Power-efficient Joint Resource Allocation and Decoding Error Probability for Multiuser Downlink MISO with Finite Block Length Codes

Jalal Jalali[§], Atefeh Rezaei[†], Ata Khalili[‡], and Jeroen Famaey[§]
[§]IDLab, University of Antwerp - imec, Sint-Pietersvliet 7, 2000 Antwerp, Belgium
[†]Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran 14115-111, Iran
[‡]Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg, Erlangen, Germany
Email: jalal.jalali@uantwerpen.be

***Abstract*—In this paper, we study a multiuser multiple-input single-output (MISO) ultra-reliable low-latency (URLLC)-enabled intelligent reflecting surface (IRS) system, where a multi-antenna access point (AP) transmits information symbols to a set of URLLC users by taking into account short packet transmission and low-latency wireless communication. In particular, the total system transmits power is minimized by jointly optimizing active and passive beamformers at the AP and the IRS, respectively. An efficient algorithm based on alternating optimization (AO) is proposed to solve the main optimization problem iteratively. Firstly, we adopt the difference of convex functions (DC) and successive convex approximation (SCA) to obtain a sub-optimal solution for the active beamformers at the AP. Secondly, a penalty-based approach is adopted together with the SCA technique to handle the unit-modulus constraints at the IRS. Moreover, an explicit objective is proposed to provide a better convergence. Simulation results exhibit the performance of the proposed algorithm compared to other baseline schemes.**

***Index Terms*—Ultra-reliable low-latency communication (URLLC), intelligent reflecting surface (IRS), alternating optimization (AO).**

I. Introduction

Intelligent reflecting surface (IRS) has recently attracted considerable research interest to enrich both the spectral- and energy-efficiency of future networks with its simplistic deployments [1]. Specifically, as a planar metasurface with an extensive array of passive reflecting elements, IRS is able to control the radio propagation environment by dynamically adjusting the amplitudes and phases of incident signals. The IRS typically operates in a full-duplex mode without increasing noise amplification and demanding active radio-frequency (RF) chains for signal transmission/reception and self-interference cancellation, making it a cost-effective candidate for beyond fifth-generation (B5G) communication systems [2]. The active beamformers at the base station (BS) and passive beamformers at the IRS can alleviate the maximization of the spectral efficiency (SE) and the weighted sum data rate of the network, as was shown in [3]. To achieve green communication, the wireless power transfer and simultaneous information and power transfer (SWIPT) and IRS can also help maximize the network's energy efficiency (EE) indicator by jointly optimizing the phase shifts at the IRS and active beamformers at the transmitter [4].

On the other hand, ultra-reliable low-latency communication (URLLC) has been identified as one of the critical issues in the B5G wireless communication systems to realize short packet transmission and low-latency wireless communication, especially for sensitive applications such as health-care, autonomous driving, and tactical Internet [5], [6]. However, the conventional Shannon capacity formula can not be adopted for the URLLC systems under the short package regime [7]. In [8], a global optimal resource allocation for a URLLC system was obtained, where the bandwidth, power allocation, and antenna arrangement were optimized to minimize the weighted sum of downlink (DL) and uplink average power consumption. The authors in [9] designed the active BS beamforming vectors to maximize the sum data rate performance of a multiple-input single-output (MISO) orthogonal frequency division multiple access (OFDMA)-URLLC system.

Additionally, an IRS platform could benefit the URLLC users [10], [11] in confounding the latency dilemma. In [10], an IRS-aided OFDMA-URLLC system was considered to maximize the weighted sum data rate by jointly optimizing the active beamforming vectors and phase shifts at the BS and the IRS, respectively. [11] studied an IRS-aided mobile edge computing system to minimize the latency by jointly optimizing the edge computing resources, computation offloading, and beamforming matrices at both BS and the IRS. The average decoding error probability (DEP) and achievable data rate of an IRS-aided low-latency systems were analyzed in [12]. In [13], authors exploited the user grouping concept to minimize the total latency in IRS-enabled networks with URLLC users.

To the best of our knowledge, the gain of deploying the IRS platform in URLLC, considering both the DEP and average traffic load, has not been investigated in the literature. Furthermore, since IRS boosts the network's quality of service (QoS), it is inspiring to consider a high data rate URLLC service for more complicated use cases, such as a high signal-to-interference-plus-noise ratio (SINR) region. We consider the resource allocation algorithm design for a DL MISO URLLC-enabled IRS system to address the aforementioned issues. Our system model uses a multi-antenna access point (AP) to serve

This work was supported by the CHIST-ERA grant SAMBAS (CHIST-ERA-20-SICT-003), with funding from FWO, ANR, NKFIH, and UKRI.

multiple single-antenna URLLC receivers using a smart reconfigurable reflector. Consequently, we focus on minimizing the total transmission power in the proposed system, which provides valuable insights for the system design. This paper presents the following main contributions:

- We aim at minimizing the total transmit power of the system by jointly optimizing active and passive beamformers at the AP and IRS, respectively, and DEP subject to the minimum required data rate for each URLLC user. In particular, the traffic of the URLLC users with finite blocklength data rate is modeled as a chance constraint which enables the network to assigns proper resources to meet the aggregate traffic load.
- We exploit an alternating optimization (AO) resource allocation algorithm to solve the formulated optimization problem iteratively. For the active beamformers at the AP, we first define a lower bound of the SINR and then apply the difference of convex functions (DC) and successive convex approximation (SCA) technique to obtain a sub-optimal solution. Second, a penalty-based approach is adopted along with the SCA technique to handle the unit-modulus constraints at the IRS. In addition, to provide a better convergence, an explicit objective is proposed to design more efficient phase shifts.
- The simulation results reveal that deploying an IRS and a multi-antenna AP can increase the performance gain of the URLLC systems in terms of low latency and high reliability. Besides, results show that the IRS is more energy-efficient than the furnishing multiple antennas at the AP.

