On Generalized Downlink Beamforming with NOMA

Jinho Choi

Abstract: **Recently, nonorthogonal multiple access (NOMA) has been studied to increase the spectral efficiency of downlink in a multiuser system by exploiting the notion of superposition coding with successive interference cancellation (SIC). NOMA can be employed with downlink beamforming for downlink transmissions from a base station (BS) equipped with an antenna array, which results in NOMA beamforming. In this paper, we formulate a multiuser NOMA beamforming problem as a semidefinite programming (SDP) problem and generalize it in order to include the conventional (multiuser) beamforming. A low-complexity approach to decide SIC sets for the generalized NOMA beamforming is studied using the correlation between channel vectors for better performance. From analysis and simulation results, we show that the (generalized) NOMA beamforming can outperform the conventional beamforming, especially under limited scattering environments.**

Index Terms: **Downlink beamforming, nonorthogonal multiple access (NOMA), semidefinite programming, superposition coding.**

I. INTRODUCTION

THE notion of superposition coding (SC) [1] has been studied to find the capacity of multiple access channels in conjunction with successive interference cancellation (SIC) [2]. In [3], experimental evaluations of some SC approaches are carried out. SC can also be considered to improve the spectral efficiency by exploiting different powers of multiple signals in a multiple access channel. The resulting multiple access scheme is often called nonorthogonal multiple access (NOMA) and recently studied in [4]–[7]. In [8], multiuser superposition transmission (MUST) schemes are proposed to implement NOMA within standards. A power allocation method is investigated for a practical MUST scheme in [9]. In [7], [10], multiple input multiple output (MIMO) for NOMA is studied to see how NOMA can be applied to MIMO systems. In addition, in [11], NOMA is studied for downlink coordinated two-point systems. A performance analysis is presented in [12] and a power allocation problem for NOMA is studied in [13]. Furthermore, it is shown that NOMA can be employed without SIC in [14] at the cost of insignificant performance degradation.

To increase the spectral efficiency of downlink in a multiuser system, multiuser downlink beamforming can also be considered. Among various multiuser downlink beamforming approaches, multiuser downlink beamforming problems with individual quality-of-service (QoS) constraints using the signal-to-interference-plus-noise ratio (SINR) are widely studied and solved by exploiting the uplink-downlink duality [15]–[17]. They can also be solved using semidefinite programming (SDP) [18]–[20], which has been used for other beamforming problems such as multicast beamforming [19].

While there are various approaches for multiuser downlink beamforming without NOMA, only few beamforming schemes are studied with NOMA. For example, zero-forcing (ZF) approaches are considered in [5], [7] and random beams are used in [21], [22]. In [23], a minorization-maximization algorithm (MMA) is employed for form beams to maximize the sum rate in NOMA. In [24], beamforming with limited feedback is studied. In this paper, in order to improve the performance of multiuser beamforming with NOMA and exploit the power domain as well as spatial domain, we consider optimal NOMA beamforming problems by formulating an SDP problem with individual SINR constraints for downlink transmissions in a single-cell system under the assumption that a base station (BS) has perfect channel state information (CSI). This problem differs from the conventional multiuser beamforming problem due to an extended set of SINR constraints and the presence of SIC for NOMA, which changes the SINR formulation in deriving optimal beams. In addition, the problem in this paper differs from that in [23] as we have SINR constraints to guarantee certain QoS.

Interestingly, we show that the optimal performance of NOMA beamforming can be worse than that of the conventional multiuser beamforming (without NOMA) when the correlation between channels is sufficiently low. In other words, the performance improvement by exploiting the power and space domains using SC with SIC may not be better than that by only exploiting the space domain for channel vectors when the interference is not significant due to a low spatial correlation. To avoid this problem, we consider a generalization of NOMA beamforming using variable SIC sets that can take into account the spatial correlation between channel vectors.

From the analysis and simulation results in this paper, we can see that the generalized NOMA beamforming can outperform the conventional multiuser beamforming, especially under limited scattering environments, which would be typical in millimeter-wave based cellular systems [25]–[27]. Thus, the generalized NOMA beamforming might be well-suited to beamforming in millimeter-wave based cellular systems to improve the spectral efficiency. While we consider optimal design for generalized NOMA beamforming in this paper, we note that it is also possible to consider random beamforming for NOMA in millimeter-wave systems as in [28].

While we mainly focus on the optimization for NOMA beam-

Manuscript received November 14, 2016; approved for publication by Sinan Gezici, Division II Editor, February 7, 2017.

This work was supported by the Gwangju Institute of Science and Technology (GIST) Research Institute (GRI) in 2017.

The author is with School of Electrical Engineering and Computer Science, GIST, Korea, email: jchoi0114@gist.ac.kr.

Digital object identifier: 10.1109/JCN.2017.000056

forming with a generalization in a narrow sense by introducing each user's SIC set in this paper, there could be different generalizations of NOMA (not just NOMA beamforming) that include radio resource (e.g., power, bandwidth, and beam) allocation issues for multicarrier systems as in [22]. Thus, we could consider optimization problems for NOMA beamforming as well as other radio resource allocation for multicarrier systems, which would be a further research topic.

The main contributions of the paper are i) to formulate an optimization problem for optimal NOMA beamforming as an SDP problem; ii) to generalize NOMA beamforming with a low complexity approach to decide SIC sets based on the correlation between channel vectors through the analysis of the impact of the correlation on the generalized NOMA beamforming. From the analysis and simulation results in this paper, we can see that the generalized NOMA beamforming can outperform the conventional multiuser beamforming, especially under limited scattering environments, which would be typical in millimeter-wave based cellular systems [25]–[27].

The rest of the paper is organized as follows. In Section II, we present a system model for downlink beamforming in a single-cell with the conventional multiuser beamforming problem. We formulate a downlink multiuser beamforming problem with NOMA in Section III. We generalize the NOMA beamforming and take into account the spatial correlations between channel vectors to form SIC sets for the generalized NOMA beamforming in Section IV. After presenting simulation results in Section V, we finally conclude the paper with some remarks in Section VI.

Notation: Matrices and vectors are denoted by upper- and lower-case boldface letters, respectively. The superscripts $*$, T, and H denote the complex conjugate, transpose, Hermitian transpose, respectively. The p-norm of a vector $\mathbf{a}$ is denoted by $||\mathbf{a}||_p$ (If $p = 2$, the norm is denoted by $||\mathbf{a}||$ without the subscript). $\mathrm{Tr}(\mathbf{X})$ denotes the trace of a square matrix $\mathbf{X}$. $\mathbb{E}[\cdot]$ and $\mathrm{Var}(\cdot)$ denote the statistical expectation and variance, respectively. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

II. MULTIUSER DOWNLINK BEAMFORMING

In this section, we present a system model for multiuser downlink beamforming in a single-cell system that consists of a BS and multiple users. In addition, we formulate a well-known multiuser beamforming problem to minimize the total transmission power with SINR constraints to take into account individual QoS requirements [15]–[17].

A. System Model

Suppose that there are K users in a cell and a BS is equipped with an antenna array of L elements. Throughout the paper, we assume that each user is equipped with a single receive antenna and denote by $\mathbf{h}_k^{\mathrm{H}}$ the channel vector from the transmit antenna array at the BS to user k, which is a $1 \times L$ vector. Let s_k and $\mathbf{w}_k$ denote the data symbol and beamforming vector for user k in downlink transmissions, respectively. For convenience, it is assumed that $\mathbb{E}[s_k] = 0$ and $\mathbb{E}[|s_k|^2] = 1$ for all k. Note that the transmission power to user k is decided by $\mathbf{w}_k$, which is $||\mathbf{w}_k||^2$ and denoted by P_k (i.e., $P_k = ||\mathbf{w}_k||^2$). The transmitted signal from the BS to all K users is given by

$$\mathbf{x} = \sum_{j=1}^{K} \mathbf{w}_j s_j. \tag{1}$$

At user k, the received signal becomes

$$r_k = \mathbf{h}_k^{\mathrm{H}} \mathbf{x} + n_k = \mathbf{h}_k^{\mathrm{H}} \left(\sum_{j=1}^{K} \mathbf{w}_j s_j \right) + n_k, \tag{2}$$

where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the background noise.

