

Efficient Codebook Design and User Scheduling for Large STAR-RIS-aided Downlink URLLC Transmission

Mostafa Darabi, Ata Khalili, Lutz Lampe, and Robert Schober

Abstract—To ensure ultra-reliable, low-latency communications (URLLC) despite blockages, we propose using simultaneously transmitting and reflecting (STAR) reconfigurable intelligent surfaces (RISs). These surfaces can transmit and reflect signals, expanding coverage on both sides of the RIS. However, phase shifter adjustments in STAR-RIS are limited by coupling constraints. We introduce a codebook design for phase-shift adjustment that accounts for these constraints. Leveraging this design, we jointly optimize base station beamforming and STAR-RIS phase-shifting to maximize the number of admitted URLLC users in a downlink multi-user communication system. Numerical results show that the proposed STAR-RIS codebook surpasses existing designs and effectively admits more URLLC users.

Index Terms—URLLC, reconfigurable intelligent surfaces, simultaneously transmitting and reflecting, 6G cellular networks.

I. INTRODUCTION

Traditional reconfigurable intelligent surfaces (RISs) reflect or transmit signals to only one side. To expand spatial coverage, recent work has introduced simultaneously transmitting and reflecting (STAR) RISs, enabling signals to be transmitted and reflected to both sides concurrently [1]. Previous works have utilized STAR-RISs to enhance the coverage area [2] and reduce power consumption [3] in downlink transmission systems. These studies assumed that phase-shift coefficients for transmission and reflection could be adjusted independently, permitting arbitrary complex electric and magnetic impedance values. However, this assumption overlooks the constraints on purely passive STAR-RISs, where the impedances are purely reactive, making the assumption impractical. This has been corrected in [4], where the authors proposed a practical coupled phase-shift model.

Passive RISs need to be sufficiently large to compensate for the extensive end-to-end path loss resulting from multiplying the individual path losses of the base station (BS)-to-RIS and the RIS-to-user channels. A large number of RIS elements results in greater signalling overhead to obtain channel state information (CSI) and higher computational complexity for phase-shift optimization. To address these issues, RISs can be divided into a few sub-arrays, known as tiles, with the phase shifts of each tile determined by a pre-designed vector. Such a vector of phase shifts is referred to as a codeword [5]. While codebook optimization has been explored for conventional RISs [5], [6], it has not yet been applied to STAR-RISs.

Massive ultra-reliable and low-latency communication (m-URLLC) is a use case in sixth-generation (6G) cellular

networks, combining low-latency requirements with massive connectivity. However, efficiently serving URLLC requests while admitting a large number of users is a challenging task. Adopting RISs can enhance the coverage area and increase the number of admitted URLLC users [7]. It is a natural question to ask what further benefits can be gained by adopting STAR-RISs to improve *service coverage* and *user admission* for URLLC.

In this letter, we propose using STAR-RIS to maximize the number of admitted URLLC users. We first present an optimized codebook design for the phase shifts of STAR-RISs with coupling constraints. Using this codebook, we jointly optimize STAR-RIS phase shifts and BS beamforming for URLLC user admission under quality-of-service (QoS) and BS power budget constraints. We present an iterative algorithm to efficiently solve the resulting non-convex program. We note that this is the first work to design an optimized phase-shift codebook under the practical assumption of coupled phase shifts and to optimize URLLC user admission in STAR-RIS-assisted networks. Simulation results demonstrate the effectiveness of our proposed codebook design in saving BS power and improving the admission rate for URLLC users compared to state-of-the-art linear and quadratic codebook designs.

II. SYSTEM MODEL

To study the advantages of STAR-RISs in a typical URLLC setup, we consider a single-cell downlink multi-user multiple-input single-output (MISO) system. The BS is equipped with M antennas and serves the URLLC users with the support of a STAR-RIS. The STAR-RIS operates in the *energy-splitting* regime [3], reflecting and transmitting the incident signal simultaneously, and employs *uni-cast* transmission. The direct links between the BS and the URLLC users are assumed to be heavily blocked. User admission for a set of K URLLC users is performed, where K_1 users are positioned in front of the RIS, and K_2 users are located at the back of the RIS.

A. STAR-RIS Structure

We first briefly describe the STAR-RIS structure and its components. We employ codebooks to reduce the CSI estimation overhead and computation time associated with the RIS phase shifts. A set of phase-shift configurations required for illuminating a desired area is referred to as a codebook, where each phase-shift configuration is a codeword. We assume that the RIS consists of Q tiles, each equipped with a codebook containing L codewords. Each tile consists of two surfaces, one for reflection and the other for transmission, with N_x elements along the x-axis and N_y elements along the y-axis, spaced d_x and d_y apart, respectively. Consequently, each

