IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 17, 2021, accepted June 17, 2021, date of publication June 28, 2021, date of current version July 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3092754

Joint Antenna and User Scheduling in the Massive MIMO System Over Time-Varying Fading Channels

YUAN-XIN ZHU[1],*, DO-YUP KIM[2],*, (Graduate Student Member, IEEE), AND JANG-WON LEE[2], (Senior Member, IEEE)
[1]AI Cloud Group, Quality Assurance Department, Baidu Inc., Beijing 100193, China
[2]Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Jang-Won Lee (jangwon@yonsei.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant through the Korea Government (MSIT) under Grant 2019R1A2C2084870, and in part by the Unmanned Vehicles Core Technology Research and Development Program through the NRF and Unmanned Vehicle Advanced Research Center (UVARC) through the Ministry of Science and Information Communication Technologies (ICT), Republic of Korea, under Grant NRF-2020M3C1C1A02086427.

This work did not involve human subjects or animals in its research.

*Yuan-Xin Zhu and Do-Yup Kim contributed equally to this work.

ABSTRACT Massive multiple-input multiple-output (MIMO) technology, mainly equipped with dozens or even hundreds of antennas at transmitter and/or receiver, is one of the most important technologies in 5G/6G era due to its capability of achieving high transmission rate. However, since each transmit antenna typically needs a complete radio frequency (RF) chain, a large number of RF chains need to be installed accordingly, resulting in high economic cost, high hardware complexity, and high power consumption. To resolve these problems, architectures where a small number of RF chains are installed have been proposed in recent years, and an antenna selection technique that activates antennas only as many as the number of RF chains has been envisioned as one of solutions. In this paper, we study the joint antenna and user scheduling problem for the downlink massive MIMO system over time-varying fading channels to maximize the weighted average sum rate while ensuring users' minimum average data rate requirements. To solve the problem, we first develop an opportunistic joint antenna and user scheduling algorithm (OJAUS) using the dual and stochastic subgradient methods, which makes it possible to schedule antennas and users without any underlying distributions of the fading channels. However, it requires solving a joint antenna and user selection (JAUS) problem to maximize the instantaneous weighted sum rate in every time slot. Thus, we additionally develop a simple heuristic JAUS algorithm with low computational complexity, called JAUS-LCC, which is executed in every time slot within OJAUS. Finally, through simulation results, we first show that our JAUS-LCC provides near-optimal performance despite requiring very low computational complexity, and then show that our OJAUS with JAUS-LCC well guarantees given minimum average data rate requirements.

INDEX TERMS Antenna selection, multiple-input multiple-output (MIMO), quality-of-service (QoS), time-varying fading channels, user scheduling, weighted sum rate maximization, zero-forcing (ZF) precoding.

I. INTRODUCTION

With the advent of the next generation mobile communication era, such as the fifth generation (5G), beyond 5G (B5G), and the sixth generation (6G), much higher user connectivity (i.e., massive access) and spectral efficiency have become required [1]–[4]. In this circumstance, massive multiple-input multiple-output (MIMO) technology, in which a large number of antennas are generally installed at the base station (BS), has been attracting great attention as a core technology of future communication systems in both industry and academia [5]–[7]. However, to freely control each antenna, one complete radio frequency (RF) chain should be provided for each antenna. Thus, a massive MIMO system ideally

The associate editor coordinating the review of this manuscript and approving it for publication was Kang Chen.

requires as many RF chains as the number of antennas, resulting in high hardware manufacturing and driving costs, high complexity, and high power consumption. Due to these shortcomings, it is practically very challenging to implement such an ideal massive MIMO system.

In this regard, many interesting studies have been conducted over the past few years [8]–[12]. One main direction is to employ a hybrid-analog/digital-beamforming-based architecture [8], [9], where massively many transmit antennas are connected with a limited number of RF chains via analog phase shifters, so that digital precoding and analog beamforming are combined exclusively in the baseband and RF band, respectively. Thereby, high hardware manufacturing costs caused by massively many RF chains can be addressed [10]. However, in this approach, all transmit antennas need to be always activated, so still high hardware driving costs, high complexity, and high power consumption are incurred. The other main direction is to employ an antenna-selection-based architecture [11], [12], where only a few of transmit antennas (usually as many as the number of RF chains) are activated and mapped one-to-one to RF chains. This architecture is based on the fact that, in general, not all antennas contribute equally to performance as they usually have different channel conditions [13], [14]. In general, this approach has very slightly lower performance than the hybrid analog/digital beamforming approach. However, since only a few antennas are activated, this antenna selection approach has much lower hardware driving costs and complexity. Thereby, power consumption and latency can also be typically reduced. It is worth noting that the low energy consumption and low latency are ones of the key performance indicators (KPI) of 5G and B5G [4], which can be addressed by the antenna selection technology. Hence, in this paper, we focus on the antenna selection technology that can reduce the complexity and implementation overhead while maintaining reasonably high performance.

Recently, many studies on antenna selection have been conducted [15]–[23]. An optimal antenna selection solution can be obtained using exhaustive search algorithm (ES) [15], but its computational complexity is too high to be used in practice. Hence, a variety of heuristic antenna selection algorithms with less computational complexity than that of ES have been proposed. In [16], [17], the authors have presented greedy-based antenna selection algorithms that select antennas in a greedy manner in the direction of minimizing the loss of their performance metrics. In [18], under a scenario where one BS with multiple antennas serves only one user, the authors have proposed an antenna selection algorithm to maximize the channel capacity using the reinforcement learning. In [19], the authors have considered a specific architecture consisting of a set of substructures in which every two antennas are exclusively connected to one RF chain. With such an architecture, they have developed an algorithm that activates only one of the two antennas for each RF chain so that the sum capacity is maximized. In [20]–[22], antenna selection algorithms based on channel norms between transmit and receive antennas have been proposed, which greatly reduce the computational complexity but incur significant sum rate losses. In [23], the authors have proposed channel correlation-based worst-first discarding (CWF) and channel correlation-based best-first selection (CBF) algorithms for antenna selection. However, in all the above works [16]–[23], the authors have dealt with only the antenna selection under the assumption that either multiple users or a single user being served are given in advance.

More recently, with recognizing the importance of user selection in multi-user massive MIMO systems, joint antenna and user selection (JAUS) has been studied [24]–[26]. The authors in [24] have proposed a heuristic JAUS algorithm for sum rate maximization, in which starting with all antennas activated, the antennas are deactivated one by one in a greedy manner until the number of active antennas equals the number of RF chains, while user selection is conducted using the semi-orthogonal user selection algorithm (SUS) [27] that selects a user subset in such a way that their channel orthogonality becomes maximized. In [25], the authors have developed another heuristic JAUS algorithm for maximizing the sum rate per unit energy consumption, in which users with superior channels are selected first, and then antennas with the least contributions to the sum rate are iteratively deleted until the number of active antennas equals the number of RF chains. In [26], the authors have developed a probability-based JAUS algorithm that finds the probabilities that each antenna and each user will be activated, respectively, with the aim of maximizing the sum capacity. Although these works [24]–[26] have studied the JAUS problem, there are still several critical limitations. First, they all have considered the sum rate as the objective function to be maximized. Although the sum rate maximization is one of the very important and fundamental problems, it cannot reflect the relative importance, priority, and fairness between users. Second, they have not considered any quality-of-service (QoS) requirements. Thereby, users with poor channel conditions may have very low data rate performance (even zero) as focusing only on maximizing the sum rate. In this regard, it is very important to consider the weights and/or QoS constraints in the JAUS problem.

In addition, although there are many studies on either antenna selection, user selection, or both, most of the existing studies including [15]–[26] have focused on optimization problems from a snapshot perspective. That is, the antenna and/or user *selection* under fixed channel conditions has been dealt with, rather than antenna and/or user *scheduling* with considering time-varying fading channels. Accordingly, the characteristics of time-varying fading channels have not been able to be fully leveraged in the existing studies even though resource scheduling with leveraging them is very effective in increasing various system performance (such as the network throughput, the harvested energy amount, etc.) especially in wireless network systems [28]–[31]. Meanwhile, unlike the conventional MIMO, massive MIMO generally requires much higher computational complexity in performing scheduling due to the large number of antennas and additional restrictions caused by the limited number of RF chains. Such the complexity issue should be addressed in scheduling in the massive MIMO system.

Motivated by the above observations, in this paper, we study the joint antenna and user scheduling problem for the downlink in the massive MIMO system over time-varying fading channels. The goal of the problem is to maximize the weighted average sum rate while ensuring each user's minimum average data rate requirement. To solve the problem, we first develop an opportunistic joint antenna and user scheduling algorithm (OJAUS) using the dual and stochastic subgradient methods, which allows us not only to take full advantage of the time-varying fading channels, but also to solve the problem without knowing underlying distributions of the fading channels. However, OJAUS necessitates solving a JAUS problem for maximizing the instantaneous weighted sum rate at every time slot. To the best of our knowledge, there are many existing studies attempting to optimize antenna selection and/or user selection for maximizing the instantaneous (equally-weighted) sum rate, but none to optimize them jointly for maximizing the instantaneous (unequally) weighted sum rate. Hence, we additionally develop a simple heuristic JAUS algorithm with low computational complexity (JAUS-LCC) to maximize the instantaneous weighted sum rate based on the greedy algorithm. Through simulation results, we show that JAUS-LCC provides near-optimal performance despite requiring very low computational complexity, and then confirm that OJAUS in which JAUS-LCC runs internally in every time slot meets given QoS requirements well. It is worth noting that, to the best of our knowledge, this is the first work to jointly optimize antenna and user scheduling to maximize the weighted average sum rate while taking QoS constraints into account in the massive MIMO system over time-varying fading channels.

The remainder of the paper is organized as follows. In Section II, we provide the system model. In Section III, we formulate the joint antenna and user scheduling problem and develop OJAUS to solve it. In Section IV, we develop JAUS-LCC that will be embedded in OJAUS. Simulation results are presented in Section V, followed by the conclusion in Section VI.

A. NOTATION

Scalars, vectors, matrices, and sets are denoted by lower- or upper-case italic letters, lower-case boldface letters, upper-case boldface letters, and calligraphic letters, respectively. The set of complex numbers is denoted by $\mathbb{C}$. The expectation and conjugate transpose operators are denoted by $\mathbb{E}[\cdot]$ and $(\cdot)^H$, 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 zero vector of size n is denoted by $\mathbf{0}_K$, and the identity matrix of size n is denoted by $\mathbf{I}_n$. A diagonal matrix whose ith diagonal element is the ith element of a vector $\mathbf{x}$ is denoted by $\text{diag}(\mathbf{x})$. The absolute value of a real/complex number x, the ℓ^2-norm of a vector $\mathbf{x}$, and the cardinality of a set $\mathcal{X}$ are denoted by $|x|$, $\|\mathbf{x}\|$, and $|\mathcal{X}|$, respectively. Unless otherwise specified, the base of the logarithm is considered to be 2.

