Precoding Design and Sum Rate upper bound of RSMA Using Interference Nulling

Elaheh Sadeghabadi and Steven Blostein
Department of Electrical and Computer Engineering Queens' University, Canada

***Abstract*—Rate-splitting multiple access (RSMA) is a new multiple access technique that improves limitations of conventional multiple access techniques. RSMA manages interference by precoding at the transmitter and successive interference cancellation (SIC) at the receivers. Several precoding design approaches are considered in the literature that propose algorithms to solve a joint optimization problem. In this work, we focus on a special class of RSMA that nulls the interference of streams that must not be decoded, i.e., RSMA with interference nulling (RSMA-IN). We consider a sum rate maximization problem subject to a total power constraint and interference nulling (IN). To address the performance of precoding design algorithms relative to the optimal sum rate of RSMA-IN, we propose a convex upper bound problem for RSMA-IN sum rate and compare its performance with that of the precoding design algorithms. Further, to take advantage of the cases where the upper bound problem achieves the optimal sum rate, we propose an upper bound aided (UBA) algorithm to both enhance the performance and reduce the complexity of existing precoding design algorithms. Simulation results show that the performance of the UBA algorithm is close to the optimal sum rate of RSMA-IN.**

***Index Terms*—Rate-splitting multiple access, Broadcast channel, Precoding design, Interference nulling.**

I. Introduction

Rate-splitting (RS) is a new multiple access technique proposed for the physical (PHY) layer of future communication networks that improves the limitations of existing multiple access techniques such as space-division multiple access (SDMA) and nonorthogonal multiple access (NOMA) [1], [2].

The RSMA idea is based on message splitting and combining at the transmitter and successive interference cancellation (SIC) at the receiver [1]. In the generalized version of RSMA, the number of streams and SIC layers grow exponentially with the number of user receivers. In addition, choosing an optimal receiver SIC decoding order is computationally prohibitive even for modest numbers of users [3]. To reduce the RSMA complexity, 1-layer rate splitting (RS) and 2-layer RS [1], are proposed where the base station (BS) sends a small number of streams such that there is a unique decoding order. However, these RSMA schemes are too limited to exploit RSMA's potential performance. In [3], to improve the tradeoff between complexity and performance, we proposed *hierarchical RS* that maintains the no-ordering property, while the numbers of streams and SIC layers are allowed to grow linearly with the number of users. The focus of this paper is on precoding design for RSMA in general and hierarchical RS in particular.

RSMA manages interference by precoding at the transmitter and SIC at the receiver. RSMA precoding design is analyzed for various scenarios to optimize objective functions such as sum rate (SR) [3], [4], weighted sum rate (WSR) [1], [4], [5], energy efficiency (EE) [4]–[7], and worst-case user rate [8]. Most of these works propose algorithms based on successive convex approximation (SCA) [5], [7] and weighted minimum mean square error (WMMSE) [1], [5], [8] to solve the optimization problem, which are shown to converge to Karush-Kuhn-Tucker (KKT) points [9]. However, KKT conditions are neither necessary nor sufficient for a local optimum. Therefore, the solutions of these algorithms are not necessarily a local optimum of the optimization problem, which may result in performance loss.

In this paper, similar to [3], precoding is designed in two stages. Precoding plays two roles in an RSMA system: managing unintended interference, i.e., the part of interference that must be treated as noise and managing the part of interference that has to be decoded. Hence, one approach involves separating these two interference management tasks into a two-stage precoding design. In the first stage, precoding design approaches in SDMA can be used to manage unintended interference. In [3], interference nulling (IN) is proposed for the first stage of precoding, which nulls the part of interference that will not be decoded. This approach can be considered as a generalization of zero-forcing (ZF) [10] to RSMA. In the special case where the base station (BS) only sends (unsplit) private streams, IN reduces to ZF. The second precoding stage can be designed by solving a SR maximization problem. IN in RSMA (RSMA-IN) is a promising approach simplifying signal-to-interference-plus-noise-ratio (SINR) as well as sum rate expressions and enables us to formulate a precoding design algorithm as in [3] that solves decoupled optimization problems instead of a joint optimization problem to find beamforming vectors sequentially from the last stream to the first stream. Here, we refer to this algorithm as *sequentially descending precoder design* (SDPD).

