Transmit Power Minimization with QoS Guarantees in STAR-RIS-aided RSMA Networks

Donghyun Lee, Chihyun Song, The Vi Nguyen, Junsuk Oh, Yunseong Lee, Dongwook Won, Wonjong Noh, *Senior Member, IEEE*, and Sungrae Cho, *Senior Member, IEEE*

***Abstract*—Rate-splitting multiple access (RSMA) and simultaneously transmitting and reflecting (STAR) reconfigurable intelligent surface (RIS) have emerged as promising technologies for multiple access and relaying in sixth-generation (6G) networks. This paper investigates a STAR-RIS-aided RSMA system that minimizes transmit power while satisfying users' quality of service (QoS) requirements under spatially correlated channels. The formulated non-convex optimization problem jointly optimizes the precoding vector, common message ratio, phase shift matrix, and transmission and reflection ratio. To solve this, the problem is decomposed into three subproblems: (i) precoding and common message ratio control at the base station (BS), (ii) phase shift matrix design, (iii) transmission/reflection ratio control at the STAR-RIS. These subproblems are transformed into convex forms via penalty functions and linear approximations. A stationary solution is obtained using interior-point initialization, successive convex approximation, and alternating optimization algorithms. Through extensive simulations under various channel environments, including spatially correlated scenarios, the proposed method achieves much lower transmit power while meeting QoS requirements, outperforming conventional (STAR-)RIS-aided RSMA, non-orthogonal multiple access, and space-division multiple access. The performance gain becomes more significant when the number of STAR-RIS elements or BS antennas is limited.**

***Index Terms*—Rate-splitting multiple access, simultaneously transmitting and reflecting reconfigurable intelligent surface, transmit power minimization.**

I. Introduction

NEXT-generation wireless networks are expected to support a wide range of mission-critical applications and services, including high-performance holography, tactile-based communications, and immersive virtual experiences, and ubiquitous device interconnectivity [1]. These services demand high-spectral efficiency (SE) and energy efficiency (EE), which can be achieved through dense base station (BS) deployment, massive device connectivity, and the utilization of high-frequency and wideband channels. However, merely extending existing technologies is insufficient to meet the stringent performance requirements of future applications.

To this end, new multiple access (MA) schemes have recently been proposed, such as non-orthogonal multiple access (NOMA) and rate-splitting multiple access (RSMA). Space-division multiple access (SDMA) reduces inter-user interference (IUI) through sophisticated beamforming techniques. NOMA mitigates IUI via superposition coding and successive interference cancellation (SIC) in the power domain. RSMA, as a generalized framework bridging NOMA and SDMA, employs common and private messages to manage IUI. The common message partially decodes the IUI, leveraging the benefits of NOMA, while private messages are user-specific and are directed to the intended users, exploiting the spatial separation in SDMA.

Recently, new signal relaying technologies such as reconfigurable intelligent surface (RIS) and simultaneously transmitting and reflecting RIS (STAR-RIS) have emerged as promising solutions. RIS controls wireless channels between transmitters and receivers by adjusting signal reflections through multiple low-cost passive elements [2]. This enables an additional degree of freedom (DoF) and supports the realization of programmable wireless environments [2]. However, conventional RIS can only reflect incident signals, which limits deployment flexibility [3] and restricts cell coverage. To address this, a novel architecture called STAR-RIS has been introduced, which allows simultaneous transmission and reflection of signals toward users [4]. This capability enables deployment not only at cell boundaries but also within cell interiors, thereby extending coverage [5]. A successful prototype development by NTT Docomo has demonstrated the practical feasibility of STAR-RIS.

Furthermore, RSMA has been explored beyond conventional physical-layer designs. For instance, it has been studied in conjunction with semantic communication that jointly considers communication and computation resources, highlighting its potential for future wireless networks [6], [7]. Concurrently, RIS-assisted architectures have been extended to IIoT scenarios through distributed RIS-enabled semantic communications and joint resource optimization, further demonstrating their wide applicability [8].

Motivated by these advancements, the integration of STAR-RIS and RSMA is expected to play a crucial role in next-generation wireless networks, as it can enhance network

This work was supported in part by the IITP (Institute of Information Communications Technology Planning Evaluation) - ITRC (Information Technology Research Center) (IITP-2026-RS-2022-00156353, 40%) grant funded by the Korea government (Ministry of Science and ICT) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. RS-2023-00209125 / RS-2023-NR076382). The editor coordinating the review of this article was B. Smida. (Corresponding authors: Wonjong Noh; Sungrae Cho.)

D. Lee, C. Song, T.-V. Nguyen, J. Oh, D. Won, and S. Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea. *(e-mail: {dhlee, chsong, tvnguyen, jsoh, dwwon}@uclab.re.kr; srcho@cau.ac.kr)*

Y. Lee is with the School of Computer Science Engineering and Converged Technology, Duksung Women's University, Seoul 01369, Republic of Korea. *(e-mail: yunslee@duksung.ac.kr)*

W. Noh is with the School of Software, Hallym University, Chuncheon 24252, Republic of Korea. *(e-mail: wonjong.noh@hallym.ac.kr)*

connectivity, improve SE and EE, and enable intelligent control over wireless propagation with expanded coverage. Nevertheless, several challenges persist, including waveform and resource control for signal enhancement and interference mitigation [9], physical layer security concerns [10], and hardware complexity issues [11].

In this context, this paper investigates the performance benefits of jointly integrating STAR-RIS and RSMA. Specifically, we propose waveform and resource control strategies that minimize power consumption while ensuring the QoS requirements of each user in STAR-RIS-aided RSMA networks.

