

Synergizing Beyond Diagonal Reconfigurable Intelligent Surface and Rate-Splitting Multiple Access

Hongyu Li, *Graduate Student Member, IEEE*, Shanpu Shen, *Senior Member, IEEE*,
and Bruno Clerckx, *Fellow, IEEE*

Abstract—This work focuses on the synergy of rate-splitting multiple access (RSMA) and beyond diagonal reconfigurable intelligent surface (BD-RIS) to enlarge the coverage, improve the performance, and save on antennas. Specifically, we employ a multi-sector BD-RIS modeled as a prism, which can achieve highly directional full-space coverage, in a multiuser multiple input single output communication system. With the multi-sector BD-RIS aided RSMA model, we jointly design the transmit precoder and BD-RIS matrix under the imperfect channel state information (CSI) conditions. The robust design is performed by solving a stochastic average sum-rate maximization problem. With sample average approximation and weighted minimum mean square error-rate relationship, the stochastic problem is transformed into a deterministic one with multiple blocks, each of which is iteratively designed. Simulation results show that multi-sector BD-RIS aided RSMA outperforms space division multiple access schemes. More importantly, synergizing multi-sector BD-RIS with RSMA is an efficient strategy to reduce the number of active antennas at the transmitter and the number of passive antennas in BD-RIS.

Index Terms—Beyond diagonal reconfigurable intelligent surface, rate-splitting multiple access, robust design.

I. INTRODUCTION

The next generation of wireless communications is expected to improve ongoing techniques in 5G and explore novel techniques to support high data rate transmission and massive connectivity [1]. Specifically in the physical (PHY) layer, rate-splitting multiple access (RSMA) [2], [3] and reconfigurable intelligent surface (RIS) [4] are two promising techniques, which can improve the wireless communication performance, such as spectral efficiency, energy efficiency, and coverage.

RSMA is a PHY-layer non-orthogonal transmission strategy, which enables more flexible inter-user interference management compared to other existing multiple access techniques, such as spatial division multiple access (SDMA) and orthogonal multiple access (OMA) [2]. The key principle of (1-layer) RSMA is summarized as follows. 1) At the transmitter, user messages are split into common and private parts, where the

common parts from different users are combined and encoded into one common stream, while the private parts are individually encoded into private streams. 2) At the receiver, each user first decodes the common stream and performs successive interference cancellation (SIC) to remove the common stream, and then decodes its own private stream. Such a technique has been proved to achieve higher degree of freedom (DoF) [5], spectral and energy efficiency [6] than SDMA, and be robust to channel state information (CSI) uncertainty [5] and mobility [7]. The higher efficiency, universality, flexibility, robustness, and reliability of RSMA over existing multiple access strategies have been demonstrated in more than forty different applications and scenarios relevant to 6G [2], [3].

Meanwhile, RIS is an emerging PHY-layer technique, which enables flexible manipulations of wireless communication environments in an energy-efficient manner [4]. The key advantages of RIS come from the following two perspectives. 1) RIS is a planar surface consisting of a large number of nearly-passive elements with low power consumption. 2) RIS has tuneable elements which can reconfigure incident signals. Such a technique has been widely employed into various schemes, such as wireless power transfer (WPT) [8], simultaneous wireless information and power transfer (SWIPT) [9], integrated sensing and communication (ISAC) [10], to show the benefits of RIS in improving sum output current [8], [9] and radar output signal-to-interference-plus-noise (SINR) [10].

Conventional RIS utilized in aforementioned research [8]–[10] is modeled as a *diagonal* phase shift matrix, assuming an architecture where each RIS element is connected to ground by its own load without interacting with other elements. However, new RIS architectures have recently appeared, namely beyond diagonal RIS (BD-RIS), whose mathematical model is not limited to diagonal matrices. Until now, there are three categories of BD-RIS. The *first* category is BD-RIS with group/fully-connected architectures modeled as block diagonal matrices [11]–[16]. In this category, group/fully-connected BD-RIS with reflective mode is first proposed and modeled in [11], followed by the study of discrete-value group/fully-connected architectures [12] and optimal group/fully-connected design [13] of BD-RIS. To enlarge the coverage of RIS, simultaneous transmitting and reflecting (STAR) RIS or intelligent omni-surface (IOS) [14] is recently introduced and implemented. Then STAR-RIS/IOS is proved to be a special case of group-connected architecture with group size 2 and further generalized in [15]. Multi-sector BD-RIS is proposed in [16] to further enhance the achievable performance of RIS while guaranteeing the full-space coverage. The *second* category is BD-RIS with dynamically group-connected architectures modeled as

Manuscript received 10 March, 2023; revised 18 August, 2023; accepted 29 December, 2023. The associate editor coordinating the review of this article and approving it for publication was Theodoros Tsiftsis. (*Corresponding author: Shanpu Shen*).

Hongyu Li is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: c.li21@imperial.ac.uk).

Shanpu Shen is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. (email: Shanpu.Shen@liverpool.ac.uk).

Bruno Clerckx is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. and with Silicon Austria Labs (SAL), Graz A-8010, Austria (e-mail: b.clerckx@imperial.ac.uk; bruno.clerckx@silicon-austria.com).

permuted block diagonal matrix [17]. Different from the first category, where the architectures are fixed regardless of CSI, a dynamically group-connected architecture adapting to CSI is proposed in [17] based on a dynamic grouping strategy to further enhance the performance. The *third* category is BD-RIS with non-diagonal phase shift matrix [18]. In this category, a novel architecture, where signals impinging on one element can be reflected from another element, is modeled, analyzed, and proved to achieve better performance than conventional RIS with diagonal phase shift matrices.

Due to the potential advantages of RSMA and RIS, the synergy among them has recently gained much attention [19], [20]. The synergy among the conventional RIS and 1-layer RSMA in multiuser systems is most commonly considered, whose benefits are shown by focusing on various metrics, such as sum-rate maximization [21], energy efficiency maximization [22], [23], outage probability analysis [24]–[26], transmit power minimization [27], and max-min fairness [28]. In addition, the interplay between conventional RIS and RSMA beyond 1-layer architectures, such as 2-layer hierarchical RSMA [29], [30], is also studied. While the abovementioned RIS aided RSMA work [22]–[30] is restricted to perfect CSI conditions, a more practical imperfect CSI scenario is analyzed in [31]. However, the above results are restricted to the integration of RSMA and conventional RIS, which limits both the coverage and the achievable performance. To realize full-space coverage in wireless networks, there are a few results on STAR-RIS aided RSMA considering spatially correlated channels [32], hardware impairments [33], PHY layer security [34], and the coupling of transmission and reflection phases [35]. Meanwhile, to enhance the spectrum efficiency of wireless communications, there is only one work [36] considering the employment of BD-RIS with fully-connected architecture and reflective mode in RSMA transmission network, where the sum-rate enhancement compared to SDMA schemes is highlighted. Nevertheless, none of the existing work investigates the integration of BD-RIS and RSMA to simultaneously 1) enlarge the coverage, 2) boost the rate performance [36], and 3) reduce the required number of active/passive antennas to improve energy efficiency, which motivates this work.

The contributions of this work are summarized as follows.

First, we consider the integration of RSMA and multi-sector BD-RIS in a multiuser multiple input single output (MUMISO) system to simultaneously achieve the abovementioned three goals. The first two goals are achieved by adopting the multi-sector BD-RIS, which is modeled as a prism consisting of multiple sectors with each covering part of the horizontal space. More importantly, the multi-sector BD-RIS enables highly directional beams and thus can provide high performance gains, thanks to the use of high-gain antennas with narrow beamwidth. To achieve the third goal, we adopt 1-layer RSMA at the transmitter, which provides additional DoF over conventional multiuser linear precoding, enables better inter-user interference management, and has the potential to save on the number of both active and passive antennas when synergizing with BD-RIS.

Second, we propose a beamforming design algorithm to maximize the average sum-rate for multi-sector BD-RIS aided

RSMA under imperfect CSI conditions. To solve the stochastic problem, we first approximate it as a deterministic optimization by sample average approximation (SAA) methods [37], and then transform the deterministic problem into a more tractable three-block optimization based on the weighted minimum mean square error (WMMSE)-rate relationship [38]. Specifically for the design of the BD-RIS matrix block, due to the newly introduced constraint of the multi-sector BD-RIS, the approaches for conventional RIS cases are not feasible. Therefore, we propose a novel algorithm by first transforming the non-smooth objective into a smooth one, and then solving the smooth problem by manifold methods. Initialization, complexity analysis, and convergence analysis of the proposed design are also provided.

Third, we present simulation results to evaluate the ergodic sum-rate performance of the multi-sector BD-RIS aided RSMA compared to the SDMA scheme from the perspective of 1) the radiation pattern of BD-RIS antennas, 2) CSI conditions, and 3) the quality of service (QoS) threshold. Simulation results show that multi-sector BD-RIS aided RSMA always outperforms multi-sector BD-RIS aided SDMA in either of the abovementioned perspectives.

Fourth, with the same ergodic sum-rate requirement and QoS thresholds, we show the required numbers of active and passive antennas can be reduced compared to the SDMA scheme by synergizing BD-RIS with RSMA. Such reduction of antenna numbers leads to a potential energy efficiency gain.

Organization: Section II illustrates the multi-sector BD-RIS aided RSMA model. Section III provides the robust beamforming design for the multi-sector BD-RIS aided RSMA. Section IV evaluates the performance of the proposed design and Section V concludes this work.

Notations: Boldface lower- and upper-case letters indicate column vectors and matrices, respectively. \mathbb{C} , \mathbb{R} , and \mathbb{R}^+ denote the set of complex numbers, real numbers, and non-negative real numbers, respectively. $\mathbb{E}\{\cdot\}$ represents statistical expectation. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the conjugate, transpose, conjugate-transpose operations, and inversion, respectively. $\Re\{\cdot\}$ denotes the real part of a complex number. \mathbf{I}_L denotes an $L \times L$ identity matrix. $\mathbf{0}$ denotes an all-zero matrix. $\mathcal{CN}(\mathbf{0}, \Sigma)$ denotes the circularly symmetric complex Gaussian (CSCG) distribution with mean $\mathbf{0}$ and covariance matrix Σ . $\|\mathbf{A}\|_F$ denotes the Frobenius norm of matrix \mathbf{A} . $\|\mathbf{a}\|_2$ denotes the ℓ_2 -norm of vector \mathbf{a} . $|a|$ denotes the norm of variable a . $\text{vec}(\cdot)$ denotes the vectorization of a matrix. $\text{diag}(\cdot)$ denotes a diagonal matrix. $\text{Tr}\{\cdot\}$ denotes the summation of diagonal elements of a matrix. $\nabla(\cdot)$ denotes the first-order Euclidean gradient operation. $\frac{\partial(\cdot)}{\partial(\cdot)}$ denotes the first-order partial gradient operation. $[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{:,j}$, $[\mathbf{A}]_{i,j}$, and $[\mathbf{a}]_i$ denote the i -th row, the j -th column, the (i,j) -th element of matrix \mathbf{A} , and the i -th element of vector \mathbf{a} , respectively.

II. MULTI-SECTOR BD-RIS AIDED RSMA

In this section, we briefly review the concept of multi-sector BD-RIS, describe the multi-sector BD-RIS aided MUMISO system, explain the architecture of 1-layer RSMA, and illustrate the modeling of CSI uncertainty.

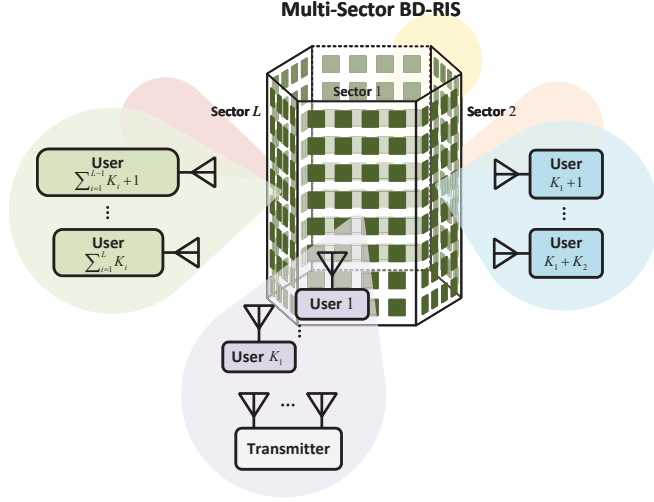


Fig. 1. A paradigm of a multi-sector BD-RIS aided MU-MISO system.