Notations: Matrices and vectors are denoted by boldface capital letters $\mathbf{A}$ and lower case letters $\mathbf{a}$, respectively. For a square matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^H$, Rank($\mathbf{A}$), Tr($\mathbf{A}$), and $||\mathbf{A}||_*$ are transpose, Hermitian conjugate transpose, rank of a matrix, trace, norm of a matrix, respectively. $\mathbf{A} \succeq \mathbf{0}$ shows a positive semidefinite matrix. $\mathbf{I}_N$ denotes the N-by-N identity matrix. diag$(\cdot)$ is the diagonalization operation. Diag($\mathbf{A}$) indicates a vector whose elements are extracted from the main diagonal elements of matrix $\mathbf{A}$. The absolute value of a complex scalar, and the Euclidean norm of a complex vector are expressed by $|\cdot|$ and $\|\cdot\|$, respectively. $\sim$ $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes the distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. The largest eigenvalue of matrix $\mathbf{X}$ is denoted by $\lambda_{\max}(\mathbf{X})$. $Q^{-1}(\cdot)$ stands for the inverse of the Gaussian Q-function. Pr$(X > a)$ denotes the probability that the random variable X assumes a particular value strictly greater than a. Finally, $\mathbb{C}^{M\times N}$ represents an $M \times N$ dimensional complex matrix and $\nabla_{\mathbf{x}}$ expresses the gradient vector with respect to $\mathbf{x}$.

II. System Model

We consider a DL MISO system with N-element IRS, M-antenna AP, and K single-antenna users with the set of $\mathcal{K} = \{1, ..., K\}$ as shown in Fig. 1. Assume that L_k information bits are assigned to user k, the AP encodes these information bits into a block code with the length of m_d (symbols) which is given by $z_k[l], l \in \{1, 2, ..., m_d\}$. Subsequently, the transmit signal at the AP can be expressed as $\mathbf{s}[l] = \sum_{k\in\mathcal{K}} \mathbf{w}_k z_k[l]$, where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ represents the beamforming vector for user k. Denote the baseband equivalent channel responses [1] from the AP-to-IRS, IRS-to-user k, and AP-to-user k as $\mathbf{H} \in \mathbb{C}^{N\times M}$, $\mathbf{h}_k^{\text{IU}} \in \mathbb{C}^{N\times 1}$, and $\mathbf{h}_k^{\text{AU}} \in \mathbb{C}^{M\times 1}$, respectively. Also, define $\boldsymbol{\Theta} = \text{diag}(\beta_1 e^{j\alpha_1}, \beta_2 e^{j\alpha_2}, ..., \beta_N e^{j\alpha_N})$ as the reflection-coefficients matrix at the IRS, where $\beta_n \in [0, 1]$ and $\alpha_n \in (0, 2\pi]$, $\forall n \in \{1, ..., N\}$ are the reflection amplitude[2] and phase shift of the n-th reflection coefficient at the IRS, respectively. By defining $\mathbf{h}_k^H \triangleq (\mathbf{h}_k^{\text{IU}})^H \boldsymbol{\Theta}\mathbf{H} + (\mathbf{h}_k^{\text{AU}})^H$, $\forall k$, as the equivalent channel link, the received signal at user k can be written as:

$$y_k[l] = \mathbf{h}_k^H \mathbf{s}[l] + n_k[l] \triangleq \sum_{k\in\mathcal{K}} \mathbf{h}_k^H \mathbf{w}_k z_k[l] + n_k[l], \forall k, l, \quad (1)$$

where $n_k[l] \sim \mathcal{CN}(0, \sigma_k^2)$ is the noise received at user k with variance σ_k^2. Then, the SINR at user k can be expressed as:

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \mathbf{w}_k\right|^2}{\sum\limits_{i\neq k, i\in K} \left|\mathbf{h}_k^H \mathbf{w}_i\right|^2 + \sigma_k^2}, \forall k. \quad (2)$$

It should be noted that in URLLC systems, the data blocks must be finite and have a short length to guarantee low-latency and high-reliability wireless communication. The precise approximation for the achievable data rate of each user is given by [7]:

$$R_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta}) = F_k(\mathbf{w}_k, \boldsymbol{\Theta}) - G_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta}), \quad (3)$$

where

$$F_k(\mathbf{w}_k, \boldsymbol{\Theta}) = \log_2(1 + \gamma_k), \forall k, \quad (4)$$

$$G_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta}) = Q^{-1}(\epsilon_k)\sqrt{\frac{1}{m_d} V_k}, \forall k. \quad (5)$$

