Optimization of Rate-Splitting Multiple Access in Beyond Diagonal RIS-assisted URLLC Systems

Mohammad Soleymani, *Member, IEEE*, Ignacio Santamaria, *Senior Member, IEEE*, Eduard Jorswieck, *Fellow, IEEE*, and Bruno Clerckx, *Fellow, IEEE*

Abstract—This paper proposes a general optimization framework for rate splitting multiple access (RSMA) in beyond diagonal (BD) reconfigurable intelligent surface (RIS) assisted ultra-reliable low-latency communications (URLLC) systems. This framework can provide a suboptimal solution for a large family of optimization problems in which the objective and/or constraints are linear functions of the rates and/or energy efficiency (EE) of users. Using this framework, we show that RSMA and RIS can be mutually beneficial tools when the system is overloaded, i.e., when the number of users per cell is higher than the number of base station (BS) antennas. Additionally, we show that the benefits of RSMA increase when the packets are shorter and/or the reliability constraint is more stringent. Furthermore, we show that the RSMA benefits increase with the number of users per cell and decrease with the number of BS antennas. Finally, we show that RIS (either diagonal or BD) can highly improve the system performance, and BD-RIS outperforms regular RIS.

Index Terms—Beyond diagonal reconfigurable intelligent surface, energy efficiency, MISO broadcast channels, rate splitting multiple access, spectral efficiency, ultra-reliable low-latency communications.

I. INTRODUCTION

The sixth generation (6G) of communication systems should be around 100 times more reliable than 5G networks and at the same time, provide around 10 times lower latency comparing to 5G networks [1], [2]. Moreover, it is expected that 6G networks become around 100 times more energy efficient and around 10 times more spectral efficient than 5G systems [1], [2]. To fulfill such ambitious goals, 6G should employ promising technologies such as reconfigurable intelligent surface (RIS) and rate splitting multiple access (RSMA) [3]–[6]. In this paper, we propose a general optimization framework to improve spectral efficiency (SE) and energy efficiency (EE)

of ultra-reliable and low-latency communication (URLLC) systems by employing RIS and RSMA. Moreover, we investigate whether and how RIS and RSMA can be beneficial in URLLC systems, and show how their possible benefits can vary in different operational points depending on the latency and reliability constraints.

A. Related works

To support low latency, we cannot operate in very large packet length regimes and have to employ shorter packet lengths in which the Shannon rates are not accurate anymore. In [7], it was shown that the rate for single-input single-output (SISO) point-to-point channels transmitting Gaussian signals can be approximated as

$$r = C - Q^{-1}(\epsilon)\sqrt{\frac{V}{n_t}}, \tag{1}$$

where C is the Shannon rate, n_t is the packet length in bits, Q^{-1} is the inverse of the Gaussian Q-function, ϵ is the decoding error probability of the message, and V is the channel dispersion. The finite-block-length (FBL) rate approximation in (1) is known as the normal approximation (NA). The accuracy of the NA at different operational points has been vastly discussed in [7]–[9]. As can be easily verified from (1), the FBL rates are smaller than Shannon rates. In addition, as expected, the shorter the packet length is and/or the more stringent the reliability constraint is, the lower rate can be achieved. Indeed, we have to transmit at a lower rate to ensure a more reliable communication with low latency. Resource allocation and transmission schemes for FBL regimes based on the NA have been studied in [10]–[12]. In [10], the authors proposed power optimization and beamforming schemes for a broadcast channel (BC). In [11], the authors proposed schemes to maximize the weighted sum rate of a multiple-input single-output (MISO) orthogonal frequency division multiple access (OFDMA) URLLC system. Moreover, the paper [12] proposed schemes to maximize the minimum rate and minimum EE of users in a cell-free massive multiple-input multiple-output (MIMO) URLLC system.

One of the targets of 6G is to significantly improve the SE and EE, which can be even more important in URLLC systems. A promising technology that will enable us to meet this target is RIS, which has been shown to enhance the performance of various interference-free and interference-limited systems with Shannon rates and/or the NA in (1) [3], [4], [13]–[33]. In [13], [14], it was shown that RIS can improve

the SE and/or EE of MISO BCs. In [15], a non-orthogonal multiple access (NOMA) based scheme was proposed for a multi-cell MISO RIS-assisted BC, and it was shown that RIS can significantly increase the minimum rate and EE of the network. The superiority of RIS in multi-cell MIMO BCs was studied in [16], [17] when treating interference as noise (TIN) is employed. In [18]–[20], it was shown that RIS can highly enhance the SE and EE of a multi-cell MIMO BC when the transceivers suffer from I/Q imbalance. In [21], it was shown that RIS can decrease the interference leakage of K-user MIMO interference channels.

All the aforementioned papers [13]–[21] on RIS considered Shannon rates. The use of RIS in FBL regimes has been investigated in [22]–[33]. The authors in [22] showed that RIS can improve the performance of URLLC unmanned aerial vehicle (UAV). In [23]–[25], it was shown that RIS can improve the performance of SISO/MISO BCs with TIN in URLLC systems. The authors in [26] showed that RIS and NOMA can decrease the average decoding rate and at the same time increase the throughput of a two-user SISO BC. In [27], it was shown that RIS can enhance the performance of a MISO point-to-point URLLC system with a factory automation scenario. Finally, [33] showed that RIS can increase the sum rate of a multi-cell MISO orthogonal frequency division multiplexing (OFDM) BC. For a more detailed review on RIS, we refer the reader to [3], [4].

There are different RIS technologies and architectures. In most of the early works, the matrix modeling the RIS coefficients is assumed to be diagonal. However, the system performance can be improved by relaxing the diagonality assumption and using beyond diagonal (BD) RIS architectures [34]. In a BD-RIS, each RIS element can be connected to other elements through a circuit [35], [36]. There are three different possibilities for BD-RIS based on the connectivity of RIS elements: single-connected, group-connected and fully-connected architectures [36]. Indeed, a regular passive RIS can be considered as a special case of BD-RIS, which can be referred to as single-connected architecture. In a fully-connected BD-RIS, all the BD-RIS elements are connected to each other, while in a group-connected BD-RIS, each element is connected to only a group of elements, which reduces the implementation complexities. The superiority of BD-RIS over a regular RIS has been studied in [36]–[40]. For instance, in [37], the authors proposed a scheme that results in a non-diagonal phase shift RIS matrix, showing that BD-RIS can improve the performance of single- and multi-user MISO BCs. Moreover, it was shown in [38] that BD-RIS can outperform RIS in dual-function radar-communication systems. Additionally, [36] showed that BD-RIS (with the group- and fully-connected architectures) can enhance sum rate of a MISO BC. It is shown in [39] that BD-RIS can increase the signal-to-noise ratio (SNR) of SISO and MISO systems.

Another promising technology to enhance SE and EE is RSMA, which includes as particular cases many other technologies such as TIN, NOMA, multicasting, broadcasting, and spatial division multiple access (SDMA) [5], [6]. In rate splitting (RS) schemes, there are two types of messages:

common and private. Each private message is intended for only a specific user, while common messages are decoded by all or by a group of users depending the employed RSMA scheme [5]. Indeed, there are different RSMA schemes based on the number/format of common messages. The simplest RSMA scheme is the 1-layer RS, which is very practical and efficient [41]–[46]. In 1-layer RS, there is only one common message, which is decoded by all the users, while treating the private messages as noise. Moreover, each user decodes its own private message after decoding and canceling the common message from the received signal. Note that when interference is very weak, the optimal strategy for sum rate maximization or in terms of the generalized degrees of freedom is to treat the interference as noise [47], [48]. Furthermore, if interference is strong, the interfering signal should be decoded and canceled from the received signal, which is widely known as successive interference cancellation (SIC) [49]. RSMA bridges TIN and SIC, which makes RSMA very flexible and powerful. Based on the interference level at users, RSMA can switch between TIN and SIC, without the need to order users, thus reducing the design complexities. For a more detailed overview on RSMA, we refer the reader to [5], [6].

The performance of RSMA with FBL has been studied in [50]–[58] for systems without RIS. In [50], the authors proposed a flexible RSMA scheme for a MISO BC with FBL and showed that their proposed scheme outperforms SDMA and NOMA. In [51], [52], it was shown that RSMA can achieve the same minimum or sum rate as NOMA and SDMA with a smaller packet length, meaning that RSMA may reduce latency for a given target rate. In [53], the authors proposed resource allocation schemes for RSMA in a MISO URLLC BC, showing that RSMA provides a higher effective throughput than NOMA. Additionally, they showed that RSMA can reduce latency and enhance reliability. In [54], a robust beamforming design was proposed for a MISO BC with 1-layer RS, and it was shown that RSMA can significantly increase the minimum rate of the users in the considered scenario. In [58], it was shown that a 1-layer RS scheme can enhance the performance of a SISO UAV-assisted BC with imperfect channel state information (CSI) and imperfect SIC by considering both infinite and finite block length regimes.

Finally, we summarize some of the most related works in Table I. As indicated, RIS and RSMA are powerful tools to improve SE and EE of various systems. However, there are only a limited number of papers that considered the performance of RSMA in RIS-assisted URLLC systems (i.e., [59]–[61]). In [59], 1-layer RS schemes were proposed for a SISO system aided by RIS and UAV in which RIS is mounted at a UAV relay that operates in a full-duplex mode with the decode and forward strategy. Then it was shown that RSMA can outperform NOMA and OMA schemes in both infinite and finite block length regimes. The authors in [60] proposed 1-layer RS schemes for both infinite and finite block length regimes in a SISO STAR-RIS-assisted BC with spatially-correlated channels. In [61], it was shown that RSMA can increase the global EE of a MISO RIS-assisted BC with FBL regime.

To sum up, to the best of our knowledge, the paper

TABLE I: A brief comparison of the most related works.

	RIS	FBL	MU communications	Multiple-antenna Systems	EE metrics	RSMA	SE metrics	BD-RIS
This paper	✓	✓	✓	✓	✓	✓	✓	✓
[10], [11]		✓	✓	✓			✓	
[12]		✓	✓	✓	✓		✓	
[16]	✓		✓	✓			✓	
[14], [15], [17]	✓		✓	✓	✓		✓	
[18]	✓		✓	✓	✓	✓	✓	
[22], [33]	✓	✓	✓				✓	
[23], [26], [29]	✓	✓	✓				✓	
[27], [28]	✓	✓		✓			✓	
[24]	✓	✓	✓				✓	
[50]–[55]		✓	✓	✓		✓	✓	
[59]	✓	✓	✓			✓	✓	
[56]–[58]		✓	✓			✓	✓	
[61]	✓	✓	✓	✓	✓	✓		
[60]	✓	✓	✓			✓	✓	

[61] is the only paper on RSMA in multiple-antenna RIS-assisted URLLC systems. Moreover, there is only one paper considering EE metrics in URLLC systems with RSMA, i.e., [61], in which, it was shown that RSMA can increase the global EE of a single-cell MISO RIS-assisted BC. Thus, the performance of RSMA in multi-antenna RIS-assisted systems should be further studied. Additionally, more energy-efficient RSMA schemes should be developed for URLLC systems.

B. Motivation

RSMA is a very effective and flexible tool to manage interference that encompasses a large variety of multiple-access technologies such as NOMA, TIN, SDMA, broadcasting and multi-casting [5], [43]. Moreover, RIS enables optimizing environment by modulating channels, which can be employed to neutralize interference and/or to improve coverage. Hence, one might expect that the RSMA benefits are reduced when RIS is employed since RIS can manage interference in some scenarios especially in multiple-antenna systems. However, in [18], [62]–[66], it was shown that RSMA and RIS can be mutually beneficial tools in overloaded systems with Shannon rate. Unfortunately, in FBL regimes, the Shannon rates are not accurate, which makes optimizing parameters more complicated and can bring some new challenges/tradeoffs. Moreover, the solutions for Shannon rates cannot be used in FBL regime, which further motivates developing specific RSMA techniques for RIS-assisted URLLC systems.

In this work, we study the role of RIS and RSMA in multiple-antenna URLLC systems, and particularly, provide an answer to the following question: *What is the impact of RIS on the performance of RSMA in multiple-antenna URLLC systems?* We show that the impact of RIS on RSMA is a dichotomy depending on the network load. More specifically, RIS can enhance the benefits of RSMA in overloaded systems, but the RSMA benefits decrease (or even become negligible) by optimizing RIS components in underloaded systems. Moreover, in this work, we investigate the impact of packet length, which is related to the latency constraint, and the reliability constraint on the RSMA performance. We show that the benefits of RSMA increase when packets are shorter or when the maximum tolerable decoding probability is smaller.

