STAR-RIS Enhanced Finite Blocklength Transmission for Uplink NOMA Networks

Suyu Lv, Xiaodong Xu, *Senior Member, IEEE,* Shujun Han, *Member, IEEE,* Yuanwei Liu, *Senior Member, IEEE* Ping Zhang, *Fellow, IEEE* and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*—A simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) assisted uplink non-orthogonal multiple access (NOMA) framework for finite blocklength (FBL) transmission is proposed. Considering the different communication requirements of Internet of Things devices (IoTDs), a novel design to achieve high-rate and low-error is proposed. Two operating protocols for STAR-RIS are considered, namely *energy splitting* (ES) and *mode switching* (MS). 1) For STAR-RIS with ES, an alternating optimization (AO) algorithm is proposed to handle the highly-coupled mixed integer programming problem. More particularly, a low-complexity received-signal-strength-based device pairing scheme is proposed. Based on the given device pair, the closed-form solutions for the power allocation problem are obtained. The transmitting and reflecting coefficient optimization problem is solved by exploiting the successive convex approximation and semidefinite relaxation methods. 2) For STAR-RIS with MS, a double-layer penalty-based (DLPB) algorithm is proposed to tackle the newly introduced binary amplitude constraints. Numerical results reveal that: i) the proposed AO and DLPB algorithms can converge within a few iteration times; ii) the FBL transmission performance can be improved by employing the proposed STAR-RIS framework compared with conventional transmitting/reflecting-only RISs; iii) NOMA is capable of enhancing FBL rate while guaranteeing the reliability constraints compared with orthogonal multiple access.**

***Index Terms*—Finite blocklength (FBL) transmission, Internet of things (IoT), non-orthogonal multiple access (NOMA), simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS).**

I. Introduction

The flourishing Internet of Things (IoT) technology will provide an enabling platform for the ubiquitous smart devices and sensors, making it an integral part for the realization of intelligent communication systems. It is predicted that by 2030, about 50 billion IoT devices (IoTDs) will be in service worldwide [1], which will form a tremendous network of interconnected devices from wearables to driverless vehicles. Profiting from this, our world is developing towards a more intelligent direction, covering various important fields such as smart home, smart healthcare, intelligent agriculture and intelligent manufacturing [2]–[4].

On the one hand, IoTDs serving different industries also have different communication objectives and requirements, such as large bandwidth, high data rate, stringent reliability and end-to-end latency [5], [6]. On the other hand, the demand for massive IoTDs connections also makes the scarcity of spectrum resources increasingly prominent. Therefore, the novel design for satisfying various objectives and massive connections of IoT networks is becoming increasingly urgent. Furthermore, IoTDs may be deployed in a variety of complex environments, and the presence of high-rises, trees and other objects may cause the obstruction of line-of-sight (LoS) link, especially in urban or indoor environments. In order to overcome these challenges, reconfigurable intelligent surface (RIS) can be recognized as a potential green solution to improve communication quality by intelligently adjusting the wireless radio environment with extremely low energy consumption.

Traditionally, RISs consist of an array of low-cost passive reflecting elements, in which each element independently reflects the incident signal after adjusting the electromagnetic wave phase. However, the limitation of conventional RISs lies in its reflecting-only ability, which means that the transmitter and the receiver should be located on the same side of RIS, resulting in *half-space* coverage constraints. For solving this problem, a new paradigm named simultaneously transmitting and reflecting RIS (STAR-RIS) has been proposed, where the incident signals can be reflected and transmitted to both sides of the surface concurrently, thus achieving a *full-space* intelligent propagation environment [7]–[10]. Three practical protocols for operating STAR-RISs in wireless networks, namely energy splitting (ES), mode switching (MS), and time switching (TS), were proposed by the authors in [7].

A. Prior Works

Strict latency requirements and massive access requirements are key challenges for IoT. Aiming at realizing extremely low packet error probability and ultra-low latency for IoT networks, finite blocklength (FBL) transmission is viewed as a feasible solution, which has attracted considerable attention in recent years [11]–[15]. Moreover, non-orthogonal multiple access (NOMA) is considered a key remedy to provide massive connectivity by allowing multiple devices to share the same time-frequency resources simultaneously, and eliminating inter-device interferences at the receivers through

Suyu Lv and Shujun Han are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail: lvsuyu@bupt.edu.cn; hanshujun@bupt.edu.cn).

Xiaodong Xu and Ping Zhang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen 518066, Guangdong, China (e-mail: xuxiaodong@bupt.edu.cn, pzhang@bupt.edu.cn).

Yuanwei Liu and Arumugam Nallanathan are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk; a.nallanathan@qmul.ac.uk).

successive interference cancellation (SIC) technology [16], [17]. For improving user fairness and spectral efficiency of IoT networks, the authors in [18] introduced NOMA into FBL transmission, and investigated the tradeoff among the throughput, latency and reliability. Considering the joint decoding of short packets in NOMA-based networks, the authors in [19] studied the optimization of decoding error probability and power allocation factor. To address the problem that wireless signals are easily blocked by densely deployed equipment in factory automation scenarios, the authors in [20]–[22] proposed to utilize RIS to provide an alternative link, thus enhancing the FBL transmission rate and reliability. The authors in [23] combined RIS and NOMA into IoT systems to meet the needs of high energy efficiency, low power consumption and wide coverage. Aiming at maximizing the network sum rate or energy efficiency for RIS-assisted NOMA systems, the joint resource optimization problems were investigated in [24]–[26]. However, although the combination of RIS and NOMA has obvious advantages in enhancing capacity, it still has the limitations of half-space smart environment.

STAR-RISs have natural advantages in NOMA-based communication systems because of their distinctive ability to expand the channel differences by reconstructing the wireless propagation environment on both sides, which encourages the research on the integration of STAR-RIS and NOMA technologies [27]–[30]. The studies in [27] and [28] confirmed the superiority of STAR-RIS-NOMA in improving system sum rate compared with the conventional RIS and STAR-RIS-OMA. A simultaneous-signal-enhancement-and-cancellation-based design was proposed for STAR-RIS enhanced NOMA-joint-transmission-coordinated-multipoint networks in [29], aiming at eliminating the inter-cell interferences while boosting the desired signals simultaneously. Under FBL regime, the authors in [30] provided block error rate (BLER) analysis for STAR-RIS-aided downlink NOMA communication system. However, when applied to uplink communication systems, the incident signals come from both sides of the STAR-RISs, which will inevitably cause energy loss and information leakage. Focusing on uplink NOMA systems with the assistance of STAR-RISs, the authors in [31]–[33] investigated the hardware and signal model for the dual-sided STAR-RISs, the transmit power minimization problem and the secrecy transmission problem, respectively. In order to explore the potential of exploiting STAR-RIS to enhance throughput in uplink NOMA systems, the authors in [34] theoretically derived the ergodic sum rate under channel estimation errors and hardware impairments, while the authors in [35] investigated the average rate maximization problem with movable STAR-RIS. However, these researches are only aimed at one single performance metric, which cannot be applicable to IoT networks with diverse communication requirements.

B. Motivation and Contributions

On the one hand, massive IoT devices will generate a huge number of short data packets for uplink transmission, which requires high-efficiency multiple access to alleviate the pressure of scarce spectrum resources. On the other hand, IoT devices have various communication demands, which requires novel system design to satisfy specific needs in different application scenarios. Although the STAR-RIS-assisted NOMA system has received widespread attention, the research on dual-sided STAR-RIS for uplink transmission is still in its infancy, which motivates us to develop this paper. More specifically, there is no research considering that users on STAR-RIS's two sides have different communication requirements, especially for the uplink-transmission-dominant networks with FBL constraints. Therefore, we focus on the FBL transmission performance enhancement in STAR-RIS-aided uplink NOMA IoT networks, aiming at maximizing the achievable sum rate of the IoTDs on the reflection space, while ensuring the reliability of the IoTDs on the transmission space. The main contributions of our works are listed as follows:

1) We propose a STAR-RIS-assisted uplink NOMA communication framework for IoT networks, where IoTDs on reflection space (R-IoTDs) are transmission performance-oriented, aiming to enhance achievable FBL rate, while the IoTDs on transmission space (T-IoTDs) aim at guaranteeing certain reliability under FBL constraints.
2) For STAR-RIS with ES protocol, we propose an alternating optimization (AO) algorithm to perform the resource allocation in an iterative manner, aiming to achieve high-rate and low-error (HRLE) for R-IoTDs and T-IoTDs respectively. Specifically, we first propose a received signal strength (RSS) based device pairing scheme, then obtain the closed-form expressions for the IoTDs' transmission power allocation problem. Finally, the STAR-RIS coefficient optimization problem is solved by the successive convex approximation (SCA) and semi-definite relaxation (SDR) methods.
3) For STAR-RIS with MS protocol, we propose a double-layer penalty-based (DLPB) algorithm to handle the newly introduced binary constraints. More particularly, the inner layer is for solving the resource allocation optimization problem with relaxed binary amplitude coefficient constraints, while the outer layer is for the binary amplitude constraint recovery.
4) Numerical results reveal that: i) the proposed AO and DLPB algorithms can achieve convergence within a few iterations; ii) the proposed STAR-RIS framework can obtain a higher sum FBL rate compared with the conventional transmitting/reflecting-only RISs; iii) NOMA scheme outperforms the OMA scheme in improving the achievable FBL rate while guaranteeing reliability constraints; iv) ES protocol of STAR-RIS has better FBL transmission performance than MS protocol.

The organization for the rest of this paper is as follows. In section II, we propose a STAR-RIS-assisted uplink NOMA framework for FBL transmission. In section III, we propose an AO algorithm for STAR-RIS with ES protocol to achieve HRLE design. Based on the AO algorithm, we propose a DLPB algorithm for STAR-RIS with MS protocol in section IV. The simulation results and conclusions are given in sections V and VI, respectively. The notations used in this paper are presented in **TABLE I**.

TABLE I
NOTATIONS

Notations	Meanings
IoTD m	m-th IoTD on *R-Space*
IoTD n	n-th IoTD on *T-Space*
$\mathcal{M}/\mathcal{N}$	set of IoTDs on *R/T-Space*
M/N	number of IoTDs on *R/T-Space*
K	number of STAR-RIS elements
$x_{m,n}$	indicator of device pairing
h_0	path loss at the reference distance $d = 1m$
α	path loss exponent
κ	Rician factor
p_m/p_n	transmission power of IoTD m/IoTD n
L_m/L_n	short-packet blocklength
$\varepsilon_m/\varepsilon_n$	tolerable decoding error probability
$R_{n,\text{target}}$	targeted transmission rate of IoTD n
$\boldsymbol{\Theta}_r/\boldsymbol{\Theta}_t$	reflection/transmission-coefficient matrix of STAR-RIS
$\text{diag}(\mathbf{x})$	diagonal matrix whose diagonal elements are $\mathbf{x}$
$\text{Diag}(\mathbf{X})$	vector whose elements are the diagonal elements of $\mathbf{X}$
$Q(\cdot)$	Gaussian Q-function
$[\cdot]^+$	$\max[\cdot, 0]$

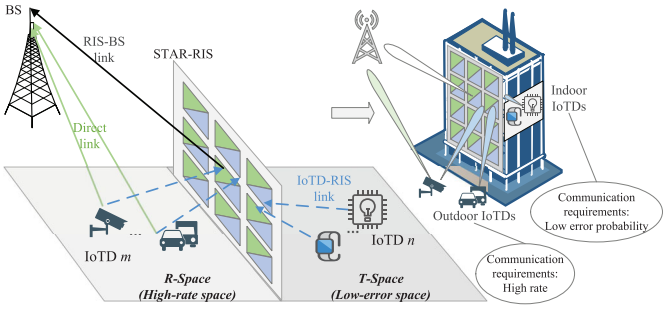


Fig. 1. STAR-RIS assisted uplink NOMA transmission for IoT networks.

