to find the other *candidate* user based on (13) is $O(N)$. Then, thanks to the closed-form power allocation formulas in Theorem 4, the two *candidate* users' power allocation and the corresponding weighted sum rate can be calculated in $O(1)$. Consequently, since the number of cases in which any one user is selected as the last SIC user is at most N, the overall computational complexity of Algorithm 2 is $O(N^2)$.

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

---

Algorithm 2: Solution to Problem (S_k)

---

1 **for** *each User $\varphi_k \in \mathcal{N}$* **do**

2 | Suppose that User φ_k is the last SIC user.

3 | Select the other *candidate* user using (13).

4 | Obtain their power allocation using Theorem 4.

5 | Obtain $\phi_k(\bar{P}_k; \varphi_k)$ using (17).

6 Obtain φ_k^* and $\phi_k^*(\bar{P}_k)$ using (18) and (19), respectively.

7 Let Users φ_k^* and $\psi_k(\varphi_k^*)$ be the optimal *candidate* users.

8 Obtain $\mathbf{p}_k^{\dagger}$ using Theorem 4.

9 Derive $\{\mathbf{p}_k^*, \mathbf{q}_k^*\}$ from $\mathbf{p}_k^{\dagger}$ using Theorem 2.

10 **return** $\{\mathbf{p}_k^*, \mathbf{q}_k^*, \phi_k^*(\bar{P}_k)\}$.

---

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 (19), we can obtain $\phi_k^*(\bar{P}_k)$ as

$$
\phi_k^*(\bar{P}_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \leq C_k^2(\bar{P}_k), \\ \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\varphi_k}}\right), & \text{otherwise,} \end{cases}
\tag{20}
$$

where $C_k^2(\bar{P}_k)$ is defined in (16), and

$$
C_k^3 = \tilde{w}_{\psi_k} B_k \log_2\left(\frac{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}{\eta_{k,\varphi_k} - \eta_{k,\psi_k}} \cdot \frac{\eta_{k,\psi_k}}{\tilde{w}_{\psi_k}}\right) + \tilde{w}_{\varphi_k} B_k \log_2\left(\frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}} \cdot \frac{\tilde{w}_{\varphi_k}}{\eta_{k,\varphi_k}}\right).
\tag{21}
$$

Note that, in (20), the case where $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \leq C_k^1$ 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 (20) is difficult due to $C_k^2(\bar{P}_k)$ that varies with $\bar{P}_k$, we reformulate it equivalently as

$$
\phi_k^*(\bar{P}_k) = \begin{cases} \tilde{w}_{\psi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\psi_k}}\right) + C_k^3, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \bar{P}_k \geq C_k^4, \\ \tilde{w}_{\varphi_k} B_k \log_2\left(1 + \dfrac{\bar{P}_k}{\eta_{k,\varphi_k}}\right), & \text{otherwise,} \end{cases}
\tag{22}
$$

where

$$
C_k^4 = \frac{\tilde{w}_{\psi_k} \eta_{k,\varphi_k} - \tilde{w}_{\varphi_k} \eta_{k,\psi_k}}{\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k}}.
\tag{23}
$$

The proof of the equivalence of (20) and (22) is provided in Appendix D of the arXiv version [45]. From (22), we can see that if $\tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} \geq 1$, $\phi_k^*(\bar{P}_k)$ 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 (22) is a continuously differentiable concave function of $\bar{P}_k$ on $[0, \infty)$.*

Proof. Due to the page limit, we leave the proof in Appendix E of the arXiv version [45]. □

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 [46].

