Non-Orthogonal Multiple Access in Distributed Antenna Systems for Max-Min Fairness and Max-Sum-Rate

DONGJAE KIM[1], (Member, IEEE), AND MINSEOK CHOI[ID][2], (Member, IEEE)
[1]Artificial Intelligence Convergence Research Center for Regional Innovation, Korea Maritime and Ocean University, Busan 49112, South Korea
[2]Department of Telecommunication Engineering, Jeju National University, Jeju 63241, South Korea

Corresponding author: Minseok Choi (ejaqmf@jejunu.ac.kr)

ABSTRACT The distributed antenna systems (DAS) and non-orthogonal multiple access (NOMA) are the key technologies to boost the data rate of the cellular system in terms of small cell and multiple access, respectively. To meet the high data rate requirements for 5G and beyond, we suggest a framework of using NOMA in DAS. In the proposed scheme, RRUs which are geographically distributed in the cell serve the cell-edge users within their own coverage. Meanwhile, the macro BS, which covers the entire cell region with relatively high transmit power, supports both the cell-center and the cell-edge users with identical resources by using NOMA. Compared to the conventional DAS where the macro BS and the RRU serve the cell-center and the cell-edge users, respectively, the proposed framework also boosts the data rate of the cell-center user by improving the reliability of successive interference cancellation (SIC) at the cell-center user. In the proposed framework, this paper also proposes the optimal power allocation rules maximizing the user fairness in two different cases of instantaneous channel gain information (CGI) and channel distribution information (CDI) known at the transmitter. Also, the power allocation methods maximizing the sum-rate with a minimum rate constraint and the weighted sum-rate are presented in the case of CGI known at the transmitter. Simulation results show that the proposed framework of using NOMA in DAS can boost data rates more in a variety of system environments compared to the conventional NOMA or DAS.

INDEX TERMS Non-orthogonal multiple access (NOMA), distributed antenna systems (DAS), user fairness, sum-rate maximization, power allocation.

I. INTRODUCTION

Supporting a large number of users with high data rates is one of the most rewarding challenges for next generation wireless communications. A promising method to comply with this demand is to densely deploy small cells in a cellular network, which improves the area spectral efficiency [1]. The distributed antenna systems (DAS) has attracted considerable attention in both academia and industry to deploy the small cells in cellular networks [2]–[4]. In the DAS, multiple remote radio units (RRUs) are geographically distributed in a cell and cover subsets of the entire region of the macro cell. Cooperation of RRUs with the macro BS enables the DAS to have larger capacity [2] and to reduce inter-cell interference

and transmit power consumption [4], compared to the conventional co-located system. The advantages of the DAS in massive multiple-input multiple-output (MIMO) antenna systems are shown in [5]. Also, some studies show the potential for coexistence of DAS and other small cell techniques, e.g., picocells [6] and femtocells [7], in heterogeneous networks (HetNets). The additional diversity and scaling behavior of the massive DAS are also explored in [8].

In terms of multiple access, non-orthogonal multiple access (NOMA) is regarded as a promising candidate to improve the system throughput for 5G and beyond. Power-domain NOMA serves multiple users in the same time/frequency/code with different power levels relying on successive interference cancellation (SIC) performed at the receivers [9], [10]; therefore, NOMA provides higher spectral efficiency and system throughput compared to orthogonal

The associate editor coordinating the review of this manuscript and approving it for publication was Miguel López-Benítez[ID].

multiple access (OMA). Many recent studies have applied NOMA to various technologies including MIMO [11], wireless caching [12], relay networks [13], delay-sensitive networks [14], and blind modulation classification [15]. In addition, cooperative NOMA is proposed in [16], [17] where users can be used as relays to improve data rates.

Mostly, the existing researches have shown the advantages of DAS and NOMA separately; on the other hand, this paper proposed the framework of using NOMA in DAS and shows that synergy can be created by collaboration of NOMA and DAS. One of the example scenarios of our framework is a vehicle-to-everything (V2X) network. Throughput maximization and high spectral efficiency are key challenges for VANETs also due to proliferation of vehicular applications and crowded licensed spectrum [18]. The roadside units (RSUs) in the V2X network that collects traffic information and repeatedly provides services to vehicles can act as RRUs in the DAS [19], [20]. NOMA also provides the high spectral efficiency in high-mobility vehicular networks as explored in [21], [22]; therefore, the proposed framework of using NOMA in DAS would enhance the throughput of the vehicular network. In similar to cooperative transmissions for vehicular networks [23], this paper allows cooperation of RRUs with the macro BS to maximize the sum-rate as well as the user fairness.

A. CONTRIBUTIONS

This paper proposes the framework of using NOMA in DAS to boost the system throughput for both user fairness and sum-rate problems. The conventional DAS enhances data rates of cell-edge users with the help of nearby RRUs; however, in this case, orthogonal resources have to be allocated to RRUs and macro BS for mitigating interference among them. Therefore, motivated from the fact that relatively high power is available at the macro BS compared to RRUs so that the macro BS can cover the whole cell, we employ NOMA signaling at the macro BS while enjoying the advantages of cooperation from RRUs. Then, the macro BS and RRUs do not have to use orthogonal resources so that the system throughput can be improved more. The conventional DAS allows two different transmission schemes, *blanket transmission* and *selective transmission* [2]. Here, we focus on the selective transmission scheme because RRUs in the blanket transmission mode should transmit signals to all cell-edge users but the power budget of the RRU is limited; therefore, it is more reasonable to select one RRU under the best channel condition for improving the data rate of the cell-edge user.

The proposed framework using NOMA in DAS is strongly motivated by the fact that the cooperation signal from the RRU is helpful for the cell-center user as well as the cell-edge user. The cooperation signal from the RRU directly improves the data rate of the cell-edge user. In addition, the cooperation signal from the RRU can be used for the cell-center user to perform SIC much better with the appropriate power allocation. Therefore, we also propose the optimal power allocation for the proposed framework of using NOMA in DAS. The

proposed scheme can be applied to any system in which any entity can help cell-edge users and cooperate with the BS, e.g., HetNets where a macro BS and femto/pico BSs coexist.

The main contributions of this paper are as follows:

- This paper proposes the framework of using NOMA in DAS and observe the synergy effects created by collaboration of NOMA and DAS. The cooperation signal from the RRU can be used for the cell-center user to perform SIC much better in the proposed framework.
- We derive the closed-form expressions of data rates in the proposed framework of using NOMA in DAS. In addition, this paper proposes the optimal power allocation rules in a variety of problem settings: 1) max-min fairness, 2) sum-rate maximization with a minimum data rate constraint, and 3) weighted sum-rate maximization.
- The two different CSI at the transmitter (CSIT) cases are studied under the framework of using NOMA in DAS: 1) instantaneous channel gain information (CGI) known at the transmitter, and 2) channel distribution information CDI) known at the transmitter. More specifically, we solve the max-min fairness problem in both CSIT cases, and the max-sum-rate problem is handled in the case of CDI known at the transmitter. Also, the impacts of the imperfect CSI are studied when finding the optimal power allocation.
- The simulation results show that the proposed framework of using NOMA in DAS with the optimal power allocation rule can boost the system throughput more compared to conventional NOMA or DAS only. The impacts of the level of CSIT and imperfect CSI are also observed and the proposed framework still outperforms the conventional ones even with the lack of CSI.

B. RELATED WORKS

The sum-rate and user fairness are important metrics of NOMA, which have been studied actively in recent years. The ergodic sum-rate of MIMO-NOMA with two users is studied in [24], and the sum-rate performance of NOMA with randomly deployed users is analyzed in [25]. In [26], user scheduling for NOMA is studied to maximize the sum-rate. The user fairness issues of NOMA are discussed in [10] with instantaneous or statistical CSI at the BS. The power allocation for NOMA based on proportional fairness scheduling is studied for both max-sum-rate and max-min-rate problems in [27]. The joint optimization of power and channel allocations for NOMA is studied in [28] for various criteria including the fairness and the sum-rate. Overall, the above studies have shown that NOMA can improve fairness and sum-rate performances compared to OMA; however, they do not apply NOMA signaling to the DAS.

There have been some studies on NOMA in coordinated multipoint (CoMP) systems. In [29], NOMA is used in a coordinated two-point system to support a cell-edge user. The authors of [30] propose the opportunistic BS selection scheme for NOMA in a multi-cell CoMP system. While the

previous studies of [29], [30] focus on using NOMA in the system where macro BSs cooperate with each other under the fixed power allocation, this paper considers using NOMA in the DAS where RRUs (with relatively small power budget) and the macro BS (with relatively large power budget) coexist. Also, we propose the optimal power allocation rules for NOMA in the DAS in the cases of CGI and CDI known at the transmitter.

The authors of [31] propose a hybrid HetNet framework where small cells employ NOMA and the massive MIMO is deployed at the macro BS. This is different from our framework, where the macro BS employs NOMA and distributed RRUs send cooperation signals for supporting weak cell-edge users. The power controls for NOMA in HetNets are proposed in [32], [33]; however, each user is served by only one BS. This assumption is different from our model where users could be served by both the macro BS and RRUs.

C. ORGANIZATION AND NOTATIONS

This paper is organized as follows. Section II describes the system model including a cellular architecture of the DAS, the channel model, and NOMA signaling. In Section III, the closed-form expressions of data rates in our system model are derived. The optimal power allocation rules for the max-min fairness and the sum-rate maximization problems are proposed in Section IV and Section V, separately. Section VI investigates the effects of the imperfect CSI on the proposed framework and power allocation rules. The numerical results are shown in Section VII, and finally, Section VIII concludes the paper.

Notation: $\mathbf{X}^*$ and $\mathbf{X}^T$ denote conjugate and transpose of matrix $\mathbf{X}$, respectively. $\mathbb{E}[\cdot]$ denotes the expectation and $CN(\mu, \sigma^2)$ denotes complex Gaussian distribution with mean μ and variance σ^2.

II. SYSTEM MODEL

This section introduces the cellular architecture of the DAS and the framework of using NOMA in DAS. In addition, we describe the two different CSIT cases: 1) instantaneous CGI known at the transmitter and 2) only CDI known at the transmitter. The important notations describing the system model introduced in this section are summarized in Table 1.