The SDPD algorithm provides suboptimal sum rate performance. To examine the closeness of SDPD performance to that of the optimal RSMA-IN sum rate, we derive a convex problem for the SR upper bound and examine the closeness of SDPD's sum rate to the upper bound. The

This work was sponsored by Huawei Technologies Canada.

upper bound problem is obtained for both the general version of RSMA and hierarchical RS. It can be shown that the upper bound problem always exists for hierarchical RS [3]. Also, taking advantage of the cases that the upper bound problem has the same solution as the SR maximization problem, we provide an *upper bound aided* (UBA) algorithm to enhance existing precoding designs including SDPD [3] and augmented weighted mean square error (AWMSE) [1]. As described in Sections IV and V, the proposed UBA algorithm not only improves the performance of precoding design algorithms but also reduces their complexity by only requiring a single joint convex optimization rather than an iterative sequence of optimizations. Note that the objective of this paper is to compare the performance of precoding design algorithms for the same set of common messages. To examine RSMA performance for special cases of common messages including SDMA, NOMA, and multicasting, see [1].

The remaining sections are organized as follows: Section II specifies the system model. The convex upper bound problem is proposed for both the general case and hierarchical RS in Section III followed by the proposed UBA algorithm in Section IV. Section V presents the simulation results, and Section VI concludes the paper.

Notation: Bold upper and lower case letters denote matrices and vectors, respectively. Sets are represented by calligraphic font $\mathcal{S}$. $|.|$ and $\|.\|$ respectively represent absolute value of a scalar and Euclidean norm of a vector. $(.)^H$ and $\mathrm{Tr}(.)$ indicate conjugate-transpose and trace of a matrix, respectively. $\mathbb{R}$ and $\mathbb{C}$ represent real and complex fields of numbers.

II. System Model

Consider a downlink system including a BS with M antennas and K single antenna users, and assume that perfect channel state information (CSI) is available at both BS and users. The channel matrix is given by $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{M\times K}$, where $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ is the channel vector between the BS and user k.

Suppose that $\mathcal{K} = \{1, \ldots, K\}$ is the set of users' indices. The BS uses RSMA, which means that users' messages are split into partial messages. Then partial messages of a set of users $\mathcal{I}$ are combined and passed through an encoder to produce stream $s_{\mathcal{I}}$. For example, partial message of user 1 and partial message of user 2 are encoded into stream $s_{\{1,2\}}$. A subset of users $\mathcal{I}$ whose partial messages are combined and encoded into the stream $s_{\mathcal{I}}$ is called a *superuser*. If $|\mathcal{I}| = r$, the stream $s_{\mathcal{I}}$ is called r-order stream. If $r = 1$, $s_{\mathcal{I}}$ is a private stream and must be decoded only by one user. On the other hand, if $r > 1$, the stream $s_{\mathcal{I}}$ is a common stream and must be decoded by all users in $\mathcal{I}$. All possible scenarios of splitting messages include all nonempty subsets of $\mathcal{K}$, which is denoted by $\mathcal{S} = \{\mathcal{I} | \mathcal{I} \subseteq \mathcal{K}, |\mathcal{I}| \neq 0\}$ and has $2^K - 1$ superusers. In the low complexity RSMA schemes, the BS uses a subset of $\mathcal{S}$, i.e., $\mathcal{G}$ to split users' messages, where $|\mathcal{G}| = N \leq 2^K - 1$, which means that the BS sends N streams to the users. Note that $\mathcal{S}$ and $\mathcal{G}$ are sets of superusers. For example, for three users, $\mathcal{S} = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ and $\mathcal{G}$ can be $\{\{1\}, \{2\}, \{3\}, \{1,2,3\}\}$. The uncoded message of user $u \in \mathcal{K}$, is denoted by W_u, which is split into partial messages $\left\{W_u^{\mathcal{I}} | \mathcal{I} \in \mathcal{G}, u \in \mathcal{I}\right\}$. The partial messages of the users with the same superscript are encoded into the same stream. That is, the partial messages $\left\{W_{k'}^{\mathcal{I}} | k' \in \mathcal{I}\right\}$ are encoded into $s_{\mathcal{I}}$.

The BS sends a superposition of precoded streams to users, and the received signal by user u can be expressed as

$$y_u = \mathbf{h}_u^H \sum_{\mathcal{I}\in\mathcal{G}} \mathbf{q}_{\mathcal{I}} s_{\mathcal{I}} + n_u, \tag{1}$$

where $n_u \sim \mathcal{CN}(0,1)$, and $s_{\mathcal{I}}$ is the stream corresponding to superuser $\mathcal{I}$. The beamforming vector of stream $s_{\mathcal{I}}$ is denoted by $\mathbf{q}_{\mathcal{I}}$. Channel input covariance matrices are defined as $\mathbf{Q}_{\mathcal{I}} = \mathbb{E}\left\{\mathbf{q}_{\mathcal{I}} s_{\mathcal{I}} s_{\mathcal{I}}^* \mathbf{q}_{\mathcal{I}}^H\right\}$. Consider inputs $s_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}$ to be zero-mean and unit-variance. Also, we assume that each user's messages that are split into different streams are independent of one another, and different streams are independent of one another.