M. Darabi and L. Lampe are with the University of British Columbia, Canada (email: mostafadarabi, lampe@ece.ubc.ca). A. Khalili and R. Schober are with the Institute for Digital Communications, Friedrich-Alexander-University, Germany (email: ata.khalili, robert.schober@fau.de). This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

tile includes a total of $N = 2N_x N_y$ elements. The phase shifts applied by the (n_x, n_y) -th reflection and transmission elements of the q -th RIS tile are denoted by $\omega_{n_x, n_y}^{q,r}$ and $\omega_{n_x, n_y}^{q,t}$, respectively. Moreover, the reflection and transmission phase-shift matrices for the q -th tile are represented by Θ_q^r and Θ_q^t , respectively. These matrices can be expressed as $\Theta_q^m = a_{q,m} \sum_{l=1}^L b_{q,l} \text{diag}(\phi_{q,l}^m)$, $m \in \{r, t\}$, where $\text{diag}(\cdot)$ creates a diagonal matrix from the elements of a given vector and vector $\phi_{q,l}^m \in \mathbb{C}^{\frac{N}{2} \times 1}$ contains the reflection/transmission phase shifts $\omega_{n_x, n_y}^{q,m}$ associated with the l -th codeword of the q -th tile. The binary indicator $b_{q,l} \in \{0, 1\}$ determines the selection of the l -th codeword for the q -th tile. The factors $a_{q,r}$ and $a_{q,t} \in \mathbb{R}_{\geq 0}$ are the amplitudes of the reflected and transmitted signals for the q -th tile, respectively. These amplitudes are not determined by a codeword but through online optimization. Assuming a passive and lossless STAR-RIS, the amplitudes and phase shifts of the (n_x, n_y) -th elements of both sides of the q -th tile are coupled by [4]

$$a_{q,r}^2 + a_{q,t}^2 = 1, \quad \forall q, \quad (1)$$

$$\cos(\omega_{n_x, n_y}^{q,r} - \omega_{n_x, n_y}^{q,t}) = 0, \quad \forall q, n_x, n_y. \quad (2)$$

B. Signal Model

The channels between the BS and the q -th RIS tile and the channel between the q -th RIS tile and the k -th URLLC user are denoted by $\mathbf{G}_q \in \mathbb{C}^{\frac{N}{2} \times M}$ and $\mathbf{h}_{q,k} \in \mathbb{C}^{\frac{N}{2}}$, respectively. The digital precoding vector for the message from the BS to the k -th URLLC user is denoted by $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$. Then, the received signal-to-interference-plus-noise ratio (SINR) for the k -th URLLC user in the front/back of the RIS is given by $\gamma_k = \frac{|\mathbf{f}_k^H \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^K |\mathbf{f}_k^H \mathbf{w}_i|^2 + \sigma_k^2}$, where $\mathbf{f}_k^H = \sum_{q=1}^Q \mathbf{h}_{q,k}^H \Theta_q^m \mathbf{G}_q$ and σ_k^2 is the variance of the receiver noise at the k -th URLLC user. Adopting short packet communication for URLLC applications, the achievable rate for the k -th URLLC user in the front/back of STAR-RIS can be written as [8]

$$R_k = \log_2(1 + \gamma_k) - \log_2(e) \sqrt{\frac{V(\gamma_k)}{S}} Q^{-1}(\epsilon), \quad (3)$$

where ϵ is the decoding packet error probability, S is the blocklength of the URLLC streams, $Q^{-1}(\cdot)$ denotes the inverse Gaussian Q-function, and $V(\gamma_k) = 1 - (1 + \gamma_k)^{-2}$.

The packet arrival for the k -th URLLC user is assumed to follow a Poisson point process with parameter λ_k . Each URLLC packet has a length of X_k bits. For the k -th URLLC user, Λ_k is the number of packets arriving in a mini slot of duration Δt . Therefore, Λ_k is a Poisson random variable with mean $\lambda_k \Delta t$. In order to preserve the QoS for the k -th URLLC user, the outage probability in a mini slot is upper bounded by ν as $\Pr(\frac{X_k \Lambda_k}{\Delta t} > R_k) \leq \nu$, i.e., $R_k \geq \frac{X_k}{\Delta t} F_{\Lambda_k}^{-1}(1 - \nu)$, where $F_{\Lambda_k}^{-1}(\cdot)$ is the inverse cumulative distribution function of the packet arrival of the k -th URLLC user.

III. CODEBOOK DESIGN

In this section, we present a codebook design tailored for energy-splitting STAR-RISs with coupled phase shifts which allows for reflection and transmission of signals from any angle of arrival (AoA) towards a specified angle of departure

(AoD). We follow a similar approach as in [6] for the case of a conventional RIS and aim to maximize STAR-RIS response functions. We formulate the reflection and transmission response functions for the q -th STAR-RIS tile as [5]

$$g_q^m(\Psi_i, \Psi_r) = \bar{g} a_{q,m} \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} e^{j\bar{k}d_x A_x^m(\Psi_i, \Psi_m) n_x} e^{j\bar{k}d_y A_y^m(\Psi_i, \Psi_m) n_y} e^{j\omega_{n_x, n_y}^{q,m}}, \quad (4)$$