In addition, ϵ_k is the decoding error, m_d indicates the blocklength, and V_k denotes the channel dispersion which is given by $V_k = a^2\big(1-(1+\gamma_k)^{-2}\big)$, where $a = \log_2(e)$. Generally, the traffic load of URLLC user k can be described as a random variable like $L_k = \nu_k \Omega_k$, where ν_k and Ω_k are the packet size and the packet arrival rate, respectively. To guarantee the QoS of user k, the probability that the traffic load exceeds the allocated total data rate for user k should be less than or equal to the maximum tolerable probability of failure, ζ, on supporting the traffic load [14]. Consequently, this constraint can be written as:

$$\Pr\Big(L_k > R_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta})\Big) \leq \zeta,\ 0 < \zeta < 1,\ \forall k. \quad (6)$$

In particular, for each user k, we consider the packet size ν_k which is constant and the packet arrival rate Ω_k that follows the Poisson distribution with parameter Λ_k. Let $F_{\Omega_k}(\cdot)$ denote the cumulative distribution function (CDF) of the

[1]It is assumed that the channel state information (CSI) and the delay requirements of all users are perfectly known at the AP (see [2]–[4], [9]).

[2]For reflection efficiency maximization, the amplitudes of all passive elements are assumed to be one [2] i.e., $\beta_n = 1$, $\forall n$.

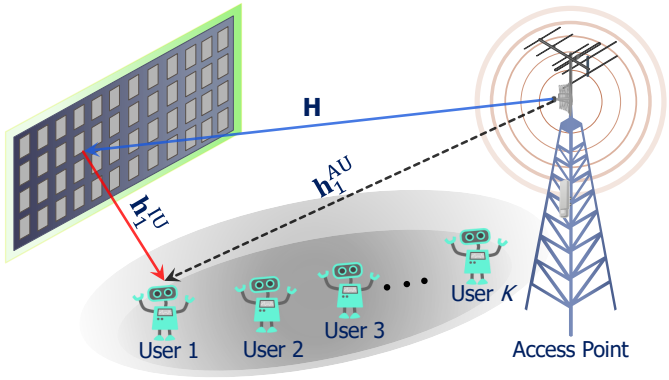

Fig. 1. A multiuser DL MISO URLLC-enabled IRS system.

packet arrival rate for user k. By performing some algebraic manipulations, we can express inequality (6) as:

$$R_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta}) \geq \nu_k F_{\Omega_k}^{-1}(1-\zeta), \forall k, \tag{7}$$

where $F_{\Omega_k}^{-1}$ is the inverse of CDF F_{Ω_k}.

III. Problem Formulation

In this section, we aim at minimizing the total transmit power of the considered system by jointly optimizing the active beamformers at the AP, the phase shifts at the IRS, and DEP. Accordingly, the problem of minimizing the total transmit power of the system can be mathematically formulated as:

$$\begin{aligned}
(\text{P1}): \ & \underset{\mathbf{w}_k, \boldsymbol{\Theta}, \epsilon_k}{\text{minimize}} \ \sum_{k\in\mathcal{K}} \|\mathbf{w}_k\|^2 && \text{(8a)}\\
\text{s.t.}: \ & R_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta}) \geq \nu_k F_{\Omega_k}^{-1}(1-\zeta), \forall k, && \text{(8b)}\\
& |\boldsymbol{\Theta}_{nn}| = 1, \qquad \forall n, && \text{(8c)}\\
& \epsilon_k \leq \epsilon_{k,\max}, \qquad \forall k, && \text{(8d)}
\end{aligned}$$

where (8b) is the minimum rate requirement of user k and (8c) insures that the diagonal phase shift matrix has N unit-modulus elements on its main diagonal. Finally, (8d) guarantees the reliability of each URLLC user, where $\epsilon_{\max}$ denotes the maximum error rate.

IV. Proposed Solution

(P1) is a non-convex optimization problem due to the highly coupled optimization variables. In general, there is no well-organized method to solve (P1). However, we propose an alternating optimization (AO) with low computational complexity to achieve a sub-optimal solution. In the first sub-problem, the successive convex approximation (SCA) technique and difference of convex functions (DC) are applied to design the active beamformers at the AP. While in the second one, the phase shifts at the IRS are optimized by exploiting the penalty approach and SCA technique. Also, a new objective function was proposed to avoid the feasibility problem.