II. SYSTEM MODEL AND RATE ANALYSIS

We focus on the downlink in the massive MIMO system, where one BS equipped with N RF chains and M antennas serves K users equipped with a single antenna in a single cell. We assume that $N < M$, and denote an index set of the BS antennas by $\mathcal{M} = \{1, 2, \ldots, M\}$ and that of the users by $\mathcal{K} = \{1, 2, \ldots, K\}$. According to the literature [8], [9], the following two hybrid beamforming architectures are typically adopted in the massive MIMO system: the fully-connected architecture and the partially-connected architecture. In this paper, we adopt the fully-connected architecture where each RF chain is connected to all antennas since it generally offers a higher array gain as well as higher flexibility in the design of narrow beam, compared to the partially-connected architecture. An illustration of the system model that we are considering in this paper is shown in Fig. 1.

We assume a time-slotted system over block fading channels, where each channel gain changes from one time slot to another but remains constant during a time slot. We denote the channel gain between BS antenna m and user k in time slot t by $h^t_{m,k}$, which incorporates the effects of both large-scale fading and small-scale fading. More precisely, the channel gain is modeled as $h^t_{m,k} = (\gamma_k)^{1/2}\tilde{h}_{m,k}$, where $\tilde{h}_{m,k}$ is the small-scale fading and γ_k is the large-scale fading depending only on k. This modeling is reasonable if the distance between the BS antennas is much smaller than the distance between the BS and the user [32]. Then, we define the fading process between BS antenna m and user k by $\{h^t_{m,k}, t = 1, 2, \ldots\}$ and assume that the fading process is stationary and ergodic. In turn, we define the channel vector for each user k in time slot t and the channel matrix in time slot t by $\mathbf{h}^t_k = (h^t_{m,k})_{\forall m \in \mathcal{M}}$ and $\mathbf{H}^t = [\mathbf{h}^t_1, \ldots, \mathbf{h}^t_K]$, respectively. We assume that underlying distributions of the fading process are unknown to the BS since it is practically challenging to obtain such information a priori, but an instantaneous channel matrix is perfectly known to the BS at the beginning of each time slot, as in [15]–[17], [33].[1]

We now model the transmitted and received signal vectors. The transmitted signal vector, $\mathbf{x}^t \in \mathbb{C}^{M\times 1}$, at the BS to all the K users in time slot t is given by

$$\mathbf{x}^t = \sum_{k\in\mathcal{K}} \mathbf{w}^t_k \sqrt{p^t_k}\, q^t_k, \tag{1}$$

[1]In general, channel estimation is not straightforward especially in the massive MIMO system with a limited number of RF chains. However, in this paper, to facilitate the study of antenna and user scheduling from a system-level point of view, we assume perfect instantaneous channel estimation. For readers interested in channel estimation in the massive MIMO system with a limited number of RF chains, we refer to [34]–[37] and references therein. For example, a general channel estimation problem is dealt with in [34], angle-domain channel estimation/tracking schemes are proposed in [35], [36], and a low-complexity channel estimation scheme is proposed in [37]. In the same vein, we do not take into account the overhead consumption of channel estimation in this paper, and instead, refer to [14] for readers interested in this.

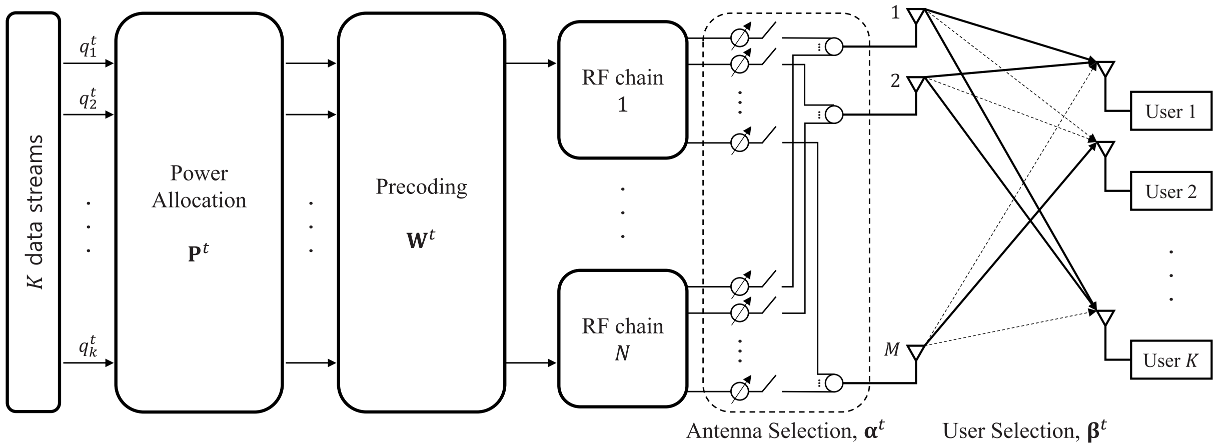


FIGURE 1. **An Illustration of the system model with transmit antenna selection.**

where $\mathbf{w}_k^t = (w_{m,k}^t)_{\forall m \in \mathcal{M}}$ is the precoding vector, p_k^t is the power allocation variable, and q_k^t is the unity-energy information signal for user k in time slot t. For convenience, the transmitted signal vector in (1) can be rewritten as

$$\mathbf{x}^t = \mathbf{W}^t \mathbf{P}^t \mathbf{q}^t, \tag{2}$$

where $\mathbf{W}^t = [\mathbf{w}_1^t, \ldots, \mathbf{w}_K^t]$ is the precoding matrix, $\mathbf{P}^t = \mathrm{diag}((p_1^t)^{1/2}, \ldots, (p_K^t)^{1/2})$ is the diagonal power allocation matrix, and $\mathbf{q}^t = (q_k^t)_{\forall k \in \mathcal{K}}$ is the information signal vector in time slot t. The transmit power is constrained by the total transmission power budget, P_T, of the BS as

$$\mathbb{E}\left[(\mathbf{x}^t)^H \mathbf{x}^t\right] = \sum_{k \in \mathcal{K}} \left\|\mathbf{w}_k^t\right\|^2 p_k^t \le P_T. \tag{3}$$

The received signal, y_k^t, at user k in time slot t is given by

$$\begin{aligned} y_k^t &= (\mathbf{h}_k^t)^H \mathbf{x}^t + n_k^t \\ &= \underbrace{(\mathbf{h}_k^t)^H \mathbf{w}_k^t \sqrt{p_k^t}\, q_k^t}_{\text{desired signal}} + \underbrace{(\mathbf{h}_k^t)^H \sum_{\substack{k' \in \mathcal{K} \\ k' \neq k}} \mathbf{w}_{k'}^t \sqrt{p_{k'}^t}\, q_{k'}^t}_{\text{interference signal}} + \underbrace{n_k^t}_{\substack{\text{noise} \\ \text{signal}}}, \end{aligned} \tag{4}$$

where $n_k^t \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive white Gaussian noise at user k in time slot t. For convenience, similarly to (2), the received signal vector in time slot t can be expressed as

$$\mathbf{y}^t = (\mathbf{H}^t)^H \mathbf{x}^t + \mathbf{n}^t, \tag{5}$$

where $\mathbf{y}^t = (y_k^t)_{k \in \mathcal{K}}$ and $\mathbf{n}^t = (n_k^t)_{k \in \mathcal{K}}$.

In this paper, we adopt the zero-forcing (ZF) precoding scheme, which is one of the most popular linear precoding schemes due to its low computational complexity and interference cancellation property [38], [39]. Accordingly, the precoding matrix, $\mathbf{W}^t$, in time slot t is defined by

$$\mathbf{W}^t = \mathbf{H}^t \left[(\mathbf{H}^t)^H \mathbf{H}^t\right]^{-1}. \tag{6}$$

We assume that $\mathbf{H}^t$ has a full row-rank so that $(\mathbf{H}^t)^H \mathbf{H}^t$ is invertible as in many literature on this field [40]–[42]. Since it is assumed that the instantaneous channel matrix is perfectly known to the BS at that time slot, the BS can obtain the corresponding precoding matrix with it according to (6). The kth column of $\mathbf{W}^t$ obtained by (6) corresponds to $\mathbf{w}_k^t$ in (1), which satisfies that $(\mathbf{h}_j^t)^H \mathbf{w}_k^t = 0$ if $j \neq k$ for all $1 \le j \le K$, and $(\mathbf{h}_j^t)^H \mathbf{w}_k^t = 1$ if $j = k$. Then, by plugging the obtained $\mathbf{w}_k^t$'s into (4), we have the received signal at user k in time slot t as

$$y_k^t = (\mathbf{h}_k^t)^H \mathbf{w}_k^t \sqrt{p_k^t}\, q_k^t + n_k^t = \sqrt{p_k^t}\, q_k^t + n_k^t, \tag{7}$$

in which the undesired interference signals are suppressed by the ZF precoding.

We now define a antenna selection vector in time slot t by $\boldsymbol{\beta}^t = (\beta_m^t)_{\forall m \in \mathcal{M}}$, where

$$\beta_m^t = \begin{cases} 1, & \text{if the } m\text{ th antenna of the BS is} \\ & \text{selected to be activated in time slot } t, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

It is worth noting that the fewer antennas activated, the lower the computational complexity and hardware driving cost can be derived [43]. In contrast, one RF chain can support at most one data stream, which should be transmitted to the corresponding user through at least one antenna. In this regard, we adopt a method of activating as few antennas as possible, but at least as many as the number of RF chains. Accordingly, we have the antenna selection constraint as

$$\sum_{m \in \mathcal{M}} \beta_m^t = N, \quad \forall t. \tag{9}$$

Similarly, we define a user selection vector in time slot t by $\boldsymbol{\alpha}^t = (\alpha_k^t)_{\forall k \in \mathcal{K}}$, where

$$\alpha_k^t = \begin{cases} 1, & \text{if user } k \text{ is selected} \\ & \text{to be served data in time slot } t, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Note that the number of users who can be served data at the same time cannot exceed the number of RF chains in a general scenario where each user is provided with a data stream independent of each other [44]. Hence, we assume that U users ($U \leq N$) are selected in each time slot. Accordingly, we have the user selection constraint as

$$\sum_{k \in \mathcal{K}} \alpha_k^t = U, \quad \forall t. \tag{11}$$

Now, taking the antenna and user selection vectors into account, we can tailor the transmit power constraint of the BS in (3) as

$$\sum_{k \in \mathcal{K}} \alpha_k^t \left\| \mathbf{w}_k^t \right\|^2 p_k^t \leq P_T, \quad \forall t, \tag{12}$$

and the received signal at user k in time slot t in (7) as

$$y_k^t = \alpha_k^t \left[\operatorname{diag}(\boldsymbol{\beta}^t)\mathbf{h}_k^t\right]^H \mathbf{w}_k^t \sqrt{p_k^t}\, q_k^t + n_k^t. \tag{13}$$

From (13), the received signal-to-noise ratio (SNR), ρ_k^t, at user k in time slot t is given by

$$\rho_k^t = \frac{\alpha_k^t p_k^t \left|[\operatorname{diag}(\boldsymbol{\beta}^t)\mathbf{h}_k^t]^H \mathbf{w}_k^t\right|^2}{\sigma_k^2}. \tag{14}$$

Note that since we have adopted the ZF precoding scheme, the terms related to the interference signals are removed in both (13) and (14).