C. Contribution

The main goal of this work is to investigate the overall system performance as well as the specific role of RIS and RSMA in URLLC systems. We not only show that RIS and/or RSMA can significantly improve the system performance, but also clarify the role of RIS and RSMA in such improvements. To this end, we propose an optimization framework for RSMA to enhance SE and EE of MISO (BD-)RIS-assisted URLLC systems, which can be applied to every interference-limited system with 1-layer RS. As shown in Table I and discussed in Section I-A, there are a limited number of works on RSMA in RIS-assisted URLLC systems, and to the best of our knowledge, the SE of RSMA in multiple-antenna RIS-assisted systems with FBL regimes has not been studied yet. Thus, it is required to develop a general optimization framework for RSMA in multiple-antenna RIS-assisted URLLC systems that can provide a solution for a large family of optimization problems, including various SE and EE metrics such as the minimum weighted rate, weighted sum rate, minimum weighted EE, and global EE. In this paper, we address this issue by proposing a framework to solve every optimization problem in which the objective and/or constraints are linear functions of the rates and/or EE of users and/or the received powers. Note that in [25], we proposed resource allocation schemes for STAR-RIS-assisted URLLC systems with TIN. However, in this work, we extend the results in [25] to RSMA and consider another technology for RIS (i.e., BD-RIS). Indeed, not only the transmission strategy at the BSs is different from [25], but also the RIS technologies vary in these two papers.

To clarify the role of RIS/RSMA, we define two operational regimes based on the number of users and the number of BS antennas. We call the system underloaded if the number of BS antennas is higher than the number of users per cell. Otherwise, we call the system overloaded. We show that RIS and RSMA are mutually beneficial tools in overloaded systems; however, the benefits of RSMA decrease by employing RIS in underloaded systems. The reason is that the interference level is lower in underloaded systems than in overloaded systems. Hence, interference in underloaded systems can be managed in a simpler way by SDMA and optimizing channels through RIS. However, in overloaded systems, we require

more powerful interference-management techniques such as RSMA to mitigate interference. We also show that RIS with TIN may even perform worse than RSMA without employing RIS in an overloaded system, which shows the importance of RSMA. To sum up, the role of RSMA is to manage interference especially in overloaded systems. The role of RIS is mainly to improve the coverage in both underloaded and overloaded systems, as well as to partly manage interference in underloaded systems.

We, moreover, aim at investigating the impact of the reliability and latency constraints on the performance of RSMA. We show that the benefits of RSMA in overloaded systems increase when the reliability constraint is more stringent and/or when the packet lengths are shorter. This shows that RSMA can enhance reliability and ensure a lower latency. We also show that the benefits of RSMA increase with the number of users per cell. The reason is that the interference level increases with the number of users per resources, which makes RSMA as an interference-management technique more beneficial. Additionally, we show that the benefits of RSMA decrease with the number of antennas at BSs. When the number of BS antennas increases, indeed the number of resources increases, which makes it easier to manage interference by SDMA. We also show that RIS can significantly improve the EE and SE of the system even with a relatively low number of RIS elements per users.

In this paper, we also develop optimization techniques for BD-RIS with a group-connected architecture of group size two. To the best of our knowledge, this is the first wok that studies BD-RIS in URLLC systems. We consider two feasibility sets for optimizing the BD/diagonal RIS elements and show that RIS (either regular or beyond diagonal) can significantly improve the system performance. Moreover, we show that BD-RIS with group-connected architecture of group size two can outperform a regular RIS.

D. Paper outline

This paper is organized as follows. Section II presents the system model and formulates the problem. Section III proposes schemes to optimize the beamforming vectors. Section IV provides solutions for optimizing the BD-RIS elements. Section V presents some numerical results. Finally, Section VI concludes the paper.

II. SYSTEM MODEL

We propose an optimization framework for 1-layer RS, which can be applied to any interference-limited (BD-)RIS-assisted URLLC system and is able to provide a solution for a large family of optimization problems in which the objective/utility function and/or constraints are linear functions of the rates and/or EE of users and/or the received power. As an illustrative example, we consider a multicell broadcast channel (BC) with L multiple-antenna base stations (BSs) with N_{BS} transmit antennas, as shown in Fig. 1. We assume that each BS serves K single-antenna users, and there are $M \geq L$ BD-RIS and/or regular RISs with N_{RIS} components each. Moreover, perfect, global and instantaneous CSI is assumed
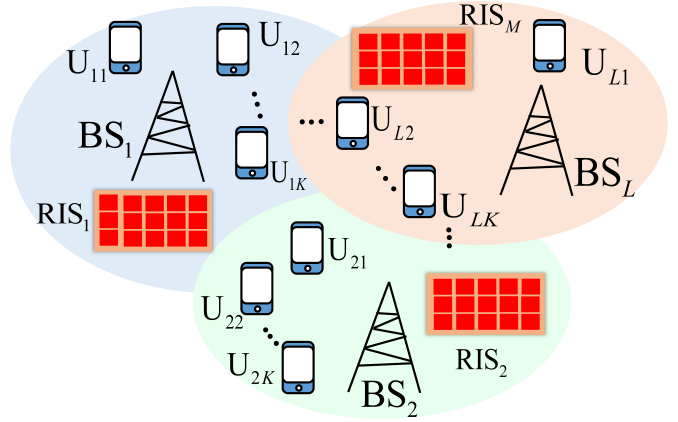

Fig. 1: A multicell broadcast channel with RIS.

to be available at all transceivers similar to many other works on RIS [13]–[18]. It should also be emphasized that this is a common assumption when proposing resource allocation schemes for URLLC systems [10], [11], [25], [33], [50]. In this case, it is assumed that the coherence time of the channels is large, and the system remains static for a long time. Thus, after acquiring CSI, the obtained beamforming vectors, RSMA parameters and RIS elements solutions can be used over many time slots. Therefore, the overhead associated with channel estimation (as well as that associated with obtaining the solution) is typically not considered to increase latency. However, such schemes may not be applicable for fast fading systems. Additionally, it should be noted that considering perfect CSI can provide an upper bound for the benefits of all the schemes. If such benefits are minor even with perfect CSI, then it may happen that RSMA and RIS cannot be effective in URLLC systems with imperfect CSI.

A. RIS model

In this paper, we consider single-sector BD-RISs in the reflective mode with group-connected architecture of group size two. In this case, the channel between BS i and user k associated to BS l, denoted by u_{lk}, is

$$\mathbf{h}_{lk,i}(\{\boldsymbol{\Theta}\}) = \underbrace{\sum_{m=1}^{M} \mathbf{f}_{lk,m} \boldsymbol{\Theta}_m \mathbf{G}_{mi}}_{\text{Links through RIS}} + \underbrace{\mathbf{d}_{lk,i}}_{\text{Direct link}} \in \mathbb{C}^{1 \times N_{BS}}, \quad (2)$$

where $\mathbf{d}_{lk,i} \in \mathbb{C}^{1 \times N_{BS}}$ is the direct link between BS i and u_{lk}, $\mathbf{G}_{mi} \in \mathbb{C}^{N_{RIS} \times N_{BS}}$ is the channel matrix between BS i and BD-RIS m, $\mathbf{f}_{lk,m} \mathbb{C}^{1 \times N_{RIS}}$ is the channel vector between BD-RIS m and u_{lk}, $\{\boldsymbol{\Theta}\} = \{\boldsymbol{\Theta}_m\}$ is the set containing all the BD-RIS components. For a regular RIS, $\boldsymbol{\Theta}_m$ is a diagonal matrix, given by $\boldsymbol{\Theta}_m = \text{diag}\left(\theta_{m_1}, \theta_{m_2}, \cdots, \theta_{m_{N_{RIS}}}\right)$, where θ_{m_i} is the coefficient corresponding to the i-th element of RIS m. However, in a BD-RIS, the diagonality assumption is relaxed, and $\boldsymbol{\Theta}_m$ is a symmetric non-diagonal matrix. For BD-RIS with group-connected architecture of group size two, $\boldsymbol{\Theta}_m$ is a block-diagonal matrix as

$$\boldsymbol{\Theta}_m = \text{diag}(\boldsymbol{\Theta}_{m_1}, \boldsymbol{\Theta}_{m_2}, \cdots, \boldsymbol{\Theta}_{m_G}), \quad \boldsymbol{\Theta}_{m_g} = \boldsymbol{\Theta}_{m_g}^T, \quad (3)$$

where $\boldsymbol{\Theta}_{m_g}$ for all m, g is a 2-by-2 symmetric matrix, and $G = \frac{N_{RIS}}{2}$. Note that without loss of generality, we assume that N_{RIS} is an even number since it is assumed that each RIS element is connect to another RIS element. If N_{RIS} is an odd number, there is one RIS element, which is single connected, and our analysis can easily include this case as well.

There can be two different constraints for the symmetric matrices $\boldsymbol{\Theta}_{m_g}$s. First, we have the convex constraint $\boldsymbol{\Theta}_{m_g}\boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}$ for all m, g, which results in the following feasibility set:

$$\mathcal{T}_U = \{\boldsymbol{\Theta}_{m_g} = \boldsymbol{\Theta}_{m_g}^T, \boldsymbol{\Theta}_{m_g}\boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}, \forall m, g\} \qquad (4)$$

Second, we have $\boldsymbol{\Theta}_{m_g}\boldsymbol{\Theta}_{m_g}^H = \mathbf{I}$, which yields

$$\mathcal{T}_I = \{\boldsymbol{\Theta}_{m_g} = \boldsymbol{\Theta}_{m_g}^T, \boldsymbol{\Theta}_{m_g}\boldsymbol{\Theta}_{m_g}^H = \mathbf{I}, \forall m, g\} \qquad (5)$$

Note that $\mathcal{T}_I \subset \mathcal{T}_U$. Indeed, $\mathcal{T}_U$ includes $\mathcal{T}_I$ as a special case and should not perform worse than $\mathcal{T}_I$. Note that if u_{lk} (or BS i) is not in the reflection space of RIS m, then we have $\mathbf{f}_{lk,m} = \mathbf{0}$ (or $\mathbf{G}_{mi} = \mathbf{0}$). In other words, in order to get a signal through a reflective BD-RIS, the transceivers should be in the reflection space of the BD-RIS. For more details on different architectures of BD-RIS, we refer the reader to [36]. It should be noted that realizing a BD-RIS is more challenging than a regular RIS since the BD-RIS elements should be connected through a proper circuit design. Additionally, optimizing the BD-RIS elements is more complicated than optimization of a regular RIS due to the unitary and symmetric constraints. However, BD-RIS is more general than a regular RIS, and an optimal BD-RIS scheme never performs worse than any scheme with regular RIS. Hereafter, for notational simplicity, we drop the dependency of the channels on $\boldsymbol{\Theta}_m$ and represent the channels as $\mathbf{h}_{lk,i}$ for all i, l, k.

Note that there are more assumptions regarding the coefficients of regular RISs, as mentioned in [3, Sec. II.B]. However, to the best of our knowledge, there is no other feasibility set for BD-RIS coefficients (apart from $\mathcal{T}_U$ and $\mathcal{T}_I$) in the literature. Indeed, we are not aware of any implementation of BD-RIS in practice, and the feasibility sets $\mathcal{T}_U$ and $\mathcal{T}_I$ are mainly based on the fact that the BD-RIS is assumed to operate in a passive mode. Such feasibility sets may provide an upper bound for the performance of BD-RIS, which can be useful to clarify the limitations of BD-RIS and/or to investigate how much gain can be obtained by moving beyond the diagonality assumption for the RIS coefficients.

B. Signal model

We consider the 1-layer RS to manage intra-cell interference. In the 1-layer RS, each BS transmits one common message, to be decoded by all its associated users, in addition to K private messages intended for each individual user as

$$\mathbf{x}_l = \underbrace{\mathbf{x}_{c,l}s_{c,l}}_{\text{Common message}} + \underbrace{\sum_{k=1}^{K}\mathbf{x}_{p,lk}s_{p,lk}}_{\text{Private messages}} \in \mathbb{C}^{N_{BS}\times 1}, \qquad (6)$$

where $s_{c,l} \sim \mathcal{CN}(0,1)$ is the common message of BS l, and $s_{p,lk} \sim \mathcal{CN}(0,1)$ is the private message intended for u_{lk}.

Moreover, $\mathbf{x}_{c,l}$ and $\mathbf{x}_{p,lk}$ are, respectively, the beamforming vectors corresponding to the common message $s_{c,l}$ and private message $s_{p,lk}$. Note that $s_{c,l}$ and $s_{p,lk}$ for all l, k are independent and identically distributed proper Gaussian signals. Moreover, the transmission power of the common message at BS l (private message intended for u_{lk}) is $\mathbf{x}_{c,l}^H\mathbf{x}_{c,l}$ ($\mathbf{x}_{p,lk}^H\mathbf{x}_{p,lk}$).