B. Downlink Beamforming with SINR Constraints

Throughout the paper, we assume that the channel vectors, $\{\mathbf{h}_k\}$, are available at the BS. This can be accomplished exploiting the channel reciprocity in time division duplexing (TDD) [29] or CSI feedback [30], [31]. However, we do not discuss this issue as it is beyond the scope of the paper.

From (2), we can find the SINR at user k as follows:

$$\gamma_k = \frac{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_j|^2 + \sigma_k^2}. \tag{3}$$

To decide the beamforming vectors, we can consider the following optimization problem to minimize the total (or sum) transmission power (i.e., $\sum_{k=1}^{K} P_k = \sum_{k=1}^{K} ||\mathbf{w}_k||^2$) with SINR constraints for individual QoS requirements:

$$\begin{aligned} &\min_{\{\mathbf{w}_k\}} \textstyle\sum_{k=1}^{K} ||\mathbf{w}_k||^2 \\ &\gamma_k \geq \Gamma_k, \text{ for all } k, \end{aligned} \tag{4}$$

where Γ_k is the target SINR for user k. For convenience, this problem is referred to as **standard problem**. The resulting beamforming is referred to as the conventional (multiuser) beamforming.

There have been various approaches to find the solution of **standard problem** [15]–[17]. It is also known that semidefinite relaxation can be applied to **standard problem** in order to convert it into an SDP problem [18], [20]. While the solution of **standard problem** is well-known and can be easily found, its application might be limited to the case where the number of transmit antennas, L, is larger than or equal to the number of users, K, i.e., $L \geq K$, for sufficiently high target SINRs. This condition can be easily verified using the uplink-downlink duality [16], [32]. For the virtual uplink channel with a set of linearly independent channel vectors, $\{\mathbf{h}_k\}$, we can find beamforming vectors, $\{\mathbf{w}_k\}$, to meet sufficiently high target SINRs if $L \geq K$. On the other hand, if $L < K$ (i.e., overloaded cases), only low target SINRs may allow feasible solutions.

III. DOWNLINK BEAMFORMING IN NOMA SYSTEMS

In NOMA systems, it is possible to transmit signals to multiple users without beamforming using SC and SIC. Thus, we can consider downlink beamforming for the case of $L \geq K$ as well as that of $L < K$ (i.e., overloaded cases) in NOMA systems. In this section, we formulate a downlink beamforming problem for NOMA systems and show that it can be converted into an SDP problem.

A. NOMA Systems

In NOMA systems, we assume that all the signals are coded signals[1] at certain rates that can be decoded if the SINRs at users meet required target SINRs. This assumption is considered throughout the paper. From this assumption, we can see that $\mathbf{x}$ in (1) is a superposition coded signal.

In order to illustrate that a NOMA system can be used for the case of $K > L$, we consider the case of $K = 2$ and $L = 1$. In addition, let $w_k = \sqrt{P_k}$. The received signals are given by

$$r_1 = h_1^* \sqrt{P_1} s_1 + h_1^* \sqrt{P_2} s_2 + n_1$$
$$r_2 = h_2^* \sqrt{P_1} s_1 + h_2^* \sqrt{P_2} s_2 + n_2.$$

We can assume that user 1 is closer to the BS than user 2. Thus, we expect that P_2 might be higher than P_1. From this, user 1 can decode s_2 first and subtract it to decode s_1. At user 1, the SINR of user 2's signal is given by $\gamma_{1,2} = \frac{|h_1|^2 P_2}{|h_1|^2 P_1 + \sigma_1^2}$. If $\gamma_{1,2} \ge \Gamma_2$, user 1 can decode s_2 and subtract it as $z_1 = r_1 - h_1^* \sqrt{P_2} \hat{s}_2$, which is used to decode s_1. Here, $\hat{s}_2$ is a decoded signal of s_2 and it is assumed that $\hat{s}_2 = s_2$ if $\gamma_{1,2} \ge \Gamma_2$. The resulting scheme is called SIC. After SIC, the signal-to-noise ratio (SNR) at user 1 becomes $\gamma_{1,1} = |h_1|^2 P_1 / \sigma_1^2$. To decode s_1, we need $\gamma_{1,1} \ge \Gamma_1$. It is noteworthy that P_1 does not need to be high to compete with P_2 as s_2 is removed. Consequently, we may have a lower transmission power than a conventional system that does not employ SC and SIC. This is the main advantage of NOMA systems.

At user 2, no cancellation is assumed. Thus, the SINR becomes

$$\gamma_{2,2} = \frac{|h_2|^2 P_2}{|h_2|^2 P_1 + \sigma_2^2}.$$

Thus, for successful decoding at user 2, we need $\gamma_{2,2} \ge \Gamma_2$. As a result, we have the following constraint to decide P_1 and P_2 for successful decoding:

$$\begin{bmatrix} |h_1|^2 & 0 \\ -\Gamma_2 |h_1|^2 & |h_1|^2 \\ -\Gamma_2 |h_2|^2 & |h_2|^2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \ge \begin{bmatrix} \Gamma_1 \sigma_1^2 \\ \Gamma_2 \sigma_1^2 \\ \Gamma_2 \sigma_2^2 \end{bmatrix}. \quad (5)$$

It is desirable to find P_1 and P_2 that satisfy (5) with the minimum total power, $P_1 + P_2$. The solution for the minimum total power is given in [33]. In the next subsection, based on the above approach, we formulate a beamforming problem for NOMA systems with $K \ge 2$ users.

B. Beamforming with NOMA

As in (4), we can formulate an optimization problem to decide beamforming vectors with SC and SIC. For convenience, we assume that user k can decode the signals to user j, $j = k+1, \cdots, K$. In addition, the descending order is used for decoding. For example, if $K = 3$, user 1 is to decode s_3 first and then s_2. In this case, the SINR constraints are as follows:

$$\begin{aligned} &|\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_3|^2 \ge \Gamma_3 (\sigma_1^2 + |\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_1|^2 + |\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_2|^2) \\ &|\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_2|^2 \ge \Gamma_2 (\sigma_1^2 + |\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_1|^2) \\ &|\mathbf{h}_1^{\mathrm{H}} \mathbf{w}_1|^2 \ge \Gamma_1 \sigma_1^2. \end{aligned} \quad (6)$$

[1]Although s_k represents a data symbol to user k, it is assumed to be an element of a coded signal sequence of sufficiently long length.

As in (6), the SINR constraints can be found for any K when SC and SIC are employed. A total power minimization problem is given by

$$\begin{aligned} &\min_{\{\mathbf{w}_k\}} \sum_{k=1}^K ||\mathbf{w}_k||^2 \\ &\text{subject to } ||\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_j||^2 \ge \Gamma_j \left(\sigma_k^2 + \sum_{l=1}^{j-1} ||\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_l||^2 \right), \\ &\qquad j = k, \cdots, K; \ k = 1, \cdots, K. \end{aligned} \quad (7)$$

In (7), the set of SINR constraints is necessary for user k to decode the users' signals with the indices lager than k for SIC, while the signals to the users with the indices smaller than k are present as interference. Let $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}}$ and relax the rank-1 constraint for $\mathbf{W}_k$. Then, the minimum total power problem in (7) can be re-formulated as follows:

$$\begin{aligned} &\min_{\{\mathbf{W}_k\}} \sum_{k=1}^K \mathrm{Tr}(\mathbf{W}_k) \\ &\text{subject to} \\ &\text{(a) } \mathbf{h}_k^{\mathrm{H}} \mathbf{W}_j \mathbf{h}_k \ge \Gamma_j \left(\sigma_k^2 + \mathbf{h}_k^{\mathrm{H}} \left(\sum_{l=1}^{j-1} \mathbf{W}_l \right) \mathbf{h}_k \right), \\ &\qquad j = k, \cdots, K; \ k = 1, \cdots, K, \\ &\text{(b) } \mathbf{W}_k \succeq \mathbf{0}, \ k = 1, \cdots, K. \end{aligned} \quad (8)$$

For convenience, this problem is referred to as **NOMA problem**. In addition, the beamforming based on (8) is referred to as NOMA beamforming. This NOMA problem in (8) differs from the problem in [23] where the sum rate maximization is considered with a total transmission power constraints. In (8), we have the SINR constraints, which are required to guarantee QoS in terms of data rates.