A. CELLULAR ARCHITECTURE OF DISTRIBUTED ANTENNA SYSTEM

Consider a general multi-cell DAS illustrated in Fig. 1, where each cell includes a macro BS and multiple RRUs. The macro BS located at the center of the cell covers the whole macro cell while RRUs cover their own local areas including the cell-edge. Note that Fig. 1 shows an example scenario in which a cell consists of a macro BS and six RRUs and there are interfering cells around the home cell; however, the number of RRUs in a cell could depend on various system environments. Let P_m and P_r denote the power budgets of the macro BS and each RRU, respectively, and suppose that every RRU has the identical power budget of P_r. Then, the total power budget

TABLE 1. System parameters.

|	
P_m	Power budget of macro BS
P_r	Power budget of RRUs
S	Number of RRUs in each cell
L	Number of interfering cells
η	Pathloss exponent
$x_{i,j}$	Data symbol for user j transmitted from cell i
P_j	Power allocation of macro BS for $x_{i,j}$
r_j	Received signal at user j
f_j	Inter-cell interference to user j
n_j	Noise signal at user j
y_j	Alamouti decoded signal at user j
Z_j	Data rate for decoding the user 1's signal at user j
R_j	Data rate of the user j's signal
R_{sic}	Data rate threshold for reliable SIC
$d_{i,j}^{(k)}$	Distance from RRU i at cell k to user j at home cell
$L_{i,j}^{(k)}$	Pathloss from RRU i at cell k to user j at home cell
$h_{i,j}^{(k)}$	Channel gain between RRU i at cell k and user j at home cell
σ_n^2	Noise variance
$\sigma_{f_j}^2$	Variance of interference power at user j
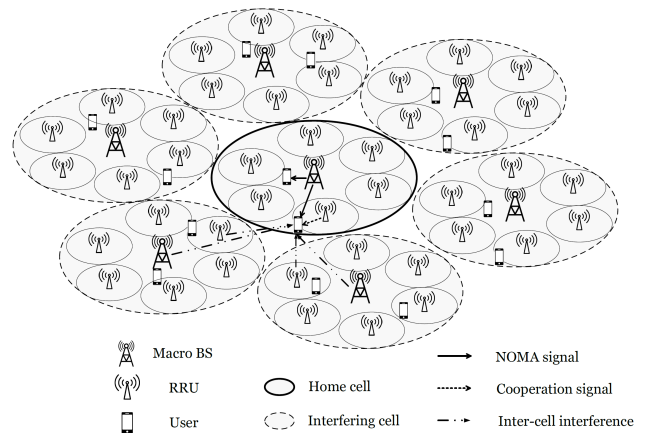
FIGURE 1. NOMA transmission in multi-cell DAS.

within a cell becomes $P = P_m + SP_r$, where S is the number of RRUs in each cell. In general, the macro BS has larger power than RRUs, i.e., $P_m > P_r$, and we assume that there are L dominant interfering cells around the home cell.

Within each cell, N users are divided into K clusters, and M users are grouped for each cluster, i.e., $N = MK$. We assume that the total bandwidth B is divided to K subbands and each cluster is served by using orthogonal subbands for mitigating the inter-cluster interference. In practical scenarios using NOMA, serving only two or three users in each subband is reasonable as in [26], because complexity and error propagation of SIC increase as the number of users sharing the identical band grows. The 3GPP LTE Advanced [34] also adopts pairing of two users (or four users optionally) for NOMA signaling. Accordingly, this paper considers a practical scenario with $M = 2$ which indicates $N = 2K$. The advanced user pairing methods in [26], [28] can be used

to form user clusters for NOMA signaling; however, user pairing is out of scope of this paper.

Within each subband, we suppose that one user is close to the macro BS (cell-center user) while another one is relatively far from the macro BS (cell-edge user). For simplicity of analysis, we consider a single cluster (i.e., two users) in each cell, but our analysis can be easily extended to a general multi-cluster scenario when the power budget for each cluster is sufficiently large. The frequency reuse factor is assumed to be one, which means that all the cells share the identical frequency band.

B. CHANNEL MODEL

Let the home cell be the center cell in which our target users are located described in Fig. 1. The Rayleigh fading channel from RRU i in cell k to user j in the home cell is denoted as $h_{i,j}^{(k)} = \sqrt{L_{i,j}^{(k)}} g_{i,j}^{(k)}$, where $L_{i,j}^{(k)}$ and $g_{i,j}^{(k)}$ are pathloss and fast fading components, respectively, for all $i \in \{0, 1, \cdots, S\}$, $j \in \{1, 2\}$ and $k \in \{0, 1, \cdots\}$. The macro BS is indexed as $i = 0$ and the home cell is indexed as $k = 0$. Here, the slow fading component $L_{i,j}^{(k)}$ is modeled as $L_{i,j}^{(k)} = 1/[d_{i,j}^{(k)}]^\eta$, where $d_{i,j}^{(k)}$ is the distance between RRU i in cell k to user j in the home cell and η is the pathloss exponent. The fast fading component follows the complex Gaussian distribution, i.e., $g_{i,j}^{(k)} \sim CN(0, 1)$. For simplicity of notation, we drop the cell index k of the home cell afterwards, i.e., $h_{i,j} = h_{i,j}^{(0)}$, $L_{i,j} = L_{i,j}^{(0)}$, and $g_{i,j} = g_{i,j}^{(0)}$.

We consider two CSIT cases: 1) instantaneous CGI and 2) only CDI known at the transmitter [35]. The CGI is the norm of the instantaneous channel gain, which can be obtained by using feedback from receivers. Meanwhile, the CDI is also known as the statistical CSI. In general, the user who experiences a better signal-to-interference-plus-noise ratio (SINR) is considered as the strong user, and another one becomes the weak user. Depending on the CSIT case, the macro BS determines whether the user is the strong one or the weak one. In the case of the CGI known at the transmitter, the instantaneous channel gain is used to determine the strong and the weak users; on the other hand, the expected channel power becomes a decision factor for determining the strong user when the CDI is known at the transmitter.

1) INSTANTANEOUS CGI KNOWN AT THE TRANSMITTER
When the transmitter knows the instantaneous CGI, user j' is determined as the strong user as follows:

$$j' = \arg \max_{j \in \{1,2\}} \frac{|h_{0,j}|^2}{\sigma_{f_j}^2 + \sigma_n^2}, \qquad (1)$$

where $\sigma_{f_j}^2$ denotes the interference power at user j which will be specified in Section III, and σ_n^2 is the noise power. Here, the strong user does not always imply the cell-center user because $L_{0,1} < L_{0,2}$ is not equivalent to $\frac{|h_{0,1}|^2}{\sigma_{f_1}^2 + \sigma_n^2} < \frac{|h_{0,2}|^2}{\sigma_{f_2}^2 + \sigma_n^2}$ due to the randomness of channel gains and the effect of inter-cell interference.

2) ONLY CDI KNOWN AT THE TRANSMITTER
In the case of CDI known at the transmitter, the following rule determines user j' as the strong user:

$$j' = \arg \max_{j \in \{1,2\}} \frac{L_{0,j}}{\sigma_{f_j}^2 + \sigma_n^2}. \qquad (2)$$

Here, the cell-edge user is more vulnerable to inter-cell interference than the cell-center user so that $L_{0,1} < L_{0,2}$ generally implies $\sigma_{f_1}^2 > \sigma_{f_2}^2$. Therefore, the above rule (2) can be reduced to $j' = \arg \max_{j \in \{1,2\}} L_{0,j}$. Thus, the cell-center and the cell-edge users always become the strong and the weak users, respectively.

For both CSIT cases, the weak user always decodes its signal directly, and the strong user decodes its own data after performing SIC. From now on, we denote the weak user as user 1, and the strong user as user 2. We first assume the perfect channel estimation, and the effects of imperfect CSIT will be studied in Section VI.

C. NONORTHOGONAL MULTIPLE ACCESS
The single selection scheme, where each user is served by one RRU or by the macro BS with the strongest channel condition, is one of the standard transmission schemes in the DAS [2]. This paper is also based on the single selection scheme; however, it is different that the macro BS supports both strong and weak users by using NOMA in our framework. In general, the macro BS has relatively larger power compared to RRUs, and it is available to employ NOMA signaling for supporting both users. Therefore, user 1 simultaneously receives the NOMA signal from the macro BS and the cooperation signal from the RRU having the strongest channel condition.

When the CGI is available at the transmitter, the RRU q is selected to serve user 1 where $q = \arg\max_{p \in \{1,2...,S\}} |h_{p,1}|^2$. On the other hand, in the case of CDI known at the transmitter, the RRU q serves user 1 where $q = \arg\max_{p \in \{1,2...,S\}} L_{q,1}$. Since the single RRU is selected, other RRUs could server other users with orthogonal bands in a general multi-user scenario. To summarize, we specify the roles of the macro BS and the RRU as follows: 1) The macro BS serves both weak and strong users by using NOMA, and 2) the RRU cooperates with the macro BS by serving the user in its own coverage as in the conventional single selection scheme in DAS.

Denote $x_{i,j}$ as the data symbol which is sent from RRU i to user j at the home cell, for all $i \in \{0, 1, \ldots, S\}$, and $j \in \{1, 2\}$. Again, note that the macro BS is indexed as $i = 0$. Then, when RRU q in the home cell is selected to serve user 1, the received signals of users 1 and 2 are given by

$$r_1 = h_{0,1}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + h_{q,1}\sqrt{P_r}x_{q,1} + f_1 + n_1 \qquad (3)$$

$$r_2 = h_{0,2}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + h_{q,2}\sqrt{P_r}x_{q,1} + f_2 + n_2 \qquad (4)$$

where r_j, f_j, n_j are the received signal, inter-cell interference, noise at user $j \in \{1, 2\}$ of the home cell, respectively. The macro BS allocates power levels of P_1 and P_2 to users 1 and

2, respectively, satisfying $P_m = P_1 + P_2$. The inter-cell interference to user j in the home cell is written as

$$f_j = \sum_{k=1}^{L} \Big[\underbrace{h_{0,j}^{(k)}\Big(\sqrt{P_1^{(k)}}x_{0,1}^{(k)} + \sqrt{P_2^{(k)}}x_{0,2}^{(k)}\Big)}_{\text{NOMA signals from cell } k}$$
$$+ \underbrace{h_{q_k,j}^{(k)}\sqrt{P_r}x_{q_k,1}^{(k)}}_{\text{RRU signal from cell } k} \Big] \quad (5)$$

where $P_1^{(k)}$ and $P_2^{(k)}$ are the allocated power levels of the BS for users in cell k, and $x_{i,j}^{(k)}$ is the symbol transmitted from cell k. Since the single selection scheme is adopted, only one RRU from each interfering cell shares the frequency band with users 1 and 2 of the home cell, and its index is denoted as $q_k \in \{1, 2, \ldots, S\}$ in the k-th cell. Suppose that $\mathbb{E}[|n_1|^2] = \mathbb{E}[|n_2|^2] = \sigma_n^2$ and $\mathbb{E}[|x_{i,j}|^2] = 1$ for all $i \in \{0, 1, \cdots, S\}$ and $j \in \{1, 2\}$. User 1 decodes $x_{0,1}$ and $x_{q,1}$ directly; on the other hand, user 2 performs SIC for canceling $x_{0,1}$ and $x_{q,1}$ from the received signal first, and then decodes $x_{0,2}$. This receiver process will be explained more clearly in Section III.