II. System Model

We consider an uplink NOMA-based IoT communication scenario with STAR-RIS-assisted, as illustrated in Fig. 1, where single-antenna IoTDs transmit signals to a single-antenna destination base station (BS), denoted as $\boldsymbol{D}$. The STAR-RIS splits the whole communication system into two parts. One half space and $\boldsymbol{D}$ are located on the same side of STAR-RIS, and the signals from IoTDs can be reflected to $\boldsymbol{D}$, called *R-Space*. The other half space and $\boldsymbol{D}$ are located on the opposite side of STAR-RIS, and the signals from IoTDs need to be transmitted to $\boldsymbol{D}$ passing through STAR-RIS, called *T-Space*. The outdoor IoTDs on *R-Space* and indoor IoTDs on *T-Space* are abbreviated as R-IoTDs and T-IoTDs, whose sets are denoted by $\mathcal{M}$ ($|\mathcal{M}| = M$) and $\mathcal{N}$ ($|\mathcal{N}| = N$) respectively.

In this paper, we consider that the IoTDs on *R-Space* and *T-Space* have different communication requirements. Specifically, since STAR-RIS can be deployed on the glass in buildings, the IoTDs on *R-Space* can be considered outdoor devices with high transmission rate requirements. The IoTDs on *T-Space* can be considered indoor devices, such as smart sensors in intelligent homes. This kind of device has low requirements for latency and transmission rate, and only needs to transmit the signals to the destination BS within a tolerable error probability[1]. Therefore, *R-Space* and *T-Space* can be also referred to as *high-rate space* and *low-error space* respectively, as illustrated in Fig. 1.

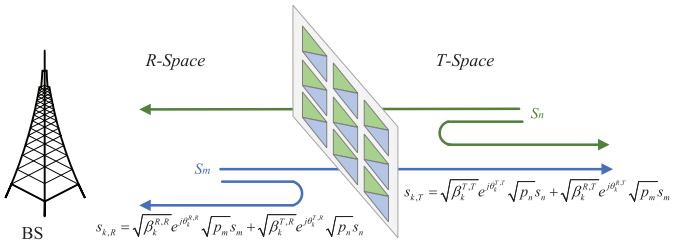


Fig. 2. Dual-sided incident signal model.

A. STAR-RIS Model

STAR-RIS consists of a uniform planar array (UPA) with $K = K_x K_y$ elements, where K_x and K_y are the number of elements along the x-axis and y-axis, respectively. Since uplink transmission services are dominant in IoT networks, we consider signals are transmitted from IoTDs on both sides to $\boldsymbol{D}$ with the assistance of STAR-RIS. Denote the transmitted data symbols of IoTD m ($m \in \mathcal{M}$) and IoTD n ($n \in \mathcal{N}$) as s_m and s_n, respectively. Disregarding the effect of path loss and fading, the signals of IoTD m and IoTD n radiating from the k-th element towards *R-Space* and towards *T-Space*, as shown in Fig. 2, can be given as follows

$$s_{k,R} = \sqrt{\beta_k^{R,R}} e^{j\theta_k^{R,R}} \sqrt{p_m} s_m + \sqrt{\beta_k^{T,R}} e^{j\theta_k^{T,R}} \sqrt{p_n} s_n, \quad \text{(1a)}$$

$$s_{k,T} = \sqrt{\beta_k^{R,T}} e^{j\theta_k^{R,T}} \sqrt{p_m} s_m + \sqrt{\beta_k^{T,T}} e^{j\theta_k^{T,T}} \sqrt{p_n} s_n, \quad \text{(1b)}$$

where p_m and p_n denote the transmission power of IoTD m and IoTD n respectively. $\beta_k^{R,R}, \beta_k^{T,T}$ $(\beta_k^{R,T}, \beta_k^{T,R})$ denote the amplitude coefficient of reflection for *R/T-Space* (transmission from *R/T-Space* to *T/R-Space*) of the dual-sided STAR-RIS. Specifically, $\beta_k^{R,R}, \beta_k^{T,T}, \beta_k^{R,T}, \beta_k^{T,R} \in [0,1]$ for STAR-RIS operating with energy splitting (ES) protocol, and $\beta_k^{R,R}, \beta_k^{T,T}, \beta_k^{R,T}, \beta_k^{T,R} \in \{0,1\}$ for STAR-RIS operating with mode switching (MS) protocol, satisfying the principle of energy conservation, i.e., $\beta_k^{R,T} + \beta_k^{R,R} = 1$ and $\beta_k^{T,R} + \beta_k^{T,T} = 1$ [7]. Similarly, $\theta_k^{R,R}, \theta_k^{T,T} \in [0, 2\pi)$ $(\theta_k^{R,T}, \theta_k^{T,R} \in [0, 2\pi))$ denote the phase shift coefficient of the dual-sided STAR-RIS.

Symmetrical dual-sided STAR-RIS elements are considered [31]. Thus, we have $\beta_k^{R,R} = \beta_k^{T,T}$, $\beta_k^{R,T} = \beta_k^{T,R}$, $\theta_k^{R,R} = \theta_k^{T,T}$ and $\theta_k^{R,T} = \theta_k^{T,R}$, which are abbreviated as β_k^r, β_k^t, θ_k^r and θ_k^t respectively hereinafter, denoting the amplitude coefficient of reflection/transmission and the phase shift coefficient of reflection/transmission. Accordingly, (1) can be reformulated as

$$s_{k,R} = \sqrt{\beta_k^r} e^{j\theta_k^r} \sqrt{p_m} s_m + \sqrt{\beta_k^t} e^{j\theta_k^t} \sqrt{p_n} s_n, \quad \text{(2a)}$$

$$s_{k,T} = \sqrt{\beta_k^t} e^{j\theta_k^t} \sqrt{p_m} s_m + \sqrt{\beta_k^r} e^{j\theta_k^r} \sqrt{p_n} s_n, \quad \text{(2b)}$$

[1]Although the scenario considered in this article is specific, it provides the possibility for STAR-RIS to support users with different communication requirements. Based on the proposed framework, more STAR-RIS-aided wireless systems with various communication requirements can be designed.

where $\beta_k^r, \beta_k^t \in [0,1]$ for ES, $\beta_k^r, \beta_k^t \in \{0,1\}$ for MS, $\beta_k^r + \beta_k^t = 1$, and $\theta_k^r, \theta_k^t \in [0, 2\pi)$. Note that $\boldsymbol{D}$ is located at *R-Space*, thus the signal $s_{k,R}$ is the main consideration. Denote the reflection-coefficient matrix and the transmission-coefficient matrix of STAR-RIS as $\boldsymbol{\Theta}_r \in \mathbb{C}^{K\times K}$ and $\boldsymbol{\Theta}_t \in \mathbb{C}^{K\times K}$ respectively, given by

$$\boldsymbol{\Theta}_r = \mathrm{diag}\left(\sqrt{\beta_1^r}e^{j\theta_1^r}, \cdots, \sqrt{\beta_k^r}e^{j\theta_k^r}, \cdots, \sqrt{\beta_K^r}e^{j\theta_K^r}\right), \tag{3a}$$

$$\boldsymbol{\Theta}_t = \mathrm{diag}\left(\sqrt{\beta_1^t}e^{j\theta_1^t}, \cdots, \sqrt{\beta_k^t}e^{j\theta_k^t}, \cdots, \sqrt{\beta_K^t}e^{j\theta_K^t}\right). \tag{3b}$$

B. Channel Model

Uplink NOMA scheme is adopted, where one R-IoTD and one T-IoTD are considered to share a same resource block. Define a binary matching coefficient $x_{m,n}$, where $x_{m,n} = 1$ indicates that m and n are on a same NOMA cluster, sharing the same time-and-frequency resource block, $x_{m,n} = 0$, otherwise. Without loss of generality, we consider that IoTD m and IoTD n are on a same NOMA cluster. To investigate the pairing strategy of one device on each side of STAR-RIS, we consider the case that $M = N$ in this paper.

The direct channel from IoTD m to $\boldsymbol{D}$ is denoted as h_m, which is modeled as Rician fading channels,

$$h_{m,d} = \sqrt{h_0 (d_{m,d})^{-\alpha}}\left(\sqrt{\frac{\kappa}{1+\kappa}}h_{m,d}^{LoS} + \sqrt{\frac{1}{1+\kappa}}h_{m,d}^{NLoS}\right), \tag{4}$$

where h_0 denotes the path loss at the reference distance $d =$ 1m, $d_{m,d}$ is the distance between IoTD m and $\boldsymbol{D}$, $\alpha \geq 2$ is the path loss exponent, and κ is the Rician factor. $h_{m,d}^{LoS}$ is the deterministic LoS component, $h_{m,d}^{NLoS}$ is the non-LoS (NLoS) component which is modeled as Rayleigh fading. The channels from IoTD m to STAR-RIS, from IoTD n to STAR-RIS and from STAR-RIS to $\boldsymbol{D}$ are denoted by $\mathbf{g}_{m,s}$, $\mathbf{g}_{n,s}$ and $\mathbf{g}_{s,d}$ respectively, which can be generated by Rician fading mode similarly. Therefore, we can obtain the equivalent cascaded m-RIS-$\boldsymbol{D}$ channel and n-RIS-$\boldsymbol{D}$ channel as

$$c_m = h_{m,d} + \mathbf{g}_{m,s}^H \boldsymbol{\Theta}_r \mathbf{g}_{s,d}, \tag{5a}$$

and

$$c_n = \mathbf{g}_{n,s}^H \boldsymbol{\Theta}_t \mathbf{g}_{s,d}. \tag{5b}$$

With given $\langle m, n\rangle$, the received superimposed signals at $\boldsymbol{D}$ can be given by

$$y = c_m\sqrt{p_m}s_m + c_n\sqrt{p_n}s_n + \omega, \tag{6}$$

where $\omega \sim \mathcal{CN}\left(0, \sigma^2\right)$ is the additive white Gaussian noise (AWGN) at $\boldsymbol{D}$. The signal-to-interference plus noise ratio (SINR) for decoding s_z, $z \in \{m, n\}$ at $\boldsymbol{D}$ is

$$\gamma_z = \frac{p_z|c_z|^2}{\lambda_z p_{\overline{z}}|c_{\overline{z}}|^2 + \varrho_z\left(1-\lambda_z\right)p_{\overline{z}}|c_{\overline{z}}|^2 + \sigma^2}. \tag{7}$$

If $z = m$, then $\overline{z} = n$, else $\overline{z} = m$. $\lambda_z \in \{0,1\}$ indicates the decoding order of s_m and s_n at $\boldsymbol{D}$, $\lambda_m + \lambda_n = 1$. If $p_m|c_m|^2 \geq p_n|c_n|^2$, then $\lambda_m = 1$, $\lambda_n = 0$, else $\lambda_m = 0$, $\lambda_n = 1$. $\varrho_z \in [0,1]$ denotes the residual interference factor caused by imperfect SIC. Specifically, $\varrho_z = 0$ means perfect interference cancellation, $\varrho_z = 1$ means the failure of interference cancellation.[2]

C. Performance Metrics for FBL Transmission

FBL transmission is considered in the IoT scenario. The data generated by IoT devices usually only contains 20-250 bytes [36], so FBL will increase the proportion of intended data in the transmitted packet, thus improving the spectral efficiency. Furthermore, FBL transmission is an effective technology to reduce the transmission delay and processing delay, thus realizing the high timeliness of information for delay-sensitive IoT applications. However, FBL has a negative effect on the channel coding gain. In this regard, Shannon's theory based on an infinite blocklength (IFBL) assumption will overestimate the achievable rate and underestimate the error probability of the systems operating under FBL constraints. Therefore, aiming at satisfying the communication requirements with FBL constraints, the following performance metrics are considered.

1) *Achievable FBL rate:* Considering the given packet length L_m and tolerable decoding error probability ε_m, the maximal rate for IoTD m transmitting the message to $\boldsymbol{D}$ reliably can be characterized by the achievable FBL rate, given by [37]

$$R_m = \log_2\left(1+\gamma_m\right) - \kappa_m\sqrt{1 - \frac{1}{\left(1+\gamma_m\right)^2}}, \tag{8}$$

where $\kappa_m = \frac{Q^{-1}(\varepsilon_m)}{\sqrt{L_m}\ln 2}$ is a constant value when the tolerable decoding error probability ε_m and blocklength L_m are given, $Q^{-1}[\cdot]$ denotes the inverse of Gaussian Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$.