Note that in (8), we use semidefinite relaxation as the rank-1 constraint is not imposed. This problem (i.e., **NOMA problem**) can be solved as an SDP problem [20]. However, a rank-1 solution may not be available for a large K. To see this, we need to consider the relationship between the ranks of the optimal solutions of (8), denoted by $\hat{\mathbf{W}}_k$, and the number of the constraints, which is $N_{\mathrm{C}} = \sum_{j=1}^K (K - j + 1) = K(K+1)/2$. From [34, Theorem 3.2], we have

$$\sum_{k=1}^K \mathrm{rank}^2(\hat{\mathbf{W}}_k) \le N_{\mathrm{C}} = \frac{K(K+1)}{2}. \quad (9)$$

Thus, for example, if $K = 3$, the rank of $\hat{\mathbf{W}}_k$ can be 2. To find the rank-1 solution, we may need to use successive approximations in [35], [36].

The problem in (8) is a quadratically constrained quadratic program (QCQP) problem. The worst case complexity is known to be $O((K^2 + L^2)^{3.5})$ [37]. It is also shown in [35] that the complexity order per iteration of successive approximations, which can be used to solve (8) is $O(L^3)$. In [36], it is also shown that the number of iterations is not large (e.g., few tens). Thus, the complexity to solve (8) mainly depends on L. Thus, for large arrays, the computational complexity can be prohibitively high. Thus, we may need to consider user clustering [5], [22] to simplify beamforming problems.

C. Optimal User Order

In NOMA systems, the user order is important. In general, user k has higher SNR than user k', if $k < k'$. Thus, the transmission power to user k' is higher than that to user k, which allows user k to decode the signal to user k' prior to decoding its signal provided that the target SINR for successful decoding is the same for both users k and k'. Thus, if the same target SINR is assumed, the user order can be based on SNRs as the transmission power is inversely proportional to the SNR. However, it does not take into account the interference and may not be optimal.

Based on the formulation in (8), it is possible to find an optimal user order. For all possible user orders, we can find the solutions of **NOMA problem** in (8). The user order that has the minimum total transmission power can be considered as the optimal order.

Denote by $u(k)$ the user index to be decoded at the $(K-k+1)$th stage. Let $\mathcal{U} = \{\mathbf{u} = [u(1) \cdots u(K)]^{\mathrm{T}}\}$ be the set of all the possible user orders. Then, **NOMA problem** in (8) can be extended as follows:

$$\min_{\mathbf{u}\in\mathcal{U}} \min_{\{\mathbf{W}_{u(k)}\}} \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_{u(k)})$$

subject to

$$\text{(a) } \mathbf{h}_{u(k)}^{\mathrm{H}} \mathbf{W}_{u(j)} \mathbf{h}_{u(k)} \geq \Gamma_{u(j)} \left(\sigma_{u(k)}^2 + \mathbf{h}_{u(k)}^{\mathrm{H}} \left(\sum_{l=1}^{j-1} \mathbf{W}_{u(l)} \right) \mathbf{h}_{u(k)} \right),$$
$$j = k, \cdots, K; \; k = 1, \cdots, K,$$
$$\text{(b) } \mathbf{W}_k \succeq \mathbf{0}, \; k = 1, \cdots, K. \quad (10)$$

Unfortunately, since there are $|\mathcal{U}| = K!$ possible user orders, for a large K, finding the optimal user order may require a prohibitively high computational complexity, which is one of the main difficulties in NOMA. Thus, for a large K, with taking into account the different target SINRs, we may consider the following order:

$$\frac{||\mathbf{h}_{u(1)}||^2}{\Gamma_{u(1)}} \geq \cdots \geq \frac{||\mathbf{h}_{u(K)}||^2}{\Gamma_{u(K)}}. \quad (11)$$

This ordering is similar to that in [38] apart from the normalization with the target SINRs. If there is no interference, the order in (11) is the descending order of required transmission powers, which is typically used in NOMA, provided that $\sigma_k^2 = \sigma^2$ for all k. In the presence of interference, this order is not optimal. While finding the optimal order is an important open problem, we employ the user order in (11) for simplicity throughout the paper unless stated otherwise, and let $u(k) = k$.

IV. GENERALIZED NOMA BEAMFORMING

In Section III.B, we formulated an optimization problem for NOMA beamforming. In this section, we show that NOMA beamforming can outperform conventional beamforming when the correlation between channels is high. On the other hand, conventional beamforming performs better than NOMA beamforming for low correlation. Based on this observation, we propose a generalized NOMA beamforming to take into account the correlation between channels through SIC sets.

A. Impact of Channel Correlation

To see the impact of the correlation on beamforming, we consider special cases with $K = 2$.

Property 1: Suppose that $\mathbf{h}_1 \perp \mathbf{h}_2$ with $K = 2$. Then, the total transmission power of the NOMA beamforming in (8) is greater than that of the conventional beamforming in (4). In particular, the total power difference is $\frac{1}{||\mathbf{h}_1||^2}\Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2)$. In other words, the NOMA beamforming consumes $\frac{1}{||\mathbf{h}_1||^2}\Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2)$ more transmission power than the conventional beamforming.

Proof: See Appendix A.

According to Property 1, we can see that NOMA beamforming can be worse than the conventional one if $\mathbf{h}_1$ and $\mathbf{h}_2$ are orthogonal to each other. Actually, it is pointless to perform NOMA beamforming if two channels are orthogonal as there is no interference. In general, if the correlation between channel vectors is low, we should use the conventional beamforming approach. We now consider the opposite case where $\mathbf{h}_1 \propto \mathbf{h}_2$.

Property 2: Suppose that $\mathbf{h}_2 = \alpha \mathbf{h}_1$, where $|\alpha| < 1$. Then, there is no feasible solution of **standard problem** (i.e., in the conventional beamforming) if $\Gamma_1\Gamma_2 > 1$. On the other hand, NOMA beamforming has a feasible solution. That is, the solution of **NOMA problem** has a finite total transmission power (the amount of the finite total transmission power is given in Appendix B as a closed-form expression).

Proof: See Appendix VI.B.

Based on Properties 1 and 2, we can see that the correlation between channel vectors plays a crucial role in deciding a beamforming strategy. This motivates us to consider a generalization in the following subsection.

B. Generalization

For a generalization, we define index sets $\mathcal{I}_k$ for users, $k = 1, \cdots, K$. At user k, the signals to user $l \in \mathcal{I}_k$ are to be decoded using SIC and the signals to user $q \in \mathcal{I}_k^c$ are considered as inter-user interference. Thus, k should be an element of $\mathcal{I}_k$, i.e., $k \in \mathcal{I}_k$. For convenience, $\mathcal{I}_k$ is referred to as the SIC set at user k. With the SIC set, $\mathcal{I}_k$, we can consider a generalized NOMA beamforming scheme and find optimal beamforming vectors from the following optimization problem:

$$\min_{\{\mathbf{W}_k\}} \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k)$$

subject to

$$\text{(a) } \mathbf{h}_k^{\mathrm{H}} \mathbf{W}_j \mathbf{h}_k \geq \Gamma_j \left(\sum_{l\in\mathcal{I}_k, l<j} \mathbf{h}_k^{\mathrm{H}} \mathbf{W}_l \mathbf{h}_k + \sum_{q\in\mathcal{I}_k^c} \mathbf{h}_k^{\mathrm{H}} \mathbf{W}_q \mathbf{h}_k + \sigma_k^2 \right)$$
$$j \in \mathcal{I}_k; k = 1, \cdots, K,$$
$$\text{(b) } \mathbf{W}_k \succeq \mathbf{0}, \; k = 1, \cdots, K. \quad (12)$$

For convenience, the problem in (12) is referred to as **generalized NOMA (GNOMA) problem**. The resulting NOMA beamforming is a generalization that can include the conventional and NOMA beamforming approaches. To see this, let $\mathcal{I}_k = \{k\}$. Then, the problem in (12) becomes that in (4), i.e., **standard problem**. On the other hand, if $\mathcal{I}_k = \{k, \cdots, K\}$, the problem in (12) becomes **NOMA problem** in (8).