where $\bar{g} = \frac{4\pi d_x d_y}{\lambda^2}$, λ denotes the wavelength, and \bar{k} is the angular wavenumber. Moreover, $A_x^m(\Psi_i, \Psi_m) = \sin(\vartheta_i) \cos(\varphi_i) + \sin(\vartheta_m) \cos(\varphi_m)$ and $A_y^m(\Psi_i, \Psi_m) = \sin(\vartheta_i) \sin(\varphi_i) + \sin(\vartheta_m) \sin(\varphi_m)$, where $\Psi_i = (\vartheta_i, \varphi_i)$, $\Psi_r = (\vartheta_r, \varphi_r)$, and $\Psi_t = (\vartheta_t, \varphi_t)$ represent the AoA of the incident wave and the AoD of the reflected and transmitted waves, respectively, with elevation angle ϑ and azimuth angle φ . For notational simplicity, we define $v_{n_x, n_y}^{q,m} = e^{j\omega_{n_x, n_y}^{q,m}}$.

One way to design a codebook is to discretize Ψ_i , Ψ_r , and Ψ_t into intervals, and for each interval of Ψ_i , design a dedicated codeword to reflect the signal towards specific intervals of Ψ_r and Ψ_t . However, the response function in (4) depends on Ψ_i , Ψ_r , and Ψ_t solely through A_x^m and A_y^m , $m \in \{r, t\}$. Therefore, discretization of A_x^m and A_y^m is preferable instead. The values of A_x^m and A_y^m fall within the range from -2 to 2 and can be uniformly quantized into L_x^m and L_y^m intervals in x - and y -directions, respectively. Thus, each tile has a codebook which comprises $L = L_x^r L_y^r L_x^t L_y^t$ codewords, and each codeword is designed for a specific interval in the x - and y -directions, represented by $\mathcal{B}_{x,l_x}^m, \forall l_x \in \{1, \dots, L_x\}$, and $\mathcal{B}_{y,l_y}^m, \forall l_y \in \{1, \dots, L_y\}$, $m \in \{r, t\}$, respectively.

To extend the coverage area, we aim to maximize the minimum amplitude of the tile response function on both sides of the STAR-RIS in the codeword design. We optimize the phase shifts of the codewords and leave the amplitude adjustment to online optimization (see Section IV). All tiles employ the same codebook. Thus, without loss of generality, we simplify the reflection and transmission response functions $g_q^m(\Psi_i, \Psi_r)$ and phase shifts $v_{n_x, n_y}^{q,m}$ by dropping the index q . Accordingly, we formulate the following optimization problem to design the codeword for intervals $\mathcal{B}_{x,l_x}^m, \forall l_x \in \{1, \dots, L_x\}$, and $\mathcal{B}_{y,l_y}^m, \forall l_y \in \{1, \dots, L_y\}$, $m \in \{r, t\}$:

$$\begin{aligned} \mathcal{P}_1 : & \text{maximize } \alpha, \\ & \alpha, \mathbf{v}_{l_x, l_y}^r, \mathbf{v}_{l_x, l_y}^t \\ \text{s.t. C1: } & |g^m(\Psi_i, \Psi_r)|^2 \geq \alpha, \forall A_x^m \in \hat{\mathcal{B}}_{x,l_x}^m, A_y^m \in \hat{\mathcal{B}}_{y,l_y}^m, m \in \{r, t\}, \\ \text{C2: } & \cos(\angle v_{n_x, n_y}^r - \angle v_{n_x, n_y}^t) = 0, \forall n_x, n_y, \\ \text{C3: } & |v_{n_x, n_y}^r| = |v_{n_x, n_y}^t| = 1, \forall n_x, n_y, \end{aligned} \quad (5)$$

where \mathbf{v}_{l_x, l_y}^m is the vector containing all $v_{n_x, n_y}^m, \forall n_x, n_y$, for codeword for intervals \mathcal{B}_{x,l_x}^m and \mathcal{B}_{y,l_y}^m . Here, \angle denotes the phase of a complex number. To facilitate the optimization, we discretize the continuous intervals \mathcal{B}_{x,l_x}^m and \mathcal{B}_{y,l_y}^m into P_x^m and P_y^m discrete points, respectively. These points are then collected into the two discrete sets $\hat{\mathcal{B}}_{x,l_x}^m$ and $\hat{\mathcal{B}}_{y,l_y}^m$. Constraint C2 ensures the coupling between the phase shifts of the reflected and transmitted signals, and C3 guarantees that the optimization variables have unit magnitude. We note that the optimization in [6] for the tile response function of a traditional

RIS does not include the constraint caused by the coupled phase shifts.