A. Related Work

1) NOMA and RSMA with RIS: Several studies have investigated RIS-aided NOMA [12]–[14] and RIS-aided RSMA [15]–[17] systems. Shehab *et al.* [12] studied sum rate maximization in RIS-aided downlink NOMA using a deep reinforcement learning (DRL)-based phase shift optimization algorithm to solve the resulting non-convex problem. Wang *et al.* [13] analyzed the trade-off between sum rate and transmit power in RIS-aided downlink NOMA, and proposed an alternating optimization (AO)-based algorithm for joint beamforming and phase shift optimization. Li *et al.* [14] assessed the phase shift matrix optimization of RIS according to the transmit power of RIS-aided uplink NOMA users. A closed-form phase shift was derived through semidefinite programming (SDP) and majorization-minimization. Khisa *et al.* [15] proposed a cooperative RIS-aided RSMA system, where users close to the BS relay the common stream to distant users, assuming that nearby users experience more favorable channel conditions. The time domain was divided into two phases to enable this cooperative transmission. Yang *et al.* [16] studied EE maximization by jointly optimizing the RIS phase shift and BS beamforming in a RIS-aided RSMA system. Truong *et al.* [17] explored an RSMA-based downlink system assisted by active RIS with low-resolution DACs and phase quantization. The work focused on EE maximization by jointly optimizing transmit power, RIS coefficients, and quantization levels.

2) NOMA and RSMA with STAR-RIS: Several studies have extended to STAR-RIS-aided NOMA systems. Fang *et al.* [18] introduced a zero-forcing (ZF) precoding strategy combined with a Dinkelbach method-based phase shift design. An AO algorithm was used to iteratively optimize both beamforming and STAR-RIS phase shifts until convergence. This approach effectively enhanced the EE of the STAR-RIS-aided downlink NOMA system. Zhao *et al.* [19] introduced an aerial active STAR-RIS-assisted IoT NOMA network, where a UAV-mounted surface enhances coverage through dynamic reflection and transmission control. An AO-based algorithm was employed to jointly optimize the UAV trajectory and resource allocation for sum rate maximization.

Finally, some studies have been conducted on STAR-RIS-aided RSMA systems [20]–[23]. Katwe *et al.* [20] investigated SE maximization in STAR-RIS-assisted uplink RSMA systems. AO and fractional programming-based algorithms were employed to optimize power allocation, decoding order, user fairness, and beamforming design at the STAR-RIS. Meng *et al.* [21] focused on sum rate maximization in a STAR-RIS-aided RSMA system. A DRL-based algorithm was proposed to jointly optimize the beamforming vector, common stream rate allocation, and transmission and reflection coefficients. Maghrebi *et al.* [22] investigated a cooperative RSMA framework assisted by an active STAR-RIS, where a relay user forwards the common message to improve connectivity. An AO-based algorithm was developed to jointly optimize transmit beamforming and STAR-RIS coefficients for sum rate maximization.

B. Motivations, Contributions and Organization

Most existing works [12], [14], [20], [21] have focused on SE for network optimization. However, with the rapid growth in mobile traffic and increasing demand for sustainable communication, EE has become a critical concern. In this context, transmit power minimization is fundamentally different from sum rate maximization [24]: instead of trading users' rates to boost the aggregate throughput, it is a feasibility-driven design that must satisfy all per-user QoS constraints, and the optimal operating point typically lies on the boundary of these constraints. As a result, the design becomes sensitive to channel conditions that shrink the feasible set, e.g., correlation-induced channel alignment that weakens multi-user separability.

Although transmit power minimization under target QoS constraint has been investigated in [25] and [15], these studies considered only conventional RIS combined with NOMA [25] and RSMA [15]. In addition, few studies assumed correlated channels in STAR-RIS aided systems, which naturally arises in practice [26]. Dhok *et al.* [27] studied STAR-RIS-assisted RSMA systems over spatially correlated channels, but it does not address the EE problem. While [15] and [22] focused on improving SE and EE through cooperative RSMA frameworks, they did not consider the inherent trade-offs of cooperation, such as increased transmission delay and time resource loss due to the BS being unable to transmit during user relaying. Motivated by this gap, this work focuses on the design of a STAR-RIS-aided RSMA system that minimizes transmit power over spatially correlated channels while satisfying user-specific QoS requirements without relying on user-side relaying. The main contributions of this paper are summarized as follows:

- We consider a STAR-RIS-aided single-layer multiple-input single-output RSMA system operating under an energy-splitting (ES) protocol. Here, spatial correlation is assumed between RIS elements and BS antennas based on their geometric configuration. A transmit power minimization problem is formulated under users' QoS constraints by jointly optimizing the precoding vector, common stream-rate allocation, phase shift matrix, and transmission/reflection (T/R) ratio.
- The formulated problem is non-convex and involves highly coupled variables. To address this, we decompose the problem into three subproblems, (i) joint precoding and common message ratio control, (ii) phase shift matrix control, and (iii) T/R ratio control. These subproblems are

solved alternately in an iterative manner. Additionally, we introduce an interior point-based initialization method to obtain a feasible starting point for the AO algorithm.

- For the first subproblem, we optimize the precoding vector and common message ratio for a given RIS phase shift and T/R ratio using an SCA method. The remaining subproblems are cast as feasibility checks, where the goal is to find a phase shift or T/R ratio satisfying the target QoS. We apply penalty and SCA-based techniques to solve them. In particular, the proposed penalty-based method for the phase shift matrix subproblem achieves lower computational complexity than SDP-based approach in [14], [15], especially when the number STAR-RIS elements is large.
- We analyze that the proposed approach has polynomial computational complexity and converges to a stationary point stably and rapidly. Extensive simulations demonstrate that the proposed scheme achieves the lowest transmit power while satisfying QoS, outperforming legacy (STAR-)RIS-aided RSMA, NOMA, and SDMA. The performance gain is more pronounced when the number of STAR-RIS elements or BS antennas is limited.