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 $k \in \mathcal{K}$,*

$$
\bar{P}_k^* = \begin{cases} \left[\mu^* \tilde{w}_{\psi_k} B_k - \eta_{k,\psi_k}\right]_0^{P_{\max,k}}, & \text{if } \tilde{w}_{\varphi_k}/\tilde{w}_{\psi_k} < 1 \text{ and } \mu^* > C_k^5, \\ \left[\mu^* \tilde{w}_{\varphi_k} B_k - \eta_{k,\varphi_k}\right]_0^{P_{\max,k}}, & \text{otherwise,} \end{cases}
\tag{24}
$$

where

$$
C_k^5 = \frac{\eta_{k,\varphi_k} - \eta_{k,\psi_k}}{B_k(\tilde{w}_{\varphi_k} - \tilde{w}_{\psi_k})},
\tag{25}
$$

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

Proof. Due to the page limit, we leave the proof in Appendix F of the arXiv version [45]. □

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 (24). 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.

We now analyze the computational complexity of Algorithm 3. In each iteration, the computation of $f(\mu_{\text{new}})$ dominates the others, and its computational complexity is $\mathcal{O}(K)$ due to the sum over $\mathcal{K}$. Thus, letting κ be the number of iteration until convergence, the overall computational

---

Algorithm 3: Solution to Problem (M)

---

1 Let $f(\mu) = \sum_{k \in \mathcal{K}} \bar{P}_k^*(\mu) - P_{\max}$, where $\bar{P}_k^*(\mu)$ is obtained by (24) with $\mu^* = \mu$.

2 Initialize μ_L to zero, and μ_U to a sufficiently large value.

3 **repeat**

4 $\mu_{\text{new}} \leftarrow (\mu_L + \mu_U)/2$.

5   **if** $f(\mu_L) \cdot f(\mu_{\text{new}}) < 0$ **then** $\mu_U \leftarrow \mu_{\text{new}}$.

6   **if** $f(\mu_U) \cdot f(\mu_{\text{new}}) < 0$ **then** $\mu_L \leftarrow \mu_{\text{new}}$.

7 **until** *convergence*

8 Obtain $\bar{\mathbf{P}}^*$ by plugging $\mu^* = (\mu_L + \mu_U)/2$ into (24).

9 **return** $\bar{\mathbf{P}}^*$.

---

complexity of Algorithm 3 can be given by $O(\kappa K)$. Note that the exact value of κ cannot be derived rigorously, but is usually considered to have an order of $\log K$ [47], [48].

Consequently, the total computational complexity of Algorithm 1, running Algorithm 2 K times and Algorithm 3 once in each iteration, can be expressed as $O(\xi K(N^2 + \kappa))$, where ξ is the number of iterations until convergence.

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

$$\tilde{w}_i^t = \frac{1}{R_{\text{EMA},i}^t}, \tag{26}$$

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

$$R_{\text{EMA},i}^{t+1} = \left(1 - \frac{1}{\tau}\right) R_{\text{EMA},i}^t + \frac{1}{\tau} R_i^t, \tag{27}$$

where τ is the time-averaging window coefficient.

This remark shows that our Joint-SAPA-LCC algorithm can be easily extended to the proportional fair scheduling algorithm. However, it is worth noting that the purpose of the proportional fair scheduling is to reduce the performance differences among all users, so it cannot guarantee the QoS constraints in (7). Hence, in the next section, we develop a scheduling algorithm that guarantees given QoS constraints by utilizing the Joint-SAPA-LCC algorithm.

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_1^t) for each time slot. To this end, we first take advantage of the well-known property that if the fading process is stationary and ergodic, the long-term time average converges almost surely to the expectation for almost all realizations of the fading process [42], [49]. Thereby, by denoting a random vector representing the channel vector in a generic time slot by $\mathbf{h}$ and replacing the superscript t of the decision variables with $\mathbf{h}$, we can reformulate Problem (P1) equivalently as

$$(P2) \quad \underset{\mathbf{p^h}, \, \mathbf{q^h}, \, \forall \mathbf{h}}{\text{maximize}} \quad \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} w_i R_i(\mathbf{p^h}, \mathbf{q^h}; \mathbf{h}) \right]$$

$$\text{subject to} \quad \mathbb{E}_{\mathbf{h}}\left[R_i(\mathbf{p^h}, \mathbf{q^h}; \mathbf{h}) \right] \geq \bar{R}_{\min,i}, \ \forall i \in \mathcal{N},$$

$$\mathbf{p^h} \in \mathcal{P}, \ \mathbf{q^h} \in Q, \ \forall \mathbf{h}.$$

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 $\mathbf{p^h}$ and $\mathbf{q^h}$ 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

