

# STAR-RIS-aided RSMA for the URLLC multi-user MIMO Downlink

Mohammad Soleymani<sup>1</sup>, Ignacio Santamaria<sup>2</sup>, Eduard Jorswieck<sup>3</sup>, Robert Schober<sup>4</sup>, Lajos Hanzo<sup>5</sup>

<sup>1</sup>University of Paderborn, Germany, <sup>2</sup>Universidad de Cantabria, Spain, <sup>3</sup> Technische Universität Braunschweig, Germany

<sup>4</sup>Friedrich Alexander University of Erlangen-Nuremberg, Germany, <sup>5</sup>University of Southampton, United Kingdom

Email: mohammad.soleymani@upb.de, santamal@unican.es, e.jorswieck@tu-bs.de, robert.schober@fau.de, lh@ecs.soton.ac.uk

**Abstract**—Rate splitting multiple access (RSMA) is intrinsically amalgamated with simultaneously transmitting and reflecting (STAR) reconfigurable intelligent surfaces (RIS) to enhance energy efficiency (EE) of the finite block length (FBL) multiple-input multiple-output (MIMO) downlink. An alternating optimization-based algorithm is proposed to jointly optimize the transmit beamforming matrices, STAR-RIS configurations, and rate-splitting parameters. STAR-RIS attains 360-degree full-plane coverage, while RSMA provides a prominent gain by efficiently managing interference. Numerical results reveal a strong synergy between RSMA and STAR-RIS, demonstrating significant EE gains over reflective RIS and spatial division multiple access (SDMA).

**Index Terms**—Fractional matrix programming, interference management, multi-user MIMO, RSMA, STAR-RIS, URLLC.

## I. INTRODUCTION

Next generation (NG) wireless communications will support a large number of devices with various services and improved sustainability in terms of energy and resource efficiency. The authors of [1] provide a comprehensive overview of the visions, requirements, key technologies, and testbeds on the road to 6G, emphasizing the role of advanced network architectures and intelligent communication systems. Gong et al. in [2] explore the theoretical foundations and enabling technologies of holographic multiple-input multiple-output (MIMO) communications, outlining future research directions for this promising 6G technology. Vaezi et al. in [3] offer a detailed survey of advancements in 5G for cellular, wide-area, and non-terrestrial Internet of Things (IoT), outlining the path toward 6G networks. The paper in [3] highlights the challenges and opportunities in integrating massive IoT connectivity, and low-latency communications for the NG Internet of Everything (IoE).

An attractive technology to face challenges of NG networks is the reconfigurable intelligent surface (RIS) [4], [5]. Wu et al. [4] offer a comprehensive tutorial on intelligent reflecting surface (IRS)-aided wireless communications, detailing the system models, signal processing techniques and their performance analysis. In contrast, the authors of [5] introduce the concept of smart radio environments empowered by RIS, discussing their operation, current research status and future challenges. Additionally, there are clear benefits of passive RIS in improving energy efficiency (EE) [6]–[11]. Lyu et al. in [6] explore hybrid RIS architectures in cell-free networks

to enhance EE, while [7] compares different RIS setups in multiple-input single-output (MISO) broadcast channels. Our former work in [8] investigates integrating rate-splitting with RIS in MIMO systems under hardware impairments. Furthermore, [9] analyzes the trade-off between spectral efficiency (SE) and EE in active RIS-assisted RSMA networks, and [10] contrasts active and nearly passive RIS designs under global reflection constraints to optimize EE in RIS-aided systems.

For practical low-latency communications, finite block-length (FBL) coding must be applied. The fundamental limits for ultra-reliable low-latency communications (URLLC) are established by analyzing channel coding rates within the FBL regime in [12]. By contrast, in [13], tight performance bounds and their asymptotic approximations are derived using Laplace integral techniques for short-packet transmission scenarios.

The benefits of RIS in enhancing the communication performance under FBL constraints, particularly for ultra-reliable low-latency communications (URLLC), are highlighted in [14]–[25]. RISs are shown to significantly improve the SE versus EE trade-off in dynamic propagation environments. Several works, such as [14], [16], and [18], investigate RIS-aided URLLC scenarios and demonstrate performance gains in short-packet transmissions through optimized beamforming and user grouping. Advanced architectures like simultaneously transmitting and reflecting (STAR)-RIS and aerial RISs are advocated in [20] and [23], respectively, for their ability to provide improved reliability and coverage. Moreover, the integration of rate-splitting multiple access (RSMA) with RIS, as examined in [21] and [22], is shown to enhance robustness against channel impairments and improve throughput under stringent latency and reliability requirements. These studies established RIS as a key enabler for achieving efficient and reliable communication in the FBL regime.

As a future advance, the authors of [26] surveyed RSMA, detailing its theoretical foundations, its advantages in managing multi-user interference, and its superiority compared to conventional multiple access techniques such as spatial division multiple access (SDMA) and non-orthogonal multiple access (NOMA). The paper also surveys recent developments and practical applications across various network scenarios and outlines key future research directions, establishing RSMA as a promising paradigm for NG networks.

The RSMA approach can also be applied in low-latency

communication situations using the FBL rate expressions that take coding length and error probabilities into account [27]. The max-min fairness of RSMA in FBL communication scenarios is investigated in [27], showing that RSMA significantly enhances both the rate fairness and reliability in short-packet transmissions, making it well-suited for URLLC applications in NG wireless networks.