Based on (14), the maximum achievable data rate of user k in time slot t can be defined by

$$\begin{aligned} R_k(\boldsymbol{\alpha}^t, \boldsymbol{\beta}^t, \mathbf{p}^t; \mathbf{H}^t) &= \log\left(1 + \rho_k^t\right) \\ &= \alpha_k^t \log\left(1 + \frac{p_k^t \left|\left[\operatorname{diag}(\boldsymbol{\beta}^t)\mathbf{h}_k^t\right]^H \mathbf{w}_k^t\right|^2}{\sigma_k^2}\right), \end{aligned} \tag{15}$$

where $\mathbf{p}^t = (p_k^t)_{\forall k \in \mathcal{K}}$. From (15), the average data rate of user k can be defined by

$$\bar{R}_k = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R_k(\boldsymbol{\alpha}^t, \boldsymbol{\beta}^t, \mathbf{p}^t; \mathbf{H}^t). \tag{16}$$

We consider that each user k has its minimum average data rate requirement, $\bar{r}_k$, and accordingly, we have the QoS constraints for all the users as

$$\bar{R}_k \geq \bar{r}_k, \quad \forall k \in \mathcal{K}. \tag{17}$$

III. JOINT ANTENNA AND USER SCHEDULING

In this section, based on the system model and rate analysis developed in Section II, we first formulate a stochastic optimization problem for joint antenna and user scheduling, and then develop OJAUS to solve it.

A. OPTIMIZATION PROBLEM FORMULATION

In our study, the goal is to maximize the weighted average sum rate while ensuring QoS constraints given in (17) via joint antenna and user scheduling taking into account the time-varying fading channels. Accordingly, based on (8), (9), (10), (11), (12), (16), and (17), we formulate our joint antenna and user scheduling problem as

$$\begin{aligned} \mathbf{P1}: \quad & \underset{\boldsymbol{\alpha}^t, \boldsymbol{\beta}^t, \mathbf{p}^t, \forall t}{\text{maximize}} \sum_{k \in \mathcal{K}} \eta_k \bar{R}_k \\ & \text{subject to } \bar{R}_k \geq \bar{r}_k, \quad \forall k \in \mathcal{K}, \\ & \qquad \sum_{k \in \mathcal{K}} \alpha_k^t \left\| \mathbf{w}_k^t \right\|^2 p_k^t \leq P_T, \quad \forall t, \\ & \qquad \sum_{k \in \mathcal{K}} \alpha_k^t = U, \quad \forall t, \\ & \qquad \sum_{m \in \mathcal{M}} \beta_m^t = N, \quad \forall t, \\ & \qquad \alpha_k^t \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall t, \\ & \qquad \beta_m^t \in \{0,1\}, \quad \forall m \in \mathcal{M}, \forall t, \\ & \qquad p_k^t \geq 0, \quad \forall k \in \mathcal{K}, \forall t, \end{aligned}$$

where η_k is the weight factor of user k. Since the channel fading is assumed to be ergodic, the long-term time average converges almost surely to the expectation for almost all realizations of the fading process. Hence, by denoting a channel matrix in a generic time slot by $\mathbf{H}$ and replacing the superscript t in decision variables with $\mathbf{H}$, we can reformulate Problem **P1** equivalently as

$$\begin{aligned} \mathbf{P2}: \quad & \underset{\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{p}}}{\text{maximize}} \ \mathbb{E}\left[\sum_{k \in \mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H})\right] \\ & \text{subject to } \mathbb{E}\left[R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H})\right] \geq \bar{r}_k, \quad \forall k \in \mathcal{K}, \\ & \qquad \sum_{k \in \mathcal{K}} \alpha_k^{\mathbf{H}} \left\| \mathbf{w}_k^{\mathbf{H}} \right\|^2 p_k^{\mathbf{H}} \leq P_T, \ \forall \mathbf{H} \in \mathcal{H}, \\ & \qquad \sum_{k \in \mathcal{K}} \alpha_k^{\mathbf{H}} = U, \ \forall \mathbf{H} \in \mathcal{H}, \\ & \qquad \sum_{m \in \mathcal{M}} \beta_m^{\mathbf{H}} = N, \ \forall \mathbf{H} \in \mathcal{H}, \\ & \qquad \alpha_k^{\mathbf{H}} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall \mathbf{H} \in \mathcal{H}, \\ & \qquad \beta_m^{\mathbf{H}} \in \{0,1\}, \quad \forall m \in \mathcal{M}, \forall \mathbf{H} \in \mathcal{H}, \\ & \qquad p_k^{\mathbf{H}} \geq 0, \quad \forall k \in \mathcal{K}, \forall \mathbf{H} \in \mathcal{H}. \end{aligned}$$

where $\bar{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^{\mathbf{H}})_{\forall \mathbf{H} \in \mathcal{H}}$, $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta}^{\mathbf{H}})_{\forall \mathbf{H} \in \mathcal{H}}$, $\bar{\mathbf{p}} = (\mathbf{p}^{\mathbf{H}})_{\forall \mathbf{H} \in \mathcal{H}}$, and $\mathcal{H}$ is the support of $\mathbf{H}$. Note that solving Problem **P2** is equivalent to solving Problem **P1**. More precisely, for each time slot t with an arbitrary channel realization $\mathbf{H}$, an optimal solution $(\boldsymbol{\alpha}^t)^*$, $(\boldsymbol{\beta}^t)^*$, and $(\mathbf{p}^t)^*$ is obtained according to

$$(\boldsymbol{\alpha}^t)^* = (\boldsymbol{\alpha}^{\mathbf{H}})^*, \ (\boldsymbol{\beta}^t)^* = (\boldsymbol{\beta}^{\mathbf{H}})^*, \ \text{and } (\mathbf{p}^t)^* = (\mathbf{p}^{\mathbf{H}})^*, \tag{18}$$

where $(\boldsymbol{\alpha}^{\mathbf{H}})^*$, $(\boldsymbol{\beta}^{\mathbf{H}})^*$, and $(\mathbf{p}^{\mathbf{H}})^*$ are obtained from the optimal solution to Problem **P2**.

B. OPPORTUNISTIC JOINT ANTENNA AND USER SCHEDULING ALGORITHM (OJAUS)

In this subsection, we discuss how to solve Problem **P2** with the following challenges. First, in many practical applications, underlying distributions of $\mathbf{H}$ are generally unknown. Second, the problem is a nonconvex optimization problem mainly owing to the existence of the integer variables. To cope with such challenges, we use the dual and stochastic subgradient descent methods, by which we only need to observe the realized channel matrix $\mathbf{H}$ for each time slot without knowing the underlying distributions of $\mathbf{H}$.

To this end, we first define the Lagrangian function of Problem **P2** by

$$\begin{aligned} L(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{p}}, \boldsymbol{\lambda}) &= \mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H})\right] \\ &\quad + \sum_{k\in\mathcal{K}} \lambda_k \left(\mathbb{E}\left[R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H})\right] - \bar{r}_k\right) \\ &= \mathbb{E}\left[\sum_{k\in\mathcal{K}} (\eta_k + \lambda_k) R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H})\right] - \sum_{k\in\mathcal{K}} \lambda_k \bar{r}_k, \end{aligned} \tag{19}$$

where $\boldsymbol{\lambda} = (\lambda_k)_{\forall k\in\mathcal{K}}$ is the nonnegative Lagrangian multiplier vector corresponding to the QoS constraints in (17). Then, based on (19), we can define the dual problem associated with Problem **P2** by

$$\mathbf{D} : \underset{\boldsymbol{\lambda}\succeq \mathbf{0}_K}{\text{minimize}} \quad F(\boldsymbol{\lambda}),$$

where $\succeq$ is the elementry-wise inequality, and the objective function is obtained as

$$\begin{aligned} F(\boldsymbol{\lambda}) = \underset{\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{p}}}{\text{maximize}} \ & L(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{p}}, \boldsymbol{\lambda}) \\ \text{subject to} \ & \sum_{k\in\mathcal{K}} \alpha_k^{\mathbf{H}} \left\|\mathbf{w}_k^{\mathbf{H}}\right\|^2 p_k^{\mathbf{H}} \le P_T, \quad \forall \mathbf{H} \in \mathcal{H}, \\ & \sum_{k\in\mathcal{K}} \alpha_k^{\mathbf{H}} = U, \quad \forall \mathbf{H} \in \mathcal{H}, \\ & \sum_{m\in\mathcal{M}} \beta_m^{\mathbf{H}} = N, \quad \forall \mathbf{H} \in \mathcal{H}, \\ & \alpha_k^{\mathbf{H}} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \ \forall \mathbf{H} \in \mathcal{H}, \\ & \beta_m^{\mathbf{H}} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, \ \forall \mathbf{H} \in \mathcal{H}, \\ & p_k^{\mathbf{H}} \ge 0, \quad \forall k \in \mathcal{K}, \ \forall \mathbf{H} \in \mathcal{H}. \end{aligned} \tag{20}$$