At the receiver side, user $u, \forall u \in \mathcal{K}$ only decodes streams containing a message intended for user u and treats other streams as noise. Streams are decoded by SIC. Therefore, the SIC decoding order must be determined. Decoding order is such that i-order streams are decoded before j-order streams such that $i > j$, and for streams of the same order, the decoding order must be optimized jointly with the beamforming vectors to maximize performance, e.g., weighted sum rate. Assume that streams are decoded in the following order: $\pi : s_{\mathcal{I}_1} \to s_{\mathcal{I}_2} \to \cdots \to s_{\mathcal{I}_N}$, where $s_{\mathcal{I}_1}$ is decoded first and $\forall i < j, |\mathcal{I}_i| > \left|\mathcal{I}_j\right|$. The interfering streams in decoding stream $s_{\mathcal{I}_i}$ at user u ($u \in \mathcal{I}_i$) are streams not intended for user u, i.e., $\left\{s_{\mathcal{I}_j} | \forall j \in \{1, \ldots, N\}, u \notin \mathcal{I}_j\right\}$ and streams intended for user u, but not yet decoded, i.e., $\left\{s_{\mathcal{I}_j} | \forall j \in \{i+1, \ldots, N\}, u \in \mathcal{I}_j\right\}$. Thus, the SINR of stream $s_{\mathcal{I}_i}$ at user u is as follows:

$$\gamma_{\mathcal{I}_i}^u = \frac{\mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_i} \mathbf{h}_u}{1 + \sum\limits_{\substack{j=i+1,\\ u\in\mathcal{I}_j}}^{N} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_j} \mathbf{h}_u + \sum\limits_{\substack{j=1,\\ u\notin\mathcal{I}_j}}^{N} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_j} \mathbf{h}_u}. \tag{2}$$

The stream $s_{\mathcal{I}}$ must be successfully decoded by any user in $\mathcal{I}$. The achievable rate of stream $s_{\mathcal{I}}$ is $R_{\mathcal{I}} = \min_{u\in\mathcal{I}} R_{\mathcal{I}}^u$, where $R_{\mathcal{I}}^u = \log_2\left(1 + \gamma_{\mathcal{I}}^u\right)$. User $u_{\mathcal{I}} = \arg\left\{\min_{u\in\mathcal{I}} \gamma_{\mathcal{I}}^u\right\}$ is called the *weakest user* of stream $s_{\mathcal{I}}$. Suppose that each stream has only one weakest user[1]. Based on this definition, the achievable sum rate is as follows:

$$R_{\mathrm{sum}} = \sum_{\mathcal{I}\in\mathcal{G}} R_{\mathcal{I}} = \sum_{\mathcal{I}\in\mathcal{G}} \min_{u\in\mathcal{I}} R_{\mathcal{I}}^u = \sum_{\mathcal{I}\in\mathcal{G}} R_{\mathcal{I}}^{u_{\mathcal{I}}}. \tag{3}$$

Considering RSMA-IN, the interference of streams not

[1]If a stream has multiple users with the lowest SINR, one of these users can be chosen arbitrarily as the weakest user.

intended for user $u, \forall u \in \mathcal{K}$ must be zero, i.e.,

$$\sum_{\substack{j=1,\\ u\notin \mathcal{I}_j}}^{N} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_j} \mathbf{h}_u = 0, \forall u \in \mathcal{K}. \quad (4)$$

Since $\mathbf{Q}_{\mathcal{I}} \succeq 0, \forall \mathcal{I} \in \mathcal{G}$, (4) can be expressed as

$$\mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u = 0, \forall \mathcal{I} \in \mathcal{G}, \forall u \in \mathcal{K} \backslash \mathcal{I}. \quad (5)$$

From (5), the SINR of ith stream at user u is as follows:

$$\gamma_{\mathcal{I}_i}^u = \frac{\mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_i} \mathbf{h}_u}{1 + \sum\limits_{\substack{j=i+1,\\ u\in \mathcal{I}_j}}^{N} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}_j} \mathbf{h}_u}. \quad (6)$$

IN is beneficial in deriving an upper bound for the achievable sum rate of RSMA-IN and designing an algorithm to compute sum rate. As explained in Section III, using IN and under certain conditions, R_{sum} can be a concave function of covariance matrices, enabling derivation of a convex upper bound problem for SR maximization. In addition, for fixed power allocation, from Equation (6), the SINR of stream $s_{\mathcal{I}_i}$ only depends on the beamforming vectors of streams with indices $j > i$, which is useful in decoupling the optimization problem and deriving a suboptimal algorithm to compute sum rate as in the SDPD algorithm [3].

