Min-SINR Maximization for mmWave Massive MISO-NOMA System With Randomly Directional Beamforming

FUYUAN XU AND HAILIN ZHANG, (Member, IEEE)
State Key Laboratory of Integrated Service Network, Xidian University, Xi’an 710071, China
Corresponding author: Fuyuan Xu (xufuyuan_fly@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1200202, in part by the National Natural Science Foundation of China under Grant 61671347 and Grant 61771368, and in part by the 111 Project under Grant B08038.

ABSTRACT Based on the high-direction characteristic of millimeter wave (mmWave) transmission, randomly directional beamforming (RDB) can be used for the mmWave massive multiple-input single-output (MISO) non-orthogonal multiple access (NOMA) system to reduce the number of radio frequency (RF) chains. However, the impact of RDB on the signal-to-interference-plus-noise ratio (SINR) of each user is related to the corresponding beamforming gain and interference from other users. Thus, in this paper, we investigate the max-min SINR among all users to evaluate user fairness. In particular, we focus on the single-cell downlink mmWave MISO-NOMA system with RDB, where single-antenna users are divided into multiple NOMA clusters according to their azimuth angles. We formulate the minimum achievable SINR maximization problem associated with power allocation and propose the sum of power allocation coefficients based iterative algorithm (SPACIA) to find the max-min SINR. We also prove that the max-min SINR monotonically decreases as the number of paired users in an NOMA cluster increases as well as the number of beams in the cell with large-scale base station (BS) antenna array increases. Moreover, we derive the upper bound of the max-min SINR. Simulation results verify our theoretical analyses and demonstrate that the proposed algorithm guarantees user fairness, thus outperforming existing schemes.

INDEX TERMS Massive MISO, max-min SINR, mmWave, NOMA, randomly directional beamforming.

I. INTRODUCTION

Recently, non-orthogonal multiple access (NOMA), millimeter wave (mmWave) communication, and massive multiple-input multiple-output (MIMO) have been considered as three promising technologies in future wireless networks to meet demands for high data rate and massive connectivity [1]–[3]. For the three technologies, mmWave with short wavelengths enables large-scale antenna array to be integrated in limited space [4]–[6], while sufficient large-scale antenna array gain compensates for high path loss in mmWave transmission [7]–[9]. Moreover, the strong channel correlation of paired users in NOMA system can be satisfied by highly directional nature of mmWave transmission [10]–[13]. Inspired by the potential advantages, it is desirable to combine the three technologies in future wireless networks [4], [10]–[21].

The mmWave massive MIMO-NOMA system relies on analog processing or hybrid processing to reduce hardware cost and power consumption [18], [22]. Fortunately, analog precoding/combining or hybrid precoding/combining can be close to the optimal precoding/combining for sum rate optimization [8], [19], [23]–[27]. In particular, randomly directional beamforming (RDB), which is a form of analog precoding, can be applied to the mmWave massive multiple-input single-output (MISO) NOMA system where one base station (BS) is communicating with multiple single-antenna users. According to the RDB principle [10], [27], radio frequency (RF) signals at the BS can be processed by a set of random orthonormal vectors before being transmitted to users. Moreover, the uniform random single-path (UR-SP) channel model can be utilized. In the way, the BS randomly generates...
multiple beams, which are steered at different azimuth angles to cover the whole cell. Then, the users with channel vectors close to the same beam direction are paired together and served by the corresponding beam. However, RDB has a different impact on each user’s signal-to-interference-plus-noise ratio (SINR). Specifically, the beamforming gain of the user depends on how close its channel vector is to its serving beam direction, and interference suffered by the user is related to the azimuth angle differences between its channel vector and interference beam directions. As a result, compared with other users, RDB has a positive impact on SINRs of users with higher beamforming gain and lower interference. Moreover, the highly directional characteristic of mmWave transmission increases the probability that the user falls into the side lobes [28]–[30]. Thus, it is important to guarantee user fairness in the system with RDB.

As a typical criterion, the max-min fairness has been extensively studied. For example, some authors investigate the max-min fairness by proposing schemes related to beamforming, power allocation, or user pairing [11], [31]–[38]. In particular, in the mmWave massive MISO-NOMA system, the authors of [11] and [34] investigated the max-min user rate based on the assumptions that all scheduled users were in an NOMA cluster and each user was served by a steering beam. However, the assumptions are hard to meet when the number of the users is large. Furthermore, for other works in the mmWave massive MISO-NOMA system, the authors of [10] investigated the sum rate and outage probability in the scenario with RDB and azimuth angle-based user pairing (AAUP), where users in the cell were divided into multiple NOMA clusters and users in a cluster were served by a steering beam. Nevertheless, they do not consider user fairness. The authors of [4] pointed out that the number of clusters in the cell and the number of users in a cluster had an impact on the sum rate, whereas it was hard to find the elbow point. The authors of [39] analyzed variations of the user rate with the number of antennas and the number of beams, respectively. However, the effect of the number of paired users on the user rate is not taken into consideration.

Different from the aforementioned works, in this paper, we maximize the minimum achievable SINR among all users and investigate how the max-min SINR varies with system parameters in the single-cell downlink mmWave massive MISO-NOMA system with RDB, where users are paired based on their azimuth angles. The main contributions of this paper are summarized as follows:

1) We formulate the max-min SINR problem for power allocation in the single-cell downlink mmWave massive MISO-NOMA system, which is non-convex and hard to solve. Through variable substitution and optimal decoding order assignment for intra-cluster users, we convert the problem into an equivalent standard convex problem.

2) We derive the closed-form expressions of power allocation coefficients associated with the max-min SINR by investigating the constraints of the converted convex problem. Then, the sum of power allocation coefficients based iterative algorithm (SPACIA) is proposed to search for the max-min SINR. We further prove that the proposed algorithm achieves better performance than conventional orthogonal multiple access (OMA) scheme in terms of user fairness.

3) We prove that the max-min SINR decreases as the number of paired users in a cluster increases. We also prove when the number of BS antennas tends to infinity and the transmit SNR is high, the max-min SINR, with the fixed number of users in a cluster, decreases as the number of beams increases in the cell. In addition, we derive the upper bound of the max-min SINR when the maximum azimuth angle difference between each user’s channel vector and its serving beam direction tends 0.

The rest of this paper is organized as follows. Section II presents the single-cell downlink mmWave massive MISO-NOMA system model. In Section III, we formulate the max-min SINR problem and then propose SPACIA algorithm to obtain the optimal solution. In Section IV, we compare the max-min SINR in our system with that in the conventional OMA system and investigate the impact of system parameters on the max-min SINR. Section V presents simulation results. Finally, the paper concludes with Section VI.
have \( N_s \leq N \leq M \) \cite{5, 8, 41}. Without loss of generality, we assume \( N = N_s \).

### A. CHANNEL MODEL

Owing to the highly directional characteristic of mmWave transmission, in the mmWave channel there exist very few multipaths, which consist of a line-of-sight (LoS) path and few non-LoS (NLoS) paths \cite{4, 10, 25, 27}. Because the gain of the LOS path can be 20 dB stronger than that of any NLoS path, the LoS path is the dominant path. However, when the LoS path does not exist, the strongest NLoS path becomes the dominant path. Additionally, it is feasible to consider only one dominant path for mmWave channel. Thus, as the authors of \cite{4, 10, 25, 27} consider, in the paper we adopt the UR-SP channel model with only one dominant path. Consequently, the channel vector from the BS to the \( j \)-th user in the \( i \)-th cluster is expressed as

\[
\mathbf{h}_{ij} = \sqrt{M} \frac{a_{ij}}{\sqrt{1 + d_{ij}}} \mathbf{a}(\varphi_{ij}),
\]

where \( a_{ij} \sim CN(0, 1) \) denotes the complex gain of the dominant path, \( d_{ij} \) is the distance from the BS to the user, \( \alpha \) is the path loss exponent, \( 1 \leq i \leq N_s \), and \( 1 \leq j \leq K_i \). Moreover, \( \varphi_{ij} \) is the azimuth angle of departure (AoD). The response vector of the uniform linear array (ULA) with \( M \) elements, denoted by \( \mathbf{a}(\varphi_{ij}) \), is a directional or steering vector. Thus, \( \mathbf{a}(\varphi_{ij}) \) is written as follows

\[
\mathbf{a}(\varphi_{ij}) = \frac{1}{\sqrt{M}} \left[ 1, e^{-j2\pi \frac{D}{\lambda} \sin(\varphi_{ij})}, \ldots, e^{-j2\pi (M-1) \frac{D}{\lambda} \sin(\varphi_{ij})} \right]^H,
\]

where \( D \) is the distance between two adjacent antenna elements, \( \lambda \) is the wavelength \cite{42}, and \( [\cdot]^H \) denotes conjugate transpose. Without loss of generality, we assume \( D/\lambda = 1/2 \). Correspondingly, the normalized direction \( \theta_{ij} \) is

\[
\theta_{ij} = \frac{2D}{\lambda} \sin(\varphi_{ij}) = \sin(\varphi_{ij}).
\]

### B. SIGNAL MODEL

As shown in Fig. 2, we assume the \( i \)-th data stream bears the desired signals transmitted by the BS to users of the \( i \)-th NOMA cluster after power allocation, where \( 1 \leq i \leq N_s \). Passing through the RF chain, the signals of each cluster are processed by an analog precoding vector. Then, the weighted sum of the signals of \( N_s \) clusters is transmitted to all users in the cell. Thus, the received signal at the user consists of desired signal, noise, intra-cluster interference, and inter-cluster interference. Moreover, the number of scheduled users in different clusters may be different.