A. Multi-Sector BD-RIS

The multi-sector BD-RIS is modeled as a polygon consisting of L sectors, each of which has M antennas, as shown in Fig. 1. The multi-sector BD-RIS is modeled as multiple antennas connected to a group-connected reconfigurable impedance network [16]. In this subsection, we briefly explain the key of multi-sector BD-RIS from the following two perspectives.

1) The multi-sector BD-RIS consists of M cells with an index set $\mathcal{M} = \{1, \dots, M\}$. Particularly, cell m , $\forall m \in \mathcal{M}$ contains antennas $m, M + m, \dots, (L - 1)M + m$, which are connected to each other by reconfigurable impedance components to support the multi-sector mode.

2) Each multi-sector BD-RIS antenna has a uni-directional radiation pattern with beamwidth $2\pi/L$, which only covers $1/L$ azimuth space¹. In each cell, L antennas are located at the edge of the polygon such that L antennas cover the whole azimuth space. The larger the number of sectors L , the narrower the beamwidth of each BD-RIS antenna, and thus the higher the gain of each BD-RIS antenna.

To summarize, benefiting from group-connected impedance network and antenna array arrangements, the multi-sector BD-RIS can realize enlarged wireless coverage, more flexible wave manipulation, and enhanced channel gain compared to conventional RIS.

Mathematically, the multi-sector BD-RIS is characterized by L matrices, $\Phi_l \in \mathbb{C}^{M \times M}$, $\forall l \in \mathcal{L} = \{1, \dots, L\}$, corresponding to L sectors, which are sub-matrices of the scattering matrix $\Phi \in \mathbb{C}^{LM \times LM}$ for the LM -port reconfigurable impedance network, i.e., $\Phi_l = [\Phi]_{(l-1)M+1:LM, 1:M}$. They satisfy a combined unitary constraint given by [16]

$$\sum_l \Phi_l^H \Phi_l = \mathbf{I}_M. \quad (1)$$

More details about the modeling, design, and performance evaluation of the multi-sector BD-RIS can be found in [16].

¹In practice, the uni-directional radiation pattern can be realized by patch antennas.

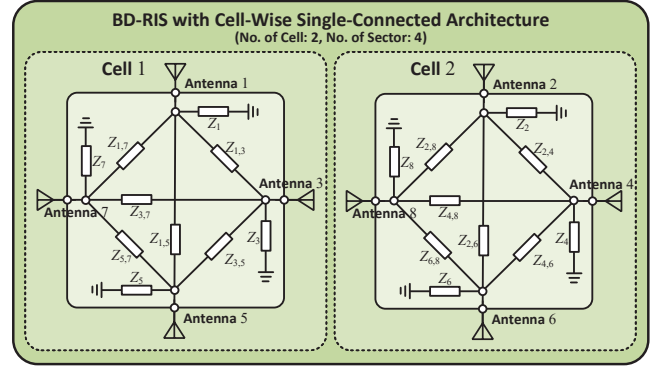


Fig. 2. Diagram of cell-wise single-connected BD-RIS with 4 sectors and 2 cells, each of which has 4 antennas.

B. Multi-Sector BD-RIS Aided MU-MISO

Consider a MU-MISO system operating in the downlink as illustrated in Fig. 1, where the transmitter equipped with N antennas communicates to a set of single-antenna users $\mathcal{K} = \{1, \dots, K\}$ with the aid of an M -cell multi-sector BD-RIS². In this work, we assume the transmitter is within the coverage of sector 1 of the multi-sector BD-RIS and K_l users in $\mathcal{K}_l = \{\sum_{i=1}^{l-1} K_i + 1, \sum_{i=1}^{l-1} K_i + 2, \dots, \sum_{i=1}^l K_i\}$ are located within the coverage of sector l , $\forall l \in \mathcal{L}$ of the multi-sector BD-RIS³, $\cup_l \mathcal{K}_l = \mathcal{K}$, as illustrated in Fig. 1. For the multi-sector BD-RIS aided downlink MU-MISO, the signal received by user k , $\forall k \in \mathcal{K}_l$ covered by sector l , $\forall l \in \mathcal{L}$ is

$$y_k = \mathbf{h}_k^H \Phi_l \mathbf{G} \mathbf{x} + n_k, \quad (2)$$

where $\mathbf{h}_k \in \mathbb{C}^M$ and $\mathbf{G} \in \mathbb{C}^{M \times N}$, respectively, are the channel between the multi-sector BD-RIS and user k , $\forall k \in \mathcal{K}$, and the channel between the transmitter and the BD-RIS, $\mathbf{x} \in \mathbb{C}^N$ is the transmit signal vector, and $n_k \sim \mathcal{CN}(0, \sigma_k^2)$, $\forall k \in \mathcal{K}$ is the additive white Gaussian noise (AWGN). The transmit signal vector \mathbf{x} is subject to the power constraint $\mathbb{E}\{\mathbf{x}^H \mathbf{x}\} \leq P$, where P is the maximum transmit power.

As discussed in [11], [15], [16], reconfigurable impedance networks with different circuit topologies have scattering matrices with different mathematical characteristics. In this work, we focus on a cell-wise single-connected (CW-SC) architecture of the multi-sector BD-RIS, where the inner-cell antennas are connected to each other by reconfigurable impedance components while the inter-cell antennas are not connected to each other⁴. To facilitate understanding, we provide an example of the BD-RIS with 4 sectors, 2 cells, and the CW-SC architecture in Fig. 2. The CW-SC architecture of the multi-sector BD-RIS results in diagonal matrices of Φ_l , $\forall l \in \mathcal{L}$,

²In this work, we ignore the direct links between the transmitter and users for simplicity, while the proposed design in the following section is still feasible when direct links exist.

³In this work, we consider the scenario that the relative location between each user and each sector of BD-RIS is fixed for simplicity such that there is no user selection issue. We also assume the sidelobe of each BD-RIS antenna is negligible compared to the mainlobe such that both transmitter and each user are covered by only one sector of the BD-RIS.

⁴The reconfigurable impedance components to construct the circuit topology can be realized by using tunable inductance and capacitance, such as varactors with continuous values or PIN diodes with discrete states.

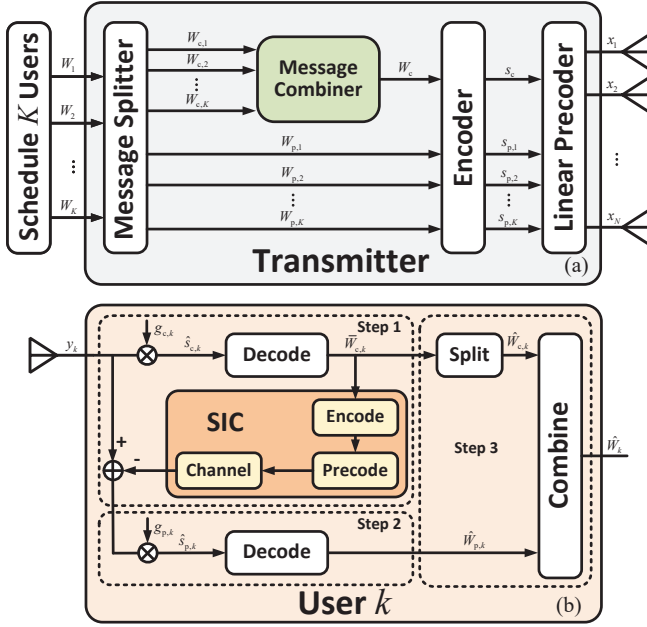


Fig. 3. (a) Transmitter and (b) receiver architectures with 1-layer RSMA.

each of which can be described as $\Phi_l = \text{diag}(\phi_{l,1}, \dots, \phi_{l,M})$ with $\phi_{l,m} \in \mathbb{C}$, $\forall l \in \mathcal{L}$, $\forall m \in \mathcal{M}$. Therefore, the constraint (1) can be re-written as

$$\sum_l |\phi_{l,m}|^2 = 1, \forall m, \quad (3)$$

which indicates that the transmit wave impinging on one sector of BD-RIS is flexibly split and reflected by other sectors, where the proportion of energy split to different sectors is determined by constraint (3). With $\Phi_l = \text{diag}(\phi_{l,1}, \dots, \phi_{l,M})$, we can rewrite equation (2) as

$$y_k = \phi_l^T \mathbf{Q}_k^H \mathbf{x} + n_k, \quad (4)$$

where $\mathbf{Q}_k = \mathbf{G}^H \text{diag}(\mathbf{h}_k) \in \mathbb{C}^{N \times M}$, $\forall k \in \mathcal{K}$ denotes the cascaded channel between the transmitter and user k , $\phi_l = [\phi_{l,1}, \dots, \phi_{l,M}]^T \in \mathbb{C}^M$, $\forall l \in \mathcal{L}$.

C. Rate-Splitting

In this work, we adopt an 1-layer RSMA architecture at the transmitter. Fig. 3 illustrates the transceiver architectures with 1-layer RSMA, where the splitting, combining, encoding, and precoding of transmit signals at the transmitter and the decoding, splitting, and combining of received signal at the user side will be described in detail.

Transmitter: The message W_k intended to user k , $\forall k \in \mathcal{K}$ is split into a common part $W_{c,k}$ and a private part $W_{p,k}$. As shown in Fig. 3(a), the common parts $W_{c,k}$ from all K users are merged into a common message W_c and further encoded into the common stream s_c , which will be decoded by all users, by a common codebook. K private parts are encoded into private streams $s_{p,1}, \dots, s_{p,K}$ and individually decoded by different users. The common stream and K private streams, $\mathbf{s} = [s_c, s_{p,1}, \dots, s_{p,K}]^T \in \mathbb{C}^{K+1}$, are linearly precoded by $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_{p,1}, \dots, \mathbf{p}_{p,K}] \in \mathbb{C}^{N \times (K+1)}$, where $\mathbf{p}_c \in \mathbb{C}^N$ is

the common precoder and $\mathbf{p}_{p,k} \in \mathbb{C}^N$ is the private precoder for user k . The resulting transmit signal is

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{p}_c s_c + \sum_k \mathbf{p}_{p,k} s_{p,k}. \quad (5)$$

Assuming $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{K+1}$, the transmit power constraint becomes $\mathbb{E}\{\mathbf{x}^H \mathbf{x}\} = \text{Tr}(\mathbf{P}\mathbf{P}^H) = \|\mathbf{P}\|_F^2 \leq P$.

Receiver: Substituting (5) into (4), we obtain the received signal at user k covered by sector l , $\forall k \in \mathcal{K}_l, \forall l \in \mathcal{L}$ as

$$y_k = \phi_l^T \mathbf{Q}_k^H \mathbf{p}_c s_c + \phi_l^T \mathbf{Q}_k^H \sum_{k'} \mathbf{p}_{p,k'} s_{p,k'} + n_k. \quad (6)$$

Therefore, the average received power of user k covered by sector l for given CSI is given by

$$\begin{aligned} \tau_{c,k} = \mathbb{E}\{|y_k|^2\} &= |\phi_l^T \mathbf{Q}_k^H \mathbf{p}_c|^2 \\ &+ \underbrace{|\phi_l^T \mathbf{Q}_k^H \mathbf{p}_{p,k}|^2}_{=\tau_{p,k}=\tau_{c,k}} + \underbrace{\sum_{k' \neq k} |\phi_l^T \mathbf{Q}_k^H \mathbf{p}_{p,k'}|^2}_{=\tau_{p,k}} + \sigma_k^2. \end{aligned} \quad (7)$$

At the user side, the signal process, taking user k covered by sector l , $\forall k \in \mathcal{K}_l, \forall l \in \mathcal{L}$ as an example, is summarized as following three steps and also illustrated in Fig. 3(b).

Step 1: User k covered by sector l decodes the equalized common stream $\hat{s}_{c,k} = g_{c,k} y_k$ with equalizer $g_{c,k} \in \mathbb{C}$ into $\hat{W}_{c,k}$ by regarding the interference induced by all K private streams as noise, i.e., $\tau_{p,k}$ or $\tau_{c,k}$ in (7). After performing SIC, $\hat{W}_{c,k}$ is subtracted from the received signal.

Step 2: User k decodes the equalized private stream $\hat{s}_{p,k} = g_{p,k} y_k$ with equalizer $g_{p,k} \in \mathbb{C}$ into $\hat{W}_{p,k}$ by regarding the interference induced by remaining $K - 1$ private streams as noise, i.e., $\tau_{p,k}$ in (7).

Step 3: User k splits its contribution to the common message, $\hat{W}_{c,k}$, from the decoded $\hat{W}_{c,k}$, and combines $\hat{W}_{c,k}$ with $\hat{W}_{p,k}$ to reconstruct \hat{W}_k .