III. Convex Upper Bound Problem for Sum Rate in RSMA-IN

In this section, an upper bound for the sum rate of RSMA-IN is derived. From (3) and (5), the SR maximization problem for RSMA-IN subject to a total power constraint P_{tot} can be expressed as

$$\mathcal{R}_{\text{RSMA-IN}}(\pi) = \underset{\mathbf{Q}_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}}{\text{maximize}} \quad \sum_{\mathcal{I}\in\mathcal{G}} \min_{u\in\mathcal{I}} R_{\mathcal{I}}^u \quad (7a)$$

$$\text{Subject to} \;\; \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u = 0, \quad \forall \mathcal{I} \in \mathcal{G}, \forall u \in \mathcal{K} \backslash \mathcal{I}, \quad (7b)$$

$$\mathbf{Q}_{\mathcal{I}} \succeq 0, \qquad \forall \mathcal{I} \in \mathcal{G}, \quad (7c)$$

$$\sum_{\mathcal{I}\in\mathcal{G}} \operatorname{Tr}(\mathbf{Q}_{\mathcal{I}}) \leq P_{\text{tot}}. \quad (7d)$$

Defining an auxiliary variable $\bar{R}$ and using epigraph form, Problem (7) is equivalent to the following problem:

$$\mathcal{R}_{\text{RSMA-IN}}(\pi) = \underset{\mathbf{Q}_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}, \bar{R}\in\mathbb{R}}{\text{maximize}} \quad \bar{R} \quad (8a)$$

$$\text{Subject to} \;\; R_{\text{sum}}^i = \sum_{\mathcal{I}\in\mathcal{G}} R_{\mathcal{I}}^{u_{\mathcal{I},i}} \geq \bar{R}, \quad \forall i = 1, \ldots, \prod_{\mathcal{I}\in\mathcal{G}} |\mathcal{I}|, \quad (8b)$$

$$\mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u = 0, \quad \forall \mathcal{I} \in \mathcal{G}, \forall u \in \mathcal{K} \backslash \mathcal{I}, \quad (8c)$$

$$\mathbf{Q}_{\mathcal{I}} \succeq 0, \qquad \forall \mathcal{I} \in \mathcal{G}, \quad (8d)$$

$$\sum_{\mathcal{I}\in\mathcal{G}} \operatorname{Tr}(\mathbf{Q}_{\mathcal{I}}) \leq P_{\text{tot}}, \quad (8e)$$

where $u_{\mathcal{I},i} \in \mathcal{I}$ is the weakest user of stream $s_{\mathcal{I}}$ in case i. The weakest user of stream $s_{\mathcal{I}}$ can take $|\mathcal{I}|$ possible values, and as a result, $\prod_{\mathcal{I}\in\mathcal{G}} |\mathcal{I}|$ possible sum rate cases exist in (8b). The objective function and constraints of Problem (8), except for the constraints in (8b) are linear functions of optimization variables. To have a convex problem, the left side of the inequalities in (8b) must be a concave function of covariance matrices. However, they are not necessarily a concave function of $\mathbf{Q}_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}$. Removing the nonconcave constraints in (8b), the relaxed problem provides an upper bound for the sum rate in Problem (8).

Since SINR functions in (2) are rational functions, the rate of stream $s_{\mathcal{I}}$ at user u, $R_{\mathcal{I}}^u, \forall \mathcal{I} \in \mathcal{G}, \forall u \in \mathcal{I}$, is not a concave function. However, sum of nonconcave functions is not necessarily nonconcave. One advantage of IN is that private streams' SINRs in (6) are linear functions of covariance matrices, and hence, their corresponding rates are concave functions. Without IN, none of the sum rate cases cannot be concave, because without IN, the nonconcave part in a private stream's rate cannot be removed with another term in the sum rate expression. Therefore, IN is necessary, but not sufficient, for R_{sum}^i in (8b) to be a concave function. Note that by removing all constraints in (8b), the relaxed problem becomes a feasibility check problem. Therefore, at least one concave sum rate case in (8b) must exist to find an upper bound for sum rate. In the following, necessary and sufficient conditions for R_{sum}^i to be a concave function are derived for both the general case of RSMA-IN and hierarchical RS.

A. Sum-Rate Upper Bound for the General Case

To find an upper bound of sum rate by solving a convex optimization problem, it is required to know when R_{sum} is a concave function of matrices $\mathbf{Q}_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}$. Consider that $\mathcal{G}_u = \{\mathcal{I} | u_{\mathcal{I}} = u, \forall \mathcal{I} \in \mathcal{G}\}$ is the set of superusers that their weakest user is user u. Since each stream has only one weakest user, $\{\mathcal{G}_1, \mathcal{G}_2 \ldots, \mathcal{G}_K\}$ is a partition of $\mathcal{G}$, i.e., $\mathcal{G}_u$s ($\forall u \in \mathcal{K}$) have null intersection and their union is the set of all streams $\mathcal{G}$. Theorem 1 provides a condition on streams such that R_{sum} is a concave function. Due to lack of space, the proofs of theorems and corollaries are provided in [11].