In the generalized NOMA beamforming, it is possible to optimize the SIC sets, $\{\mathcal{I}_k\}$, for better performance as follows:

$$\begin{array}{c}\min_{\{\mathcal{I}_k\}} \min_{\{\mathbf{W}_k\}} \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k) \\ \text{subject to the SINR constraints in (12).}\end{array} \quad (13)$$

Then, this optimization leads to the generalized NOMA beamforming approach that can always provide a better performance than (at least the same as) any of the conventional and NOMA beamforming approaches. However, it suffers from a high computational complexity due to the outer optimization with respect to $\{\mathcal{I}_k\}$ in (13), which is a combinatorial optimization. To see this, consider an example of $K = 3$. Noting that k has to be an element of $\mathcal{I}_k$, we can have the following possible sets for $\mathcal{I}_k$ when $k = 1$: $\mathcal{I}_1 = \{1\}, \{1, 2\}, \{1, 3\}$, or $\{1, 2, 3\}$. Thus, for any K, the number of the possible sets for $\mathcal{I}_k$ can be given by $N_k = \sum_{i=0}^{K-k} \binom{K-k}{i} = 2^{K-k}$. This implies that there are $\prod_{k=1}^{K} N_k = 2^{\sum_{k=1}^{K}(K-k)} = 2^{\frac{K(K-1)}{2}}$ possible combinations for $\{\mathcal{I}_k\}$. In addition, if the user order is to be optimized in conjunction with the optimization of SIC sets, the number of possible combinations becomes

$$N_{\mathrm{comb}} = K! \times \prod_{k=1}^{K} N_k = K! \times 2^{\frac{K(K-1)}{2}}. \quad (14)$$

Certainly, it is computationally prohibitive to choose the best SIC sets when K is not small. For example, if $K = 5$, then there are about 10^3 possible combinations for all the possible SIC sets, $\{\mathcal{I}_k\}$.

C. Criteria for SIC Sets

In order to find good SIC sets for the generalized NOMA beamforming with low computational complexity, we may need to consider the correlation between channel vectors, $\{\mathbf{h}_k\}$. In particular, for a low correlation, it is desirable to have $\mathcal{I}_k = \{k\}$ (i.e., conventional beamforming). However, if the correlation is high, we may need to have $\mathcal{I}_k = \{1, \cdots, K\}$ (i.e., NOMA beamforming). Thus, based on the correlation between channel vectors, the SIC sets can be decided with low complexity. To this end, we consider the following criterion to decide $\mathcal{I}_k$:

$$\psi_{k,j} = |\rho(\mathbf{h}_k, \mathbf{h}_j)|, \quad (15)$$

where

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^{\mathrm{H}} \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||}. \quad (16)$$

The absolute correlation, $|\rho|$, is bounded as $0 \le |\rho| \le 1$ due to the Cauchy-Schwarz inequality. With $\psi_{k,j}$, we can decide $\mathcal{I}_k$ as

$$\mathcal{I}_k = \{j \mid \psi_{k,j} \ge \delta, \ j = k+1, \cdots, K\} \cup \{k\}, \quad (17)$$

where $\delta \in [0, 1]$ is a design parameter. If $\delta = 0$, we have $\mathcal{I}_k = \{1, \cdots, K\}$. That is, the generalized NOMA beamforming becomes the NOMA[2] beamforming in (8). However, if $\delta = 1$, we have $\mathcal{I}_k = \{k\}$ (provided that $\mathbf{h}_1, \cdots, \mathbf{h}_K$ are linearly independent), which means that the generalized NOMA beamforming becomes the conventional beamforming in (4).

In order to see the behavior of the generalized NOMA beamforming with $\{\mathcal{I}_k\}$ in (17) for a fixed δ, we first consider the probability that $\mathcal{I}_k$ has more than one element (note that $\mathcal{I}_k$ should have its user index k) under rich scattering environments (the case of limited scattering will be discussed later).

Property 3: Suppose that the $\mathbf{h}_k$'s are independent. In addition, $\mathbf{h}_k \sim \mathcal{CN}(0, \beta_k \mathbf{I})$, where $\beta_k > 0$ (the corresponding channel environment is referred to as the rich scattering environment). Then, the probability that $|\mathcal{I}_k| > 1$ is bounded as

$$\begin{aligned} &\Pr(|\mathcal{I}_k| > 1) = \Pr(\psi_{k,j} \ge \delta, \text{ for any } j \in \{k+1, \cdots, K\}) \\ &\ge P_k(\delta) \\ &= 1 - \left(2C_L \sum_{t=0}^{L-2} \binom{L-2}{t} \frac{(-1)^t}{2t+1} \left(1 - \delta^{2t+1}\right) \right)^{K-k}, \end{aligned} \quad (18)$$

where $C_L = \frac{\Gamma\left(L - \frac{1}{2}\right)}{\Gamma(L-1)\sqrt{\pi}}$. For a sufficiently large δ, let $\epsilon = 1 - \delta \ll 1$. The lower bound in (18), $P_k(\delta)$, is approximated as

$$P_k(\delta) \approx 1 - \left(1 - C_L (2\epsilon)^{L-1}\right)^{K-k} \approx (K-k) C_L (2\epsilon)^{L-1}. \quad (19)$$

Note that $P_k(\delta)$ is a lower-bound on the probability of high correlation between two channel vectors, which decreases exponentially with L.

Proof: See Appendix C.

The results in Property 3 show that the probability that the SIC set can include other signals decreases exponentially with L for a large δ (or $\epsilon = 1-\delta \ll 1$). In addition, from (19), we can see that $P_k(\delta)$ decreases exponentially with L, while it increases linearly with K. This implies that when both L and K grow at a fixed ratio, the impact of L on $P_k(\delta)$ is more dominant than that of K. In other words, the probability of high correlation decreases quickly when both L and K grow at a fixed ratio.

Note that the number of constraints in **GNOMA problem** in (12) for the generalized NOMA beamforming is

$$N_{\mathrm{C}} = \sum_{k=1}^{K} |\mathcal{I}_k|. \quad (20)$$

From (9), we can have the rank-1 solution of (12) if $N_{\mathrm{C}} < K+3$ [34]. According to Property 3, for a large δ, we have a low probability that $|\mathcal{I}_k| > 1$. Thus, we may have an overwhelming probability of rank-1 solution of **GNOMA problem** for a sufficiently large δ.

In above, we have considered the rich scattering environment for the generalized NOMA beamforming. As opposed to rich scattering environments, we now consider limited scattering environments, which might be applied to millimeter-wave communications [26], [27]. In a small-cell with millimeter-wave communications, a channel may consist of the line-of-sight (LoS)

[2]The resulting NOMA beamforming is the ordered one as we assume in (11).

path and few non-LoS paths due to a high path loss [27], [39]. In this case, the channel vectors can be highly correlated. For tractable analysis, we consider a special case where there is only LoS path without non-LoS paths and find the following result.