Based on the above signal process, the instantaneous SINRs of the common stream and private streams at user k covered by sector l , $\forall k \in \mathcal{K}_l, \forall l \in \mathcal{L}$ are given by

$$\gamma_{c,k} = |\phi_l^T \mathbf{Q}_k^H \mathbf{p}_c|^2 \tau_{c,k}^{-1}, \quad \gamma_{p,k} = |\phi_l^T \mathbf{Q}_k^H \mathbf{p}_{p,k}|^2 \tau_{p,k}^{-1}. \quad (8)$$

Assuming Gaussian inputs, the instantaneous common rates and private rates for user k are

$$R_{c,k} = \log_2(1 + \gamma_{c,k}), \quad R_{p,k} = \log_2(1 + \gamma_{p,k}). \quad (9)$$

To guarantee the common message W_c is successfully decoded by all K users, the common rate should not exceed $\min_{\forall k} R_{c,k}$. Given the architecture of RSMA as illustrated above, the common rate $\min_{\forall k} R_{c,k}$ is split into K parts, yielding $\min_{\forall k} R_{c,k} = \sum_k C_k$. Each part $C_k \in \mathbb{R}^+$ corresponds to the common part of the achievable rate for user k . Finally, the total rate for user k is a summation of common part and private part, i.e., $R_k = C_k + R_{p,k}$, $\forall k \in \mathcal{K}$.

D. Channel State Information

Perfect CSI acquisition in RIS aided systems is challenging due to the passive property of RIS. Therefore, in this work, we assume the transmitter has an imperfect instantaneous CSI $\hat{\mathbf{Q}} = [\hat{\mathbf{Q}}_1, \dots, \hat{\mathbf{Q}}_K] \in \mathbb{C}^{N \times MK}$. To model

the CSI uncertainty, the perfect cascaded channel matrix $\mathbf{Q} = [\mathbf{Q}_1, \dots, \mathbf{Q}_K] \in \mathbb{C}^{N \times MK}$ is expressed as a summation of the channel estimate and channel estimation errors, that is $\mathbf{Q} = \hat{\mathbf{Q}} + \tilde{\mathbf{Q}}$, where the estimation error matrix is defined as $\tilde{\mathbf{Q}} = [\tilde{\mathbf{Q}}_1, \dots, \tilde{\mathbf{Q}}_K] \in \mathbb{C}^{N \times MK}$. The channel estimation error could come from the noise and limited uplink training in time division duplex (TDD) settings based on the typical MMSE criterion [39]⁵. In this case, the channel estimation error generally follows the CSCG distribution, that is $\text{vec}(\tilde{\mathbf{Q}}_k) \sim \mathcal{CN}(\mathbf{0}, \delta_k^2 \mathbf{I}_{MN})$, $\forall k \in \mathcal{K}$, where δ_k^2 measures the estimation inaccuracy of the cascaded channel between the transmitter and user k .

III. BEAMFORMING DESIGN WITH SAMPLE AVERAGE APPROXIMATION

In this section, we formulate the sum-rate maximization problem for BD-RIS aided RSMA based on imperfect CSI conditions and propose efficient algorithms to jointly design the transmit precoder and BD-RIS.

A. Problem Formulation

Beamforming design for RIS aided systems highly relies on the CSI among the transmitter, RIS, and users. When perfect CSI is available at the transmitter, beamforming design can be developed according to perfect channels such that the instantaneous sum-rate, $\sum_k R_k$, is maximized, and thus the long-term ergodic sum-rate given by $\mathbb{E}_{\mathbf{Q}}\{\sum_k R_k\}$. However, when the transmitter has only partial instantaneous CSI as illustrated in Section II-D, beamforming design that maximizes instantaneous sum-rate as in perfect CSI case is not possible, which makes it difficult to measure the long-term ergodic sum-rate. One robust approach to tackle this issue is to perform the beamforming design based on the fact that the transmitter has access to the average rate, which captures the statistical property of the channel estimation error with a given estimate. The average rate is defined as follows.

Definition 1: The average rate is a measure of the expected performance over a distribution of the channel estimation error, e.g., the distribution of $\tilde{\mathbf{Q}}$, with determined estimate $\hat{\mathbf{Q}}$ [5]. Mathematically, the average common and private rates for user k , $\forall k \in \mathcal{K}$, depend on the channel estimate $\hat{\mathbf{Q}}$, and are, respectively, defined as $\bar{R}_{c,k} = \mathbb{E}_{\mathbf{Q}|\hat{\mathbf{Q}}}\{R_{c,k}|\hat{\mathbf{Q}}\}$ and $\bar{R}_{p,k} = \mathbb{E}_{\mathbf{Q}|\hat{\mathbf{Q}}}\{R_{p,k}|\hat{\mathbf{Q}}\}$, where $R_{c,k}$ and $R_{p,k}$ are instantaneous rates determined by the perfect channel \mathbf{Q} as illustrated in (9). \square

With Definition 1, we can connect the ergodic rates $\mathbb{E}_{\mathbf{Q}}\{R_{c,k}\}$, $\mathbb{E}_{\mathbf{Q}}\{R_{p,k}\}$ with average rates $\bar{R}_{c,k}$, $\bar{R}_{p,k}$ as

$$\mathbb{E}_{\mathbf{Q}}\{R_{c,k}\} = \mathbb{E}_{\hat{\mathbf{Q}}}\{\mathbb{E}_{\mathbf{Q}|\hat{\mathbf{Q}}}\{R_{c,k}|\hat{\mathbf{Q}}\}\} = \mathbb{E}_{\hat{\mathbf{Q}}}\{\bar{R}_{c,k}\}, \quad (10a)$$

$$\mathbb{E}_{\mathbf{Q}}\{R_{p,k}\} = \mathbb{E}_{\hat{\mathbf{Q}}}\{\mathbb{E}_{\mathbf{Q}|\hat{\mathbf{Q}}}\{R_{p,k}|\hat{\mathbf{Q}}\}\} = \mathbb{E}_{\hat{\mathbf{Q}}}\{\bar{R}_{p,k}\}, \quad (10b)$$

⁵The existing popular and effective channel estimation strategy, where the cascaded channel is estimated by pre-defining the RIS pattern during the uplink training process to minimize MSE, can be easily extended to the multi-sector BD-RIS case with CW-SC architecture. The key and main difference compared with conventional RIS case is the RIS pattern design due to the new constraint of multi-sector BD-RIS. A simple choice is to set identical amplitude for non-zero elements of the BD-RIS matrix, i.e., $|\phi_{l,m}| = \frac{1}{\sqrt{L}}$. Then orthogonal matrices, such as discrete Fourier transform (DFT) matrix and Hadamard matrix, can be adopted to construct the RIS pattern.

which indicate that ergodic rates can be characterized by the expectation of average rates over varied channel estimate $\hat{\mathbf{Q}}$. Accordingly, maximizing the ergodic sum-rate $\mathbb{E}_{\mathbf{Q}}\{\sum_k R_k\}$ with partial CSI is transformed into maximizing the average sum-rate for each $\hat{\mathbf{Q}}$, given by $\sum_k \bar{C}_k + \sum_k \bar{R}_{p,k}$, where $\bar{C}_k \in \mathbb{R}^+$, $\forall k \in \mathcal{K}$ referring to the common part of the average rate should satisfy $\sum_k \bar{C}_k = \min_{\forall k} \bar{R}_{c,k}$. Therefore, the average sum-rate maximization problem can be formulated as

$$\max_{\{\mathbf{P}, \bar{\mathbf{c}}\}, \{\phi_l, \forall l\}} \sum_k \bar{C}_k + \sum_k \bar{R}_{p,k} \quad (11a)$$

$$\text{s.t.} \quad \sum_l |\phi_{l,m}|^2 = 1, \forall m, \quad (11b)$$

$$\|\mathbf{P}\|_F^2 \leq P, \quad (11c)$$

$$\sum_{k'} \bar{C}_{k'} \leq \bar{R}_{c,k}, \forall k, \quad (11d)$$

$$\bar{\mathbf{c}} \succeq \mathbf{0}, \quad (11e)$$

$$\bar{R}_k \geq R_{\text{th},k}, \forall k, \quad (11f)$$

where $\phi_l = [\phi_{l,1}, \dots, \phi_{l,M}]^T \in \mathbb{C}^M$, $\forall l \in \mathcal{L}$, $\bar{\mathbf{c}} = [\bar{C}_1, \dots, \bar{C}_K]^T \in \mathbb{R}^K$, and $R_{\text{th},k} \in \mathbb{R}^+$ is the threshold rate for user k , $\forall k \in \mathcal{K}$, to guarantee the QoS. Problem (11) is a challenging stochastic optimization different from existing optimization problems due to 1) the new constraint (11b), 2) the constraint of common rates (11d), and 3) the QoS constraint (11e), where the BD-RIS matrix and the precoder are always coupled. One reliable solution is to 1) transform problem (11) into a deterministic one based on SAA [37], and to 2) deal with the deterministic problem by well-known WMMSE-rate relationship [38]. Details of the two-step transformation will be illustrated in Section III-B.

B. Two-Step Transformation

1) *SAA:* To facilitate SAA, we give the following definition.

Definition 2: Given an estimate $\hat{\mathbf{Q}}_k$ of user k , a sample of i.i.d. channel realizations is constructed based on the set $\mathcal{Q}_k = \{\mathbf{Q}_k^a = \hat{\mathbf{Q}}_k + \tilde{\mathbf{Q}}_k^a | \hat{\mathbf{Q}}_k, \forall a \in \mathcal{A} = \{1, \dots, A\}\}$, where $\tilde{\mathbf{Q}}_k^a$ follows the CSCG distribution, i.e., $\text{vec}(\tilde{\mathbf{Q}}_k^a) \sim \mathcal{CN}(\mathbf{0}, \delta_k^2 \mathbf{I}_{MN})$, $\forall a \in \mathcal{A}$. Then the SAA of average common and private rates are, respectively, defined as $\hat{R}_{c,k} = \frac{1}{A} \sum_a R_{c,k}^a$ and $\hat{R}_{p,k} = \frac{1}{A} \sum_a R_{p,k}^a$, $\forall k \in \mathcal{K}$. \square

The SAA of average rate in Definition 2 is a tight approximation of the average rate with sufficient large number of samples⁷, i.e., $\lim_{A \rightarrow \infty} \hat{R}_{c,k} = \bar{R}_{c,k}$, $\lim_{A \rightarrow \infty} \hat{R}_{p,k} = \bar{R}_{p,k}$, $\forall k \in \mathcal{K}$, which motivates us to approximate the stochastic problem (11) into a deterministic problem

$$\begin{aligned} & \max_{\{\mathbf{P}, \bar{\mathbf{c}}\}, \{\phi_l, \forall l\}} \sum_k \bar{C}_k + \sum_k \hat{R}_{p,k} \\ & \text{s.t.} \quad (11b), (11c), (11e), \\ & \quad \sum_{k'} \bar{C}_{k'} \leq \hat{R}_{c,k}, \forall k, \\ & \quad \bar{C}_k + \hat{R}_{p,k} \geq R_{\text{th},k}, \forall k. \end{aligned} \quad (12)$$

⁶In the following illustrations, notations with superscript “ $(\cdot)^a$ ” have the same expressions as that without “ $(\cdot)^a$ ”, except that the term \mathbf{Q}_k is replaced by \mathbf{Q}_k^a , $\forall k \in \mathcal{K}$, $\forall l \in \mathcal{L}$, $\forall a \in \mathcal{A}$.

⁷In this paper, we fix $A = 50$ to save on the computational time without sacrificing performance based on simulations using the proposed algorithm and the results in [5].

Problem (12) can be further transformed into a more tractable form based on the following step.