A. First-stage: Optimizing $\mathbf{w}_k$ and ϵ_k with Fixed $\boldsymbol{\Theta}$

At this stage, we assume that the passive reflecting elements at the IRS, i.e., $\boldsymbol{\Theta}$ are fixed to design the active beamformers, $\mathbf{w}_k$, at the AP and DEP, ϵ_k. By adopting semidefinite programming (SDP), we have $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ and $\mathbf{H}_k = \mathbf{h}_k\mathbf{h}_k^H$, $\forall k$. In the following, (P1) can be rewritten as:

$$\begin{aligned}
(\text{P2}): \ & \underset{\mathbf{W}_k, \epsilon_k}{\text{minimize}} \ \sum_{k\in\mathcal{K}} \text{Tr}(\mathbf{W}_k) && \text{(9a)}\\
\text{s.t.}: \ & F_k(\mathbf{W}_k) - G_k(\mathbf{W}_k, \epsilon_k) \geq \nu_k F_{\Omega_k}^{-1}(1-\zeta), \ \forall k, && \text{(9b)}\\
& \text{Rank}(\mathbf{W}_k) \leq 1, \forall k, , && \text{(9c)}\\
& \mathbf{W}_k \succeq \mathbf{0}, \qquad \forall k, && \text{(9d)}\\
& \text{(8d)},
\end{aligned}$$

where γ_k in $F_k(\mathbf{W}_k)$ and $G_k(\mathbf{W}_k)$ can be expressed as:

$$\gamma_k = \frac{\text{Tr}(\mathbf{H}_k\mathbf{W}_k)}{\sum\limits_{i\in K, i\neq k} \text{Tr}(\mathbf{H}_k\mathbf{W}_i) + \sigma_k^2}, \forall k. \tag{10}$$

It is notable that constraint (9b) in (P2) is not concave. To deal with it, we propose a set of auxiliary variables μ_k, $\forall k$, to provide a lower bound of the SINR. Accordingly, we can write the SINR in (10) as below:

$$0 \leq \mu_k \leq \gamma_k = \frac{f_k(\mathbf{W}_k)}{g_k(\mathbf{W}_k)}, \forall k, \tag{11}$$

where the nominator and denominator of (11) can be expressed as:

$$f_k(\mathbf{W}_k) = \text{Tr}(\mathbf{H}_k\mathbf{W}_k), \tag{12}$$

$$g_k(\mathbf{W}_k) = \sum_{i\in K, i\neq k} \text{Tr}(\mathbf{H}_k\mathbf{W}_i) + \sigma_k^2, \tag{13}$$

respectively. Finally, by exploiting the lower bound in (11), the optimization problem in the first stage can be restated as:

$$\begin{aligned}
(\text{P3}): \ & \underset{\mathbf{W}_k, \mu_k, \epsilon_k}{\text{minimize}} \ \sum_{k\in\mathcal{K}} \text{Tr}(\mathbf{W}_k) && \text{(14a)}\\
\text{s.t.}: \ & \mu_k \geq 0, \qquad \forall k, && \text{(14b)}\\
& \mu_k \leq \frac{f_k(\mathbf{W}_k)}{g_k(\mathbf{W}_k)}, \forall k, && \text{(14c)}\\
& R_k(\epsilon_k, \mu_k) \geq \nu_k F_{\Omega_k}^{-1}(1-\zeta), \ \forall k, && \text{(14d)}\\
& \text{(8d)}, \text{(9c)}, \text{(9d)},
\end{aligned}$$

where $R_k(\epsilon_k, \mu_k) = F_k(\mu_k) - G_k(\epsilon_k, \mu_k)$ in constraint (14d), and the terms $F_k(\mu_k)$ and $G_k(\epsilon_k, \mu_k)$ are given by:

$$F_k(\mu_k) = \log(1+\mu_k), \forall k, \tag{15}$$

$$G_k(\epsilon_k, \mu_k) = Q^{-1}(\epsilon_k)\sqrt{\frac{a^2}{m_d}\left(1-(1+\mu_k)^{-2}\right)}, \forall k. \tag{16}$$

(P3) is still a non-convex optimization problem. To overcome this, we first modify the optimization problem and represent it as the canonical form which is required for the DC forms. Consequently, we apply first-order Taylor expansion to get a

Algorithm 1 Iterative SCA Algorithm

Input: Set iteration number $t = 0$, maximum number of iterations $T_{\max}$, and initialize the decoding error $\epsilon_k = \epsilon_k^0$, the auxiliary variable $\mu_k = \mu_k^0$, and the active beamformers as $\mathbf{W}_k = \mathbf{W}_k^{(0)}$.

1: **repeat**
2: Calculate $\tilde{G}_k(\mu_k)$ and $\tilde{\mathcal{Q}}_k(\Upsilon_k)$ as stated in (27) and (28), respectively.
3: Solve (P5) to obtain $\{\epsilon_k^t, \mu_k^t, \mathbf{W}_k^t\}$.
4: Set $t = t + 1$.
5: **until** $t = T_{\max}$
6: **Return** $\Upsilon_k^* = \{\epsilon_k^t, \mu_k^t, \mathbf{W}_k^t\} = \{\epsilon_k^*, \mu_k^*, \mathbf{W}_k^*\}$.