2) *Decoding error probability:* For IoTD n on the *T-Space*, with a given packet length L_n and target rate $R_{n,\mathrm{target}}$, the decoding error probability can be calculated as

$$\varepsilon_n = Q\left(\frac{\log_2\left(1+\gamma_n\right) - R_{n,\mathrm{target}}}{\sqrt{1 - \frac{1}{(1+\gamma_n)^2}}} \cdot \sqrt{L_n}\ln 2\right). \tag{9}$$

Note that for the IoTDs on the same subchannel, the equation $L_m = L_n$ should hold. In this paper, we aim to maximize the sum FBL rate of R-IoTDs while guaranteeing the decoding error constraints of T-IoTDs, namely, high rate and low error (HRLE).

III. HRLE Design with ES Protocol

In this section, we will propose an alternating optimization algorithm to realize high-rate for R-IoTDs and low-error for T-IoTDs for STAR-RIS with ES protocol. Specifically, HRLE

[2] We can observe from expression (7) that the number of devices sharing a same resource block is restricted in the considered system, and the BS only performs interference cancellation once for each NOMA cluster, which will not cause excessive accumulated residual interference. Moreover, with the development of SIC technology and the improvement of hardware capabilities, the value of residual interference ϱ_z can be extremely small. Therefore, to evaluate upper bound of the FBL transmission performance, we consider perfect SIC in this paper, which means $\varrho_z = 0$ in (7). We might extend our work to imperfect SIC in the future.

will be achieved through jointly optimizing device pairing, the transmission power of IoTDs, and the amplitude response and phase shift coefficient of STAR-RIS, with the mathematical form given by

$$(\text{P1}): \max_{\{\mathbf{x},\mathbf{p}_r,\mathbf{p}_t,\boldsymbol{\Theta}_r,\boldsymbol{\Theta}_t\}} \sum_{m\in\mathcal{M},n\in\mathcal{N}} x_{m,n}R_m, \tag{10a}$$

$$\text{s.t. } x_{m,n}\in\{0,1\},\forall m\in\mathcal{M},n\in\mathcal{N}, \tag{10b}$$

$$\sum_{m\in\mathcal{M}} x_{m,n}\le 1, \sum_{n\in\mathcal{N}} x_{m,n}\le 1, \tag{10c}$$

$$\begin{cases} \lambda_m=1,\lambda_n=0, \text{if } p_m|c_m|^2\ge p_n|c_n|^2, \\ \lambda_m=0,\lambda_n=1, \text{if } p_m|c_m|^2< p_n|c_n|^2, \end{cases} \tag{10d}$$

$$p_z\le P_{z,\max},\forall z\in\{r,t\}, \tag{10e}$$

$$\beta_k^r,\beta_k^t\in[0,1],\beta_k^r+\beta_k^t=1,\forall k\in\mathcal{K}, \tag{10f}$$

$$\theta_k^r,\theta_k^t\in[0,2\pi),\forall k\in\mathcal{K}, \tag{10g}$$

$$\varepsilon_n\le\varepsilon_{n,\max},\forall n\in\mathcal{N}, \tag{10h}$$

where (10b)-(10c) are the constraints for device pairing, (10d) is the decoding order constraint, (10e) is the maximum transmission rate of IoTDs, (10f) is the amplitude constraints for STAR-RIS with ES protocol, (10g) is the phase shift coefficient constraint of STAR-RIS, and (10h) is the maximum tolerable decoding error probability of IoTDs on the *T-Space*.

The proposed optimization problem is intractable due to the high coupling of optimization variables and the non-convexity of the objective function and the constraints. Furthermore, the device pairing index $x_{m,n}$ can only be 0 or 1, which makes the proposed optimization problem a tough mixed integer programming (MIP) problem. In this section, we propose an alternating optimization based algorithm to solve (10) in an iterative manner.

A. Received-Signal-Strength-Based Device Pairing Scheme

In this subsection, we propose an algorithm to solve the device pairing problem with the given transmission power of IoTDs, and amplitude response and phase shift coefficient of STAR-RIS. The optimality proof of the proposed scheme is given under the condition of high SINR. Under such circumstances, the product of transmission power and equivalent cascaded channel gain, $p_z|c_z|^2, z\in\{r,t\}$, is determined values. Actually, $p_z|c_z|^2$ is the received signal strength of s_z at ***D***. The device pairing problem can be written as

$$(\text{P1.1}): \max_{\{\mathbf{x}\}} \sum_{m\in\mathcal{M},n\in\mathcal{N}} x_{m,n}R_m, \tag{11a}$$

$$\text{s.t. } (10b)-(10d). \tag{11b}$$

Remark 1. *If **D** first decodes IoTD n's signal and then decodes IoTD m's signal, the device pairing scheme will not have an effect on the achievable rate of IoTD m since s_n as interference has been eliminated according to the SIC principle.*

Theorem 1. *Without loss of generality, we consider that $p_{1_r}|c_{1_r}|^2\ge p_{2_r}|c_{2_r}|^2\ge\cdots\ge p_M|c_M|^2$, $p_{1_t}|c_{1_t}|^2\ge p_{2_t}|c_{2_t}|^2\ge\cdots\ge p_N|c_N|^2$. Under the high SINR conditions, if **D** first decodes IoTD m's signal and then decodes IoTD n's signal, the optimal device pairing scheme that maximizes the sum transmission rate of IoTDs on R-Space is that IoTD $m\in\mathcal{M}_u$ with maximum $p_m|c_m|^2$ and IoTD $n\in\mathcal{N}_u$ with minimum $p_n|c_n|^2$ are assigned to the same cluster. That is, the optimal device pairing scheme should be*

$$\Pi_{opt}=\{(1_r,N),(2_r,N-1),\cdots,(M,1_t)\}. \tag{12}$$

Proof. See **Appendix A**. ■

Algorithm 1 RSS-Based Device Pairing Scheme

Input: $\mathcal{M}_u$, $\mathcal{N}_u$.
Output: $\mathbf{x}^*$.

```
repeat
    Find m = max_{m∈M_u} p_m|c_m|^2, n = min_{n∈N_u} p_n|c_n|^2;
    if p_m|c_m|^2 ≥ p_n|c_n|^2 then
        Allocate IoTD m and IoTD n to the same pair;
        Let x_{m,n} = 1, M_u = M_u\m, N_u = N_u\n;
        λ_m = 1, λ_n = 0;
    else
        Find m = min_{m∈M_u} p_m|c_m|^2;
        Allocate IoTD m and IoTD n to the same pair;
        Let x_{m,n} = 1, M_u = M_u\m, N_u = N_u\n;
        λ_m = 0, λ_n = 1;
    end if
until M_u = ∅, N_u = ∅.
```

Based on **Remark 1** and **Theorem 1**, we propose the received signal strength (RSS) based device pairing scheme as concluded in **Algorithm 1**. Specifically, in the case that the decoding order is $\lambda_m=1$ and $\lambda_n=0$, in order to maximize the sum transmission rate of IoTDs on *R-Space*, $m\in\mathcal{M}_u$ with maximum $p_m|c_m|^2$ and $n\in\mathcal{N}_u$ with minimum $p_n|c_n|^2$ is matched into a same pair according to **Theorem 1**. On the contrary, if the decoding order is $\lambda_m=0$ and $\lambda_n=1$, $m\in\mathcal{M}_u$ with minimum $p_m|c_m|^2$ and $n\in\mathcal{N}_u$ with minimum $p_n|c_n|^2$ is matched into a same pair since the transmission rate is irrelevant to the device pairing scheme based on **Remark 1**, while minimizing the interference to n at the same time. Other unpaired devices in the system can be transmitted in an orthogonal or non-orthogonal manner, which is not within the scope of this study.

B. Power Allocation Algorithm

In this subsection, we solve the uplink transmission power allocation problem under fixed device pairing, and STAR-RIS coefficient, which can be expressed as

$$(\text{P1.2}): \max_{\{\mathbf{p}_r,\mathbf{p}_t\}} \sum_{m\in\mathcal{M},n\in\mathcal{N}} x_{m,n}R_m, \tag{13a}$$

$$\text{s.t. } (10d),(10e),(10h). \tag{13b}$$

The optimization goal is to maximize the sum transmission rate of IoTDs on the *R-Space*, and the decoding order will have a significant impact on the rate calculation. Therefore, according to the different decoding orders, we will discuss the power allocation scheme in the following two cases.

Theorem 2. *To maximize the sum transmission rate of IoTDs on the R-Space while satisfying the maximum decoding error constraints of IoTDs on the T-Space, the minimum SINR of IoTD n should meet the following conditions,*

$$\gamma_n^* = \min\left\{ \underset{\gamma_n}{solve}\left[\log_2(1+\gamma_n) - \frac{Q^{-1}(\varepsilon_{n,\max})}{\sqrt{L_n}\ln 2} - \frac{Q^{-1}(\varepsilon_{n,\max})}{\sqrt{L_n}\ln 2}\sqrt{1-\frac{1}{(1+\gamma_n)^2}} - R_{n,\text{target}} = 0\right]\right\}^+. \tag{14}$$

Case 1: When IoTD m is the stronger user and IoTD n is the weaker user, that is when the decoding order is $\lambda_m = 1$ and $\lambda_n = 0$, the closed-form expressions of optimal transmission power for IoTD m and IoTD n can be given by

$$p_m^* = P_{m,\max}, \tag{15a}$$

$$p_n^* = \min\left\{\frac{\gamma_n^*\sigma^2}{|c_n|^2}, P_{n,\max}\right\}. \tag{15b}$$

Case 2: When IoTD m is the weaker user and IoTD n is the stronger user, that is when the decoding order is $\lambda_m = 0$ and $\lambda_n = 1$, the optimal transmission power of IoTD n and IoTD m can be expressed as

$$p_n^* = P_{n,\max}, \tag{16a}$$

$$p_m^* = \min\left\{\frac{P_{n,\max}|c_n|^2}{\gamma_n^*|c_m|^2} - \frac{\sigma^2}{|c_m|^2}, P_{m,\max}\right\}. \tag{16b}$$

Proof. *Case 1:* If IoTD m is the strong user, its signal is decoded by the BS first, where IoTD n's signal is regarded as interferences. Therefore, in order to maximize the achievable rate of IoTD m, the transmission power of IoTD m should be as large as possible, that is,

$$p_m^* = P_{m,\max}. \tag{17}$$

At the same time, the transmission power of IoTD n should be the minimum value under the condition that the maximum decoding error probability constraint is met, and should not exceed its maximum power. Then we have

$$p_n^* = \min\left\{\frac{\gamma_n^*\sigma^2}{|c_n|^2}, P_{n,\max}\right\}. \tag{18}$$

Case 2: If IoTD m is the weaker user, the IoTD n's signal has been eliminated when the BS decodes s_m. Thus, IoTD n's transmission power will not affect IoTD m's achievable rate. However, to increase the achievable rate of IoTD m by enhancing the transmission power while still meeting the error probability constraint of IoTD n, the transmission power of IoTD n should be as large as possible, that is,

$$p_n^* = P_{n,\max}. \tag{19}$$

Moreover, IoTD m's power should be the maximum value that brings interference to IoTD n but still meets IoTD n's minimum SINR constraint, i.e.,

$$\gamma_n^* = \frac{p_n^*|c_n|^2}{p_m|c_m|^2 + \sigma^2}. \tag{20}$$

Though some algebraic transformation and substituting $p_n^* = P_{n,\max}$ into (20), and taking the maximum power of IoTDs into consideration, we can obtain

$$p_m^* = \min\left\{\frac{P_{n,\max}|c_n|^2}{\gamma_n^*|c_m|^2} - \frac{\sigma^2}{|c_m|^2}, P_{m,\max}\right\}. \tag{21}$$

■

Remark 2. *In a single iteration, constraint (10h) may not be satisfied even if IoTD n transmits signals with its maximum power, which can be caused by poor channel conditions or the unoptimized STAR-RIS coefficient. This problem can be solved by iterative optimization, and the optimization problem will eventually converge to the optimal solution meeting the constraints.*