Prior to discussing how to solve the maximization problem in (20), we focus on the fact that $L(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{p}}, \boldsymbol{\lambda})$ in (19) is separable for each channel matrix realization. Hence, for a given $\boldsymbol{\lambda}$, we can solve the maximization problem in (20) by separately solving the following subproblem for each $\mathbf{H}$:

$$\begin{aligned} \mathbf{D}^{\mathbf{H}} : \underset{\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}}{\text{maximize}} \ & \sum_{k\in\mathcal{K}} (\eta_k + \lambda_k) R_k(\boldsymbol{\alpha}^{\mathbf{H}}, \boldsymbol{\beta}^{\mathbf{H}}, \mathbf{p}^{\mathbf{H}}; \mathbf{H}) \\ \text{subject to} \ & \sum_{k\in\mathcal{K}} \alpha_k^{\mathbf{H}} \left\|\mathbf{w}_k^{\mathbf{H}}\right\|^2 p_k^{\mathbf{H}} \le P_T, \\ & \sum_{k\in\mathcal{K}} \alpha_k^{\mathbf{H}} = U, \\ & \sum_{m\in\mathcal{M}} \beta_m^{\mathbf{H}} = N, \\ & \alpha_k^{\mathbf{H}} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \\ & \beta_m^{\mathbf{H}} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, \\ & p_k^{\mathbf{H}} \ge 0, \quad \forall k \in \mathcal{K}. \end{aligned}$$

Algorithm 1 OJAUS

Initialize: $t = 1$ and $\boldsymbol{\lambda}^t = \mathbf{0}$.
for each time slot t **do**
  Obtain the solution to Problem $\mathbf{D}^{\mathbf{H}}$ with $\mathbf{H} = \mathbf{H}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$ using Algorithm 2.
  Transmit the signal generated by the obtained solution.
  Calculate $\boldsymbol{\lambda}^{t+1}$ according to (21).
  $t = t + 1$.
end for

Since the expectation has disappeared in Problem $\mathbf{D}^{\mathbf{H}}$, we can solve it without knowing the underlying distributions of $\mathbf{H}$. In addition, Problem $\mathbf{D}^{\mathbf{H}}$ can be interpreted as an instantaneous weighted sum rate maximization problem with weight $\eta_k + \lambda_k$ for each user k. Hence, we will develop an algorithm to solve Problem $\mathbf{D}^{\mathbf{H}}$ in the next section, so hereinafter, we develop an algorithm to solve Problem **D** under the assumption that Problem $\mathbf{D}^{\mathbf{H}}$ can be solved.

In solving Problem **D**, there is still a challenge in that although Problem $\mathbf{D}^{\mathbf{H}}$ is solvable for any given $\boldsymbol{\lambda}$ and $\mathbf{H}$, the underlying distributions of $\mathbf{H}$ are still required to solve Problem **D**. To resolve this challenge, we use the stochastic subgradient method based on the fact that Problem **D** is a convex stochastic optimization problem. Accordingly, we update the Lagrangian multiplier of Problem **D** according to

$$\boldsymbol{\lambda}^{t+1} = \max\left\{0, \ \boldsymbol{\lambda}^t - \zeta^t \mathbf{v}^t\right\}, \tag{21}$$

where $\boldsymbol{\lambda}^t$ and ζ^t are the Lagrangian multiplier vector and the step size, respectively, in time slot t, and $\mathbf{v}^t = (v_k^t)_{\forall k\in\mathcal{K}}$ is the stochastic subgradient of $F(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$. By Danskin's theorem [45], we can determine $\mathbf{v}^t$ as

$$v_k^t = R_k^t - \bar{r}_k, \quad \forall k \in \mathcal{K}, \tag{22}$$

where R_k^t is the instantaneous data rate of user k in time slot t, which is obtained by selecting antennas and users and allocating power to them according to the solution to Problem $\mathbf{D}^{\mathbf{H}}$ with $\mathbf{H} = \mathbf{H}^t$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$. When the Lagrangian multiplier vector is updated according to (21), it converges to the optimal solution, $\boldsymbol{\lambda}^*$, to Problem **D** if the step size, ζ^t, meets the following conditions [46]:

$$\zeta^t \ge 0, \ \sum_{t=1}^{\infty} \zeta^t = \infty, \quad \text{and} \ \sum_{t=1}^{\infty} (\zeta^t)^2 < \infty. \tag{23}$$

The pseudocode of OJAUS is provided in Algorithm 1.

Before finishing this section, we discuss the optimality of our scheduling problem. According to optimization theory, if a primal problem is not a convex problem, there is generally a duality gap between the primal problem and its dual problem, resulting in loss of optimality in the process of converting to a dual problem. However, there is no duality gap between Problem **P2** and Problem **D** although Problem **P2** is not a convex problem as mentioned before.

Theorem 1: The duality gap is equal to 0 between Problem **P2** and Problem **D**.

Proof: See Appendix. □

By Theorem 1, we confirm that OJAUS guarantees the optimality as long as Problem $\mathbf{D^H}$ is optimally solved.

IV. JOINT ANTENNA AND USER SELECTION (JAUS)

In Section III, we have developed OJAUS under the assumption that Problem $\mathbf{D^H}$ is solvable. Hence, in this section, we develop a heuristic JAUS algorithm with low computational complexity, called JAUS-LCC, that solves Problem $\mathbf{D^H}$.

A. JAUS ALGORITHM WITH LOW COMPUTATIONAL COMPLEXITY (JAUS-LCC)

As mentioned before, Problem $\mathbf{D^H}$ is the JAUS problem to maximize the instantaneous weighted sum rate for a given realization of $\mathbf{H}$. Thus, we omit the superscript $\mathbf{H}$ in decision variables for notational brevity. Then, by plugging (15) into Problem $\mathbf{D^H}$, we can rewrite it equivalently as

$$\begin{aligned}
\mathbf{Q}: \quad & \underset{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{p}}{\text{maximize}} \sum_{k\in\mathcal{K}} \omega_k \alpha_k \log\left(1+\frac{p_k\left|[\mathrm{diag}(\boldsymbol{\beta})\mathbf{h}_k]^H\mathbf{w}_k\right|^2}{\sigma_k^2}\right)\\
& \text{subject to} \sum_{k\in\mathcal{K}} \alpha_k \|\mathbf{w}_k\|^2 \; p_k \le P_T,\\
& \qquad \sum_{k\in\mathcal{K}} \alpha_k = U,\\
& \qquad \sum_{m\in\mathcal{M}} \beta_m = N,\\
& \qquad \alpha_k \in \{0,1\}, \quad \forall k \in \mathcal{K},\\
& \qquad \beta_m \in \{0,1\}, \quad \forall m \in \mathcal{M},\\
& \qquad p_k \ge 0, \quad \forall k \in \mathcal{K}.
\end{aligned}$$

where $\omega_k = \eta_k + \lambda_k$ is the effective weight of user k.[2] It is worth noting that there are many existing studies dealing with the JAUS problem to maximize the sum rate. However, to the best of our knowledge, although the weighted sum rate is one of the most important performance metrics in wireless communications, nothing has been done to maximize it. Hence, the existing algorithms are usually not suitable for solving Problem **Q**, and also provide limited performance, even if they are applicable.

To solve Problem **Q**, we first investigate an optimal power allocation for given user and antenna selection vectors. To this end, we introduce an index set, $\mathcal{K}_{\text{sel}}$, of selected users and an index set, $\mathcal{M}_{\text{sel}}$, of the selected antennas as

$$\mathcal{K}_{\text{sel}} = \{k \in \mathcal{K} : \alpha_k = 1\}, \tag{24}$$
$$\mathcal{M}_{\text{sel}} = \{m \in \mathcal{M} : \beta_m = 1\}. \tag{25}$$

Then, since the ZF precoding is employed, the optimal power allocation for given $\mathcal{K}_{\text{sel}}$ and $\mathcal{M}_{\text{sel}}$ can be obtained by solving

$$\begin{aligned}
& \underset{p_k,\,\forall k\in\mathcal{K}_{\text{sel}}}{\text{maximize}} \sum_{k\in\mathcal{K}_{\text{sel}}} \omega_k \log\left(1+\frac{p_k}{\sigma_k^2}\right)\\
& \text{subject to} \sum_{k\in\mathcal{K}_{\text{sel}}} \|\mathbf{w}_k\|^2 p_k \le P_T,\\
& \qquad p_k \ge 0, \quad \forall k \in \mathcal{K}_{\text{sel}}.
\end{aligned} \tag{26}$$

According to Karush–Kuhn–Tucker (KKT) conditions, we can obtain the optimal solution to the problem in (26) as [47], [48]

$$p_k^* = \max\left\{0, \frac{\omega_k \mu}{\|\mathbf{w}_k\|^2} - \sigma_k^2\right\}, \quad \forall k \in \mathcal{K}_{\text{sel}}, \tag{27}$$

where μ is determined such that

$$\sum_{k\in\mathcal{K}_{\text{sel}}} \|\mathbf{w}_k\|^2 p_k^* = P_T. \tag{28}$$

Note that no power is allocated to non-selected users, i.e., $p_k^* = 0$ for all $k \in \mathcal{K} \setminus \mathcal{K}_{\text{sel}}$.

We now focus on solving Problem **Q** under the assumption that the power allocation is done by (27). Then, the problem is a nonlinear integer programming (NIP) problem that can be optimally solved by ES [15]. However, in spite of its optimality, it undergoes extremely high computational complexity and thus, it is virtually impossible to be exploited in practical applications. Hence, in this paper, we develop a heuristic JAUS-LCC using the concept of the greedy algorithm [49]. Although optimality is not guaranteed, it provides good performance comparable to optimal one despite requiring relatively low computational complexity. Its main idea is to delete users and antennas one by one in a greedy manner in each iteration from whole user and antenna sets until the desired numbers of antennas and users remain. More specifically, it starts by setting all users and all antennas to be selected, i.e.,

$$\mathcal{K}_{\text{sel}} = \mathcal{K} \text{ and } \mathcal{M}_{\text{sel}} = \mathcal{M}. \tag{29}$$

Accordingly, the user and antenna selection vectors are set, respectively, as

$$\boldsymbol{\alpha} = \mathbf{1}_K \text{ and } \boldsymbol{\beta} = \mathbf{1}_M, \tag{30}$$

where $\mathbf{1}_K$ and $\mathbf{1}_M$ are all-ones vectors of size K and M, respectively.

Then, as the first step, under the assumption that the antenna selection vector, $\boldsymbol{\beta}$, is fixed, we update the index set, $\mathcal{K}_{\text{sel}}$, of selected users and the corresponding user selection

[2] For convenience, we also refer to the *effective weight* simply as a *weight* if there is no confusion.

vector, $\boldsymbol{\alpha}$. To this end, we first find one user k^* from $\mathcal{K}_{\text{sel}}$ who provides the smallest weighted data rate, i.e.,

$$k^* = \underset{k\in\mathcal{K}_{\text{sel}}}{\text{argmin}}\ \omega_k \log\left(1+\frac{p_k\left|[\text{diag}(\boldsymbol{\beta})\mathbf{h}_k]^H\mathbf{w}_k\right|^2}{\sigma_k^2}\right), \tag{31}$$

where p_k's are given by the solution to the problem in (26) for currently given $\mathcal{K}_{\text{sel}}$ and $\mathcal{M}_{\text{sel}}$. After finding k^*, we update $\mathcal{K}_{\text{sel}}$ by deleting user k^* from it, i.e.,

$$\mathcal{K}_{\text{sel}} = \mathcal{K}_{\text{sel}} \setminus \{k^*\}, \tag{32}$$

and $\alpha_{k^*} = 0$ accordingly.