The signal vector for \( N_s \) clusters, denoted by \( \mathbf{s} \), can be written as

\[
\mathbf{s} = \begin{bmatrix} \sqrt{\rho_i} s_1 \\ \vdots \\ \sqrt{\rho_i} s_i \\ \vdots \\ \sqrt{\rho_i} s_{N_s} \end{bmatrix} = \begin{bmatrix} \sqrt{\rho_1} s_1 \\ \sqrt{\rho_1} s_i \\ \vdots \\ \sqrt{\rho_1} s_{N_s} \end{bmatrix} + \begin{bmatrix} \sqrt{\rho_2} s_{N_s+1} \\ \vdots \\ \sqrt{\rho_2} s_{2N_s} \\ \vdots \\ \sqrt{\rho_2} s_{M} \end{bmatrix} + \begin{bmatrix} \sqrt{\rho_3} s_{2M+1} \\ \vdots \\ \sqrt{\rho_3} s_{3M} \\ \vdots \\ \sqrt{\rho_3} s_{3M} \end{bmatrix} + \cdots
\]

\[
\cdots + \begin{bmatrix} \sqrt{\rho_{N_s}} s_{N_s M+1} \\ \vdots \\ \sqrt{\rho_{N_s}} s_{N_s M+2} \\ \vdots \\ \sqrt{\rho_{N_s}} s_{N_s M+N_s} \end{bmatrix}
\]

where \( s_i \) is the sum of the signals for the \( i \)-th cluster and \( \rho_i \) denotes the power allocated to the \( i \)-th cluster. Moreover, \( s_{i,j} \) is the desired signal and \( p_{i,j} \) denotes the intra-cluster power allocation coefficient of the \( j \)-th user in the \( i \)-th cluster, where \( 1 \leq i \leq N_s \) and \( 1 \leq j \leq K_i \). Thus, for the \( i \)-th cluster, the intra-cluster power allocation coefficient vector is written as

\[
\mathbf{\beta}_i = [\beta_{i,1} \cdots \beta_{i,K_i}]^H.
\]

The inter-cluster power allocation vector is

\[
\mathbf{p} = [\rho_1 \cdots \rho_{N_s}]^H.
\]

Let \( U = \{U_{i,j}\} \) denotes the set of all scheduled users in the cell, where \( U_{i,j} \) represents the \( j \)-th user in the \( i \)-th cluster. Since \( U_{i,j} \) is equipped with the single antenna, the received signal at \( U_{i,j} \) can be expressed as

\[
y_{i,j} = \mathbf{h}_{ij}^H \mathbf{W}_{i,j} + n_{i,j} = \mathbf{h}_{ij}^H \left[ \mathbf{w}_1 \cdots \mathbf{w}_{N_s} \right] \begin{bmatrix} \sqrt{\rho_1} s_1 \\ \vdots \\ \sqrt{\rho_1} s_{N_s} \end{bmatrix} + n_{i,j}
\]

\[
= \mathbf{h}_{ij}^H \sqrt{\rho_1} \mathbf{w}_s s_i + \mathbf{h}_{ij}^H \sum_{l \neq i} \sqrt{\rho_1} \mathbf{w}_l s_l + n_{i,j}
\]

\[
= \mathbf{h}_{ij}^H \sqrt{\rho_1} \mathbf{w}_s \beta_{i,s} + \mathbf{h}_{ij}^H \mathbf{w}_i \sum_{k \neq i} \sqrt{\rho_2} \mathbf{w}_k \beta_{i,k} + n_{i,j}
\]

\[
= \mathbf{h}_{ij}^H \sqrt{\rho_1} \mathbf{w}_s \beta_{i,s} + \mathbf{h}_{ij}^H \mathbf{w}_i \sum_{k \neq i} \sqrt{\rho_2} \mathbf{w}_k \beta_{i,k} + n_{i,j}
\]

\[
= \mathbf{h}_{ij}^H \sqrt{\rho_1} \mathbf{w}_s \beta_{i,s} + \mathbf{h}_{ij}^H \mathbf{w}_i \sum_{k \neq i} \sqrt{\rho_2} \mathbf{w}_k \beta_{i,k} + n_{i,j}
\]

\[
\text{desired signal} + \text{intra-cluster interference} + \text{inter-cluster interference}
\]

where \( \mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_{N_s}] \) is the \( M \times N_s \) analog precoding matrix, \( \mathbf{w}_i \) is the \( M \times 1 \) analog precoding vector for RDB, \( n_{i,j} \sim CN(0, \sigma^2) \) represents the additive Gaussian noise, and \( \sigma^2 \) is the noise power.
Based on (7), we write the equivalent channel gain as [4]

$$g_{i,j} = \frac{p_i |h_{i,j}^H w_i|^2}{\sum_{l \neq i} p_l |h_{i,l}^H w_l|^2 + \sigma^2},$$

(8)

where $| \cdot |$ denotes the absolute value. Without loss of generality, in the $i$-th cluster, we assume the equivalent channel gain becomes weaker and weaker from the first user to the $K_i$-th user. Therefore, the equivalent channel gains of users are sorted as

$$g_{i,1} \geq g_{i,2} \cdots \geq g_{i,K_i}.$$  

(9)

When successive interference cancellation (SIC) is adopted at users, the interference caused by weaker users at stronger users can be eliminated. As a result, when $j = 1$, the SINR of the $j$-th user is

$$\text{SINR}_{i,j}^j = \frac{p_i |h_{i,j}^H w_i|^2 \beta_{i,j}}{\sum_{k < j} p_i |h_{i,j}^H w_i|^2 + \sigma^2} = g_{i,j} \beta_{i,j}. $$

(10)

When $1 < j \leq K_i$, the SINR of the $j$-th user becomes

$$\text{SINR}_{i,j}^j = \frac{p_i |h_{i,j}^H w_i|^2 \beta_{i,j}}{p_i |h_{i,j}^H w_i|^2 \sum_{k < j} \beta_{i,k} + \sum_{l \neq i} p_l |h_{i,j}^H w_l|^2 + \sigma^2} = \frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k < j} \beta_{i,k} + 1}. $$

(11)

As shown in (10) and (11), the SINR of each user is related to desired signal, intra-cluster interference, inter-cluster interference, and noise. Moreover, the SINR of each user varies with respect to the intra-cluster power allocation coefficients. In Section III, in order to ensure the max-min fairness, we will find the optimal intra-cluster power allocation coefficients and the corresponding max-min SINR.

In addition, in the $i$-th cluster, when $1 \leq j < j' \leq K_i$, the SINR of the $j'$-th user decoded by the $j$-th user can be expressed as

$$\text{SINR}_{i,j}^{j'} = \frac{p_i |h_{i,j}^H w_i|^2 \beta_{i,j'}}{p_i |h_{i,j}^H w_i|^2 \beta_{i,j} \sum_{k < j'} \beta_{i,k} + \sum_{l \neq i} p_l |h_{i,j}^H w_l|^2 + \sigma^2} = \frac{g_{i,j} \beta_{i,j'}}{g_{i,j} \sum_{k < j'} \beta_{i,k} + 1}. $$

(12)

### III. PROBLEM FORMULATION AND SOLUTION SCHEME

In the section, firstly, we formulate the max-min SINR problem. Secondly, we convert the original non-convex problem into an equivalent convex one. Thirdly, we propose SPACIA algorithm to solve the convex problem.

### A. PROBLEM FORMULATION

In this paper, as the authors of [10] consider, we utilize $w_i = a(\bar{\theta}_i)$ and allocate the power equally to each cluster. $\bar{\theta}_i$ is the normalized central azimuth angle of the $i$-th cluster and it is given by

$$\bar{\theta}_i = \frac{\varepsilon + 2(n - 1)}{N_s}, 1 \leq n \leq N_s$$

(13)

where $\varepsilon$ is a random variable that is uniformly distributed between $-1$ and $1$. In this way, the cell is divided into $N_s$ sectors and RDB is achieved by the analog precoding vector $a(\bar{\theta}_i)$.