C. STAR-RIS Coefficient Optimization For ES

The STAR-RIS amplitude response and phase shift coefficient optimization problem will be solved in this subsection, with the given device pairing scheme and IoTDs' transmission power. The mathematical form of STAR-RIS coefficient optimization can be written as

$$(\text{P1.3}): \max_{\{\boldsymbol{\Theta}_r, \boldsymbol{\Theta}_t\}} \sum_{m\in\mathcal{M}, n\in\mathcal{N}} x_{m,n} R_m, \tag{22a}$$

$$\text{s.t.} \quad (10d), (10f), (10g), (10h). \tag{22b}$$

Due to the non-convexity of the objective function and constraints, (P3) is a non-convex optimization problem. It can be easily proved that (8) is the difference between two monotonically increasing concave functions, which makes it difficult to judge its concavity or convexity. Utilizing first-Taylor Expansion, we rewrite the function $g(\gamma_m) = \sqrt{1-\frac{1}{(1+\gamma_m)^2}}$ included in the second term of (8) into a linear function as follows,

$$\tilde{g}(\gamma_m, \tilde{\gamma}_m) = g(\tilde{\gamma}_m) + g'(\tilde{\gamma}_m)(\gamma_m - \tilde{\gamma}_m), \tag{23}$$

where $g'(\gamma_m) = \frac{1}{(1+\gamma_m)^2\sqrt{(1+\gamma_m)^2-1}}$ is the first derivative of $g(\gamma_m)$, $\tilde{\gamma}_m$ is the feasible point. Then the achievable transmission rate of IoTD m can be given by

$$\tilde{R}_m = \log_2(1+\gamma_m) - \kappa_m\left[g(\tilde{\gamma}_m) + g'(\tilde{\gamma}_m)(\gamma_m - \tilde{\gamma}_m)\right], \tag{24}$$

which is an increasing concave function with respect to γ_m. However, $\tilde{R}_m$ is still non-concave to the optimization variables $\boldsymbol{\Theta}_r$ and $\boldsymbol{\Theta}_r$. Thus, we deal with this problem through the following methods.

Let $\mathbf{u}_z = \text{Diag}(\boldsymbol{\Theta}_z) = \left[\sqrt{\beta_1^z}e^{j\theta_1^z}, \cdots, \sqrt{\beta_k^z}e^{j\theta_k^z}, \cdots, \sqrt{\beta_K^z}e^{j\theta_K^z}\right]^T$, $z \in \{r, t\}$ denote the amplitude response and phase shift coefficient vector of the STAR-RIS. Introduce a variable $\boldsymbol{\Phi}_m = \text{diag}\left(\mathbf{g}_{m,s}^H\right)\mathbf{g}_{s,d} \in \mathbb{C}^{K\times 1}$, then the cascaded channel from IoTD m to the destination BS via STAR-RIS can be rewritten as $\mathbf{g}_{m,s}^H\boldsymbol{\Theta}_r\mathbf{g}_{s,d} = \mathbf{u}_r^H\boldsymbol{\Phi}_m$. Thus, the equivalent combined channel gain can be expressed as

$$\begin{aligned}&\left|h_{m,d} + \mathbf{g}_{m,s}^H\boldsymbol{\Theta}_r\mathbf{g}_{s,d}\right|^2 = \left|h_{m,d} + \mathbf{u}_r^H\boldsymbol{\Phi}_m\right|^2 \\ &= \mathbf{u}_r^H\boldsymbol{\Phi}_m\boldsymbol{\Phi}_m^H\mathbf{u}_r + \mathbf{u}_r^H\boldsymbol{\Phi}_m h_{m,d} + h_{m,d}\boldsymbol{\Phi}_m^H\mathbf{u}_r + |h_{m,d}|^2.\end{aligned} \tag{25}$$

Let $\mathbf{Q}_m = \begin{bmatrix} \mathbf{\Phi}_m \mathbf{\Phi}_m^H & \mathbf{\Phi}_m h_{m,d} \\ h_{m,d} \mathbf{\Phi}_m^H & 0 \end{bmatrix} \in \mathbb{C}^{(K+1)\times(K+1)}$, $\bar{\mathbf{u}}_r = \begin{bmatrix} \mathbf{u}_r \\ 1 \end{bmatrix} \in \mathbb{C}^{(K+1)\times 1}$, then we have $\left|h_{m,d} + \mathbf{g}_{m,s}^H \mathbf{\Theta}_r \mathbf{g}_{s,d}\right|^2 = \bar{\mathbf{u}}_r^H \mathbf{Q}_m \bar{\mathbf{u}}_r + |h_{m,d}|^2$. Introduce another variable $\mathbf{U}_r = \bar{\mathbf{u}}_r \bar{\mathbf{u}}_r^H \in \mathbb{C}^{(K+1)\times(K+1)}$, thus we can get $\bar{\mathbf{u}}_r^H \mathbf{Q}_m \bar{\mathbf{u}}_r = \mathrm{Tr}\left(\mathbf{Q}_m \bar{\mathbf{u}}_r \bar{\mathbf{u}}_r^H\right) = \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)$, satisfying $\mathbf{U}_r \succeq 0$ and $\mathrm{rank}\left(\mathbf{U}_r\right) = 1$. Thus,

$$|c_m|^2 = |h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right). \tag{26}$$

Similarly,

$$|c_n|^2 = \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right), \tag{27}$$

where $\mathbf{Q}_n = \mathbf{\Phi}_n \mathbf{\Phi}_n^H \in \mathbb{C}^{K\times K}$ with $\mathbf{\Phi}_n = \mathrm{diag}\left(\mathbf{g}_{n,s}^H\right) \mathbf{g}_{s,d} \in \mathbb{C}^{K\times 1}$, $\mathbf{U}_t = \mathbf{u}_t \mathbf{u}_t^H \in \mathbb{C}^{K\times K}$ satisfying $\mathbf{U}_t \succeq 0$ and $\mathrm{rank}\left(\mathbf{U}_t\right) = 1$. Furthermore, due to the coupling relationship between the amplitude coefficients of reflection and transmission, the following conditions should be met,

$$\mathbf{U}_r(k,k) + \mathbf{U}_t(k,k) = 1, \forall k \in \mathcal{K}. \tag{28}$$

Accordingly, the received SINRs of IoTD m's and n's signal at the BS can be respectively reshaped as

$$\gamma_m = \frac{p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right]}{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2}, \tag{29a}$$

$$\gamma_n = \frac{p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_r\right)}{\lambda_n p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right] + \sigma^2}. \tag{29b}$$

To handle the non-convexity of $\tilde{R}_m$, we introduce two slack variables $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_m, \cdots, \alpha_M]^T$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_m, \cdots, \delta_N]^T$, denote $\tilde{\boldsymbol{\gamma}} = [\tilde{\gamma}_1, \cdots, \tilde{\gamma}_m, \cdots, \tilde{\gamma}_M]^T$, and reformulate the objective function as the following form

$$\begin{aligned} f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right) = \sum_{m\in\mathcal{M}} \{ & \log_2\left(1+\alpha_m\right) \\ & -\kappa_m \left[g\left(\tilde{\gamma}_m\right) + g'\left(\tilde{\gamma}_m\right)\left(\delta_m - \tilde{\gamma}_m\right)\right] \}, \end{aligned} \tag{30}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ should satisfy the following constraints,

$$\alpha_m \le \frac{p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right]}{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2}, \tag{31a}$$

$$\delta_m \ge \frac{p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right]}{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2}. \tag{31b}$$

Lemma 1. *Given $\tilde{\boldsymbol{\gamma}}$, $f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right)$ is a tight lower bound of the original objective function $\sum_{m\in\mathcal{M}} R_m$.*

Proof. Since $g\left(\gamma_m\right) = \sqrt{1 - \frac{1}{\left(1+\gamma_m\right)^2}}$ is concave, based on the first-order condition for a concave function, we have $g\left(\tilde{\gamma}_m\right) + g'\left(\tilde{\gamma}_m\right)\left(\gamma_m - \tilde{\gamma}_m\right) \ge g\left(\gamma_m\right)$, which means $\log_2\left(1+\gamma_m\right) - \kappa_m\left[g\left(\tilde{\gamma}_m\right) + g'\left(\tilde{\gamma}_m\right)\left(\gamma_m - \tilde{\gamma}_m\right)\right] \le \log_2\left(1+\gamma_m\right) - \kappa_m g\left(\gamma_m\right) = R_m$. It is easy to tell that the right side of inequality (31a) and (31b) is actually γ_m. Thus, when the constraints of (31a) and (31b) are satisfied, $f\left(\alpha_m, \delta_m, \mathbf{U}_r, \mathbf{U}_t, \tilde{\gamma}_m\right) \le R_m$, i.e., $f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right) \le \sum_{m\in\mathcal{M}} R_m$. Moreover, when $\gamma_m = \tilde{\gamma}_m$, $\alpha_m = \gamma_m$ and $\delta_m = \gamma_m$, we have $f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right) = \sum_{m\in\mathcal{M}} R_m$, which implies that $f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right)$ is a tight lower bound of $\sum_{m\in\mathcal{M}} R_m$. ■

Therefore, the original optimization problem (P3) can be rewritten as

$$(\text{P1.3.1}) \quad \max_{\{\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t\}} f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right), \tag{32a}$$

$$\text{s.t.} \quad (10h), (28), (31), \tag{32b}$$

$$\begin{cases} \lambda_m = 1, \lambda_n = 0, & \text{if } p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) \\ & \le p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right], \\ \lambda_m = 0, \lambda_n = 1, & \text{if } p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) \\ & > p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right], \end{cases} \tag{32c}$$

$$\mathbf{U}_z \succeq 0, z \in \{r, t\}, \tag{32d}$$

$$\mathrm{rank}\left(\mathbf{U}_z\right) = 1, z \in \{r, t\}. \tag{32e}$$

The decoding error constraint $(10h)$ is intractable due to the existence of Gaussian Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. Since the Gaussian Q-function is monotonically decreasing with x, $\varepsilon \le Q(x)$ is equivalent to $x \ge Q^{-1}(\varepsilon)$. Thus, with given maximum tolerable decoding error probability $\varepsilon_{n,\max}$ and targeted transmission rate $R_{n,\text{target}}$, we recast $(10h)$ into the following form

$$\frac{\log_2\left(1+\gamma_n\right) - R_{n,\text{target}}}{\sqrt{1 - \frac{1}{\left(1+\gamma_n\right)^2}}} \ge \frac{Q^{-1}\left(\varepsilon_{n,\max}\right)}{\sqrt{L_n} \ln 2} \triangleq \kappa_n, \tag{33}$$

which is equivalent to

$$\log_2\left(1+\gamma_n\right) - R_{n,\text{target}} \ge \kappa_n \sqrt{1 - \frac{1}{\left(1+\gamma_n\right)^2}}. \tag{34}$$

To tackle the non-convexity of (34), we first construct approximate linear upper bounds of $\sqrt{1 - \frac{1}{\left(1+\gamma_n\right)^2}}$ by utilizing first-order Taylor expansion

$$\begin{aligned} \mathcal{F}\left(\gamma_n, \hat{\gamma}_n\right) &\triangleq \sqrt{1 - \frac{1}{\left(1+\hat{\gamma}_n\right)^2}} + \frac{1}{\left(1+\hat{\gamma}_n\right)^2 \sqrt{\hat{\gamma}_n^2 + 2\hat{\gamma}_n}} \left(\gamma_n - \hat{\gamma}_n\right) \\ &\ge \sqrt{1 - \frac{1}{\left(1+\gamma_n\right)^2}}, \end{aligned} \tag{35}$$

where $\hat{\gamma}_n$ is the optimal solution obtained in the previous iteration. Thus, we can rewrite (34) as

$$\log_2\left(1+\gamma_n\right) - R_{n,\text{target}} \ge \kappa_n \mathcal{F}\left(\gamma_n, \hat{\gamma}_n\right), n \in \mathcal{N}. \tag{36}$$