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

$$= \mathbb{E}_{\mathbf{h}}\left[\sum_{i \in \mathcal{N}} (w_i + \lambda_i) R_i(\mathbf{p^h}, \mathbf{q^h}; \mathbf{h}) \right] - \sum_{i \in \mathcal{N}} \lambda_i \bar{R}_{\min,i}, \tag{28}$$

where $\bar{\mathbf{p}} = (\mathbf{p^h})_{\forall \mathbf{h}}$, $\bar{\mathbf{q}} = (\mathbf{q^h})_{\forall \mathbf{h}}$, 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

$$F(\boldsymbol{\lambda}) = \underset{\mathbf{p^h} \in \mathcal{P}, \, \mathbf{q^h} \in Q, \, \forall \mathbf{h}}{\text{maximize}} \quad L(\bar{\mathbf{p}}, \bar{\mathbf{q}}, \boldsymbol{\lambda}). \tag{29}$$

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. Due to the page limit, we leave the proof in Appendix G of the arXiv version [45]. □

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 (28) 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 (29) can be solved by separately solving the subproblem for each channel realization, defined by

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

Since the expectation has disappeared in Problem (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 (D$^\mathbf{h}$) turns into a deterministic optimization problem for Joint-SAPA that aims to maximize the instantaneous weighted sum rate with weight $w_i + \lambda_i$ 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 (D$^\mathbf{h}$) can be solved for given $\mathbf{h}$ and $\boldsymbol{\lambda}$, the underlying distributions of the fading process are still fundamentally needed to solve Problem (D). Nevertheless, thanks to the fact that Problem (D) is a form of convex stochastic optimization problems, we can solve it without resorting on the distributions using

---

Algorithm 4: Opportunistic MC-NOMA Scheduling

---

1   Initialize: $t = 1$ and $\boldsymbol{\lambda}^t = \mathbf{0}$.

2   **for** *each time slot t* **do**

3   Solve Problem (D$^\mathbf{h}$) with $\mathbf{h} = \mathbf{h}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$ using the Joint-SAPA-LCC algorithm (Algorithm 1).

4   Transmit the signal generated by the obtained solution.

5   Calculate $\boldsymbol{\lambda}^{t+1}$ according to (30) and (31).

6       $t \leftarrow t + 1$.

---

the stochastic subgradient method [50], where the Lagrange multiplier vector, $\boldsymbol{\lambda}$, is iteratively updated by

$$\boldsymbol{\lambda}^{t+1} = \left[\boldsymbol{\lambda}^t - \zeta^t \mathbf{v}^t \right]^+, \tag{30}$$

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 [51], the stochastic subgradient, $\mathbf{v}^t$, can be obtained by

$$v_i^t = R_i^t - \bar{R}_{\min,i}, \ \ \forall i \in \mathcal{N}, \tag{31}$$

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 (D$^\mathbf{h}$) with $\mathbf{h} = \mathbf{h}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$. With the update process of (30), 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 [52], i.e.,

$$\zeta^t \geq 0, \ \ \sum_{t=1}^{\infty} \zeta^t = \infty, \ \text{ and } \ \sum_{t=1}^{\infty} (\zeta^t)^2 < \infty. \tag{32}$$

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

It is worth noting that, due to the stationary ergodic fading process, Problem (P1) and Problem (P2) are equivalent with probability one, and according to Theorem 6, there is no duality gap between Problem (P2) and Problem (D). Therefore, there is no loss of optimality in solving Problem (P1) by our algorithms as long as Problem (D$^\mathbf{h}$), i.e., Problem (P$_1^t$), is optimally solved. Accordingly, we can expect that our opportunistic MC-NOMA scheduling provides near-optimal performance by showing via simulation that the Joint-SAPA-LCC algorithm provides near-optimal performance.