\mathcal{P}_1 is non-smooth, non-convex and further complicated by the phase-shift coupling constraint C2. An effective approach to address an optimization problem with non-linearly coupled optimization variables within non-convex constraints is the penalty dual decomposition (PDD) approach. In this method, the coupling constraints are incorporated into the objective as penalty terms, resulting in an augmented Lagrangian problem, which can be efficiently solved through iterative methods [9]. Following the PDD framework proposed in [10], we introduce auxiliary variables $\bar{v}_{n_x, n_y}^r = v_{n_x, n_y}^r$ and $\bar{v}_{n_x, n_y}^t = v_{n_x, n_y}^t$ and partition the set of optimization variables into $\{\alpha, \mathbf{v}_{l_x, l_y}^r, \mathbf{v}_{l_x, l_y}^t\}$ and $\{\bar{\mathbf{v}}_{l_x, l_y}^r, \bar{\mathbf{v}}_{l_x, l_y}^t\}$, where $\bar{\mathbf{v}}_{l_x, l_y}^r$ and $\bar{\mathbf{v}}_{l_x, l_y}^t$ are vectors containing all \bar{v}_{n_x, n_y}^r and $\bar{v}_{n_x, n_y}^t, \forall n_x, n_y$, respectively. This partitioning enables an iterative approach in which the following two subproblems are solved alternately:

$$\mathcal{P}_2 : \underset{\alpha, \mathbf{v}_{l_x, l_y}^r, \mathbf{v}_{l_x, l_y}^t}{\text{maximize}} \quad \alpha - \frac{1}{2\rho} \sum_{\substack{n_x=0, n_y=0 \\ m \in \{r, t\}}}^{N_x-1, N_y-1} \|\bar{v}_{n_x, n_y}^m - v_{n_x, n_y}^m + \rho \lambda_{n_x, n_y}^m\|^2, \\ \text{s.t. C1}, \quad (6)$$

and for each n_x and n_y , we solve

$$\mathcal{P}_3 : \underset{\bar{v}_{n_x, n_y}^r, \bar{v}_{n_x, n_y}^t}{\text{minimize}} \quad \frac{1}{2\rho} (\|\bar{v}_{n_x, n_y}^r - v_{n_x, n_y}^r + \rho \lambda_{n_x, n_y}^r\|^2 \\ + \|\bar{v}_{n_x, n_y}^t - v_{n_x, n_y}^t + \rho \lambda_{n_x, n_y}^t\|^2), \\ \text{s.t. } \widetilde{\text{C2}} : \cos(\angle \bar{v}_{n_x, n_y}^r - \angle \bar{v}_{n_x, n_y}^t) = 0, \quad \forall n_x, n_y, \\ \widetilde{\text{C3}} : |\bar{v}_{n_x, n_y}^r| = |\bar{v}_{n_x, n_y}^t| = 1, \quad \forall n_x, n_y, \quad (7)$$

where ρ and $\lambda_{n_x, n_y}^m, m \in \{r, t\}$, are the penalty factor and Lagrangian dual variables, respectively.

The optimization problem in (6) can be solved using the semi-definite programming and difference of convex (DC) methods similar to [6]. For the solution of the optimization problem in (7), we define $\nu_{n_x, n_y}^r = -v_{n_x, n_y}^r + \rho \lambda_{n_x, n_y}^r$ and $\nu_{n_x, n_y}^t = -v_{n_x, n_y}^t + \rho \lambda_{n_x, n_y}^t$ and rewrite (7) as

$$\mathcal{P}_4 : \underset{\bar{v}_{n_x, n_y}^r, \bar{v}_{n_x, n_y}^t}{\text{minimize}} \quad \text{Re}((\nu_{n_x, n_y}^r)^* \bar{v}_{n_x, n_y}^r) + \text{Re}((\nu_{n_x, n_y}^t)^* \bar{v}_{n_x, n_y}^t) \\ \text{s.t. } \widetilde{\text{C2}}, \widetilde{\text{C3}}, \quad (8)$$

where $\text{Re}(\cdot)$ is the real part of a complex number. $\widetilde{\text{C2}}$ is equivalent to $v_{n_x, n_y}^r = \pm j v_{n_x, n_y}^t$. By substituting this constraint into the objective function and performing some manipulations, the optimal solution to \mathcal{P}_4 can be derived as [10]

$$\underset{(\bar{v}_{n_x, n_y}^r, \bar{v}_{n_x, n_y}^t) \in \chi_{n_x, n_y}}{\text{argmin}} \quad \text{Re}((\nu_{n_x, n_y}^r)^* \bar{v}_{n_x, n_y}^r) + \text{Re}((\nu_{n_x, n_y}^t)^* \bar{v}_{n_x, n_y}^t), \quad (9)$$

where χ_{n_x, n_y} is a set defined as

$$\chi_{n_x, n_y} = \left\{ \left(e^{j(\pi - \angle \tau_{n_x, n_y}^+)} e^{j(\frac{3}{2}\pi - \angle \tau_{n_x, n_y}^+)}, \right. \right. \\ \left. \left. \left(e^{j(\pi - \angle \tau_{n_x, n_y}^-)} e^{j(\frac{3}{2}\pi - \angle \tau_{n_x, n_y}^-)} \right) \right\}, \quad (10)$$

and $\tau_{n_x, n_y}^\pm = (\nu_{n_x, n_y}^t)^* \pm j (\nu_{n_x, n_y}^r)^*$.

Algorithm 1 summarizes the proposed codeword design for given intervals $\hat{\mathcal{B}}_{x, l_x}^m$ and $\hat{\mathcal{B}}_{y, l_y}^m$ in the x - and y -directions.