convex approximation of the non-convex terms. In particular, constraint (14c) can be represented as

$$\begin{aligned} &\mu_k g_k(\mathbf{W}_k) \leq f_k(\mathbf{W}_k) \\ \Rightarrow &\mu_k \mathcal{A}(\mathbf{W}_k) \leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \forall k, \end{aligned} \tag{17}$$

where $\mathcal{A}(\mathbf{W}_k) = \sum_{i \in K, i \neq k} \text{Tr}(\mathbf{H}_k \mathbf{W}_i)$. Nevertheless, (17) is a non-convex constraint since it is the product of two optimization variables, i.e., $\mathbf{W}_i$ and μ_k, $\forall i, k$. However, it can be decoupled by adopting the following form [9], [15]:

$$\mu_k \mathcal{A}(\mathbf{W}_k) = \mathcal{P}_k(\mu_k, \mathbf{W}_k) - \mathcal{Q}_k(\mu_k, \mathbf{W}_k), \tag{18}$$

where

$$\mathcal{P}_k(\mu_k, \mathbf{W}_k) = \frac{1}{2}\left(\mu_k + \mathcal{A}(\mathbf{W}_k)\right)^2, \forall k, \tag{19}$$

$$\mathcal{Q}_k(\mu_k, \mathbf{W}_k) = \frac{1}{2}\left(\mu_k\right)^2 + \left(\mathcal{A}(\mathbf{W}_k)\right)^2, \forall k. \tag{20}$$

By denoting $\Upsilon_k = \{\epsilon_k, \mu_k, \mathbf{W}_k\}$ as a set of optimization variables, (P3) can be recast as follows:

$$\text{(P4)}: \underset{\Upsilon_k}{\text{minimize}} \quad \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \tag{21a}$$

$$\text{s.t.}: \; F_k(\mu_k) - G_k(\mu_k) \geq \nu_k F_{\Omega_k}^{-1}(1-\zeta), \; \forall k, \tag{21b}$$

$$\mathcal{P}_k(\Upsilon_k) - \mathcal{Q}_k(\Upsilon_k) \leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \forall k, \tag{21c}$$

$$(8d), (9c), (9d), (14b).$$

One obstacle, for solving the above optimization problem in (P4) is the incorporating $Q^{-1}(\cdot)$ function. To tackle this issue, we introduce the following *Lemma* to approximate it.

Lemma 1: For $0 < \epsilon_k < 1$, an approximation of $Q^{-1}(\epsilon_k)$ is given by:

$$Q^{-1}(\epsilon_k) \approx \sqrt{\frac{\pi}{2}}(B - C\epsilon_k), \tag{22}$$

where B and C are defined as:

$$B = \left(1 + \frac{\pi}{12} + \frac{7\pi^2}{480} + \frac{127\pi^3}{40320} + ...\right), \tag{23}$$

$$C = \left(1 + \frac{\pi}{2} + \frac{7\pi^2}{48} + \frac{127\pi^3}{2880} + ...\right). \tag{24}$$

By adopting the Lemma 1, the data rate constraint function in (21b), can be rewritten as:

$$F_k(\mu_k) - \sqrt{\frac{\pi}{2\ m_d}}\left(1 - \frac{1}{(1+\gamma_k)^2}\right)(B - C\epsilon_k) \geq R_{\min}. \tag{25}$$

However, (25) is still non-convex due to incorporating the channel dispersion. Let assume that we are in the high SINR regime and the channel dispersion can be approximated by:

$$V_k = \left(1 - \frac{1}{(1+\gamma_k)^2}\right) \approx 1. \tag{26}$$

Consequently, equation (25) can be restated as follows:

$$F_k(\mu_k) - \overbrace{\sqrt{\frac{\pi}{2\ m_k}}(B - C\epsilon_k)}^{=\tilde{G}_k(\epsilon_k, \mu_k)} \geq R_{\min}. \tag{27}$$

It should be noted that the constraint (21c) belongs to the class of DC problems [15]. Thus, the SCA technique can be directly applied to approximate the non-convex problem in each iteration. To this end, we use first-order Taylor expansion to obtain a globally lower-bound of function $\mathcal{Q}_k(\Upsilon_k)$, $\forall k$. At iteration t, the lower-bounds of these functions are given by:

$$\begin{aligned} \mathcal{Q}_k(\Upsilon_k) \geq \tilde{\mathcal{Q}}_k(\Upsilon_k) \triangleq\ & \mathcal{Q}_k(\Upsilon_k^t) + \partial_{\mu_k}^T \mathcal{Q}_k\left(\Upsilon_k^t\right)(\mu_k - \mu_k^t) \\ & + \text{Tr}\Big(\nabla_{\mathbf{W}_k}^H \mathcal{Q}_k\left(\Upsilon_k^t\right)\left(\mathbf{W}_k - \mathbf{W}_k^t\right)\Big), \forall k, \end{aligned} \tag{28}$$

respectively. Then, by dropping the non-convex rank-one constraint (9c), (P4) with any given local point at iteration t can be approximated as:

$$\text{(P5)}: \underset{\Upsilon_k}{\text{minimize}} \quad \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \tag{29a}$$

$$\text{s.t.}: \; F_k(\mu_k) - \tilde{G}_k(\epsilon_k, \mu_k) \geq R_{\min}, \; \forall k, \tag{29b}$$

$$\mathcal{P}_k(\Upsilon_k) - \tilde{\mathcal{Q}}_k(\Upsilon_k) \leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \forall k. \tag{29c}$$

$$(8d), (9d), (14b).$$

The optimization problem (P5) is now a convex optimization problem that can be efficiently solved by standard convex optimization solvers such as CVX. The iterative SCA algorithm for (P5) is given in **Algorithm 1**.