Additionally, we assume $P_T$ is the total power of the system. With the equal power among clusters, the maximum usable power for each cluster is

$$P_{\text{max}} = \frac{P_T}{N_s}.$$  

(14)

while the actually used power in the $i$-th cluster is

$$p_i = \max_{j=1}^{K_i} \beta_{i,j}.$$  

(15)

Consequently, the signals of the $i$-th cluster are transmitted along the beam vector $w_i = a(\bar{\theta}_i)$ due to the highly directional nature of mmWave transmission [5], [10], [26]. However, the SINR of each user is related to how close its channel vector is to its serving beam direction, which challenges user fairness. Thus, with the given $w_i = a(\bar{\theta}_i)$ and the equal power among clusters, we aim to maximize the minimum achievable SINR by intra-cluster power allocation to evaluate user fairness. Considering the power budget, lower SINR threshold of users, and power constraint for SIC at users, we formulate the max-min SINR problem for intra-cluster power allocation as follows.

$$\text{P1} : \max_{1 \leq i \leq N_s, 1 \leq j \leq K_i} \text{SINR}_{i,j}^j$$

s.t. $\text{SINR}_{i,j}^j \geq \Gamma,$

$$\sum_{j=1}^{K_i} \beta_{i,j} \leq 1, \beta_{i,j} > 0,$$

$$\sum_{i=1}^{N_s} p_i \leq P_T, \quad p_i > 0,$$

$$\beta_{i,1} \leq \beta_{i,2} \cdots \leq \beta_{i,K_i}.$$  

(16)

Constraint (16a) together with (16b) represents that the lower SINR threshold, denoted by $\Gamma$, is the same for each user to decode itself or be decoded by other better users for SIC. Constraint (16c) guarantees the power constraint in one cluster. Constraint (16d) ensures the total power constraint. Constraint (16e) denotes the power constraint for SIC at users. Problem P1 is non-convex and therefore it is hard to solve. However, it can be converted into an equivalent convex problem.
B. SOLUTION SCHEME

We convert problem P1 into an equivalent convex problem based on the following three aspects.

Firstly, constraint (16b) can be removed. We assume the desired signals are decoded from the $K_i$-th user with the weakest equivalent channel gain to the first one with the strongest equivalent channel gain at users, which is the optimal decoding order verified in [16], [17]. As a result, based on the condition that $\text{SINR}_{i,j}^l \geq \Gamma (1 < j \leq K_i)$, which is a subset of constraint (16a), we can derive (16b). Thus, constraint (16b) is redundant and can be removed. For the detailed proof, please refer to [4].

Secondly, problem P1 can be converted into an equivalent maximization problem through variable substitution. This is because P1 is quasi-concave, as presented in Proposition 1.

Proposition 1: Problem P1 is quasi-concave.
Proof: When the objective function is quasi-concave and the constraints are convex, a maximization problem is quasi-concave [32]. Obviously, in problem P1 the objective function $\min \text{SINR}_{i,j}^l \geq \gamma$ is equivalent to $\text{SINR}_{i,j}^l \geq \gamma$, where $1 \leq i \leq N_s$, $1 \leq j \leq K_i$, and $\gamma$ is an auxiliary variable to replace $\min \text{SINR}_{i,j}^l$ of problem P1 in the following work. Since $\text{SINR}_{i,j}^l \geq \gamma$ is linear with respect to $\beta_{i,j}$, the objective function is quasi-concave. Moreover, the constraints of P1 are convex because they are all linear inequalities associated with $\beta_{i,j}$; Therefore, problem P1 is quasi-concave.

Thirdly, problem P1 can be decomposed into $N_s$ decoupled subproblems. This is because the inter-cluster power allocation vector $\mathbf{p}$ and the precoding matrix $\mathbf{W}$ are known. Without loss of generality, we focus on the max-min SINR in one cluster, e.g., the $i$-th cluster, where $1 \leq i \leq N_s$.

Consequently, the original problem P1 is equivalently converted into a standard convex problem P2.

\[
P2 : \max_{1 \leq j \leq K_i, \beta_{i,j} \geq \gamma} \gamma \\
\text{s.t. (16a), (16c), (16d), (16e),} \\
\text{ } \\
\text{SINR}_{i,j}^l \geq \gamma. \tag{17a}
\]

For problem P2, $\min \text{SINR}_{i,j}^l$ of problem P1 is replaced by the auxiliary variable $\gamma$. In addition, when the max-min SINR achieves, constraints (16c) and (17a) are strictly equal, respectively. More details are provided in Lemma 1.

Lemma 1: When $\gamma$ is optimal, constraint (16c) becomes
\[
\sum_{j=1}^{K_i} \beta_{i,j} = 1, \tag{18}
\]
and constraint (17a) satisfies
\[
\text{SINR}_{i,j}^l = \gamma. \tag{19}
\]
Proof: Please see Appendix A.

Based on Lemma 1, we derive the optimal power allocation coefficients for the $i$-th cluster, which are detailed in Lemma 2.

Lemma 2: The optimal intra-cluster power allocation coefficients for the $i$-th cluster are
\[
\beta_{i,j} = \gamma \left( \sum_{k<j} \beta_{i,k} + \frac{\mathbf{h}_{ij}^H \mathbf{w}_j}{\gamma} \right) + \gamma \left( \sum_{k>j} \beta_{i,k} + \frac{\mathbf{h}_{ij}^H \mathbf{w}_j}{\gamma} \right), \quad 2 \leq j \leq K_i. \tag{20}
\]
Proof: $\rho = \frac{p_{\text{max}}}{\sigma^2}$ represents the transmit SNR. Substituting (10) and (11) into (19), we solve $\beta_{i,j}$ when $1 \leq j \leq K_i$. Then, equation (20) is obtained.

From (9), we have the equivalent channel gains of users are sorted as $g_{i,1} \geq g_{i,2} \cdots \geq g_{i,K_i}$. According to the NOMA principle, if the intra-cluster power allocation coefficients follow the order as $\beta_{i,1} \leq \beta_{i,2} \cdots \leq \beta_{i,K_i}$, SIC can be carried out at users successfully [16], [43], [44]. Thus, in Proposition 2, we prove $\beta_{i,1} \leq \beta_{i,2} \cdots \leq \beta_{i,K_i}$ in (20).

Proposition 2: For the $i$-th cluster, the power allocation coefficients in (20) satisfy $\beta_{i,1} \leq \beta_{i,2} \cdots \leq \beta_{i,K_i}$.
Proof: If $K_i \geq j' > j \geq 1$, we have $g_{i,j'} \geq g_{i,j}$. Then, based on (20), we obtain
\[
\beta_{i,j'} - \beta_{i,j} = \gamma \left( \sum_{k<j'} \beta_{i,k} + \frac{1}{g_{i,j'}} \right) - \gamma \left( \sum_{k<j} \beta_{i,k} + \frac{1}{g_{i,j}} \right) = \gamma \left( \sum_{k<j} \beta_{i,k} + \frac{1}{g_{i,j}} - \sum_{k<j'} \beta_{i,k} - \frac{1}{g_{i,j'}} \right) = \gamma \left( \sum_{k<j'} \beta_{i,k} - \sum_{k<j} \beta_{i,k} \right) + \left( \frac{1}{g_{i,j'}} - \frac{1}{g_{i,j}} \right) \geq \gamma \left( \sum_{k<j'} \beta_{i,k} - \sum_{k<j} \beta_{i,k} \right) > 0,
\]
which means the intra-cluster power allocation coefficients in (20) satisfy $\beta_{i,1} \leq \beta_{i,2} \cdots \leq \beta_{i,K_i}$ and thus SIC can be carried out at users. This also means constraint (16e) is redundant and may not be considered in the paper.

Lemma 1 ensures the max-min fairness in the system. Furthermore, in Lemma 2 the closed-form expressions of power allocation coefficients are given to effectively eliminate user differences caused by highly directional beams. Based on Lemma 2, we obtain the sum of power allocation coefficients in Theorem 1.
The sum of power allocation coefficients in the $i$-th cluster is expressed as

$$f(\gamma) = \sum_{j=1}^{K_i} \frac{\gamma (1 + \gamma)^{K_i-1}}{\gamma^{2}} \left( \sum_{l \neq i}^{N_l} \frac{\| \mathbf{h}^H_l \mathbf{w}_l \|^2}{\gamma^{2}} + \frac{1}{\rho} \right).$$

Proof: Please see Appendix B.