2) *WMMSE-Rate Relationship*: Let $\hat{s}_{c,k}^a = g_{c,k}^a y_k^a$ be the estimate of s_c with equalizer $g_{c,k}^a$ from user k covered by sector l , $\forall k \in \mathcal{K}_l$, $\forall l \in \mathcal{L}$, $\forall a \in \mathcal{A}$. The estimate of $s_{p,k}$ with equalizer $g_{p,k}^a$ after removing the common stream is given by $\hat{s}_{p,k}^a = g_{p,k}^a (y_k^a - \phi_l^T(\mathbf{Q}_k^a)^H \mathbf{p}_{c,s_c})$, $\forall k \in \mathcal{K}_l$, $\forall l \in \mathcal{L}$, $\forall a \in \mathcal{A}$. Define MSE functions regarding the common and private streams for user k covered by sector l at channel sample a as $\epsilon_{c,k}^a = \mathbb{E}\{|\hat{s}_{c,k}^a - s_c|^2\} = |g_{c,k}^a|^2 \tau_{c,k}^a - 2\Re\{g_{c,k}^a \phi_l^T(\mathbf{Q}_k^a)^H \mathbf{p}_c\} + 1$ and $\epsilon_{p,k}^a = \mathbb{E}\{|\hat{s}_{p,k}^a - s_{p,k}|^2\} = |g_{p,k}^a|^2 \tau_{p,k}^a - 2\Re\{g_{p,k}^a \phi_l^T(\mathbf{Q}_k^a)^H \mathbf{p}_{p,k}\} + 1$. Then optimal unconstrained equalizers for the common stream and private streams which achieve minimum MSEs are given by

$$(g_{c,k}^a)^* = \frac{\mathbf{p}_c^H \mathbf{Q}_k^a \phi_l^*}{\tau_{c,k}^a}, \quad (g_{p,k}^a)^* = \frac{\mathbf{p}_{p,k}^H \mathbf{Q}_k^a \phi_l^*}{\tau_{p,k}^a}, \quad (13)$$

and the corresponding minimum MSEs are $(\epsilon_{o,k}^a)^* = \min_{g_{o,k}^a} \epsilon_{o,k}^a = (\tau_{o,k}^a)^{-1} l_{o,k}^a = (1 + \gamma_{o,k}^a)^{-1}$, $\forall o \in \{c, p\}$, $\forall k \in \mathcal{K}$, $\forall a \in \mathcal{A}$. Introducing weights $\lambda_{c,k}^a \in \mathbb{R}^+$ and $\lambda_{p,k}^a \in \mathbb{R}^+$ associating common and private MSEs for user k , we can establish the WMMSE-rate relationship $R_{o,k}^a = \max_{\lambda_{o,k}^a} \log_2 \lambda_{o,k}^a - \lambda_{o,k}^a (\epsilon_{o,k}^a)^* + 1$, $\forall o \in \{c, p\}$, $\forall k \in \mathcal{K}$, $\forall a \in \mathcal{A}$, where the maximum of the right-hand problem is achieved when weights $\lambda_{c,k}$ and $\lambda_{p,k}$ satisfy

$$(\lambda_{c,k}^a)^* = 1 + \gamma_{c,k}^a, \quad (\lambda_{p,k}^a)^* = 1 + \gamma_{p,k}^a. \quad (14)$$

Based on the above derivations, we reformulate problem (12) as a three-block optimization

$$\begin{aligned} & \max_{\substack{\{\lambda_o^a, \mathbf{g}_o^a, \forall o, \forall a\} \\ \{\mathbf{P}, \bar{\mathbf{c}}\}, \{\phi_l, \forall l\}}} \sum_k \left(\bar{C}_k + \frac{1}{A} \sum_a (\log_2 \lambda_{p,k}^a - \lambda_{p,k}^a \epsilon_{p,k}^a) \right) \\ & \text{s.t.} \quad (11b), (11c), (11e), \\ & \quad \frac{1}{A} \sum_a (\log_2 \lambda_{c,k}^a - \lambda_{c,k}^a \epsilon_{c,k}^a + 1) \\ & \quad \geq \sum_{k'} \bar{C}_{k'}, \forall k, \quad (15a) \\ & \quad \bar{C}_k + \frac{1}{A} \sum_a (\log_2 \lambda_{p,k}^a - \lambda_{p,k}^a \epsilon_{p,k}^a + 1) \\ & \quad \geq R_{th,k}, \forall k, \quad (15b) \end{aligned}$$

where $\lambda_o^a = [\lambda_{o,1}^a, \dots, \lambda_{o,K}^a]^T$ and $\mathbf{g}_o^a = [g_{o,1}^a, \dots, g_{o,K}^a]^T$, $\forall o \in \{c, p\}$, $\forall a \in \mathcal{A}$. Although problem (15) is not convex in a joint design manner, the sub-problem regarding one block with fixing other blocks is either convex, such as the design of blocks $\{\lambda_o^a, \mathbf{g}_o^a, \forall o, \forall a\}$ and $\{\mathbf{P}, \bar{\mathbf{c}}\}$, or more tractable, such as the design of block $\{\phi_l, \forall l \in \mathcal{L}\}$, than directly dealing with the original objective. Specifically, the unconstrained solution to block $\{\lambda_o^a, \mathbf{g}_o^a, \forall o, \forall a\}$ with fixed two other blocks have been given in equations (13) and (14), respectively. Solutions to blocks $\{\mathbf{P}, \bar{\mathbf{c}}\}$ and $\{\phi_l, \forall l\}$ will be respectively given in Sections III-C and III-D.

C. Solution to Block $\{\mathbf{P}, \bar{\mathbf{c}}\}$

With fixed equalizers, weights, and BD-RIS matrix, the sub-problem regarding block $\{\mathbf{P}, \bar{\mathbf{c}}\}$ is given by

$$\begin{aligned} & \min_{\{\mathbf{P}, \bar{\mathbf{c}}\}} \frac{1}{A} \sum_k \sum_a \lambda_{p,k}^a \epsilon_{p,k}^a - \sum_k \bar{C}_k \\ & \text{s.t.} \quad (11c), (11e), (15a), (15b). \end{aligned} \quad (16)$$

With the following definitions

$$\psi_{c,k}^a = \lambda_{c,k}^a (g_{c,k}^a)^* \mathbf{Q}_k^a \phi_l^*, \quad \psi_{p,k}^a = \lambda_{p,k}^a (g_{p,k}^a)^* \mathbf{Q}_k^a \phi_l^*, \quad (17a)$$

$$\Psi_{c,k}^a = \frac{\psi_{c,k}^a (\psi_{c,k}^a)^H}{\lambda_{c,k}^a}, \quad \Psi_{p,k}^a = \frac{\psi_{p,k}^a (\psi_{p,k}^a)^H}{\lambda_{p,k}^a}, \quad (17b)$$

$$\xi_{c,k}^a = \log_2 \lambda_{c,k}^a - \lambda_{c,k}^a |g_{c,k}^a|^2 \sigma_k^2 - \lambda_{c,k}^a + 1, \quad (17c)$$

$$\xi_{p,k}^a = \log_2 \lambda_{p,k}^a - \lambda_{p,k}^a |g_{p,k}^a|^2 \sigma_k^2 - \lambda_{p,k}^a + 1, \quad (17d)$$

and their averages over A samples, i.e., $\psi_{c,k} = \frac{1}{A} \sum_a \psi_{c,k}^a$, $\psi_{p,k} = \frac{1}{A} \sum_a \psi_{p,k}^a$, $\Psi_{c,k} = \frac{1}{A} \sum_a \Psi_{c,k}^a$, $\Psi_{p,k} = \frac{1}{A} \sum_a \Psi_{p,k}^a$, $\xi_{c,k} = \frac{1}{A} \sum_a \xi_{c,k}^a$, $\xi_{p,k} = \frac{1}{A} \sum_a \xi_{p,k}^a$, $\forall k \in \mathcal{K}$, problem (16) can be rewritten as

$$\begin{aligned} & \min_{\bar{\mathbf{c}}, \mathbf{P}, \{\psi_{p,k}, \forall k\}} \sum_k \left(\bar{C}_k + \sum_{k'} \mathbf{p}_{p,k'}^H \Psi_{p,k} \mathbf{p}_{p,k'} \right. \\ & \quad \left. - 2\Re\{\psi_{p,k}^H \mathbf{p}_{p,k}\} \right) \end{aligned} \quad (18a)$$

$$\text{s.t.} \quad \|\mathbf{p}_c\|_2^2 + \sum_k \|\mathbf{p}_{p,k}\|_2^2 \leq P, \quad (18b)$$

$$\begin{aligned} & \mathbf{p}_c^H \Psi_{c,k} \mathbf{p}_c + \sum_{k'} \mathbf{p}_{p,k'}^H \Psi_{c,k} \mathbf{p}_{p,k'} \\ & - 2\Re\{\psi_{c,k}^H \mathbf{p}_c\} \leq \sum_{k'} \bar{C}_{k'} + \xi_{c,k}, \forall k, \end{aligned} \quad (18c)$$

$$\bar{\mathbf{c}} \preceq \mathbf{0}, \quad (18d)$$

$$\begin{aligned} & \sum_{k'} \mathbf{p}_{p,k'}^H \Psi_{p,k} \mathbf{p}_{p,k'} - 2\Re\{\psi_{p,k}^H \mathbf{p}_{p,k}\} \\ & + \bar{C}_k \leq \xi_{p,k} - R_{th,k}, \forall k, \end{aligned} \quad (18e)$$

where $\bar{\mathbf{c}} = [\bar{C}_1, \dots, \bar{C}_K]^T = -\bar{\mathbf{c}}$. Problem (18) is a convex second-order cone program (SOCP) and can be solved by interior-point methods with computational complexity bounded by $\mathcal{O}((NK + N + K)^{3.5})$ [40].

D. Solution to Block $\{\phi_l, \forall l\}$

Recall that (15b) in problem (15) is the QoS constraint, which can be guaranteed in the design of block $\{\mathbf{P}, \bar{\mathbf{c}}\}$. To simplify the design of block $\{\phi_l, \forall l\}$, we temporarily ignore constraint (15b). Then, without the QoS constraint (15b), we can remove the term $\bar{\mathbf{c}}$ by putting constraint (15a) into the objective. In this case, the original problem (15) can be simplified as the following form

$$\begin{aligned} & \max_{\substack{\{\lambda_o^a, \mathbf{g}_o^a, \forall o, \forall a\} \\ \mathbf{P}, \{\phi_l, \forall l\}}} \left\{ \min_{\forall k} \frac{1}{A} \sum_a (\log_2 \lambda_{c,k}^a - \lambda_{c,k}^a \epsilon_{c,k}^a + 1) \right\} \\ & + \frac{1}{A} \sum_k \sum_a (\log_2 \lambda_{p,k}^a - \lambda_{p,k}^a \epsilon_{p,k}^a) \end{aligned} \quad (19a)$$

$$\text{s.t.} \quad (11b), (11c), (11e). \quad (19b)$$

With given equalizers, weights, and the transmit precoder, the sub-problem regarding the multi-sector BD-RIS is

$$\begin{aligned} \min_{\{\phi_l, \forall l\}} & \left\{ \max_{\forall k} \frac{1}{A} \sum_a (\lambda_{c,k}^a \epsilon_{c,k}^a - \log_2 \lambda_{c,k}^a - 1) \right\} \\ & + \frac{1}{A} \sum_a \sum_k \lambda_{p,k}^a \epsilon_{p,k}^a \\ \text{s.t.} & \sum_l |\phi_{l,m}|^2 = 1, \forall m. \end{aligned} \quad (20)$$

With the following definitions

$$\mathbf{v}_{c,k}^a = (\mathbf{Q}_k^a)^T \mathbf{p}_c^*, \quad \mathbf{v}_{p,k,k'}^a = (\mathbf{Q}_k^a)^T \mathbf{p}_{p,k'}^*, \quad (21a)$$

$$\begin{aligned} \mathbf{X}_{c,k}^a &= \lambda_{c,k}^a |g_{c,k}^a|^2 \mathbf{v}_{c,k}^a (\mathbf{v}_{c,k}^a)^H \\ & + \lambda_{c,k}^a |g_{c,k}^a|^2 \sum_{k'} \mathbf{v}_{p,k,k'}^a (\mathbf{v}_{p,k,k'}^a)^H, \end{aligned} \quad (21b)$$

$$\mathbf{X}_{p,k}^a = \lambda_{p,k}^a |g_{p,k}^a|^2 \sum_{k'} \mathbf{v}_{p,k,k'}^a (\mathbf{v}_{p,k,k'}^a)^H, \quad (21c)$$

$$\mathbf{x}_{c,k}^a = \lambda_{c,k}^a (g_{c,k}^a)^* \mathbf{v}_{c,k}^a, \quad \mathbf{x}_{p,k}^a = \lambda_{p,k}^a (g_{p,k}^a)^* \mathbf{v}_{p,k,k}^a, \quad (21d)$$

and their averages over A samples, i.e., $\mathbf{X}_{c,k} = \frac{1}{A} \sum_a \mathbf{X}_{c,k}^a$, $\mathbf{X}_{p,k} = \frac{1}{A} \sum_a \mathbf{X}_{p,k}^a$, $\mathbf{x}_{c,k} = \frac{1}{A} \sum_a \mathbf{x}_{c,k}^a$, $\mathbf{x}_{p,k} = \frac{1}{A} \sum_a \mathbf{x}_{p,k}^a$, $\forall k \in \mathcal{K}$, problem (20) can be rearranged as

$$\begin{aligned} \min_{\{\phi_l, \forall l\}} & \left\{ \max_{\forall k} (\phi_l^H \mathbf{X}_{c,k} \phi_l - 2\Re\{\phi_l^H \mathbf{x}_{c,k}\} - \xi_{c,k}) \right\} \\ & + \sum_l (\phi_l^H \bar{\mathbf{X}}_{p,l} \phi_l - 2\Re\{\phi_l^H \bar{\mathbf{x}}_{p,l}\}) \\ \text{s.t.} & \sum_l |\phi_{l,m}|^2 = 1, \forall m, \end{aligned} \quad (22)$$

where $\bar{\mathbf{X}}_{p,l} = \sum_{k' \in \mathcal{K}_l} \mathbf{X}_{p,k'}$ and $\bar{\mathbf{x}}_{p,l} = \sum_{k' \in \mathcal{K}_l} \mathbf{x}_{p,k'}$, $\forall l \in \mathcal{L}$. We observe from problem (22) that the objective regarding each cell of the multi-sector BD-RIS is coupled with each other, while the constraints for different cells are independent of each other. This fact motivates us to successively design each cell while fixing the rest cells. Specifically, the design of one cell with fixed others will be given as follows.