In this paper, we demonstrate the advantages of STAR-RIS-aided RSMA in FBL scenarios, considering both fairness and EE. The concept of STAR-RIS was introduced in [28], [29] to improve the network performance for users on both sides of the RIS. The existing literature, including [25], address this question only partly. It is shown in [30]–[32] that the combination of STAR-RIS with user fairness for spatially separated user groups significantly improves performance. Furthermore, RSMA shows improvements in scenarios where the system is highly loaded and hence does not have sufficient degrees of freedom to mitigate interference. In [25], we proposed RSMA schemes to enhance SE and EE of MU-MIMO URLLC system, assisted by conventional reflective RIS. In this paper, we extend the results of [25] to STAR-RIS-assisted systems, revealing the beneficial synergy of RSMA and STAR-RIS. In particular, **our main contributions are:**

- We formulate the system model of a mode-switching-aided STAR-RIS-assisted MIMO broadcast channel (BC) and 1-layer rate splitting. The problem is to maximize the minimum individual EE under minimum rate constraints, sum transmit power constraints, and STAR-RIS mode-switching constraints.
- We propose an alternating optimization (AO) based algorithm for optimizing both the beamforming and STAR-RIS scattering matrices. The approach is based on the optimization framework proposed in [33] for fraction matrix programming problems, and we extend the solution proposed in [25] to STAR-RIS-aided systems.
- Finally, we answer the question raised above by showing that reliable and energy-efficient network operation requires both STAR-RIS and RSMA for maximizing EE fairness.

*Notations:* We represent scalar/vector/matrix by  $x/\mathbf{x}/\mathbf{X}$ . A zero-mean complex Gaussian vector  $\mathbf{x}$  with covariance matrix  $\mathbf{X}$  is denoted by  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{X})$ . Moreover,  $Q^{-1}$ ,  $\mathbf{I}$ ,  $\text{Tr}(\mathbf{X})$ ,  $|\mathbf{X}|$ , and  $\Re(x)$  denote the inverse Q-function for Gaussian signals, an identity matrix, the trace of  $\mathbf{X}$ , the determinant of  $\mathbf{X}$ , and the real part of  $x$ , respectively.

## II. SYSTEM MODEL AND PROBLEM STATEMENT

We consider a MIMO BC, where a base station (BS) equipped with  $N_{BS}$  transmit antennas (TA) aims to support  $K$  users, each equipped with  $N_u$  receive antennas (RA). To enhance the communication performance, a STAR-RIS with  $M$  elements is deployed to assist the BS, as shown in Fig. 1. Thus, the channel between the BS and user  $k$  is modeled as [34]

$$\mathbf{H}_k = \mathbf{D}_k \Theta^{r/t} \mathbf{D} + \mathbf{G}_k, \quad (1)$$

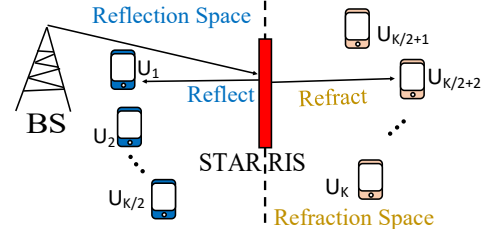


Fig. 1. Network model: STAR-RIS-assisted downlink transmission.

where  $\mathbf{G}_k$  represents the channels between the BS and user  $k$ ,  $\mathbf{D}_k$  models the channels between user  $k$  and the STAR-RIS, and  $\mathbf{D}$  is the channel between the BS and the STAR-RIS. The matrices  $\Theta^r = \text{diag}(\theta_1^r, \theta_2^r, \dots, \theta_M^r)$  and  $\Theta^t = \text{diag}(\theta_1^t, \theta_2^t, \dots, \theta_M^t)$  represent the scattering matrices of the STAR-RIS corresponding to the reflection and transmission (refraction) modes, respectively. If a user is located in the reflection (or transmission) half-space, its channel can only be optimized through  $\Theta^r$  (or  $\Theta^t$ ). We assume a nearly passive RIS architecture, where the amplitude constraints of the RIS elements have to satisfy  $|\theta_m^r|^2 + |\theta_m^t|^2 \leq 1$ , for all  $m$ . We consider a mode-switching scheme for STAR-RIS, as it improves the performance with lower complexities than the energy-splitting scheme. More specifically, we divide the STAR-RIS elements into two groups with an equal number of elements per group. The elements of the first group operate only in a transmission mode, while the elements of the second group operate in a reflection mode, i.e.,

$$\theta_m^r = 0, \quad |\theta_m^t|^2 \leq 1, \quad \text{if} \quad 1 \leq m \leq \frac{M}{2}, \quad (2a)$$

$$\theta_m^t = 0, \quad |\theta_m^r|^2 \leq 1, \quad \text{if} \quad \frac{M}{2} + 1 \leq m \leq M. \quad (2b)$$

### A. Signal Model

The BS leverages a 1-layer rate-splitting (RS) design, where the message of each user is divided into two parts: common and private. The common message is intended to be detected by all users, whereas the private messages are detected only by their respective intended users. More specifically, the signal transmitted from the BS is given by  $\mathbf{x} = \mathbf{W}_c \mathbf{t}_c + \sum_{k=1}^K \mathbf{W}_k \mathbf{t}_k$ , where  $\mathbf{t}_c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$  denotes the common data stream,  $\mathbf{W}_c$  is its beamforming matrix,  $\mathbf{t}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$  is the private data stream intended for user  $k$ , and  $\mathbf{W}_k$  is its corresponding beamforming matrix. Note that the common message aggregates the common parts of all users data. Moreover, we assume that the BS employs FBL coding for all transmitted data streams. The codeword lengths for the common and private messages are  $n_c$  and  $n_p$ , respectively.