III. DATA RATE DERIVATION WITH THE ALAMOUTI SCHEME

In recent years, many studies applied the Alamouti scheme to NOMA systems to derive a data rate gain of NOMA as well as a diversity gain of the Alamouti scheme [29], [36], [37]. For cooperation of the macro BS and an RRU, we also adopt the Alamouti scheme and derive the data rate of each user. We let $x_{0,1}(1) = a(1)$, $x_{0,1}(2) = -a^*(2)$, $x_{q,1}(1) = a(2)$, $x_{q,1}(2) = a^*(1)$, $x_{0,2}(1) = b(1)$, $x_{0,2}(2) = b(2)$ where $x_{i,j}(t)$ is the data symbol at time t. The symbols a and b denote the data symbols for users 1 and 2, respectively.

The Alamouti scheme decodes two received symbols in consecutive time slots, i.e., $r_1(1)$ and $r_1(2)$, as follows: $[y_1(1)\ y_1(2)]^{\mathrm{T}} = \begin{bmatrix} h_{0,1}^*\sqrt{P_1} & h_{q,1}\sqrt{P_r} \\ h_{q,1}^*\sqrt{P_r} & -h_{0,1}\sqrt{P_1} \end{bmatrix}[r_1(1)\ r_1^*(2)]^{\mathrm{T}}$. Specifically, the decoded symbols of $y_1(1)$ and $y_1(2)$ are obtained as

$$y_1(1) = (|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)a(1) + |h_{0,1}|^2\sqrt{P_1P_2}b(1)$$
$$+h_{0,1}^*h_{q,1}\sqrt{P_2P_r}b^*(2) + h_{0,1}^*\sqrt{P_1}f_1(1)$$
$$+h_{q,1}\sqrt{P_r}f_1^*(2) + \tilde{n}_1(1) \quad (6)$$
$$y_1(2) = (|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)a(2) - |h_{0,1}|^2\sqrt{P_1P_2}b^*(2)$$
$$+h_{0,1}^*h_{q,1}\sqrt{P_2P_r}b(1) + h_{q,1}^*\sqrt{P_r}f_1(1)$$
$$-h_{0,1}^*\sqrt{P_1}f_1^*(2) + \tilde{n}_1(2). \quad (7)$$

and $\tilde{n}_1(t) \sim CN(0, (|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)\sigma_n^2)$. In (6), the signal component is $(|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)a(1)$ and the remaining components indicate intra-cell interference, inter-cell interference and noise components respectively, as follows:

$$y_1^{\text{sig}}(1) = (|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)a(1)$$
$$y_1^{\text{intra}}(1) = |h_{0,1}|^2\sqrt{P_1P_2}b(1) + h_{0,1}^*h_{q,1}\sqrt{P_2P_r}b^*(2)$$
$$y_1^{\text{inter}}(1) = h_{0,1}^*\sqrt{P_1}f_1(1) + h_{q,1}\sqrt{P_r}f_1^*(2)$$

$$y_1(1) = y_1^{\text{sig}}(1) + y_1^{\text{intra}}(1) + y_1^{\text{inter}}(1) + \tilde{n}_1(1).$$

Then, the SINR for decoding user 1's signal from $y_1(1)$ can be obtained as

$$SINR_1 = \mathbb{E}\left[\frac{|y_1^{\text{sig}}(1)|^2}{|y_1^{\text{intra}}(1) + y_1^{\text{inter}}(1) + \tilde{n}_1(1)|^2}\right] \quad (8)$$

$$= \frac{\mathbb{E}_a[|y_1^{\text{sig}}(1)|^2]}{\mathbb{E}_b[|y_1^{\text{intra}}(1)|^2] + \mathbb{E}_{f_j}[|y_1^{\text{inter}}(1)|^2] + \sigma_n^2} \quad (9)$$

$$= \frac{(|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)^2}{(|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r)(|h_{0,1}|^2P_2 + \sigma_{f_1}^2 + \sigma_n^2)} \quad (10)$$

$$= \frac{|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r}{|h_{0,1}|^2P_2 + \sigma_{f_1}^2 + \sigma_n^2}, \quad (11)$$

where $\mathbb{E}_a$ and $\mathbb{E}_b$ indicates the expectation with respect to data symbols a and b, and $\mathbb{E}_{f_j}$ is the expectation with respect to random variables in (5), i.e., $h_{0,j}^{(k)}$, $x_{0,1}^{(k)}$, $x_{0,2}^{(k)}$, and $x_{q_k,1}^{(k)}$ for all k. Since $h_{0,j}^{(k)}$, $x_{0,1}^{(k)}$, $x_{0,2}^{(k)}$, and $x_{q_k,1}^{(k)}$ are independent for all k, we can compute the expected value of $f_1(1)$ by taking averages of each component in (5) independently; therefore, the variance of inter-cell interference is obtained by

$$\sigma_{f_1}^2 = \sum_{k=1}^{L} \left(L_{0,1}^{(k)}P_m + L_{q_k,1}^{(k)}P_r \right). \quad (12)$$

Note that $f_1(1)$ and $f_1(2)$ have the same variance of $\sigma_{f_1}^2$.

Similarly, the SINR for decoding user 1's signal from $y_1(2)$ becomes exactly the same as (11). Accordingly, from both (6) and (7), the data rate for decoding its own signal at user 1 becomes

$$Z_1 = \log_2\left(1 + \frac{|h_{0,1}|^2P_1 + |h_{q,1}|^2P_r}{|h_{0,1}|^2P_2 + \sigma_{f_1}^2 + \sigma_n^2}\right). \quad (13)$$

Now we define the channel power to noise ratio as $\Gamma_{i,j} = \frac{|h_{i,j}|^2}{\sigma_n^2}$. Then, Z_1 can be rewritten as follows:

$$Z_1 = \log_2\left(1 + \frac{\Gamma_{0,1}P_1 + \Gamma_{q,1}P_r}{\Gamma_{0,1}P_2 + D_1}\right), \quad (14)$$

where $D_1 = \sigma_{f_1}^2/\sigma_n^2 + 1$.

The decoded signals at user 2 are similarly expressed by

$$[y_2(1)\ y_2(2)]^{\mathrm{T}} = \begin{bmatrix} h_{0,2}^*\sqrt{P_1} & h_{q,2}\sqrt{P_r} \\ h_{q,2}^*\sqrt{P_r} & -h_{0,2}\sqrt{P_1} \end{bmatrix}[r_2(1)\ r_2^*(2)]^{\mathrm{T}},$$
$$\quad (15)$$

and specifically,

$$y_2(1) = (|h_{0,2}|^2P_1 + |h_{q,2}|^2P_r)a(1) + |h_{0,2}|^2\sqrt{P_1P_2}b(1)$$
$$+h_{0,2}^*h_{q,2}\sqrt{P_2P_r}b^*(2) + h_{0,2}^*\sqrt{P_1}f_2(1)$$
$$+h_{q,2}\sqrt{P_r}f_2^*(2) + \tilde{n}_2(1) \quad (16)$$
$$y_2(2) = (|h_{0,2}|^2P_1 + |h_{q,2}|^2P_r)a(2) - |h_{0,2}|^2\sqrt{P_1P_2}b^*(2)$$
$$+h_{0,2}^*h_{q,2}\sqrt{P_2P_r}b(1) + h_{q,2}^*\sqrt{P_r}f_2(1)$$
$$-h_{0,2}^*\sqrt{P_1}f_2^*(2) + \tilde{n}_2(2), \quad (17)$$

where $\tilde{n}_2(t) \sim CN(0, (|h_{0,2}|^2 P_1 + |h_{q,2}|^2 P_r)\sigma_n^2)$. Similar to $\sigma_{f_1}^2$, inter-cell interference f_2 follows the zero-mean complex Gaussian distribution with the variance

$$\sigma_{f_2}^2 = \sum_{k=1}^{L} \left(L_{0,2}^{(k)} P_m + L_{qj,2}^{(k)} P_r \right). \tag{18}$$

From (16) and (17), the data rate for decoding the signal of user 1 at user 2, becomes

$$Z_2 = \log_2 \left(1 + \frac{\Gamma_{0,2} P_1 + \Gamma_{q,2} P_r}{\Gamma_{0,2} P_2 + D_2} \right), \tag{19}$$

where $D_2 = \sigma_{f_2}^2/\sigma_n^2 + 1$. Since the data symbols for user 1 should be decoded at both user sides, the data rate of the signal for user 1 can be finally written as

$$R_1 = \min(Z_1, Z_2), \tag{20}$$

when the instantaneous CGI is known at the transmitter.

Note that in conventional NOMA, $P_r = 0$; therefore, $Z_1 < Z_2$ is always satisfied so that $R_1 = Z_1$. On the other hand, $Z_1 < Z_2$ is not always guaranteed in our framework of using NOMA in the DAS, because of the cooperation signal from the RRU; therefore, we have to compare Z_1 and Z_2 for determining R_1 and it also gives an impact on finding the optimal power allocation. If only CDI is known at the transmitter, the expected data rates should be considered as follows: $\mathbb{E}[R_1] = \mathbb{E}[\min(Z_1, Z_2)]$.

After user 2 performs SIC to cancel $x_{0,2}$ and $x_{q,1}$ from $r_2(1)$ and $r_2(2)$, the desired symbols of user 2, i.e., $x_{0,1}$, can be decoded directly. Then, the data rate of user 2 becomes

$$R_2 = \log_2 \left(1 + \frac{\Gamma_{0,2} P_2}{D_2} \right) \tag{21}$$

in the case of the CGI known at the transmitter. With only CDI at the transmitter, the expected data rate has to be obtained as $\mathbb{E}[R_2] = \mathbb{E}[\log_2(1 + \frac{\Gamma_{0,2} P_2}{D_2})]$.

IV. MAX-MIN FAIRNESS PROBLEM

In this section, the optimal power allocation rule is proposed for the max-min fairness problem in our framework. The two different CSIT cases are studied: 1) instantaneous CGI and 2) only CDI known at the transmitter.

A. CGI KNOWN AT THE TRANSMITTER

Recall that $\Gamma_{0,1}/D_1 < \Gamma_{0,2}/D_2$ from (1) i.e., user 1 and user 2 are denoted as the weak and strong users, respectively. Let P_1^{opt} be the optimal power allocation for the weak user, and

$P_m - P_1^{opt}$ is allocated to the strong user. The max-min fairness problem is formulated as

$$\max_{P_1 \in [0, P_m]} \min\{R_1(P_1), R_2(P_1)\} \tag{22}$$

$$\text{subject to } Z_2(P_1) \geq R_{sic} \tag{23}$$

where R_{sic} is the data rate threshold to guarantee the reliability of SIC operation at user 2.