Notation: The operators $\odot$ and $\succeq$ represent the Hadamard product and componentwise inequality, respectively. $\mathcal{R}(x)$ denotes the real part of a complex number x. Given a set A with multiple elements, $\bar{x}$ indicates all elements of A except x. For example, if $A = \{x, y, z\}$, then $\bar{x} = \{y, z\}$.

II. System Model and Problem Formulation

We consider a STAR-RIS-aided single-layer multiple-input single-output RSMA system. The BS is equipped with a uniform linear array (ULA) of M antennas and serves $2K$ single-antenna users. The STAR-RIS is also equipped with a ULA of N elements. Prior studies [2], [21] have shown that STAR-RIS achieves the highest signal-to-interference-plus-noise ratio (SINR) when deployed near the BS or users. Based on this, it is reasonable to place the STAR-RIS at the cell edge to assist users located farther from the BS, as illustrated in Fig. 1.

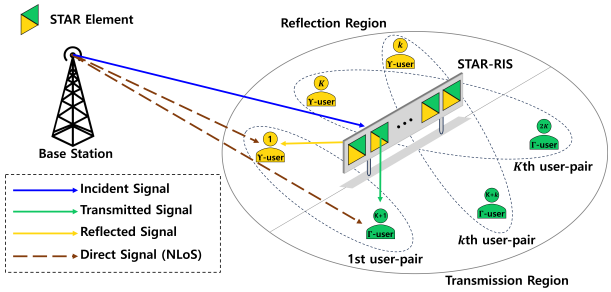


Fig. 1: STAR-RIS-aided system model.

As illustrated in Fig. 1, all $2K$ users are assumed to be located at the cell edge. The K users in the reflection region are denoted as $\Upsilon := \{1, \ldots, K\}$ and the remaining K users in the transmission region are denoted as $\Gamma := \{K+1, \ldots, 2K\}$. The Υ-users are supported via reflected signals, while the Γ-users are served through transmitted signals from the STAR-RIS. It is assumed that each Υ-user is paired one-to-one with a Γ-user for STAR-RIS-assisted transmission [27]. For convenience, in each pair, the Υ-user and Γ-user are labeled as users r and t, respectively. Here, all users can be paired using a low-complexity user scheduling algorithm such as [28].

However, in this work, we focus on a single user pair to identify the link-level core controllability and effectiveness of the proposed control. In particular, combined with some user-pairing algorithm, hybrid multiple access schemes, such as OFDMA-NOMA [28] and OFDMA-RSMA [29], can be employed to manage multiple user pairs while reusing the single-pair control architecture within each subchannel. This hierarchical structure ensures that the techniques developed and validated under the single-pair setting can be efficiently adapted to more complex systems with minor algorithmic modifications.

In the RSMA system, for a given user pair, the message of user $\chi \in \{t, r\}$ is divided into two parts: a common message $W_{c,\chi}$ and a private message $W_{p,\chi}$. The common messages from all users $\{W_{c,r}, W_{c,t}\}$ are combined into a single common message W_C and encoded into a common stream s_c. The private messages $W_{p,r}$ and $W_{p,t}$ are encoded into private streams s_r and s_t, respectively. Thus, the signal transmitted by the BS for a user pair is given by

$$\mathbf{x} = \mathbf{p}_c s_c + \mathbf{p}_t s_t + \mathbf{p}_r s_r, \tag{1}$$

where $\mathbf{p_c} \in \mathbb{C}^{\mathbf{M\times 1}}$ is the precoding vector for the common stream, and $\mathbf{p}_\chi \in \mathbb{C}^{\mathbf{M\times 1}}, \forall \chi \in \{\mathbf{t}, \mathbf{r}\}$ are the precoding vectors for the private streams.

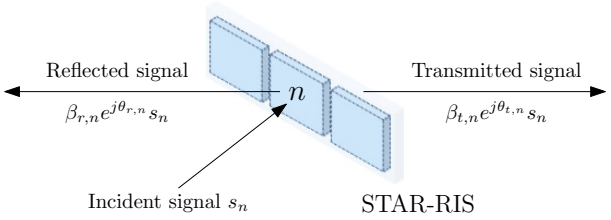


Fig. 2: ES protocol of the STAR-RIS.

For the STAR-RIS, we adopt the ES protocol [4], as illustrated in Fig. 2. The incident signal s_n to the n-th STAR-RIS element is divided into the transmitted and reflected components according to the T/R ratio $\beta_{t,n}^2$ and $\beta_{r,n}^2$, respectively. To satisfy the energy conservation law, the sum of power coefficients satisfies $\beta_{t,n}^2 + \beta_{r,n}^2 = 1$. The reflected and transmitted signals are then phase-shifted by $\phi_{r,n} = e^{j\theta_{r,n}}$ and $\phi_{t,n} = e^{j\theta_{t,n}}$, where $\theta_{r,n}, \theta_{t,n} \in [0, 2\pi)$, and are delivered to the corresponding user.