Upon transmission of  $\mathbf{x}$ , the signal received by user  $k$  is  $\mathbf{y}_k = \mathbf{H}_k \mathbf{W}_c \mathbf{t}_c + \sum_{j=1}^K \mathbf{H}_k \mathbf{W}_j \mathbf{t}_j + \mathbf{n}_k$ , where  $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$  is the additive white Gaussian noise. User  $k$  first detects  $\mathbf{t}_c$ , treating all other signals as noise, which results in the maximum achievable rate of [12], [24]

$$r_{kc} = \log \left| \mathbf{I} + \left( \sigma^2 \mathbf{I} + \sum_i \Gamma_{ki} \right)^{-1} \Gamma_{kc} \right|$$

$$- \frac{Q^{-1}(\epsilon_c)}{\sqrt{n_c}} \sqrt{2\text{Tr} \left( \mathbf{\Gamma}_{kc} \left[ \sigma^2 \mathbf{I} + \mathbf{\Gamma}_{kc} + \sum_i \mathbf{\Gamma}_{ki} \right]^{-1} \right)}, \quad (3)$$

where  $\mathbf{\Gamma}_{ki} = \mathbf{H}_k \mathbf{W}_i \mathbf{W}_i^H \mathbf{H}_k^H$ ,  $\mathbf{\Gamma}_{kc} = \mathbf{H}_k \mathbf{W}_c \mathbf{W}_c^H \mathbf{H}_k^H$ , and  $\epsilon_c$  is the maximum tolerable detection error rate for  $\mathbf{t}_c$ . All users should be able to detect the common message, which is only possible when the transmission rate of  $\mathbf{t}_c$  satisfies  $r_c \leq \min_k r_{kc}$ .

After detecting  $\mathbf{t}_c$ , user  $k$  removes it from  $\mathbf{y}_k$  and then detects  $\mathbf{t}_k$ , treating all the remaining signals as noise. Therefore, the maximum achievable rate of  $\mathbf{t}_k$  is

$$r_{kp} = \log \left| \mathbf{I} + \left( \sigma^2 \mathbf{I} + \sum_{i \neq k} \mathbf{\Gamma}_{ki} \right)^{-1} \mathbf{\Gamma}_{kk} \right| - \frac{Q^{-1}(\epsilon_p)}{\sqrt{n_p}} \sqrt{2\text{Tr} \left( \mathbf{\Gamma}_{kk} \left[ \sigma^2 \mathbf{I} + \sum_i \mathbf{\Gamma}_{ki} \right]^{-1} \right)}, \quad (4)$$

where  $\epsilon_p$  is the maximum tolerable detection error rate for  $\mathbf{t}_k$ . Note that the total detection error rate of each user can be upper bounded as  $\epsilon_c + \epsilon_p$ , as shown in [22, Eq. (18)].

The rate of user  $k$  is the summation of  $r_{kp}$  and the portion of the rate it receives from detecting the common message, i.e.,  $r_k = r_{kp} + q_k$ , where  $\sum_k q_k = r_c$ . Finally, the EE of user  $k$  is [35]

$$e_k = \frac{r_k}{p_k(\mathbf{W}_c, \mathbf{W}_k)}, \quad (5)$$

where  $p_k(\{\mathbf{W}_c, \mathbf{W}_k\})$  is the power consumed to transmit data to user  $k$ , given by [25]

$$p_k(\mathbf{W}_c, \mathbf{W}_k) = P_C + \beta(\text{Tr}(\mathbf{W}_c \mathbf{W}_c^H)/K + \text{Tr}(\mathbf{W}_k \mathbf{W}_k^H)), \quad (6)$$

where  $\beta$  is the power efficiency of the transmitter at the BS, and  $P_C$  is the static power consumed to operate the network divided by  $K$ , given by [36].

### B. Problem Statement

We aim for maximizing the minimum EE of users, which can be formulated as

$$\max_{\mathbf{W}_c, \{\mathbf{W}_k\}_{\forall k}, \mathbf{q}, \mathbf{\Theta}^r, \mathbf{\Theta}^t} \min_k \frac{r_{pk} + q_k}{p_k(\mathbf{W}_c, \mathbf{W}_k)} \quad (7a)$$

$$\text{s.t.} \quad r_{pk} + q_k \geq \tilde{r}_k^{th}, \quad (7b)$$

$$\sum_k q_k \geq \min_k(r_{kc}), \quad (7c)$$

$$\text{Tr} \left( \mathbf{W}_c \mathbf{W}_c^H + \sum_{i=1}^K \mathbf{W}_i \mathbf{W}_i^H \right) \leq P, \quad (7d)$$

$$(1), \text{ and } (2), \quad (7e)$$

where  $P$  is the BS power budget, and  $\{\mathbf{q}\} = \{q_1, q_2, \dots, q_K\}$ . The optimization problem in (7) is non-convex. The objective function of (7) is a fractional function, and its optimization variables include matrices. Therefore, (7) is a fractional matrix programming problem. The constraints in (7d) and (7e) are convex. However, (7b) and (7c) are non-convex, since  $r_{pk}$  and  $r_{ck}$  are non-concave for all  $k$ .