The received signal for the user u_{lk} is

$$y_{lk} = \underbrace{\mathbf{h}_{lk,l}\mathbf{x}_{c,l}s_{c,l}}_{\text{Desired C. signal}} + \underbrace{\mathbf{h}_{lk,l}\mathbf{x}_{p,lk}s_{p,lk}}_{\text{Desired P. signal}} + \mathbf{h}_{lk,l}\underbrace{\sum_{j=1,j\neq k}^{K}\mathbf{x}_{p,lj}s_{p,lj}}_{\text{Intracell interference}}$$

$$+ \underbrace{\sum_{i=1,i\neq l}^{L}\mathbf{h}_{lk,i}\mathbf{x}_i}_{\text{Intercell interference}} + \underbrace{n_{lk}}_{\text{Noise}}, \qquad (7)$$

where $\mathbf{h}_{lk,i} \in \mathbb{C}^{1\times N_{BS}}$ is the channel between BS l and u_{lk}, given by (2), and $n_{lk} \sim \mathcal{CN}(0,\sigma^2)$ is additive white Gaussian noise, which is independent of the transmitted signals.

C. Rate and energy-efficiency expressions

Each user first decodes the common message, treating all other signals as noise. Thus, the rate of decoding $s_{c,l}$ at u_{lk} is [7], [12, Eq. (8)]

$$\bar{r}_{c,lk} = \underbrace{\log(1 + \gamma_{c,lk})}_{\text{Shannon Rate}} - \underbrace{Q^{-1}(\epsilon_c)\sqrt{\frac{V_{c,lk}}{n_c}}}_{\delta_{c,lk}(\{\mathbf{x}\},\{\boldsymbol{\Theta}\})}, \qquad (8)$$

where $V_{c,lk}$ is the channel dispersion for decoding $s_{c,l}$ at u_{lk}, n_c is the packet length of the common message in bits, ϵ_c is the decoding error probability of the common message, and $\gamma_{c,lk}$ is the corresponding SINR given by

$$\gamma_{c,lk} = \frac{|\mathbf{h}_{lk,l}\mathbf{x}_{c,l}|^2}{\sigma^2 + \sum_{i\neq l}|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2 + \sum_{ij}|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2}. \qquad (9)$$

The optimal channel dispersion is [7]

$$V_{c,lk}^{opt} = 1 - \frac{1}{(1 + \gamma_{c,lk})^2}, \qquad (10)$$

which is not achievable by Gaussian signals in the presence of interference [67]. An achievable channel dispersion for Gaussian signals in interference-limited systems is [67]

$$V_{c,lk} = 2\frac{\gamma_{c,lk}}{1 + \gamma_{c,lk}}. \qquad (11)$$

The common message $s_{c,l}$ should be decodable for all the users associated to BS l. Hence, the transmission rate of $s_{c,l}$ must be smaller than or equal to the minimum achievable rate of decoding $s_{c,l}$ at the users associated to BS l, i.e.,

$$r_l \leq \min_k\{\bar{r}_{c,lk}\} \triangleq r_{c,l}. \qquad (12)$$

Each user decodes and cancels the common message. After this, it decodes its own private message, treating the remaining signals as noise. Thus, the decoding rate of $s_{p,lk}$ at u_{lk} is [7], [12, Eq. (8)]

$$r_{p,lk} = \underbrace{\log(1 + \gamma_{p,lk})}_{\text{Shannon Rate}} - \underbrace{Q^{-1}(\epsilon_p)\sqrt{\frac{V_{p,lk}}{n_p}}}_{\delta_{p,lk}(\{\mathbf{x}\},\{\boldsymbol{\Theta}\})}, \qquad (13)$$

where ϵ_p is the decoding error probability of the private message, n_p is the packet length of the private message in bits, and $\gamma_{p,lk}$ is the SINR for decoding $s_{p,lk}$ given by

$$\gamma_{p,lk} = \frac{|\mathbf{h}_{lk,l}\mathbf{x}_{p,lk}|^2}{\sigma^2 + \sum_{i\neq l}|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2 + \sum_{[ij]\neq[lk]}|\mathbf{h}_{lk,l}\mathbf{x}_{p,lj}|^2}, \quad (14)$$

where $\sum_{[ij]\neq[lk]}|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 = \sum_{ij}|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 - |\mathbf{h}_{lk,l}\mathbf{x}_{p,lk}|^2$. Similarly, an achievable channel dispersion is [67]

$$V_{p,lk} = 2\frac{\gamma_{p,lk}}{1 + \gamma_{p,lk}}. \quad (15)$$

Finally, the rate of u_{lk} is

$$r_{lk} = r_{p,lk} + r_{c,lk}, \quad (16)$$

where $r_{c,lk} \geq 0$ is the portion of the rate that u_{lk} gets from the common message. Note that $\sum_k r_{c,lk} \leq r_{c,l}$, where $r_{c,l}$ is given by (12).

Finally, the EE of u_{lk} is defined as [68]

$$e_{lk} = \frac{r_{lk}}{p_c + \eta\left(\mathbf{x}_{p,lk}^H\mathbf{x}_{p,lk} + \mathbf{x}_{c,l}^H\mathbf{x}_{c,l}/K\right)}, \quad (17)$$

where η^{-1} is the power efficiency of the BSs, and p_c is the constant power consumption to transmit data to each user, given by [17, Eq. (27)].

D. Discussion on reliability and latency constraints

The reliability constraint is modeled by the decoding error probabilities, ϵ_c and ϵ_p. The total decoding error probability for a user can be upper bounded as

$$\epsilon_t = \epsilon_c + (1 - \epsilon_c)\epsilon_p \approx \epsilon_c + \epsilon_p. \quad (18)$$

In general, the error probability of decoding private and/or common messages can be different at each user, based on their service requirements. To simplify the notations, we consider a symmetric system in which the decoding error probability of the private messages are the same at all users. However, this framework can be easily modified for asymmetric scenarios.

The latency constraint can be also translated to a rate constraint. The reason is that, if the latency for a packet with length n bits should be less than T seconds, then its transmission rate should be higher than $r \geq \frac{n}{\beta T}$ (b/s/Hz), where β is the used bandwidth. Thus, the rate of u_{lk} should be $r_{lk} \geq r_{lk}^{th} = \frac{n_p + n_c}{\beta T}$ (b/s/Hz), where T is the latency constraint.

Note that ϵ_t and r_{lk}^{th} are upper bounds for the decoding error probability and the latency, respectively, since it may happen that u_{lk} receives its rate from only the common message [43]. In this case, $\epsilon_t = \epsilon_c$ and $r_{lk}^{th} = \frac{n_c}{\beta T}$. However, in this paper, we consider the upper bounds to ensure that the latency and reliability constraints are met.

E. Problem statement

We consider a general optimization problem, similar to, e.g., [5, Eq. (29)], as

$$\max_{\{\mathbf{x}\}\in\mathcal{X},\{\boldsymbol{\Theta}\}\in\mathcal{T},\mathbf{r}_c} f_0\left(\{\mathbf{x}\},\{\boldsymbol{\Theta}\}\right) \quad (19a)$$

$$\text{s.t.} \quad f_i\left(\{\mathbf{x}\},\{\boldsymbol{\Theta}\}\right) \geq 0, \forall i, \quad (19b)$$

$$r_{lk} \geq r_{lk}^{th}, \forall l,k, \quad (19c)$$

$$\sum_{k=1}^K r_{c,lk} \leq \min_k\{\bar{r}_{c,lk}\} \triangleq r_{c,l}(\{\mathbf{x}\}), \forall lk, \quad (19d)$$

$$r_{c,lk} \geq 0, \forall l,k, \quad (19e)$$

where $\{\mathbf{x}\}$ is the set of the beamforming vectors, $\mathcal{X}$ is the feasibility set for the beamforming vectors, $\mathbf{r}_c = \{r_{c,lk}, \forall lk\}$ is the set of the common rates, $\mathcal{T}$ is the feasibility set for RIS components, which can be either $\mathcal{T}_I$ or $\mathcal{T}_U$, f_is are linear functions of rates/EEs and/or concave/convex/linear functions of beamforming vectors and/or channels, (19c) is the latency constraint, (19d) is the decodability constraint of each common message, (19e) is because of non-negative rates. The variables $\mathbf{r}_c$, $\{\mathbf{x}\}$ and $\{\boldsymbol{\Theta}\}$ are the optimization parameters. Indeed, we jointly optimize the RSMA parameters, beamforming vectors, transmission powers and RIS elements. Finally, the feasibility set $\mathcal{X}$ is

$$\mathcal{X} = \left\{\mathbf{x}_{p,lk}, \mathbf{x}_{c,l} : \mathbf{x}_{c,l}^H\mathbf{x}_{c,l} + \sum_k \mathbf{x}_{p,lk}^H\mathbf{x}_{p,lk} \leq p_l, \forall l\right\}, \quad (20)$$

where p_l is the power budget of BS l.

The general optimization problem (19) includes a large family of optimization problems for enhancing SE and EE of the system. For instance, (19) includes the minimum weighted rate maximization (MWRM), weighted sum rate maximization (WSRM), minimum weighted EE maximization (MWEEM), global EE maximization (GEEM) problems, to mention a few.

III. PROPOSED OPTIMIZATION FRAMEWORK FOR OPTIMIZING BEAMFORMING VECTORS

Our proposed optimization framework is an iterative optimization technique, based on majorization minimization (MM) and alternating optimization (AO). That is, we first fix RIS components to $\{\boldsymbol{\Theta}^{(t-1)}\}$ and solve (19) over the beamforming vectors $\{\mathbf{x}\}$ to obtain $\{\mathbf{x}^{(t)}\}$. We then alternate the optimization parameters and solve (19) for fixed beamforming vectors $\{\mathbf{x}^{(t)}\}$ to obtain $\{\boldsymbol{\Theta}^{(t)}\}$. We continue this procedure until a convergence metric is met. To obtain a feasible initial point, we employ the approach in [25, Appendix A] and maximize the minimum of $\gamma_{p,lk}$ and $\gamma_{c,lk}$ for all l,k until the rates $r_{p,lk}$ and $r_{c,lk}$ for all l,k are positive. The reason is that, the rates with FBL can be negative if the SINR is very low as shown in [25, Lemma 2], which is not a feasible initial point. Note that it is not ensured that AO and/or MM converge to a global optimal solution. However, to the best of our knowledge, there is no work on multi-user and multiple-antenna RIS-assisted systems that obtains the global optimal solution of the general optimization problem in (19), even with a simpler scenario, i.e., considering Shannon rates and/or TIN. Additionally, most of the existing works on RIS (e.g., [13]–[16]) employ AO to jointly optimize the transmission parameters and RIS elements. Hence, we employ AO to tackle (19).

In this section, we present our schemes to optimize beamforming vectors as well as the RSMA parameters. To update $\{\mathbf{x}\}$, we solve (19) for fixed $\{\boldsymbol{\Theta}^{(t-1)}\}$, which is a complicated non-convex optimization problem. To this end, we employ an MM-based approach. That is, we first find suitable surrogate

functions for the rates and then, solve the corresponding surrogate optimization problem. The surrogate functions should be concave lower bounds for the rates, satisfying the three conditions in [69, Lemma 3] to ensure a convergence to a stationary point of (19). Note that our proposed framework for systems without RIS converges to a stationary point of (19) since the surrogate functions fulfill these conditions. We will discuss the optimality conditions of our framework for systems with (BD-)RIS in the next section. In the following lemma, we present concave lower bounds for $r_{p,lk}$ and $\bar{r}_{c,lk}$ that satisfy the conditions in [69, Lemma 3].