The received signal at user χ is expressed as

$$y_\chi = \left(\mathbf{h}_\chi^H \mathbf{\Theta}_\chi \mathbf{G} + \mathbf{d}_\chi^H\right)\mathbf{x} + n_\chi, \quad \forall \chi \in \{t, r\}, \tag{2}$$

where $\mathbf{G} \in \mathbb{C}^{N\times M}$ denotes the channel gain between the BS and the STAR-RIS. The matrix $\mathbf{\Theta}_\chi = \mathrm{diag}\{\beta_{\chi,1}\phi_{\chi,1}, \ldots, \beta_{\chi,n}\phi_{\chi,N}\} \in \mathbb{C}^{N\times N}$ represents the phase shift matrix at the STAR-RIS, where $\phi_{\chi,n} = e^{j\theta_{\chi,n}}, \forall n \in \{1, ..., N\}$ and $\theta_n \in [0, 2\pi)$. In addition, $\mathbf{h}_\chi \in \mathbb{C}^{N\times 1}$ denotes the channel gain from the STAR-RIS to user χ, $\mathbf{d}_\chi \in \mathbb{C}^{M\times 1}$ denotes the direct channel gain from the BS to the user χ, and

$n_\chi \sim \mathcal{CN}(0, \sigma_{n,\chi}^2)$ represents additive white Gaussian noise at user χ. We assume a block fading channel model, where the channel remains constant within each coherence block and varies independently across blocks. The perfect channel state information (CSI) is assumed to be available at both the BS and users.

A. Channel Model

The presence of multiple BS antennas and STAR-RIS elements leads to spatial correlation, which depends on the signal wavelength and the antenna or element spacing. This spatial correlation significantly affects the achievable rate. To model spatial correlation between antennas, we adopt a distance-based sinc correlation model under the assumption of isotropic scattering. This model is physically justified, as uniformly distributed multipath components over angles lead to a correlation that depends on the element spacing normalized by the wavelength, resulting in a sinc function. Such models have been widely adopted in the RIS literature and have also been considered for STAR-RIS under correlated fading [26], [30]. The position of the m-th BS antenna element is given by $r_m = d_{ant} \cdot (m-1), m \in \{1, 2, \ldots, M\}$, where d_{ant} represents BS antenna spacing. Then, the (m, l)-th entry of the BS correlation matrix $\mathbf{R}_{\mathrm{BS}} \in \mathbb{C}^{\mathbf{M}\times\mathbf{M}}$ is defined as:

$$[\mathbf{R}_{\mathrm{BS}}]_{m,l} = \mathrm{sinc}\left(\frac{2}{\lambda} \cdot |r_m - r_l|\right), \tag{3}$$

where $\mathrm{sinc}(x) = \frac{\sin \pi x}{\pi x}$ denotes the normalized sinc function. Furthermore, $\lambda = \frac{c}{f_c}$ denotes the carrier wavelength, where c is the speed of light and f_c is the carrier frequency. We generate spatially correlated fading channels via singular value decomposition $\mathbf{R}_{\mathrm{BS}} = \mathbf{U}_{\mathrm{BS}}\mathbf{S}_{\mathrm{BS}}\mathbf{U}_{\mathrm{BS}}^H$. The spatially correlated MIMO channel is then obtained by applying the square root of the correlation matrix, $\mathbf{R}_{\mathrm{BS}}^{1/2} = \mathbf{U}_{\mathrm{BS}}\sqrt{\mathbf{S}_{\mathrm{BS}}}\mathbf{U}_{\mathrm{BS}}^H$, to the i.i.d. non-line-of-sight (NLoS) components. The same correlation modeling approach is applied to the STAR-RIS, using a spatial correlation matrix $\mathbf{R}_{\mathrm{RIS}}$.

To investigate the impact of spatial correlation under various channel conditions, each channel is modeled as a Kronecker correlation model based on a Rician fading channel as

$$\mathbf{G} = L_1\left(\sqrt{\frac{\varepsilon_g}{\varepsilon_g+1}}\hat{\mathbf{G}} + \sqrt{\frac{1}{\varepsilon_g+1}}\mathbf{R}_{RIS}^{1/2}\tilde{\mathbf{G}}\mathbf{R}_{\mathrm{BS}}^{1/2}\right), \tag{4}$$

$$\mathbf{h}_\chi = L_{2,\chi}\left(\sqrt{\frac{\varepsilon_{h,\chi}}{\varepsilon_{h,\chi}+1}}\hat{\mathbf{h}}_\chi + \sqrt{\frac{1}{\varepsilon_{h,\chi}+1}}\mathbf{R}_{\mathrm{RIS}}^{1/2}\tilde{\mathbf{h}}_\chi\right), \tag{5}$$

$$\mathbf{d}_\chi = L_{d,\chi}\left(\sqrt{\frac{\varepsilon_{d,\chi}}{\varepsilon_{d,\chi}+1}}\hat{\mathbf{d}}_\chi + \sqrt{\frac{1}{\varepsilon_{d,\chi}+1}}\mathbf{R}_{\mathrm{BS}}^{1/2}\tilde{\mathbf{d}}_\chi\right), \tag{6}$$

where $\mathbf{R}_{\mathrm{RIS}} \in \mathbb{C}^{N\times N}$ denotes the spatial correlation matrices for the STAR-RIS. Furthermore, ε denotes the Rician factor, which indicates the power ratio between the LoS and NLoS components. The terms with $\hat{(\cdot)}$ and $\tilde{(\cdot)}$ represent the LoS component and NLoS components, respectively. The NLoS components are modeled as i.i.d. complex Gaussian vectors or matrices with zero mean and unit variance, i.e., each entry follows $\mathcal{CN}(0, 1)$. The scaling factors L_1, $L_{2,\chi}$ and $L_{d,\chi}$ account for path loss. Both the BS and the STAR-RIS employ uniform linear arrays (ULAs), and their array response vectors are given by:

$$\mathbf{a}_M(\rho) = \left[1, e^{j2\pi\frac{d_{ant}}{\lambda_{wav}}\sin\rho}, \ldots, e^{j2\pi\frac{d_{ant}}{\lambda_{wav}}(M-1)\sin\rho}\right]^H, \tag{7}$$

$$\mathbf{a}_N(\rho) = \left[1, e^{j2\pi\frac{d_{RIS}}{\lambda_{wav}}\sin\rho}, \ldots, e^{j2\pi\frac{d_{RIS}}{\lambda_{wav}}(N-1)\sin\rho}\right]^H, \tag{8}$$

where ρ denotes the angle of arrival (AoA) or angle of departure (AoD). Furthermore, d_{RIS} and λ_{wav} represent the element spacing at the STAR-RIS, and the signal wavelength, respectively. Accordingly, the LoS components are given by [31, p. 253]

$$\hat{\mathbf{G}} = \mathbf{a}_N(\rho_{\mathrm{AoA,RIS}})\mathbf{a}_M^T(\rho_{\mathrm{AoD,BS}}), \tag{9}$$

$$\hat{\mathbf{h}}_\chi = \mathbf{a}_N(\rho_{\mathrm{AoD,RIS},\chi}), \tag{10}$$

$$\hat{\mathbf{d}}_\chi = \mathbf{a}_M(\rho_{\mathrm{AoD,BS},\chi}), \tag{11}$$

where $\rho_{\mathrm{AoA,RIS}}$ and $\rho_{\mathrm{AoD,BS}}$ denote the AoA at the STAR-RIS and AoD at the BS, respectively. The parameters $\rho_{\mathrm{AoD,RIS},\chi}$ and $\rho_{\mathrm{AoD,BS},\chi}$ refer to the AoDs from the STAR-RIS and BS to user χ, respectively.

The correlation depends on the normalized element spacing $\frac{d_{\mathrm{RIS}}}{\lambda_{wav}}$ and antenna spacing $\frac{d_{ant}}{\lambda_{wav}}$, which affect the channels of the STAR-RIS elements and BS antennas. In general, spatial correlation is assumed to be negligible when the antenna or element spacing is set greater than or equal to $0.5\lambda_{wav}$.

B. Rate-Splitting Multiple Access Rates

To simplify notation in (2), let $\mathbf{v}_\chi := \mathbf{h}_\chi^H\mathbf{\Theta}_\chi\mathbf{G} + \mathbf{d}_\chi^H$. Then, the SINR of the received common stream s_c at user $\chi \in \{t, r\}$ is given by:

$$\mathrm{SINR}_{c,\chi}(\mathbf{\Theta}_\chi, \mathbf{p}) = \frac{|\mathbf{v}_\chi\mathbf{p}_c|^2}{|\mathbf{v}_\chi\mathbf{p}_r|^2 + |\mathbf{v}_\chi\mathbf{p}_t|^2 + \sigma_\chi^2}, \tag{12}$$

and the SINRs of the private stream $s_{p,\chi}$ for user t and r are given by:

$$\mathrm{SINR}_{p,t}(\mathbf{\Theta}_t, \mathbf{p}) = \frac{|\mathbf{v}_t\mathbf{p}_t|^2}{|\mathbf{v}_t\mathbf{p}_r|^2 + \sigma_t^2}, \tag{13}$$

$$\mathrm{SINR}_{p,r}(\mathbf{\Theta}_r, \mathbf{p}) = \frac{|\mathbf{v}_r\mathbf{p}_r|^2}{|\mathbf{v}_r\mathbf{p}_t|^2 + \sigma_r^2}, \tag{14}$$

where $\mathbf{p} = [\mathbf{p}_c^H, \mathbf{p}_r^H, \mathbf{p}_t^H]^H \in \mathbb{C}^{3M\times 1}$ is the concatenated precoding vector for all streams. The achievable rates for common and private streams at user χ are defined as:

$$R_{c,\chi}(\mathbf{\Theta}_\chi, \mathbf{p}) = \log_2(1 + \mathrm{SINR}_{c,\chi}), \tag{15}$$

$$R_{p,\chi}(\mathbf{\Theta}_\chi, \mathbf{p}) = \log_2(1 + \mathrm{SINR}_{p,\chi}), \tag{16}$$

respectively. To ensure decodability of the common stream, both users must decode it; thus, the common rate $R_c \triangleq \min\{R_{c,t}, R_{c,r}\}$.. We split R_c as $R_c = C_t + C_r$, where C_χ is the portion assigned to user χ. Accordingly, the achievable rate of user χ is $C_\chi + R_{p,\chi}$.

C. Problem Formulation

This work aims to minimize the total transmit power of the BS while satisfying the QoS requirements of each user in the STAR-RIS-aided RSMA system. To fully exploit the potential of STAR-RIS-aided RSMA systems, it is essential to jointly optimize the four key parameters: the precoding vector, the common message ratio, the phase shift matrix, and the T/R ratio vector. These variables are mutually coupled in their impact on the signal quality, interference suppression, and EE. The optimization problem is formulated as:

$$\textbf{(P1)}: \min_{\mathbf{p},\mathbf{c},\boldsymbol{\phi},\boldsymbol{\beta}} \quad ||\mathbf{p}_c||^2 + ||\mathbf{p}_t||^2 + ||\mathbf{p}_r||^2, \tag{17a}$$

$$\text{s.t. } C_\chi + R_{p,\chi} \geq R_\chi^{th}, \qquad \forall \chi \in \{t, r\}, \tag{17b}$$