We first formulate the sub-problem regarding cell m , $\forall m \in \mathcal{M}$, with fixed other cells as

$$\begin{aligned} \min_{\bar{\phi}_m} & \left\{ \max_{\forall k} \underbrace{\nu_{c,k,m} |\phi_{l,m}|^2 - 2\Re\{\phi_{l,m}^* \chi_{c,k,m}\}}_{=f_{c,k}(\phi_{l,m})} - \xi_{k,m} \right\} \\ & + \sum_l \underbrace{\nu_{p,l,m} |\phi_{l,m}|^2 - 2\Re\{\phi_{l,m}^* \chi_{p,l,m}\}}_{=f_{p,l}(\phi_{l,m})} \\ \text{s.t.} & \|\bar{\phi}_m\|_2^2 = 1, \end{aligned} \quad (23)$$

where the fresh notations are respectively defined as

$$\nu_{c,k,m} = [\mathbf{X}_{c,k}]_{m,m}, \quad \nu_{p,l,m} = [\bar{\mathbf{X}}_{p,l}]_{m,m}, \quad (24a)$$

$$\chi_{c,k,m} = [\mathbf{x}_{c,k}]_m - \sum_{n \neq m} [\mathbf{X}_{c,k}]_{m,n} \phi_{l,n}, \quad (24b)$$

$$\chi_{p,l,m} = [\bar{\mathbf{x}}_{p,l}]_m - \sum_{n \neq m} [\bar{\mathbf{X}}_{p,l}]_{m,n} \phi_{l,n}, \quad (24c)$$

$$\begin{aligned} \xi_{k,m} &= \xi_{c,k} - \sum_{n \neq m} \sum_{n' \neq m} [\mathbf{X}_{c,k}]_{n,n'} \phi_{l,n}^* \phi_{l,n'} \\ & + 2\Re\left\{ \sum_{n \neq m} \phi_{l,n}^* [\mathbf{x}_{c,k}]_n \right\}, \end{aligned} \quad (24d)$$

and $\bar{\phi}_m = [\phi_{1,m}, \dots, \phi_{L,m}]$, $\forall m \in \mathcal{M}$, $\forall k \in \mathcal{K}_l$, $\forall l \in \mathcal{L}$. Problem (23) is a non-smooth optimization with a non-convex

constraint. To solve this problem, we 1) introduce the well-known LogSumExp (LSE) function to transform the non-smooth part of the objective function into a smooth form, and 2) adopt the manifold theory [41] to transform the constrained problem on the Euclidean space into an unconstrained optimization on the manifold space. Each of the two steps is described as follows.

Step 1: LSE Approximation. To facilitate the LSE approximation, we give the following definition.

Definition 3: Given a set $\mathbb{F} = \{f_{c,k} \in \mathbb{R} | \forall k \in \mathcal{K}\}$, an LSE function is defined as the logarithm of the sum of the exponentials of $f_{c,k}$, $\forall k$, i.e., $\log \sum_k \exp(f_{c,k})$, which is a smooth approximation of the maximum $\max_{\forall k} f_{c,k}$ with a lower bound $\max_{\forall k} f_{c,k}$ and an upper bound $\max_{\forall k} f_{c,k} + \log |\mathcal{K}|$, i.e., $\max_{\forall k} f_{c,k} \leq \log \sum_k \exp(f_{c,k}) \leq \max_{\forall k} f_{c,k} + \log |\mathcal{K}|$.

Proof: Let $f_c^{\max} = \max_{\forall k} f_{c,k}$. Then we have the inequality $\exp(f_c^{\max}) \leq \sum_k \exp(f_{c,k}) \leq |\mathcal{K}| \exp(f_c^{\max})$. Performing the logarithm to the inequality gives the result $\max_{\forall k} f_{c,k} \leq \log \sum_k \exp(f_{c,k}) \leq \max_{\forall k} f_{c,k} + \log |\mathcal{K}|$. \square

With Definition 3, we have the following corollary to obtain an approximation to $\max_{\forall k} f_{c,k}(\phi_{l,m})$ in problem (23).

Corollary 1: $\max_{\forall k} f_{c,k}(\phi_{l,m})$ has the following bounds

$$\begin{aligned} \max_{\forall k} f_{c,k}(\phi_{l,m}) & \stackrel{(a)}{\leq} \underbrace{\varepsilon \log \sum_k \exp\left(\frac{f_{c,k}(\phi_{l,m})}{\varepsilon}\right)}_{=f(\bar{\phi}_m)} \\ & \stackrel{(b)}{\leq} \max_{\forall k} f_{c,k}(\phi_{l,m}) + \varepsilon \log |\mathcal{K}|, \end{aligned} \quad (25)$$

with $\varepsilon > 0$, where (a) achieves its equality when $|\mathcal{K}| = 1$ and (b) achieves its equality when $f_{c,1}(\phi_{l,m}) = \dots = f_{c,K}(\phi_{l,m})$.

Proof: Based on Definition 1, by replacing each $f_{c,k}$ with $\frac{f_{c,k}}{\varepsilon}$, $\forall k \in \mathcal{K}$, we have the inequality $\max_{\forall k} \frac{f_{c,k}}{\varepsilon} \leq \log \sum_k \exp\left(\frac{f_{c,k}}{\varepsilon}\right) \leq \max_{\forall k} \frac{f_{c,k}}{\varepsilon} + \log |\mathcal{K}|$. Multiplying by ε gives the inequality (25). \square

In Corollary 1, $\bar{f}(\bar{\phi})$ refers to the LSE function of $\max_{\forall k} f_{c,k}(\phi_{l,m})$ with ε a scalar to make the bounds tighter. In addition, we can easily deduce that the smaller the ε , the tighter the LSE approximation. To get some insights of how the ε influences the tightness of the bounds, we compare the values of $f_c^{\max} = \max_{\forall k} f_{c,k}(\phi_{l,m})$, its LSE approximation $\bar{f}(\bar{\phi}_m)$, and the upper bound $f_c + \varepsilon \log |\mathcal{K}|$ with different ε in Fig. 4, which verifies that a smaller ε leads to a smaller gap between f_c^{\max} and the corresponding LSE function.

With Corollary 1, problem (23) can be approximated as

$$\bar{\phi}_m^* = \arg \min_{\|\bar{\phi}_m\|_2^2=1} \underbrace{\sum_l f_{p,l}(\phi_{l,m}) + \bar{f}(\bar{\phi}_m)}_{=f(\bar{\phi}_m)}. \quad (26)$$

Now problem (26) has a smooth objective function $f(\bar{\phi}_m)$ and only one constraint, which can be solved by the manifold method [41] as illustrated below.

Step 2: Manifold Construction. We give the following definition to construct the feasible region of problem (26), i.e., $\|\bar{\phi}_m\|_2^2 = 1$, as a manifold [41], [42].

Definition 4: An L -dimensional complex sphere manifold is defined as $\mathbb{M} = \{\bar{\phi}_m \in \mathbb{C}^{1 \times L} : \|\bar{\phi}_m\|_2 = 1\}$, where $\bar{\phi}_m$ is a point on the complex sphere manifold \mathbb{M} . Given a point $\bar{\phi}_m$ on \mathbb{M} , a tangent vector refers to one possible moving direction

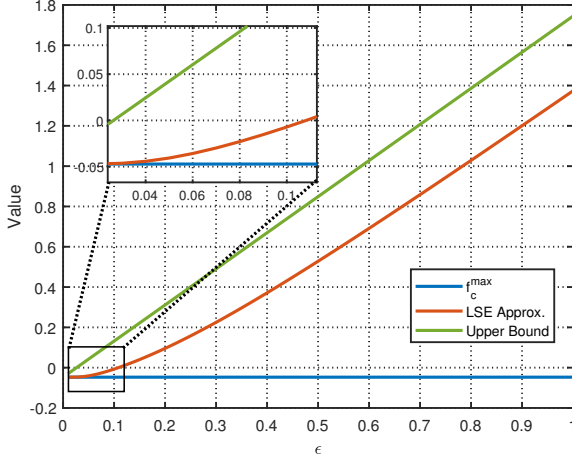


Fig. 4. The impact of ε to the tightness of LSE approximation.

of this point. The set of these tangent vectors regarding point $\bar{\phi}_m$ forms the tangent space, which is mathematically defined as $\mathbb{T}_{\bar{\phi}_m} \mathbb{M} = \{\mathbf{u}_m \in \mathbb{C}^{1 \times L} : \Re\{\text{Tr}(\bar{\phi}_m^H \mathbf{u}_m)\} = 0\}$. Among all these tangent vectors on $\mathbb{T}_{\bar{\phi}_m} \mathbb{M}$, Riemannian gradient is defined as the tangent vector with the steepest-descent direction of the objective function. \square

With Definition 4, we say $\bar{\phi}_m$ is a point on the complex sphere manifold \mathbb{M} , yielding problem (26) an unconstrained optimization on \mathbb{M} , which can be solved by many classic algorithms with necessary projections. The algorithm we use here to solve problem (26) is the Riemannian conjugate gradient (RCG) algorithm [41], which is a modified version of the well-known CG algorithm by changing the search space from the Euclidean space to a manifold as detailed below.

The key of RCG is to determine the descent direction of the objective function. In RCG, the Riemannian gradient, which can be obtained from the Euclidean gradient, is required to calculate the descent direction. Therefore, we first calculate the Euclidean gradient of the objective function

$$\begin{aligned} f(\bar{\phi}_m) &= \sum_l f_{p,l}(\phi_{l,m}) + \bar{f}(\bar{\phi}_m) \\ &= \sum_l f_{p,l}(\phi_{l,m}) + \varepsilon \log \sum_k \exp\left(\frac{f_{c,k}(\phi_{l,m})}{\varepsilon}\right), \end{aligned} \quad (27)$$

which is given by

$$\begin{aligned} \nabla f(\bar{\phi}_m) &= \left[\frac{\partial f(\bar{\phi}_m)}{\partial \phi_{1,m}}, \dots, \frac{\partial f(\bar{\phi}_m)}{\partial \phi_{L,m}} \right] \\ &= \left[\nabla f_{p,1}(\phi_{1,m}) + \frac{\partial \bar{f}(\bar{\phi}_m)}{\partial \phi_{1,m}}, \right. \\ &\quad \left. \dots, \nabla f_{p,L}(\phi_{L,m}) + \frac{\partial \bar{f}(\bar{\phi}_m)}{\partial \phi_{L,m}} \right], \end{aligned} \quad (28a)$$

$$\nabla f_{p,l}(\phi_{l,m}) = 2\nu_{p,l,m}\phi_{l,m} - 2\chi_{p,l,m}, \quad (28b)$$

$$\frac{\partial \bar{f}(\bar{\phi}_m)}{\partial \phi_{l,m}} = \frac{\sum_{k \in \mathcal{K}_l} \exp\left(\frac{f_{c,k}(\phi_{l,m})}{\varepsilon}\right) \nabla f_{c,k}(\phi_{l,m})}{\sum_{k \in \mathcal{K}} \exp\left(\frac{f_{c,k}(\phi_{l,m})}{\varepsilon}\right)}, \quad (28c)$$

$$\nabla f_{c,k}(\phi_{l,m}) = 2\nu_{c,k,m}\phi_{l,m} - 2\chi_{c,k,m}. \quad (28d)$$

Algorithm 1 RCG: Riemannian Conjugate Gradient

Input: $\nu_{c,k,m}, \nu_{p,l,m}, \chi_{c,k,m}, \chi_{p,l,m}, \xi_{c,k,m}, \xi_{p,l,m}, \forall k, \forall l, \bar{\phi}_m, \varepsilon, m, \zeta, v_{\max}$.