Property 4: Suppose that the BS has a uniform linear array (ULA) with antenna spacing d and the channel vector is given by $\mathbf{h}_k = \sqrt{\beta_k}\mathbf{a}(\theta_k)$, where $\sqrt{\beta}_k$ is the large scale fading coefficient, θ_k is the angle-of-departure (AoD) to user k, and $\mathbf{a}(\theta)$ is the array response vector (ARV) that is given by $\mathbf{a}(\theta) = \frac{1}{\sqrt{L}}[1\ e^{-i2\pi\frac{d}{\lambda}\sin\theta} \cdots e^{-i2\pi(L-1)\frac{d}{\lambda}\sin\theta}]^{\mathrm{T}}$, where $i = \sqrt{-1}$ and λ is the wavelength. In addition, assume that the θ_k's are uniformly distributed over $(-\pi/2, \pi/2)$ and independent (the resulting channel environment is referred to as the limited scattering environment). Then, the probability that $\psi_{k,j}$ is greater than or equal to $1/\sqrt{2}$ any $j \in \{k+1, \cdots, K\}$ is lower-bounded as

$$\Pr\left(\psi_{k,j} \geq \frac{1}{\sqrt{2}} \text{ for any } j \in \{k+1,\cdots,K\}\right) \geq 1 - \left(1 - \frac{4\sin^{-1}\frac{\Delta_{\mathrm{HP}}}{2}}{\pi}\right)^{K-k} \approx (K-k)\frac{2\Delta_{\mathrm{HP}}}{\pi}, \quad (21)$$

where $\Delta_{\mathrm{HP}} \approx 0.981\frac{\lambda}{Ld}$ [40] if $L \geq 10$.

Proof: See Appendix D.

Property 4 shows that the probability of high correlation decreases with L. However, the rate is slow as it follows $O(1/L)$ in the limited scattering environment. Thus, if K and L are of the same order, we can see that the probability of high correlation (e.g., $\psi_{k,j} \geq 1/\sqrt{2} = 0.7071$) can be high.

According to Properties 3 and 4, we can see that the correlation between channel vectors depends on the scattering environment. Thus, for a fixed δ, the SIC sets, $\{\mathcal{I}_k\}$, may have more elements in the limited scattering environment than those in the rich scattering environment. Furthermore, the performance of the generalized NOMA beamforming would depend on the scattering environment. This implies that the value of δ needs to be chosen carefully depending on a given set of parameters (e.g., L, K, and target SINRs) as well as the scattering environment.

We may find the optimal value of δ as follows:

$$\begin{aligned}&\min_{\delta\in[0,1]} \min_{\{\mathbf{W}_k\}} \textstyle\sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k)\\ &\text{subject to the SINR constraints in (12).}\end{aligned} \quad (22)$$

In (22), $\mathcal{I}_k$ is decided for a given δ as in (17). Note that since the generalized NOMA beamforming becomes the conventional beamforming or the NOMA beamforming if $\delta = 1$ or $\delta = 0$, respectively, the optimization with respect to δ over $[0, 1]$ can provide a performance that is better than (or equal to) those of the conventional beamforming and the NOMA beamforming at the expense of high computational complexity. For a reasonably low computational complexity, δ can be found from a finite set as $\delta \in \{0, \mu, \cdots, 1-\mu, 1\}$, where $\mu \ll 1$, while efficient ways to optimize δ might be considered as a further research topic.

It is noteworthy to see that for the outer optimization, a combinatorial optimization (with respect to $\{\mathcal{I}_k\}$) in (13) is replaced with a single-parameter one (with respect to δ) in (22), which can reduce the computational complexity significantly at the cost of degraded performance. In particular, the complexity order for the outer optimization changes from an exponential one (in K) to constant (i.e., $O(1/\mu)$). Thus, the generalized NOMA beamforming approach becomes computationally feasible by (22).

V. SIMULATION RESULTS

In this section, we present simulation results for various scattering environments with the following Rician channel model:

$$\mathbf{h}_{k,\zeta} = \sqrt{\beta_k}\left(\sqrt{\frac{\zeta}{1+\zeta}}\mathbf{a}(\theta_k) + \sqrt{\frac{1}{1+\zeta}}\mathbf{u}_k\right),$$

where $\mathbf{u}_k \sim \mathcal{CN}(0, \frac{1}{L}\mathbf{I})$ is a CSCG random vector, θ_k is the AoD to user k, and ζ is the Rician factor. Here, $\mathbf{a}(\theta)$ is the ARV for a ULA of half-wavelength spacing. Note that if $\zeta = 0$, we have rich scattering as in Property 3. On the other hand, if $\zeta \to \infty$, we have limited scattering as in Property 4. It is also assumed that the users are uniformly distributed within a cell. The radius of cell is normalized and the large scaling fading factor is decided as $\beta_k = 1/d_k^{\eta}$, where d_k is the distance between the BS and user k ($0 < d_k \leq 1$) and η is the path loss exponent, which is set to 3.8. For convenience, we assume that $\Gamma_k = \Gamma$ and $\sigma_k^2 = 1$ for all k. The total transmission power is used for the performance measure in this section. Note that we do not use any successive approximations [35], [36] in simulations to find a rank-1 solution when the solutions are not rank-1. Thus, the minimum total power from simulation results would be lower than that from a rank-1 solution. However, as will be shown later, since we can obtain a rank-1 solution with a high probability, simulation results might be reasonable.

For the generalized NOMA beamforming with optimized δ in simulations, we consider (22), with a finite number of δ as $\delta \in \{0, 0.1, \cdots, 0.9, 1\}$. Among the results with a finite number of δ, we choose the best one.

We consider the performance of beamforming in terms of the total transmission power under the rich (i.e., $\zeta = 0$) and limited (i.e., $\zeta \to \infty$) scattering environments for different target SINR, Γ, and show the results in Fig. 1, where we have $(L, K) = (4, 4)$ in the rich scattering environment and $(L, K) = (8, 4)$ in the limited scattering environment. We can see that NOMA beamforming provides almost the same performance in both rich and limited scattering environments. On the other hand, the performance of the conventional beamforming depends on the scattering environment. In particular, in the limited scattering environment, the performance of the conventional beamforming becomes poor due to a high spatial correlation. The generalized NOMA beamforming with optimized δ outperforms the others at the cost of increasing complexity. Since we consider $\delta \in \{0, 0.1, \cdots, 0.9, 1\}$, its complexity is 11 times higher than the generalized NOMA in this case.

As mentioned earlier, due to the semidefinite relaxation in (12), the solution may not be rank-1. In Fig. 2, we show the probability of rank-1 solution. It is shown that if $L \geq K = 4$ with $\delta = 0.6$, the solution is rank-1 with a high probability (nearly 1). If a rank-1 solution is not obtained, we can use successive approximations in [35], [36] to obtain a rank-1 solution.

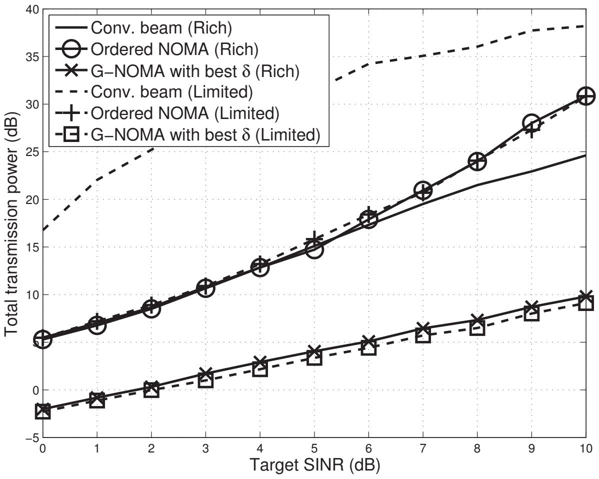

Fig. 1. Total transmission power versus target SINR with $(L, K) = (4, 4)$ for rich scattering environment and $(L, K) = (8, 4)$ for limited scattering environment.