Algorithm 1 Proposed optimization-based codeword design

- 1: **Initialization:** Set up intervals $\hat{\mathcal{B}}_{x, l_x}^m$ and $\hat{\mathcal{B}}_{y, l_y}^m$, initialize $\alpha, v_{n_x, n_y}^r, v_{n_x, n_y}^t, \bar{v}_{n_x, n_y}^r, \bar{v}_{n_x, n_y}^t$, and $0 < c < 1$
 - 2: **Repeat**
 - 3: **Repeat**
 - 4: update $\{\alpha, \mathbf{v}_{l_x, l_y}^r, \mathbf{v}_{l_x, l_y}^t\}$ by solving \mathcal{P}_2
 - 5: update $\{\bar{v}_{n_x, n_y}^r, \bar{v}_{n_x, n_y}^t\}$ by (9) for each n_x and n_y
 - 6: **Until** convergence
 - 7: Set $\lambda_{n_x, n_y}^m = \lambda_{n_x, n_y}^m + \frac{1}{\rho} (\bar{v}_{n_x, n_y}^m - v_{n_x, n_y}^m)$, $m \in \{r, t\}$ and $\rho = c\rho$
 - 8: **Until** $\varepsilon = \max_{m \in \{r, t\}} (\sum_{n_y=0}^{N_y-1} \sum_{n_x=0}^{N_x-1} |v_{n_x, n_y}^m - \bar{v}_{n_x, n_y}^m|)$ drops under a pre-determined limit
 - 9: **Return** v_{n_x, n_y}^r and $v_{n_x, n_y}^t, \forall n_x, n_y$
-

By employing the DC method for solving \mathcal{P}_2 in (6) and optimally solving \mathcal{P}_3 in (7), Algorithm 1 iteratively generates a sequence of improved feasible points until convergence to a locally optimal solution [6]. The computational complexities of updating $\{\alpha, \mathbf{v}_{l_x, l_y}^r, \mathbf{v}_{l_x, l_y}^t\}$ in (6) and $\{\bar{\mathbf{v}}_{l_x, l_y}^r, \bar{\mathbf{v}}_{l_x, l_y}^t\}$ in (7) are $\mathcal{O}([P_x^r + P_x^t + P_y^r + P_y^t]N^3)$ and $\mathcal{O}(2N) + \mathcal{O}(4N)$, respectively, where $\mathcal{O}(\cdot)$ is the big-O notation [6], [10].

IV. URLLC USER ADMISSION SCHEME

We employ the proposed codebook design from Section III to maximize the number of admitted URLLC users for the given QoS requirements and BS power budget. This is achieved through the following optimization problem:

$$\mathcal{P}_5 : \underset{\mathbf{a}_r, \mathbf{a}_t, \mathbf{B}, \mathbf{W}, \mathbf{u}}{\text{maximize}} \quad \sum_{k=1}^K \mu_k u_k \quad (11) \\ \text{s.t. C1: } R_k \geq u_k R_k^{\min}, \quad \forall k, \\ \text{C2: } \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_{\max}, \quad \text{C3: } a_{q,r}^2 + a_{q,t}^2 = 1, \quad \forall q, \\ \text{C4: } 0 \leq a_{q,r}, a_{q,t} \leq 1, \quad \forall q, \quad \text{C5: } b_{q,l} \in \{0, 1\}, \quad \forall q, l, \\ \text{C6: } \sum_{l=1}^L b_{q,l} = 1, \quad \forall q, \quad \text{C7: } u_k \in \{0, 1\}, \quad \forall k,$$

where $R_k^{\min} = \frac{X_k}{\Delta t} F_{\Lambda_k}^{-1}(1 - \nu)$, $\mathbf{a}_r = [a_{1,r}, a_{2,r}, \dots, a_{Q,r}]^T$, $\mathbf{a}_t = [a_{1,t}, a_{2,t}, \dots, a_{Q,t}]^T$, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_Q]$, $\mathbf{b}_q = [b_{q,1}, b_{q,2}, \dots, b_{q,L}]^T$, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$, and $\mathbf{u} = [u_1, u_2, \dots, u_K]^T$. u_k is a binary admission variable and is equal to one if the k -th URLLC user is admitted. Moreover, μ_k is adopted as a continuous coefficient in the objective function to maintain fairness among URLLC users. C1 ensures the satisfaction of the QoS for the URLLC users. C2 limits the transmission power of the BS to P_{\max} , and C3 and C4 are imposed to control the amplitude of the STAR-RIS. Finally, C5 – C6 guarantee that each tile uses one codeword, and C7 renders the user selection variable binary.

Due to the integer variables of user admission and codebook selection, the coupling between the optimization variables, and the fractional rate function in C1, the optimization problem in (11) is a mixed integer non-convex problem. Similar to [8], we employ an alternating optimization (AO) approach

and divide the optimization variables into subsets $\{\mathbf{W}, \mathbf{u}\}$ and $\{\mathbf{a}_r, \mathbf{a}_t, \mathbf{B}, \mathbf{u}\}$ optimized alternately as follows.