Obviously, $f(\gamma)$ is an increasing function with respect to $\gamma$. When $f(\gamma) = 1$, we have $\gamma = \gamma^*$, where $\gamma^*$ is optimal. Therefore, SPACIA algorithm is proposed to find $\gamma^*$ between the upper bound $\gamma_{\text{max}}$ and the lower bound $\gamma_{\text{min}}$, where initial $\gamma_{\text{max}}$ is set to be $\rho \times \max \left( \| \mathbf{h}^H_{ij} \mathbf{w}_i \|^2 \right)$ and initial $\gamma_{\text{min}}$ is set to be the lower SINR threshold $1 / \rho$. In steps 4 to 9, if $f \left( \gamma_{\text{max}} + \gamma_{\text{min}} / 2 \right) < 1$, $\gamma_{\text{min}}$ increases; otherwise, $\gamma_{\text{max}}$ decreases. The iterations stop until $|f(\gamma) - 1| \leq \varepsilon$, where $\varepsilon$ is the search accuracy that is small enough. More details about the algorithm are shown by the pseudo-code in Algorithm 1.

Algorithm 1 The Sum of Power Allocation Coefficients Based Iterative Algorithm (SPACIA)

1. Initialize $\bar{\gamma}_i$, $\mathbf{h}_{ij}$, $P_t$, $N_t$, $\gamma_{\text{min}}$, $\Delta$, $\rho$, $\zeta$, $\sigma^2$, $\varepsilon$.
2. Calculate the initial upper bound $\gamma_{\text{max}}$.
3. Repeat
4. $\gamma = \left( \gamma_{\text{max}} + \gamma_{\text{min}} / 2 \right)$
5. Calculate $f(\gamma)$ based on (22)
6. if $f(\gamma) > 1$
7. $\gamma_{\text{max}} = \gamma$
8. else
9. $\gamma_{\text{min}} = \gamma$
10. End until $|f(\gamma) - 1| \leq \varepsilon$
11. Output $\gamma^* = \gamma$.
12. Calculate the optimal $\beta_{ij}$ based on (20).

The searching time required by Algorithm 1 is $T = \log_2 \left( \left( \gamma_{\text{max}} - \gamma_{\text{min}} \right) / \varepsilon \right)$ [11]. Hence, the computational complexity of Algorithm 1 is $O(T)$.

IV. PERFORMANCE ANALYSIS

In the section, firstly, we demonstrate the max-min SINR in our proposed system outperforms that in the conventional OMA system. Secondly, we investigate the variation of the max-min SINR with the number of paired users in a cluster. Thirdly, we investigate the impact of the number of beams on the max-min SINR, which reflects the asymptotic behavior of the max-min SINR when the number of BS antennas tends infinity and the transmit SNR is high. Finally, we derive the upper bound of the max-min SINR. Correspondingly, the detailed conclusions are presented in the following four corollaries.

Corollary 1: With multiple users in a cluster, our proposed algorithm achieves better performance than the conventional OMA scheme in terms of the max-min SINR.

Proof: In the conventional OMA system, such as time division multiple access (TDMA) system, time resources can be equally allocated to each user in a cluster. Thus, in the $i$-th cluster, if $\gamma_{i,\text{OMA}}$ is the SINR of the $j$-th user, the data rate of the $j$-th user, denoted by $R_{j,\text{OMA}}$, is written as

$$R_{j,\text{OMA}} = \frac{1}{K_i} \log_2 \left( 1 + \left( \frac{\| \mathbf{h}^H_j \mathbf{w}_j \|^2}{\sum_{l \neq i}^{N_l} \| \mathbf{h}^H_l \mathbf{w}_l \|^2 + \frac{1}{\rho}} \right) \right) \frac{1}{K_i}$$

$$= \log_2 \left( 1 + \left( \frac{\| \mathbf{h}^H_j \mathbf{w}_j \|^2}{\sum_{l \neq i}^{N_l} \| \mathbf{h}^H_l \mathbf{w}_l \|^2 + \frac{1}{\rho}} \right) \frac{1}{K_i} \right)$$

Equation (23) can be rewritten as

$$(1 + \gamma_{i,\text{OMA}})^{K_i} - 1 = \frac{\| \mathbf{h}^H_j \mathbf{w}_j \|^2}{\sum_{l \neq i}^{N_l} \| \mathbf{h}^H_l \mathbf{w}_l \|^2 + \frac{1}{\rho}} = g_{i,j}.$$  

Substituting (24) into (22), we have

$$f \left( \gamma_{i,\text{OMA}} \right) = \sum_{j=1}^{K_i} \frac{\gamma_{i,\text{OMA}} (1 + \gamma_{i,\text{OMA}})^{K_i-j}}{\sum_{l \neq i}^{N_l} \| \mathbf{h}^H_l \mathbf{w}_l \|^2 + \frac{1}{\rho}}$$

$$= \sum_{j=1}^{K_i} g_{i,j} \gamma_{i,\text{OMA}} (1 + \gamma_{i,\text{OMA}})^{K_i-j}$$

$$= \sum_{j=1}^{K_i} \left( 1 + \gamma_{i,\text{OMA}} \right)^{K_i} - 1$$

$$= \frac{g_{i,j}}{g_{i,K_i}},$$  

where (a) follows from $g_{i,j} \geq g_{i,K_i}$ (1 $\leq j \leq K_i$) since $U_{i,K_i}$ is the user with the worst equivalent channel gain. When $j = K_i$, inequality (25) becomes

$$f \left( \gamma_{i,K_i,\text{OMA}} \right) \leq \frac{g_{i,K_i}}{g_{i,K_i}} = 1 = f \left( \gamma^* \right).$$

Because $f(\gamma)$ is an increasing function with respect to $\gamma$, we have

$$\gamma_{i,K_i,\text{OMA}} \leq \gamma^*.$$  

Thus, in the conventional OMA system, there exists at least one user $U_{i,K_i}$ with the SINR that is not greater than $\gamma^*$. The proof is complete.

Corollary 1 implies that in the mmWave massive MIMO system, the performance of our proposed NOMA scheme is
better than the conventional OMA scheme in terms of user fairness. This ensures that all users are fairly served even though they have different equivalent channel gains and suffer different interference.

**Corollary 2:** The max-min SINR is a decreasing function with respect to the number of paired users in a cluster when other parameters are fixed.

**Proof:** If a new user \( U_{l,K_{new}} \) is scheduled in the \( i \)-th cluster, the total number of users increases from \( K_i \) to \( K_i + 1 \). Correspondingly, we assume that the max-min SINR changes from \( \gamma^* \) to \( \eta^* \). The equivalent channel gain of the new user satisfies

\[
g_1 \geq g_{K_{new}} \geq g_{K_i + 1}.
\]

(28)

Based on (9) and (28), we have \( 1 \leq K_{new} \leq K_i + 1 \). When \( 1 < K_{new} < K_i + 1 \), we prove \( \gamma^* > \eta^* \) as follows

\[
\sum_{j=1}^{K_i} \frac{\gamma^*(1 + \gamma^*)K_i - j}{g_{i,j}} = f(\gamma^*) = 1 = f(\eta^*)
\]

\[
= \sum_{j=1}^{K_i + 1} \eta^*(1 + \eta^*)K_i - j \frac{g_{i,j}}{g_{i,j}}
\]

\[
= \sum_{j=1}^{K_i + 1} \eta^*(1 + \eta^*)K_i - j \frac{g_{i,j}}{g_{i,j}}
\]

\[
\frac{f}_{j=1}^{K_{new} - 1} \eta^*(1 + \eta^*)K_i - j \frac{g_{i,j}}{g_{i,j}}
\]

\[
\frac{f}_{j=K_{new}}^{j=K_i + 1} \eta^*(1 + \eta^*)K_i - j \frac{g_{i,j}}{g_{i,j}}
\]

\[
> \sum_{j=1}^{K_i} \eta^*(1 + \eta^*)K_i - j \frac{g_{i,j}}{g_{i,j}}.
\]

(29)

Because \( f(\gamma^*) \) is an increasing function with respect to \( \gamma^* \), we have \( \gamma^* > \eta^* \) when \( 1 < K_{new} < K_i + 1 \). In addition, when \( K_{new} = 1 \) or \( K_{new} = K_i + 1 \), following steps similar to those in (29), we also have \( \gamma^* > \eta^* \). Thus, Corollary 2 is true.

**Corollary 3:** When the number of BS antennas tends to infinity and the transmit SNR is high, the max-min SINR, with the fixed number of users in a cluster, decreases as the number of beams increases in the cell.

**Proof:** This is reasonable because with the number of beams increasing, inter-cluster interference increases, which is dominant in reducing the max-min SINR. For the detailed proof, please refer to Appendix C.

**Remark 1:** Assume the number of users in a cluster is not fixed. Based on Appendix C, we have that with the increase of the number of beams, whether the max-min SINR decreases depends on the number of users in the cluster. For example, when the total number of served users in the cell is fixed, more beams make users less likely to fall into the side lobes. This strengthens the desired signal gains of users, which may dominate the increase in the max-min SINR.