## III. PROPOSED SOLUTION

We solve (7) by harnessing an iterative algorithm based on alternating optimization (AO) and the fractional matrix programming (FMP) solver proposed in [33]. We commence from a feasible initial point  $\mathbf{W}_c^{(1)}$ ,  $\{\mathbf{W}_k^{(1)}\}_{\forall k}$ ,  $\mathbf{\Theta}^{r^{(1)}}$  and  $\mathbf{\Theta}^{t^{(1)}}$ . In each iteration, we first optimize the beamforming matrices, when  $\mathbf{\Theta}^r$  and  $\mathbf{\Theta}^t$  are kept fixed. We then alternate the optimization variables and update  $\mathbf{\Theta}^r$  and  $\mathbf{\Theta}^t$ , while keeping the beamforming matrices fixed. In the following subsections, we provide the solutions for updating the beamforming matrices, STAR-RIS coefficients, and rate splitting parameters.

### A. Beamforming Optimization

In this subsection, we propose a new solution for updating the beamforming matrices, when the RIS coefficients are kept fixed in iteration  $j$ . In this case, (7) is simplified to the following FMP problem

$$\max_{\mathbf{W}_c, \{\mathbf{W}_k\}_{\forall k}, \mathbf{q}} \min_k \frac{r_{pk} + q_k}{p_k(\mathbf{W}_c, \mathbf{W}_k)} \quad (8a)$$

$$\text{s.t.} \quad (7b), (7c), (7d). \quad (8b)$$

To solve (8), we employ the FMP solver proposed in [33]. To this end, we first calculate concave lower bounds for  $r_{pk}$  and  $r_{ck}$  in the following lemma.

*Lemma 1 ([25]):* The following inequalities hold for all feasible  $\mathbf{W}_c, \{\mathbf{W}_k\}_{\forall k}$

$$r_{pk} \geq \tilde{r}_{pk}^{(j)} = 2\Re\{\text{Tr}(\mathbf{A}_k \mathbf{W}_k^H \bar{\mathbf{H}}_k^H)\} + 2 \sum_{i \neq k} \Re\{\text{Tr}(\mathbf{A}_{ki} \mathbf{W}_i^H \bar{\mathbf{H}}_k^H)\} + a_k - \text{Tr} \left( \mathbf{B}_{pk} \left( \sigma^2 \mathbf{I} + \sum_i \bar{\mathbf{H}}_k \mathbf{W}_i \mathbf{W}_i^H \bar{\mathbf{H}}_k^H \right) \right),$$

$$r_{ck} \geq \tilde{r}_{ck}^{(j)} = a_{ck} + 2\Re\{\text{Tr}(\mathbf{A}_{ck} \mathbf{W}_c^H \bar{\mathbf{H}}_k^H)\} + 2 \sum_i \Re\{\text{Tr}(\mathbf{A}_{cki} \mathbf{W}_i^H \bar{\mathbf{H}}_k^H)\} - \text{Tr} \left( \mathbf{B}_{ck} \left( \sigma^2 \mathbf{I} + \bar{\mathbf{H}}_k \mathbf{W}_c \mathbf{W}_c^H \bar{\mathbf{H}}_k^H + \sum_i \bar{\mathbf{H}}_k \mathbf{W}_i \mathbf{W}_i^H \bar{\mathbf{H}}_k^H \right) \right),$$

where  $\bar{\mathbf{H}}_k = \mathbf{H}_k (\mathbf{\Theta}^{r^{(j)}})^{-1} (\mathbf{\Theta}^{t^{(j)}})^{-1}$ , and we have:

$$a_k = \ln \left| \mathbf{I} + \left( \sigma^2 \mathbf{I} + \sum_{i \neq k} \bar{\mathbf{\Gamma}}_{ki} \right)^{-1} \bar{\mathbf{\Gamma}}_{kk} \right| - \text{Tr} \left( \left( \sigma^2 \mathbf{I} + \sum_{i \neq k} \bar{\mathbf{\Gamma}}_{ki} \right)^{-1} \bar{\mathbf{\Gamma}}_{kk} \right) - \frac{Q^{-1}(\epsilon_p)}{2\sqrt{n}} \left( \sqrt{v_k} + \frac{2I}{\sqrt{v_k}} \right),$$

$$a_{ck} = \ln \left| \mathbf{I} + \left( \sigma^2 \mathbf{I} + \sum_i \bar{\mathbf{\Gamma}}_{ki} \right)^{-1} \bar{\mathbf{\Gamma}}_{kc} \right|$$