B. Second-stage: Optimizing $\boldsymbol{\Theta}$ with Fixed $\mathbf{w}_k$ and ϵ_k

The main difficulties of optimizing the phase shifts at the IRS is constraint (8c). To be more specific, constraint (8c) is a unit-module constraint, which makes solving the problem intractable. Accordingly, we first define $\boldsymbol{\theta} = (e^{j\alpha_1}, ..., e^{j\alpha_N})^H \in \mathbb{C}^{N \times 1}$ and $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}^T \ \tau]^T \in \mathbb{C}^{(N+1)\times 1}$, respectively, where $\tau \in \mathbb{C}$ is a dummy variable with $|\tau| = 1$. To facilitate the solution design, we also define $\mathbf{V} = \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H \in \mathbb{C}^{(N+1)\times(N+1)}$. Thus, we obtain:

$$\left|\left((\mathbf{h}_k^{\text{IU}})^H \boldsymbol{\Theta} \mathbf{H} + (\mathbf{h}_k^{\text{AU}})^H\right)\mathbf{w}_k\right|^2 \triangleq \text{Tr}(\mathbf{V}\mathbf{X}_k\mathbf{W}_k\mathbf{X}_k^H) = \text{Tr}(\mathbf{W}_k\mathbf{Y}_k), \tag{30}$$

where

$$\mathbf{X}_k = \begin{bmatrix} \left(\text{diag}\left((\mathbf{h}_k^{\text{IU}})^H\right)\mathbf{H}\right)^T & (\mathbf{h}_k^{\text{AU}})^* \end{bmatrix}^T, \tag{31}$$

$$\mathbf{Y}_k = \mathbf{X}_k^H \mathbf{V} \mathbf{X}_k. \tag{32}$$

Since the objective function in (P1) is independent of $\mathbf{V}$, it turns into a feasibility problem. To dispose of the feasibility

problem, we resort to another technique in order to obtain $\boldsymbol{\Theta}$ by studying the subsequent optimization problem:

$$\text{(P6)}: \underset{\mathbf{V},\alpha_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \tag{33a}$$

$$\text{s.t.:} \quad \text{Tr}(\mathbf{W}_k\mathbf{Y}_k) - \sum_{i\in K, i\neq k} \text{Tr}(\mu_k^*\mathbf{W}_i\mathbf{Y}_k) \geq \mu_k^*\sigma_k^2 + \alpha_k, \forall k, \tag{33b}$$

$$\text{Diag}(\mathbf{V}) = \mathbf{1}_{N+1}, \tag{33c}$$

$$\mathbf{V} \succeq \mathbf{0}, \tag{33d}$$

$$\text{Rank}(\mathbf{V}) \leq 1. \tag{33e}$$

Thus, the value of $\boldsymbol{\Theta}$ can be obtained to maximize the SINR margin from the minimum required values specified in (P1). It is worth mentioning that constraint (33c) is forced to guarantee the unit-modulus constraints. Generally, (P6) usually yield a solution with a rank higher than one due to constraint (33c). To handle it, we rewrite constraint (33e) in a mathematically tractable form via the DC method. Thus, the equivalent form of constraint (33c) can be represented as:

$$||\mathbf{V}||_* - ||\mathbf{V}||_2 \leq 0. \tag{34}$$

Note that $||\mathbf{V}||_* = \sum_i \sigma_i \geq ||\mathbf{V}||_2 = \max_i\{\sigma_i\}$ holds for any given $\mathbf{V} \in \mathbb{H}^{N\times N}$, where σ_i is the i-th singular value of $\mathbf{V}$. The equality holds if and only if $\mathbf{V}$ achieves rank one i.e., $\text{Rank}(\mathbf{V}) = 1$ [16], [17]. Now, we take the first-order Taylor approximation of $||\mathbf{V}||_2$ as:

$$||\mathbf{V}||_2 \geq ||\mathbf{V}^{(t)}||_2 + \text{Tr}\Big(\lambda_{\max}(\mathbf{V}^{(t)})\lambda_{\max}^H(\mathbf{V}^{(t)})(\mathbf{V}-\mathbf{V}^t)\Big). \tag{35}$$

By resorting to (35), a convex approximation can be obtained for (34) which is given by:

$$||\mathbf{V}||_* - ||\mathbf{V}^{(t)}||_2 - \text{Tr}\Big(\lambda_{\max}(\mathbf{V}^{(t)})\lambda_{\max}^H(\mathbf{V}^{(t)})(\mathbf{V}-\mathbf{V}^t)\Big) \leq 0. \tag{36}$$

Finally, with the convex constraint (36) at hand, the optimization problem in the $(t+1)$-iteration can be written as follows:

$$\text{(P7)}: \underset{\mathbf{V},\alpha_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \tag{37a}$$

$$\text{s.t.}: \text{(33b)–(33d)}, \text{(36)}. \tag{37b}$$

The optimization problem (P7) is now convex and can be efficiently solved by CVX [4]. Eventually, the Final iterative-based AO algorithm is provided in **Algorithm 2**.