1) *Optimization of Beamforming at the BS and User Selection*: We relax the integer variables u_k to continuous variables and substitute C7 with the equivalent constraints

$$\text{C7a: } 0 \leq u_k \leq 1, \forall k, \quad \text{C7b: } \sum_{k=1}^K u_k - u_k^2 \leq 0.$$

Constraint C7a is in the form of DC functions and can be convexified via DC programming. Applying the Taylor series approximation (TSA) at points $u_k^{(i_1)}$, C7b can be rewritten as

$$\text{C7c: } \sum_{k=1}^K u_k - (u_k^{(i_1)})^2 - 2u_k^{(i_1)}(u_k - u_k^{(i_1)}) \leq 0,$$

where i_1 is the iteration index for successive convex approximation (SCA).

Since the rate expression in C1 is a non-convex function, we introduce the auxiliary variables Γ_k and β_k as a lower bound on the SINR and an upper bound on the received interference and noise for k -th URLLC user, respectively. Thus, defining $D = \frac{Q^{-1}(\epsilon)}{\sqrt{S}} \log_2(e)$, C1 can be replaced by

$$\text{C1a: } \log_2(1 + \Gamma_k) - D\sqrt{V(\Gamma_k)} \geq u_k R_k^{\min}, \forall k,$$

$$\text{C1b: } \frac{|\mathbf{f}_k^H \mathbf{w}_k|^2}{\beta_k} \geq \Gamma_k, \forall k, \quad \text{C1c: } \sum_{\substack{i=1, \\ i \neq k}}^K |\mathbf{f}_k^H \mathbf{w}_i|^2 + \sigma_k^2 \leq \beta_k, \forall k.$$

C1a is in the form of DC functions and can be convexified by employing the TSA of $\sqrt{V(\Gamma_k)}$ as

$$\widetilde{\text{C1a:}} \log_2(1 + \Gamma_k) - D \left(\left[1 - (1 + \Gamma_k^{(i_1)})^{-2} \right]^{-\frac{1}{2}} \times \left[(1 + \Gamma_k^{(i_1)})^{-3} (\Gamma_k - \Gamma_k^{(i_1)}) - (1 + \Gamma_k^{(i_1)})^{-2} + 1 \right] \right) \geq u_k R_k^{\min}, \forall k.$$

Furthermore, the left-hand side of C1b can be approximated around the point $(\mathbf{w}_k^{(i_1)}, \beta_k^{(i_1)})$ using a convex surrogate function:

$$\widetilde{\text{C1b:}} \frac{2\text{Re} \left((\mathbf{w}_k^{(i_1)})^H \mathbf{f}_k \mathbf{f}_k^H \mathbf{w}_k \right)}{\beta_k^{(i_1)}} - \frac{|\mathbf{f}_k^H \mathbf{w}_k^{(i_1)}|^2 \beta_k}{(\beta_k^{(i_1)})^2} \geq \Gamma_k.$$

Hence, the first sub-problem can be rewritten as

$$\begin{aligned} \mathcal{P}_6 : \text{maximize} \quad & \sum_{k=1}^K \mu_k u_k \\ \text{w.r.t. } \mathbf{W}, \mathbf{u}, \Gamma_k, \beta_k & \\ \text{s.t. } \widetilde{\text{C1a}}, \widetilde{\text{C1b}}, \text{C1c}, \text{C2}, \text{C7a}, \text{C7c}. & \end{aligned} \quad (12)$$

2) *Optimization of STAR-RIS Design and User Selection*:

We relax the integer variables $b_{q,l}$ to continuous variables and substitute C5 with the equivalent constraints

$$\text{C5a: } 0 \leq b_{q,l} \leq 1, \forall q, l, \quad \text{C5b: } \sum_{q=1}^Q \sum_{l=1}^L b_{q,l} - b_{q,l}^2 \leq 0.$$

Constraint C5b is in the form of a DC function and can be convexified as

$$\text{C5c: } \sum_{q=1}^Q \sum_{l=1}^L b_{q,l} - (b_{q,l}^{(i_2)})^2 - 2b_{q,l}^{(i_2)}(b_{q,l} - b_{q,l}^{(i_2)}) \leq 0,$$

where i_2 is the SCA iteration index. Furthermore, we deal with the coupling between the variables of $\mathbf{a}_m, m \in \{r, t\}$ and \mathbf{B} using the big-M formulation technique. Since $b_{q,l}$ is a binary variable and $a_{q,r}$ and $a_{q,t}$ are between 0 and 1, we can introduce the new variables $c_{q,l}^r = a_{q,r} b_{q,l}, \forall q, l, c_{q,l}^t = a_{q,t} b_{q,l}, \forall q, l$, and add the linear constraints

$$\text{C8: } 0 \leq c_{q,l}^r \leq b_{q,l}, \quad \text{C9: } c_{q,l}^r \leq a_{q,r},$$

$$\text{C10: } c_{q,l}^r \geq a_{q,r} - (1 - b_{q,l}), \quad \text{C11: } 0 \leq c_{q,l}^t \leq b_{q,l},$$