Output: $\bar{\phi}_m^*$.

- 1: Initialize $\bar{\phi}_m^0 = \bar{\phi}_m, \beta^0 = -\nabla_{\mathbb{M}} f(\bar{\phi}_m^0), v = 0$.
- 2: **repeat**
- 3: Update $v = v + 1$.
- 4: Calculate stepsize ς^{v-1} by Armijo backtracking [41].
- 5: Set $\bar{\phi}_m^v$ by retraction (31).
- 6: Update Polak-Ribière formula μ^v by (30).
- 7: Update search direction β^v by (29).
- 8: **until** $\|\nabla_{\mathbb{M}} f(\bar{\phi}_m^v)\|_2 \leq \zeta$ or $v \geq v_{\max}$

Algorithm 2 Multi-Sector BD-RIS: Manifold Based Solution

Input: $\mathbf{X}_{c,k}, \bar{\mathbf{X}}_{p,l}, \mathbf{x}_{c,k}, \bar{\mathbf{x}}_{p,l}, \xi_{c,k}, \forall k, \phi_l, \forall l, \varepsilon, \zeta, v_{\max}$.

Output: $\phi_l^*, \forall l$.

- 1: Initialize $\phi_l^0 = \phi_l, \forall l \in \mathcal{L}, v = 0$.
- 2: **repeat**
- 3: Update $v = v + 1$.
- 4: **for** $m \in \mathcal{M}$ **do**
- 5: Update $\chi_{c,k,m}^v, \chi_{p,l,m}^v$, and $\xi_{c,k,m}^v, \forall k \in \mathcal{K}$ by (24).
- 6: Update $\bar{\phi}_m^v$ by the RCG algorithm.
- 7: **end for**
- 8: **until** $\sum_l \|\phi_l^v - \phi_l^{v-1}\|_2^2 \leq \zeta$ or $v \geq v_{\max}$

Then, we obtain the Riemannian gradient, which projects the Euclidean gradient onto the tangent space, i.e., $\nabla_{\mathbb{M}} f(\bar{\phi}_m) = \text{Pr}_{\bar{\phi}_m}(\nabla f(\bar{\phi}_m)) = \nabla f(\bar{\phi}_m) - \Re\{\text{Tr}(\bar{\phi}_m^H \nabla f(\bar{\phi}_m))\bar{\phi}_m\}$. The descent direction at the v -th iteration of RCG is thus

$$\beta^v = -\nabla_{\mathbb{M}} f(\bar{\phi}_m^v) + \mu^v \text{Pr}_{\bar{\phi}_m}(\beta^{v-1}), \quad (29)$$

where μ^v is the Riemannian version of the Polak-Ribière formula [41], which is given by

$$\mu^v = \frac{\langle \nabla_{\mathbb{M}} f(\bar{\phi}_m^v), \nabla_{\mathbb{M}} f(\bar{\phi}_m^v) - \text{Pr}_{\bar{\phi}_m^v}(\nabla_{\mathbb{M}} f(\bar{\phi}_m^{v-1})) \rangle}{\langle \nabla_{\mathbb{M}} f(\bar{\phi}_m^{v-1}), \nabla_{\mathbb{M}} f(\bar{\phi}_m^{v-1}) \rangle}, \quad (30)$$

where $\langle \mathbf{A}, \mathbf{B} \rangle = \Re\{\text{Tr}(\mathbf{A}^H \mathbf{B})\}$. The $v+1$ -th update is given by a retraction function mapping point $\bar{\phi}_m^v$ on the tangent space $\mathbb{T}_{\bar{\phi}_m^v} \mathbb{M}$ to the manifold \mathbb{M} , that is [41]

$$\bar{\phi}_m^{v+1} = \frac{\bar{\phi}_m^v + \varsigma^v \beta^v}{\|\bar{\phi}_m^v + \varsigma^v \beta^v\|_2}, \quad (31)$$

where ς^v is the stepsize obtained by Armijo backtracking [41]. The RCG procedure is summarized in Algorithm 1.

Summary: With solutions to each cell of the multi-sector BD-RIS, the procedure of the successive design is straightforward and summarized in Algorithm 2. The complexity of Algorithm 2 is given by $\mathcal{O}(I_1 L^{1.5})$, where I_1 is the number of iterations of Algorithm 2.

E. Algorithm and Analysis

1) *Framework of the Algorithm:* The above three blocks are updated in an iterative manner until convergence as summarized in Algorithm 3, where $\hat{R}^v = \min_{\forall k} \hat{R}_{c,k}^v + \sum_k \hat{R}_{p,k}^v$.

Algorithm 3 Joint Design: Block Coordinate Descent

Input: $\mathbb{Q}_k, \sigma_k, R_{\text{th},k}, \forall k, \forall l, P, \zeta, v_{\max}$.

Output: $\phi_l^*, \forall l, \mathbf{P}^*$.

- 1: Initialize $\phi_l^0, \mu_l = \mathbf{0}, \forall l, \mathbf{P}^0, v = 0$.
 - 2: **repeat**
 - 3: Update $v = v + 1$.
 - 4: Set weights $\{\lambda_o^a, \forall o, \forall a\}^{v-1}$ by (14).
 - 5: Set equalizers $\{\mathbf{g}_o^a, \forall o, \forall a\}^{v-1}$ by (13).
 - 6: Update BD-RIS matrix $\{\phi_l^v, \forall l\}$ by Algorithm 2.
 - 7: Update precoder \mathbf{P}^v by solving SOCP (18).
 - 8: **until** $\frac{|\hat{R}^v - \hat{R}^{v-1}|}{\hat{R}^{v-1}} \leq \zeta$ or $v \geq v_{\max}$
-

2) *Initialization:* In Algorithm 3, $\phi_l, \forall l \in \mathcal{L}$ are initialized as $\phi_l^0 = \frac{1}{\sqrt{L}}[e^{j\varphi_{l,1}}, \dots, e^{j\varphi_{l,M}}]^T$, where $\varphi_{l,m}$ is randomly selected within the range $[0, 2\pi]$, $\forall l \in \mathcal{L}, \forall m \in \mathcal{M}$. The transmit power allocated to the common streams is initialized as $P_c = 0.7P$, and thus that allocated to the private streams is $P_p = 0.3P$. Accordingly, the common precoder is initialized as $\mathbf{p}_c = \sum_a \sum_k \mathbf{Q}_k^a \phi_l^*$, followed by a normalization $\mathbf{p}_c^0 = \frac{\sqrt{P_c} \mathbf{p}_c}{\|\mathbf{p}_c\|_2}$. The private precoder is initialized as $\mathbf{p}_{p,k} = \sum_a \mathbf{Q}_k^a \phi_l^*$, followed by $\mathbf{p}_{p,k} = \sqrt{\frac{P_p}{K}} \frac{\mathbf{p}_{p,k}}{\|\mathbf{p}_{p,k}\|_2}, \forall k \in \mathcal{K}$.

3) *Complexity Analysis:* The complexity of Algorithm 3 mainly comes from the update of three blocks (steps 4-7). Specifically, updating block $\{\lambda_o^a, \mathbf{g}_o^a, \forall o, \forall a\}$ requires complexity $\mathcal{O}(MNKL A)$. Updating block $\{\mathbf{P}, \bar{\mathbf{c}}\}$ by solving SOCP has complexity $\mathcal{O}((NK + N + K)^{3.5})$, while updating block $\{\phi_l, \forall l\}$ by Algorithm 2 requires $\mathcal{O}(I_1 L^{1.5})$. The overall complexity of Algorithm 3 is thus $\mathcal{O}(I(NK + N + K)^{3.5})$, where I is the number of iterations of Algorithm 3 to achieve the convergence threshold.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed robust beamforming design for multi-sector BD-RIS aided RSMA under different CSI conditions. Particularly, we first illustrate the simulation settings. Then, we show convergence of the proposed algorithm. Finally, we evaluate the achievable ergodic sum-rate performance of multi-sector BD-RIS aided RSMA by analyzing the impact of the radiation pattern of BD-RIS antennas, CSI uncertainty, QoS threshold, and numbers of active/passive antennas.

A. Simulation Setup

The perfect channels $\mathbf{Q}_k = \sqrt{\zeta_k^{-1}} \check{\mathbf{G}}^H \text{diag}(\check{\mathbf{h}}_k), \forall k \in \mathcal{K}$ are generated as a combination of small-scale fading and large-scale fading. Specifically, the small-scale fading of transmitter-RIS and RIS-user channels is modeled as Rician fading, i.e.,

$$\check{\mathbf{G}} = \sqrt{\frac{\kappa}{\kappa + 1}} \check{\mathbf{G}}^{\text{LoS}} + \sqrt{\frac{1}{\kappa + 1}} \check{\mathbf{G}}^{\text{NLoS}}, \quad (32a)$$

$$\check{\mathbf{h}}_k = \sqrt{\frac{\kappa_k}{\kappa_k + 1}} \check{\mathbf{h}}_k^{\text{LoS}} + \sqrt{\frac{1}{\kappa_k + 1}} \check{\mathbf{h}}_k^{\text{NLoS}}, \forall k, \quad (32b)$$

where $\{\kappa, \kappa_k, \forall k\}$ denotes the Rician factor, and channels with superscript “LoS” and “NLoS”, respectively, denote line-of-sight (LoS) and non-LoS (NLoS) channels. ζ_k is the path loss

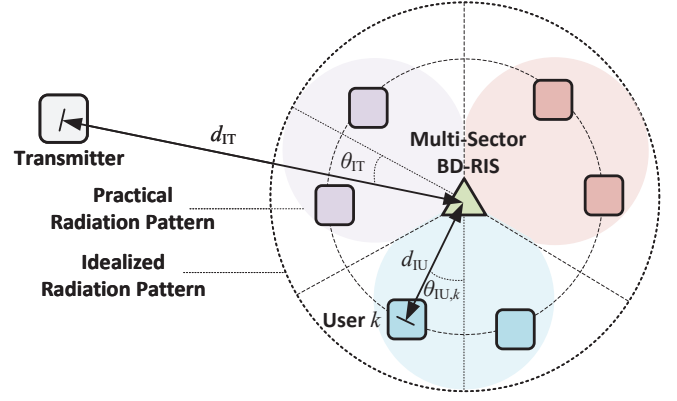


Fig. 5. Top view of the locations among the transmitter, 3-sector BD-RIS, and users, and the illustration of idealized/practical radiation pattern.

component measuring the large-scale fading. The large-scale fading components have different expressions when BD-RIS antennas have different radiation patterns. Here we consider two cases, that is each BD-RIS antenna has either idealized radiation pattern or practical radiation pattern [16], which are illustrated in Fig. 5 from the top view, yielding the following path loss model:

$$\zeta_k = \begin{cases} \varrho(1 - \cos \frac{\pi}{L})^2, & \text{Idealized Pattern,} \\ \frac{\varrho}{(\alpha_L + 1)^2 (\cos \theta_{IT} \cos \theta_{IU,k})^{\alpha_L}}, & \text{Practical Pattern,} \end{cases} \quad (33)$$

where $\varrho = \frac{4^3 \pi^4 d_{IT}^{\epsilon_{IT}} d_{IU,k}^{\epsilon_{IU}}}{\lambda^4 G_T G_U}$ is calculated by transmitter-RIS distance d_{IT} , RIS-user distance $d_{IU,k}$, path loss exponents⁸ ϵ_{IT} and ϵ_{IU} , wavelength of transmit signal λ , transmit antenna gain G_T and receive antenna gain G_R . α_L is a parameter related to the beamwidth of BD-RIS antennas, θ_{IT} is the elevation angle between the transmit antenna and BD-RIS antenna, and $\theta_{IU,k}$ is the elevation angle from the BD-RIS antenna to user k . Sectors of the multi-sector BD-RIS are modeled as $M_x \times M_y = M$ uniform planar arrays and the transmitter is modeled as a uniform linear array with half-wavelength inter-antenna spacing. Locations among the transmitter, multi-sector BD-RIS, and users are shown in Fig. 5.