We first state Lemma 1, and the optimal power allocation that solves the problem of (22)–(23) can be obtained by using Lemma 1 as described in Theorem 1.

Lemma 1: R_1 is an increasing function of P_1, and R_2 is a decreasing function of P_1 for $P_1 \leq P_m$.

Proof: For $P_1 \leq P_m$, we obtain

$$\frac{\partial R_2}{\partial P_1} = \frac{1}{\ln 2} \frac{-\Gamma_{0,2}}{\Gamma_{0,2} P_2 + D_2} < 0, \tag{24}$$

$$\frac{\partial Z_1}{\partial P_1} = \frac{1}{\ln 2} \frac{\Gamma_{0,1}}{\Gamma_{0,1} P_2 + D_1} > 0, \tag{25}$$

$$\frac{\partial Z_2}{\partial P_1} = \frac{1}{\ln 2} \frac{\Gamma_{0,2}}{\Gamma_{0,2} P_2 + D_2} > 0. \tag{26}$$

Since Z_1 and Z_2 are increasing functions of P_1, the data rate of user 1, $R_1 = \min(Z_1, Z_2)$, is also an increasing function of P_1. ∎

Theorem 1: The optimal power allocation P_1^{opt} for the max-min fairness problem of (22)–(23) becomes

$$P_1^{opt} = \begin{cases} P_1^\star, & Z_2(P_1^\star) \geq R_{sic} \\ P_1^{sic}, & Z_2(P_1^\star) < R_{sic}, \; P_1^{sic} \in [0, P_m] \\ \text{outage}, & \text{otherwise} \end{cases} \tag{27}$$

where

$$P_1^\star = \begin{cases} P_{R_1}, & P_{R_1} \in [0, P_m] \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

$$P_1^{sic} = \frac{(2^{R_{sic}} - 1)(\Gamma_{0,2} P_m + D_2) - \Gamma_{q,2} P_r}{2^{R_{sic}} \Gamma_{0,2}} \tag{29}$$

and $P_{R_1} = \max(P_{Z_1}, P_{Z_2})$. P_{Z_1} and P_{Z_2} are given in (30) and (31), respectively, shown at the bottom of the page.

Proof: We first obtain $P_1^\star$ which is the solution without the SIC constraint (23). For any $P_1 \leq P_m$, $Z_1(P_m) > 0$, $Z_2 > 0$, $Z_1(-\frac{\Gamma_{q,1}}{\Gamma_{0,1}} P_r) = 0$ and $Z_2(-\frac{\Gamma_{q,2}}{\Gamma_{0,2}} P_r) = 0$ are satisfied; therefore, we have $R_1(P_m) > 0$ and $R_1(\max\{-\frac{\Gamma_{q,1}}{\Gamma_{0,1}} P_r, -\frac{\Gamma_{q,2}}{\Gamma_{0,2}} P_r\}) = 0$. Similarly, we also have $R_2(P_m) = 0$ and $\lim_{P_1 \to -\infty} R_2(P_1) = \infty$. It means that $R_1(P_1)$ and $R_2(P_1)$ functions of P_1 should intersect each other in $P_1 \in [0, P_m]$. Therefore, according to Lemma 1, $\min(R_1, R_2)$ is maximized, when $R_1 = R_2$ for any $P_1 \leq P_m$.

$$P_{Z_1} = P_m - \frac{1}{2\Gamma_{0,1}\Gamma_{0,2}} \cdot \left\{ -[\Gamma_{0,1} D_2 + \Gamma_{0,2} D_1] + \sqrt{[\Gamma_{0,1} D_2 + \Gamma_{0,2} D_1]^2 + 4\Gamma_{0,1}\Gamma_{0,2}(\Gamma_{0,1} P_m + \Gamma_{q,1} P_r) D_2} \right\} \tag{30}$$

$$P_{Z_2} = P_m - \frac{-D_2 + \sqrt{D_2^2 + (\Gamma_{0,2} P_m + \Gamma_{q,2} P_r) D_2}}{\Gamma_{0,2}} \tag{31}$$

Let $Z_1 = R_2$ and $Z_2 = R_2$ are satisfied when $P_1 = P_{Z_1}$ and $P_1 = P_{Z_2}$, respectively. By solving quadratic equations, the closed-form expressions of P_{Z_1} and P_{Z_2} are obtained as in (30) and (31), respectively. It can be easily shown that P_{R_1} is an unique solution of $R_1(P_1) = R_2(P_1)$ for $P_1 \leq P_m$.

Note that the optimal solution should be in the range of $[0, P_m]$, i.e., $P_1^{opt} \in [0, P_m]$. First, when $P_{R_1} \in [0, P_m]$, we just have $P_1^{opt} = P_{R_1}$. Second, if $P_{R_1} < 0$, according to Lemma 2, $R_2(a) \leq R_2(0) \leq R_2(P_{R_1})$ and $R_1(P_{R_1}) \leq R_1(0) \leq R_1(a)$ are satisfied for any $a \in [0, P_m]$. Accordingly, $\min(R_1(a), R_2(a)) \leq \min(R_1(0), R_2(0))$ for any $a \in [0, P_m]$ so that $P_1^{opt} = 0$. Lastly, the case of $P_{R_1} > P_m$ could not happen because $P_{Z_1}, P_{Z_2} < P_m$ from (30) and (31). Thus, without the SIC constraint of (23), the optimal solution can be represented as (28).

Now, consider the SIC constraint of (23), i.e., $Z_2 \geq R_{sic}$. Let the solution to $Z_2 = R_{sic}$ as P_1^{sic}, which is obtained as (29). If $Z_2(P_1^\star) \geq R_{sic}$, the optimal solution is still $P_1^\star$. On the other hand, if $Z_2(P_1^\star) < R_{sic}$ and $P_1^{sic} \in [0, P_m]$, we have $P_1^\star < P_1^{sic}$ and P_1^{sic} becomes the optimal power allocation because max-min(R_1, R_2) is a decreasing function of P_1 for $P_1 > P_1^\star$. Otherwise, the SIC constraint cannot be satisfied and the system outage occurs. Thus, the optimum solution becomes (27). ∎

Remark: Power allocated to user 1 should be at least greater than or equal to the threshold value of P_1^{sic} to guarantee the reliability of SIC by satisfying (23). It can be seen from (29) that P_1^{sic} increases as R_{sic} grows so that the optimal solution P_1^{opt} also increases as shown in (27).

B. CDI KNOWN AT THE TRANSMITTER

Recall that $\frac{L_{0,1}}{\sigma_{f_1}^2 + \sigma_n^2} < \frac{L_{0,2}}{\sigma_{f_2}^2 + \sigma_n^2}$ from (2), and we know the order of distances from the macro BS to users when the CDI is known at the transmitter. Therefore, the cell-center and the cell-edge users are considered as the strong and weak users, respectively. In this scenario, we would like to maximize the minimum expected data rate among users. Again, let P_1^{opt} be the optimal power allocation for user 1 (i.e., the weak user). The max-min fairness problem is then formulated as follows:

$$\max_{P_1 \in [0, P_m]} \min\{\mathbb{E}[R_1(P_1)], \mathbb{E}[R_2(P_1)]\} \tag{32}$$

$$\text{subject to } \mathbb{E}[Z_2(P_1)] \geq R_{sic}. \tag{33}$$

After performing SIC, the average data rate of user 1 (i.e., strong user) is written as $\mathbb{E}[R_2] = \mathbb{E}[\log_2(1 + \frac{\Gamma_{0,2}P_2}{D_2})]$. Define $C_l(x)$ as the ergodic capacity of an i.i.d MISO channel with l transmit antennas given by [42]

$$C_l(x) = \frac{e^{1/x}}{\ln 2} \sum_{k=0}^{l-1} E_{k+1}\left(\frac{1}{x}\right) \tag{34}$$

where $E_n(x) = \int_1^\infty \frac{e^{-xt}}{t^n} dt$ is an exponential integral. According to [42], the closed-form expression of the expected data rate for user 2 can be obtained as

$$\mathbb{E}[R_2] = C_1\left(\frac{L_{0,2}P_2}{\sigma_{f_2}^2 + \sigma_n^2}\right). \tag{35}$$

Similar to (20), the expected data rate of user 1 can be written as

$$\mathbb{E}[R_1] = \mathbb{E}[\min\{Z_1, Z_2\}]. \tag{36}$$

In the conventional NOMA, the closed-form expression of (36) can be obtained by introducing a parameter which is the minimum of the channel gain of users, as in [43]. However, in this work, the derivation in [43] cannot be directly applied because of the cooperation signal sent from the RRU. Instead, we define the upper bound on $\mathbb{E}[R_1]$, denoted by $\mathbb{E}[R_1^{UB}]$, as

$$\mathbb{E}[\min\{Z_1, Z_2\}] \leq \mathbb{E}[R_1^{UB}] \triangleq \min\{\mathbb{E}[Z_1], \mathbb{E}[Z_2]\}. \tag{37}$$

The expected data rate for detecting the desired signal at user 1 is obtained by

$$\mathbb{E}[Z_1] = \mathbb{E}\left[\log_2\left(1 + \frac{|h_{0,1}|^2 P_1 + |h_{q,1}|^2 P_r}{|h_{0,1}|^2 P_2 + \sigma_{f_1}^2 + \sigma_n^2}\right)\right] \tag{38}$$

$$= \mathbb{E}\left[\log_2\left(1 + \frac{|h_{0,1}|^2 P_m + |h_{q,1}|^2 P_r}{\sigma_{f_1}^2 + \sigma_n^2}\right)\right]$$

$$- \mathbb{E}\left[\log_2\left(1 + \frac{|h_{0,1}|^2 P_2}{\sigma_{f_1}^2 + \sigma_n^2}\right)\right]$$

$$= \frac{L_{0,1}P_m}{L_{0,1}P_m - L_{q,1}P_r} C_1\left(\frac{L_{0,1}P_m}{\sigma_{f_1}^2 + \sigma_n^2}\right) - C_1\left(\frac{L_{0,1}P_2}{\sigma_{f_1}^2 + \sigma_n^2}\right)$$

$$+ \frac{L_{q,1}P_r}{L_{q,1}P_r - L_{0,1}P_m} C_1\left(\frac{L_{q,1}P_r}{\sigma_{f_1}^2 + \sigma_n^2}\right). \tag{39}$$

Similarly, $\mathbb{E}[Z_2]$ can be derived as

$$\mathbb{E}[Z_2] = \mathbb{E}\left[\log_2(1 + \frac{|h_{0,2}|^2 P_1 + |h_{q,2}|^2 P_r}{|h_{0,2}|^2 P_2 + \sigma_{f_1}^2 + \sigma_n^2})\right] \tag{40}$$

$$= \frac{L_{0,2}P_m}{L_{0,2}P_m - L_{q,2}P_r} C_1\left(\frac{L_{0,2}P_m}{\sigma_{f_1}^2 + \sigma_n^2}\right) - C_1\left(\frac{L_{0,2}P_2}{\sigma_{f_1}^2 + \sigma_n^2}\right)$$

$$+ \frac{L_{q,2}P_r}{L_{q,2}P_r - L_{0,0}^{(2)}P_m} C_1\left(\frac{L_{q,2}P_r}{\sigma_{f_1}^2 + \sigma_n^2}\right) \tag{41}$$

Now, we would like to maximize the upper bound on the objective function in (32), which is

$$\min\{\mathbb{E}[R_1(P_1)], \mathbb{E}[R_2(P_1)]\} \leq \min\{\mathbb{E}[R_1^{UB}], \mathbb{E}[R_2]\}. \tag{42}$$

Denote the power allocation that maximizes the upper bound of (42) as P_1^{UB}. Although P_1^{UB} is not optimal for the original max-min problem of (32)–(33), we use this suboptimal power allocation in simulation because the closed-form expression of $\mathbb{E}[R_1]$ is difficult to obtain; nevertheless, Section VII shows that the power allocation rule based on the relaxed problem still achieves a sufficiently large performance gain compared to the existing schemes. In order to obtain P_1^{UB}, we first state the following lemma.