$$C_t + C_r \leq R_{c,\chi}, \qquad \forall \chi \in \{t, r\}, \tag{17c}$$

$$|\phi_{\chi,n}| = 1, \quad \forall \chi \in \{t, r\}, \forall n \in \{1, ..., N\}, \tag{17d}$$

$$\beta_{r,n}^2 + \beta_{t,n}^2 = 1, \quad \forall n \in \{1, ..., N\}, \tag{17e}$$

where $\mathbf{c} = [C_r, C_t] \succeq 0$, and $\succeq$ denotes the component-wise inequality. In addition, $\boldsymbol{\phi} = [\boldsymbol{\phi}_\chi]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_\chi]$, where $\boldsymbol{\phi}_\chi = [\phi_{\chi,1}, ..., \phi_{\chi,N}]$ and $\boldsymbol{\beta}_\chi = [\beta_{\chi,1}, \beta_{\chi,2}, \ldots, \beta_{\chi,N}], \chi \in \{t, r\}$. In problem **(P1)**, equation (17a) denotes a total consumption power at the BS as an objective function. Constraint (17b) ensures the QoS requirement for each user χ, where R_χ^{th} is the minimum required data rate at user χ. Constraint (17c) guarantees that the sum of the allocated common rates does not exceed the achievable common rate for all users, ensuring successful decoding of the common message. Constraint (17d) imposes the unit-modulus constraint on the STAR-RIS phase shift for each user χ at the n-th STAR-RIS element. Constraint (17e) enforces energy conservation by requiring the sum of squared transmission and reflection coefficients to be one.

The optimization problem (**P1**) is non-convex problem due to the non-convexity of constraints (17b), (17c), (17d), and (17e), and the strong coupling among the control variables. To address this challenge, we decompose the transmit power minimization problem (**P1**) into subproblems and propose an iterative algorithm based on SCA and AO techniques.

III. The Proposed Control Algorithm

A. Joint Optimization of Precoding and Rate-Splitting Vectors

Given the phase shift vectors in $\boldsymbol{\phi}$ and T/R ratio vectors in $\boldsymbol{\beta}$, the goal is to optimize the precoding vectors and the common stream rate-splitting vector at the BS. Accordingly, the problem **(P1)** is reformulated as follows:

$$\textbf{(P2)}: \min_{\mathbf{p},\mathbf{c}} \quad ||\mathbf{p}_c||^2 + ||\mathbf{p}_t||^2 + ||\mathbf{p}_r||^2, \tag{18}$$

$$\text{s.t.(17b), (17c).}$$

To solve this, we adopt the SCA method. It solves the non-convex problem by iteratively constructing and solving convex surrogate subproblems. In the following, we first approximate the non-convex constraints accordingly.

Convex approximation for constraint (17b)*:* To address the non-convexity of constraint (17b), a slack variable $\boldsymbol{\gamma}_p = [\gamma_{p,\chi}]$, $\forall \chi \in \{t, r\}$, is introduced to represent the SINRs of the private streams [15]. Using this, constraint (17b) can be reformulated as:

$$C_\chi + \log_2(1 + \gamma_{p,\chi}) \geq R_\chi^{th}, \tag{19}$$

$$\frac{|\mathbf{v}_\chi \mathbf{p}_\chi|^2}{|\mathbf{v}_\chi \mathbf{p}_{\bar{\chi}}|^2 + \sigma_\chi^2} \geq \gamma_{p,\chi}, \tag{20}$$

where $\bar{\chi} \in \{t, r\}, \bar{\chi} \neq \chi$. It can be verified that constraint (19) is convex by proving that the left-hand side is concave. This can be proved by employing the following lemma.

Lemma 1. *Define* $\mathcal{G}(C_\chi, \gamma_{p,\chi}) = C_\chi + \log_2(1 + \gamma_{p,\chi})$*, for* $\gamma_{p,\chi} \geq 0$*. Then* $\mathcal{G}(C_\chi, \gamma_{p,\chi})$ *is concave.*

Proof. $\log_2(1+\gamma_{p,\chi})$ is concave for $\gamma_{p,\chi} \geq 0$ and C_χ is affine; hence $\mathcal{G}(C_\chi, \gamma_{p,\chi})$ is concave. □

On the other hand, constraint (20) is also non-convex due to the fractional form of the left-hand side. To address this, we introduce another slack variable $\boldsymbol{\xi}_p = [\xi_{p,\chi}]$, $\forall \chi \in \{t, r\}$, and reformulate (20) as follows:

$$\frac{|\mathbf{v}_\chi \mathbf{p}_\chi|^2}{\xi_{p,\chi}} \geq \gamma_{p,\chi}, \tag{21}$$

$$\xi_{p,\chi} \geq |\mathbf{v}_\chi \mathbf{p}_{\bar{\chi}}|^2 + \sigma_\chi^2. \tag{22}$$

Constraint (22) ensures that the denominator is greater than the power of the interference-plus-noise term, enabling a tractable reformulation. It can be observed that constraint (22) is convex, while (21) remains non-convex due to the (non-concave) fractional quadratic term on the left-hand side. Fortunately, this term can be approximated using the general form, $f(u, v) = \frac{|v|^2}{u}, \forall v \in \mathbb{C}, \forall u \in \mathbb{R}^+$, which admits a convex lower bound via a first-order approximation around a feasible point $(u^{[i]}, v^{[i]})$:

$$f(u, v) \geq F\left(u, v; u^{[i]}, v^{[i]}\right) = \frac{2\mathcal{R}\left\{(v^{[i]})^H v\right\}}{u^{[i]}} - \left(\frac{|v^{[i]}|}{u^{[i]}}\right)^2 u, \tag{23}$$

where $\mathcal{R}(z)$ denotes the real part of a complex number z. Applying (23) to (21), we obtain:

$$\frac{|\mathbf{v}_\chi \mathbf{p}_\chi|^2}{\xi_{p,\chi}} \geq \frac{2\mathcal{R}\left\{(\mathbf{p}_\chi^{[i]})^H \mathbf{v}_\chi^H \mathbf{v}_\chi \mathbf{p}_\chi\right\}}{\xi_{p,\chi}^{[i]}} - \frac{\left|\mathbf{v}_\chi \mathbf{p}_\chi^{[i]}\right|^2 \xi_{p,\chi}}{\left(\xi_{p,\chi}^{[i]}\right)^2}, \tag{24}$$

where $(\mathbf{p}_\chi^{[i]}, \xi_{p,\chi}^{[i]})$ are the values of $(\mathbf{p}_\chi, \xi_{p,\chi})$ at the i-th iteration of the SCA. Therefore, the convex approximation of constraint (21) is:

$$\frac{2\mathcal{R}\left\{(\mathbf{p}_\chi^{[i]})^H \mathbf{v}_\chi^H \mathbf{v}_\chi \mathbf{p}_\chi\right\}}{\xi_{p,\chi}^{[i]}} - \frac{\left|\mathbf{v}_\chi \mathbf{p}_\chi^{[i]}\right|^2 \xi_{p,\chi}}{\left(\xi_{p,\chi}^{[i]}\right)^2} \geq \gamma_{p,\chi}, \tag{25}$$

which is a convex constraint.

Convex approximation for constraint (17c)*:* To manage this constraint, we introduce slack variables $\boldsymbol{\gamma}_c = [\gamma_{c,\chi}]$, $\forall \chi \in \{t, r\}$ representing the SINRs of the common stream

and $\boldsymbol{\xi}_c = [\xi_{c,\chi}]$, $\forall \chi \in \{t, r\}$. Accordingly, constraint (17c) can be reformulated as follows:

$$C_t + C_r \leq \log_2\left(1 + \gamma_{c,\chi}\right), \tag{26}$$

$$\frac{\left|\mathbf{v}_\chi \mathbf{p}_c\right|^2}{\xi_{c,\chi}} \geq \gamma_{c,\chi}, \tag{27}$$

$$\xi_{c,\chi} \geq \left|\mathbf{v}_\chi \mathbf{p}_\chi\right|^2 + \left|\mathbf{v}_\chi \mathbf{p}_{\bar{\chi}}\right|^2 + \sigma_\chi^2. \tag{28}$$

Here, the convex approximation of (27) can be expressed as, by applying the same approximation technique as in constraint (21),

$$\frac{2\mathcal{R}\left\{\left(\mathbf{p}_c^{[i]}\right)^H \mathbf{v}_\chi^H \mathbf{v}_\chi \mathbf{p}_c\right\}}{\xi_{c,\chi}^{[i]}} - \frac{\left|\mathbf{v}_\chi \mathbf{p}_c^{[i]}\right|^2 \xi_{c,\chi}}{\left(\xi_{c,\chi}^{[i]}\right)^2} \geq \gamma_{c,\chi}, \tag{29}$$

where $(\mathbf{p}_c^{[i]}, \xi_{c,\chi}^{[i]})$ denotes the value of $(\mathbf{p}_c, \xi_{c,\chi})$ at the i-th iteration of the SCA.

Finally, problem **(P2)** can be reformulated as:

$$\textbf{(P3)}: \min_{\mathbf{p},\mathbf{c},\boldsymbol{\gamma}_c,\boldsymbol{\gamma}_p,\boldsymbol{\xi}_c,\boldsymbol{\xi}_p} \quad ||\mathbf{p}_c||^2 + ||\mathbf{p}_t||^2 + ||\mathbf{p}_r||^2, \tag{30}$$
$$\text{s.t. (19), (22), (25), (26), (28), (29).}$$

Problem **(P3)** is convex and can be solved by standard solvers such as CVX [32]. As summarized in Alg. 1, we apply an SCA procedure that initializes the precoders, common rate, and slack variables, and then iteratively solves a sequence of convex surrogates to update all optimization variables until the transmit power convergence.

Algorithm 1 Algorithm for Jointly Solving Precoding and Rate-Splitting Vectors

1: **Input:** phase shift vectors $\boldsymbol{\phi}_\chi$, T/R ratios $\boldsymbol{\beta}_\chi$, threshold R_χ^{th}, tolerance ϵ_1, maximum iteration $I_{\max}$, and set the current iteration index $i = 0$.
2: **Initialize:** $\mathbf{p}^{[0]}, \mathbf{c}^{[0]}, \boldsymbol{\gamma}_c^{[0]}, \boldsymbol{\gamma}_p^{[0]}, \boldsymbol{\xi}_c^{[0]}, \boldsymbol{\xi}_p^{[0]}$, set $P^{[0]} \triangleq \|\mathbf{p}_c^{[0]}\|^2 + \|\mathbf{p}_t^{[0]}\|^2 + \|\mathbf{p}_r^{[0]}\|^2$.
3: **repeat**
4: $i \leftarrow i + 1$;
5: Solve **(P3)** to obtain optimal solution $\mathbf{p}^*, \mathbf{c}^*, \boldsymbol{\gamma}_c^*, \boldsymbol{\gamma}_p^*, \boldsymbol{\xi}_c^*, \boldsymbol{\xi}_p^*$;
6: Update $\mathbf{p}^{[i]} \leftarrow \mathbf{p}^*, \mathbf{c}^{[i]} \leftarrow \mathbf{c}^*, \boldsymbol{\gamma}_c^{[i]} \leftarrow \boldsymbol{\gamma}_c^*, \boldsymbol{\gamma}_p^{[i]} \leftarrow \boldsymbol{\gamma}_p^*, \boldsymbol{\xi}_c^{[i]} \leftarrow \boldsymbol{\xi}_c^*, \boldsymbol{\xi}_p^{[i]} \leftarrow \boldsymbol{\xi}_p^*$;
7: $P^{[i]} \triangleq \|\mathbf{p}_c^{[i]}\|^2 + \|\mathbf{p}_t^{[i]}\|^2 + \|\mathbf{p}_r^{[i]}\|^2$;
8: **until** $|P^{[i]} - P^{[i-1]}| \leq \epsilon_1$ or $i \geq I_{\max}$
9: **Output:** $\mathbf{p}^{opt}, \mathbf{c}^{opt}$