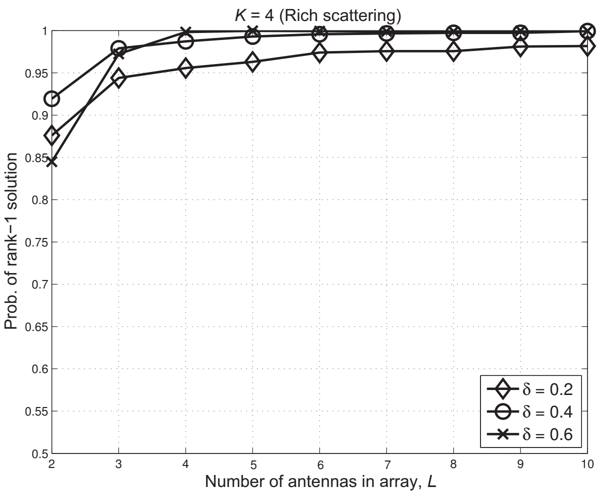

Fig. 2. Probability of rank-1 solution for the generalized NOMA beamforming for various values of L under rich scattering environments ($K = 4$ and $\Gamma = 6$ dB).

Fig. 3 shows the total transmission powers of the different beamforming methods for various values of ζ when $L = K = 4$ and $\Gamma_k = \Gamma = 6$ dB. It is shown that as the Rician factor increases (i.e., the spatial channels are more correlated), the conventional beamforming scheme performs worse than the others and its total transmission power increases, while the NOMA beamforming schemes have similar total transmission powers regardless of ζ. Note that for a large ζ (i.e., a limited scattering environment), the conventional beamforming scheme may not have feasible solutions or require a significantly high transmission power. We exclude them in obtaining the total transmission power of the conventional beamforming scheme in Fig. 3.

In order to see the impact of L on the performance under different scattering environments, we consider the generalized NOMA beamforming with optimized δ when $K = 4$ and $\Gamma = 6$ dB in Fig. 4. Under the rich scattering environment, with a reasonable performance gap (the difference between the total transmission powers of the conventional and generalized NOMA beamforming schemes), we can see that the generalized NOMA beamforming can be used even if $L < K$ (e.g., $L = 2,\ 3$), while the conventional beamforming cannot be employed (as there is no feasible solution). Under the limited scattering environment, the performance gap becomes significant due to high spatial correlations between channel vectors. Furthermore, this gap does not decrease quickly although L increases. From this, we can see that the generalized NOMA beamforming needs to be employed when the channels are highly correlated.

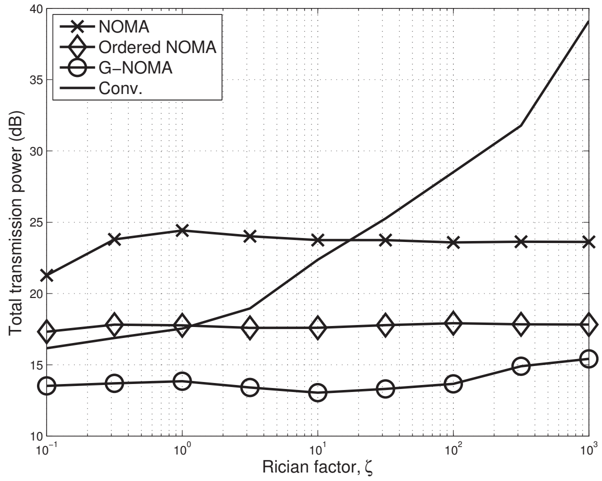

Fig. 3. Total transmission power versus Rician factor when $L = K = 4$ and $\Gamma_k = \Gamma = 6$ dB (for the generalized NOMA beamforming, we assume $\delta = 0.6$).

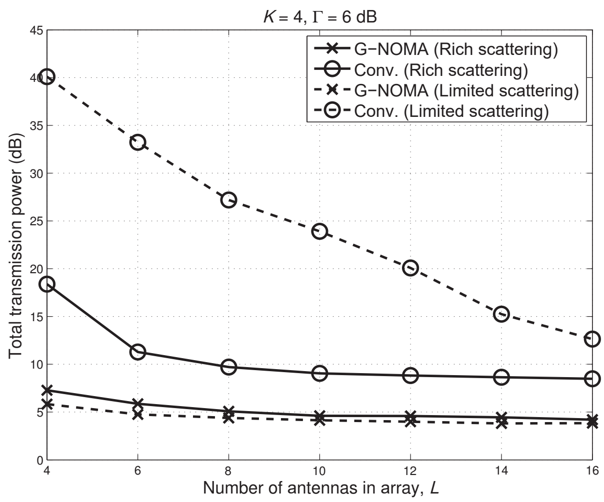

Fig. 4. Total transmission power versus L ($K = 4$ and $\Gamma = 6$ dB) under the rich and limited scattering environments.

Finally, we present simulation results when K is varying for a fixed L ($L = 4$ (12) under the rich (limited, resp.) scattering environment) with $\Gamma = 6$ dB, which are shown in Fig. 5. We also consider the generalized NOMA beamforming with optimized δ. Under the rich scattering environment, the conventional beamforming can be used up to $K = 4$ users when $L = 4$ without outage events. For the case of $K = 5$, a total transmission power of the conventional beamforming in Fig. 5 is obtained over the cases where feasible solutions exist among 1,000 runs. If $K > 5$, the conventional beamforming does not have feasible solutions, which means it cannot be employed for overloaded cases in general. On the other hand, the generalized NOMA beamforming can be employed for a wide range of K (from 2 to 7) and provide a better performance than the conventional beamforming.

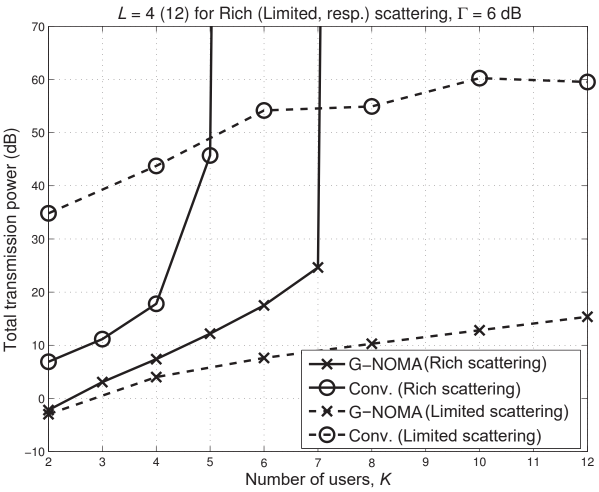

Fig. 5. Total transmission power versus K under the rich and limited scattering environments with $L = 4$ and $L = 12$, respectively, when $\Gamma = 6$ dB.

Under the limited scattering environment, the conventional beamforming suffers from high correlation and has a much higher total transmission power than that of the generalized NOMA beamforming as shown in Fig. 5. Furthermore, the probability of outage (i.e., the probability that there is no feasible solution) increases with K as shown in Fig. 6. This demonstrates that the generalized NOMA beamforming needs to be employed when the channels are highly correlated with a reasonably large number of users.

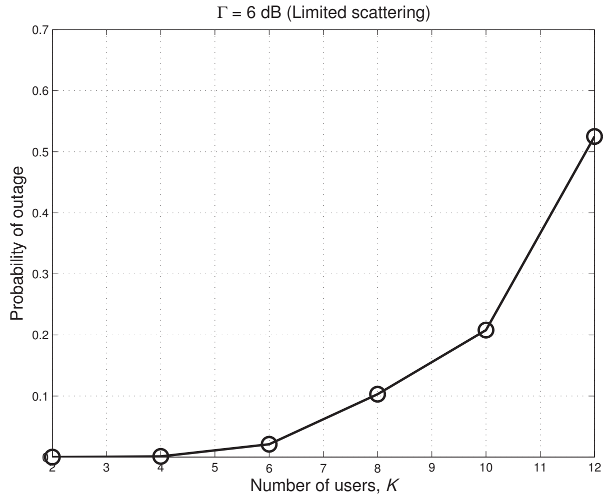

Fig. 6. The probability of outage versus K when $L = 12$ and $\Gamma = 6$ dB under the limited scattering environments.