$$\begin{aligned}
& -\text{Tr}\left(\left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \bar{\Gamma}_{kc} - \frac{Q^{-1}(\epsilon_p)}{2\sqrt{n}} \left(\sqrt{v_{ck}} + \frac{2I}{\sqrt{v_{ck}}}\right), \right. \\
& \mathbf{A}_k = \left(\sigma^2\mathbf{I} + \sum_{i \neq k} \bar{\Gamma}_{ki}\right)^{-1} \bar{\mathbf{H}}_k \bar{\mathbf{W}}_k, \\
& \mathbf{A}_{ck} = \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \bar{\mathbf{H}}_k \bar{\mathbf{W}}_c, \\
& \mathbf{A}_{ki} = \frac{Q^{-1}(\epsilon_p)}{\sqrt{nv_k}} \left(\sigma^2\mathbf{I} + \sum_n \bar{\Gamma}_{kn}\right)^{-1} \bar{\mathbf{H}}_k \bar{\mathbf{W}}_i, \\
& \mathbf{A}_{cki} = \frac{Q^{-1}(\epsilon_c)}{\sqrt{nv_{ck}}} \left(\sigma^2\mathbf{I} + \bar{\Gamma}_{ck} + \sum_n \bar{\Gamma}_{kn}\right)^{-1} \bar{\mathbf{H}}_k \bar{\mathbf{W}}_i, \\
& \mathbf{B}_k = \left(\sigma^2\mathbf{I} + \sum_{i \neq k} \bar{\Gamma}_{ki}\right)^{-1} - \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \\
& \quad + \frac{Q^{-1}(\epsilon_p)}{\sqrt{nv_k}} \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \left(\sigma^2\mathbf{I} + \sum_{i \neq k} \bar{\Gamma}_{ki}\right) \\
& \quad \times \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1}, \\
& \mathbf{B}_{ck} = \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} - \left(\sigma^2\mathbf{I} + \bar{\Gamma}_{kc} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \\
& \quad + \frac{Q^{-1}(\epsilon_p)}{\sqrt{nv_{ck}}} \left(\sigma^2\mathbf{I} + \bar{\Gamma}_{kc} + \sum_i \bar{\Gamma}_{ki}\right)^{-1} \left(\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki}\right) \\
& \quad \times \left(\sigma^2\mathbf{I} + \bar{\Gamma}_{kc} + \sum_i \bar{\Gamma}_{ki}\right)^{-1},
\end{aligned}$$

where  $I = \min(N_{BS}, N_u)$ ,  $\bar{\Gamma}_{ki} = \bar{\mathbf{H}}_k \bar{\mathbf{W}}_i \bar{\mathbf{W}}_i^H \bar{\mathbf{H}}_k^H$ ,  $\bar{\Gamma}_{kc} = \bar{\mathbf{H}}_k \bar{\mathbf{W}}_c \bar{\mathbf{W}}_c^H \bar{\mathbf{H}}_k^H$ ,  $\bar{\mathbf{W}}_c = \mathbf{W}_c^{(j)}$ ,  $\bar{\Gamma}_{ki} = \bar{\mathbf{H}}_k \bar{\mathbf{W}}_i \bar{\mathbf{W}}_i^H \bar{\mathbf{H}}_k^H$ ,  $v_k = 2\text{Tr}\left((\sigma^2\mathbf{I} + \sum_i \bar{\Gamma}_{ki})^{-1} \bar{\Gamma}_{kk}\right)$ ,  $v_{ck} = 2\text{Tr}\left((\sigma^2\mathbf{I} + \bar{\Gamma}_{kc} + \sum_i \bar{\Gamma}_{ki})^{-1} \bar{\Gamma}_{kc}\right)$ , and  $\bar{\mathbf{W}}_k = \mathbf{W}_k^{(j)}$  for all  $i, k$ .

Upon using [33, Lemma 2] and Lemma 1, we have

$$\max_{\mathbf{W}_c, \{\mathbf{W}_k\}_{\forall k}, \mathbf{q}, \mathbf{z}} \min_k \left\{ 2\eta_k^{(j)} z_k - \eta_k^{(j)^2} p_k(\mathbf{W}_c, \mathbf{W}_k) \right\} \quad (9a)$$

$$\text{s.t.} \quad \tilde{r}_{pk} + q_k \geq \tilde{r}_k^{th}, \quad (9b)$$

$$\sum_k q_k \geq \min_k(\tilde{r}_{kc}), \quad (9c)$$

$$\tilde{r}_{pk} + q_k - z_k^2 \geq 0, \quad (9d)$$

$$(7d), \quad (9e)$$

where  $\mathbf{z} = \{z_1, z_2, \dots, z_k\}$  is the set of auxiliary variables, and  $\eta_k^{(j)} = \frac{\sqrt{r_k(\mathbf{W}_c^{(j)}, \{\mathbf{W}_k^{(j)}\}_{\forall k})}}{p_k(\mathbf{W}_c^{(j)}, \mathbf{W}_k^{(j)})}$ . The updated beamforming matrices  $\mathbf{W}_c^{(j+1)}$  and  $\mathbf{W}_k^{(j+1)}$  are calculated for all  $k$  by solving (9).