Theorem 1. *R_{sum} is a concave function of $\mathbf{Q}_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}$ for any channel realization, if and only if all interfering streams in decoding stream $s_{\mathcal{I}}, \forall \mathcal{I} \in \mathcal{G}_u$ are in $\mathcal{G}_u, \forall u \in \mathcal{K}$.*

Note that IN is not mentioned as an assumption for Theorem 1. However, it is straightforward to show that without IN, the concavity condition in Theorem 1 cannot be satisfied. Therefore, IN is a necessary condition.

The idea to find SR upper bound of RSMA-IN is to remove constraints that are not concave. Since $\mathcal{G}_u^j, \forall u \in \mathcal{K}$ is a partition of $\mathcal{G}$, the jth sum rate case R_{sum}^j can be expressed as $\sum_{u\in\mathcal{K}} \sum_{\mathcal{I}\in\mathcal{G}_u^j} R_{\mathcal{I}}^{u_{\mathcal{I},j}}$. Expanding $R_{\mathcal{I}}^{u_{\mathcal{I},j}} = \log\left(1 + \gamma_{\mathcal{I}}^{u_{\mathcal{I},j}}\right)$ based on (6) and using Theorem 1, we can show that $\sum_{\mathcal{I}\in\mathcal{G}_u^j} R_{\mathcal{I}}^{u_{\mathcal{I},j}} = \log\left(1 + \sum_{\mathcal{I}\in\mathcal{G}_u^j} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u\right)$, where $\mathcal{G}_u^j, \forall u \in \mathcal{K}$ refer to the sets in the jth sum rate case in (8b). Therefore, the relaxed convex problem is as follows:

$$\mathcal{R}_{\text{RSMA-IN}}^{\text{upper}}(\pi) = \underset{\substack{\mathbf{Q}_{\mathcal{I}},\forall \mathcal{I}\in\mathcal{G},\\ \bar{R}\in\mathbb{R}}}{\text{maximize}} \quad \bar{R} \tag{9a}$$

$$\text{Subject to} \;\; R_{\text{sum}}^{j} = \sum_{u\in\mathcal{K}} \log\left(1+\sum_{\mathcal{I}\in\mathcal{G}_u^j} \mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u\right) \geq \bar{R}, \quad \forall j \in C, \tag{9b}$$

$$\mathbf{h}_u^H \mathbf{Q}_{\mathcal{I}} \mathbf{h}_u = 0, \quad \forall u \in \mathcal{K}\backslash\mathcal{I}, \forall \mathcal{I} \in \mathcal{G}, \tag{9c}$$

$$\mathbf{Q}_{\mathcal{I}} \succeq 0, \qquad \forall \mathcal{I} \in \mathcal{G}, \tag{9d}$$

$$\sum_{\mathcal{I}\in\mathcal{G}} \text{Tr}\left(\mathbf{Q}_{\mathcal{I}}\right) \leq P_{\text{tot}}, \tag{9e}$$

where $C \subseteq \left\{1,\ldots,\prod_{\mathcal{I}\in\mathcal{G}}|\mathcal{I}|\right\}$ refers to the set of indices corresponding to concave constraints in (8b).

B. Sum-Rate Upper Bound for Hierarchical Streams

In this section, upper bound problem (9) is examined for special type of streams, called *hierarchical streams*, which are defined as follows:

Definition 1 (Hierarchical Streams)**.** *A set of superusers $\mathcal{G} \subseteq \mathcal{S}$ is called hierarchical, if for any two superusers $\mathcal{I}, \mathcal{J} \in \mathcal{G}$, where $|\mathcal{I}| \leq |\mathcal{J}|$, one of the following properties holds: i)$\mathcal{I} \cap \mathcal{J} = \mathcal{I}$, or ii)$\mathcal{I} \cap \mathcal{J} = \emptyset$.*

The set of streams corresponding to hierarchical superusers are called hierarchical streams, and the class of RSMA that exploits hierarchical streams is called hierarchical RS.

An example of hierarchical streams for 7 users is $\mathcal{G} = \left\{\{1\},\ldots,\{7\},\{1,2\},\{3,4,5\},\{6,7\},\{1,2,3,4,5\},\mathcal{K}\right\}$. Also, in [3, Algorithm1], an algorithm is proposed to generate hierarchical streams.

Hierarchical streams have attractive properties, as shown next. First, for hierarchical streams, there is a unique decoding order [3], [11]. Therefore, hierarchical RS does not require exhaustive search for the optimal decoding order that maximizes sum rate.