**Corollary 4:** When the maximum azimuth angle difference between each user’s channel vector and its serving beam direction, denoted by \( \Delta_i = \max_j (\hat{\theta}_i - \theta_{i,j}) \), tends 0, the max-min SINR is expressed as

\[
y = \left( \frac{1}{\rho} \sum_{i=1}^{M-1} \frac{1}{\sum_{m=0}^{\lceil \beta_i \rceil} \frac{g_i - \gamma_i}{\rho + (g_i - \gamma_i)^2}} \right)^{\frac{1}{K_i}} + 1.
\]

(30)

Furthermore, if \( M = n N_s \) for \( n = 1, 2, 3, \ldots \), (30) becomes

\[
y = \left( \frac{1}{\rho} \sum_{i=1}^{M-1} \frac{g_i - \gamma_i}{\rho + (g_i - \gamma_i)^2} \right)^{\frac{1}{K_i}} + 1.
\]

(31)

**Proof:** For the detailed proof, please refer to Appendix D.

**Remark 2:** As \( \Delta_i \) decreases, the average beamforming gain for each user increases. When \( \Delta_i \) equals 0, the beamforming gain of each user achieves the highest value. Furthermore, if \( M = n N_s \) for \( n = 1, 2, 3, \ldots \), inter-cluster interference becomes 0. Therefore, the upper bound of the max-min SINR is given by (31).

**V. SIMULATION RESULTS**

In this section, simulation results are provided to illustrate the performance of our proposed algorithm in terms of the max-min SINR as well as show the relationships between the max-min SINR and system parameters, such as the number of paired users, beams, and BS antennas. The BS is located in the center of the cell with the radius of 50 m. All users are uniformly distributed in the cell. The path loss exponent is set as \( \alpha = 2 \) and the noise power is set as \( \sigma^2 = -174 + 10 \log(B) \) in dBm, where the bandwidth is set as \( B = 100\text{MHz} \). Active users in the same sector can be scheduled in an NOMA cluster and served by a randomly directional beam. Each performance curve is obtained by taking the average of 5000 realizations.

Fig. 3 shows the based-10 logarithm of the sum of power allocation coefficients \( f(\gamma) \) against iterations, where \( M = 64, N_s = 16, \) and \( K_i = 2 \). The transmission power of a cluster is set to be 20 dBm and 30 dBm, respectively. The upper bound is set as \( \gamma = \rho \times \max_j \left( \left| H_{i,j}^H w_i \right|^2 \right) \). The lower bound \( \gamma_{min} \) is set to be the lower SINR threshold \( \Gamma \). For each curve, we observe that after dozens of iterations, the based-10 logarithm of \( f(\gamma) \) tends to 0, when the max-min SINR \( \gamma^* \) is obtained. Thus, SPACIA algorithm is convergent and effective. In addition, the higher transmission power of a cluster is, the more iterations are needed. The reason is that \( \gamma_{min} \) increases with the transmission power of the cluster,
which makes iterations increase. This also implies the algorithm complexity is affected by the search bounds.

Fig. 4a plots the SINR of the worst user over conventional OMA scheme, NOMA with SPACIA algorithm, NOMA with power allocation algorithm for system capacity maximization in [4], and NOMA with analog beamforming and power allocation algorithm for the max-min fairness in [11]. Fig. 4b depicts the sum of power allocation coefficients $f(\gamma)$ over the aforementioned schemes, where $\gamma$ is also the SINR of the user with the worst equivalent channel gain. In Fig. 4a and Fig. 4b, $M = 64$, $N_s = 4$, and $K_i = 2$. It is worth mentioning that Fig. 4a and Fig. 4b are consistent because $f(\gamma)$ is an increasing function with respect to $\gamma$. We observe that SPACIA algorithm has the best performance. Compared with SPACIA algorithm, the algorithm in [4] makes more power be allocated to the best user and less power be allocated to the users left in the same cluster. Moreover, the SINR obtained by the algorithm in [4] is an increasing function with respect to $c$, where $c = \Gamma / \gamma^*$. $\Gamma$ denotes the lower SINR threshold that is the same for all users, and $\gamma^*$ is the max-min SINR obtained by SPACIA algorithm. This is reasonable because greater ratio $c$ makes the worst user obtain more power. Only when $c = 1$, the SINR of the worst user in [4] equals $\gamma^*$.

Using the scheme in [11], we have each user is actually served by a beam, which makes inter-beam interference increase in our system and thus reduces the max-min SINR. Moreover, user fairness performance achieved by SPACIA algorithm is better than that achieved by OMA scheme, which verifies Corollary 1.

Fig. 5 illustrates the max-min SINR against the transmission power of a cluster under different number of paired users in a cluster, where $M = 64$ and $N_s = 16$. For each curve, when the transmission power of a cluster increases to a certain value, the max-min SINR achieves the stable value. This is because as the transmission power of a cluster increases, $1 / \rho$ decreases. When $1 / \rho$ is small enough, the max-min SINR is determined by the ratio of the desired signal and interference. In addition, for the same transmission power, as the number of paired users increases, the max-min SINR decreases. Thus, Corollary 2 is verified.

Fig. 6 shows the max-min SINR against the transmission power of a cluster under different number of beams in the cell, where $M = 64$ and $K_i = 2$. We observe that the max-min SINR decreases as the number of beams increases. The conclusion in Corollary 3 is verified. The beam width is a decreasing function with respect to the number of BS antennas. The increased number of beams decreases the spatial direction distance of the adjacent beams, which increases inter-cluster interference and thus makes the max-min SINR decrease.

Fig. 7 illustrates the max-min SINR against transmission power of a cluster, where $M = 64$ and the number of served users in the cell is 64. It is worth mentioning that different from Fig. 6, in Fig. 7 the number of paired users in a cluster is inversely proportional to the number of beams and thus the total number of served users in the cell is fixed. We observe that the max-min SINR increases as the number of beams increases in Fig. 7. This is because when the total number of users is fixed, more beams make users less likely to fall into the side lobes, which strengthens the desired signal.
gain. However, in Fig. 6 the max-min SINR decreases as the number of beams increases. Therefore, user fairness in the cell with more beams is not always worse than that with less beams. With the number of beams increasing in the cell, whether the max-min SINR decreases depends on the number of users in a cluster, which is explained in Remark 1.

Fig. 8 shows the max-min SINR against the base-2 logarithm of the number of BS antennas, where the number of served users in the cell is 64. The transmission power in a cluster is set to be 20 dBm when \( N_s = 32 \) and \( K_i = 2 \), and 23 dBm when \( N_s = 16 \) and \( K_i = 4 \). This ensures that the total power of the cell is the same for all the curves. The beam width is a decreasing function of the number of BS antennas. For each curve, as the number of BS antennas increases, the max-min SINR first increases due to the decrease of inter-cluster interference, and it then decreases because more users fall into the side lobes. From any two curves with the same \( \Delta_i \) that denotes the maximum azimuth angle difference between each user’s channel vector and its serving beam direction, we observe when the BS antenna array is small-scale, the system with less beams has higher SINR because inter-cluster interference is lower. However, when the BS antenna array is large-scale, the system with more beams has higher SINR because less users fall into the side lobes. From any two curves with different \( \Delta_i \) but the same \( N_s \) and \( K_i \), we observe the max-min SINR increases as \( \Delta_i \) decreases. This is because the increase of the beamforming gain dominates the increase of the max-min SINR, which is explained in Remark 2.

VI. CONCLUSION
This paper focused on the max-min SINR in the single-cell downlink mmWave massive MISO-NOMA system with RDB. We formulated the minimum SINR maximization problem associated with power allocation. In order to solve the problem, based on the closed-form expressions of intra-cluster power allocation coefficients, we proposed SPACIA algorithm, which outperforms the conventional OMA scheme in terms of the max-min fairness. Furthermore, we investigated variations of the max-min SINR with respect to system parameters. Specifically, we proved the max-min SINR decreases as the number of paired users in an NOMA cluster increases. We proved when the number of BS antennas tends to infinity and the transmit SNR is high, the max-min SINR is determined by joint the number of beams in the cell and the number of users in a cluster. We also derived the upper bound of the max-min SINR related to the number of BS antennas and the number of users in a cluster. Simulation results verified our theoretical analyses and demonstrated our proposed algorithm outperforms existing schemes. In the future, we plan to further investigate the max-min SINR in the mmWave massive MIMO-NOMA with the intelligent reflecting surface.

APPENDIXES
APPENDIX A
PROOF OF LEMMA 1
We denote by \( \gamma^* \) the optimal SINR and prove Lemma 1 with contradiction as follows.