Lemma 1. *Concave and quadratic lower bounds for $r_{p,lk}$ and $\bar{r}_{c,lk}$ are, respectively,*

$$
\begin{aligned}
r_{p,lk} \geq \tilde{r}_{p,lk} = & a_{p,lk} + 2c_{lk}\Re\left\{\left(\mathbf{h}_{lk,l}\mathbf{x}_{p,lk}^{(t-1)}\right)^*\mathbf{h}_{lk,l}\mathbf{x}_{p,lk}\right\} \\
& + \sum_{i\neq l} f_{p,lk}d_{lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}\right)^*\mathbf{h}_{lk,i}\mathbf{x}_{c,i}\right\} \\
& + \sum_{[ij]\neq[lk]} f_{p,lk}d_{lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t-1)}\right)^*\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}\right\} \\
& - b_{p,lk}d_{lk}\left(\sum_{ij}|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_{i\neq l}|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2\right), \quad (21)
\end{aligned}
$$

$$
\begin{aligned}
\bar{r}_{c,lk} \geq \tilde{r}_{c,lk} = & a_{c,lk} + 2d_{lk}\Re\left\{\left(\mathbf{h}_{lk,l}\mathbf{x}_{c,l}^{(t-1)}\right)^*\mathbf{h}_{lk,l}\mathbf{x}_{c,l}\right\} \\
& + \sum_{ij} e_{lk}f_{c,lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t-1)}\right)^*\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}\right\} \\
& + \sum_{i\neq l} e_{lk}f_{c,lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}\right)^*\mathbf{h}_{lk,i}\mathbf{x}_{c,i}\right\} \\
& - b_{p,lk}e_{lk}\left(\sum_{ij}|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_{i}|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2\right), \quad (22)
\end{aligned}
$$

where t is the number of the current iteration, $a_{c,lk}$, $b_{c,lk}$, $a_{p,lk}$, $b_{p,lk}$, c_{lk}, d_{lk}, e_{lk}, $f_{p,lk}$, and $f_{c,lk}$ are constants, given by

$$
\begin{aligned}
a_{p,lk} = & \log\left(1 + \gamma_{p,lk}^{(t-1)}\right) - \gamma_{p,lk}^{(t-1)} + (f_{p,lk} - b_{p,lk})d_{lk}\sigma^2 \\
& - \frac{Q^{-1}(\epsilon_p)}{\sqrt{n_p}}\left(\frac{\sqrt{V_{p,lk}^{(t-1)}}}{2} + \frac{1}{\sqrt{V_{p,lk}^{(t-1)}}}\right),
\end{aligned}
$$

$$
\begin{aligned}
a_{c,lk} = & \log\left(1 + \gamma_{c,lk}^{(t-1)}\right) - \gamma_{c,lk}^{(t-1)} + (f_{c,lk} - b_{c,lk})e_{lk}\sigma^2 \\
& - \frac{Q^{-1}(\epsilon_c)}{\sqrt{n_c}}\left(\frac{\sqrt{V_{c,lk}^{(t-1)}}}{2} + \frac{1}{\sqrt{V_{c,lk}^{(t-1)}}}\right),
\end{aligned}
$$

$$
b_{p,lk} = \gamma_{p,lk}^{(t)} + \frac{\zeta_{p,lk}^{(t-1)}f_{p,lk}}{2}, \quad b_{c,lk} = \gamma_{c,lk}^{(t)} + \frac{\zeta_{c,lk}^{(t-1)}f_{c,lk}}{2},
$$

$$
f_{p,lk} = \frac{2Q^{-1}(\epsilon_p)}{\sqrt{n_p V_{p,lk}^{(t-1)}}}, \quad f_{c,lk} = \frac{2Q^{-1}(\epsilon_c)}{\sqrt{n_c V_{c,lk}^{(t-1)}}},
$$

$$
c_{lk} = \left(\sigma^2 + \sum_{i\neq l}|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}|^2 + \sum_{[ij]\neq[lk]}|\mathbf{h}_{lk,l}\mathbf{x}_{p,lj}^{(t-1)}|^2\right)^{-1},
$$

$$
d_{lk} = \left(\sigma^2 + \sum_{ij}\left|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t-1)}\right|^2 + \sum_{i\neq l}\left|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}\right|^2\right)^{-1},
$$

$$
e_{lk} = \left(\sigma^2 + \sum_{ij}\left|\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t-1)}\right|^2 + \sum_{i}\left|\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}\right|^2\right)^{-1},
$$

where $\gamma_{c,lk}^{(t-1)}$, $V_{c,lk}^{(t-1)}$, $\gamma_{p,lk}^{(t-1)}$ and $V_{p,lk}^{(t-1)}$ are, respectively, obtained by replacing $\{\mathbf{x}^{(t-1)}\}$ in (9), (11), (14) and (15). Moreover, $\zeta_{p,lk}^{(t-1)} = d_{lk}c_{lk}^{-1}$ and $\zeta_{c,lk}^{(t-1)} = e_{lk}d_{lk}^{-1}$.

Proof. Please refer to Appendix A. $\square$

Substituting $\tilde{r}_{p,lk}$s and $\tilde{r}_{c,lk}$s in f_is gives the surrogate functions $\tilde{f}_i$s and consequently, the following surrogate optimization problem

$$
\max_{\{\mathbf{x}\}\in\mathcal{X},\mathbf{r}_c} \tilde{f}_0\left(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\}\right) \tag{23a}
$$

$$
\text{s.t.} \quad \tilde{f}_i\left(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\}\right) \geq 0, \forall i, \tag{23b}
$$

$$
\tilde{r}_{lk} = \tilde{r}_{p,lk} + r_{c,lk} \geq r_{lk}^{th}, \forall l, k, \tag{23c}
$$

$$
\sum_{k=1}^{K} r_{c,lk} \leq \min_k\{\tilde{r}_{c,lk}\} \triangleq \tilde{r}_{c,l}(\{\mathbf{x}\}), \forall l, \tag{23d}
$$

$$
r_{c,lk} \geq 0, \forall l, k. \tag{23e}
$$

This optimization problem is convex for spectral efficiency metrics, which can be efficiently solved by numerical tools. We can employ Dinkelbach-based algorithms to find the solution of (23) for energy efficiency metrics such as GEE and weighted minimum EE. Due to a space restriction, we do not provide the solutions here and refer the reader to [18], [68] for more details.

IV. PROPOSED OPTIMIZATION FRAMEWORK FOR OPTIMIZING BD-RIS COMPONENTS

In this section, we solve (19) for fixed $\{\mathbf{x}^{(t)}\}$ to update the BD-RIS components. To this end, we employ an MM-based algorithm to obtain a suboptimal solution for the complicated optimization problem. For fixed $\{\mathbf{x}^{(t)}\}$, (19) is non-convex because of two reasons. First, the rates are not concave in $\{\boldsymbol{\Theta}\}$. Second, the set $\mathcal{T}_I$ is non-convex due to the constraint $\boldsymbol{\Theta}\boldsymbol{\Theta}^H = \mathbf{I}$. To solve (19) for fixed $\{\mathbf{x}^{(t)}\}$, we have to handle these two issues. To this end, we first obtain suitable surrogate functions for the rates. Then we try to convexify $\mathcal{T}_I$. The rates have a similar structure with respect to the channels and beamforming vectors. Thus, we can employ an approach similar to Lemma 1 to find concave lower bounds for the rates with respect to $\{\boldsymbol{\Theta}\}$. In the following corollary, we state the concave lower bound for $\bar{r}_{c,lk}$. Due to a strict space restriction, we do not provide the concave lower bound for $r_{p,lk}$ since it is straightforward to obtain it from Lemma 1 and Corollary 1.

Corollary 1. *A concave lower bound for $\bar{r}_{c,lk}$ is*

$$
\begin{aligned}
\bar{r}_{c,lk} \geq \hat{r}_{c,lk} = & a_{c,lk} + 2d_{lk}\Re\left\{\left(\mathbf{h}_{lk,l}^{(t-1)}\mathbf{x}_{c,lk}^{(t)}\right)^*\mathbf{h}_{lk,l}\mathbf{x}_{c,lk}^{(t)}\right\} \\
& + \sum_{ij} e_{lk}f_{c,lk}\Re\left\{\left(\mathbf{h}_{lk,i}^{(t-1)}\mathbf{x}_{p,ij}^{(t)}\right)^*\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t)}\right\}
\end{aligned}
$$

$$+ \sum_{i \neq l} e_{lk} f_{c,lk} \Re \left\{ \left(\mathbf{h}_{lk,i}^{(t-1)} \mathbf{x}_{c,i}^{(t)} \right)^* \mathbf{h}_{lk,i} \mathbf{x}_{c,i}^{(t)} \right\}$$

$$- b_{p,lk} e_{lk} \left(\sum_{ij} \left| \mathbf{h}_{lk,i} \mathbf{x}_{p,ij}^{(t)} \right|^2 + \sum_i \left| \mathbf{h}_{lk,i} \mathbf{x}_{c,i}^{(t)} \right|^2 \right), \quad (24)$$

where $\mathbf{h}_{lk,i}^{(t-1)} = \mathbf{h}_{lk,i} \left(\boldsymbol{\Theta}^{(t-1)} \right)$, and the other parameters are defined as in Lemma 1.

Let us call the concave lower bound for $r_{p,lk}$ as $\hat{r}_{p,lk}$. Substituting $\hat{r}_{p,lk}$s and $\hat{r}_{c,lk}$s in f_is yield the surrogate functions $\hat{f}_i$s as well as the following problem

$$\max_{\{\boldsymbol{\Theta}\} \in \mathcal{T}, \mathbf{r}_c} \hat{f}_0 \left(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\} \right) \tag{25a}$$

$$\text{s.t.} \quad \hat{f}_i \left(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\} \right) \geq 0, \forall i, \tag{25b}$$

$$\hat{r}_{lk} = \hat{r}_{p,lk} + r_{c,lk} \geq r_{lk}^{th}, \forall l, k, \tag{25c}$$

$$\sum_{k=1}^{K} r_{c,lk} \leq \min_k \{\hat{r}_{c,lk}\} \triangleq \hat{r}_{c,l}(\{\boldsymbol{\Theta}\}), \forall lk, \tag{25d}$$

$$r_{c,lk} \geq 0, \forall l, k, \tag{25e}$$

which is convex only for $\mathcal{T}_U$. The proposed framework converges to a stationary point of (19) when $\mathcal{T}$ is a convex set. Note that $\mathcal{T}_U$ contains the convex constraint $\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}$, which may not be suitable for implementing in some existing numerical solvers. In the following, we propose a suboptimal approach to rewrite the constraint $\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}$ as a series of inequality constraints on scalar optimization parameters, which are referred to as disciplinary convex constraints that can be easily handled in numerical solvers. Moreover, we propose an approach to convexify $\mathcal{T}_I$ in Section IV-A.

1) Making $\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}$ a disciplinary convex constraint: The constraint $\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H \preceq \mathbf{I}$ can be equivalently expressed as $\mathbf{T} = \mathbf{I} - \boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H \succeq \mathbf{0}$. The matrix $\mathbf{T}$ can be written as

$$\mathbf{T} = \mathbf{I} - \boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H = \mathbf{I} - \boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^*$$
$$= \begin{bmatrix} 1 - |\theta_{11}|^2 - |\theta_{12}|^2 & -\theta_{11}^* \theta_{12} - \theta_{12}^* \theta_{22} \\ -(\theta_{11}^* \theta_{12} + \theta_{12}^* \theta_{22})^* & 1 - |\theta_{12}|^2 - |\theta_{22}|^2 \end{bmatrix}. \quad (26)$$

Since $\mathbf{T}$ is a 2×2 Hermitian matrix, $\mathbf{T}$ is PSD if and only if the following constraints hold

$$|\theta_{11}|^2 + |\theta_{12}|^2 \leq 1, \tag{27}$$

$$|\theta_{12}|^2 + |\theta_{22}|^2 \leq 1, \tag{28}$$

$$\zeta_{m_g} = \left(1 - |\theta_{11}|^2 - |\theta_{12}|^2 \right) \left(1 - |\theta_{12}|^2 - |\theta_{22}|^2 \right)$$
$$- |\theta_{11}^* \theta_{12} + \theta_{12}^* \theta_{22}|^2 \geq 0. \tag{29}$$