The second advantage of hierarchical streams is related to the SR upper bound problem tightness. As we next establish, for hierarchical streams, under a certain condition, the upper bound problem is optimal. Hence, if we observe that this condition occurs with high probability for different channel realizations, then the upper bound is tight for hierarchical streams.

Theorem 2 proposes the necessary and sufficient condition for the tightness of the SR upper bound problem of hierarchical RS. Here, "tightness" refers to the case that Problem (9) has the same solution as the SR maximization problem.

Theorem 2. *For hierarchical streams, the upper bound problem (9) is tight for any channel realization if and only if for any two superusers $\mathcal{I}, \mathcal{J} \in \mathcal{G}$, where $|\mathcal{I}| \leq |\mathcal{J}|$, one of the following properties holds: i) $u_{\mathcal{J}} = u_{\mathcal{I}}$, or ii) $u_{\mathcal{J}} \in \mathcal{J}\backslash\mathcal{I}$, where $u_{\mathcal{I}}$ and $u_{\mathcal{J}}$ are the weakest users of streams $s_{\mathcal{I}}$ and $s_{\mathcal{J}}$, respectively.*

In Section V, we observe that the tightness probability of the upper bound problem is high for hierarchical RS.

As for the third advantage, from Theorem 2, it can be shown that the upper bound problem in (9) always provides an upper bound for the sum rate of hierarchical streams. If C in the constraints (9b) is empty, Problem (9) becomes a feasibility check problem, and no upper bound is achieved for the sum rate. It can be proven that C is not empty for hierarchical streams.

Corollary 1. *Problem (9) always provides an upper bound for the sum rate of hierarchical streams.*

Since according to Corollary 1, the set C is nonempty for hierarchical streams, Problem (9) is therefore not a feasibility check problem.

IV. Upper Bound Aided Algorithm

In this section, to enhance the performance of precoding design algorithms including SDPD [3] and AWMSE [1], the UBA algorithm is proposed, which takes adavantage of the upper bound problem in (9). The SDPD algorithm is presented in [3]. In addition, the details of applying AWMSE algorithm to find the sum rate of RSMA-IN can be found in [11]. To evaluate these algorithms, their performances are compared with the SR upper bound in Section V.

According to Theorems 1 and 2, in some cases the upper bound problem in (9) can be tight . In this section, the UBA algorithm is introduced to take advantage of Problem (9) in the tight case. At first, the UBA algorithm solves the upper bound problem in (9), which is a convex optimization problem. Then, it examines whether or not the solution is tight. To check the tightness, the sum rate in (3) is calculated based on the covariance matrices obtained by the upper bound problem. If the derived sum rate is not less than the optimal value of Problem (9), it means that the solution of (9) is tight. In the case that the upper bound problem is not tight, SDPD or AWMSE can be used to design beamforming vectors.

In the tight case, UBA algorithm has the complexity of solving one convex optimization problem. Otherwise, the complexity of applying SDPD or AWMSE is also added. The convex problem in (9) can be solved by interior-point methods with polynomial-time complexity [12].

V. Simulation Results

In this section, the sum rate performance of the precoding design algorithms is evaluated and compared with the upper bound problem in (9) for hierarchical RS. To produce hierarchical streams, the user grouping algorithm in [3, Algorithm 1] is applied. This algorithm finds a subset of all possible common streams by hierarchical user grouping and assigns a common stream to each group. Since the hierarchical user grouping approach creates groups of users of the same size that do not have a user in common, this approach does not require user ordering. In Fig. 1, the average number of streams is represented by $\bar{N}$. Users' channels are modeled by a one-ring scattering model [14]. The BS antennas have

Algorithm 1: Upper Bound Aided (UBA)

Inputs : Channel matrix $\mathbf{H}$ and decoding order $s_{\mathcal{I}_1} \to s_{\mathcal{I}_2} \to \cdots \to s_{\mathcal{I}_N}$

Outputs: Beamforming vectors $\mathbf{q}_{\mathcal{I}_i}, \forall i \in \{1, \ldots, N\}$

Solve Problem (9);
Compute sum rate in (3);
Check the tightness of Problem (9) by comparing its optimal value with the sum rate;
if *The solution of (9) is tight* **then**
    Find $\mathbf{q}_{\mathcal{I}}$ from $\mathbf{Q}_{\mathcal{I}}$ by randomization for all $\mathcal{I} \in \mathcal{G}$ [13];
else
    Use another precoding design algorithm, e.g., SDPD or AWMSE;
end

uniform linear array structure, and users' angles with respect to the BS antennas are distributed uniformly over $[0, 2\pi)$. The same angle spread is considered for all users, denoted by $\Delta \in [0, \frac{\pi}{2})$. The greater the angle spread, the lower the correlation between users' channels. To consider the worst-case scenario, the same path-loss is considered for all users. The total transmit power is proportional to the number of users. Roughly speaking, the average received power in each user is $\frac{P_{\text{tot}}}{K}$, and the noise power is one. Therefore, the average received signal-to-noise-ratio (SNR) in dB is $10 \log_{10} P_{\text{tot}}$.