Firstly, we prove (18). We assume \( \sum_{j=1}^{K_i} p_{i,j} < 1 \) when \( \gamma^* \) is achieved. The power left can be allocated to each user by multiplying intra-cluster power allocation coefficient vector \( \beta_i \) by a constant \( c \), where \( c > 1 \). Correspondingly, after reallocating the power left, we denote the SINR of \( j \)-th user
by $\text{SINR}_{i,j,r}^j$, where $1 \leq j \leq K_i$. As shown in (32), $\text{SINR}_{i,j,r}^j$ is higher than the original $\text{SINR}_{i,j}^j$, which means the increase in the SINR of each user. Therefore, $\gamma^*$ is not the optimal value, which contradicts the condition that $\gamma^*$ is the optimal value. Thus, equation (18) is true.

$$\text{SINR}_{i,j,r}^j = \frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k<j} \beta_{i,k} + 1} > \frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k<j} \hat{\beta}_{i,k} + 1}$$

$$= \text{SINR}_{i,j}^j. \quad (32)$$

Secondly, we prove (19). According to (10) and (11), when all other parameters are fixed, $\text{SINR}_{i,j}^j$ is an increasing function of $\beta_{i,j}$ and a decreasing function of $\beta_{i,k}$, where $k < j$. In addition, we observe that in a cluster, the first user is not interfered by intra-cluster users, whereas the $K_i$-th user is interfered by all other intra-cluster users. Because $\max_{1 \leq i \leq N_c, K_i \geq 1} \min_{j \neq \hat{j}} \text{SINR}_{i,j}^j = \gamma^*$, there must exist at least one user $U_{i,j}$ with $\text{SINR}_{i,j}^j = \gamma^*$. We assume there exists another user $U_{i,j'}$ with $\text{SINR}_{i,j'}^j > \gamma^*$, where $j' \neq \hat{j}$. Subtracting a very slight value $\Delta \beta_{i,j'}$ from the transmit power coefficient $\beta_{i,j'}$, inequality $\text{SINR}_{i,j}^j > \gamma^*$ still holds. We reallocate the reduced power $\Delta \beta_{i,j'}$ to $U_{i,j}$ for $j \neq j'$ and then denote the SINR of $j$-th user by $\text{SINR}_{i,j,r}^j$, which is compared with the original $\text{SINR}_{i,j}^j$ as follows.

Case 1: When $j' = K_i$, $U_{i,j'}$ does not interfere with any other intra-cluster users. By multiplying power allocation coefficient $\hat{\beta}_{i,j}$ ($j \neq \hat{j}$) by a constant $c$ ($c > 1$), $\Delta \beta_{i,j'}$ is allocated to $U_{i,j}$. The SINR of $U_{i,j}$ increases when $j \neq \hat{j}$, as shown in (33).

$$\text{SINR}_{i,j,r}^j = \begin{cases} 
\frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k<j} \beta_{i,k} + 1} > \text{SINR}_{i,j}^j, & 1 < j < K_i, \\
\frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k<j} \hat{\beta}_{i,k} + 1} > \text{SINR}_{i,j}^j, & j = 1.
\end{cases} \quad (33)$$

Case 2: When $j' = 1$, $U_{i,j'}$ interferes with all other intra-cluster users. If $\Delta \beta_{i,j'}$ is only allocated to $U_{i,K_i}$, the SINR of $U_{i,j}$ increases when $j \neq \hat{j}$, as shown in (34).

$$\text{SINR}_{i,j,r}^j = \begin{cases} 
\frac{g_{i,j} \beta_{i,j}}{g_{i,j} \sum_{k<j} \beta_{i,k} - \Delta \beta_{i,j'}} + 1 > \text{SINR}_{i,j}^j, & 1 < j < K_i, \\
\frac{g_{i,j} \left( \sum_{k<j} \beta_{i,k} - \Delta \beta_{i,j'} \right) + 1}{g_{i,j} \left( \hat{\beta}_{i,j} + \Delta \beta_{i,j'} \right) + 1} > \text{SINR}_{i,j}^j, & j = K_i.
\end{cases} \quad (34)$$

Case 3: When $1 < j' < K_i$, one part of $\Delta \beta_{i,j'}$, denoted by $\Delta \beta_{i,j_K}$, is allocated to $U_{i,K_i}$, and the other part $\Delta \beta_{i,j'} - \Delta \beta_{i,j_K}$ is allocated to $U_{i,j}$ ($1 \leq j < j'$) through multiplying each of their power allocation coefficients by a constant $c$ ($c > 1$). The SINR of $U_{i,j}$, where $1 \leq j < j'$, can be expressed by (33). The SINR of $U_{i,j}$, where $j' < j \leq K_i$, can be expressed by (34). Hence, the SINR of $U_{i,j}$ also increases when $j \neq \hat{j}$.

From the above three cases, we observe that the SINR of each user increases when $j \neq \hat{j}$. $\gamma^*$ is not the optimal SINR, which contradicts the condition that $\gamma^*$ is the optimal SINR. Thus, there is no $U_{i,j}$ with $\text{SINR}_{i,j}^j > \gamma^*$. Equation (19) is true.

**APPENDIX B**

**PROOF OF THEOREM 1**

We prove Theorem 1 with mathematical induction. When $K_i = 1$, it is easily verified

$$\sum_{j=1}^{1} \beta_{i,j} = \frac{\gamma \left( \sum_{j=1}^{N_i} \left| h_{i,j}^H w_i \right|^2 + \frac{1}{\rho} \right)}{\left| h_{i,j}^H w_i \right|^2}. \quad (35)$$

When $K_i = m > 1$, we assume

$$\sum_{j=1}^{m} \beta_{i,j} = \sum_{j=1}^{m} \beta_{i,j} + \beta_{i,m+1} = (1 + \gamma) \sum_{j=1}^{m} \beta_{i,j} + \beta_{i,m+1} + \gamma \left( \sum_{j=1}^{m} \left| h_{i,j}^H w_i \right|^2 + \frac{1}{\rho} \right) \quad (36)$$

When $K_i = m + 1$,

$$\sum_{j=1}^{m+1} \beta_{i,j} = \sum_{j=1}^{m} \beta_{i,j} + \beta_{i,m+1} + \gamma \left( \sum_{j=1}^{m+1} \left| h_{i,j}^H w_i \right|^2 + \frac{1}{\rho} \right) \quad (37)$$

82006 VOLUME 8, 2020

F. Xu, H. Zhang: Min-SINR Maximization for mmWave Massive MISO-NOMA System
Thus, the proof is complete.

**APPENDIX C**

**PROOF OF COROLLARY 3**

Motivated by [27], [39], [48], [49], we prove Corollary 3 through replacing $|h_{i,j}^H w_i|^2$ with its expectation $E\{ |h_{i,j}^H w_i|^2 \}$ over the azimuth angle difference $\hat{\theta} = \tilde{\theta}_i - \tilde{\theta}_j$ and the distance $d_{i,j}$ from the user to the BS, where $E\{ \cdot \}$ denotes the expectation. Based on (1), (2), and (13), we have

$$|h_{i,j}^H w_i|^2 = M \frac{a_{i,j}^2}{M + 1 + d_{i,j}^a} |a^{H} (\tilde{\theta}_i) a (\tilde{\theta}_j)|^2$$

$$= \frac{a_{i,j}^2}{M (1 + d_{i,j}^a)} \left[ 1 - e^{-j\pi M \hat{\theta}} \right]^2$$

$$= \frac{a_{i,j}^2}{M (1 + d_{i,j}^a)} \left( 1 - e^{-j\pi M \hat{\theta}} \right) \left( 1 - e^{j\pi \hat{\theta}} \right)$$

$$= \frac{a_{i,j}^2}{M (1 + d_{i,j}^a)} \left( \sum_{t=0}^{M-1} e^{-j\pi t \hat{\theta}} \right) \left( \sum_{t=0}^{M-1} e^{j\pi t \hat{\theta}} \right)$$

$$= \frac{a_{i,j}^2}{M (1 + d_{i,j}^a)} \left( \sum_{t=0}^{M-1} e^{-j\pi t \hat{\theta}} \right) \left( \sum_{t=0}^{M-1} e^{j\pi t \hat{\theta}} \right)$$