$$\text{C12: } c_{q,l}^t \leq a_{q,t}, \quad \text{C13: } c_{q,l}^t \geq a_{q,t} - (1 - b_{q,l}),$$

for all q and l . Then, the second sub-problem is written as

$$\begin{aligned} \mathcal{P}_7 : \text{maximize} \quad & \sum_{k=1}^K \mu_k u_k \\ \text{w.r.t. } \mathbf{a}_r, \mathbf{a}_t, \mathbf{B}, \mathbf{u}, \mathbf{C}_r, \mathbf{C}_t & \\ \text{s.t. } \widetilde{\text{C3}}, \text{C4}, \text{C5a}, \text{C5c}, \text{C6}, \text{C8} - \text{C13}, & \end{aligned} \quad (13)$$

where $\mathbf{C}_r = [c_1^r, \dots, c_L^r]$, $\mathbf{C}_t = [c_1^t, \dots, c_L^t]$, $\mathbf{c}_r^t = [c_{1,l}^r, \dots, c_{Q,l}^r]^T$ and $\mathbf{c}_t^t = [c_{1,l}^t, \dots, c_{Q,l}^t]^T$.

Both \mathcal{P}_6 and \mathcal{P}_7 are convex problems and can be solved optimally. Since the cost functions in (12) and (13) are monotonically non-decreasing in each iteration, and the total number of admissible URLLC users is upper bounded by K , the proposed AO approach converges to a stationary solution of \mathcal{P}_5 . The computational complexity of solving (12) and (13) are $\mathcal{O}([KM]^4[2K + (K + 1)M])$ and $\mathcal{O}([QL + Q^2L^2 + 2KQ]^4[K(1 + 2Q) + Q(1 + L) + 4Q^2L^2])$, respectively.

V. SIMULATION RESULTS

In this section, we assess the effectiveness of the URLLC user admission scheme using STAR-RIS with the new codebook design. We consider three codebook designs available in the literature for comparison: the optimized codebook design introduced in [6], and the linear and quadratic codebook designs proposed in [5]. The design in [6] is for conventional RISs and thus can only be applied to the idealized case of a STAR-RIS without coupling constraints, i.e., the phase-shift coupling constraints at the STAR-RIS are ignored. In the linear and quadratic codebooks, the phase shifts follow a linear or quadratic function of the RIS element index. As a result, these codebooks can be used for STAR-RIS with coupling constraints. For concreteness, we consider a scenario where the BS is equipped with 32 antennas and serves 40 single-antenna URLLC users. 20 URLLC users are located in front of the RIS, and 20 URLLC users are located behind it, all randomly placed within a 100-meter cell radius. To ensure fairness among the URLLC users, we select μ_k in (11) according to the channel quality and average arrival rate as $\mu_k \propto \frac{\lambda_k}{\log_2(h_k/\sigma_k^2)}$, where $h_k = \sum_{q=1}^Q \|\mathbf{h}_{k,q}\|_2^2$ and $\|\cdot\|_2$ denotes the L_2 -norm. We model the path loss as $35.3 + 37.6 \log_{10}(d)$ dB, where d denotes the distance between the BS and the RIS or between the RIS and a URLLC user. The carrier frequency and the total bandwidth are 5 GHz and 1.2 MHz, respectively. The noise power spectral density is set to -174 dBm/Hz. The channel fading coefficients between the BS and the RIS tiles as well as those between the RIS tiles and the users are modelled as independent and identically distributed Rician random variables, each with a Rician factor

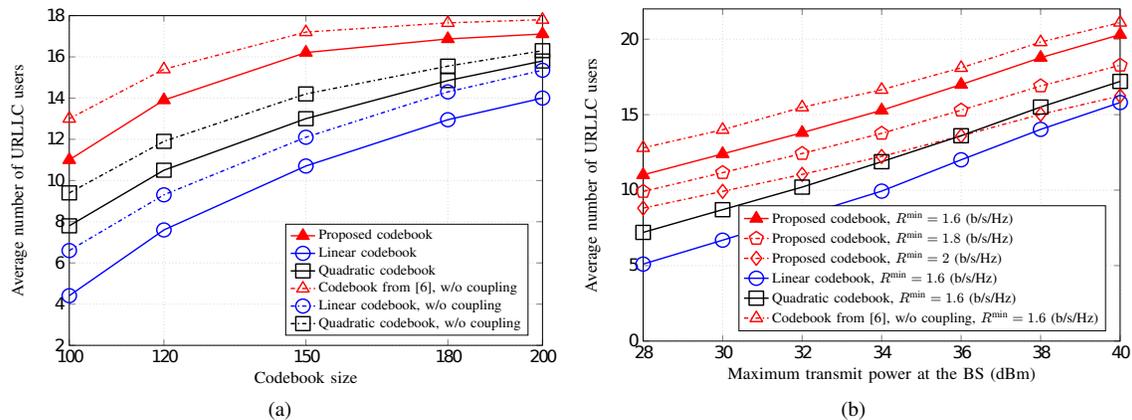


Fig. 1. (a) Average number of admitted URLLC users versus codebook size for a maximum transmit power of 35 dBm at the BS. (b) Average number of admitted URLLC users versus maximum transmit power at the BS and different QoS levels for URLLC users.

of 3 dB. Additionally, the direct links between the BS and the URLLC users are assumed to be obstructed. The target packet error probability of the URLLC users in (3) and the outage probability are set to 10^{-6} and 10^{-5} , respectively. The STAR-RIS has 3600 elements on each side, spaced at half a wavelength, and is divided into 9 tiles.