To measure the likelihood of tightness of the SR upper bound problem for hierarchical RS for a given channel realization, 200 Monte Carlo channel realizations based on one-ring scattering model in [14] are generated in a system with 7 users. Table I shows the SR upper bound tightness probability for 2-layer RS and hierarchical RS with more than 2 layers [1], [3], respectively. These tables represent the effect of total transmit power, which is proportional to the average received SNR, and the number of streams on the tightness probability of the SR upper bound problem. To consider the worst case, we produce channels without modeling path loss, and as a result, disparity of channel strengths, is not high. From Tables Ia and Ib, as SNR increases, the tightness probability of the upper bound problem decreases. In fact, as SNR and transmit power increase, the power constraint in (8e) is easier to satisfy, which intensifies the importance of nonconvex constraints in Problem (8) and causes reduction in tightness probability of the upper bound problem. From [3], the number of streams for hierarchical RS with more than 2 layers is greater than 2-layer RS. Therefore, from Tables Ia and Ib, the higher the number of streams, the lower the tightness probability of the SR upper bound problem. This is because as the number of streams increases, the number of nonconvex constraints in (8b) also increases. In general, from Tables Ia and Ib and considering that these tables show the worst case, the SR upper bound problem tightness occurs with high probability for hierarchical RS.

TABLE I: Tightness probability of SR upper bound problem

(a) Hierarchical RS (2-layer)

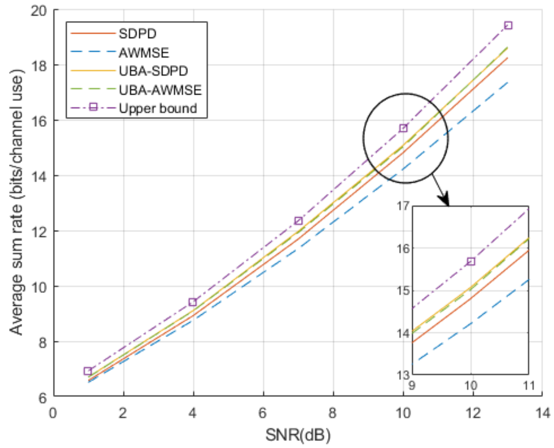
SNR / Δ	1dB	4dB	7dB	10dB	13dB
$\pi/12$	0.88	0.85	0.81	0.77	0.77
$\pi/6$	0.76	0.73	0.76	0.78	0.79

(b) Hierarchical RS (layers$\geq$ 2)

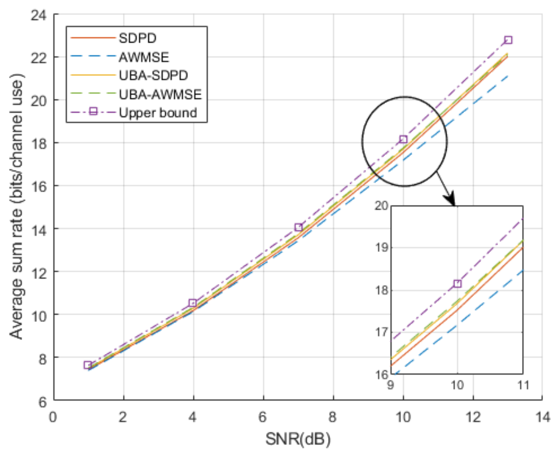
SNR / Δ	1dB	4dB	7dB	10dB	13dB
$\pi/12$	0.62	0.56	0.50	0.47	0.44
$\pi/6$	0.55	0.53	0.53	0.54	0.53