Now, as the second step, under the assumption that the user selection vector, $\boldsymbol{\alpha}$, is fixed, we update the index set, $\mathcal{M}_{\text{sel}}$, of selected antennas and the corresponding antenna selection vector, $\boldsymbol{\beta}$. To this end, we first define the weighted sum rate when antenna m^- is deactivated by

$$\text{WSR}(m^-) = \sum_{k\in\mathcal{K}_{\text{sel}}} \omega_k \log\left(1+\frac{p_k\left|[\text{diag}(\tilde{\boldsymbol{\beta}})\mathbf{h}_k]^H\mathbf{w}_k\right|^2}{\sigma_k^2}\right), \tag{33}$$

where $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_m)_{\forall m\in\mathcal{M}}$ is defined such that $\tilde{\beta}_m = \beta_m$ for all $m \in \mathcal{M} \setminus \{m^-\}$ and $\tilde{\beta}_{m^-} = 0$. In (33), the power allocation is done according to the solution to the problem in (26) for currently given $\mathcal{K}_{\text{sel}}$ and $\mathcal{M}_{\text{sel}}$. Then, we find one antenna m^* from $\mathcal{M}_{\text{sel}}$ that provides the largest weighted sum rate when it is deactivated, i.e.,

$$m^* = \underset{m^-\in\mathcal{M}_{\text{sel}}}{\text{argmax}}\ \text{WSR}(m^-). \tag{34}$$

After finding m^*, we update $\mathcal{M}_{\text{sel}}$ by deleting antenna m^* from it, i.e.,

$$\mathcal{M}_{\text{sel}} = \mathcal{M}_{\text{sel}} \setminus \{m^*\}, \tag{35}$$

and $\beta_{m^*} = 0$ accordingly.

We move on to the next iteration, and delete one user and one antenna in the same way until the number of antennas becomes N and that of users becomes U. If either the cardinality of $\mathcal{K}_{\text{sel}}$ or that of $\mathcal{M}_{\text{sel}}$ reaches its desired number first, we continue the deletion process only for the one that have not reached its desired number. For example, suppose that $M = 10$, $K = 6$, $N = 4$, and $U = 4$ so that M-N $> K - U$. Then, both antennas and users are deleted one by one during the first two iterations, and then only the antenna is deleted one by one during the next four iterations. The pseudocode of JAUS-LCC is presented in Algorithm 2.

B. COMPUTATIONAL COMPLEXITY ANALYSIS

In this subsection, we analyze the computational complexity of our JAUS-LCC. To facilitate the analysis, we assume that $M \gg N$. This assumption makes sense because the number of antennas is much greater than that of RF chains in general in the massive MIMO system.

Algorithm 2 JAUS-LCC

```
Initialize: K_sel and M_sel using (29), and α and β using (30).
repeat
    if |K_sel| > U then                     ▷ 1st step – user deletion
        Find the user index, k*, using (31).
        Update the selected user set, K_sel, using (32).
        Accordingly, α_{k*} = 0.
    end if
    if |M_sel| > N then                     ▷ 2nd step – antenna deletion
        for each m⁻ ∈ M_sel do
            Obtain WSR(m⁻) according to (33).
        end for
        Find the antenna index, m*, using (34).
        Update the selected antenna set, M_sel, using (35).
        Accordingly, β_{m*} = 0.
    end if
until |K_sel| = U and |M_sel| = N
```

Our JAUS-LCC can be largely divided into two parts in each iteration: user deletion part and antenna deletion part. In the user deletion part, we first calculate the ZF precoding matrix using (6), obtain the power allocation solution by solving the problem in (26), and then compare the weighted data rates of users to find k^* according to (31). With the assumption that $M \gg N$, their computational complexities are given by $\mathbb{O}(K^3)$, $\mathbb{O}(K \log K)$ [48], and $\mathbb{O}(MK)$, respectively. Hence, the overall computational complexity of the user deletion part can be expressed as $\mathbb{O}(K(M + K^2))$. Next, in the antenna deletion part, we first calculate the weighted sum rate when an arbitrary antenna is excluded using (33), whose computational complexity is given by $\mathbb{O}(K(M+K^2))$, and then compare all cases to find m^* according to (34), whose computational complexity is given by $\mathbb{O}(M)$. Hence, the overall computational complexity of the antenna deletion part is expressed as $\mathbb{O}(MK(M + K^2))$. As a result, since the user deletion part and antenna deletion part are repeated $K - U$ times and $M - N \approx M$ times, respectively, the total computational complexity of JAUS-LCC is given by $\mathbb{O}(K(K - U)(M + K^2)) + \mathbb{O}(M^2K(M + K^2)) = \mathbb{O}((M^2 + (K - U))K(M + K^2)) \approx \mathbb{O}(M^2K(M + K^2))$, where the last approximation holds if $M \gg K$, which is commonly assumed in massive MIMO studies.

It is worth noting that JAUS based on ES [15], called JAUS-ES, gives an optimal JAUS solution, but using it is challenging in practice due to its excessive computational complexity. More specifically, JAUS-ES requires to compare all possible JAUS combinations of size $\binom{M}{N}\binom{K}{U}$, and the weighted sum rate for each JAUS combination should be evaluated with the computational complexity of $\mathbb{O}(K(M + K^2))$. Thus, its total computational complexity is given by $\mathbb{O}(\binom{M}{N}\binom{K}{U}K(M+K^2))$. This computational complexity analysis confirms that our JAUS-LCC has significantly lower computational complexity than JAUS-ES, especially in the massive MIMO system with large M.

TABLE 1. Simulation parameter values.

Parameter	Value
Cell radius	500 m
Total transmission power budget	43 dBm
Large-scale path loss	$128.1 + 37.6\log_{10}(d_{km})$ dB
Small-scale fading	Rayleigh fading
Shadowing standard deviation	8 dB
Antenna gain of the BS	15 dBi
Antenna gain of each user	0 dBi
System bandwidth	5 MHz
Noise spectral density	−174 dBm/Hz

V. SIMULATION RESULTS

In this section, we present simulation results to evaluate the performance of our proposed algorithms. Throughout the simulations, we consider the downlink in the single-cell massive MIMO system where K single-antenna users are being served by one BS equipped with N RF chains and M transmit antennas. We assume that N antennas out of M antennas and U users out of K users are selected. Unless otherwise specified, the total transmission power budget, P_T, of the BS is set to 43 dBm, and users are randomly deployed between 30 m and 500 m away from the BS. As in many massive MIMO studies [50]–[52], the large-scale path loss is set to $128.1 + 37.6\log_{10}(d_{km})$ dB, where d_{km} is the propagation distance from the BS in kilometer, and the antenna gains of the BS and each user are set to 15 dBi and 0 dBi, respectively. We also consider the small-scale fading modeled as the Rayleigh fading with unit variance and the shadow fading with a standard deviation of 8 dB. In turn, the bandwidth and the noise spectral density are set to 5 MHz and −174 dBm/Hz, respectively. We summarize the simulation parameter values in Table 1.

A. EVALUATION OF JAUS-LCC

In this subsection, we investigate the performance of our JAUS-LCC that aims at maximizing the weighted sum rate by solving Problem **Q**. To this end, we compare it with performances of the following baseline algorithms: (i) JAUS-ES [15], (ii) JAUS-CWF where antenna selection is done by CWF based on channel correlation [23], and user selection is done by our user deletion process, (iii) JAUS-CBF where antenna selection is done by CBF based on channel correlation [23], and user selection is done by our user deletion process, (iv) JAUS-SUS where antenna selection is done by our antenna deletion process, and user selection is done by SUS based on channel orthogonality [24], [27], (v) JAUS-Norm where N antennas with the highest channel norms with users are selected, and U users with the highest channel norms with antennas are selected [22], and (vi) JAUS-Random where N antennas and U users are randomly selected. All the following simulation results are obtained by average of 5000 independent trials, in which users' locations are independently and randomly generated in the cell, the channel gains are accordingly generated, and the weights of users are randomly given between 0 and 1.

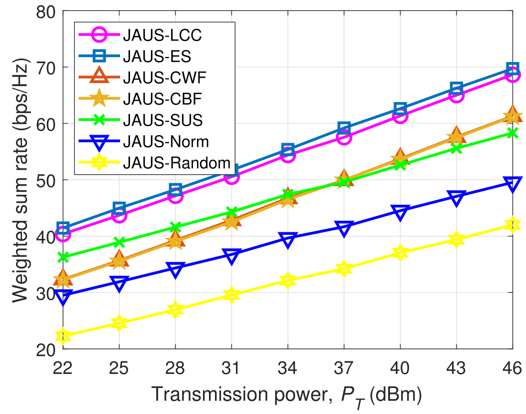


FIGURE 2. Weighted sum rates with varying the transmit power.

Fig. 2 shows the weighted sum rate performance with varying the total transmission power budget, P_T, of the BS from 22 dBm to 46 dBm. In this simulation, the numbers of BS antennas, RF chains, and users in the system are set to $M = 10$, $N = 5$, and $K = 10$, respectively. We assume that the five users are selected, i.e., $U = 5$. In the figure, the weighted sum rate performance of all the JAUS algorithms is increasing as the total transmission power budget, P_T, of the BS increases. Meanwhile, our JAUS-LCC not only provides near-optimal performance (i.e., very close to that of JAUS-ES that provides optimal performance), but also outperforms the others. This demonstrates that it is very effective to select antennas and users by deleting one by one alternately as in our JAUS-LCC. The reason why JAUS-CWF, JAUS-CBF, JAUS-SUS, and JAUS-Norm perform much worse than ours is that they focus only on the correlation, orthogonality, and norm of channels, respectively, while our JAUS-LCC tries to maximize the weighted sum rate by considering not only the channels of users but also their weights. As a result, our JAUS-LCC is very practical not only because it gives near-optimal performance comparable to JAUS-ES, but also because it has relatively very low computational complexity.