Note that since $\boldsymbol{\Theta}_{m_g}$ is symmetric, we have $\boldsymbol{\Theta}_{m_g}^H = \boldsymbol{\Theta}_{m_g}^*$. The constraints (27) and (28) are convex. Moreover, the constraint (29) can be simplified to

$$\zeta_{m_g} = \underbrace{1 - 2|\theta_{12}|^2 - |\theta_{11}|^2 - |\theta_{22}|^2 - |\theta_{12}|^2 |\theta_{11}^* + \theta_{22}|^2}_{\text{Concave Part}}$$
$$+ \underbrace{|\theta_{12}|^4 + |\theta_{11}|^2 |\theta_{22}|^2 + |\theta_{12}|^2 |\theta_{22}|^2 + |\theta_{11}|^2 |\theta_{12}|^2}_{\text{Convex Part}} \geq 0, \quad (30)$$

which is not a convex constraint since ζ_{m_g} is not a concave function. However, we can apply CCP to convexify this constraint. That is, we keep the concave part of ζ_{m_g} and find

a linear lower bound for the convex part of ζ_{m_g} by the first order Taylor expansion as

$$\zeta_{m_g} \geq \hat{\zeta}_{m_g} = 1 - 2|\theta_{12}|^2 - |\theta_{11}|^2 - |\theta_{22}|^2$$
$$- |\theta_{12}|^2 |\theta_{11}^* + \theta_{22}|^2 + 4|\theta_{12}^{(t-1)}|^2 \Re \left(\theta_{12}^{(t-1)} \theta_{12}^* \right) - 3|\theta_{12}^{(t-1)}|^4$$
$$+ 2|\theta_{22}^{(t-1)}|^2 \Re \left(\theta_{11}^{(t-1)} \theta_{11}^* \right) + 2|\theta_{11}^{(t-1)}|^2 \Re \left(\theta_{22}^{(t-1)} \theta_{22}^* \right)$$
$$- 3|\theta_{11}^{(t-1)}|^2 |\theta_{22}^{(t-1)}|^2 + 2|\theta_{22}^{(t-1)}|^2 \Re \left(\theta_{12}^{(t-1)} \theta_{12}^* \right)$$
$$+ 2|\theta_{12}^{(t-1)}|^2 \Re \left(\theta_{22}^{(t-1)} \theta_{22}^* \right) - 3|\theta_{12}^{(t-1)}|^2 |\theta_{22}^{(t-1)}|^2$$
$$+ 2|\theta_{11}^{(t-1)}|^2 \Re \left(\theta_{12}^{(t-1)} \theta_{12}^* \right) + 2|\theta_{12}^{(t-1)}|^2 \Re \left(\theta_{11}^{(t-1)} \theta_{11}^* \right)$$
$$- 3|\theta_{11}^{(t-1)}|^2 |\theta_{12}^{(t-1)}|^2 \geq 0. \quad (31)$$

Unfortunately, even though (31) is a convex constraint, it is not still a disciplinary constraint because of the convex function $|\theta_{12}|^2 |\theta_{11}^* + \theta_{22}|^2$. To address this issue, we employ the first order Taylor expansion to approximate $-|\theta_{12}|^2 |\theta_{11}^* + \theta_{22}|^2$ as a linear function of θ_{11}, θ_{12} and θ_{22}, which yields $\tilde{\zeta}_{m_g}$. The constraint $\tilde{\zeta}_{m_g} \geq 0$ is a disciplinary convex constraint, which can be easily implemented in existing numerical solvers. Finally, by inserting the corresponding constraints into (25), we have the following disciplinary convex problem

$$\max_{\{\boldsymbol{\Theta}\}, \mathbf{r}_c} \hat{f}_0 \left(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\} \right) \tag{32a}$$

$$\text{s.t.} \ (25b) - (25e), \tag{32b}$$

$$\tilde{\zeta} > 1 - \epsilon, (27), (28), \ \forall m, g. \tag{32c}$$

Let us call the solution of (32) as $\boldsymbol{\Theta}_{m_g}^{(\star)}$. Unfortunately, fulfilling the constraints $\tilde{\zeta}_{m_g} \geq 0$, (27) and (28) does not guarantee $\boldsymbol{\Theta}_{m_g}^{(\star)} \boldsymbol{\Theta}_{m_g}^{(\star)H} \preceq \mathbf{I}$. To ensure obtaining a feasible point, we first check if $\boldsymbol{\Theta}_{m_g}^{(\star)} \boldsymbol{\Theta}_{m_g}^{(\star)H} \preceq \mathbf{I}$ holds. If the largest eigenvalue of $\boldsymbol{\Theta}_{m_g}^{(\star)} \boldsymbol{\Theta}_{m_g}^{(\star)H}$ is greater than 1, i.e., $\lambda_{m_g} > 1$, then we choose $\hat{\boldsymbol{\Theta}}_{m_g} = \frac{\boldsymbol{\Theta}_{m_g}^{(\star)}}{\sqrt{\lambda_{m_g}}}$. Finally, we update $\boldsymbol{\Theta}_{m_g}$ as

$$\{\boldsymbol{\Theta}^{(t)}\} = \begin{cases} \{\hat{\boldsymbol{\Theta}}\} & \text{if} \quad f_0 \left(\{\hat{\boldsymbol{\Theta}}\} \right) > f_0 \left(\{\boldsymbol{\Theta}^{(t-1)}\} \right) \\ \{\boldsymbol{\Theta}^{(t-1)}\} & \text{Otherwise.} \end{cases} \tag{33}$$

This updating rule ensures generating a sequence of nondecreasing f_0, which guarantees the convergence, but we do not make any claim on the optimality of the algorithm in this case.

A. Convexifying $\mathcal{T}_I$

The constraints $\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H = \mathbf{I}$ and $\boldsymbol{\Theta}_{m_g} = \boldsymbol{\Theta}_{m_g}^T$ can be simplified as

$$\boldsymbol{\Theta}_{m_g} \boldsymbol{\Theta}_{m_g}^H = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{bmatrix} \begin{bmatrix} \theta_{11}^* & \theta_{12}^* \\ \theta_{12}^* & \theta_{22}^* \end{bmatrix}$$
$$= \begin{bmatrix} |\theta_{11}|^2 + |\theta_{12}|^2 & \theta_{11}^* \theta_{12} + \theta_{12}^* \theta_{22} \\ (\theta_{11}^* \theta_{12} + \theta_{12}^* \theta_{22})^* & |\theta_{12}|^2 + |\theta_{22}|^2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which results in

$$|\theta_{11}|^2 + |\theta_{12}|^2 = 1, \tag{34}$$

$$|\theta_{12}|^2 + |\theta_{22}|^2 = 1, \tag{35}$$

$$\theta_{11}^*\theta_{12} + \theta_{12}^*\theta_{22} = 0. \tag{36}$$

The constraints (34)-(36) are equivalent to $|\theta_{11}| = |\theta_{22}|$, and

$$\theta_{11}^*\theta_{12} + \theta_{12}^*\theta_{22} = |\theta_{12}|e^{j\angle\theta_{12}}\left(\theta_{11}^* + \theta_{22}e^{-j2\angle\theta_{12}}\right) = 0, \tag{37}$$

which yields $\theta_{22} = \theta_{11}^* e^{j(2\angle\theta_{12}+\pi)}$. As can be verified through (34)-(37), the phases and amplitudes of the coefficients for a group connected BD-RIS are highly dependent. For instance, if an amplitude of one of the coefficients is known, the amplitude of the other two coefficients can be uniquely obtained. Additionally, the phases are related to each other according to $\angle\theta_{22} = -\angle\theta_{11}+2\angle\theta_{12}+\pi$. The constraints (34) and (35) are not convex. To convexify (34), we rewrite it as the following two constraints:

$$|\theta_{11}|^2 + |\theta_{12}|^2 \le 1, \tag{38}$$

$$|\theta_{11}|^2 + |\theta_{12}|^2 \ge 1. \tag{39}$$

The constraint (38) is a convex constraint, but (39) is not since $|\theta_{11}|^2 + |\theta_{12}|^2$ is a jointly convex function in $|\theta_{11}|^2$ and $|\theta_{12}|^2$. Thus, we can employ CCP to convexify $|\theta_{11}|^2 + |\theta_{12}|^2 \ge 1$ and relax it for a faster convergence as

$$|\theta_{11}^{(t-1)}|^2 + 2\Re\left(\theta_{11}^{(t-1)}(\theta_{11} - \theta_{11}^{(t-1)})^*\right) + |\theta_{12}^{(t-1)}|^2$$
$$+ 2\Re\left(\theta_{12}^{(t-1)}(\theta_{12} - \theta_{12}^{(t-1)})^*\right) \ge 1-\epsilon, \tag{40}$$

where $\epsilon > 0$. Similarly, we can convexify (35) by considering the following constraints:

$$|\theta_{22}|^2 + |\theta_{12}|^2 \le 1 \tag{41}$$

$$|\theta_{22}^{(t-1)}|^2 + 2\Re\left(\theta_{22}^{(t-1)}(\theta_{22} - \theta_{22}^{(t-1)})^*\right)$$
$$+ |\theta_{12}^{(t-1)}|^2 + 2\Re\left(\theta_{12}^{(t-1)}(\theta_{12} - \theta_{12}^{(t-1)})^*\right) \ge 1-\epsilon, \tag{42}$$

In addition to (34) and (35), the constraint (36) (or $\theta_{22} = \theta_{11}^* e^{j(2\angle\theta_{12}+\pi)}$) is not convex neither. A suboptimal way to convexify this constraint is to fix the phase of θ_{12}. For instance, if θ_{12} is real (or pure imaginary), then $\theta_{22} = -\theta_{11}^*$ (or $\theta_{22} = \theta_{11}^*$). Note that when $\theta_{12} = 0$, the BD-RIS with group-connected architecture of group size two is equivalent to the diagonal RIS, and the constraint (36) is automatically satisfied. Thus, in this case, θ_{22} is independent of θ_{11}, and there is no need to consider $\theta_{22} = -\theta_{11}^*$ (or $\theta_{22} = \theta_{11}^*$). As a result, this algorithm never performs worse than the diagonal RIS. Finally, the corresponding optimization problem in this case is

$$\max_{\{\boldsymbol{\Theta}\}, \mathbf{r}_c} \hat{f}_0(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\}) \tag{43a}$$

$$\text{s.t. } (25b) - (25e), \tag{43b}$$

$$(38), (40) - (42), \forall m, g, \tag{43c}$$

$$\theta_{22} = -\theta_{11}^* \text{ if } \theta_{12} \ne 0, \forall m, g, \tag{43d}$$

which is convex and can be efficiently solved. Let us call the solution of (43) as $\boldsymbol{\Theta}_{m_g}^{(\star)}$. We first normalize $\boldsymbol{\Theta}_{m_g}^{(\star)}$ to satisfy (34) and (35), which results in $\hat{\boldsymbol{\Theta}}_{m_g}$. Finally, we update $\boldsymbol{\Theta}_{m_g}$ based on the rule in (33), which ensures the convergence, but we do not make any claim on the optimality for our framework with $\mathcal{T}_I$.
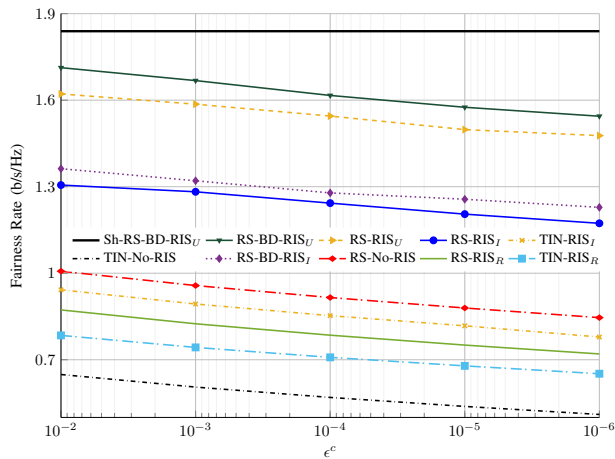
B. Computational complexity analysis

In this subsection, we provide computational complexity analysis only for the MWRM problem with the feasibility set $\mathcal{T}_U$ due to a strict page limitation. However, it can be straightforward to extend the analysis to other optimization problems with feasibility set $\mathcal{T}_I$. Note that the actual computational complexity of our algorithms depends on their implementation. Here, we provide an approximation for the number of multiplications to obtain a solution for our schemes.
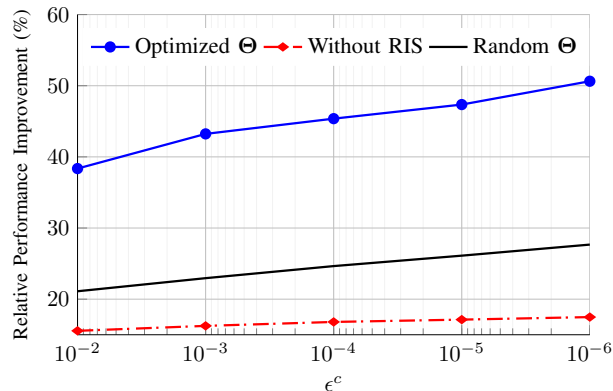
Our proposed framework is iterative, and in each iteration, two convex problems should be solved to update the beamforming vectors, RSMA parameters and BD-RIS coefficients. Firstly, we have to solve (23) to update $\{\mathbf{x}\}$. According to [70, Chapter 11], the number of Newton iterations to solve a convex optimization problem is proportional to the square root of the number of constraints in the problem. Thus, the number of Newton iterations to solve (23) grows with the square root of the total number of users in the system, i.e., $\sqrt{LK}$. To solve each Newton iteration, we have to obtain $2LK$ concave lower bounds for the rates. Each concave lower bound has a quadratic structure, and the number of multiplications to compute each is proportional to $N_{BS}LK$. Thus, we can approximate the computational complexity for solving (23) by $\mathcal{O}\left(N_{BS}L^2K^2\sqrt{KL}\right)$. To update $\{\boldsymbol{\Theta}\}$, we have to solve (32), which has $3KL + 1.5MN_{RIS}$ constraints. To solve each iteration, we have to compute LK channels, which approximately requires $MN_{RIS}^2N_{BS}$ multiplications per channel. Furthermore, we have to compute $2LK$ rates, which approximately requires $2N_{BS}L^2K^2$ multiplications ($N_{BS}LK$ multiplications for each rate). Hence, the overall computational complexity of solving (32) can be approximated as $\mathcal{O}\left(LKN_{BS}\left(MN_{RIS}^2 + 2LK\right)\sqrt{3KL + 1.5MN_{RIS}}\right)$. Considering N as the maximum number of iterations for our algorithm, the overall computational complexity can be approximated as N times the summation of the computational complexities of updating $\{\mathbf{x}\}$ and $\{\boldsymbol{\Theta}\}$.