Then, we have

$$E\left\{ |h_{i,j}^H w_i|^2 \right\}$$

$$= E \left\{ \frac{a_{i,j}^2}{M (1 + d_{i,j}^a)} \left( \sum_{t=1}^{M-1} (M - t) e^{-j\pi t \hat{\theta}} + e^{j\pi t \hat{\theta}} \right) \right\}$$

$$= \int_0^R \frac{1}{R} \times \frac{a_{i,j}^2}{1 + r_{i,j}^a} d r_{i,j}$$

$$\times \int_{\frac{1}{M} (l-\frac{1}{2}) \frac{\pi}{N_s}}^{\frac{1}{M} (l-\frac{1}{2}) \frac{\pi}{N_s}} N_s$$

$$\times \frac{1}{M} \left( M + \sum_{t=1}^{M-1} (M - t) (e^{-j\pi t \hat{\theta}} + e^{j\pi t \hat{\theta}}) \right) d \hat{\theta}$$

$$= A \times \int_{\frac{1}{M} (l-\frac{1}{2}) \frac{\pi}{N_s}}^{\frac{1}{M} (l-\frac{1}{2}) \frac{\pi}{N_s}} N_s$$

$$\times \left( M + \sum_{t=1}^{M-1} (M - t) (e^{-j\pi t \hat{\theta}} + e^{j\pi t \hat{\theta}}) \right) d \hat{\theta}$$

Let $e^{-j\pi \hat{\theta}} = z$, and (39) can be rewritten as (40), as shown at the bottom of this page, where the term $A$ is

$$A = \int_0^R \frac{1}{R} \times \frac{a_{i,j}^2}{1 + r_{i,j}^a} d r_{i,j}.$$  

We define a new function $\Psi (x)$ for algebraic manipulation as follows

$$\Psi (x) = \int_0^x \frac{\sin t}{t} dt.$$  

When $x \to -\infty$, $\Psi (x) = -\pi/2$; when $x \to +\infty$, $\Psi (x) = \pi/2$. When $M \to +\infty$, the term $B$ of (40) is
expressed as (43), as shown at the bottom of this page, where \( \kappa = M / N_s \) denotes the ratio of the BS antenna number to the beam number. When \( M \to +\infty \), we have \( \kappa \to +\infty \). This is because the number of beams \( N_s \) is limited by RF chains. Moreover, when \( M \to +\infty \), the term \( C \) of (40) is expressed as

\[
C = \frac{N_s}{\pi M} \sum_{i=1}^{M-1} \frac{-t}{t} \left( 2 \cos \left( \frac{2\pi (l - i) t}{N_s} \right) \times \sin \frac{\pi t}{N_s} \right)
\]

\[
= -\frac{N_s}{\pi} \sum_{i=1}^{M-1} \frac{1}{M} \left( 2 \cos \left( \frac{2\pi (l - i) \kappa x}{M} \right) \times \sin \frac{\pi \kappa x}{M} \right)
\]

\[
\Rightarrow \frac{M}{\to \infty} \frac{-N_s}{\pi} \int_0^1 \left( 2 \cos \left( \frac{2\pi (l - i) \kappa x}{M} \right) \times \sin \frac{\pi \kappa x}{M} \right) dx
\]

\[
= -\frac{N_s}{\pi} \int_0^1 \left( \sin \pi \kappa (1 + 2(l - i)) x + \sin \pi \kappa (1 - 2(l - i)) x \right) dx
\]

\[
\Rightarrow 1, \text{each term of (44) tends 0 and therefore (44) tends 0. Consequently, substituting (43) and (44) into (40), we have (45), as shown at the bottom of this page, when } l \neq i, \Psi [\pi \kappa (1 + 2(l - i))] + \Psi [\pi \kappa (1 - 2(l - i))] = 0; \text{ otherwise, } \Psi(\pi \kappa) = \pi / 2. \text{ In addition, when the transmit SNR } \rho \text{ is high, } 1/\rho \to 0. \text{ Substituting (45) into (22), we have }
\]

\[
1 = \sum_{j=1}^{K_i} \frac{\gamma (1 + \gamma)^{K_i-j} \left( E \left\{ \sum_{i=1}^{N_s} \left| h_{ij}^H w_i \right|^2 \right\} + \frac{1}{\rho} \right)}{A (1 + N_s)}
\]

\[
= \sum_{j=1}^{K_i} \frac{\gamma (1 + \gamma)^{K_i-j} (N_s - 1) A}{A (1 + N_s)}
\]

\[
= \frac{N_s - 1}{N_s + 1} \sum_{j=1}^{K_i} \gamma (1 + \gamma)^{K_i-j}
\]

\[
= \frac{N_s - 1}{N_s + 1} (1 + \gamma)^{K_i-1}.
\]

Furthermore, we rewrite (46) as

\[
(1 + \gamma)^{K_i-1} = \frac{N_s + 1}{N_s - 1} = 1 + \frac{2}{N_s - 1}.
\]

Finally, we have

\[
\gamma = \left( \frac{2N_s}{N_s - 1} \right)^{1/K_i} - 1.
\]

Thus, \( \gamma \) is a decreasing function of the number of beams \( N_s \) and the number of users \( K_i \) in the \( i \)-th cluster, respectively. We complete the proof.

\[
B = \frac{N_s}{\pi M} \sum_{i=1}^{M-1} \frac{M}{t} \left( 2 \cos \left( \frac{2\pi (l - i) t}{N_s} \right) \times \sin \frac{\pi t}{N_s} \right)
\]

\[
= \frac{N_s}{\pi} \sum_{i=1}^{M-1} \frac{1}{M} \left( 2 \cos \left( \frac{2\pi (l - i) \kappa x}{M} \right) \times \sin \frac{\pi \kappa x}{M} \right)
\]

\[
\Rightarrow \frac{M}{\to \infty} \frac{N_s}{\pi} \int_0^1 \frac{2 \sin \pi \kappa x \times \cos \frac{2\pi (l - i) \kappa x}{M}}{x} dx
\]

\[
= \frac{N_s}{\pi} \int_0^1 \left( \sin \pi \kappa (1 + 2(l - i)) x + \sin \pi \kappa (1 - 2(l - i)) x \right) dx
\]

\[
= \frac{N_s}{\pi} \left\{ \Psi [\pi \kappa (1 + 2(l - i))] + \Psi [\pi \kappa (1 - 2(l - i))] \right\},
\]

\[
E \left\{ \left| h_{ij}^H w_i \right|^2 \right\} = A \left( 1 + B + C \right)
\]

\[
= A \left\{ 1 + \frac{N_s}{\pi} \left\{ \Psi [\pi \kappa (1 + 2(l - i))] + \Psi [\pi \kappa (1 - 2(l - i))] \right\} \right\},
\]
\[
\sum_{j=1}^{K_i} \gamma (1 + \gamma)^{K_i-j} \left( \frac{\sum_{l \neq i} \|\mathbf{h}_{l,j}^H \mathbf{w}_i\|^2 + \frac{1}{\mu}}{M} \right)
\]

\[
\Delta \rightarrow 0 \sum_{j=1}^{K_i} \gamma (1 + \gamma)^{K_i-j} \left( \frac{\sum_{l \neq i} \left( \frac{1}{M} \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \right)^2 + \frac{1}{\mu}}{M} \right)
\]

\[
\Delta \rightarrow 0 \sum_{j=1}^{K_i} \gamma (1 + \gamma)^{K_i-j} \left( \frac{\sum_{l \neq i} \left( \frac{1}{M} \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \right)^2 + \frac{1}{\mu}}{M} \right)
\]

\[
= \sum_{j=1}^{K_i} \left[ \gamma (1 + \gamma)^{K_i-j} \left( \frac{\sum_{l \neq i} \left( \frac{1}{M} \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \right)^2 + \frac{1}{\mu}}{M} \right) \right]
\]

\[
= \frac{1}{\mu} \left[ \sum_{l \neq i} \left( \frac{1}{M} \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \right)^2 \right] + \frac{1}{\mu} \left( \gamma \left( \frac{1}{\mu} \right) \right)
\]

Replacing \( (1 + d_{ij}^a) / a_{ij}^2 \) with its expectation, (51) is converted into

\[
\gamma = \left( \frac{M}{\sum_{l \neq i} \left( \frac{1}{M} \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \right)^2 + \frac{1}{\mu} \left( \gamma \left( \frac{1}{\mu} \right) \right) \right) \left( \frac{1}{\mu} \right)
\]

Further, if \( l \neq i \) and \( M = nN_s \) for \( n = 1, 2, 3, \ldots \), we have

\[
1 \sum_{m=0}^{M-1} e^{-j\pi m(l-i)\Delta \theta_j} \left( \frac{1}{M} \right)^2 = \frac{\sin^2 \left( \frac{M}{2\pi} \frac{(l-i)\Delta \theta_j}{N_s} \right)}{M \times \sin^2 \left( \frac{1}{2} \frac{(l-i)\Delta \theta_j}{N_s} \right)} = F_M \left( \pi (l-i) \frac{2}{N_s} \right) = F_M \left( \pi (l-i) \frac{2n}{M} \right) = 0.
\]

\( F_M(x) \) denotes the Fejér kernel, which equals 0 when \( x = 2\pi k/M \) for \( k = 1, 2, 3, \ldots \). Thus, (53) can be further written as (55), as shown at the top of the next page.