Further, to handle the non-convexity of (36), we introduce the slack variables $\{\iota_{n,u}, \iota_{n,d}, \vartheta_n\}$, which satisfy the following constraints

$$2^{\iota_{n,u}} \le p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right] + \lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2, \tag{37a}$$

$$2^{\iota_{n,d}} \ge \lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2, \tag{37b}$$

$$\iota_{n,u} - \iota_{n,d} - R_{n,\text{target}} \ge \vartheta_n, \tag{37c}$$

$$\vartheta_n \ge \kappa_n \mathcal{F}\left(\gamma_n, \hat{\gamma}_n\right). \tag{37d}$$

Through some algebraic operations, we can rewrite (37d) into

$$\gamma_n \leq \frac{\vartheta_n}{\kappa_n} - (1+\hat{\gamma}_n)\left(\hat{\gamma}_n^2 + 2\hat{\gamma}_n\right) + \hat{\gamma}_n \triangleq \frac{\vartheta_n}{\kappa_n} - \varpi, \quad (38)$$

which can be approximately transformed into the following linear form

$$\begin{aligned} & p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_r\right) \\ \leq & \left(\tfrac{\tilde{\vartheta}_n}{\kappa_n} - \varpi\right)\left(\lambda_n p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \tilde{\mathbf{U}}_r\right)\right] + \sigma^2\right) \\ & + \tfrac{1}{\kappa_n}\left(\lambda_n p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \tilde{\mathbf{U}}_r\right)\right] + \sigma^2\right)\left(\vartheta_n - \tilde{\vartheta}_n\right) \\ & + \left(\tfrac{\tilde{\vartheta}_n}{\kappa_n} - \varpi\right)\lambda_n p_m \mathbf{Q}_m^H \left(\mathbf{U}_r - \tilde{\mathbf{U}}_r\right). \end{aligned} \quad (39)$$

However, (37b) is still a non-convex constraint. Utilizing first-order Taylor expansion, we transform it as

$$2^{\hat{\iota}_{n,d}} + 2^{\hat{\iota}_{n,d}} \ln 2 \left(\iota_{n,d} - \hat{\iota}_{n,d}\right) \geq \lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2. \quad (40)$$

To this end, we have transformed the original non-convex constraint (10h) into several convex constraints, i.e., (37a), (37c), (39) and (40). It is noteworthy that due to the existence of the unit rank constraints, $\mathrm{rank}\left(\mathbf{U}_z\right) = 1, z \in \{r, t\}$ in (P3.1), the optimization problem is still non-convex. Thus, by neglecting the rank one constraints, (P3.1) becomes a concave semi-definite programming (SDP) problem. In addition, by reshaping the constraints (31), we can obtain the reformulated optimization problem

$$\begin{aligned} (\text{P1.3.2}) \quad & \max_{\{\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \iota_{n,u}, \iota_{n,d}, \vartheta_n\}} f\left(\boldsymbol{\alpha}, \boldsymbol{\delta}, \mathbf{U}_r, \mathbf{U}_t, \tilde{\boldsymbol{\gamma}}\right), \quad (41a) \\ & \text{s.t.} \quad (28), (32c), (32d), (37a), (37c), (39), (40), \\ & \alpha_m \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2\right\} \\ & \leq p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right], \quad (41b) \\ & p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right] \\ & \leq \delta_m \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \mathbf{U}_t\right) + \sigma^2\right\}. \quad (41c) \end{aligned}$$

However, constraints (41b) and (41c) are still non-convex. Therefore, by utilizing the first-order Taylor expansion of multivariate functions, we can transform constraints (41b) and (41c) into the following linear formulations

$$\begin{aligned} & \tilde{\alpha}_m \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \tilde{\mathbf{U}}_t\right) + \sigma^2\right\} + \tilde{\alpha}_m \lambda_m p_n \mathbf{Q}_n^H \left(\mathbf{U}_t - \tilde{\mathbf{U}}_t\right) \\ & + \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \tilde{\mathbf{U}}_t\right) + \sigma^2\right\}\left(\alpha_m - \tilde{\alpha}_m\right) \\ \leq & p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right], \end{aligned} \quad (42a)$$

$$\begin{aligned} & p_m \left[|h_{m,d}|^2 + \mathrm{Tr}\left(\mathbf{Q}_m \mathbf{U}_r\right)\right] \\ \leq & \tilde{\delta}_m \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \tilde{\mathbf{U}}_t\right) + \sigma^2\right\} + \tilde{\delta}_m \lambda_m p_n \mathbf{Q}_n^H \left(\mathbf{U}_t - \tilde{\mathbf{U}}_t\right) \\ & + \left\{\lambda_m p_n \mathrm{Tr}\left(\mathbf{Q}_n \tilde{\mathbf{U}}_t\right) + \sigma^2\right\}\left(\delta_m - \tilde{\delta}_m\right). \end{aligned} \quad (42b)$$

Accordingly, (P3.2) is transformed into a standard concave problem, which can be efficiently solved via the well-known CVX tool. The rank-one constraint can be recovered by the Gaussian randomization method.

Algorithm 2 Proposed AO algorithm for STAR-RIS with ES Protocol

Input: $\mathbf{x}$, $\mathbf{p}_r$, $\mathbf{p}_t$, $\boldsymbol{\Theta}_r$, $\boldsymbol{\Theta}_t$, maximum iteration times $J_{\max}$, convergence threshold ξ.

Output: $\mathbf{x}^*$, $\mathbf{p}_r^*$, $\mathbf{p}_t^*$, $\boldsymbol{\Theta}_r^*$, $\boldsymbol{\Theta}_t^*$.

1: **Initialization:** $j = 1$, $\mathbf{x}^{(1)}$, $\mathbf{p}_r^{(1)}$, $\mathbf{p}_t^{(1)}$, $\boldsymbol{\Theta}_r^{(1)}$, $\boldsymbol{\Theta}_t^{(1)}$.
2: **repeat**
3: For given $\mathbf{p}_r^{(j)}$, $\mathbf{p}_t^{(j)}$, $\boldsymbol{\Theta}_r^{(j)}$ and $\boldsymbol{\Theta}_t^{(j)}$, update the device pairing scheme according to **Algorithm 1**, and obtain $\mathbf{x}^{(j+1)}$;
4: For given $\mathbf{x}^{(j+1)}$, $\boldsymbol{\Theta}_r^{(j)}$ and $\boldsymbol{\Theta}_t^{(j)}$, calculate and update $\mathbf{p}_r^{(j+1)}$, $\mathbf{p}_t^{(j+1)}$, according to (15) and (16);
5: For given $\mathbf{x}^{(j+1)}$, $\mathbf{p}_r^{(j+1)}$, $\mathbf{p}_t^{(j+1)}$, solve the convex optimization problem (P3.2) in (41) to obtain $\mathbf{U}_r^{(j+1)}$ and $\mathbf{U}_t^{(j+1)}$;
6: Using the Gaussian randomization method to recover the optimal STAR-RIS coefficient $\boldsymbol{\Theta}_r^{(j+1)}$ and $\boldsymbol{\Theta}_t^{(j+1)}$ from $\mathbf{U}_r^{(j+1)}$ and $\mathbf{U}_t^{(j+1)}$;
7: Update $\tilde{\boldsymbol{\alpha}}^{(j)} = \boldsymbol{\alpha}^{(j+1)}$, $\tilde{\boldsymbol{\delta}}^{(j)} = \boldsymbol{\delta}^{(j+1)}$, $\iota_d^{(j)} = \iota_d^{(j+1)}$, $\vartheta^{(j)} = \vartheta^{(j+1)}$, $\tilde{\gamma}_m^{(j)} = \gamma_m^{(j+1)}$;
8: $j \leftarrow j + 1$;
9: **until** $j = J_{\max}$ or $\left|\sum_{m\in\mathcal{M}} R_m^{(j)} - \sum_{m\in\mathcal{M}} R_m^{(j-1)}\right| \leq \xi$.

Lemma 2. *Denote $\boldsymbol{\alpha}^*$, $\boldsymbol{\delta}^*$, $\mathbf{U}_r^*$ and $\mathbf{U}_t^*$ are the optimal solutions to (P1.3.2), $\sum_{m\in\mathcal{M}} R_m^*$ are calculated with the obtained $\boldsymbol{\gamma}^*$, and $\sum_{m\in\mathcal{M}} R_m'$ are calculated with $\tilde{\boldsymbol{\gamma}}$. With given $\tilde{\boldsymbol{\gamma}}$, we have $\sum_{m\in\mathcal{M}} R_m^* \geq \sum_{m\in\mathcal{M}} R_m'$.*

Proof. Based on **Lemma 1**, we have $\sum_{m\in\mathcal{M}} R_m^* \geq f\left(\boldsymbol{\alpha}^*, \boldsymbol{\delta}^*, \mathbf{U}_r^*, \mathbf{U}_t^*, \tilde{\boldsymbol{\gamma}}\right)$. When $\tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\delta}} = \tilde{\boldsymbol{\gamma}}$, we have $f\left(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\delta}}, \tilde{\mathbf{U}}_r, \tilde{\mathbf{U}}_t, \tilde{\boldsymbol{\gamma}}\right) = \sum_{m\in\mathcal{M}} R_m'$. Furthermore, since $\boldsymbol{\alpha}^*$, $\boldsymbol{\delta}^*$, $\mathbf{U}_r^*$ and $\mathbf{U}_t^*$ are the optimal solutions, we have $f\left(\boldsymbol{\alpha}^*, \boldsymbol{\delta}^*, \mathbf{U}_r^*, \mathbf{U}_t^*, \tilde{\boldsymbol{\gamma}}\right) \geq f\left(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\delta}}, \tilde{\mathbf{U}}_r, \tilde{\mathbf{U}}_t, \tilde{\boldsymbol{\gamma}}\right)$. Therefore, $\sum_{m\in\mathcal{M}} R_m^* \geq \sum_{m\in\mathcal{M}} R_m'$. ■

D. Overall Algorithm for ES

We propose an alternating optimization (AO) based algorithm to perform the joint optimization for maximizing the sum FBL rate of R-IoTDs, as summarized in **Algorithm 2**. Specifically, we first propose an RSS-based device pairing scheme, then obtain the closed-form solution of power allocation. The STAR-RIS reflecting and transmitting coefficient optimization problem is solved by transforming it into a convex problem. Finally, the solution to the initial joint optimization problem is obtained by an alternative and iterative method, where the initial point of each iteration is the solution in the previous iteration.

The device pairing scheme involves finding the maximum or minimum values of M and N ($M = N$), leading to the computational complexity of $\mathcal{O}\left(2M^2\right)$. The computational complexity of solving the power allocation problem is $\mathcal{O}\left(1\right)$ with the given device pair since the closed-form solution is obtained. Thus, the computational complexity of

power allocation for all device pairs can be computed as $\mathcal{O}(2M)$. The computational complexity of solving the STAR-RIS coefficient optimization problem mainly comes from the SDP problem, which can be approximately calculated as $\mathcal{O}\left(\log\left(\frac{1}{\zeta}\right)\left(2K^{4.5}+21MK^{3.5}\right)\right)$, where $\zeta > 0$ is the convergence accuracy of SDP [38, Theorem 3.12]. Therefore, the overall polynomial complexity of the proposed algorithm can be approximately calculated as $\mathcal{O}\left(I_2\left\{2M^2+\log\left(\frac{1}{\zeta}\right)\left(2K^{4.5}+21MK^{3.5}\right)\right\}\right)$, with I_2 representing the number of iterations required for the convergence of **Algorithm 2**.

IV. HRLE Design with MS Protocol

In this section, we will solve the HRLE design problem for STAR-RIS with MS protocol. In contrast to ES scheme, the MS protocol only allows the amplitude response value of STAR-RIS to be 0 or 1 instead of any value from 0 to 1, which has the advantage of easy implementation. However, the binary constraint will also make the optimization problem intractable due to its non-convexity. The mathematical form of the HRLE problem with MS can be expressed as

$$\begin{aligned}(\mathrm{P2}):\ &\max_{\{\mathbf{x},\mathbf{p}_r,\mathbf{p}_t,\boldsymbol{\Theta}_r,\boldsymbol{\Theta}_t\}} \sum_{m\in\mathcal{M},n\in\mathcal{N}} x_{m,n}R_m, && (43a)\\ &\text{s.t. } (10b)-(10f),(10g),(10h), && (43b)\\ &\quad \beta_k^r,\beta_k^t\in\{0,1\},\beta_k^r+\beta_k^t=1,\forall k\in\mathcal{K}, && (43c)\end{aligned}$$

where (43c) is the amplitude constraint for STAR-RIS with MS protocol. By comparing (10) and (43), we can see that the difference only exists in the constraints on the STAR-RIS amplitude coefficient. Therefore, the proposed RSS-based device pairing scheme in Section III-A and the obtained closed-form solutions for power allocation in Section III-A are still applicable. Thus, we focus on addressing the STAR-RIS coefficient optimization problem in this section.