## B. RIS Optimization

This subsection presents our solution to update  $\Theta^r$  and  $\Theta^t$  when  $\mathbf{W}_c$  and  $\mathbf{W}_k$  are fixed at  $\mathbf{W}_c^{(j+1)}$  and  $\mathbf{W}_k^{(j+1)}$ . In this case, the optimization problem in (7) is simplified to

$$\max_{\mathbf{q}, \Theta^r, \Theta^t} \min_k \frac{r_{pk} + q_k}{p_k(\mathbf{W}_c^{(j+1)}, \mathbf{W}_k^{(j+1)})} \quad (10a)$$

$$\text{s.t.} \quad (7b), (7c), (2). \quad (10b)$$

This problem is indeed a maximization of the minimum weighted rate, since the transmission power  $p_k(\mathbf{W}_c^{(j+1)}, \mathbf{W}_k^{(j+1)})$  is fixed when optimizing the STAR-RIS coefficients. To solve (10), we invoke an approach based on majorization minimization. More specifically, we first derive concave lower bounds for  $r_{kc}$  and  $r_{kp}$  to transform (10) to a convex problem. We state the lower bounds in the following lemma.

*Lemma 2 ([25]):* The following inequalities hold for all feasible  $\mathbf{H}_k$  for all  $k$

$$r_{pk} \geq \hat{r}_{pk}^{(j)} = 2\Re\{\text{Tr}(\mathbf{A}_k \bar{\mathbf{W}}_k^H \mathbf{H}_k^H)\} + 2\sum_{i \neq k} \Re\{\text{Tr}(\mathbf{A}_{ki} \bar{\mathbf{W}}_i^H \mathbf{H}_k^H)\}$$

$$+ a_k - \text{Tr}\left(\mathbf{B}_{pk} \left(\sigma^2\mathbf{I} + \sum_i \mathbf{H}_k \bar{\mathbf{W}}_i \bar{\mathbf{W}}_i^H \mathbf{H}_k^H\right)\right),$$

$$r_{ck} \geq \hat{r}_{ck}^{(j)} = a_{ck}$$

$$+ 2\Re\{\text{Tr}(\mathbf{A}_{ck} \bar{\mathbf{W}}_c^H \mathbf{H}_k^H)\} + 2\sum_i \Re\{\text{Tr}(\mathbf{A}_{cki} \bar{\mathbf{W}}_i^H \mathbf{H}_k^H)\} - \text{Tr}\left(\mathbf{B}_{ck} \left(\sigma^2\mathbf{I} + \mathbf{H}_k \bar{\mathbf{W}}_c \bar{\mathbf{W}}_c^H \mathbf{H}_k^H + \sum_i \mathbf{H}_k \bar{\mathbf{W}}_i \bar{\mathbf{W}}_i^H \mathbf{H}_k^H\right)\right),$$

where the parameters are defined in the same way as in Lemma 1, except for  $\bar{\mathbf{W}}_c = \mathbf{W}_c^{(j+1)}$  and  $\bar{\mathbf{W}}_k = \mathbf{W}_k^{(j+1)}$  for all  $k$ .

Upon leveraging Lemma 2, we can transform (10) into the following convex problem

$$\max_{\mathbf{q}, \Theta^r, \Theta^t} \min_k \frac{\hat{r}_{pk} + q_k}{p_k(\mathbf{W}_c^{(j+1)}, \mathbf{W}_k^{(j+1)})} \quad (11a)$$

$$\text{s.t.} \quad \hat{r}_{pk} + q_k \geq \tilde{r}_k^{th}, \quad (11b)$$

$$\sum_k q_k \geq \min_k(\hat{r}_{kc}), \quad (11c)$$

$$(1), \text{ and } (2). \quad (11d)$$

Since (11) is convex, it is solvable by numerical tools such as CVX. The solution of (11) gives updated STAR-RIS coefficients  $\Theta^{r(j+1)}$  and  $\Theta^{t(j+1)}$ . The overall algorithm converges to a stationary point of (7) [33].

## IV. NUMERICAL RESULTS

We carry out Monte Carlo simulations to characterize the proposed energy-efficient scheme. In the simulation setup, we consider line-of-sight (LOS) propagation for the channels  $\mathbf{D}$  and  $\mathbf{D}_k$  for all users  $k$ , resulting in Ricean small-scale fading modeled according to [37, (55)], with a Ricean  $K$ -factor of 3. By contrast, the direct channels  $\mathbf{G}_k$  are assumed to experience

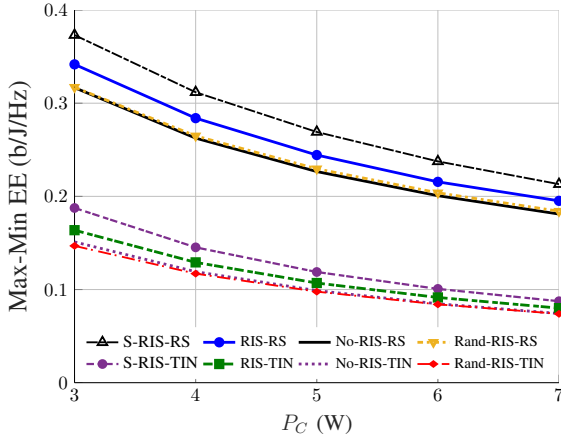


Fig. 2. Max-min EE over  $P_C$ .

non-line-of-sight (NLOS) conditions and are modeled with Rayleigh fading. Large-scale fading for all channel links is captured using the path-loss model from [36, (59)]. Additional propagation parameters, including antenna gains, bandwidth, noise power spectral density, path-loss exponent, and reference path-loss at 1 meter, are adopted from [33, Appendix G].