Lemma 2: $\mathbb{E}[R_1^{UB}]$ and $\mathbb{E}[R_1]$ are increasing functions of P_1. $\mathbb{E}[R_2]$ is a decreasing function of P_1.

Proof: It is already shown that Z_1 and Z_2 are increasing functions and R_2 is a decreasing function of P_1 in the proof of Lemma 1. The expectation operation does not affect this

proof; therefore, $\mathbb{E}[Z_1]$ and $\mathbb{E}[Z_2]$ are increasing functions and $\mathbb{E}[R_2]$ is a decreasing function of P_1, which completes the proofs for $\mathbb{E}[R_1^{UB}]$ and $\mathbb{E}[R_2]$.

According to Lemma 1, $\min\{Z_1, Z_2\}$ is an increasing function of P_1. We can rewrite $R_1(\mathbf{h}, P_1) = \min\{Z_1, Z_2\}$, where $\mathbf{h}$ is a vector consisting of all channel gains. Let the probability density function of $\mathbf{h}$ be $f_{\mathbf{H}}(\mathbf{h})$. Since R_1 is an increasing function of P_1, $R_1(\mathbf{h}, y) - R_1(\mathbf{h}, x) > 0$ for any non-negative $y > x$. Then, the following inequality is still satisfied:

$$\mathbb{E}[R_1(\mathbf{h}, y)] - \mathbb{E}[R_1(\mathbf{h}, x)]$$
$$= \int_{f_{\mathbf{H}}(\mathbf{h})>0} f_{\mathbf{H}}(\mathbf{h})[R_1(\mathbf{h}, y) - R_1(\mathbf{h}, x)]d\mathbf{h} > 0, \quad (43)$$

which completes the proof for R_1. ∎

If the SIC constraint (33) is not considered, according to Lemma 2, P_1^{UB} is obtained when $\mathbb{E}[R_2] = \min\{\mathbb{E}[Z_1], \mathbb{E}[Z_2]\}$ and $P_1^{UB} \in [0, P_m]$. However, a closed-form solution of $\mathbb{E}[R_2] = \min\{\mathbb{E}[Z_1], \mathbb{E}[Z_2]\}$ is difficult to obtain due to the integration operation in $\mathbb{E}[R_2]$, $\mathbb{E}[Z_1]$, and $\mathbb{E}[Z_2]$. Since $\mathbb{E}[R_1^{UB}]$ is an increasing function and $\mathbb{E}[R_2]$ is a decreasing function of P_1, the optimal solution P_1^{UB} can be directly obtained by using a bisection algorithm. Considering the SIC constraint, the optimal power allocation is presented in the above **Algorithm**. **Algorithm** includes the case where $\mathbb{E}[R_2]$ cannot reach $\mathbb{E}[R_1]$ for any $P_1 \in [0, P_m]$, which means that the optimal solution becomes $P_1^{UB} = 0$. Finally, the upper bound of the max-min fairness becomes $\min\{\mathbb{E}[R_1^{UB}(P_1^{UB})], \mathbb{E}[R_2(P_1^{UB})]\}$.

Note that although the bisection search has been generally used for finding the power allocations in NOMA systems [10], [46], our **Algorithm** is the first one which is applied to NOMA-assisted DASs. The computational complexity of the bisection search in **Algorithm** depends on ϵ which controls the trade-off between the accuracy and the complexity of **Algorithm**. When n iterations are in progress, the difference

---

Algorithm Bisection Method for the Problem of (32)–(33)

1: Initialize $u_1 = 0$, $v_1 = P_m$, $u_2 = 0$, $v_2 = P_m$
2: **while** $v_1 - u_1 \geq \epsilon$
3:　$P_1 = v_1 + u_1/2$
4:　**if** $\mathbb{E}[R_1(P_1)] < \mathbb{E}[R_2(P_1)] : u_1 = P_1$
5:　**else** $v_1 = P_1$
6: **end while**
7: $P_1^\star = P_1$
8: **while** $v_2 - u_2 \geq \epsilon$
9:　$P_1 = v_2 + u_2/2$
10:　**if** $\mathbb{E}[Z_2(P_1)] < R_{sic} : u_2 = P_1$
11:　**else** $v_2 = P_1$
12: **end while**
13: $P_1^{sic} = P_1$
14: **if** $\mathbb{E}[Z_2(P_1^{sic})] < R_{sic}$: Outage
15: **elseif** $P_1^{sic} > P_1^\star : P_1^{UB} = P_1^{sic}$
16: **else** $P_1^{UB} = P_1^\star$

---

between u and v is $P_m(0.5)^n$; therefore, the bisection search ends when $n \geq \log_2(P_m/\epsilon)$.

V. MAX-SUM-RATE PROBLEM
In this section, the sum-rate maximization problem with a minimum rate constraint and the weighted sum-rate maximization problem are studied, when the CGI is known at the transmitter.

A. SUM-RATE MAXIMIZATION WITH A MINIMUM RATE CONSTRAINT
The sum-rate maximization problem can be formulated as

$$\max_{P_1 \in [0, P_m]} R_1(P_1) + R_2(P_1) \quad (44)$$

$$\text{subject to } \min\{R_1(P_1), R_2(P_1)\} \geq R_t, \quad (45)$$

$$Z_2(P_1) \geq R_{sic} \quad (46)$$

where R_t is the minimum data rate constraint and R_{sic} is the data rate threshold for guaranteeing the reliability of SIC. We first state the following lemma.

Lemma 3: The sum-rate of $R_1 + R_2$ is a non-increasing function of P_1 for $P_1 \leq P_m$:
Proof: Since $\frac{\Gamma_{0,1}}{D_1} < \frac{\Gamma_{0,2}}{D_2}$ from (1), we have

$$\frac{\partial Z_1}{\partial P_1} + \frac{\partial R_2}{\partial P_1} = \frac{1}{\ln 2} \frac{\Gamma_{0,1}/D_1 - \Gamma_{0,2}/D_2}{(\frac{\Gamma_{0,1}}{D_1}P_2 + 1)(\frac{\Gamma_{0,2}}{D_2}P_2 + 1)} < 0 \quad (47)$$

$$\frac{\partial Z_2}{\partial P_1} + \frac{\partial R_2}{\partial P_1} = 0 \quad (48)$$

for $P_1 \leq P_m$. Since $Z_1 + R_2$ and $Z_2 + R_2$ are both non-increasing functions of P_1, $\min\{Z_1, Z_2\} + R_2$ is also a non-increasing function of P_1. ∎

The outage event occurs when any user fails to achieve the minimum data rate. Based on Lemma 3, the following theorem provides the solution of (44)–(46).

Theorem 2: The optimal power allocation for maximizing the sum-rate with a minimum rate constraint becomes

$$P_1^{opt} = \begin{cases} P_1^\star, & P_1^\star \in [0, P_m], P_1^\star \leq P_{R_2} \\ 0, & P_1^\star < 0 \leq P_{R_2} \\ \text{outage}, & \text{otherwise} \end{cases} \quad (49)$$

where

$$P_{R_2} = P_m - \frac{(2^{R_t} - 1)D_2}{\Gamma_{0,2}}, \quad (50)$$

$$P_1^\star = \max(P_{Z_1}, P_{Z_2}, P_1^{sic}), \quad (51)$$

$$P_{Z_1} = \frac{(2^{R_t} - 1)(\Gamma_{0,1}P_m + D_1) - \Gamma_{q,1}P_r}{2^{R_t}\Gamma_{0,1}}, \quad (52)$$

$$P_{Z_2} = \frac{(2^{R_t} - 1)(\Gamma_{0,2}P_m + D_2) - \Gamma_{q,2}P_r}{2^{R_t}\Gamma_{0,2}}, \quad (53)$$

and P_1^{sic} is the same as in (29).
Proof: Let the solutions for $R_2(P_1) = R_t$, $Z_1(P_1) = R_t$ and $Z_2(P_1) = R_t$, be P_{R_2}, P_{Z_1} and P_{Z_2}, respectively. Then, P_{R_2}, P_{Z_1}, and P_{Z_2} can be obtained as (50), (52), and (53),

respectively. From Lemma 1, Z_1 and Z_2 are increasing functions of P_1, and R_2 is a decreasing function of P_1. Using these properties, the constraints of (45) and (46) can be converted as follows:

$$P_1 \geq P_{Z_1}, \ P_1 \geq P_{Z_2}, \ P_1 \leq P_{R_2}, \ P_1 \geq P_1^{sic}. \quad (54)$$

Therefore, the constraints of (45) and (46) are summarized as

$$\max(P_{Z_1}, P_{Z_2}, P_1^{sic}) = P_1^{\star} \leq P_1 \leq P_{R_2}. \quad (55)$$

If $P_1^{\star} > P_{R_2}$ is satisfied, the outage event occurs. When $P_1^{\star} \leq P_{R_2}$, the optimum solution P_1^{opt} can be obtained by considering 3 cases depending on the value of $P_1^{\star}$. The following three cases describe this step.

Case A: When $P_1^{\star} \in [0, P_m]$, according to Lemma 3, $R_1 + R_2$ is a non-increasing function of P_1; therefore, $P_1^{opt} = P_1^{\star}$ gives the maximum sum-rate.