V. NUMERICAL RESULTS

In this section, we provide some numerical results for the MWRM and MWEEM problems. We consider $\epsilon_c = \epsilon_p$ and $n_p = n_c$. We assume that there is a line of sight (LoS) link between each BS and each (BD-)RIS as well as between each (BD-)RIS and each user, similar to e.g., [16], [17]. Thus, we assume that small-scale fading for the channels $\mathbf{f}_{lk,m}$ and $\mathbf{G}_{mi}$ for all l, k, m, i follow Rician fading as described in [16, Eq. (55)]. We consider a Rician factor of 3 for these channels [16]. Additionally, we assume that there is a non-LoS link between each BS and each user, which results in a Rayleigh small-scale fading for $\mathbf{d}_{lk,i}$ for all l, k, i [16], [17]. For large-scale fading, we employ the well-known path loss model, given in e.g., [17, Eq. (59)]. The other propagation parameters, including the antenna gains, bandwidth, noise power density, path loss components, and the path loss at the reference distance of 1 meter, are based on [17], [25]. Moreover, the simulation scenario is a two-cell MISO BC, as depicted in [25, Fig. 5], where the BSs/RISs/users positions are chosen similar to [25]. Additionally, the considered schemes
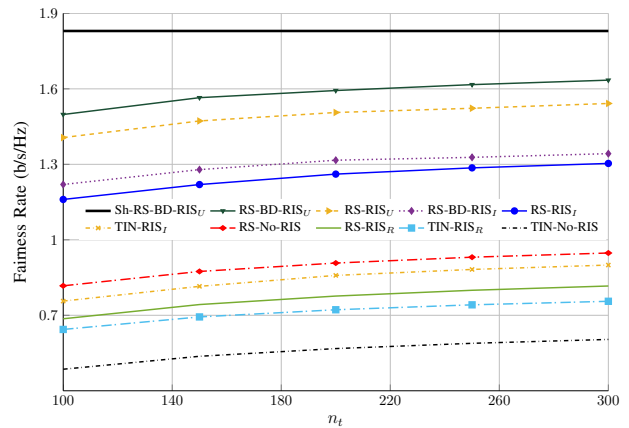
(a) Average fairness rate.

(b) Average relative performance improvement by RSMA.

Fig. 2: The average fairness rate and performance improvement by RSMA versus ϵ^c for multi-cell MISO BC with $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$ bits, $K = 6$, $N_{BS} = 5$ and $P = 10$dB.
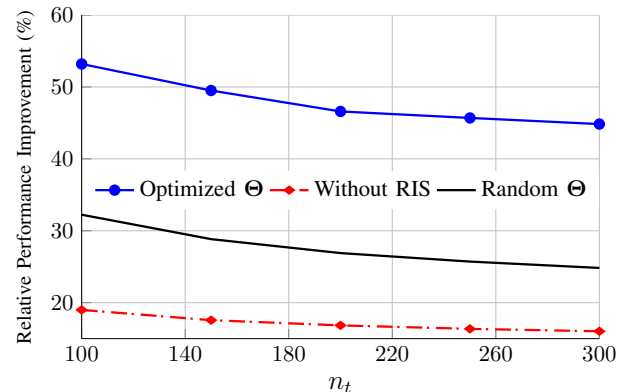

(a) Average fairness rate.

(b) Average relative performance improvement by RSMA.

Fig. 3: The average fairness rate and performance improvement by RSMA versus n_t for multi-cell MISO BC with $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 256$, $K = 6$, $\epsilon^c = 0.001$ and $P = 10$dB.

in the simulations are represented as: **RS-BD-RIS**$_U$ (or **RS-BD-RIS**$_I$): The proposed scheme for BD-RIS-assisted systems with rate splitting, BD-RIS with group-connected architecture of group size two and feasibility set $\mathcal{T}_U$ (or $\mathcal{T}_I$). **RS-RIS**$_X$ (or **TIN-RIS**$_X$): The proposed algorithm for RIS-assisted systems with rate splitting (or TIN), regular RIS and feasibility set $\mathcal{T}_X$, where X can be U or I. **RS-RIS**$_R$ (or **TIN-RIS**$_R$): The proposed algorithm for RIS-assisted systems with rate splitting (or TIN), regular RIS and random RIS elements. **RS-No-RIS** (or **TIN-No-RIS**): The RSMA (or TIN) scheme for systems without RIS. **Sh-RS-BD-RIS**$_U$: The proposed scheme for BD-RIS-assisted systems with rate splitting, BD-RIS with group-connected architecture of group size two, feasibility set $\mathcal{T}_U$ and considering Shannon rates.

A. Fairness rate

In this subsection, we present some numerical results for the fairness rate, which is defined as the maximum of the minimum rate of users. We consider the impact of the reliability constraint, packet length, number of users per cell and power budget on the performance of RSMA and RIS. Additionally, we compare the performance of regular RIS with BD-RIS with group connected architecture of group size two.

1) Impact of the reliability constraint: Fig. 2 shows the average fairness rate and performance improvement by RSMA versus ϵ^c for multi-cell MISO BC with $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$ bits, $K = 6$, $N_{BS} = 5$ and $P = 10$dB. As expected, the fairness rate for all schemes decreases when the reliability constraint is more stringent. Moreover, we can observe that the RSMA scheme with BD-RIS with group-connected architecture of group size two and the feasibility set $\mathcal{T}_U$ can outperform the other schemes. Additionally, we observe that RIS (either regular or BD) can significantly improve the system performance and enhance the RSMA benefits. Note that in this example, the number of users per cell is higher than the number of BS antennas, which means that the considered system is overloaded. Interestingly, we can observe that RSMA without RIS can outperform TIN with regular RIS, which shows the importance of employing an effective interference-management technique in overloaded systems. Finally, we observe that the benefits of RSMA increase with ϵ^{c-1}. It shows that managing interference is even more important in URLLC systems. Furthermore, we observe that the RSMA benefits with employing RIS with properly optimized elements.

2) Impact of the packet length: Fig. 3 shows the average fairness rate and performance improvement by RSMA versus n_t for multi-cell MISO BC with $N_{RIS} = 20$, $L = 2$, $M = $
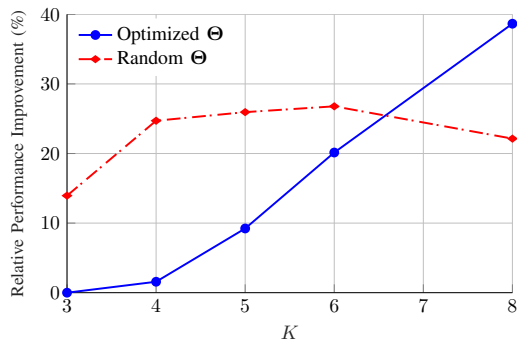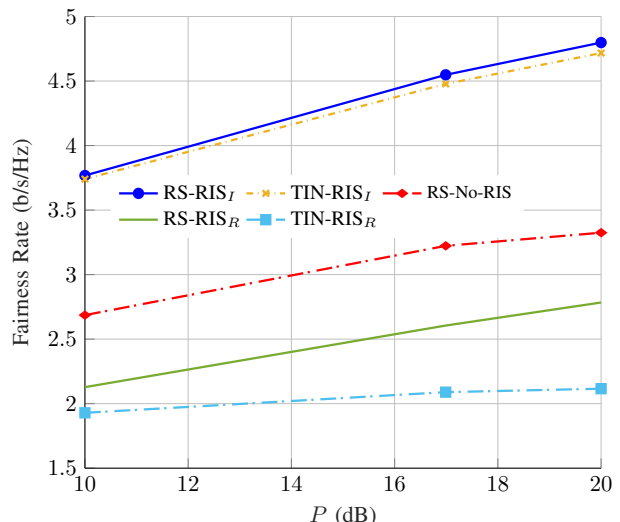
Fig. 4: The average performance improvement by RSMA versus K for multi-cell MISO RIS-assisted BC with $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$ and $P = 17$dB.
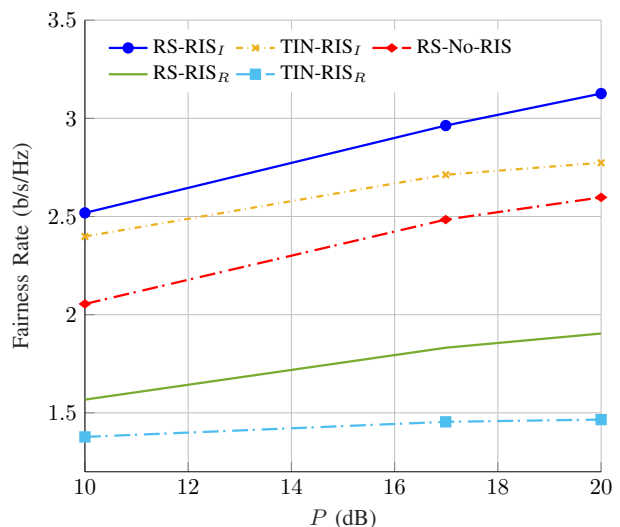
2, $n_t = 256$, $K = 6$, $\epsilon^c = 0.001$ and $P = 10$dB. As can be observed, the RSMA scheme with BD-RIS with group-connected architecture of group size two and the feasibility set $\mathcal{T}_U$ can outperform the other schemes. Moreover, we can observe that RIS and RSMA can significantly increase the average fairness rate. As expected, the average fairness rate increases with n_t for all the schemes. However, the benefits of RSMA decrease with n_t. Indeed, when the packet lengths are shorter (or the latency constraint is more stringent), RSMA can provide higher gains, which indicates that RSMA can be even more effective in URLLC systems.

3) Impact of the number of users per cell: Fig. 4 shows the impact of the number of users per cell as well as optimizing RIS components on the performance of RSMA. As can be observed, the benefits of RSMA can highly vary when optimizing RIS elements. Interestingly, we can observe that RIS may decrease RSMA benefits in underloaded systems, while it enhances the RSMA gain in overloaded systems. This happens because RIS can modulate the channels to mitigate the interference in underloaded systems. Thus, a proper design of RIS elements may decrease the RSMA benefits in underloaded systems. However, as the number of users grows, the interference becomes more severe, and RIS alone cannot completely mitigate it. As a result, we observe that the RSMA benefits monotonically increase with K when RIS elements are properly designed.

4) Impact of power budget: Fig. 5 shows the average fairness rate versus the BSs power budget P for $N_{BS} = 8$, $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$ and different K. In this figure, the number of users per cell is less than the number of BS antennas, which is referred to as underloaded systems. As can be observed, RSMA cannot provide any considerable benefits in RIS-assisted systems for $K = 4$. However, RSMA can significantly outperform TIN for $K = 5$, especially in higher SNR regimes. Indeed, it shows that RSMA can be still beneficial in underloaded systems. Additionally, we observe that RSMA can significantly outperform TIN with random RIS elements in underloaded scenarios. However, such benefits are lower (or may even vanish) when RIS elements are optimized. This result indicates that RIS can be employed as an interference-management technique in underloaded BCs. However, we should employ

(a) $K = 4$.

(b) $K = 5$.

Fig. 5: The average fairness rate versus the BSs power budget P for $N_{BS} = 8$, $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$ and K.

more advanced interference-management techniques to fully reap RIS benefits for overloaded systems, which is in line with the findings in [15], [18].