A. STAR-RIS Coefficient Optimization For MS

With given device pairs and transmission power, the transmitting and reflecting coefficient optimization problem for STAR-RIS with MS protocol can be given by

$$\begin{aligned}(\mathrm{P2.1}):\ &\max_{\{\boldsymbol{\Theta}_r,\boldsymbol{\Theta}_t\}} \sum_{m\in\mathcal{M},n\in\mathcal{N}} x_{m,n}R_m, && (44a)\\ &\text{s.t. } (10d),(10f),(10g),(10h), && (44b)\\ &\quad \beta_k^r,\beta_k^t\in\{0,1\},\forall k\in\mathcal{K}. && (44c)\end{aligned}$$

It is easy to tell that compared with the STAR-RIS coefficient optimization problem (22) with ES protocol, (44) involves the additional non-convex binary constraints (44c). Therefore, we only need to focus on dealing with this new difficulty since other non-convex terms can be handled by a similar method given in Section III-C. First, we transform the binary constraint (44c) into its equivalent form

$$\beta_k^z\left(\beta_k^z-1\right)=0,\forall z\in\{t,r\},k\in\mathcal{K}. \quad (45)$$

By adopting the similar processing method of the objective function and other constraints in Section III-C, and introducing (45) as a penalty form into the objective function in (41), the reformulated optimization problem is

$$\begin{aligned}(\mathrm{P2.2})\ &\max_{\{\boldsymbol{\alpha},\boldsymbol{\delta},\mathbf{U}_r,\mathbf{U}_t,\iota_{n,u},\iota_{n,d},\vartheta_n\}} f(\boldsymbol{\alpha},\boldsymbol{\delta},\mathbf{U}_r,\mathbf{U}_t)\\ &\qquad +\eta\sum_{z\in\{t,r\}}\sum_{k\in\mathcal{K}}\beta_k^z\left(\beta_k^z-1\right),\\ &\text{s.t. } (28),(32c),(32d),(37a),(37c),\\ &\quad (39),(40),(42a),(42b),\end{aligned} \quad (46)$$

where $\eta>0$ is the penalty factor.

Lemma 3. *The equality constraint (45) can be satisfied as* $\eta\to\infty$.

Proof. Please refer to [39]. ■

However, the resulting optimization problem (P3.1′) is still non-convex. Thus, we construct the lower bound of the penalty form by utilizing the first-order Taylor expansion

$$\begin{aligned}&\beta_k^z\left(\beta_k^z-1\right)\geq\hat{\beta}_k^z\left(\hat{\beta}_k^z-1\right)+\left(2\hat{\beta}_k^z-1\right)\left(\beta_k^z-\hat{\beta}_k^z\right)\\ &=\left(2\hat{\beta}_k^z-1\right)\beta_k^z-\left(\hat{\beta}_k^z\right)^2,\end{aligned} \quad (47)$$

where $\hat{\beta}_k^z$ is a feasible point in each iteration of the SCA method. By introducing (47) into the objective function of (46), we can obtain the reconstructed optimization problem

$$\begin{aligned}(\mathrm{P2.3})\ &\max_{\{\boldsymbol{\alpha},\boldsymbol{\delta},\mathbf{U}_r,\mathbf{U}_t,\iota_{n,u},\iota_{n,d},\vartheta_n\}} f(\boldsymbol{\alpha},\boldsymbol{\delta},\mathbf{U}_r,\mathbf{U}_t)\\ &\qquad +\eta\sum_{z\in\{t,r\}}\sum_{k\in\mathcal{K}}\left[\left(2\hat{\beta}_k^z-1\right)\beta_k^z-\left(\hat{\beta}_k^z\right)^2\right],\\ &\text{s.t. } (28),(32c),(32d),(37a),(37c),\\ &\quad (39),(40),(42a),(42b),\end{aligned} \quad (48)$$

which is convex and can be efficiently solved via standard convex problem solvers such as CVX.

B. Overall Algorithm for MS

To solve the sum FBL rate maximization problem for STAR-RIS with MS protocol, we propose a double-layer penalty-based (DLPB) algorithm as concluded in **Algorithm 3**. Specifically, the *inner loop* is AO with introduced penalty terms, where the device-pairing, power allocation and relaxed STAR-RIS coefficient are optimized in an alternative manner similar to **Algorithm 2**, while the *outer loop* is for the binary amplitude constraint recovery. Similar to the analysis given in Section III-D, the overall computational complexity can be calculated as $\mathcal{O}\left(I_3I_2\left\{2M^2+\log\left(\frac{1}{\zeta}\right)\left(2K^{4.5}+21MK^{3.5}\right)\right\}\right)$, where I_3 denotes the number of iterations required for the convergence of **Algorithm 3**.

C. Comparison of AO and DLPB

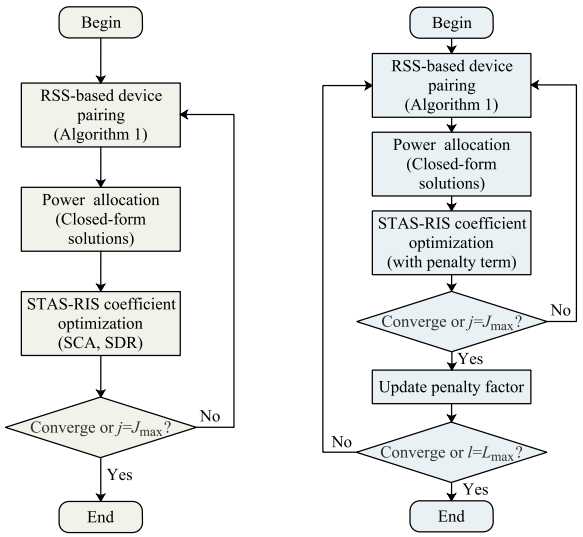
To illustrate the proposed algorithms clearly, we summarize them into schematic diagrams. The proposed AO algorithm for ES protocol is shown in Fig. 3 (a), where three sub-problems

Algorithm 3 Proposed DLPB Algorithm for STAR-RIS with MS Protocol

Input: $\mathbf{x}$, $\mathbf{p}_r$, $\mathbf{p}_t$, $\mathbf{\Theta}_r$, $\mathbf{\Theta}_t$, maximum iteration times $J_{\max}$, $L_{\max}$, convergence threshold ξ, ξ_2.

Output: $\mathbf{x}^*$, $\mathbf{p}_r^*$, $\mathbf{p}_t^*$, $\mathbf{\Theta}_r^*$, $\mathbf{\Theta}_t^*$.

1: **Initialization:** $j=1$, $l=1$, $\mathbf{x}^{(1)}$, $\mathbf{p}_r^{(1)}$, $\mathbf{p}_t^{(1)}$, $\mathbf{\Theta}_r^{(1)}$, $\mathbf{\Theta}_t^{(1)}$, penalty factor η, discount factor $0<c<1$.
2: **repeat**
3: Perform steps 2-9 in **Algorithm 2**. Specifically, step 5 becomes solving the problem (P3.2′) in (48);
4: Update $\mathbf{x}^{(1)}$, $\mathbf{p}_r^{(1)}$, $\mathbf{p}_t^{(1)}$, $\mathbf{\Theta}_r^{(1)}$, $\mathbf{\Theta}_t^{(1)}$ with the current solutions $\mathbf{x}^{(j)}$, $\mathbf{p}_r^{(j)}$, $\mathbf{p}_t^{(j)}$, $\mathbf{\Theta}_r^{(j)}$, $\mathbf{\Theta}_t^{(j)}$, $j=1$;
5: Update $\eta^{(l)}=\eta^{(l-1)}/c$;
6: $l \leftarrow l+1$;
7: **until** $l = L_{\max}$ or $\left|\max\limits_{k,z}\beta_k^{z(l)}\left(\beta_k^{z(l)}-1\right)-\max\limits_{k,z}\beta_k^{z(l-1)}\left(\beta_k^{z(l-1)}-1\right)\right| \le \xi_2$.



(a) AO algorithm for ES (b) DLPB algorithm for MS

Fig. 3. Illustration for the proposed algorithms.

are solved alternatively until the convergence condition or maximum number of iterations is reached. The proposed DLPB algorithm for MS protocol is shown in Fig. 3 (b). Specifically, the inner layer of the DLPB algorithm is similar to the AO algorithm with introduced penalty terms, while the outer layer is for the binary constraint recovery by exploiting penalty-based optimization[3].

V. Simulation Results

In this section, we evaluate the performance of our proposed AO algorithm and DLPB algorithm for the STAR-RIS-assisted uplink NOMA system via simulations[4]. Consider a three-dimensional coordinate system, where the BS and the STAR-RIS are respectively located at (0, 0, 10) meter (m) and (40, 0, 5) m. The R-IoTDs are randomly located in a circle with a radius of 5 m centered on (35, 0, 0) m, and the T-IoTDs are randomly located in a circle with a radius of 3 m centered on (45, 0, 5) m. The simulation results are obtained through an average of more than 500 channel realizations. The parameters used in the simulations are given in **TABLE II**. A stringent tolerable decoding error probability, i.e., 10^{-5}, is considered to meet the requirements of mission-critical IoT services [41]. Moreover, a relatively short blocklength, i.e., 200 channel uses, is considered to reduce the transmission latency while ensuring reliability to a certain degree. To evaluate the effectiveness of the proposed algorithms, we compare them with those schemes of information-directed branch-and-prune (IDBP) algorithm proposed in [42], conventional reflecting/transmitting-only RISs with $K/2$ elements (C-RIS), random amplitude and phase coefficient at STAR-RIS (Random-STARS), OMA scheme, fixed transmission power at IoTDs with their maximum power (Fixed-P).

TABLE II
Simulation Parameters

Parameter	Setting
noise power σ^2 (dBm)	-110
path loss at the reference distance $d=1m$	-30dB
path loss exponents $\alpha_{m,s}$, $\alpha_{n,s}$	2.5
path loss exponents $\alpha_{m,d}$	4
path loss exponents $\alpha_{s,d}$	2.2
Rician factor	3dB
number of STAR-RIS elements	20
number of IoTDs on *R/T-Space*	2
maximum transmission power of IoTDs (dBm)	23
blocklength L (channel uses)	200
tolerable decoding error probability ε	10^{-5}
convergence threshold ξ	10^{-3}

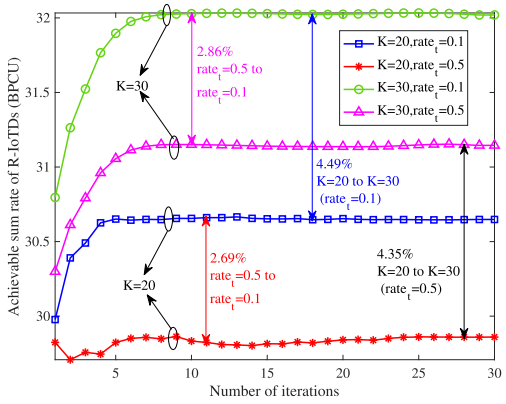


Fig. 4. Convergence performance of the AO algorithm for STAR-RIS with ES protocol.

[3]TS operating protocol is not considered is this paper because STAR-RISs with TS protocol can only provide "smart radio environment" for IoTDs on one single side at a time.

[4]In this paper, we focus on the beamforming design of small-scale STAR-RIS networks to reveal the potential of STAR-RIS-assisted uplink NOMA transmission. However, by partitioning the STAR-RIS into several tiles, a scalable algorithm can be utilized to handle the high computational complexity cased by massive elements of STAR-RIS [40], which can provide inspiration and guidance for our future works on large-scale STAR-RIS.