Thus, (53) shows the max-min SINR when the maximum azimuth angle difference between each user’s channel vector and its serving beam direction tends 0. Moreover, the output
of (55) denotes the upper bound of the max-min SINR. We complete the proof.

REFERENCES

[1] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Commun. Mag., vol. 52, no. 5, pp. 74–80, Feb. 2014.

[2] W. Cheng, X. Zhang, and H. Zhang, “Statistical-QoS driven energy-efficiency optimization over green 5G mobile wireless networks,” IEEE J. Sel. Areas Commun., vol. 34, no. 12, pp. 3092–3107, Dec. 2016.

[3] J. G. Andrews, S. Buzzi, 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.

[4] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, “Unsupervised machine learning-based user clustering in millimeter-wave-NOMA systems,” IEEE Trans. Wireless Commun., vol. 17, no. 11, pp. 7425–7440, Nov. 2018.

[5] R. W. Heath, Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter-wave MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 436–453, Apr. 2016.

[6] G. Lee, Y. Sung, and M. Kontouris, “On the performance of random beamforming in sparse millimeter wave channels,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 560–575, Apr. 2016.

[7] W. Cheng, H. Zhang, L. Liang, H. Jing, and Z. Li, “Orbital-angular-momentum embedded massive MIMO: Achieving multiplexing spectrum-efficiency for mmWave communications,” IEEE Access, vol. 6, pp. 2732–2745, 2018.

[8] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., “Spatially sparse precoding in millimeter wave MIMO systems,” IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[9] S. Sun, T. Rappaport, R. W. Heath, Jr., A. Nix, and S. Rangan, “MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?” IEEE Commun. Mag., vol. 52, no. 12, pp. 110–121, Dec. 2014.

[10] Z. Ding, P. Fan, and H. V. Poor, “Random beamforming in millimeter-wave NOMA networks,” IEEE Access, vol. 5, pp. 7676–7681, 2017.

[11] Z. Xiao, L. Zhu, Z. Gao, D. O. Wu, and X.-G. Xia, “User fairness non-orthogonal multiple access (NOMA) for millimeter-wave communications with analog beamforming,” IEEE Trans. Wireless Commun., vol. 18, no. 7, pp. 3411–3423, Jul. 2019.

[12] S. A. R. Naqvi and S. A. Hassan, “Combining NOMA and mmWave technology for cellular communication,” in Proc. IEEE 64th Veh. Technol. Conf. (VTC-Fall), Montreal, QC, Canada, Sep. 2016, pp. 1–5.

[13] Z. Ding, P. Fan, and H. V. Poor, “On the coexistence of non-orthogonal multiple access and millimeter-wave communications,” in Proc. IEEE Int. Conf. Commun. (ICC), Paris, France, May 2017, pp. 1–6.

[14] W. A. Al-Hussaibi and F. H. Ali, “Efficient user clustering, receive antenna selection, and power allocation algorithms for massive MIMO-NOMA Systems,” IEEE Access, vol. 7, pp. 31865–31882, 2019.

[15] Z. Ding and H. V. Poor, “Design of massive-MIMO-NOMA with limited feedback,” IEEE Signal Process. Lett., vol. 23, no. 5, pp. 629–633, May 2016.

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

[17] 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.

[18] L. Dai, B. Wang, M. Peng, and S. Chen, “Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer,” IEEE J. Sel. Areas Commun., vol. 37, no. 1, pp. 131–141, Jan. 2019.

[19] Z. Ding, L. Dai, R. Schober, and H. V. Poor, “NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks,” IEEE Commun. Lett., vol. 21, no. 8, pp. 1879–1882, Aug. 2017.

[20] W. Shao, S. Zhang, H. Li, N. Zhao, and O. A. Dobre, “Angle-domain NOMA over multicell millimeter wave massive MIMO networks,” IEEE Trans. Commun., vol. 68, no. 4, pp. 2277–2292, Apr. 2020, doi: 10.1109/TCOMM.2020.2968881.

[21] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, “Capacity analysis of NOMA with mmWave massive MIMO systems,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1606–1618, Jul. 2017.

[22] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” IEEE Access, vol. 4, pp. 247–267, 2016.

[23] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., “Limited feedback hybrid precoding for multi-user millimeter wave systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[24] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. Xia, “Joint Tx-Rx beamforming and power allocation for 5G millimeter-wave non-orthogonal multiple access (mmWave-NOMA) networks,” IEEE J. Sel. Topics Signal Process., vol. 67, no. 7, pp. 5114–5125, Jul. 2019.

[25] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X.-G. Xia, “Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications,” IEEE Trans. Wireless Commun., vol. 17, no. 5, pp. 2961–2974, May 2018.

[26] O. E. Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, “The capacity optimality of beam steering in large millimeter wave MIMO systems,” in Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Cesme, Turkey, Jun. 2012, pp. 100–104.

[27] G. Lee, Y. Sung, and J. Seo, “Randomly-directional beamforming in millimeter-wave multiuser MISO downlink,” IEEE Trans. Wireless Commun., vol. 15, no. 2, pp. 1086–1100, Feb. 2016.

[28] J. Li, X. Jing, Y. Zhang, and J. Mu, “Performance analysis of agile-beam NOMA in millimeter wave systems,” IEEE Access, vol. 8, pp. 6638–6649, 2020.

[29] Y. Zhang, X. Jing, and H. Gao, “An agile-beam NOMA scheme in millimeter wave networks,” IEEE Access, vol. 7, pp. 178187–178200, 2019.

[30] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, “Multi-beam NOMA for hybrid mmWave systems,” IEEE Trans. Commun., vol. 67, no. 2, pp. 1705–1719, Feb. 2019.

[31] Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, “Fairness of user clustering in MIMO non-orthogonal multiple access systems,” IEEE Commun. Lett., vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

[32] 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.

[33] 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.
[34] L. Zhu, J. Zhang, Z. Xiao, and X. Cao, “User fairness non-orthogonal multiple access (NOMA) in millimeter-wave communications,” in Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops), Beijing, China, Aug. 2018, pp. 80–84.

[35] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, “Max-min fairness linear transceiver design for a multi-user MIMO interference channel,” IEEE Trans. Signal Process., vol. 61, no. 9, pp. 2413–2423, May 2013.

[36] C. Li, C. He, L. Jiang, and F. Liu, “Robust beamforming design for max-min SINR in MIMO interference channels,” IEEE Commun. Lett., vol. 20, no. 4, pp. 724–727, Apr. 2016.

[37] C.-S. Lee and W.-H. Chung, “Max-min hybrid precoding in millimeter wave cooperative MISO systems,” in Proc. IEEE Int. Conf. Commun. (ICC), Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[38] W. Huang, Z. Lu, Y. Huang, and L. Yang, “Hybrid precoding for single carrier wideband multi-subarray millimeter wave systems,” IEEE Wireless Commun. Lett., vol. 8, no. 2, pp. 484–487, Apr. 2019.

[39] X. Hu, C. Zhong, X. Chen, W. Xu, and Z. Zhang, “Cluster grouping and power control for angle-domain mmWave MIMO NOMA systems,” IEEE J. Sel. Topics Signal Process., vol. 13, no. 5, pp. 1167–1180, Sep. 2019.

[40] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, “Optimal user scheduling and power allocation for millimeter wave NOMA systems,” IEEE Trans. Wireless Commun., vol. 17, no. 3, pp. 1502–1517, Mar. 2018.

[41] X. Song, T. Kuhn, and G. Caire, “Fully-/partially-connected hybrid beamforming architectures for mmWave MU-MIMO,” IEEE Trans. Wireless Commun., vol. 19, no. 3, pp. 1754–1769, Mar. 2020.

[42] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, “Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels,” IEEE J. Sel. Areas Commun., vol. 35, no. 6, pp. 1327–1338, Jun. 2017.

[43] 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.

[44] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, “Nonorthogonal multiple access in large-scale underlay cognitive radio networks,” IEEE Trans. Veh. Technol., vol. 65, no. 12, pp. 10152–10157, Dec. 2016.

[45] Y. Sun, Z. Ding, and X. Dai, “On the performance of downlink NOMA in multi-cell mmWave networks,” IEEE Commun. Lett., vol. 22, no. 11, pp. 2366–2369, Nov. 2018.

[46] Z. Wei, D. W. K. Ng, and J. Yuan, “NOMA for hybrid mmWave communication systems with beamwidth control,” IEEE J. Sel. Topics Signal Process., vol. 13, no. 3, pp. 567–583, Jun. 2019.

[47] H. Zhang, K. Yang, and S. Zhang, “Resource allocation based on interference alignment with clustering for data stream maximization in dense small cell networks,” IEEE Access, vol. 7, pp. 161831–161848, 2019.