Fig. 3 shows the weighted sum rate performance with varying the number of transmit antennas from 20 to 100, where the parameter settings are the same as before. Note that we omit the simulation results for JAUS-ES hereinafter since its computational complexity becomes prohibitively too high to obtain its simulation results as the parameters values (e.g., the number of antennas, users, or both) get larger. As can be seen in the figure, our JAUS-LCC provides clearly the highest weighted sum rate performance. On the other hand, we can see that the weighted sum rate performance does not significantly increase despite the increase in the number of the BS antennas. This is because the spatial diversity gain does not change since the number of antennas being activated still remains constant.

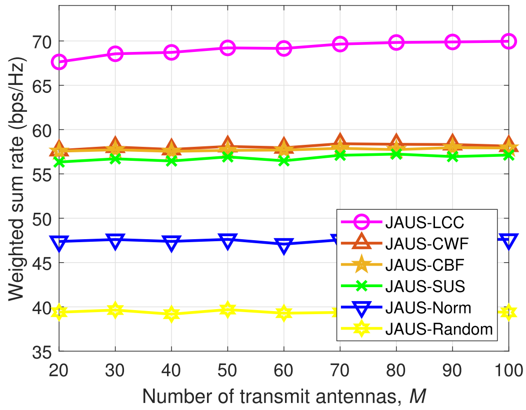


FIGURE 3. **Weighted sum rates with varying the number of transmit antennas.**

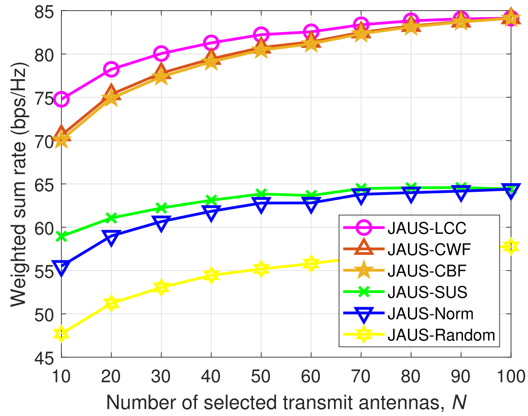


FIGURE 4. **Weighted sum rates with varying the number of selected transmit antennas.**

Fig. 4 shows the weighted sum rate performance with varying the number of selected antennas from 10 to 100, where the parameter settings are the same as before, except that the number of transmit antennas is set to 100. As expected, our JAUS-LCC provides the highest weighted sum rate performance. Also, we can see that the weighted sum rate is clearly increasing as the number of selected antennas increases. This is because the more the antennas are activated, the greater the spatial diversity gain can be achieved. On the other hand, as the number of selected antennas increases, the weighted sum rates of JAUS-CWF and JAUS-CBF gets closer and closer to that of our JAUS-LCC, and eventually become the same when all antennas are selected and activated. This is because (i) the more antennas are activated, the less effective the antenna selection will be, and (ii) both JAUS-CWF and JAUS-CBF follow our user selection scheme, whereas JAUS-SUS and JAUS-Norm follow user selection schemes considering only users' channels. In addition, as the number of active antennas gets larger, the marginal increment of the weighted sum rate gradually decreases. However, the hardware driving cost increases [26]. As a result, we can see that it is very important to determine the number of antennas to be activated in consideration of the above trade-offs in implementing a massive MIMO system.

Fig. 5 shows the weighted sum rate performance with varying the number of users from 10 to 30, where the parameter settings are as follows: $M = 50$, $N = 5$, and $U = 5$. Like the previous simulation results, our JAUS-LCC provides the highest weighted sum rate performance among all the JAUS algorithms. Also, as can be seen in the figure, the weighted sum rate increases as the number of users increases in all JAUS algorithms except JAUS-Random. This is because the multi-user diversity gain can be achieved by the user selection schemes in JAUS algorithms except JAUS-Random, in which it cannot be achieved by random user selection. Similarly to Figure. 4, as the number of users increases, the marginal increment of the weighted sum rate gradually decreases.

Lastly, since perfect knowledge of instantaneous channel state information (CSI) at the BS is unrealistic in practice,

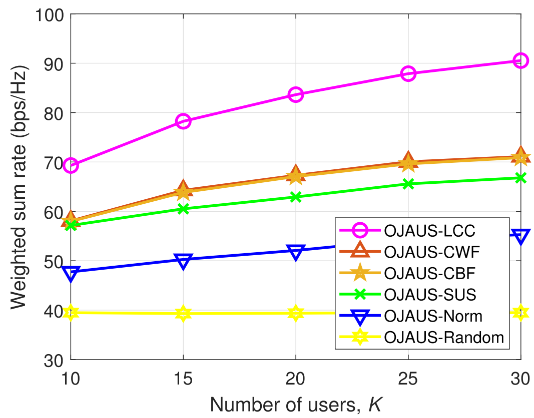
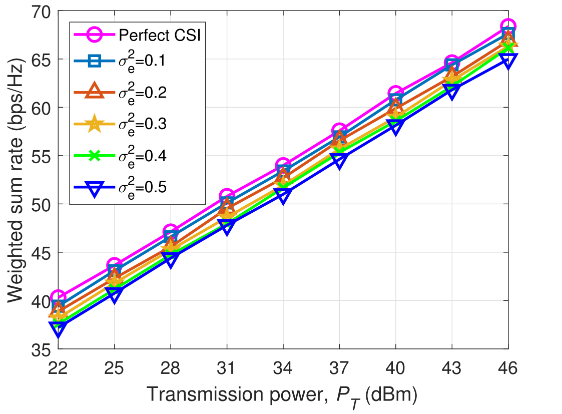
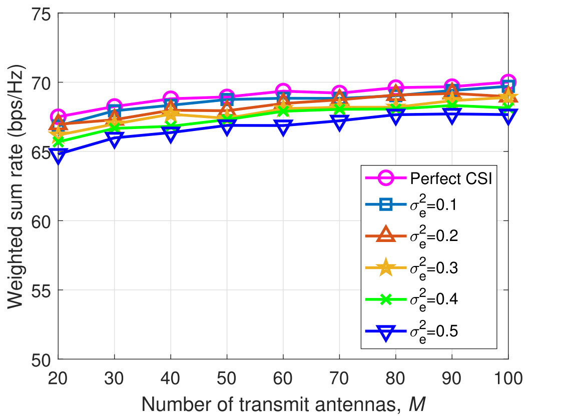


FIGURE 5. **Weighted sum rates with varying the number of users.**

we investigate the effect of the imperfection of CSI in terms of the weighted sum rate performance. To this end, as in [17], we model an estimated channel vector for user k as $\hat{\mathbf{h}}_k = \mathbf{h}_k + (\gamma_k)^{1/2}\mathbf{e}_k$, where $\mathbf{e}_k$ is the channel estimation error, each element of which follows the complex Gaussian distribution with zero mean and variance of σ_e^2. That is, in the estimated channel model being adopted, an independently generated error is imposed on each small-scale fading coefficient. Fig. 6 shows the weighted sum rate performance of our JAUS-LCC in imperfect CSI environments. The parameter settings of Figs. 6a and 6b are the same as in Figs. 2 and 3, respectively, only except that the estimated channel model is exploited. As expected, we can see that the performance of JAUS-LCC slightly deteriorates as the variance of channel estimation error increases. This is mainly due to the fact that the imperfect CSI distorts the ZF precoding as well as the power allocation, compared to the perfect CSI. Nevertheless, comparing Fig. 6a with Fig. 2 and Fig. 6b with Fig. 3, we can see that our JAUS-LCC even in an imperfect CSI environment with $\sigma_e^2 = 0.5$ provides much higher weighted sum rate performance compared to the other JAUS algorithms (except JAUS-ES) in a perfect CSI environment.



(a) Weighted sum rates of JAUS-LCC with varying P_T and σ_e^2.



(b) Weighted sum rates of JAUS-LCC with varying M and σ_e^2.

FIGURE 6. Weighted sum rates in imperfect CSI environments.

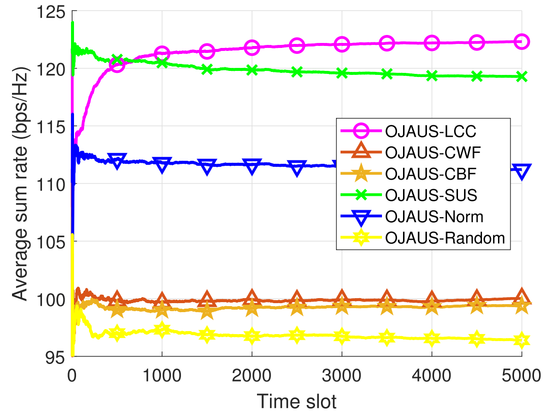
Through the simulation results above, we can conclude that our proposed algorithm, JAUS-LCC, not only provides good performance comparable to JAUS-ES, but also outperforms the other JAUS algorithms in terms of the weighted sum rate performance.

B. EVALUATION OF OJAUS

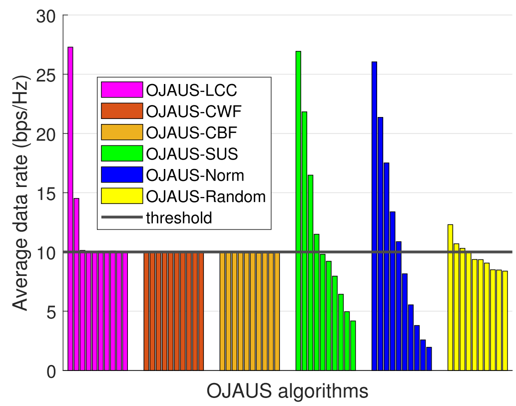
In this subsection, we finally provide the performance of our OJAUS in which JAUS-LCC is performed for each time slot, called OJAUS-LCC, taking into account the required minimum average data rates of users. To show its effectiveness, we compare it with other scheduling algorithms in which JAUS-X for $X \in$ {CWF, CBF, SUS, Norm, Random} is performed for each time slot, called OJAUS-X. To run OJAUS, we set the step size, ζ^t, in time slot t in (21) to $1/t$. The channels are realized with the same way in the previous simulations. We consider a scenario where 5 antennas out of 10 antennas and 5 users out of 10 users with unity weights are selected, and the 10 users are placed such that their distances from the BS are given as in Table 2. In this scenario, we observe the performance results over 5000 time slots.

TABLE 2. Each user's distance from the BS.