5) Comparison of RIS technologies: In Fig. 3 and Fig. 2, we compare the performance of various schemes, including regular and BD-RIS with different feasibility sets. In this part, we provide another comparison for regular RIS and BD-RIS with group-connected architecture of groups size two. Fig. 6 shows the average fairness rate versus the BSs power budget P for $N_{BS} = 5$, $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, and $\epsilon^c = 0.001$. As can be observed, the BD-RIS with group-architecture of group size two and constraint $\Theta\Theta^H \preceq \mathbf{I}$ can highly outperform a regular diagonal RIS. As indicated in Section IV, BD-RIS is a more general architecture than regular RIS that includes regular RIS as a special case. Thus, BD-RIS never performs worse than regular RIS.
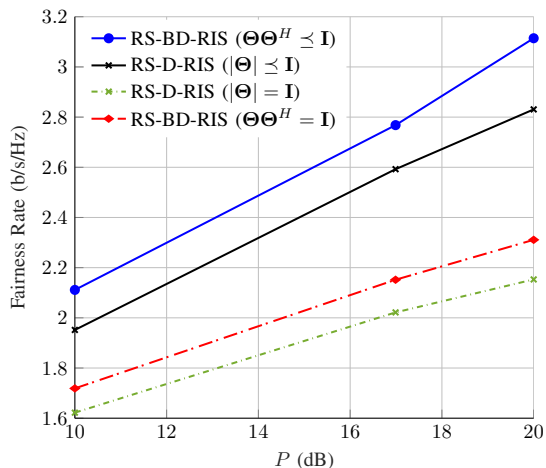
Fig. 6: The average fairness rate versus power budget for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$, and $N_{BS} = 5$.
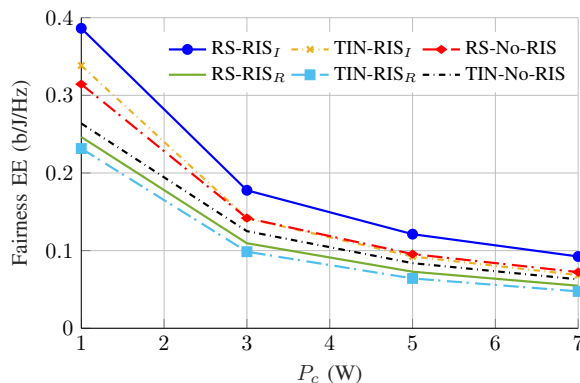

Fig. 7: The average fairness EE versus P_c for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$, $N_{BS} = 5$ and $K = 7$.

Fig. 8: The average performance improvement by RSMA versus P_c for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$, and $N_{BS} = 5$.

Fig. 9: The average performance improvement by RIS for the RSMA schemes versus P_c for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$, and $N_{BS} = 5$.

B. Fairness EE

In this subsection, we present some numerical results for the minimum EE of users, which is referred to as the fairness EE. In Fig. 7, we show the average fairness EE versus P_c for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_p = n_c = 200$, $\epsilon_c = \epsilon_p = 0.001$, $N_{BS} = 5$ and $K = 7$. As can be observed, RIS can highly increase the fairness EE for both RSMA and TIN schemes when RIS components are properly optimized. We can also observe that RSMA can outperform TIN with and without RIS. In the following, we provide an in-depth discussion of the benefits of RIS and RSMA from an EE point of view.

In Fig. 8, we show the average fairness EE improvement by RSMA for RIS-assisted systems. As can be observed, the benefits of employing RSMA increase with K. The reason is that, as the number of users grows, the interference level increases, which in turn improves the benefits of employing a powerful interference-management technique such as RS. Indeed, the more overloaded the systems is, the more gain RSMA can provide. We also observe that the RSMA benefits increase with P_c. Indeed, when the constant power consumption is higher, it is more important to properly design the system to get a better EE performance.

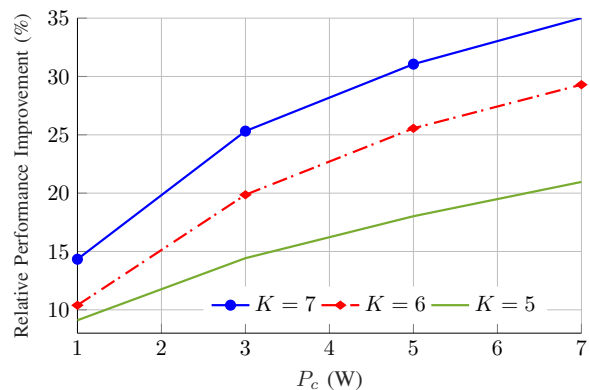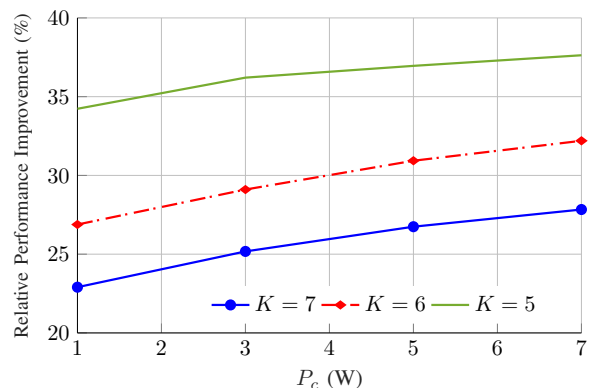Fig. 9 shows the average performance improvement by RIS

for the RSMA schemes versus P_c for $N_{RIS} = 20$, $L = 2$, $M = 2$, $n_t = 200$, $\epsilon^c = 0.001$, and $N_{BS} = 5$. As can be observed, the RIS benefits decrease with the number users for a fixed N_{RIS}. However, the RIS benefits are still significant even when the number of RIS elements per user is relatively low (less than 3 per user). This suggests that RIS may be promising in practical scenarios. We also observe that the RIS benefits increase with P_c, which is in line with the results in Fig. 8.

C. Summary

Our main findings in the numerical section can be summarized as follows:

- RSMA and RIS can significantly improve the SE and EE of the system. Moreover, the combination of the RSMA and BD-RIS with group-connected architecture of group size two outperforms the other schemes.
- The use of RIS increases the benefits of RSMA in overloaded systems. However, the relative benefits of RSMA decrease by employing RIS in underloaded systems.
- The benefits of RSMA increase with the number of users, K, since the interference level increases when there are more users in the system. However, the benefits of RIS

decrease with K when N_{RIS} is fixed. The reason is that the number of RIS elements per user decreases with K for a fixed N_{RIS}, which in turn reduces the RIS benefits.

- RSMA provides higher gains when the packet lengths are shorter or when the reliability constraint is more stringent. Therefore, RSMA is more effective in URLLC systems, especially in highly overloaded systems.

VI. Conclusion and future work

In this paper, we showed that RSMA and RIS can significantly improve the spectral and energy efficiency of MISO multi-cell BC URLLC systems. Moreover, we investigated the role of RSMA and RIS in URLLC systems, analyzing the impact on performance of different parameters, such as the reliability constraint, packet length, number of users per cell and BS power budget. We showed that the use of RIS has a different impact on the benefits of RSMA in different operational points. Specifically, RIS decreases RSMA benefits in underloaded systems, while it enhances the RSMA gain in overloaded systems. Indeed, RSMA and RIS can be mutually beneficial tools in overloaded systems. In addition, RSMA provides higher gains when the reliability constraint is more stringent and/or when the packet lengths are shorter. Finally, we showed that BD-RIS with group-connected architecture of group size two can outperform regular RIS.

As a future work, it can be interesting to investigate the performance of multi-layer RSMA schemes in URLLC systems. Moreover, the performance gap between our proposed algorithms and the global optimal solution should be studied. Additionally, proposing robust designs for RSMA in multiple-antenna BD-RIS-assisted URLLC systems can be another challenging direction for future studies. In [71]–[74], the performance RIS has been studied in the presence of imperfect CSI, which might be helpful to obtain robust RSMA designs for RIS-assisted URLLC systems. Finally, there is no implementation of BD-RIS yet, and it is important to investigate the performance of BD-RIS with more realistic models based on practical implementations.

Appendix A
Proof of Lemma 1

Due to a space restriction, we only provide a proof for the inequality in (22). It is straightforward to extend the proof to obtain the lower bound $\tilde{r}_{p,lk}$ in (21). To prove (22), we obtain a concave lower bound for the Shannon rate as well as a convex upper bound for $\delta_{c,lk}(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\})$. Employing [25, Eq. (70)], we can obtain a lower bound for the Shannon rate as

$$
\begin{aligned}
\log\left(1 + \gamma_{c,lk}\right) \geq{}& \log\left(1 + \gamma_{c,lk}^{(t-1)}\right) - \gamma_{c,lk}^{(t-1)} \\
&+ 2d_{lk}\Re\left\{\left(\mathbf{h}_{lk,l}\mathbf{x}_{c,l}^{(t-1)}\right)^* \mathbf{h}_{lk,l}\mathbf{x}_{c,l}\right\} \\
&- \gamma_{c,lk}^{(t)} e_{lk}\left(\sigma^2 + \sum_{ij} |\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_i |\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2\right),
\end{aligned}
\tag{44}
$$

which is quadratic and concave in $\{\mathbf{x}\}$. To obtain a convex upper bound for $\delta_{c,lk}(\{\mathbf{x}\}, \{\boldsymbol{\Theta}\})$, we employ the inequality in [25, Eq. (70)], which results in

$$
\begin{aligned}
\sqrt{V_{c,lk}} \leq{}& \frac{\sqrt{V_{c,lk}^{(t)}}}{2} + \frac{\gamma_{c,lk}}{\sqrt{V_{c,lk}^{(t)}}\left(1 + \gamma_{c,lk}\right)} = \frac{\sqrt{V_{c,lk}^{(t)}}}{2} \\
&+ \frac{1}{\sqrt{V_{c,lk}^{(t)}}}\left(1 - \frac{\sigma^2 + \sum_{ij} |\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_{i \neq l} |\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2}{\sigma^2 + \sum_{ij} |\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_i |\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2}\right).
\end{aligned}
\tag{45}
$$

Unfortunately, the upper bound in the right-hand side of (45) is not convex in $\{\mathbf{x}\}$. To obtain a convex upper bound for the right-hand side of (45), we employ [25, Eq. (69)], which yields

$$
\begin{aligned}
\sqrt{V_{c,lk}} \leq{}& \frac{\sqrt{V_{c,lk}^{(t)}}}{2} + \frac{1}{\sqrt{V_{c,lk}^{(t)}}}\bigg(1 - 2\sigma^2 e_{lk} \\
&- 2\sum_{ij} e_{lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}^{(t-1)}\right)^* \mathbf{h}_{lk,i}\mathbf{x}_{p,ij}\right\} \\
&- 2\sum_{i \neq l} e_{lk}\Re\left\{\left(\mathbf{h}_{lk,i}\mathbf{x}_{c,i}^{(t-1)}\right)^* \mathbf{h}_{lk,i}\mathbf{x}_{c,i}\right\} \\
&+ e_{lk}\zeta_{c,lk}^{(t-1)}\left(\sum_{ij} |\mathbf{h}_{lk,i}\mathbf{x}_{p,ij}|^2 + \sum_i |\mathbf{h}_{lk,i}\mathbf{x}_{c,i}|^2\right)\bigg).
\end{aligned}
\tag{46}
$$

Finally, we can obtain the lower bound in (22) by substituting (45) and (46) in (13).

References

[1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surv. Tutor.*, pp. 1–1, 2023.

[2] T. Gong *et al.*, "Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions," *arXiv preprint arXiv:2212.01257*, 2022.

[3] Q. Wu *et al.*, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.

[4] M. Di Renzo *et al.*, "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] 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. Surv. Tutor.*, vol. 24, no. 4, pp. 2073–2126, 2022.

[6] 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.*, vol. 41, no. 5, pp. 1265–1308, 2023.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[8] T. Erseghe, "Coding in the finite-blocklength regime: Bounds based on laplace integrals and their asymptotic approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6854–6883, 2016.

[9] ——, "On the evaluation of the Polyanskiy-Poor–Verdú converse bound for finite block-length coding in AWGN," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6578–6590, 2015.

[10] A. A. Nasir, H. D. Tuan, H. H. Nguyen, M. Debbah, and H. V. Poor, "Resource allocation and beamforming design in the short blocklength regime for URLLC," *IEEE Trans. on Wireless Commun.*, vol. 20, no. 2, pp. 1321–1335, 2020.

[11] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink MISO OFDMA-URLLC systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7184–7200, 2020.

[12] A. Nasir, H. Tuan, H. Ngo, T. Duong, and H. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5861–5871, 2021.