The channel estimation error for each channel realization $\tilde{\mathbf{Q}}_k$ is generalized such that $\text{vec}(\tilde{\mathbf{Q}}_k) \sim \mathcal{CN}(\mathbf{0}, \delta_k^2 \mathbf{I}_{MN})$, where the channel estimation uncertainty δ_k^2 is set as $\delta_k^2 = \zeta_k^{-1} \delta^2, \forall k \in \mathcal{K}$ with a scalar $\delta \in [0, 1)$. Specifically, the value $\delta = 0$ refers to the perfect CSI case. For each channel estimate $\hat{\mathbf{Q}}_k = \mathbf{Q}_k - \tilde{\mathbf{Q}}_k$, the a -th sample in the set (available at the transmitter) \mathbb{Q}_k is given by $\mathbf{Q}_k^a = \hat{\mathbf{Q}}_k + \tilde{\mathbf{Q}}_k^a$, where $\tilde{\mathbf{Q}}_k^a$ follows the same error distribution, that is $\text{vec}(\tilde{\mathbf{Q}}_k^a) \sim \mathcal{CN}(\mathbf{0}, \delta_k^2 \mathbf{I}_{MN}), \forall k \in \mathcal{K}, \forall a \in \mathcal{A}$. For each channel realization, the precoder and BD-RIS are designed based on sets $\mathbb{Q}_k, \forall k \in \mathcal{K}$ and the maximization of the average sum-rate as summarized in Algorithm 3, while the sum-rate is calculated based on the perfect instantaneous channel \mathbf{Q} . The ergodic sum-rate is plotted as an average of sum-rates for 50 perfect channel realizations. The simulation settings, unless otherwise stated, are summarized in Table I.

⁸The value of the path loss exponent is usually chosen within the range $[2, 4]$ and depends on the small-scale fading conditions.

TABLE I
SIMULATION SETTINGS

Parameters	Value
Rician Factor	$\kappa = \kappa_1 = \dots = \kappa_K = 0$ dB
Transmit/Receive Antenna Gain	$G_t = G_r = 1$
Transmit Signal Frequency	$f = 2.4$ GHz
Transmitter-RIS Distance	$d_{TI} = 100$ m
RIS-User Distance	$d_{IU,1} = \dots = d_{IU,K} = 10$ m
Path Loss Exponent	$\epsilon_{IT} = \epsilon_{IU} = 2$
Transmitter-RIS Angle	$\theta_{IT} \in [0, \frac{\pi}{L}]$
RIS-User Angle	$\theta_{IU,k} \in [0, \frac{\pi}{L}], \forall k \in \mathcal{K}$
Noise Power	$\sigma_1^2 = \dots = \sigma_K^2 = -90$ dBm
Estimation Uncertainty	$\delta \in \{0, 0.15\}$
No. of Samples	$A = 50$
No. of Sectors	$L = 3$
No. of Users per Sector	$K_1 = \dots = K_L = 2$
No. of BD-RIS antennas per Sector	$M = M_x \times M_y = 5 \times 4 = 20$
No. of transmit antennas	$N = 4$
QoS Threshold per User	$R_{th,1} = \dots = R_{th,K} = R_{th}$

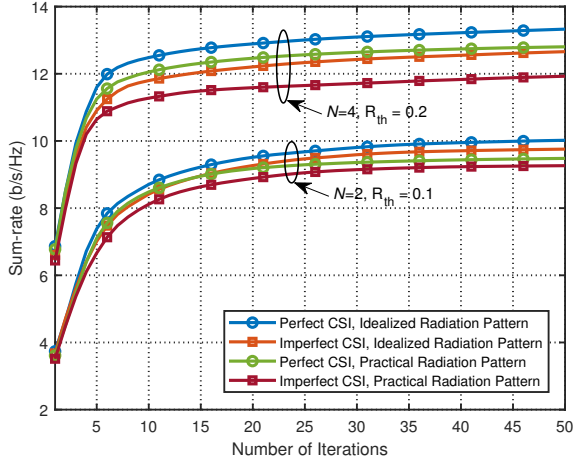


Fig. 6. Sum-rate versus the number of iterations ($P = 35$ dBm).

B. Convergence of the Algorithm

The convergence of Algorithm 3 cannot be theoretically proved since the design of BD-RIS matrix block does not lead to a global optimal solution. Instead, we provide numerical results to show the proposed algorithm can converge within limited iterations. Fig. 6 plots the sum-rate versus the number of iterations of Algorithm 3 for different CSI conditions and parameter settings. We can observe from Fig. 6 that the proposed algorithm always converges within finite iterations regardless of simulation settings, such as channel conditions (perfect and imperfect CSI), the number of transmit antennas, and QoS requirements.

C. Impact of the Radiation Pattern of BD-RIS Antennas

We first evaluate the ergodic sum-rate of multi-sector BD-RIS aided RSMA when BD-RIS antennas have either idealized or practical radiation patterns in Fig. 7. For comparison, we add the performance of multi-sector BD-RIS assisted SDMA⁹,

⁹In this work, we do not involve the NOMA scheme in the simulations since the benefits of RSMA compared to NOMA have been widely investigated in [2], [6], [43] and that in multi-antenna systems RIS aided SDMA is a much tighter benchmark than the NOMA scheme [20].

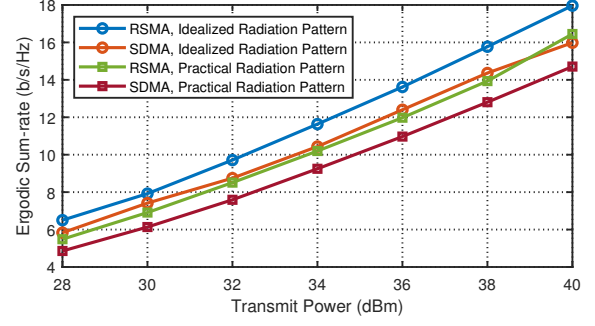


Fig. 7. Ergodic sum-rate versus transmit power with perfect CSI ($R_{th} = 0.2$).

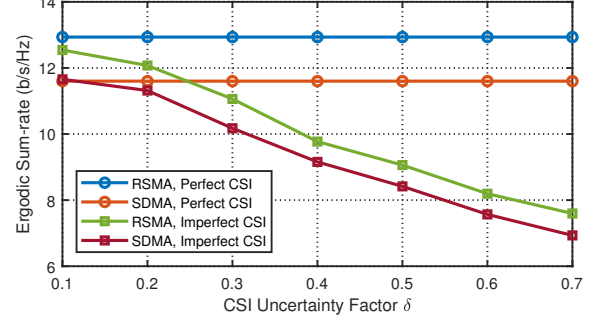


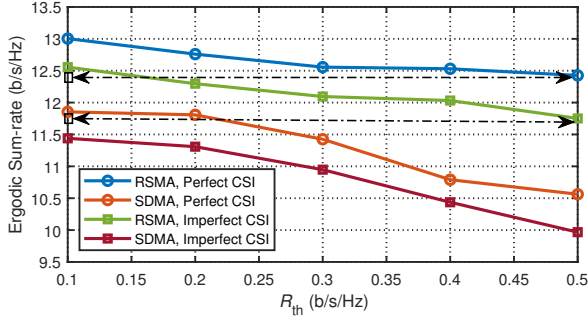
Fig. 8. Ergodic sum-rate versus CSI uncertainty δ with idealized radiation pattern ($P = 35$ dBm, $R_{th} = 0.2$).

which is achieved by switching off the common precoder at the transmitter¹⁰. We have the following observations. 1) With the same multiple access technique, the idealized radiation pattern scheme always outperforms the practical radiation pattern scheme, which is in accordance with the results in [16]. 2) With the same radiation pattern for each BD-RIS antenna, the RSMA scheme always outperforms the SDMA scheme, which demonstrates the benefits of integrating RSMA and multi-sector BD-RIS. 3) The BD-RIS aided RSMA with practical radiation pattern of BD-RIS antennas achieves similar performance to the BD-RIS aided SDMA with idealized radiation pattern for BD-RIS antennas. This phenomenon demonstrates the benefits of integrating RSMA and multi-sector BD-RIS in easing the requirements to antenna realizations.

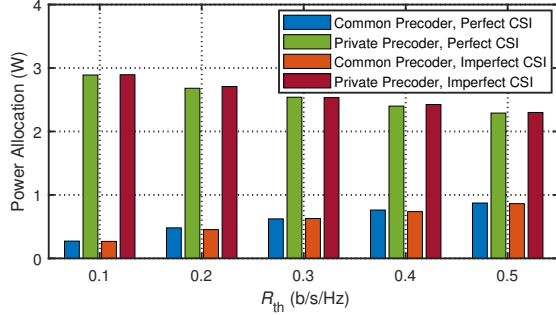
D. Impact of CSI Uncertainty

In Fig. 8, we evaluate the impact of CSI uncertainty by plotting the ergodic sum-rate versus δ when each BD-RIS has an idealized radiation pattern. We observe from Fig. 8 that the ergodic sum-rate for both RSMA and SDMA schemes decreases with increasing channel estimation errors. However, the multi-sector BD-RIS aided RSMA always outperforms multi-sector BD-RIS aided SDMA regardless of CSI conditions, which demonstrates that integrating RSMA and multi-sector BD-RIS always provides performance gain over SDMA schemes even though with limited CSI.

¹⁰In this case, the multi-sector BD-RIS design with given precoder is performed by using the closed-form solution proposed in [16].



(a) Ergodic sum-rate versus R_{th}



(b) Power allocation versus R_{th}

Fig. 9. Impact of QoS threshold with idealized radiation pattern and different CSI conditions ($P = 35$ dBm).

E. Impact of QoS Threshold

Then, in Fig. 9, we show the impact of the QoS requirement for each user to the achievable ergodic sum-rate under different CSI conditions when each multi-sector BD-RIS antenna has an idealized radiation pattern. From Fig. 9 we have the following observations. 1) The multi-sector BD-RIS aided RSMA always outperforms the SDMA scheme with various QoS thresholds and different CSI conditions. 2) The achievable ergodic sum-rate for both schemes decreases with the growth of the QoS requirement, which can be explained as follows. Since we focus on the sum-rate maximization problem, when the number of transmit antennas is smaller than the number of users ($4 = N < K = 6$), the weakest $N - K$ users tend to be turned off. In this case, increasing the QoS requirement for all users violates the aim of maximizing sum-rate. 3) The performance gap between the RSMA scheme and the SDMA scheme becomes larger with increasing QoS threshold for each user, and thus the RSMA scheme is able to achieve both higher sum-rate and better QoS than the SDMA case, e.g., the comparison between RSMA scheme with $R_{th} = 0.5$ and SDMA scheme with $R_{th} = 0.1$. This is because in RSMA scheme, the power allocated to the common precoders increases with the QoS requirement. This phenomenon demonstrates that the common precoders play an increasingly important role to guarantee the fairness while achieving better sum-rate performance.

F. Impact of Numbers of Active and Passive Antennas

In Fig. 10, we plot the ergodic sum-rate versus the total number of antennas of the multi-sector BD-RIS under different

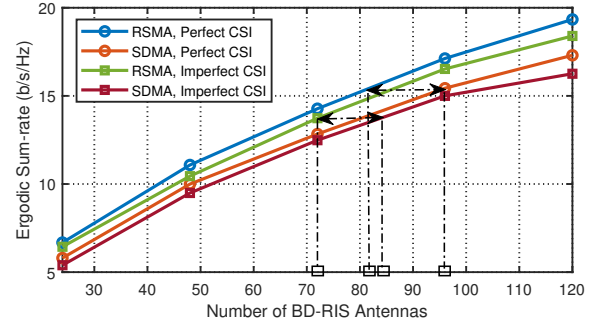


Fig. 10. Ergodic sum-rate versus the total number of antennas of BD-RIS with idealized radiation pattern ($P = 35$ dBm, $R_{th} = 0.2$).

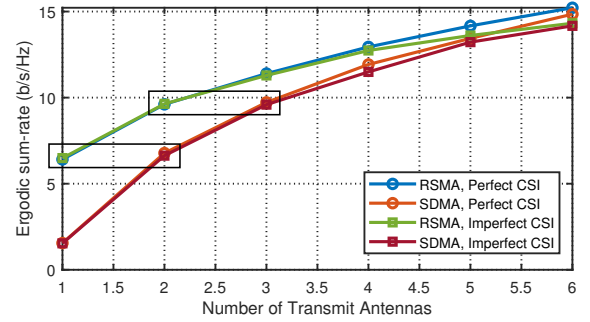


Fig. 11. Ergodic sum-rate versus the number of transmit antennas when BD-RIS antennas have idealized radiation pattern ($P = 35$ dBm, $R_{th} = 0.15$).