V. SIMULATION RESULTS

Now, we present simulation results to evaluate the performance of our proposed algorithms. We first investigate the Joint-SAPA-LCC algorithm in Section V-A and then the opportunistic MC-NOMA scheduling algorithm in Section V-B. 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 $\gamma P_{\text{max}}/K$ with $\gamma = 1.15$. The large-scale path loss is modeled by the HATA model for urban environments [53], [54]. Specifically, the path loss in dB over distance d_{km} in kilometers is set to $69.55 + 26.16 \log_{10}(f_c) - a(h_m, h_b) + b(h_b) \log_{10}(d_{km})$, where $a(h_m, h_b) = 13.82 \log_{10}(h_b) + 3.2[\log_{10}(11.75 h_m)]^2 - 4.97$, $b(h_b) = 44.9 - 6.55 \log_{10}(h_b)$, f_c is the carrier frequency, and h_m and h_b are effective antenna heights of the BS and users, respectively. The parameters are set as follows: $f_c = 900$ MHz, $h_a = 30$ m, and $h_b = 2$ m. Also, we set the antenna gains of the BS and the users to 15 dBi and 0 dBi, respectively. Then, we consider the shadow fading with a standard deviation of 8 dB and the Rayleigh small-scale fading with unit variance. The noise power spectral density, N_0, is set to -174 dBm/Hz. Thus, the noise power of User i on Subchannel k is given by $\sigma_{k,i}^2 = B_{\text{tot}} N_0/K$. In the following simulation results, the data rates are considered in units of bps/Hz. In Algorithm 4, the step size in (30) is set to $\zeta^t = 1/t$, which satisfies the conditions in (32) 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 [31], where Joint-SAPA is performed based on the FTPC and ES algorithms. The second one is the Joint-SAPA-DCP algorithm [32], where Joint-SAPA is performed based on the DCP approach. The last one is the Joint-SAPA-DP algorithm [36], 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. Note that the Joint-SAPA algorithms (including ours) do not deal with any QoS constraints. The purpose of this subsection is to verify that our Joint-SAPA-LCC algorithm provides good enough performance despite requiring very low computational complexity compared to the baseline Joint-SAPA algorithms.

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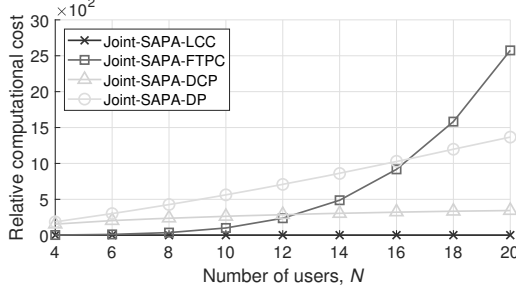
Fig. 2. The relative computational cost versus the number of users.
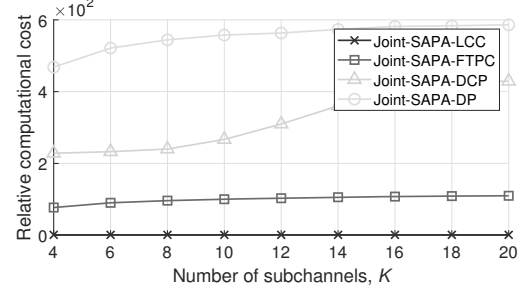

Fig. 3. The relative computational cost versus the number of subchannels.
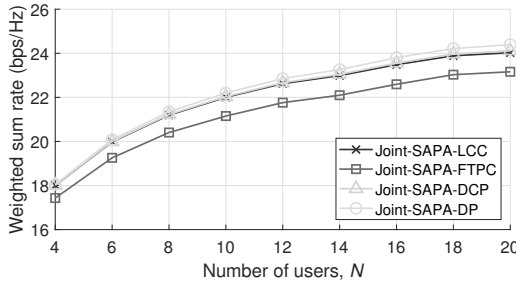