[13] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[14] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[15] M. Soleymani, I. Santamaria, E. Jorswieck, and S. Rezvani, "NOMA-based improper signaling for multicell MISO RIS-assisted broadcast channels," *IEEE Trans. Signal Process.*, vol. 71, pp. 963–978, 2023.

[16] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.

[17] M. Soleymani, I. Santamaria, and P. J. Schreier, "Improper signaling for multicell MIMO RIS-assisted broadcast channels with I/Q imbalance," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 723–738, 2022.

[18] 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, April 2023.

[19] ——, "NOMA-based improper signaling for MIMO STAR-RIS-assisted broadcast channels with hardware impairments," *IEEE Global Commun. Conf. (GLOBECOM)*, 2023.

[20] ——, "Energy-efficient rate splitting for MIMO STAR-RIS-assisted broadcast channels with I/Q imbalance," *Proc. IEEE Eu. Signal Process. Conf. (EUSIPCO)*, 2023.

[21] I. Santamaria *et al.*, "Interference leakage minimization in RIS-assisted MIMO interference channels," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2023.

[22] Y. Li, C. Yin, T. Do-Duy, A. Masaracchia, and T. Q. Duong, "Aerial reconfigurable intelligent surface-enabled URLLC UAV systems," *IEEE Access*, vol. 9, pp. 140 248–140 257, 2021.

[23] H. Xie, J. Xu, Y.-F. Liu, L. Liu, and D. W. K. Ng, "User grouping and reflective beamforming for IRS-aided URLLC," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2533–2537, 2021.

[24] M. Abughalwa, H. Tuan, D. Nguyen, H. Poor, and L. Hanzo, "Finite-blocklength RIS-aided transmit beamforming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 12 374–12 379, Nov. 2022.

[25] M. Soleymani, I. Santamaria, and E. Jorswieck, "Spectral and energy efficiency maximization of MISO STAR-RIS-assisted URLLC systems," *IEEE Access*, vol. 11, pp. 70 833–70 852, 2023.

[26] T.-H. Vu, T.-V. Nguyen, D. B. da Costa, and S. Kim, "Intelligent reflecting surface-aided short-packet non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4500–4505, 2022.

[27] H. Ren, K. Wang, and C. Pan, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 707–723, 2021.

[28] B. Zhang, K. Wang, K. Yang, and G. Zhang, "IRS-assisted short packet wireless energy transfer and communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 303–307, 2022.

[29] M. Almekhlafi, M. A. Arfaoui, M. Elhattab, C. Assi, and A. Ghrayeb, "Joint resource allocation and phase shift optimization for RIS-aided eMBB/URLLC traffic multiplexing," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1304–1319, 2022.

[30] A. Ranjha and G. Kaddoum, "URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and uav positioning," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4618–4627, 2020.

[31] ——, "URLLC-enabled by laser powered UAV relay: A quasi-optimal design of resource allocation, trajectory planning and energy harvesting," *IEEE Trans. Vehicular Technology*, vol. 71, no. 1, pp. 753–765, 2022.

[32] S. Dhok, P. Raut, P. K. Sharma, K. Singh, and C.-P. Li, "Non-linear energy harvesting in RIS-assisted URLLC networks for industry automation," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7761–7774, 2021.

[33] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell IRS-aided OFDMA-URLLC systems," in *IEEE Wireless Commun. and Netw. Conf. (WCNC)*, 2021, pp. 1–7.

[34] H. Li, S. Shen, M. Nerini, and B. Clerckx, "Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices," *arXiv preprint arXiv:2301.03288*, 2023.

[35] 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, Feb. 2022.

[36] 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, April 2023.

[37] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6367–6383, 2022.

[38] B. Wang, H. Li, Z. Cheng, S. Shen, and B. Clerckx, "A dual-function radar-communication system empowered by beyond diagonal reconfigurable intelligent surface," *arXiv preprint arXiv:2301.03286*, 2023.

[39] I. Santamaria, M. Soleymani, E. Jorswieck, and J. Gutiérrez, "SNR maximization in beyond diagonal RIS-assisted single and multiple antenna links," *IEEE Signal Process. Lett.*, vol. 30, pp. 923–926, 2023.

[40] J. C. Ruiz-Sicilia, M. Di Renzo, M. Debbah, and H. V. Poor, "Low complexity optimization for line-of-sight RIS-aided holographic communications," *arXiv preprint arXiv:2304.06330*, 2023.

[41] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, 2015.

[42] G. Lu, L. Li, H. Tian, and F. Qian, "MMSE-based precoding for rate splitting systems with finite feedback," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 642–645, 2017.

[43] 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, 2019.

[44] Y. Mao and B. Clerckx, "Beyond dirty paper coding for multi-antenna broadcast channel with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6775–6791, 2020.

[45] A. R. Flores, R. C. de Lamare, and B. Clerckx, "Linear precoding and stream combining for rate splitting in multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 890–894, April 2020.

[46] M. Soleymani, I. Santamaria, and E. Jorswieck, "Rate region of MIMO RIS-assisted broadcast channels with rate splitting and improper signaling," in *Proc. IEEE 26th Int. ITG Workshop Smart Antennas & 13th Conf. Syst., Commun., and Coding (WSA & SCC)*, 2023, pp. 1–6.

[47] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, July 2009.

[48] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1753–1767, April 2015.

[49] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786–788, Nov. 1981.

[50] Y. Wang, V. W. Wong, and J. Wang, "Flexible rate-splitting multiple access with finite blocklength," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1398–1412, May 2023.

[51] Y. Xu, Y. Mao, O. Dizdar, and B. Clerckx, "Max-min fairness of rate-splitting multiple access with finite blocklength communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6816–6821, May 2023.

[52] ——, "Rate-splitting multiple access with finite blocklength for short-packet and low-latency downlink communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 12 333–12 337, Nov. 2022.

[53] X. Ou, X. Xie, H. Lu, and H. Yang, "Resource allocation in MU-MISO rate-splitting multiple access with SIC errors for URLLC services," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 229–243, Jan. 2023.

[54] T. Li, H. Zhang, S. Guo, and D. Yuan, "Robust beamforming and rate-splitting design for next generation ultra-reliable and low-latency communications," *arXiv preprint arXiv:2209.09138*, 2022.

[55] T.-H. Vu, T. Q. Duong, T.-H. Khuat, T. H. Nguyen, and S. Kim, "Short-packet communication for cooperative UAV-based MIMO systems with rate splitting multiple access," in *Int. Conf. Commun. & Electronics (ICCE)*. IEEE, 2022, pp. 113–117.

[56] Y. Liu, B. Clerckx, and P. Popovski, "Network slicing for eMBB, URLLC, and mMTC: An uplink rate-splitting multiple access approach," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[57] Y. Karacora and A. Sezgin, "Rate-splitting enabled multi-connectivity in mixed-criticality systems," *arXiv preprint arXiv:2211.12528*, 2022.

[58] S. K. Singh, K. Agrawal, K. Singh, Y.-M. Chen, and C.-P. Li, "Performance analysis and optimization of RSMA enabled UAV-aided IBL and FBL communication with imperfect SIC and CSI," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3714–3732, 2023.

[59] S. K. Singh, K. Agrawal, K. Singh, B. Clerckx, and C.-P. Li, "RSMA for hybrid RIS-UAV-aided full-duplex communications with finite blocklength codes under imperfect SIC," *IEEE Trans. Wireless Commun.*, 2023.

[60] 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, Oct. 2022.

[61] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Rate splitting multiple access for energy efficient RIS-aided multi-user short-packet communications," in *IEEE Global Commun. Workshops (GC Wkshps)*. IEEE, 2022, pp. 644–649.

[62] X. Li, T. Wang, H. Tong, Z. Yang, Y. Mao, and C. Yin, "Sum-rate maximization for active RIS-aided downlink RSMA system," *arXiv preprint arXiv:2301.12833*, 2023.

[63] 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 Wkshps)*. IEEE, 2020, pp. 1–6.

[64] D. Shambharkar, S. Dhok, and P. K. Sharma, "Performance analysis of ris assisted rsma communication system," in *Proc. IEEE National Conf. Commun. (NCC)*. IEEE, 2022, pp. 227–232.

[65] H. Li, S. Shen, and B. Clerckx, "Synergizing beyond diagonal reconfigurable intelligent surface and rate-splitting multiple access," *arXiv preprint arXiv:2303.06912*, 2023.

[66] 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, Oct. 2022.

[67] J. Scarlett, V. Y. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 81–92, Jan. 2017.

[68] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found Trends® in Commun. Inf. Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[69] M. Soleymani, I. Santamaria, and P. J. Schreier, "Improper Gaussian signaling for the K-user MIMO interference channels with hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 632–11 645, Oct. 2020.

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

[71] J. Kim, S. Hosseinalipour, A. C. Marcum, T. Kim, D. J. Love, and C. G. Brinton, "Learning-based adaptive IRS control with limited feedback codebooks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9566–9581, 2022.

[72] Z. Peng, Z. Chen, C. Pan, G. Zhou, and H. Ren, "Robust transmission design for RIS-aided communications with both transceiver hardware impairments and imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 528–532, March 2022.

[73] P. Yang, L. Yang, and S. Wang, "Performance analysis for RIS-aided wireless systems with imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 588–592, March 2022.

[74] S. Hu, Z. Wei, Y. Cai, C. Liu, D. W. K. Ng, and J. Yuan, "Robust and secure sum-rate maximization for multiuser MISO downlink systems with self-sustainable IRS," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7032–7049, 2021.

Mohammad Soleymani was born in Arak, Iran. He received the B.Sc. degree from Amirkabir University of Technology (Tehran Polytechnic), the M.Sc. degree from Sharif University of Technology, Tehran, Iran, and the Ph.D. degree (with distinction) from the University of Paderborn, Germany, all in electrical engineering. He is currently an assistant professor (Akademischer Rat) at the Signal and System Theory Group, University of Paderborn. He was a Visiting Researcher at the University of Cantabria, Spain. His research interests include multiuser communications, wireless networking, convex optimization and statistical signal processing.

Ignacio Santamaria (M'96-SM'05) received the Telecommunication Engineer degree and the Ph.D. degree in electrical engineering from the Universidad Politecnica de Madrid (UPM), Spain, in 1991 and 1995, respectively. In 1992, he joined the Department of Communications Engineering, Universidad de Cantabria, Spain, where he is Full Professor since 2007. He has co-authored more than 250 publications in refereed journals and international conference papers, and holds two patents. He has co-authored the book D. Ramirez, I. Santamaria, and L.L. Scharf, "Coherence in Signal Processing and Machine Learning", Springer, 2022. His current research interests include signal processing algorithms and information-theoretic aspects of multiuser multiantenna wireless communication systems, multivariate statistical techniques and machine learning theories. He has been involved in numerous national and international research projects on these topics. He has been a visiting researcher at the University of Florida (in 2000 and 2004), at the University of Texas at Austin (in 2009), and at the Colorado State University (in 2015 and 2017). He has been Associate Editor of the IEEE Transactions on Signal Processing (2011-2015), and Senior Area Editor of the IEEE Transactions on Signal Processing (2013-2015). He has been a member of the IEEE Machine Learning for Signal Processing Technical Committee (2009-2014), member of the IEEE Signal Processing Theory and Methods Technical Committee (2020-2022), and member of the IEEE Data Science Initiative (DSI) steering committee (2020-2022). Prof. Santamaria co-authored a paper that received the 2012 IEEE Signal Processing Society Young Author Best Paper Award, and has received the 2022 IHP International Wolfgang Mehr Fellowship Award.

Eduard A. Jorswieck (Fellow, IEEE) received the Ph.D. degree in computer engineering from TU Berlin, in 2004. From 2006 to 2008, he was a Postdoctoral Fellow and an Assistant Professor with the Signal Processing Group, KTH Stockholm. From 2008 to 2019, he was the Chair for Communication Theory with TU Dresden. He is currently the Managing Director of the Institute of Communications Technology and the Head of the Chair for Communications Systems and a Full Professor with Technische Universität Braunschweig, Germany. His research interest lies in the broad area of communications. He has published more than 170 journal articles, 15 book chapters, one book, three monographs, and more than 320 conference papers. He is a Fellow of the IEEE. He was a recipient of the IEEE Signal Processing Society Best Paper Award. Since 2017, he has been the Editor-in-Chief of the Springer EURASIP Journal on Wireless Communications and Networking. Since 2022, he has been on the editorial board of the IEEE TRANSACTIONS ON COMMUNICATIONS. He was on the editorial boards of the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

Bruno Clerckx (Fellow, IEEE) 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 Université 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", 250 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.