VI. CONCLUDING REMARKS

In this paper, we studied NOMA beamforming that can exploit not only the power domain, but also the space domain to improve the spectral efficiency. In order to find optimal beamforming vectors, we formulated an SDP problem. To improve the performance further, we also considered the user ordering and a generalization. The resulting generalized NOMA beamforming can include the conventional (multiuser) beamforming and offer a better performance than the conventional and NOMA beamforming schemes. The impact of the spatial correlation between channel vectors has been studied. From this, we proposed a low-complexity approach to determine the SIC sets for the generalized NOMA beamforming. Simulation results showed that the generalized NOMA beamforming can outperform the conventional beamforming, especially under the limited scattering environment. Thus, we believe that the generalized NOMA beamforming can play a crucial role in improving the spectral efficiency for millimeter-wave based next generation systems.

APPENDIX

A. *Proof of Property 1*

For NOMA beamforming, the set of SINR constraints becomes

$$|\mathbf{h}_1^{\rm H}\mathbf{w}_2|^2 \geq \Gamma_2(|\mathbf{h}_1^{\rm H}\mathbf{w}_1|^2 + \sigma_1^2) \quad (23)$$

$$|\mathbf{h}_1^{\rm H}\mathbf{w}_1|^2 \geq \Gamma_1\sigma_1^2 \quad (24)$$

$$|\mathbf{h}_2^{\rm H}\mathbf{w}_2|^2 \geq \Gamma_2(|\mathbf{h}_2^{\rm H}\mathbf{w}_1|^2 + \sigma_2^2). \quad (25)$$

Since $\mathbf{h}_1$ and $\mathbf{h}_2$ are orthogonal, $\mathbf{w}_1$ can be chosen from the subspace that is orthogonal to $\mathbf{h}_2$ (otherwise, $\mathbf{w}_1$ introduces the interference to user 2 without improving the SNR at user 1). Thus, the constraint in (25) is reduced to

$$|\mathbf{h}_2^{\rm H}\mathbf{w}_2|^2 \geq \Gamma_2\sigma_2^2. \quad (26)$$

Noting that the minimum value of $|\mathbf{h}_1^{\rm H}\mathbf{w}_1|^2$ is $\Gamma_1\sigma_2^2$ from (24), we can see that a minimum power solution of $\mathbf{w}_1$ is $\mathbf{w}_1 = \frac{\sqrt{\Gamma_1\sigma_1^2}}{||\mathbf{h}_1||^2}\mathbf{h}_1$, and from (23) we also have the following inequality

$$|\mathbf{h}_1^{\rm H}\mathbf{w}_2|^2 \geq \Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2). \quad (27)$$

From (26) and (27), we can see that the optimal $\mathbf{w}_2$ is a linear combination of $\mathbf{h}_1$ and $\mathbf{h}_2$. Otherwise, it may have a larger norm of $\mathbf{w}_2$. Thus, we have

$$\mathbf{w}_2 = \tau_1\mathbf{h}_1 + \tau_2\mathbf{h}_2. \quad (28)$$

To decide τ_1 and τ_2, we can substitute (28) into (26) and (27). Then, a minimum power solution of $\mathbf{w}_2$ can be obtained using

$$|\tau_1|^2 = \frac{\Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2)}{||\mathbf{h}_1||^4} \text{ and } |\tau_2|^2 = \frac{\Gamma_2\sigma_2^2}{||\mathbf{h}_2||^4}.$$

Consequently, the minimum transmission powers of $\mathbf{w}_1$ and $\mathbf{w}_2$ for NOMA beamforming are given by

$$P_1 = \frac{\Gamma_1 \sigma_1^2}{||\mathbf{h}_1||^2} \text{ and } P_2 = \frac{\Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2)\sigma^2}{||\mathbf{h}_1||^2} + \frac{\Gamma_2\sigma_2^2}{||\mathbf{h}_2||^2}.$$

For the conventional beamforming, we only need to consider the SINR constraints in (24) and (25) or (26). Thus, the minimum transmission powers of $\mathbf{w}_1$ and $\mathbf{w}_2$ become $P_1 = \Gamma_1\sigma_1^2/||\mathbf{h}_1||^2$ and $P_2 = \Gamma_2\sigma_2^2/||\mathbf{h}_2||^2$, which shows that the total transmission power of the conventional beamforming is lower than that of NOMA beamforming and the transmission power gap is $\Gamma_2(\Gamma_1\sigma_1^2 + \sigma_2^2)/||\mathbf{h}_1||^2$. This completes the proof.

B. Proof of Property 2

Let $\mathbf{h} = \mathbf{h}_1$. Since $\mathbf{h}_2 = \alpha\mathbf{h}$, in order to have minimum transmission powers, we have

$$\mathbf{w}_1 = \sqrt{P_1}\bar{\mathbf{h}} \text{ and } \mathbf{w}_2 = \sqrt{P_2}\bar{\mathbf{h}}, \tag{29}$$

where $\bar{\mathbf{h}} = \mathbf{h}/||\mathbf{h}||$. For the conventional beamforming, the SINR constraints from (4) become

$$\begin{aligned} HP_1 &\ge \Gamma_1(HP_2 + \sigma_1^2) \\ |\alpha|^2 HP_2 &\ge \Gamma_2(|\alpha|^2 HP_1 + \sigma_2^2), \end{aligned} \tag{30}$$

where $H = ||\mathbf{h}_1||^2$. The two constraints can be combined and results in the following inequality:

$$HP_1 \ge \Gamma_1\Gamma_2 HP_1 + \Gamma_1\Gamma_2\frac{\sigma_2^2}{|\alpha|^2} + \Gamma_1\sigma_1^2.$$

A necessary condition to hold this inequality is $\Gamma_1\Gamma_2 \le 1$. Thus, if $\Gamma_1\Gamma_2 > 1$, there is no feasible solution for the conventional beamforming.

In NOMA beamforming, with (29), we can decide P_1 and P_2 using (5), where $|h_1|^2$ and $|h_2|^2$ are replaced with H and $|\alpha|^2 H$. From [33, Lemma 1], the solution is given by

$$\begin{aligned} P_1 &= \frac{\Gamma_1\sigma_1^2}{H} \\ P_2 &= \max\left\{\frac{\Gamma_2(\Gamma_1+1)\sigma_1^2}{H}, \frac{\Gamma_2\left(\Gamma_1\sigma_1^2 + \sigma_2^2\right)}{H}\right\}. \end{aligned} \tag{31}$$

C. Proof of Property 3

The correlation between random vectors $\mathbf{x}$ and $\mathbf{y}$ in (16) is given by $r = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||}$, and has the following distribution [41]

$$f(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi}}(1-r^2)^{\frac{n-4}{2}} = C_{\frac{n}{2}}(1-r^2)^{\frac{n-4}{2}}$$

for $-1 \le r \le 1$, if $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2\mathbf{I})$ and $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\mathbf{I})$ and they are independent. Here, n is the length of $\mathbf{x}$ and $\mathbf{y}$.

For two complex-valued vectors, $\mathbf{h}_k$ and $\mathbf{h}_j$, the correlation can be defined as in (16). For independent CSCG random vectors, let $\bar{\mathbf{h}}_k = \begin{bmatrix} \Re(\mathbf{h}_k) \\ \Im(\mathbf{h}_k) \end{bmatrix}$ and $\bar{\mathbf{h}}_j = \begin{bmatrix} \Re(\mathbf{h}_j) \\ \Im(\mathbf{h}_j) \end{bmatrix}$ with $n = 2L$. The correlation between $\bar{\mathbf{h}}_k$ and $\bar{\mathbf{h}}_j$ is $\bar{r} = \frac{\bar{\mathbf{h}}_k^{\mathrm{T}}\bar{\mathbf{h}}_j}{||\bar{\mathbf{h}}_k||\,||\bar{\mathbf{h}}_j||}$. It is noteworthy that ρ in (16) is a complex variable. Thus, $\bar{r} \neq \rho$. However, we have $\bar{r} = \Re(\rho)$, which implies that $|\bar{r}| \le |\rho|$.