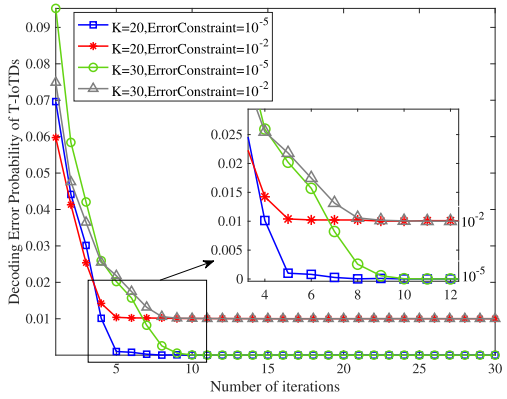


Fig. 5. Converged error probability of T-IoTDs obtained by the AO algorithm ($\text{rate}_t = 0.1$).

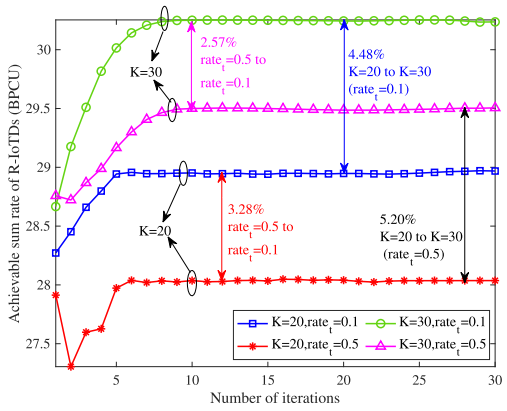


Fig. 6. Convergence performance of the DLPB algorithm for STAR-RIS with MS protocol.

The convergence performance of the proposed AO algorithm for STAR-RIS with ES protocol is evaluated in Fig. 4, where rate_t is the targeted transmission rate of T-IoTDs. It can be seen that more iteration times are required for the algorithm convergence when the number of STAR-RIS elements is 30, i.e., $K = 30$, compared with $K = 20$, where the former needs about 6 iterations while the latter needs about 10 iterations. The achievable sum rates of R-IoTDs increase by about 4.49% and 4.35% when the number of STAR-RIS elements increases from 20 to 30, which owes to the higher transmission/reflection beamforming gain brought by more elements. Furthermore, when the T-IoTDs have looser target rate requirements, more communication resources will be allocated to the R-IoTDs, resulting in a higher achievable sum rate.

In Fig. 5, we depict the converged error probability of T-IoTDs obtained by the AO algorithm with different decoding error probability constraints. Although the decoding error probability constraints are positive during the whole iteration process, the inappropriate device paring or power allocation may cause a higher decoding error probability than the target one. However, the decoding error probability of T-IoTDs will eventually converge to a value that satisfies the constraints, which verifies **Remark 2**. Moreover, more iteration times are required for $K = 30$ compared with $K = 20$.

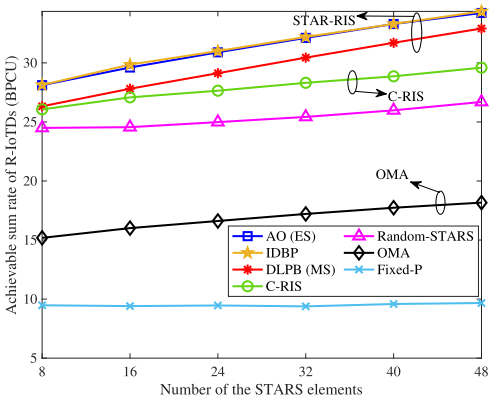


Fig. 7. The achievable sum rate of R-IoTDs versus number of STAR-RIS elements.

In Fig. 6, we evaluate the convergence performance of the proposed DLPB algorithm for STAR-RIS with MS protocol. Note that the depicted figure is the outer loop iteration, so the number of times required for convergence of the algorithm is not significantly more than that of ES protocol. Similarly, more iteration times are required for the algorithm convergence when $K = 30$ compared with $K = 20$, and the achievable rate of R-IoTDs will be improved by increasing the number of STAR-RIS elements or relaxing the target rate constraint of T-IoTDs. In addition, by comparing Fig. 6 with Fig. 4, we can tell that compared with ES protocol, there is a performance loss when the STAR-RIS adopts MS protocol, which is caused by the constrained amplitude coefficient.

The relationship between the achievable sum rate of R-IoTDs and the number of STAR-RIS elements is depicted in Fig. 7. It can be seen that the achievable sum rates obtained by the proposed AO algorithm and the IDBP algorithm are almost coincident. Since the effectiveness of IDBP algorithm has been verified by the simulation results in [42], where the performance obtained by IDBP is close to the exhaustive search with significantly reduced complexity, we can conclude that the proposed AO algorithm is also near-optimal. When the Fixed-P scheme is adopted, that is, when each IoTD sends signals at its maximum power, the sum rate almost does not change with the number of STAR-RIS elements. This shows the importance of the joint optimization and cooperation of various communication resources, otherwise, it is difficult to exploit the potential performance gains brought by more STAR-RIS elements. Except for the Fixed-P scheme, the sum rate of R-IoTDs obtained by all other schemes increases with the number of STAR-RIS elements due to the higher array gain. Compared with the conventional RISs, random STAR-RIS coefficient, OMA and Fixed-P schemes, the achievable sum rate of R-IoTDs obtained by the proposed AO algorithm for STAR-RIS with ES protocol increased by 11.75%, 23.65%, 85.96% and 226.56% respectively when $K = 24$, which verifies the effectiveness of the proposed scheme. To be specific, due to the fixed number of transmitting and reflecting elements, the conventional RISs cannot exploit the same degrees of freedom as STAR-RIS, leading to a worse ability to intelligently adjust the wireless radio environment.

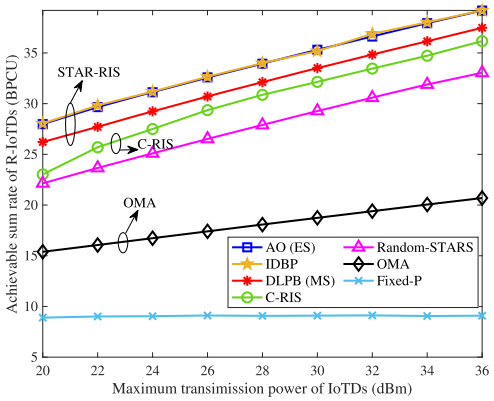


Fig. 8. The achievable sum rate of R-IoTDs versus maximum power of IoTDs.

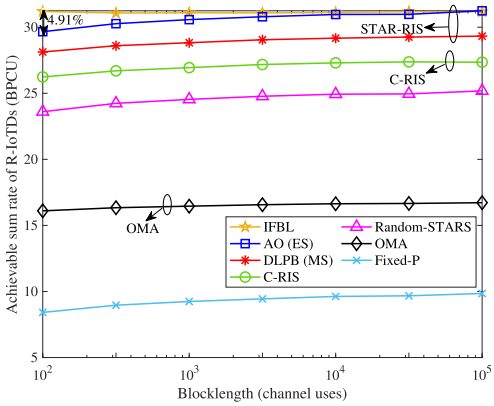


Fig. 9. The achievable sum rate of R-IoTDs versus blocklength.

Furthermore, Fig. 7 also shows that STAR-RIS with ES protocol can achieve a better transmission performance than that with MS protocol at the cost of implement complexity.

We compare the achievable sum rate of R-IoTDs versus the maximum power of IoTDs under different algorithms in Fig. 8. Regardless of the maximum transmission power of IoTDs, the proposed scheme almost coincides with IDBP algorithm, and is always superior to other comparison schemes. As expected, we can see that the achievable sum rate of R-IoTDs enhances with the increase of maximum transmission power of IoTDs, except for the Fixed-P scheme. In order to obtain better FBL transmission performance of R-IoTDs, the transmission power of T-IoTDs of other schemes will be selected as a minimum value that guarantees the decoding correctness constraints, no matter what the maximum power is. However, the Fixed-P scheme increases the transmission power of R-IoTDs and T-IoTDs at the same time, which results in a less obvious increase in the transmission rate of R-IoTDs due to higher interference. In addition, with the increase of transmission power, the performance gap between OMA scheme and the proposed scheme becomes more and more obvious, because NOMA can make better use of the difference in the received signal strength to obtain greater performance gain than OMA.

The impact of short-packet blocklength is depicted in Fig. 9. The proposed AO scheme for STAR-RIS with ES protocol always has the maximum sum rate, which confirms the superiority of the proposed scheme. Furthermore, the achievable sum rate of all schemes increases with the blocklength, and gradually approaches IFBL transmission. As shown in Fig. 9, the performance gap between FBL and the IFBL transmission is 4.91% when $L = 100$, and decreases to 0 when $L = 10^5$. Although the transmission scheme with IFBL can improve the achievable rate of the system, it will also cause greater end-to-end delay, which is unfriendly to latency-sensitive IoT services. It is not difficult to tell from Fig. 9 that the transmission rate reaches a relatively stable value when the blocklength reaches a certain level, and there is almost no use to increase the transmission rate by increasing the blocklength. The performance tradeoff between achievable transmission rate and latency encourages us to investigate the optimal blocklength, which may be our future research direction.

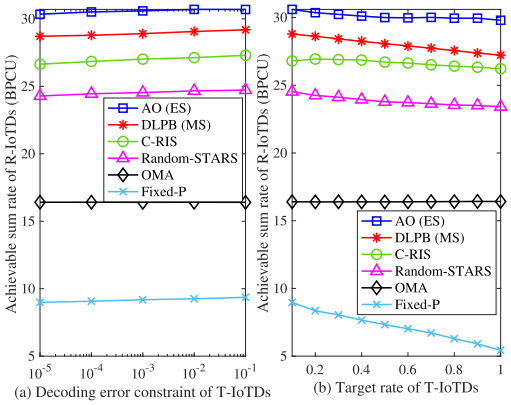


Fig. 10. The achievable sum rate of R-IoTDs versus maximum decoding error constraint and target rate of T-IoTDs.

The FBL transmission performance of R-IoTDs under different T-IoTDs' maximum decoding error constraints and required target rates is shown in Fig. 10. Except for OMA scheme, the achievable sum rates of R-IoTDs increase with T-IoTDs' maximum decoding error constraint while decrease with the required target rate. More communication resources, including transmission power and STAR-RIS amplitude response, will be needed to satisfy the stricter T-IoTDs' requirements (higher required target rate), leading to the decline of FBL transmission performance of R-IoTDs, while it is the opposite when the constraint is looser (higher tolerable decoding error probability). For OMA scheme, the change of the maximum decoding error constraint or the target rate of the T-IoTDs does not have a significant impact on the achievable sum rate of R-IoTDs, because the IoTDs on both sides of the STAR-RIS occupy different subchannels and will not have inter-device interferences. Furthermore, the proposed scheme always has the best transmission performance compared with all those schemes, confirming the effectiveness of adopting STAR-RIS and NOMA in FBL transmission systems.

VI. Conclusions

In this article, a STAR-RIS-assisted uplink NOMA transmission framework for FBL transmission was proposed, where the IoTDs on two sides of the STAR-RIS have different communication requirements. For improving the FBL trans-

mission performance of the R-IoTDs while satisfying the decoding reliability of the T-IoTDs, a sum rate maximization problem was formulated. For STAR-RIS with ES protocol, an AO algorithm was proposed to solve the highly-coupled non-convex optimization problem alternatively. For STAR-RIS with MS protocol, a DLPB algorithm was proposed to tackle the newly introduced binary amplitude constraints. The simulation results show that the proposed STAR-RIS framework outperforms the conventional transmitting/reflecting-only RISs. Furthermore, adopting NOMA can enhance the FBL transmission performance compared with the OMA schemes. In addition, the tolerable decoding error probabilities and the targeted transmission rates have a significant impact on the performance of FBL transmission, which inspires us to investigate the tradeoff between capacity, latency and reliability.