V. Numerical Results

This section demonstrates the effectiveness of the proposed algorithm for MISO URLLC-enabled IRS systems under finite blocklength codes. A rectangular area with a dimension of $(100, 100)$ meters is considered, where the AP is placed at $(0, 0)$ m, while the IRS is located at $(50, 0)$ m and all the users are assumed to be randomly located inside the rectangular area. The path loss model is given by $35.3 + 37.6\log_{10}(d_k)$ dB, where d_k indicates the distance between AP-user k in kilometer. Furthermore, the convergence tolerance is set as 10^{-2}, and a thermal noise density of -174 dBm/Hz is assumed. Besides, the value of the maximum DEP for user k is given by $\epsilon_{k,\max} = 10^{-7}$. In addition, it is assumed that the average traffic load for URLLC users is 0.1 Mbps, and the bandwidth is set to 350 kHz and $K = 4$, $N = 50$ for all simulation setups.

Algorithm 2 Iterative AO algorithm

Input: Set $i = 0$, $I_{\max}$, and initialize the phase shifts as $\boldsymbol{\Theta} = \boldsymbol{\Theta}^0$.

1: **Repeat**
2: Solve problem (P5) for given $\boldsymbol{\Theta}^i$, and obtain the optimal solutions $\{\epsilon_k^i, \mu_k^i, \mathbf{W}_k^i\}$.
3: Solve problem (P7) for given $\{\epsilon_k^i, \mu_k^i, \mathbf{W}_k^i\}$.
4: $i = i + 1$.
5: **until** $i = I_{\max}$
6: **Return** $\{\epsilon_k^*, \mu_k^*, \mathbf{w}_k^*, \boldsymbol{\Theta}^*\} = \{\epsilon_k^i, \mu_k^i, \mathbf{w}_k^i, \boldsymbol{\Theta}^i\}$.

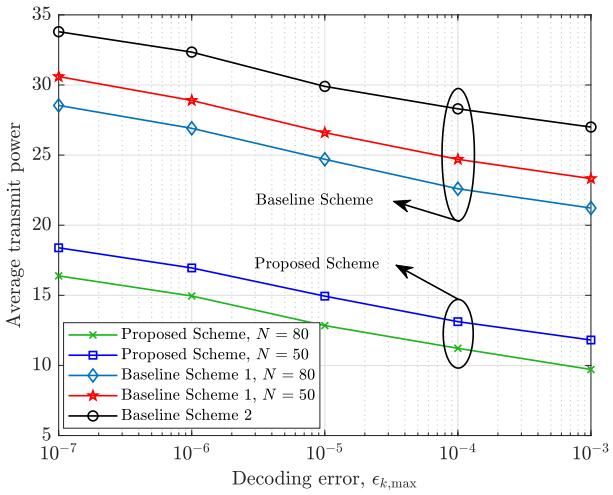

Fig. 2. Impact of decoding error, ϵ_k, on the average transmit power.

Fig. 2 shows the maximum DEP, $\epsilon_{k,\max}$ on the average transmit power for a block code with the length of 200 (symbols), i.e., $m_d = 250$. It can be seen that with increasing the decoding error, the transmit power decreases. The more reliable a network is, the more transmit power is consumed. This is because that $Q^{-1}(\epsilon_k)$ is a decreasing function with respect to ϵ_k which yields to declining $G_k(\epsilon_k, \mathbf{w}_k, \boldsymbol{\Theta})$. Therefore, with a small transmit power, the minimum data rate requirement can be satisfied, decreasing the AP's transmit power at the end. This figure also investigates that the transmit power scales down with an increasing number of reflecting elements at the AP. We also compare our proposed scheme with two baseline schemes. For baseline scheme 1, we consider fixed beamforming at the IRS, while there is no IRS in baseline scheme 2. This figure reveals that our proposed scheme has a better performance than baseline scheme 2 due to deploying IRS and jointly optimizing the beamforming matrices at the AP and IRS compared to baseline scheme 1.

IRS is a promising approach for a green wireless commu-

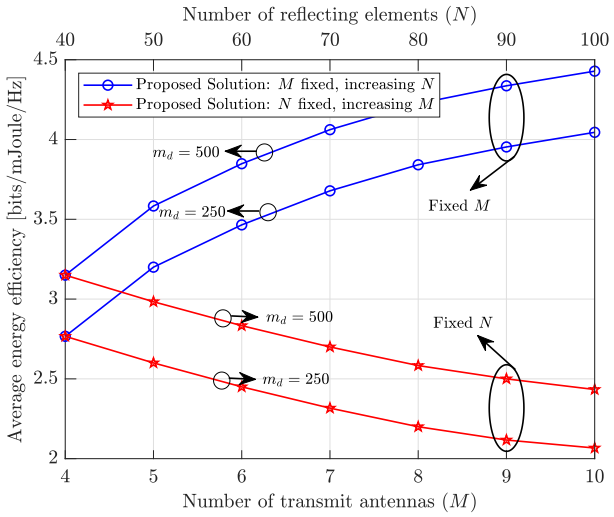

Fig. 3. Average EE vs. number of transmit antennas and reflecting elements.