Figure 1 represents the average sum rate of hierarchical RS versus average received SNR in an RSMA-IN system with $K = 7$ users and $M = 7$ antennas at the BS. Two values are considered for the angle spread $\Delta = \frac{\pi}{12}$ radians and $\Delta = \frac{\pi}{6}$ radians. The sum rate performance is measured based on SDPD, AWMSE, UBA-SDPD, and UBA-AWMSE algorithms, where UBA-SDPD and UBA-AWMSE in Algorithm 1 respectively use SDPD and AWMSE when the SR upper bound problem is not tight. AWMSE in [1] is a precoding design algorithm based on WMMSE [15] that proposes an equivalent problem for the SR maximization problem such that its objective function is a function of mean square error (MSE). At each iteration, AWMSE computes the minimum MSE, substitutes it into the equivalent problem, and then solves a joint optimization problem to find beamforming vectors. Also, the SR upper bound based on Problem (9) is provided as a comparison point with the performance of these algorithms. As observed in Fig. 1, the performance of SDPD is superior to AWMSE in terms of sum rate, which can be explained by the fact that AWMSE does not converge to the optimal point of the SR maximization problem. Based on this figure, the performance of SDPD and UBA algorithms is near to the SR upper bound, which illustrates closeness in performance of these algorithms to the optimal sum rate. In the case that the upper bound problem is not the same as the optimal sum rate, SDPD and UBA-SDPD have the same performance. Otherwise, in the tight case, UBA-SDPD uses the optimal covariance matrices to find beamforming vectors. Hence, UBA-SDPD outperforms SDPD, and similarly, UBA-AWMSE is superior to AWMSE, which agrees with the results of Fig. 1.

As for the complexity, from simulation results, the average iterations required for the convergence of SDPD is 4.19, while it is 10.7 for AWMSE algorithm. Both algorithms are iterative, where SDPD solves N decoupled optimization problems at each iteration, where N is the number of streams, and AWMSE must solve a joint convex optimization problem at each iteration to update beamforming vectors. Note that if the SR upper bound problem is tight, UBA algorithm only solves a joint convex optimization problem, and if it is not optimal, it switches to SDPD or AWMSE, which



(a) $M = 7, K = 7, \bar{N} = 12.68, \Delta = \pi/12$



(b) $M = 7, K = 7, \bar{N} = 12.68, \Delta = \pi/6$

Fig. 1: Average sum rate of hierarchical RS obtained by SDPD, AWMSE, UBA-SDPD, and UBA-AWMSE algorithms as well as SR upper bound.

are iterative algorithms. Suppose that the computational cost of solving one joint convex optimization is C, and when the upper bound problem is not tight, UBA uses Algorithm A that solves a joint convex optimization at each iteration and requires N_{iter} iterations to converge. Also, assume that the tightness probability is p. Therefore, the complexity of Algorithm A and UBA are $N_{\text{iter}}C$ and $C + (1-p)\,N_{\text{iter}}C$, respectively. For example, for 2-layer RS at an SNR of 10dB and $\Delta = \frac{\pi}{12}$, using UBA instead of AWMSE reduces the complexity from $10.7C$ to $3.46C$ on average. Hence, the UBA algorithm not only enhances sum rate performance but also reduces complexity.

VI. Conclusion

We formulated a convex upper bound problem to assess the performance of existing precoding design algorithms. Based on the upper bound problem, our proposed UBA algorithm not only enhances performance but also reduces complexity of existing precoding design algorithms. Simulation results show that the performance of the SDPD algorithm, proposed in our previous work, is close to the SR upper bound, and the performance of UBA algorithms i.e., UBA-SDPD and UBA-AWMSE are even closer to the SR upper bound of RSMA-IN, which indicates high performance of UBA algorithms. Based on the results, for hierarchical RS, the upper bound problem has the same performance as the optimal sum rate with high probability. For example, considering the worst case that there is no channel disparity, the tightness probability of the upper bound problem for 2-layer RS is 0.77 at an SNR of 10dB and the angle spread of $\frac{\pi}{12}$ radians, which results, on average, in a factor of 3.10 computation saving by using UBA-AWMSE instead of AWMSE. Hence, with high probability, UBA algorithm solves just one convex optimization problem, which substantially reduces precoding design complexity.

References

[1] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Net.*, vol. 2018, May 2018.

[2] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, March 2019.

[3] E. Sadeghabadi and S. Blostein, "Low complexity rate splitting using hierarchical user grouping," in *IEEE ICC 2021*, Montreal, Canada, June 2021.

[4] Z. Li, C. Ye, Y. Cui, S. Yang, and S. Shamai, "Rate splitting for multi-antenna downlink: Precoder design and practical implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1910–1924, Aug. 2020.

[5] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, Dec. 2019.

[6] B. Matthiesen, Y. Mao, P. Popovski, and B. Clerckx, "Globally optimal beamforming for rate splitting multiple access," in *ICASSP 2021*, Toronto, ON, Canada, June 2021.

[7] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Lisbon, Portugal, Aug. 2018.

[8] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser miso systems: A rate-splitting approach," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[10] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. on Sig. Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[11] E. Sadeghabadi and S. Blostein, "Sum rate analysis of rate splitting multiple access using interference nulling," *in preparation*.

[12] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Methods in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[13] Z. Luo, W. Ma, A. So, Y. Ye, and S. Zhang, "Semidefnite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[14] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Aug. 2013.

[15] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.