User index	1	2	3	4	5
Distance from the BS (m)	30	60	90	120	150
User index	180	210	240	270	300
Distance from the BS (m)	120	140	160	180	200



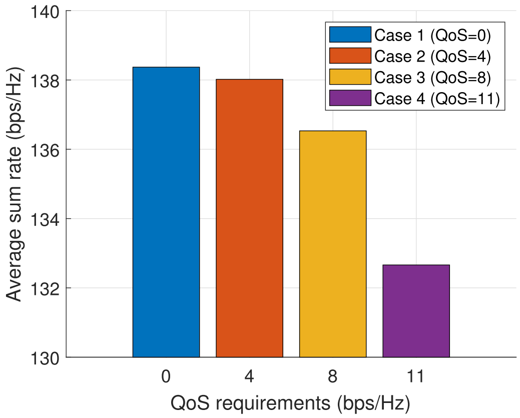
(a) The average sum rate trajectory.



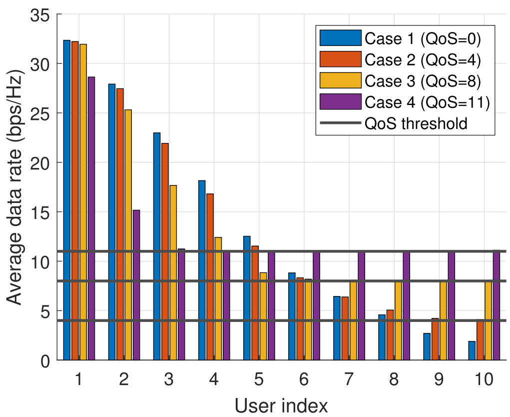
(b) The average data rate (The ith bar for each algorithm represents the ith user.).

FIGURE 7. Performance comparison results between scheduling algorithms for the case where $\bar{r}_k = 10, \forall k$.

Fig. 7 shows the average sum rate and each user's average data rate results, given a minimum average data rate requirement of each user as 10 bps/Hz. In Fig. 7a, we provide the trajectories of the average sum rate over 5000 time slots. We can first see that our OJAUS converges to a stationary point well no matter which JAUS algorithm is used. In addition, our OJAUS-LCC provides higher average sum rate performance than the others. On the other hand, in Fig. 7b, we can observe that the minimum average data rate requirements are only ensured in OJAUS-LCC, OJAUS-CWF, and OJAUS-CBF. This is mainly because of the following two points. First, according to the principle of our OJAUS, the effective weights of users whose channel gains are not good enough to meet their own minimum average data rate requirements are increased (see (21) and (22)). Second, they all follow our



(a) The average sum rate.



(b) The average data rates of users.

FIGURE 8. Performance comparison results between scheduling algorithms for the case where all users have the same QoS requirements.

proposed user selection scheme, taking into account users' channel conditions as well as effective weights (see (31) and (32)). After all, these two points work in synergy with each other, thereby ensuring the minimum average data rate requirements. Meanwhile, whereas OJAUS-CWF and OJAUS-CBF meet the minimum average data rate requirements of all users by narrow margins, our OJAUS-LCC meets them while offering significantly higher average data rates to some users close to the BS. However, OJAUS-SUS and OJAUS-Norm cannot satisfy the minimum average data rate requirements because they follow user selection techniques based only on the users' channels without considering their effective weights. Lastly, in OJAUS-Random, users are evenly selected out over a total of 5000 time slots, which results in very low average sum rate performance as well as dissatisfaction with minimum average data rate requirements.

In Fig. 8, we investigate our OJAUS-LCC in more detail for the following four different minimum average data rate requirements: 0 bps/Hz, 4 bps/Hz, 8 bps/Hz, and 11 bps/Hz. Fig. 8a shows the average sum rate, for the different four QoS requirements. As shown in the figure, the average sum rate decreases as the required average data rates increase. This is because the higher the required average data rates, the more difficult it becomes to focus on users with high channel gains who can contribute more to the sum rate performance enhancement. In other words, users with high channel gains are sacrificed to satisfy the QoS requirements of the other users with low channel gains. Meanwhile, Fig. 8b shows each user's average data rate in detail for the different QoS requirements. Similarly to Fig. 7b, our OJAUS-LCC well meets the given QoS requirements in any case.

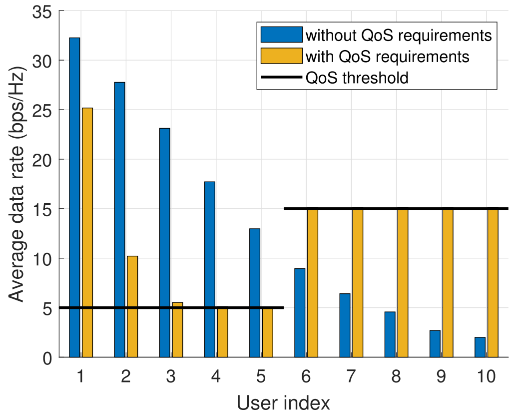


FIGURE 9. Performance comparison results between OJAUS-LCC with individual QoS requirements and OJAUS-LCC without any QoS requirements.

Lastly, we observe the performance of OJAUS-LCC for the case where the users have individually different QoS requirements. To this end, we set the required minimum average data rates of the first five users to 5 bps/Hz while those of the rest five users to 15 bps/Hz, and compare the average data rates between OJAUS-LCC with QoS requirements and that without QoS requirements through Fig. 9. The simulation results confirm that our OJAUS-LCC with QoS requirements well satisfy the individually given QoS requirements of all users. In conclusion, our proposed OJAUS not only guarantees convergence to a stationary point, but also meets given QoS requirements, whether given identically or individually.

VI. CONCLUSION

In this paper, the joint antenna and user scheduling has been investigated for the downlink of a single cell in the massive MIMO system. We aim at maximizing the weighted average sum rate while ensuring given QoS requirements. To solve the problem, we have first developed an opportunistic scheduling algorithm, called OJAUS, by which both antennas and users can be opportunistically scheduled with considering the channel conditions and the achieved QoS satisfaction of users. Then, we have developed a simple heuristic JAUS algorithm with low computational complexity, called JAUS-LCC, which is a built-in algorithm to be executed in every time slot within OJAUS. To show the effectiveness of our proposed algorithms, we have first showed that the computational complexity of our JAUS-LCC is relatively very low compared to JAUS-ES. Subsequently, we have shown that JAUS-LCC not only outperforms the baseline algorithms in terms of weighted sum rate performance but also provides near-optimal performance through the extensive simulation results, despite requiring relatively low computational complexity. In turn, we have also shown that our OJAUS-LCC well satisfies given QoS requirements. This study will be the cornerstone of future work on the development of low-complexity resource allocation/scheduling algorithms that take advantage of the physical characteristics of massive MIMO systems, e.g., channel hardening, favorable propagation, etc.

APPENDIX

To prove Theorem 1, we utilize the fact that if an optimization problem satisfies the time-sharing condition, the duality gap vanishes regardless of the convexity of the problem, which is proved in [53]. To this end, we first define the time-sharing condition for our problem.

Definition 1: Let $\bar{\mathbf{r}} = (\bar{r}_k)_{\forall k \in \mathcal{K}}$ is the vector composed of users' minimum average data rate requirements. Let $\{\bar{\boldsymbol{\alpha}}_x, \bar{\boldsymbol{\beta}}_x, \bar{\mathbf{p}}_x\}$ and $\{\bar{\boldsymbol{\alpha}}_y, \bar{\boldsymbol{\beta}}_y, \bar{\mathbf{p}}_y\}$ be optimal solutions to Problem **P2** when $\bar{\mathbf{r}}$ is given by $\bar{\mathbf{r}}_x = (\bar{r}_{x,k})_{\forall k \in \mathcal{K}}$ and $\bar{\mathbf{r}}_y = (\bar{r}_{y,k})_{\forall k \in \mathcal{K}}$, respectively. Then, Problem **P2** is said to satisfy the time-sharing condition if for any $\bar{\mathbf{r}}_x$, $\bar{\mathbf{r}}_y$, and $0 \le \theta \le 1$, there always exists a feasible solution $\{\bar{\boldsymbol{\alpha}}_z, \bar{\boldsymbol{\beta}}_z, \bar{\mathbf{p}}_z\}$ such that

$$\begin{aligned}\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_z^{\mathbf{H}}, \boldsymbol{\beta}_z^{\mathbf{H}}, \mathbf{p}_z^{\mathbf{H}}; \mathbf{H})\right] &\ge \theta\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_x^{\mathbf{H}}, \boldsymbol{\beta}_x^{\mathbf{H}}, \mathbf{p}_x^{\mathbf{H}}; \mathbf{H})\right] \\ &\quad + (1-\theta)\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_y^{\mathbf{H}}, \boldsymbol{\beta}_y^{\mathbf{H}}, \mathbf{p}_y^{\mathbf{H}}; \mathbf{H})\right], \end{aligned} \tag{36}$$

and

$$\begin{aligned}\mathbb{E}\left[R_k(\boldsymbol{\alpha}_z^{\mathbf{H}}, \boldsymbol{\beta}_z^{\mathbf{H}}, \mathbf{p}_z^{\mathbf{H}}; \mathbf{H})\right] &\ge \theta\mathbb{E}\left[R_k(\boldsymbol{\alpha}_x^{\mathbf{H}}, \boldsymbol{\beta}_x^{\mathbf{H}}, \mathbf{p}_x^{\mathbf{H}}; \mathbf{H})\right] \\ &\quad + (1-\theta)\mathbb{E}\left[R_k(\boldsymbol{\alpha}_y^{\mathbf{H}}, \boldsymbol{\beta}_y^{\mathbf{H}}, \mathbf{p}_y^{\mathbf{H}}; \mathbf{H})\right], \quad \forall k \in \mathcal{K}. \end{aligned} \tag{37}$$

Now, we prove Theorem 1 by proving that Problem **P2** satisfies the time-sharing condition, i.e., inequalities (36) and (37). First, for any $\bar{\mathbf{r}}_x$, $\bar{\mathbf{r}}_y$, and $0 \le \theta \le 1$, let us consider $\{\boldsymbol{\alpha}_z^t, \boldsymbol{\beta}_z^t, \mathbf{p}_z^t\}$ given by

$$\{\boldsymbol{\alpha}_z^t, \boldsymbol{\beta}_z^t, \mathbf{p}_z^t\} = \begin{cases} \{\boldsymbol{\alpha}_x^t, \boldsymbol{\beta}_x^t, \mathbf{p}_x^t\}, & t \le \lfloor \theta T \rfloor, \\ \{\boldsymbol{\alpha}_y^t, \boldsymbol{\beta}_y^t, \mathbf{p}_y^t\}, & t > \lfloor \theta T + 1 \rfloor, \end{cases} \tag{38}$$

where $\lfloor \cdot \rfloor$ is the floor function that gives the largest integer not exceeding its argument. Then, the first inequality (36) is always satisfied by

$$\begin{aligned}&\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_z^{\mathbf{H}}, \boldsymbol{\beta}_z^{\mathbf{H}}, \mathbf{p}_z^{\mathbf{H}}; \mathbf{H})\right] \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_z^t, \boldsymbol{\beta}_z^t, \mathbf{p}_z^t; \mathbf{H}^t) \\ &= \lim_{T\to\infty} \frac{1}{T} \left(\sum_{t=1}^{\lfloor \theta T \rfloor} \sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_x^t, \boldsymbol{\beta}_x^t, \mathbf{p}_x^t; \mathbf{H}^t) + \sum_{t=\lfloor \theta T+1 \rfloor}^{T} \sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_y^t, \boldsymbol{\beta}_y^t, \mathbf{p}_y^t; \mathbf{H}^t) \right) \\ &= \theta\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_x^{\mathbf{H}}, \boldsymbol{\beta}_x^{\mathbf{H}}, \mathbf{p}_x^{\mathbf{H}}; \mathbf{H})\right] \\ &\quad + (1-\theta)\mathbb{E}\left[\sum_{k\in\mathcal{K}} \eta_k R_k(\boldsymbol{\alpha}_y^{\mathbf{H}}, \boldsymbol{\beta}_y^{\mathbf{H}}, \mathbf{p}_y^{\mathbf{H}}; \mathbf{H})\right]. \end{aligned} \tag{39}$$