nication system. Let us define energy efficiency (EE) as the ratio of the total system data rate to the corresponding network power consumption in [bits/Joule]:

$$\mathscr{E}_{eff}(\epsilon_k, \mathbf{w}_k, \mathbf{\Theta}) = \frac{\sum\limits_{k\in\mathcal{K}} R_k(\epsilon_k, \mathbf{w}_k, \mathbf{\Theta})}{\sum\limits_{k\in\mathcal{K}} \|\mathbf{w}_k\|^2 + P_s + N_T P_d + P_c + M P_{\text{Dyn}}}, \quad (38)$$

where $P_s = 100$ mW indicates the static power consumption as required to maintain the basic circuit operations of the IRSs, $P_d = 0.33$ mW is the dynamic power dissipation per reflecting component, $P_c = 1$ W is the circuit power at the AP, and $P_{\text{Dyn}} = 100$ mW is the dynamic power consumption of the AP per antennas.

Fig. 3 illustrates the impact of the reflecting elements of IRS and transmit antennas at the AP versus the EE. It is observed that the EE increases with increasing the number of reflecting elements. On the other hand, the EE of the system decreases with increasing the number of antennas at the AP. This is because with increasing the number of antennas at the AP, the power consumption at the AP increases due to increasing the number of RF chains which degrades the performance of the system in terms of EE. One can conclude that the IRS is more efficient for green wireless communication as it does not consume more transmit power as they are passive device. Besides, increasing the number of reflecting elements provides more degrees of freedom for the network to increase the network's data rate while reducing the system's transmit power.

VI. Conclusion

This paper investigated the resource allocation for a DL multiuser MISO URLLC-aided IRS system. In particular, the resource allocation design for active/passive beamforming was formulated to minimize the total transmit power while considering the traffic load for each URLLC user as the QoS constraint with adopting the sort packet transmission. The underlying problem was non-convex. To handle this difficulty, we first employed the AO method to divide the main problem into two sub-problems, i.e., active/passive beamforming sub-problems. Then we adopted the SCA approach and a penalty-based method to solve the beamforming matrices subproblems, respectively. Simulation results investigated our proposed scheme, taking into account that the IRS could help the URLLC system meet the QoS service and reduce the transmission power significantly compared to other conventional methods. Simulation results also confirmed that the IRS could also be beneficial for EE, which shows the significant effect of the IRS on power-efficient green communication.

References

[1] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, pp. 3313–3351, May 2021.

[2] Z. Kang, C. You, and R. Zhang, "IRS-aided wireless relaying: Deployment strategy and capacity scaling," *IEEE Wirel. Commun. Lett.*, vol. 11, pp. 215–219, Feb. 2022.

[3] S. Xu, Y. Du, J. Liu, and J. Li, "Weighted sum rate maximization in IRS-BackCom enabled downlink multi-cell MISO network," *IEEE Commun. Lett.*, vol. 26, pp. 642–646, Mar. 2022.

[4] S. Zargari, A. Khalili, and R. Zhang, "Energy efficiency maximization via joint active and passive beamforming design for multiuser MISO IRS-aided SWIPT," *IEEE Wirel. Commun. Lett.*, vol. 10, pp. 557–561, Mar. 2021.

[5] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, pp. 1834–1853, Oct. 2018.

[6] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Comm. Mag.*, vol. 55, pp. 72–78, Jun. 2017.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.

[8] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wirel. Comm.*, vol. 18, pp. 402–415, Jan. 2019.

[9] 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, pp. 7184–7200, Nov. 2020.

[10] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell IRS-aided OFDMA-URLLC systems," in *Proc. IEEE WCNC*, pp. 1–7, 2021.

[11] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, pp. 2666–2682, Nov. 2020.

[12] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Average rate and error probability analysis in short packet communications over RIS-aided URLLC systems," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 10320–10334, Oct. 2021.

[13] H. Xie, J. Xu, Y.-F. Liu, L. Liu, and D. W. K. Ng, "User grouping and reflective beamforming for IRS-aided URLLC," *IEEE Wirel. Comm. Lett.*, vol. 10, pp. 2533–2537, Nov. 2021.

[14] M. Setayesh, S. Bahrami, and V. W. Wong, "Joint PRB and power allocation for slicing eMBB and URLLC services in 5G C-RAN," in *Proc. IEEE GLOBECOM*, pp. 1–6, 2020.

[15] A. Rezaei, A. Khalili, J. Jalali, H. Shafiei, and Q. Wu, "Energy-efficient resource allocation and antenna selection for IRS-assisted multicell downlink networks," *IEEE Wirel. Commun. Lett.*, vol. 11, pp. 1229–1233, Jun. 2022.

[16] S. Zargari, A. Khalili, Q. Wu, M. Robat Mili, and D. W. K. Ng, "Max-Min fair energy-efficient beamforming design for intelligent reflecting surface-aided SWIPT systems with non-linear energy harvesting model," *IEEE Trans. Veh. Technol*, vol. 70, pp. 5848–5864, Jun. 2021.

[17] A. Khalili, S. Zargari, Q. Wu, D. W. K. Ng, and R. Zhang, "Multi-objective resource allocation for IRS-aided SWIPT," *IEEE Wirel. Commu Lett.*, vol. 10, pp. 1324–1328, Jun. 2021.