Case B: When $P_1^{\star} < 0 \leq P_{R_2}$, the boundary condition $P_1 \geq 0$ should be tested. The minimum value of P_1 is picked from the interval of $[P_1^{\star}, P_{R_2}] \cap [0, P_m]$, which is $P_1^{opt} = 0$.

Case C: When $P_1^{\star} > P_m$ or $P_{R_2} < 0$, $[P_1^{\star}, P_{R_2}] \cap [0, P_m] = \phi$; therefore, there is no solution which implies the outage event. ∎

B. WEIGHTED SUM-RATE MAXIMIZATION

The weighted sum-rate maximization problem is formulated as

$$\max_{P_1 \in [0, P_m]} w_1 R_1(P_1) + w_2 R_2(P_1) \quad (56)$$

$$\text{subject to } Z_2(P_1) \geq R_{sic} \quad (57)$$

where w_1 and w_2 are the weight factors for users' data rates. The solution of the above problem can be found by dividing into the following two cases depending on w_1 and w_2.

Case A: $w_1 \leq w_2$

In this case, the objective function can be rewritten as $w_1 R_1 + w_2 R_2 = w_1(R_1 + R_2) + (w_2 - w_1)R_2$. Here, $R_1 + R_2$ and R_2 are non-increasing and decreasing functions of P_1, respectively. Accordingly, the objective function is also a non-increasing function of P_1; therefore, the optimal solution P_1^{opt} is obtained as

$$P_1^{opt} = \begin{cases} P_0^{sic}, & P_1^{sic} \leq P_m \\ \text{outage}, & \text{otherwise} \end{cases} \quad (58)$$

where $P_0^{sic} = \max\{P_1^{sic}, 0\}$.

Case B: $w_1 > w_2$

In this case, to find the optimal solution, we first obtain the region of P_1 that satisfies $Z_1 \leq Z_2$. Let the solution for $Z_1(P_1) = Z_2(P_1)$ be $P_1 = P_Z$. Then, P_Z can be obtained as

$$P_Z = \frac{\Gamma_{q,1}(\Gamma_{0,2}P_m + D_2) - \Gamma_{q,2}(\Gamma_{0,1}P_m + D_1)}{\Gamma_{0,2}(\Gamma_{q,1}P_r + D_1) - \Gamma_{0,1}(\Gamma_{q,2}P_r + D_2)} P_r. \quad (59)$$

From the equations (47) and (48), we can derive $\frac{\partial Z_1}{\partial P_1} < \frac{\partial Z_2}{\partial P_1}$ for $P_1 \leq P_m$. Accordingly, R_1 is expressed as

$$R_1(P_1) = \begin{cases} Z_2(P_1), & P_1 < P_Z \\ Z_1(P_1), & P_1 \geq P_Z \end{cases} \quad \text{for } P_1 \leq P_m. \quad (60)$$

Using the above equation, we can derive the following lemma.

Lemma 4: If $w_1 > w_2$, the weighted sum-rate of $w_1 R_1 + w_2 R_2$ is an increasing function of P_1 for $P_1 < P_Z$.

Proof: For $P_1 < P_Z$,

$$\frac{\partial(w_1 R_1 + w_2 R_2)}{\partial P_1} = w_1 \frac{\partial Z_2}{\partial P_1} + w_2 \frac{\partial R_2}{\partial P_1}$$

$$= (w_1 - w_2)\frac{\partial Z_2}{\partial P_1} > 0. \quad (61)$$

Since $w_1 - w_2 > 0$ and $\frac{\partial Z_2}{\partial P_1} > 0$ for $P_1 \leq P_m$, the derivative in (61) is always positive, which completes the proof. ∎

If $P_1 \geq P_Z$, the weighted sum-rate becomes $w_1 Z_1 + w_2 R_2$. In this case, the objective function is not monotonic, so we need to find the value that makes the derivative zero. Let P_R be the solution to the following equation,

$$w_1 \frac{\partial Z_1(P_R)}{\partial P_1} + w_2 \frac{\partial R_2(P_R)}{\partial P_1} = 0. \quad (62)$$

Then, P_R can be derived as

$$P_R = P_m - \frac{w_2 D_1/\Gamma_{0,1} - w_1 D_2/\Gamma_{0,2}}{w_1 - w_2}, \quad (63)$$

and it gives the following lemma.

Lemma 5: If $w_1 > w_2$, $w_1 Z_1 + w_2 R_2$ is increasing for $P_1 \leq P_R$ and decreasing for $P_R < P_1 \leq P_m$.

Proof: The derivative of $w_1 Z_1 + w_2 R_2$ is derived as

$$\frac{\partial(w_1 Z_1 + w_2 R_2)}{\partial P_1} = -\frac{1}{\alpha(P_1)}(w_1 - w_2)(P_1 - P_R), \quad (64)$$

where

$$\alpha(P_1) = \ln 2 \left(P_m - P_1 + \frac{D_1}{\Gamma_{0,1}}\right)\left(P_m - P_1 + \frac{D_2}{\Gamma_{0,2}}\right), \quad (65)$$

which is positive for $P_1 \leq P_m$. Therefore, the derivative is positive for $P_1 \leq P_R$ and negative for $P_R < P_1 \leq P_m$. ∎

Based on Lemma 4 and Lemma 5, the following theorem provides the optimal power allocation for $w_1 > w_2$.

Theorem 3: When w_1 is lager than w_2, the optimal power allocation of the weighted sum-rate problem becomes

$$P_1^{opt} = \begin{cases} P_1^{\star}, & P_1^{\star} \in [P_0^{sic}, P_m], \ P_1^{sic} \leq P_m \\ P_0^{sic}, & P_1^{\star} < P_0^{sic}, \ P_1^{sic} \leq P_m \\ P_m, & P_1^{\star} > P_m, \ P_1^{sic} \leq P_m \\ \text{outage}, & P_1^{sic} > P_m, \end{cases} \quad (66)$$

where $P_1^{\star} = \max(P_Z, P_R)$.

Proof: From Lemma 4 and Lemma 5, the weighted sum-rate is increasing for $P_1 \leq P_1^{\star}$ and decreasing for $P_1^{\star} < P_1 \leq P_m$. Thus, the weighted sum-rate is maximized when $P_1 = P_1^{\star}$. If the outage event does not occur, i.e., $P_1^{sic} \leq P_m$, the constraint of the problem (56) can be converted to $P_1 \in [P_0^{sic}, P_m]$. Therefore, the optimal power allocation is obtained as (66). ∎

VI. POWER ALLOCATION UNDER IMPERFECT CSIT

This section extends the CSI condition to the case of imperfect CSI. Here, we assume that the slow fading component $L_{i,j}$ is known at transmitter but the estimation of the fast fading gain is not exactly accurate. According to [47], the estimated channel $\hat{h}_{i,j}$ is modeled by

$$h_{i,j} = \hat{h}_{i,j} + z_{i,j} = \sqrt{L_{i,j}}(\hat{g}_{i,j} + \epsilon_{i,j}) \tag{67}$$

where $\epsilon_{i,j} \sim CN(0, \sigma_\epsilon^2)$ is the channel estimation error with σ_ϵ^2 being its variance, and $\hat{g}_{i,j}$ is the estimated fast fading component which is uncorrelated with $\epsilon_{i,j}$.

Applying (67) to (3) and (4), the received signals can be represented as

$$r_1 = \hat{h}_{0,1}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + \hat{h}_{q,1}\sqrt{P_r}x_{q,1}$$
$$+ z_{0,1}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + z_{q,1}\sqrt{P_r}x_{q,1} + f_1 + n_1 \tag{68}$$

$$r_2 = \hat{h}_{0,2}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + \hat{h}_{q,2}\sqrt{P_r}x_{q,1}$$
$$+ z_{0,2}(\sqrt{P_1}x_{0,1} + \sqrt{P_2}x_{0,2}) + z_{q,2}\sqrt{P_r}x_{q,1} + f_2 + n_2 \tag{69}$$

where the third and the fourth terms represent interference caused by channel estimation errors. Based on (68), the SINR for decoding user 1's signal becomes

$$SINR_1 = \frac{|\hat{h}_{0,1}|^2 P_1 + |\hat{h}_{q,1}|^2 P_r}{|\hat{h}_{0,1}|^2 P_2 + (L_{0,1}P_m + L_{q,1}P_r)\sigma_\epsilon^2 + \sigma_{f_1}^2 + \sigma_n^2}. \tag{70}$$

The data rate for decoding its own signal at user 1 becomes

$$Z_1 = \log_2\left(1 + \frac{\hat{\Gamma}_{0,1}P_1 + \hat{\Gamma}_{q,1}P_r}{\hat{\Gamma}_{0,1}P_2 + \bar{D}_1}\right), \tag{71}$$

where $\hat{\Gamma}_{i,j} = |\hat{h}_{i,j}|^2/\sigma_n^2$ and

$$\bar{D}_j = \frac{(L_{0,j}P_m + L_{q,j}P_r)\sigma_\epsilon^2 + \sigma_{f_j}^2}{\sigma_n^2}. \tag{72}$$

In the similar way, Z_2 and R_2 can be derived as

$$Z_2 = \log_2\left(1 + \frac{\hat{\Gamma}_{0,2}P_1 + \hat{\Gamma}_{q,2}P_r}{\hat{\Gamma}_{0,2}P_2 + \bar{D}_2}\right) \tag{73}$$

$$R_2 = \log_2\left(1 + \frac{\hat{\Gamma}_{0,2}P_2}{\bar{D}_2}\right). \tag{74}$$

By applying (71)–(74) to Section IV and V, the optimal power allocation rules maximizing the min-max rate and the sum-rate can be obtained even with the imperfect CSI.

VII. NUMERICAL RESULTS

This section shows the numerical results of the proposed power allocation methods for NOMA in DAS. The system model is based on Fig. 1, where the number of RRUs are assumed to be $S = 6$. The N_t-th tier of interfering cells around the home cell consists of $6N_t$ interfering cells. We consider
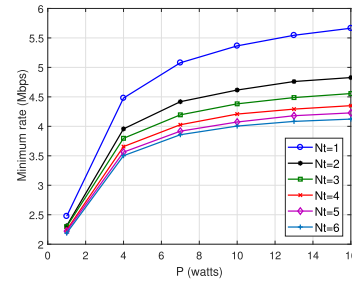

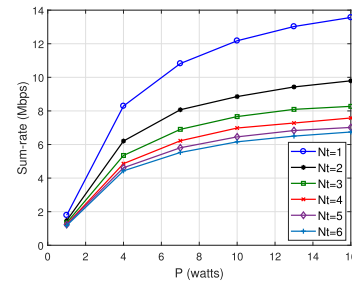**FIGURE 2.** The impacts of N_t on minimum rate with CGI.