The simulation parameters are configured as follows: the number of antennas at both the base station and each user is  $N_{BS} = N_u = 2$ ; the number of users is  $K = 4$ ; the codeword lengths are set to  $n = n_c = n_k = 256$  bits; and the error probability thresholds are  $\epsilon = 2\epsilon_c = 2\epsilon_k = 10^{-5}$ . In terms of spatial deployment, users are evenly divided between the reflection and refraction regions of the STAR-RIS, as illustrated in Fig. 1. Consequently, a conventional reflective RIS would be unable to support half of the users. Again to overcome this limitation, we adopt a mode-switching operation for the STAR-RIS.

Fig. 2 illustrates the relationship between the maximum minimum EE (max-min EE) and the static power consumption  $P_C$  across different transmission schemes. The figure compares four variants of RSMA and treating interference as noise (TIN) schemes, each integrated with either a STAR-RIS, a reflective RIS, or no RIS, and also includes a baseline with randomly configured RIS elements. Note that the TIN/SDMA schemes are developed in [24]. The results clearly demonstrate that RSMA schemes consistently outperform their TIN counterparts in terms of max-min EE across all values of  $P_C$ , affirming RSMA effectiveness in enhancing EE. Moreover, the incorporation of STAR-RIS provides additional but modest gains over conventional reflective RIS configurations, while—as expected—randomly chosen RIS parameters yield suboptimal performance. This comparison highlights the importance of STAR-RIS and the compelling synergy between RSMA and STAR-RIS in achieving robust, energy-efficient communication under FBL constraints.

Fig. 3 demonstrates the average max-min EE as a function of  $n$ . The more stringent the latency requirement is, the shorter the codeword length becomes. Fig. 3 reveals that the max-min EE attained reduces, when more restrictive latency requirement

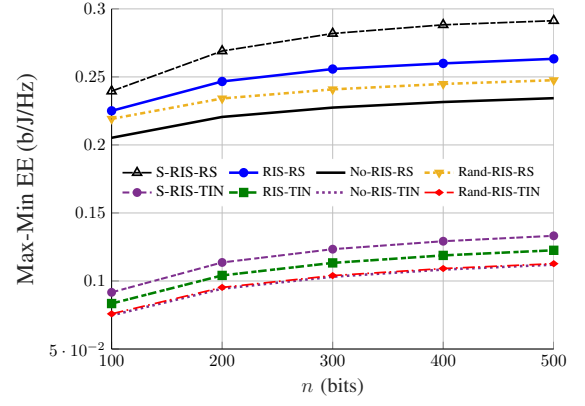


Fig. 3. Max-min EE over  $n$ .

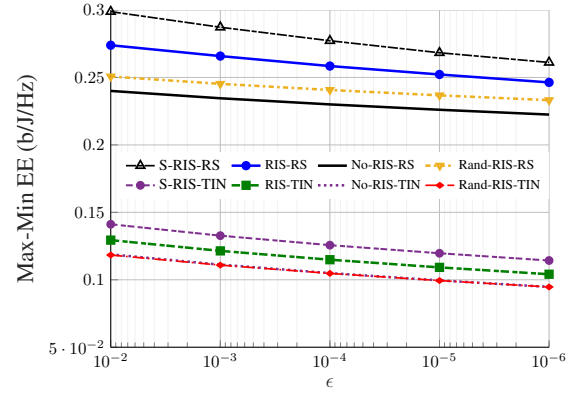


Fig. 4. Max-min EE over  $\epsilon$ .

should be fulfilled.

Fig. 4 illustrates the variation in average max-min EE as a function of  $\epsilon$ , highlighting the impact of reliability. As shown in Fig. 4, the max-min EE consistently decreases with smaller  $\epsilon$ , which correspond to more stringent reliability requirements. Each RSMA-based configuration outperforms its SDMA counterpart. Moreover, the use of STAR-RIS further improves EE, especially in RSMA schemes.

## V. CONCLUSIONS

Intrinsically integrating RSMA with STAR-RIS in MU-MIMO BCs significantly enhances EE in URLLC scenarios. By formulating and solving a max-min EE optimization problem using an AO and the FMP framework of [33], the proposed energy-efficient scheme effectively addresses the challenges posed by stringent latency and reliability requirements. Numerical results confirm the superiority of the joint RSMA and STAR-RIS approach over conventional schemes upon varying system parameters such as the static power, blocklengths, and error tolerances, highlighting its potential as a key enabler for energy-efficient and robust communication in next-generation wireless networks.