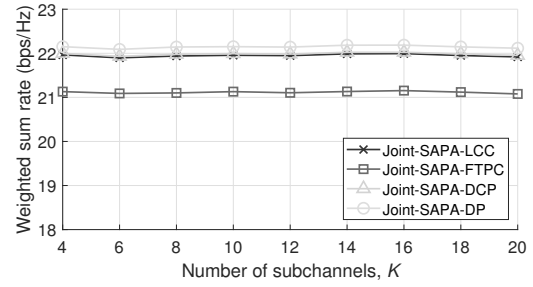Fig. 4. The weighted sum rate versus the number of users.

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

SAPA-FTPC algorithm. On the other hand, Fig. 5 shows that the weighted sum rates for all the Joint-SAPA algorithms tend to remain constant regardless of the number of subchannels. These results indicate that, as also seen in [55], [56], the effects of the number of subchannels on the weighted sum rate performance are negligible in the system where the transcievers for different subchannels operate independently. Meanwhile, 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% lower than those of the Joint-SAPA-DP and Joint-SAPA-DCP algorithms, respectively, but 4% higher than that of the Joint-SAPA-FTPC algorithm.

Next, since the Joint-SAPA-DP algorithm provides the highest weighted sum rate performance, we compare our Joint-SAPA-LCC algorithm with it in more depth for the case where $N = 10$ and $K = 10$ in Fig. 6. Fig. 6a shows the frequency histogram of the number of assigned users per subchannel. As can be seen in the figure, at most five users are assigned per subchannel in the Joint-SAPA-DP algorithm, unlike our algorithm with at most two users per subchannel. As a
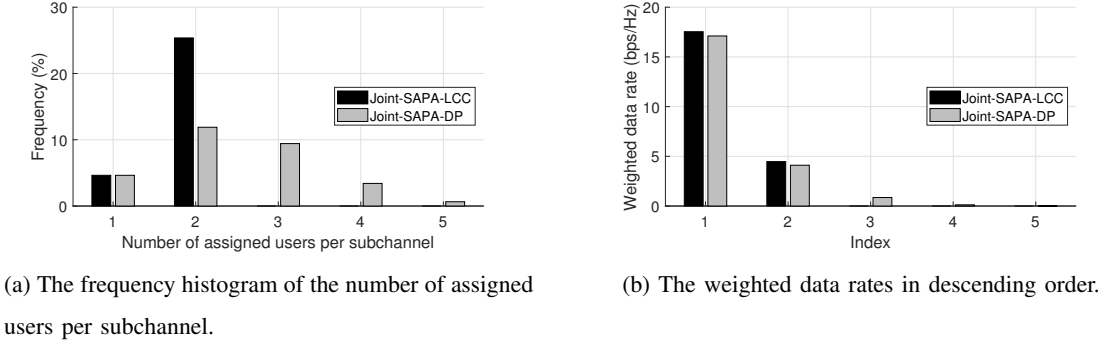
(a) The frequency histogram of the number of assigned users per subchannel.

(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 $K = 10$.

result, the user assignment patterns of the two algorithms are different, and the Joint-SAPA-DP algorithm has a larger solution space for user assignment patterns compared to our algorithm. Accordingly, the Joint-SAPA-DP algorithm has the possibility to provide higher performance than ours. On the other hand, Fig. 6b shows the achieved weighted data rates of individual assigned users in a subchannel. In this figure, the value of the bar for Index i represents the average weighted data rate of the user that achieves the ith highest data rate among the assigned users in a subchannel. We can see that more than 95 % of the weighted sum rate performance is assigned to the top two indices in the Joint-SAPA-DP algorithm, which implies that the remaining bottom three users have a very little impact on the weighted sum rate performance. Since most of the weighted sum rates correspond to the first two indices, whether in our Joint-SAPA-LCC algorithm or in the Joint-SAPA-DP algorithm, the weighted sum rate performance of our Joint-SAPA-LCC algorithm is very close to that of the Joint-SAPA-DP algorithm despite the different user assignment patterns. In summary, 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.