FIGURE 3. The impacts of N_t on sum-rate with CGI.

the third-tier of interfering cells in simulation, i.e., $N_t = 3$, so that the total number of interfering cells is $L = 36$. The path loss exponent is $\eta = 2.5$ and the cell radius is 200 m. In Figs. 2 and 3, we can see that the effects of more than third-tier interfering cells on the minimum rate and the sum-rate of the home cell are very small; therefore $N_t = 3$ is a reasonable choice. For simplicity, suppose only one cluster, i.e., $K = 1$. In addition, $B = 5$ MHz, $N_0 = -90$ dBm/Hz, $R_{sic} = 0.5$ bps/Hz, $P_m = 40$ dBm and $P_r = 30$ dBm are used. For the general case, the cell-center user is uniformly distributed within the circle with a normalized radius of 0.3, and the cell-edge user is uniformly distributed within the ring which is the region between inner and outer circles whose normalized radii are 0.8 and 1.0. All of the above environments are assumed in this section unless otherwise noted. We compare the proposed technique with the following schemes:

- Conventional in DAS: The single selection scheme is assumed for this comparison technique, which means that a cell-center user is served by the macro BS and a cell-edge user is served by the RRU with the best channel gain only. In other words, each RRU or the macro BS serve the users in its own coverage.
- OMA in DAS: Similar to the proposed scheme, the macro BS serves both the cell-center and the cell-edge users but OMA signaling is used. As in [44], a half of the total bandwidth B is allocated to both users.
- Conventional NOMA: The DAS is not considered in this scheme. The macro BS with power budget of P serves two users by NOMA with appropriate power allocations without any RRU. For fair comparisons with other schemes using DAS, the macro BS of this scheme has the power budget of $P = P_m + SP_r$.

A. MAX-MIN FAIRNESS

1) INSTANTANEOUS CGI KNOWN AT THE TRANSMITTER

Fig. 4 shows plots of the minimum rate versus the normalized distance between the cell-center and the cell-edge user; for example, when the distance is the same as the cell radius, the normalized distance becomes one. Here, the cell-center user is located at normalized distance of 0.15. Recall that the cell-edge user in this case does not always indicate the weak user. In Fig. 4, the proposed scheme outperforms all the comparison techniques. Around 0.67 of the normalized distance, the peaks of minimum rates of the schemes using DAS appear because the RRU is 0.67 of the cell radius away from the macro BS so that the signal from the RRU is the strongest at this point. However, conventional NOMA does not have any RRU; therefore, its performance decreases with the normalized distance of the cell-edge user. We can see from Fig. 4 that the conventional DAS could be valuable when the cell-edge user is near to the RRU; however, as the cell-edge user becomes farther away from the RRU, the cooperation signal from the macro BS would be helpful to improve its data rate which can be seen from performances of the proposed scheme and OMA in DAS.

In Fig. 5, the minimum rates of comparison schemes are shown with respect to the total power budget P. Here, the normalized distance from the cell-edge user to the macro BS is between 0.8 and 1 of the cell radius; therefore, the schemes
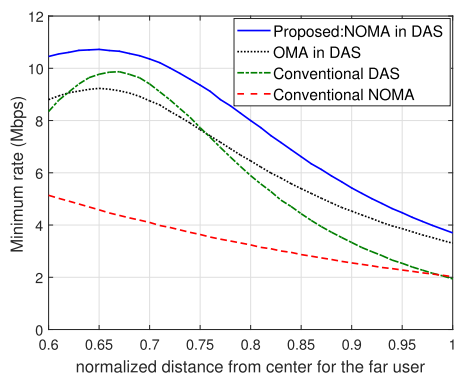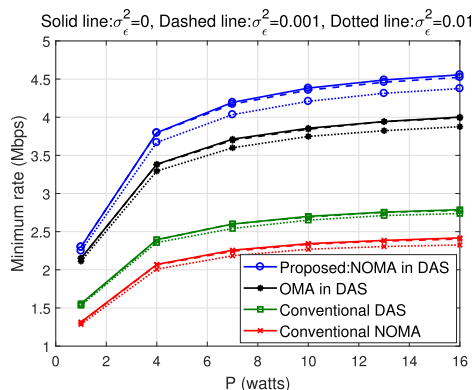
where both the macro BS and the RRU serve the cell-edge user are better than conventional DAS and conventional NOMA. The performance gap between the proposed scheme and OMA in DAS increases as P grows because the advantage of using NOMA for helping the cell-edge user rather than OMA becomes significant with large P. In addition, when P is sufficiently large, i.e., $P \geq 6$ Watts, increasing P does not give much advantage for max-min fairness because the inter-cell interference also increases.

The effects of imperfect CSI on the minimum rate are also shown in Fig. 5. The solid graphs represent the case of perfect CSI, and the dashed and the dotted graphs indicate performances obtained with imperfect channel estimation whose error variances $\sigma_\epsilon^2 = 0.001$ and $\sigma_\epsilon^2 = 0.01$, respectively. When the channel estimation error is relatively large, i.e., $\sigma_\epsilon^2 = 0.01$, the proposed scheme appears to be more sensitive to imperfect CSI than other comparison techniques; however, it still greatly outperforms others, and we can see that the proposed framework of using NOMA in DAS is quite robust to imperfect CSI while maintaining its advantages over existing systems.

2) ONLY CDI KNOWN AT THE TRANSMITTER

The minimum rate versus P when CDI is known at the transmitter is shown in Fig. 6. The upper bounds of the proposed scheme and conventional NOMA are derived by using the closed-form expressions of $\mathbb{E}[R_1^{UB}]$ and $\mathbb{E}[R_2]$ with power allocation of P_1^{UB} obtained in Section IV-B. The plots labeled as lower bound are obtained with power allocation of P_1^{UB}. Since P_1^{UB} is not the optimal solution of the max-min problem, it can be viewed as a lower bound of max-min($\mathbb{E}[R_1]$, $\mathbb{E}[R_2]$). Overall, the lower bound of the proposed scheme is much greater than other comparison techniques; therefore, the advantages of using both NOMA and DAS are still effective even with CDI at the transmitter only.
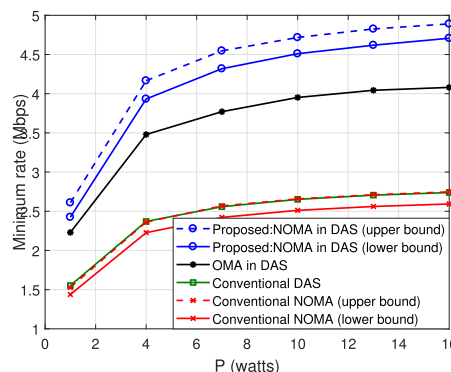

FIGURE 4. Minimum rate versus distance of the far user.

FIGURE 5. Minimum rate versus *P* with CGI.

FIGURE 6. Minimum rate versus *P* with CDI.

B. SUM-RATE MAXIMIZATION WITH A MINIMUM RATE CONSTRAINT

The plots of sum-rates are shown in Figs. 7 and 8. In these figures, we assume that the sum-rate becomes zero when the system outage occurs, i.e., $\min(R_1, R_2) < R_t$. Obviously,
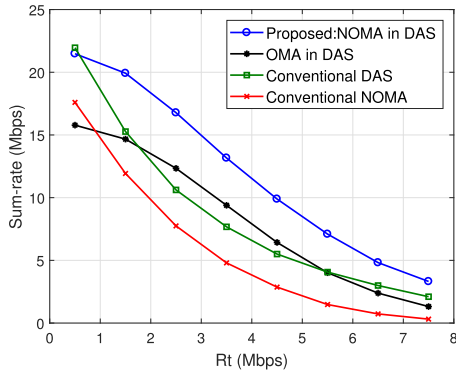
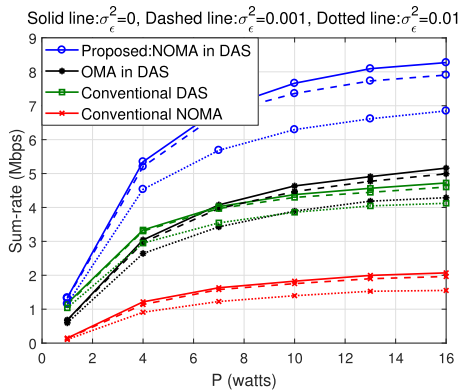**FIGURE 7.** Sum-rate versus R_t.

FIGURE 8. Sum-rate versus P.

as R_t grows, the outage event occurs more frequently and the sum-rates of all schemes decrease. We also mention that the proposed scheme has generally low outage probability which leads better sum-rates than other comparison techniques. Owing to its poor channel conditions, the cell-edge user would be more likely to experience the outage event compared to the cell-center user; therefore, the proposed scheme which greatly helps to improve the data rate of the cell-edge user provides the best sum-rate performance. Compared to the results of the max-min fairness in Section VII-A, conventional DAS provides quite comparable sum-rates to 'OMA in DAS'. The reason is that the cell-center user of 'OMA in DAS' achieve very large data rate by employing all the power budget of the macro BS of 'OMA in DAS' for serving the cell-center user.

In Fig. 8, $R_t = 1$ bps/Hz is assumed when the impacts of P are observed. Similar to Fig. 5, the proposed scheme shows much better sum-rate performances than other comparison techniques even with the imperfect CSI. The difference is that the sum-rate performance of conventional DAS is comparable to that of 'OMA in DAS', which is consistent with the results in Fig. 7. In addition, the sum-rate performance of the proposed scheme appears to be more sensitive to imperfect CSI compared to Fig. 5. This is because the sum-rate maximization problem in (44) has one more constraint of (45), compared to the max-min fairness problem in (22); therefore, outage events resulted from imperfect channel estimation occur more frequently.

C. WEIGHTED SUM-RATE MAXIMIZATION

Figs. 9 and 10 show the plots of the weighted sum-rates, where $w_1 + w_2 = 2$ is assumed. We can see that the proposed scheme outperforms the comparison techniques when w_1 which is a weight factor for the cell-edge user is large, i.e., $w_1 \geq 1.55$; however, when $w_1 < 1.55$, the conventional DAS shows the better weighted sum-rate performances than the proposed scheme. As w_1 grows, the importance of the data rate of the cell-edge user becomes more dominant; therefore, the proposed scheme becomes to provide better performances than others. On the other hand, when w_1 is small, the data rate of the cell-center user becomes more dominant, and it leads 'Conventional DAS' to achieve the better weighted sum-rate performance than the proposed scheme because the macro BS of 'Conventional DAS' employs all the power budget for serving the cell-center user. However, when $w_1 < 1.55$, the data rate of the cell-edge user is smaller than the minimum data rate constraint of the sum-rate maximization problem, i.e., $R_1 < R_t = 1$ bps/Hz; therefore, when R_t is sufficiently large, w_1 has to be larger than 1.55 and the proposed scheme would be the best solution.