## ACKNOWLEDGMENT

I. Santamaria's work was funded by MCIN/AEI/10.13039/501100011033, under Grant PID2022-137099NBC43 (MAD-

DIE) and FEDER UE, and by European Union's (EU's) Horizon Europe project 6G-SENSES under Grant 101139282. E. Jorswieck's work was supported in part by the Federal Ministry of Education and Research (BMBF, Germany) through the Program of Souverän. Digital. Vernetzt. joint Project 6G-RIC, under Grant 16KISK031, and by European Union's (EU's) Horizon Europe project 6G-SENSES under Grant 101139282. Robert Schober's work was funded by the German Ministry for Education and Research (BMBF) under the program of "Souverän. Digital. Vernetzt." joint project 6G-RIC (Project-ID 16KISK023). L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) projects under grant EP/Y026721/1, EP/W032635/1, EP/Y037243/1 and EP/X04047X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

## REFERENCES

- [1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 2, pp. 905–974, 2023.
- [2] T. Gong *et al.*, "Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 1, pp. 196–257, 2024.
- [3] M. Vaezi *et al.*, "Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road towards 6G," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 2, pp. 1117–1174, 2022.
- [4] Q. Wu *et al.*, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.
- [5] 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.
- [6] W. Lyu *et al.*, "Energy-efficient cell-free network assisted by hybrid RISs," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 718–722, 2023.
- [7] M. Soleymani *et al.*, "Energy efficiency comparison of RIS architectures in MISO broadcast channels," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 701–705, 2024.
- [8] M. Soleymani *et al.*, "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.
- [9] H. Niu *et al.*, "Active RIS assisted rate-splitting multiple access network: Spectral and energy efficiency tradeoff," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1452–1467, 2023.
- [10] R. K. Fotock, A. Zappone, and M. Di Renzo, "Energy efficiency optimization in RIS-aided wireless networks: Active versus nearly-passive RIS with global reflection constraints," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 257–272, 2024.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] M. Soleymani *et al.*, "Rate region of RIS-aided URLLC broadcast channels: Diagonal versus beyond diagonal globally passive RIS," *IEEE Wireless Commun. Lett.*, vol. 14, no. 2, pp. 320–324, 2025.
- [15] Y. Li *et al.*, "Aerial reconfigurable intelligent surface-enabled URLLC UAV systems," *IEEE Access*, vol. 9, pp. 140248–140257, 2021.
- [16] 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.
- [17] M. Soleymani, I. Santamaria, E. Jorswieck, R. Schober, and L. Hanzo, "Energy efficiency optimization of finite block length STAR-RIS-aided MU-MIMO broadcast channels," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2025.
- [18] 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.
- [19] M. Almekhlafi, M. A. Arfaoui, M. Elhattab, C. Assi, and A. Ghayeb, "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.
- [20] M. Soleymani, I. Santamaria, and E. Jorswieck, "Spectral and energy efficiency maximization of MISO STAR-RIS-assisted URLLC systems," *IEEE Access*, vol. 11, pp. 70833–70852, 2023.
- [21] S. Pala, M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Spectral-efficient RIS-aided RSMA URLLC: Toward mobile broadband reliable low latency communication (mBRLC) system," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3507–3524, 2024.
- [22] M. Soleymani, I. Santamaria, E. Jorswieck, and B. Clerckx, "Optimization of rate-splitting multiple access in beyond diagonal RIS-assisted URLLC systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 5063–5078, 2024.
- [23] M. V. Katwe, R. Deshpande, K. Singh, M.-L. Ku, and B. Clerckx, "RSMA-enabled aerial RIS-aided MU-MIMO system for improved spectral-efficient URLLC," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 3110–3127, 2025.
- [24] M. Soleymani, I. Santamaria, E. Jorswieck, R. Schober, and L. Hanzo, "Optimization of the downlink spectral-and energy-efficiency of RIS-aided multi-user URLLC MIMO systems," *IEEE Trans. Commun.*, vol. 73, no. 5, pp. 3497–3513, 2025.
- [25] M. Soleymani, I. Santamaria, E. Jorswieck, M. Di Renzo, R. Schober, and L. Hanzo, "Rate splitting multiple access for RIS-aided URLLC MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, 2025.
- [26] 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.
- [27] 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, 2023.
- [28] H. Zhang and B. Di, "Intelligent omni-surfaces: Simultaneous refraction and reflection for full-dimensional wireless communications," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 1997–2028, 2022.
- [29] Y. Liu *et al.*, "STAR: Simultaneous transmission and reflection for 360 coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, 2021.
- [30] E. Jorswieck *et al.*, "URLLC networks enabled by STAR-RIS, rate splitting, and multiple antennas," in *Proc. IEEE Int. Conf. Mobile Miniaturized Terahertz Syst. (ICMITS)*, pp. 1–5, 2025.
- [31] M. Soleymani, I. Santamaria, and E. Jorswieck, "NOMA-based improper signaling for MIMO STAR-RIS-assisted broadcast channels with hardware impairments," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 673–678, IEEE, 2023.
- [32] M. Soleymani, I. Santamaria, and E. Jorswieck, "Energy-efficient rate splitting for MIMO STAR-RIS-assisted broadcast channels with IQ imbalance," *Proc. IEEE Eu. Signal Process. Conf. (EUSIPCO)*, 2023.
- [33] M. Soleymani, E. Jorswieck, R. Schober, and L. Hanzo, "A framework for fractional matrix programming problems with applications in FBL MU-MIMO," *IEEE Trans. Wireless Commun.*, 2025.
- [34] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, 2022.
- [35] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, 2016.
- [36] M. Soleymani, I. Santamaria, and P. J. Schreier, "Improper signaling for multicell MIMO RIS-assisted broadcast channels with IQ imbalance," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 723–738, 2022.
- [37] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.