Similarly, the second inequality (37) is also always satisfied by, for all $k \in \mathcal{K}$,

$$\begin{aligned}&\mathbb{E}\left[R_k(\boldsymbol{\alpha}_z^{\mathbf{H}}, \boldsymbol{\beta}_z^{\mathbf{H}}, \mathbf{p}_z^{\mathbf{H}}; \mathbf{H})\right] \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R_k(\boldsymbol{\alpha}_z^t, \boldsymbol{\beta}_z^t, \mathbf{p}_z^t; \mathbf{H}^t) \\ &= \lim_{T\to\infty} \frac{1}{T} \left(\sum_{t=1}^{\lfloor \theta T \rfloor} R_k(\boldsymbol{\alpha}_x^t, \boldsymbol{\beta}_x^t, \mathbf{p}_x^t; \mathbf{H}^t) + \sum_{t=\lfloor \theta T+1 \rfloor}^{T} R_k(\boldsymbol{\alpha}_y^t, \boldsymbol{\beta}_y^t, \mathbf{p}_y^t; \mathbf{H}^t) \right) \\ &= \theta\mathbb{E}\left[R_k(\boldsymbol{\alpha}_x^{\mathbf{H}}, \boldsymbol{\beta}_x^{\mathbf{H}}, \mathbf{p}_x^{\mathbf{H}}; \mathbf{H})\right] \\ &\quad + (1-\theta)\mathbb{E}\left[R_k(\boldsymbol{\alpha}_y^{\mathbf{H}}, \boldsymbol{\beta}_y^{\mathbf{H}}, \mathbf{p}_y^{\mathbf{H}}; \mathbf{H})\right] \\ &\ge \theta\bar{r}_{x,k} + (1-\theta)\bar{r}_{y,k}. \end{aligned} \tag{40}$$

In conclusion, Problem **P2** can be said to satisfy the time-sharing condition, resulting in no duality gap.

REFERENCES

[1] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[2] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, 4th Quart., 2018.

[3] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.

[4] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 615–637, Mar. 2020.

[5] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, 2nd Quart., 2018.

[7] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[8] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[9] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.
[10] S. Asaad, A. M. Rabiei, and R. R. Muller, "Massive MIMO with antenna selection: Fundamental limits and applications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8502–8516, Dec. 2018.
[11] C. Jiang and L. J. Cimini, "Antenna selection for energy-efficient MIMO transmission," *IEEE Wireless Commun. Lett.*, vol. 1, no. 6, pp. 577–580, Dec. 2012.
[12] A. Ghrayeb, "A survey on antenna selection for MIMO communication systems," in *Proc. Int. Conf. Inf. Commun. Technol.*, Apr., vol. 2006, pp. 2104–2109.
[13] J.-C. Chen, "Joint antenna selection and user scheduling for massive multiuser MIMO systems with low-resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1019–1024, Jan. 2019.
[14] Y. Gao, H. Vinck, and T. Kaiser, "Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1346–1360, Mar. 2018.
[15] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "Antenna selection for MIMO-NOMA networks," in *Proc. IEEE ICC*, May 2017, pp. 1–6.
[16] M. O. K. Mendonca, P. S. R. Diniz, T. N. Ferreira, and L. Lovisolo, "Antenna selection in massive MIMO based on greedy algorithms," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1868–1881, Mar. 2020.
[17] J. Choi, J. Sung, N. Prasad, X.-F. Qi, B. L. Evans, and A. Gatherer, "Base station antenna selection for low-resolution ADC systems," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1951–1965, Mar. 2020.
[18] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using Monte Carlo tree search," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5380–5390, Oct. 2019.
[19] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Multi-switch for antenna selection in massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
[20] Y. Liu, Y. Zhang, C. Ji, W. Q. Malik, and D. J. Edwards, "A low-complexity receive-antenna-selection algorithm for MIMO–OFDM wireless systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2793–2802, Jul. 2009.
[21] Y. Zhang, C. Ji, W. Malik, D. O'Brien, and D. Edwards, "Receive antenna selection for MIMO systems over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4393–4399, Sep. 2009.
[22] W.-H. Chung and C.-Y. Hung, "Multi-antenna selection using space shift keying in MIMO systems," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–5.
[23] J. Joung and S. Sun, "Two-step transmit antenna selection algorithms for massive MIMO," in *Proc. IEEE ICC*, May 2016, pp. 1–6.
[24] M. Benmimoune, E. Driouch, W. Ajib, and D. Massicotte, "Joint transmit antenna selection and user scheduling for massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 381–386.
[25] M. Olyaee, M. Eslami, and J. Haghighat, "An energy-efficient joint antenna and user selection algorithm for multi-user massive MIMO downlink," *IET Commun.*, vol. 12, no. 3, pp. 255–260, Nov. 2018.
[26] S. Maimaiti, G. Chuai, W. Gao, K. Zhang, X. Liu, and Z. Si, "A low-complexity algorithm for the joint antenna selection and user scheduling in multi-cell multi-user downlink massive MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–14, Aug. 2019.
[27] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
[28] H.-S. Lee and J.-W. Lee, "QC2LinQ: QoS and channel-aware distributed link scheduler for D2D communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8565–8579, Dec. 2016.
[29] H.-S. Lee and J.-W. Lee, "EHLinQ: Distributed scheduler for D2D communication with RF energy harvesting," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2281–2292, Jun. 2020.
[30] K.-W. Kim, H.-S. Lee, and J.-W. Lee, "Opportunistic waveform scheduling for wireless power transfer with multiple devices," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5651–5665, Sep. 2020.
[31] 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 VTC-Spring*, Apr. 2021, pp. 1–5.
[32] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. EUSIPCO*, Sep. 2014, pp. 76–80.
[33] H. Li, M. Li, and Q. Liu, "Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 602–614, Jan. 2020.
[34] L. Pan, L. Liang, W. Xu, and X. Dong, "Framework of channel estimation for hybrid analog-and-digital processing enabled massive MIMO communications," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3902–3915, Sep. 2018.
[35] D. Fan, F. Gao, Y. Liu, Y. Deng, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.
[36] K. Xu, Z. Shen, Y. Wang, and X. Xia, "Location-aided mMIMO channel tracking and hybrid beamforming for high-speed railway communications: An angle-domain approach," *IEEE Syst. J.*, vol. 14, no. 1, pp. 93–104, Mar. 2020.
[37] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
[38] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
[39] Z. Chen, J. Li, and J. Huang, "Downlink multi-user scheduling with zero-forcing precoding in cognitive Hetnets: From performance tradeoff perspective," *IEEE Access*, vol. 6, pp. 50131–50141, Sep. 2018.
[40] S. Huang, H. Yin, H. Li, and V. C. M. Leung, "Decremental user selection for large-scale multi-user MIMO downlink with zero-forcing beamforming," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 480–483, Oct. 2012.
[41] S. Jin, X. Wang, Z. Li, K.-K. Wong, Y. Huang, and X. Tang, "On massive MIMO zero-forcing transceiver using time-shifted pilots," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 59–74, Jan. 2016.
[42] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.
[43] R. P. Antonioli, G. Fodor, P. Soldati, and T. F. Maciel, "User scheduling for sum-rate maximization under minimum rate constraints for the MIMO IBC," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1591–1595, Dec. 2019.
[44] R. Vaze and R. W. Heath, "Transmission capacity of ad-hoc networks with multiple antennas using transmit stream adaptation and interference cancellation," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 780–792, Feb. 2012.
[45] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[46] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York, NY, USA: Springer, 2011.
[47] W. Yu and J. M. Cioffi, "On constant power water-filling," in *Proc. IEEE ICC*, Jun. 2001, pp. 1665–1669.
[48] 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.
[49] D. Jungnickel and D. Jungnickel, *Graphs, Networks and Algorithms*. New York, NY, USA: Springer-Verlag, 2005.
[50] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part II: Achievable rate characterization," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2947–2959, Jun. 2018.
[51] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO HetNets," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2049–2065, May 2016.
[52] B. M. Lee, "Energy-efficient operation of massive MIMO in industrial Internet-of-Things networks," *IEEE Internet Things J.*, vol. 8, no. 9, pp. 7252–7269, May 2021.
[53] 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.

YUAN-XIN ZHU received the B.S. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2018, and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2021. Since March 2021, he has been with the AI Cloud Group, Quality Assurance Department, Baidu Inc., Beijing, China. His research interests include antenna selection, user scheduling, and quality-of-service control in wireless communication networks.

DO-YUP KIM (Graduate Student Member, IEEE) received the B.S. degree *(summa cum laude)* in electronics and communications engineering from Kwangwoon University, Seoul, South Korea, in 2016. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with Yonsei University, Seoul.

His research interests include resource allocation and scheduling, quality-of-service control, optimization, machine learning, and performance analysis in communication networks, mobile and flying ad hoc networks, and simultaneous wireless information and power transfer systems.

JANG-WON LEE (Senior Member, IEEE) received the B.S. degree in electronic engineering from Yonsei University, Seoul, South Korea, in 1994, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1996, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2004. From 2004 to 2005, he was a Postdoctoral Research Associate with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. Since September 2005, he has been with the School of Electrical and Electronic Engineering, Yonsei University, where he is currently a Professor. His research interests include communication networks, smart grid, and machine learning. He is an Editor of IEEE/ACM Transactions on Networking, *Journal of Communications and Networks*, and *ICT Express*.