Lastly, we further investigate the effects of the imperfection of instantaneous channel state information (CSI) on the weighted sum rate performance since perfect knowledge of CSI at the BS is practically impossible. To this end, as in [57], [58], we first model an estimated channel gain corresponding to $h_{k,i}$ as $\hat{h}_{k,i} = h_{k,i} + e_{k,i}$, where $e_{k,i}$ is the channel estimation error generated by the complex Gaussian distribution with zero mean and variance of σ_e^2/PL_i, and PL_i is the path loss of User i. Fig. 7 shows the weighted sum rate performance of our Joint-SAPA-LCC
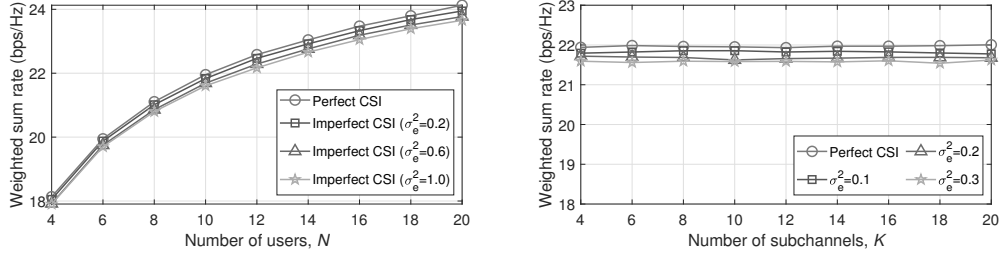
(a) The weighted data rates with varying N and σ_e^2. (b) The weighted data rates with varying K and σ_e^2.

Fig. 7. The weighted sum rates of the Joint-SAPA-LCC algorithm under the imperfect CSI environments.

algorithm in the imperfect CSI environments. The parameter settings for Figs. 7a and 7b are the same as for Figs. 4 and 5, respectively, only except that the imperfect CSI environments are applied. As expected, the weighted sum rates are slightly degraded as the channel estimation error variance increases because the imperfect CSI distorts subchannel assignment (i.e., user pairing per subchannel) as well as power allocation. Nevertheless, the performance degradation is not much, which means that the Joint-SAPA-LCC algorithm can tolerate some degree of channel estimation errors.

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 (26). The time-averaging window coefficient, τ, in (27) 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. 8 and 9 show the performance results for the first and second scenarios, respectively. In Figs. 8a and 9a, 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. 8b and 9b, 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, the proportional fair scheduling follows the principle of giving high effective weights to users with low time-averaged data rates. Accordingly, not only does it prevent users from starvation, but it 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. 9b 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
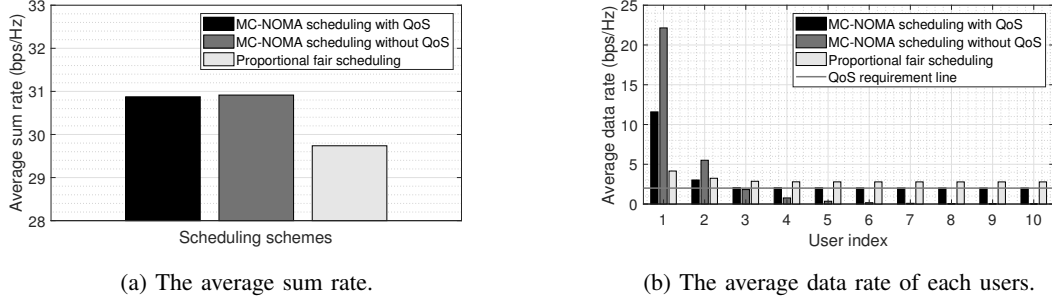
(a) The average sum rate.

(b) The average data rate of each users.

Fig. 8. Performance comparison results between scheduling algorithms when all users have the same QoS requirements.