Fig. 1(a) depicts the average number of admitted URLLC users versus the codebook size, averaged over various channel realizations and user locations, for a power budget of 35 dBm at the BS. Let us first focus on the practical case of STAR-RIS with coupling constraints. We observe that, for the proposed codebook design, more URLLC users can be served compared to the linear and quadratic designs, especially for smaller codebook sizes. This superiority stems from the maximization of the reflected/transmitted power in the design intervals \mathcal{B}_x^m and \mathcal{B}_y^m , which is absent for the linear and quadratic codebook designs. Furthermore, as the codebook size increases, the sizes of the design intervals in the x - and y -directions decrease. For these smaller intervals, also the linear and quadratic codebook designs can generate narrow beams that cover the intervals well, resulting in smaller variations in the average number of admitted URLLC users across different codebook designs. In the idealized case of STAR-RIS without coupling constraints (labelled as “w/o coupling”), more users can be admitted. Without these constraints, the design space for codewords is larger, allowing for improved phase-shift configurations. Conversely, the coupling constraints reduce the size of the feasible set for codeword optimization, leading to a performance loss.

Fig. 1(b) shows the average number of admitted URLLC users as a function of the transmit power budget at the BS, and for three different QoS levels. The codebook contains 150 codewords, with $L_x^t = 5$, $L_y^t = 3$, $L_x^r = 5$, and $L_y^r = 2$. Again, we observe a significant improvement in the number of admitted users as a result of the proposed codebook design. In particular, our codebook design can support 15 URLLC users with a QoS of 1.6 bits/s/Hz, requiring 51% and 38% less BS transmit power compared to the linear and quadratic codebook designs, respectively. Naturally, increasing the QoS level for URLLC users from 1.6 bits/s/Hz to 2 bits/s/Hz leads to a decrease in the average number of admitted URLLC users.

This is due to the stricter constraint in C1 of the optimization problem in (11). The baseline scheme in [6], which ignores the STAR-RIS phase-shift coupling, serves as an upper bound. The proposed codebook design can approach this performance bound fairly closely while accounting for practical coupling constraints, as can be seen for the QoS level of 1.6 bits/s/Hz.

VI. CONCLUSION

We proposed the first optimization-based codebook design for the phase-shift configuration of large STAR-RISs. The new design was necessary due to the coupling constraint between the phase shifts on both sides of the STAR-RIS. We incorporated the codebook approach in a user admission scheme for a STAR-RIS-aided URLLC system. Simulation results showed that the proposed codebook outperforms baseline schemes using linear and quadratic codebook designs and closely approaches an idealized performance bound.

REFERENCES

- [1] H. Zhang et al., “Intelligent Omni-Surfaces for Full-Dimensional Wireless Communications: Principles, Technology, and Implementation,” *IEEE Commun. Magazine*, vol. 60, no. 2, pp. 39–45, Feb. 2022.
- [2] C. He et al., “Capacity Maximization for Active RIS Assisted Outdoor-to-Indoor Communication System,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 1–5, 2023.
- [3] X. Mu et al., “Simultaneously Transmitting and Reflecting (STAR) RIS Aided Wireless Communications,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.
- [4] J. Xu et al., “STAR-RIS: A Correlated T&R Phase-Shift Model and Practical Phase-Shift Configuration Strategies,” *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 5, pp. 1097–1111, Aug. 2022.
- [5] V. Jamali et al., “Power Efficiency, Overhead, and Complexity Tradeoff of IRS Codebook Design—Quadratic Phase-Shift Profile,” *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2048–2052, 2021.
- [6] W. Ghanem et al., “Optimization-Based Phase-Shift Codebook Design for Large IRSs,” *IEEE Commun. Lett.*, vol. 72, pp. 635–639, Feb. 2023.
- [7] E. M. Taghavi et al., “Joint Active-Passive Beamforming and User Association in IRS-Assisted mmWave Cellular Networks,” *IEEE Trans. Vehicular Tech.*, vol. 72, no. 8, pp. 10448–10461, Aug. 2023.
- [8] M. Darabi et al., “Active IRS Design for RSMA-based Downlink URLLC Transmission,” in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, pp. 1–6, 2023.
- [9] Q. Shi et al., “Penalty Dual Decomposition Method for Nonsmooth Nonconvex Optimization—Part I: Algorithms and Convergence Analysis,” *IEEE Trans. Signal Process.*, vol. 68, pp. 4108–4122, Feb. 2020.
- [10] Z. Wang et al., “Coupled Phase-Shift STAR-RIS: A General Optimization Framework,” *IEEE Commun. Lett.*, vol. 12, no. 2, pp. 20–24, 2023.