CSI conditions when each BD-RIS antenna has an idealized radiation pattern. We observe from Fig. 10 that the ergodic sum-rate of both RSMA and SDMA schemes grows with increasing number of BD-RIS antennas, and that the RSMA scheme always achieves better performance than the SDMA scheme. More importantly, to achieve the same sum-rate, the required number of passive antennas can be reduced compared to the SDMA scheme by adopting RSMA at the transmitter. For example, in perfect CSI case, to achieve an ergodic sum-rate of 15 b/s/Hz, the required number of passive antennas can be reduced by 16%; in imperfect CSI case, to achieve an ergodic sum-rate of 13 b/s/Hz, the required number can be reduced by 14%. This observation highlights the benefit of synergizing BD-RIS and RSMA in saving on passive antennas, which is friendly to the hardware implementation of multi-sector BD-RIS.

Finally, in Fig. 11, we plot the ergodic sum-rate versus the number of transmit antennas under different CSI conditions. Similar conclusions that the RSMA scheme always outperforms the SDMA scheme can be obtained from Fig. 11. More importantly, to achieve the same sum-rate, the required number of active antennas at the transmitter can also be reduced compared to the SDMA scheme when RSMA is adopted at the transmitter. For example, an 1-antenna RSMA-enabled transmitter supports the same sum-rate performance as an 2-antenna SDMA-enabled transmitter. Such reduction of the number of transmit antennas demonstrates the benefit of synergizing BD-RIS and RSMA in reducing power consumption and cost, which leads to an energy efficiency gain.

V. CONCLUSION

In this work, we consider the integration of multi-sector BD-RIS and RSMA in a MU-MISO communication system. Specifically, we assume a multi-sector BD-RIS and an RSMA-enabled transmitter. Given that the channel estimation for RIS and BD-RIS aided scenarios is challenging, we model the channel estimation error when only partial/imperfect CSI is available at the transmitter.

With the multi-sector BD-RIS aided RSMA model and CSI uncertainty model, we propose a robust design to jointly optimize the transmit precoder and multi-sector BD-RIS for a MU-MISO system. Specifically, SAA and WMMSE-rate relationship are adopted to transform the stochastic problem into a more tractable deterministic multi-block problem. Regarding the design of the BD-RIS matrix block, we adopt the RCG algorithm after transforming the corresponding problem into a smooth optimization.

Finally, we present simulation results, which demonstrate the performance enhancement of multi-sector BD-RIS aided RSMA compared to multi-sector BD-RIS aided SDMA with different radiation pattern of BD-RIS antennas and various parameter settings, such as CSI conditions and QoS threshold for users. More importantly, with given sum-rate requirement, the numbers of active and passive antennas can be effectively reduced by synergizing RSMA with multi-sector BD-RIS. Such reduction of the number of antennas is beneficial for practical implementations.

Future research avenues include, but are not limited to the following aspects:

1) *Investigating User Selection for Multi-Sector BD-RIS*: It is interesting to consider the BD-RIS aided communication system with mobile users. In this case, imperfect CSI may lead to inaccurate selection between users and sectors of BD-RIS, such that new metrics and algorithms should be developed to effectively involve the impact of user selection.

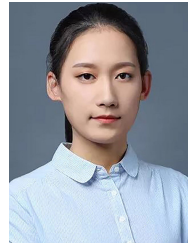
2) *Developing Different Multi-Sector BD-RIS Design Algorithms*: While the performance gain of multi-sector BD-RIS aided RSMA over SDMA schemes has been evaluated by the proposed algorithm in this work, it would also be interesting to develop other algorithms with lower complexity, better performance, and better robustness to the CSI uncertainty.

3) *Modeling the Inter-Sector Interference of Multi-Sector BD-RIS*: In this work, we assume each BD-RIS antenna has a perfect uni-directional radiation pattern such that there are no overlaps between different sectors of BD-RIS. However, in practical implementations, the BD-RIS antennas might have strong sidelobes causing overlaps, such that users will be covered by more than one sector of the BD-RIS. When taking into account this effect, the beamforming design and analysis based on the perfect model do not hold and thus new insights should be investigated for imperfect BD-RIS models.

REFERENCES

- [1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.
- [2] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Commun. Surveys & Tutorials*, 2022.
- [3] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE J. Sel. Areas Commun.*, 2023.
- [4] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.
- [5] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, 2016.
- [6] Y. Mao, B. Clerckx, and V. O. 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, 2019.
- [7] O. Dizdar, Y. Mao, and B. Clerckx, "Rate-splitting multiple access to mitigate the curse of mobility in (massive) MIMO networks," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6765–6780, 2021.
- [8] Z. Feng, B. Clerckx, and Y. Zhao, "Waveform and beamforming design for intelligent reflecting surface aided wireless power transfer: Single-user and multi-user solutions," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5346–5361, 2022.
- [9] Y. Zhao, B. Clerckx, and Z. Feng, "IRS-aided SWIPT: Joint waveform, active and passive beamforming design under nonlinear harvester model," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1345–1359, 2021.
- [10] R. Liu, M. Li, Y. Liu, Q. Wu, and Q. Liu, "Joint transmit waveform and passive beamforming design for RIS-aided DFRC systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 995–1010, 2022.
- [11] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, 2022.
- [12] M. Nerini, S. Shen, and B. Clerckx, "Discrete-value group and fully connected architectures for beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Veh. Technol.*, 2023.
- [13] —, "Closed-form global optimization of beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, 2023.
- [14] H. Zhang, S. Zeng, B. Di, Y. Tan, M. Di Renzo, M. Debbah, Z. Han, H. V. Poor, and L. Song, "Intelligent omni-surfaces for full-dimensional wireless communications: Principles, technology, and implementation," *IEEE Commun. Mag.*, vol. 60, no. 2, pp. 39–45, 2022.
- [15] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2311–2324, 2023.
- [16] —, "Beyond diagonal reconfigurable intelligent surfaces: A multi-sector mode enabling highly directional full-space wireless coverage," *IEEE J. Sel. Areas Commun.*, 2023.
- [17] —, "A dynamic grouping strategy for beyond diagonal reconfigurable intelligent surfaces with hybrid transmitting and reflecting mode," *IEEE Trans. Veh. Technol.*, 2023.
- [18] Q. Li, M. El-Hajjar, I. A. Hemadeh, A. Shojaeifard, A. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6367–6383, 2022.
- [19] A. S. de Sena, P. H. Nardelli, D. B. da Costa, P. Popovski, and C. B. Papadias, "Rate-splitting multiple access and its interplay with intelligent reflecting surfaces," *IEEE Commun. Mag.*, vol. 60, no. 7, pp. 52–57, 2022.
- [20] H. Li, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G—Part III: Interplay with reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2242–2246, 2022.
- [21] M. Wu, Z. Gao, Y. Huang, Z. Xiao, D. W. K. Ng, and Z. Zhang, "Deep learning-based rate-splitting multiple access for reconfigurable intelligent surface-aided tera-hertz massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1431–1451, 2023.
- [22] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, "Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1–6.
- [23] K. Weinberger, A. A. Ahmad, A. Sezgin, and A. Zappone, "Synergistic benefits in IRS-and RS-enabled C-RAN with energy-efficient clustering," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8459–8475, 2022.

- [24] A. Bansal, K. Singh, B. Clerckx, C.-P. Li, and M.-S. Alouini, "Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9217–9229, 2021.
- [25] D. Shambharkar, S. Dhok, A. Singh, and P. K. Sharma, "Rate-splitting multiple access for RIS-aided cell-edge users with discrete phase-shifts," *IEEE Commun. Lett.*, vol. 26, no. 11, pp. 2581–2585, 2022.
- [26] A. Bansal, N. Agrawal, and K. Singh, "Rate-splitting multiple access for UAV-based RIS-enabled interference-limited vehicular communication system," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 936–948, 2022.
- [27] K. Weinberger, A. A. Ahmad, and A. Sezgin, "On synergistic benefits of rate splitting in IRS-assisted cloud radio access networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2021, pp. 1–6.
- [28] K. Zhao, Y. Mao, Z. Yang, L. Lian, and B. Clerckx, "Reconfigurable intelligent surfaces empowered cooperative rate splitting with user relaying," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2022, pp. 1–6.
- [29] A. Bansal, K. Singh, and C.-P. Li, "Analysis of hierarchical rate splitting for intelligent reflecting surfaces-aided downlink multiuser MISO communications," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 785–798, 2021.
- [30] A. Jolly, S. Biswas, and K. Singh, "An analysis on rate-splitting multiple access for IRS aided 6G communication," *arXiv preprint arXiv:2106.04418*, 2021.
- [31] K. Weinberger and A. Sezgin, "Sacrificing CSI for a greater good: Ris-enabled opportunistic rate splitting," in *Proc. 12th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, 2022, pp. 281–285.
- [32] S. Dhok and P. K. Sharma, "Rate-splitting multiple access with STAR RIS over spatially-correlated channels," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6410–6424, 2022.
- [33] M. Soleymani, I. Santamaria, and E. Jorswieck, "Rate splitting in MIMO RIS-assisted systems with hardware impairments and improper signaling," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4580–4597, 2022.
- [34] H. R. Hashempour, H. Bastami, M. Moradikia, S. A. Zekavat, H. Behroozi, and A. L. Swindlehurst, "Secure SWIPT in STAR-RIS aided downlink MISO rate-splitting multiple access networks," *arXiv preprint arXiv:2211.09081*, 2022.
- [35] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Improved spectral efficiency in STAR-RIS aided uplink communication using rate splitting multiple access," *IEEE Trans. Wireless Commun.*, 2023.
- [36] T. Fang, Y. Mao, S. Shen, Z. Zhu, and B. Clerckx, "Fully connected reconfigurable intelligent surface aided rate-splitting multiple access for multi-user multi-antenna transmission," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2022, pp. 675–680.
- [37] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.
- [38] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.
- [39] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces—A general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, 2022.
- [40] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [41] N. Boumal, "Optimization and estimation on manifolds." 2014.
- [42] P. Petersen, *Riemannian geometry*. Springer, 2006, vol. 171.
- [43] A. Mishra, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for downlink multiuser MIMO: Precoder optimization and PHY-layer design," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 874–890, 2021.



research interests include signal processing, wireless communications, and reconfigurable intelligent surface.



and a Post-Doctoral Research Associate with the Communications and Signal Processing Group, Imperial College London, London, U.K., from 2018 to 2020. He was a Research Assistant Professor with the Department of Electronic and Computer Engineering, HKUST, from 2020 to 2023. He is currently a Lecturer (Assistant Professor) with the Department of Electrical Engineering and Electronics, University of Liverpool, U.K. His current research interests include RF energy harvesting, wireless power transfer, reconfigurable intelligent surface, Internet-of-Things, MIMO systems, and antenna design and optimization.



Bruno Clerckx is a (Full) Professor, the Head of the Wireless Communications and Signal Processing Lab, and the Deputy Head of the Communications and Signal Processing Group, within the Electrical and Electronic Engineering Department, Imperial College London, London, U.K. He is also the Chief Technology Officer (CTO) of Silicon Austria Labs (SAL) where he is responsible for all research areas of Austria's top research center for electronic based systems. He received the MSc and Ph.D. degrees in Electrical Engineering from Universit Catholique de Louvain, Belgium, and the Doctor of Science (DSc) degree from Imperial College London, U.K. Prior to joining Imperial College in 2011, he was with Samsung Electronics, Suwon, South Korea, where he actively contributed to 4G (3GPP LTE/LTE-A and IEEE 802.16m). He has authored two books on MIMO Wireless Communications and MIMO Wireless Networks, 300 peer-reviewed international research papers, and 150 standards contributions, and is the inventor of 80 issued or pending patents among which several have been adopted in the specifications of 4G standards and are used by billions of devices worldwide. His research spans the general area of wireless communications and signal processing for wireless networks. He received the prestigious Blondel Medal 2021 from France for exceptional work contributing to the progress of Science and Electrical and Electronic Industries, the 2021 Adolphe Wetrems Prize in mathematical and physical sciences from Royal Academy of Belgium, multiple awards from Samsung, IEEE best student paper award, and the EURASIP (European Association for Signal Processing) best paper award 2022. He is a Fellow of the IEEE and the IET, and an IEEE Communications Society Distinguished Lecturer.

Hongyu Li (S'18) received the B.S. degree in electronics and information engineering and the M.S. degree in information and communication engineering from Dalian University of Technology, Dalian, China, in 2018 and 2021, respectively. She is now studying toward the Ph.D. degree with the Department of Electrical and Electronic Engineering, Imperial College London, U.K. She was a visiting Ph.D. student with the Laboratory of Signals and Systems (L2S) of Paris-Saclay University, CNRS, CentraleSupélec, Paris, France, in 2023. Her current