Appendix A
Proof of Theorem 1

According to *Theorem 1*, for any subsets $\{w_{1_r}, w_{2_r}\} \in \mathcal{M}_u$ and $\{w_{1_t}, w_{2_t}\} \in \mathcal{N}_u$ satisfying $p_{1_r}|c_{1_r}|^2 \geq p_{2_r}|c_{2_r}|^2$ and $p_{1_t}|c_{1_t}|^2 \geq p_{2_t}|c_{2_t}|^2$, the optimal device clustering scheme for maximizing sum rate of IoTDs on *R-Space* is the following

$$\Pi_1 = \{(w_{1_r}, w_{2_t}), (w_{2_r}, w_{1_t})\}. \tag{49}$$

Accordingly, the sum rate of w_{1_r} and w_{2_r} can be given by

$$\begin{aligned} R_{sum} &= \log_2\left(1+\frac{p_{1_r}|c_{1_r}|^2}{p_{2_t}|c_{2_t}|^2+1}\right)+\log_2\left(1+\frac{p_{2_r}|c_{2_r}|^2}{p_{1_t}|c_{1_t}|^2+1}\right) \\ &-\kappa_{1_r}\sqrt{1-\frac{1}{\left(1+\frac{p_{1_r}|c_{1_r}|^2}{p_{2_t}|c_{2_t}|^2+1}\right)^2}}-\kappa_{2_r}\sqrt{1-\frac{1}{\left(1+\frac{p_{2_r}|c_{2_r}|^2}{p_{1_t}|c_{1_t}|^2+1}\right)^2}}. \end{aligned} \tag{50}$$

Assume the contrary, if $\Pi_1 = \{(w_{1_r}, w_{2_t}), (w_{2_r}, w_{1_t})\}$ is not the optimal set, then the device pairing scheme is given by

$$\Pi_2 = \{(w_{1_r}, w_{1_t}), (w_{2_r}, w_{2_t})\}. \tag{51}$$

Accordingly, the sum transmission rate of w_{1_r} and w_{2_r} is

$$\begin{aligned} R^{'}_{\text{sum}} &= \log_2\left(1+\frac{p_{1_r}|c_{1_r}|^2}{p_{1_t}|c_{1_t}|^2+1}\right)+\log_2\left(1+\frac{p_{2_r}|c_{2_r}|^2}{p_{2_t}|c_{2_t}|^2+1}\right) \\ &-\kappa_{1_r}\sqrt{1-\frac{1}{\left(1+\frac{p_{1_r}|c_{1_r}|^2}{p_{1_t}|c_{1_t}|^2+1}\right)^2}}-\kappa_{2_r}\sqrt{1-\frac{1}{\left(1+\frac{p_{2_r}|c_{2_r}|^2}{p_{2_t}|c_{2_t}|^2+1}\right)^2}}. \end{aligned} \tag{52}$$

Under a high SINR regime, we consider that the short-packet loss terms can be regarded as constant values, i.e., κ_{1_r} and κ_{2_r}. Then we have

$$\begin{aligned} R_{\text{sum}} - R^{'}_{\text{sum}} = &\log_2\left(p_{1_r}|c_{1_r}|^2+p_{2_t}|c_{2_t}|^2+1\right) \\ &+\log_2\left(p_{2_r}|c_{2_r}|^2+p_{1_t}|c_{1_t}|^2+1\right) \\ &-\log_2\left(p_{1_r}|c_{1_r}|^2+p_{1_t}|c_{1_t}|^2+1\right) \\ &-\log_2\left(p_{2_r}|c_{2_r}|^2+p_{2_t}|c_{2_t}|^2+1\right). \end{aligned} \tag{53}$$

To simplify writing, introduce variables $a = p_{1_r}|c_{1_r}|^2 + p_{2_t}|c_{2_t}|^2 + 1$, $b = p_{2_r}|c_{2_r}|^2 + p_{1_t}|c_{1_t}|^2 + 1$, $c = p_{1_r}|c_{1_r}|^2 + p_{1_t}|c_{1_t}|^2 + 1$ and $d = p_{2_r}|c_{2_r}|^2 + p_{2_t}|c_{2_t}|^2 + 1$. Obviously, we have $a + b = c + d$. Furthermore,

$$\begin{aligned} &|a-b|-|c-d| \\ &=\left|\left(p_{1_r}|c_{1_r}|^2-p_{2_r}|c_{2_r}|^2\right)-\left(p_{1_t}|c_{1_t}|^2-p_{2_t}|c_{2_t}|^2\right)\right| \\ &\quad-\left|\left(p_{1_r}|c_{1_r}|^2-p_{2_r}|c_{2_r}|^2\right)+\left(p_{1_t}|c_{1_t}|^2-p_{2_t}|c_{2_t}|^2\right)\right| \\ &\overset{(1)}{\leq} 0, \end{aligned} \tag{54}$$

where (1) holds due to $\left(p_{1_r}|c_{1_r}|^2-p_{2_r}|c_{2_r}|^2\right) \geq 0$ and $\left(p_{1_t}|c_{1_t}|^2-p_{2_t}|c_{2_t}|^2\right) \geq 0$, because if $x \geq 0$ and $y \geq 0$, then $|x-y| \leq |x+y|$. Thus, we can get the following inequality

$$\begin{aligned} &(a+b)^2-(a-b)^2 \geq (c+d)^2-(c-d)^2 \\ &\Rightarrow ab \geq cd. \end{aligned} \tag{55}$$

Therefore,

$$R_{\text{sum}} - R^{'}_{\text{sum}} = \log_2\frac{ab}{cd} \geq 0, \tag{56}$$

which means that the device pairing scheme $\Pi_1 = \{(w_{1_r}, w_{2_t}), (w_{2_r}, w_{1_t})\}$ is superior to $\Pi_2 = \{(w_{1_r}, w_{1_t}), (w_{2_r}, w_{2_t})\}$.

References

[1] W. Issa, N. Moustafa, B. Turnbull, N. Sohrabi, and Z. Tari, "Blockchain-based federated learning for securing internet of things: A comprehensive survey," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–43, 2023.

[2] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, 2022.

[3] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 23, no. 3, pp. 1622–1658, 2021.

[4] R. Hassan, F. Qamar, M. K. Hasan, A. H. M. Aman, and A. S. Ahmed, "Internet of things and its applications: A comprehensive survey," *Symmetry*, vol. 12, no. 10, 2020. [Online]. Available: https://www.mdpi.com/2073-8994/12/10/1674

[5] B. S. Khan, S. Jangsher, A. Ahmed, and A. Al-Dweik, "URLLC and eMBB in 5G industrial IoT: A survey," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1134–1163, 2022.

[6] A. Filali, Z. Mlika, S. Cherkaoui, and A. Kobbane, "Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2174–2187, 2022.

[7] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, 2021.

[8] 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.

[9] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, 2021.

[10] W. Khalid, Z. Kaleem, R. Ullah, T. V. Chien, S. Noh, and H. Yu, "Simultaneous transmitting and reflecting-reconfigurable intelligent surface in 6G: Design guidelines and future perspectives," *IEEE Network*, pp. 1–9, 2022.

[11] J. Östman, A. Lancho, G. Durisi, and L. Sanguinetti, "URLLC with massive MIMO: Analysis and design at finite blocklength," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6387–6401, 2021.

[12] S. He, Z. An, J. Zhu, J. Zhang, Y. Huang, and Y. Zhang, "Beamforming design for multiuser uRLLC with finite blocklength transmission," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8096–8109, 2021.

[13] X. Zhang, J. Wang, and H. V. Poor, "Statistical delay and error-rate bounded QoS provisioning for mURLLC over 6G CF M-MIMO mobile networks in the finite blocklength regime," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 652–667, 2021.

[14] Y. Xu, C. Shen, D. Cai, and G. Zhu, "Latency constrained non-orthogonal packets scheduling with finite blocklength codes," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12 312–12 316, 2020.
[15] Y. Hu, Y. Li, M. C. Gursoy, S. Velipasalar, and A. Schmeink, "Throughput analysis of low-latency IoT systems with QoS constraints and finite blocklength codes," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3093–3104, 2020.
[16] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, 2017.
[17] L. Lyu, C. Chen, N. Cheng, S. Zhu, X. Guan, and X. Shen, "NOMA-assisted on-demand transmissions for monitoring applications in industrial IoT networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12 264–12 276, 2020.
[18] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, 2018.
[19] J. Yao, Q. Zhang, and J. Qin, "Joint decoding in downlink NOMA systems with finite blocklength transmissions for ultrareliable low-latency tasks," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17 705–17 713, 2022.
[20] H. Ren, K. Wang, and C. Pan, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 707–723, 2022.
[21] 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, no. 10, pp. 10 320–10 334, 2021.
[22] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-Aho, "Joint sum rate and blocklength optimization in RIS-aided short packet URLLC systems," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1838–1842, 2022.
[23] Q. Chen, M. Li, X. Yang, R. Alturki, M. D. Alshehri, and F. Khan, "Impact of residual hardware impairment on the IoT secrecy performance of RIS-assisted NOMA networks," *IEEE Access*, vol. 9, pp. 42 583–42 592, 2021.
[24] M. Elhattab, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "RIS-assisted joint transmission in a two-cell downlink NOMA cellular system," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1270–1286, 2022.
[25] Y. Xiu, J. Zhao, W. Sun, M. D. Renzo, G. Gui, Z. Zhang, and N. Wei, "Reconfigurable intelligent surfaces aided mmWave NOMA: Joint power allocation, phase shifts, and hybrid beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8393–8409, 2021.
[26] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1057–1071, 2021.
[27] C. Wu, X. Mu, Y. Liu, X. Gu, and X. Wang, "Resource allocation in STAR-RIS-aided networks: OMA and NOMA," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7653–7667, 2022.
[28] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (STAR) RIS assisted NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 611–626, 2023.
[29] T. Hou, J. Wang, Y. Liu, X. Sun, A. Li, and B. Ai, "A joint design for STAR-RIS enhanced NOMA-CoMP networks: A simultaneous-signal-enhancement-and-cancellation-based (SSECB) design," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1043–1048, 2022.
[30] F. Karim, S. K. Singh, K. Singh, S. Prakriya, and M. F. Flanagan, "On the performance of STAR-RIS-aided NOMA at finite blocklength," *IEEE Wireless Commun. Lett.*, 2023.
[31] J. Xu, X. Mu, J. T. Zhou, and Y. Liu, "Simultaneously transmitting and reflecting (STAR)-RISs: Are they applicable to dual-sided incidence?" *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 129–133, 2023.
[32] H. Ma, H. Wang, H. Li, and Y. Feng, "Transmit power minimization for STAR-RIS-empowered uplink NOMA system," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2430–2434, 2022.
[33] Z. Zhang, J. Chen, Y. Liu, Q. Wu, B. He, and L. Yang, "On the secrecy design of STAR-RIS assisted uplink NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11 207–11 221, 2022.
[34] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, A. Shojaeifard, Y. Liu, and L. Hanzo, "Achievable rate analysis of the STAR-RIS aided NOMA uplink in the face of imperfect CSI and hardware impairments," *IEEE Trans. Commun., early access*, 2023.
[35] P. Wang, H. Wang, and Y. Fu, "Average rate maximization for mobile STAR-RIS-aided NOMA system," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1362–1366, 2023.
[36] M. Iwabuchi, A. Benjebbour, Y. Kishiyama, G. Ren, C. Tang, T. Tian, L. Gu, T. Takada, and T. Kashima, "5G field experimental trials on URLLC using new frame structure," in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.
[37] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
[38] I. Pólik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear optimization*. Springer, 2010, pp. 215–276.
[39] A. Ben-Tal and M. Zibulevsky, "Penalty/barrier multiplier methods for convex programming problems," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 347–366, 1997.
[40] D. Xu, V. Jamali, X. Yu, D. W. K. Ng, and R. Schober, "Optimal resource allocation design for large IRS-assisted SWIPT systems: A scalable optimization framework," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1423–1441, 2022.
[41] 3GPP, "Study on scenarios and requirements for next generation access technologies," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, March, Release 14.
[42] I. Z. Ahmed, H. Sadjadpour, and S. Yousefi, "A novel information-directed tree-search algorithm for RIS phase optimization in massive MIMO," in *2023 International Conference on Computing, Networking and Communications (ICNC)*, 2023, pp. 398–402.