The probability density function (pdf) of $\bar{r}$ is given by

$$f(\bar{r}) = C_L(1-\bar{r}^2)^{L-2},\ -1 \le \bar{r} \le 1. \tag{32}$$

The probability that $|\mathcal{I}_k| > 1$ is bounded as

$$\begin{aligned} \Pr(|\mathcal{I}_k| > 1) &= 1 - \prod_{j=k+1}^{K}\Pr(\psi_{k,j} < \delta) \\ &= 1 - (\Pr(|\rho| < \delta))^{K-k} \\ &\ge 1 - \left(\int_{-\delta}^{\delta} f(\bar{r})d\bar{r}\right)^{K-k}, \end{aligned} \tag{33}$$

where the inequality is due to $|\bar{r}| \le |\rho|$. It can be shown that

$$\begin{aligned} \int_{-x}^{x} f(\bar{r})d\bar{r} &= 2C_L\int_0^x (1-\bar{r}^2)^{L-2}d\bar{r} \\ &= 2C_L\sum_{t=0}^{L-2}\binom{L-2}{t}\frac{(-1)^t}{2t+1}x^{2t+1}. \end{aligned} \tag{34}$$

Substituting (34) into (33), we can obtain the lower-bound, $P_k(\delta)$, in (18).

If δ is sufficiently large (or close to 1), we have

$$\begin{aligned} \int_{-\delta}^{\delta} f(\bar{r})d\bar{r} &= 1 - 2C_L\int_{1-\epsilon}^{1}(1-\bar{r}^2)^{L-2}d\bar{r} \\ &\approx 1 - 2C_L\epsilon(2\epsilon)^{L-2} = 1 - C_L(2\epsilon)^{L-1}. \end{aligned} \tag{35}$$

Thus, the lower-bound can be approximated as in (19).

D. Proof of Property 4

For the ULA, we can show that [40]

$$\psi_{k,j} = \frac{1}{L}\left|\frac{\sin\left(L\pi\frac{d}{\lambda}(\sin\theta_j - \sin\theta_k)\right)}{\sin\left(\pi\frac{d}{\lambda}(\sin\theta_j - \sin\theta_k)\right)}\right|. \tag{36}$$

For a pair of k and j, $k \neq j$, $\psi_{k,j}$ becomes $1/\sqrt{2}$ when $|\sin\theta_j - \sin\theta_k| = \Delta_{\mathrm{HP}}$ [40, Eq.(2.100)]. Thus,

$$\Pr\left(\psi_{k,j} \ge \frac{1}{\sqrt{2}}\right) = \Pr(|\sin\theta_j - \sin\theta_k| \le \Delta_{\mathrm{HP}}).$$

Since

$$\sin\theta_j - \sin\theta_k = 2\sin\frac{\theta_j - \theta_k}{2}\cos\frac{\theta_k + \theta_j}{2} \le 2\sin\frac{\theta_j - \theta_k}{2},$$

we have

$$\begin{aligned} \Pr\left(\psi_{k,j} \ge \frac{1}{\sqrt{2}}\right) &\ge \Pr\left(|\theta_j - \theta_k| \le 2\sin^{-1}\frac{\Delta_{\mathrm{HP}}}{2}\right) \\ &= \frac{4\sin^{-1}\frac{\Delta_{\mathrm{HP}}}{2}}{\pi}. \end{aligned} \tag{37}$$

Since $\{\psi_{k,k+1}, \cdots, \psi_{k,K}\}$ are independent for given k, the lower-bound in (21) can be obtained from (37).

For a large L, Δ_{HP} can be sufficiently small such that $\sin^{-1}(\Delta_{\mathrm{HP}}/2) \approx \Delta_{\mathrm{HP}}/2$. Then, the approximation in (21) can be obtained for $\Delta_{\mathrm{HP}}/2 \ll 1$.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. NJ: John Wiley, second ed., 2006.

[2] R. Gallager, "An inequality on the capacity region of multiple access multipath channels," in *Communications and Cryptography: Two Sides and One Tapestry*, pp. 129–139, Kluwer, 1994.

[3] S. Vanka *et al.,* "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, 2012.

[4] Y. Saito *et al.,* "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Proc. VTC Spring*, pp. 1–5, June 2013.

[5] B. Kim *et al.,* "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE MILCOM*, pp. 1278–1283, Nov. 2013.

[6] A. G. Perotti and B. M. Popovic, "Non-orthogonal multiple access for degraded broadcast channels: RA-CEMA," *ArXiv e-prints*, Oct. 2014.

[7] Z. Ding, F. Adachi, and H. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 537–552, Jan 2016.

[8] 3GPP R1-154999, *TP for classification of MUST schemes*, 2015.

[9] J. Choi, "On the power allocation for a practical multiuser superposition scheme in noma systems," *IEEE Commun. Lett.*, vol. 20, pp. 438–441, Mar. 2016.

[10] Q. Sun, S. Han, C. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, pp. 405–408, Aug. 2015.

[11] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, pp. 313–316, Feb. 2014.

[12] Z. Ding, Z. Yang, P. Fan, and H. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, pp. 1501–1505, Dec. 2014.

[13] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, pp. 1647–1651, Oct. 2015.

[14] S. L. Shieh and Y. C. Huang, "A simple scheme for realizing the promised gains of downlink non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1624–1635, 2016.

[15] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiula, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998.

[16] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraint," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[17] M. Chiang, P. Hande, T. Lan, and C. W. Tan, *Power Control in Wireless Cellular Networks*. Foundations and Trends in Networking, Now Publisher Inc., 2008.

[18] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, 2010.

[19] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20–34, May 2010.

[20] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2009.

[21] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE VTC Fall*, pp. 1–5, Sept. 2013.

[22] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98.B, no. 3, pp. 403–414, 2015.

[23] M. Hanif, Z. Ding, T. Ratnarajah, and G. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, pp. 76–88, Jan. 2016.

[24] Z. Ding and H. V. Poor, "Design of massive-mimo-noma with limited feedback," *IEEE Signal Processing Letters*, vol. 23, pp. 629–633, May 2016.

[25] T. Rappaport *et al.,* "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[26] A. Sulyman *et al.,* "Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, pp. 78–86, Sept. 2014.

[27] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics in Sig. Process.*, vol. 8, pp. 831–846, Oct. 2014.

[28] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *CoRR*, vol. abs/1607.06302, 2016.

[29] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[30] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1478–1491, Sept. 2007.

[31] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2845–2866, June 2010.

[32] D. N. C. Tse and P. Viswanath, "Downlink-uplink duality and effective bandwidth," in *Proc. IEEE ISIT*, p. 52, June 2002.

[33] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, pp. 791–800, Mar. 2015.

[34] Y. Huang and D. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, pp. 664–678, Feb. 2010.

[35] L.-N. Tran, M. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, pp. 114–117, Jan. 2014.

[36] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Proc. Lett.*, vol. 22, pp. 804–808, July 2015.

[37] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Proc.*, vol. 54, pp. 2239–2251, June 2006.

[38] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 537–552, Jan. 2016.

[39] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1499–1513, Mar. 2014.

[40] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. NJ: John Wiley, 2002.

[41] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 1984.

Jinho Choi (SM'02) was born in Seoul, Korea. He received B.E. (magna cum laude) degree in electronics engineering in 1989 from Sogang University, Seoul, and M.S.E. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 1991 and 1994, respectively. He is with Gwangju Institute of Science and Technology (GIST) as a Professor. Prior to joining GIST in 2013, he was with the College of Engineering, Swansea University, United Kingdom, as a Professor/Chair in Wireless. His research interests include wireless communications and array/statistical signal processing. He authored two books published by Cambridge University Press in 2006 and 2010. Prof. Choi received the 1999 Best Paper Award for Signal Processing from EURASIP, 2009 Best Paper Award from WPMC (Conference), and is Senior Member of IEEE. Currently, he is an Editor of IEEE Trans. Communications and had served as an Associate Editor or Editor of other journals including IEEE Communications Letters, Journal of Communications and Networks (JCN), IEEE Transactions on Vehicular Technology, and ETRI journal.