Fig. 10 shows the plots of weighted sum-rates versus P, where the weights are assumed to be $w_1 = 1.7$, $w_2 = 0.3$. The results are consistent with Fig. 9, confirming that the advantages of using both NOMA and DAS are still shown.
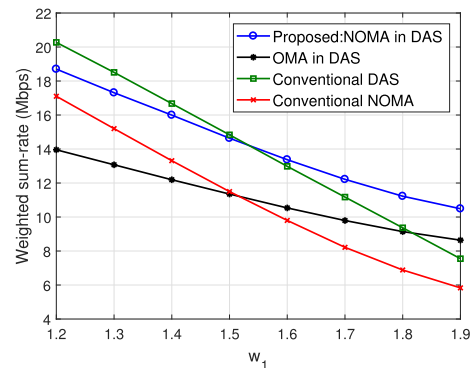

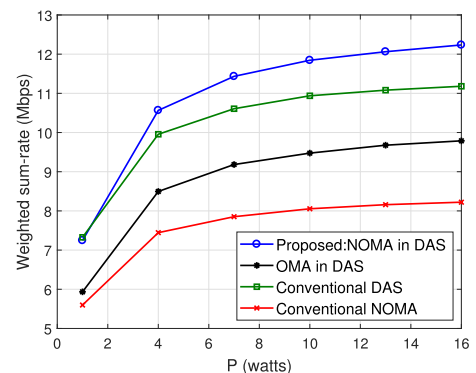
FIGURE 9. Weighted sum-rate versus w_1.

FIGURE 10. Weighted sum-rate versus P.

D. COMPOSITE FADING MODEL

Fig. 11 and Fig. 12 show that the proposed framework with the appropriate power allocation rules can provide the better minimum data rate and sum-rate compared to other comparison techniques even with the assumption of the $\kappa - \mu$ fading model [48]. By changing the values of κ and μ, we obtain performances of the proposed scheme and other comparison techniques. Here, κ means the ratio between the total power of the dominant components and the total power of the scattered waves, and μ means the number of clusters of multipath.
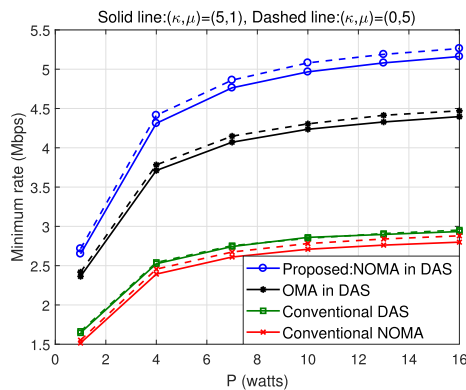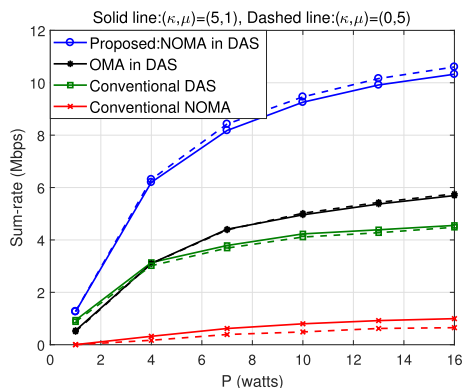

FIGURE 11. Minimum rate versus *P* for κ-μ fading channel.

FIGURE 12. Sum-rate versus *P* for κ-μ fading channel.

In Fig. 11, the minimum rate for each scheme increases with increasing κ and μ compared to Fig. 5 with $(\kappa, \mu) = (0, 1)$. This is because larger κ and μ values reduce the variance of the envelope $Var(|h|)$. The data rate uses the logarithmic function which is concave, so the smaller variance of the envelope, the higher data rate.

In Fig. 12, the proposed scheme and OMA in DAS outperform Fig. 8 with lower values of κ and μ. However, other schemes have similar or lower performances because they have higher probabilities of the outage event.

E. DISCUSSION

In summary, the proposed framework of using NOMA in DAS outperforms the comparison techniques, i.e., OMA in

DAS, conventional DAS, and conventional NOMA. Through numerical results, we verify that a combination of NOMA and DAS could create significant synergy for improving the user fairness and sum-rate performance. Especially, as the minimum data rate constraint becomes strict, the advantages of the proposed scheme increases. In addition, we can see that the proposed scheme shows significant improvements compared to existing techniques in both cases of CGI and CDI known at the transmitter. Finally, the proposed framework of using NOMA in DAS is also quite robust to the imperfect CSI.

VIII. CONCLUSION

This paper proposes the framework of using NOMA in DAS for improving data rates of cell-edge users. The key idea is that the macro BS with the larger power budget than RRUs serve both the cell-center and the cell-edge users and one of the RRUs transmits the cooperation signal to the cell-edge user. We also presents the optimal power allocation schemes in the proposed framework for the user fairness and sum-rate maximization problems in both cases of CGI and CDI known at the transmitter. Even though NOMA and DAS have been considered to efficiently use the frequency and spatial resources and to improve the users' data, simulation results show that a combination of NOMA and DAS shows much better user fairness and sum-rate performances than existing techniques.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] W. Choi and J. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, Jan. 2007.

[3] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, Apr. 2013.

[4] H. Kim, S.-R. Lee, C. Song, K.-J. Lee, and I. Lee, "Optimal power allocation scheme for energy efficiency maximization in distributed antenna systems," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 431–440, Feb. 2015.

[5] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 942–953, Oct. 2014.

[6] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo, and H. S. Dhillon, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, Jun. 2012.

[7] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[8] Z. Liu, "On the scaling behavior of the average rate performance of large-scale distributed MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4029–4043, May 2017.

[9] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[10] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[11] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[12] M. Choi, J. Kim, and J. Moon, "Wireless video caching and dynamic streaming under differentiated quality requirements," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1245–1257, Jun. 2018.

[13] C. Xue, Q. Zhang, Q. Li, and J. Qin, "Joint power allocation and relay beamforming in nonorthogonal multiple access amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7558–7562, Aug. 2017.

[14] M. Choi, J. Kim, and J. Moon, "Dynamic power allocation and user scheduling for power-efficient and delay-constrained multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4846–4858, Oct. 2019.

[15] M. Choi and J. Kim, "Blind signal classification analysis and impact on user pairing and power allocation in nonorthogonal multiple access," *IEEE Access*, vol. 8, pp. 100916–0100929, 2020.

[16] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[17] M. Choi, D.-J. Han, and J. Moon, "Bi-directional cooperative NOMA without full CSIT," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7515–7527, Nov. 2018.

[18] J. Nzouonta, N. Rajgure, G. Wang, and C. Borcea, "VANET routing on city roads using real-time vehicular traffic information," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3609–3626, Sep. 2009.

[19] Y. Li, X. Zhu, D. Jin, and D. Wu, "Multiple content dissemination in roadside-unit-aided vehicular opportunistic networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3947–3956, Oct. 2014.

[20] M. Choi, J. Kim, and J. Moon, "Adaptive detector selection for queue-stable word error rate minimization in connected vehicle receiver design," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3635–3639, Apr. 2018.

[21] H. Xiao, Y. Chen, S. Ouyang, and A. T. Chronopoulos, "Power control for clustering car-following V2X communication system with non-orthogonal multiple access," *IEEE Access*, vol. 7, pp. 68160–68171, 2019.

[22] M. Choi, D. Yoon, and J. Kim, "Blind signal classification for non-orthogonal multiple access in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9722–9734, Oct. 2019.

[23] M. Pan, P. Li, and Y. Fang, "Cooperative communication aware link scheduling for cognitive vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 760–768, May 2012.

[24] Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[25] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[26] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[27] 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.

[28] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.

[29] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[30] Y. Tian, A. R. Nix, and M. Beach, "On the performance of opportunistic NOMA in downlink CoMP networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 998–1001, May 2016.

[31] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and A. Nallanathan, "Non-orthogonal multiple access in massive MIMO aided heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[32] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, "Spectrum allocation and power control for non-orthogonal multiple access in HetNets," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5825–5837, Sep. 2017.

[33] A. Nasser, O. Muta, M. Elsabrouty, and H. Gacanin, "Interference mitigation and power allocation scheme for downlink MIMO–NOMA HetNet," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6805–6816, Jul. 2019.

[34] *Study on Downlink Multiuser Superposition Transmission for LTE*, document TSG RAN Meeting 67, 3GPP, RP-150496, Mar. 2015.

[35] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1188–1204, Mar. 2008.

[36] Z. Tang, J. Wang, J. Wang, and J. Song, "Harvesting both rate gain and diversity gain: Combination of NOMA with the Alamouti scheme," *IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2017, pp. 1–3.

[37] M. Toka and O. Kucur, "Non-orthogonal multiple access with alamouti space–time block coding," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1954–1957, Sep. 2018.

[38] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[39] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.

[40] H. Kim, S.-R. Lee, K.-J. Lee, and I. Lee, "Transmission schemes based on sum rate analysis in distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1201–1209, Mar. 2012.

[41] X. Zhang, Y. Sun, X. Chen, S. Zhou, J. Wang, and N. B. Shroff, "Distributed power allocation for coordinated multipoint transmissions in distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2281–2291, May 2013.

[42] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.

[43] J. Choi, "On the power allocation for MIMO-NOMA systems with layered transmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3226–3237, May 2016.

[44] Y. Fu, Y. Chen, and C. W. Sung, "Distributed power control for the downlink of multi-cell NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6207–6220, Sep. 2017.

[45] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[46] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α-fairness," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2357–2369, Oct. 2017.

[47] M. R. Zamani, M. Eslami, M. Khorramizadeh, and Z. Ding, "Energy-efficient power allocation for NOMA with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1009–1013, Jan. 2019.

[48] M. D. Yacoub, "The κ-μ distribution and the η-μ distribution," *IEEE Antennas Propag. Mag.*, vol. 49, no. 1, pp. 68–81, Feb. 2007.

DONGJAE KIM (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011, 2013, and 2020, respectively. He has been a Postdoctoral Researcher with the Artificial Intelligence Convergence Research Center for Regional Innovation, Korea Maritime and Ocean University, since 2021. His research interests include tranceiver design and 5G communications.

MINSEOK CHOI (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011, 2013, and 2018, respectively. He was a Visiting Postdoctoral Researcher in electrical and computer engineering with the University of Southern California (USC), Los Angeles, CA, USA, and a Research Professor in electrical engineering with Korea University, Seoul, South Korea. He has been an Assistant Professor with Jeju National University, Jeju, South Korea, since 2020. His research interests include wireless caching networks, stochastic network optimization, non-orthogonal multiple access, and 5G networks.

● ● ●