(a) The average sum rate.
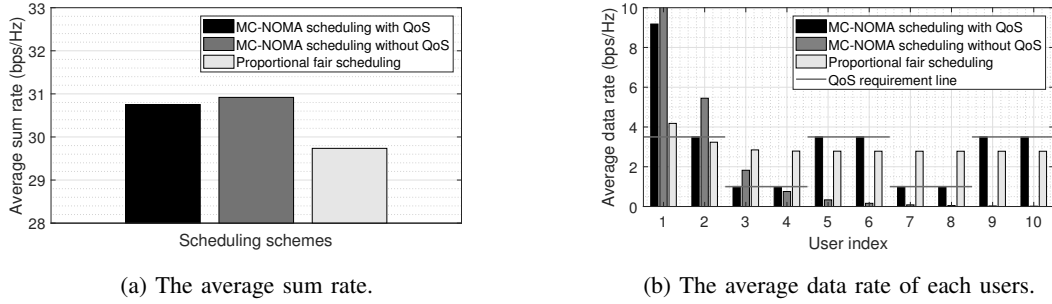
(b) The average data rate of each users.

Fig. 9. Performance comparison results between scheduling algorithms when users have individually different QoS requirements.

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. As a final remark, the issue of SIC error propagation has not been discussed in this paper. We leave this issue for our future work. This study will be the cornerstone for our future work on the development of scheduling for more complex systems, such as multi-cell MC-NOMA systems and massive MIMO MC-NOMA systems.

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] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 536–568, Jan. 2020.

[8] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[9] J. Kazemitabar and H. Jafarkhani, "Performance analysis of multiple antenna multi-user detection," in *Proc. ITA Workshop*, La Jolla, CA, USA, Feb. 2009, pp. 150–159.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] L. Salaün, C. S. Chen, and M. Coupechoux, "Optimal joint subcarrier and power allocation in NOMA is strongly NP-hard," in *IEEE ICC*, Kansas City, MO, USA, Jul. 2018, pp. 1–7.

[29] 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.

[30] ——, "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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1891–1908, 4th Quart. 2014.

[39] R. Mohammadian, A. Amini, and B. H. Khalaj, "Deterministic pilot design for sparse channel estimation in MISO/multi-user OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 129–140, Jan. 2016.

[40] P. Sure and C. M. Bhuma, "A survey on OFDM channel estimation techniques based on denoising strategies," *Eng. Sci. Technol., Int. J.*, vol. 20, no. 2, pp. 629–636, Apr. 2017.

[41] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

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

[43] 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.

[44] 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.

[45] D.-Y. Kim, H. Jafarkhani, and J.-W. Lee, "Low-complexity dynamic resource scheduling for downlink MC-NOMA over fading channels," *arXiv preprint arXiv:2103.10650*, Mar. 2021.

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

[47] P. He, L. Zhao, S. Zhou, and Z. Niu, "Water-filling: A geometric approach and its application to solve generalized radio resource allocation problems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3637–3647, Jul. 2013.

[48] S. Khakurel, C. Leung, and T. Le-Ngoc, "A generalized water-filling algorithm with linear complexity and finite convergence time," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 225–228, Apr. 2014.

[49] P. Walters, *An Introduction to Ergodic Theory*. New York: Springer-Verlag, 1982.

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

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

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

[53] 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.

[54] *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.

[55] J. Jang and K. B. Jang, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[56] B.-G. Kim and J.-W. Lee, "Joint opportunistic subchannel and power scheduling for relay-based OFDMA networks with scheduling at relay stations," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2138–2148, Jun. 2010.

[57] Z. Wei, D. W. K. Ng, J. Yuan, and H.-M. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.

[58] Y.-X. Zhu, D.-Y. Kim, and J.-W. Lee, "Joint antenna and user scheduling in the massive MIMO system over time-varying fading channels," *IEEE Access*, vol. 9, pp. 92 431–92 445, Jun. 2021.