On the other hand, it is important to initialize a feasible point for Alg. 1. Towards this end, we introduce an infeasibility indicator, denoted as $s \geq 0$, which quantifies how far the constraints in (**P3**) are from being satisfied. In other words, this parameter measures the violation level of the constraints. When $s = 0$, all constraints are guaranteed to be satisfied. Based on this idea, we can formulate the feasibility problem as follows. Let $\mathbf{x} \triangleq \{\mathbf{p}, \mathbf{c}, \boldsymbol{\gamma}_c, \boldsymbol{\gamma}_p, \boldsymbol{\xi}_c, \boldsymbol{\xi}_p\}$ collect the optimization variables in (P3). Each constraint in (P3) can be rewritten in the standard form $g_c(\mathbf{x}) \leq 0$ by moving all terms to the left-hand side. We define the constraint-violation function as $\mathcal{G}_c(\mathbf{x}) \triangleq g_c(\mathbf{x})$. For notational simplicity, each constraint is indexed by c; for instance, the constraint in (19) for user χ in denoted as $c = (13, \chi)$. This constraint can be equivalently written as:

$$g_{13,\chi}(\mathbf{x}) \triangleq R_\chi^{th} - C_\chi - \log_2(1 + \gamma_{p,\chi}) \leq 0, \ \chi \in \{t, r\}. \tag{31}$$

The remaining constraints in **(P3)**, i.e., (22), (25), (26), (28), and (29), can similarly be transformed into the $g_c(\mathbf{x}) \leq 0$, and are omitted for brevity. Based on this formulation, the feasibility problem is given by:

$$\min_{\mathbf{p},\mathbf{c},\boldsymbol{\gamma}_c,\boldsymbol{\gamma}_p,\boldsymbol{\xi}_c,\boldsymbol{\xi}_p,s} \quad s, \qquad \text{s.t. } \mathcal{G}_c(\mathbf{x}) \leq s, \ \ \forall c, \ \ s \geq 0, \tag{32}$$

where s is an infeasibility indicator. When the optimal value satisfies $s = 0$, all constraints in **(P3)** are satisfied. The initialization procedure is outlined in Alg. 2, where the infeasibility parameter s is iteratively reduced.

Algorithm 2 Initialization for Alg. 1

1: **Input:** threshold R_χ^{th}, tolerance $\tilde{\epsilon}_1 = 10^{-5}$, and set the current iteration index $\tilde{i} = 0$.
2: **Initialize:** randomly generate $\boldsymbol{\phi}_\chi^{[0]} \in \mathbb{C}^{N\times 1}$ satisfying constraint (17d), $\boldsymbol{\beta}_\chi^{[0]} = [\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2}]^T \in \mathbb{R}^{N\times 1}$, and $C_\chi^{[0]} = R_\chi^{th}/2$, $\mathbf{p}^{[0]} \in \mathbb{C}^{3M\times 1}$, $\boldsymbol{\gamma}_p^{[0]}, \boldsymbol{\xi}_p^{[0]}, \boldsymbol{\gamma}_c^{[0]}, \boldsymbol{\xi}_c^{[0]}$ are found by setting (19), (22), (26), and (28) to be equality.
3: **repeat**
4: $\tilde{i} \leftarrow \tilde{i} + 1$;
5: Solve problem (32) to obtain optimal solution $\mathbf{p}^*, \mathbf{c}^*, \boldsymbol{\gamma}_c^*, \boldsymbol{\gamma}_p^*, \boldsymbol{\xi}_c^*, \boldsymbol{\xi}_p^*, s^*$;
6: Update $\mathbf{p}^{[\tilde{i}]} \leftarrow \mathbf{p}^*, \mathbf{c}^{[\tilde{i}]} \leftarrow \mathbf{c}^*, \boldsymbol{\gamma}_c^{[\tilde{i}]} \leftarrow \boldsymbol{\gamma}_c^*, \boldsymbol{\gamma}_p^{[\tilde{i}]} \leftarrow \boldsymbol{\gamma}_p^*, \boldsymbol{\xi}_c^{[\tilde{i}]} \leftarrow \boldsymbol{\xi}_c^*, \boldsymbol{\xi}_p^{[\tilde{i}]} \leftarrow \boldsymbol{\xi}_p^*, s^{[\tilde{i}]} \leftarrow s^*$;
7: **until** $|s^*| \leq \tilde{\epsilon}_1$
8: **Output:** Feasible initial $\mathbf{p}, \mathbf{c}, \boldsymbol{\gamma}_c, \boldsymbol{\gamma}_p, \boldsymbol{\xi}_c, \boldsymbol{\xi}_p, \boldsymbol{\phi}, \boldsymbol{\beta}$.

B. Optimization of the Phase Shift

In this subsection, for given $\mathbf{p}$, $\mathbf{c}$, and $\boldsymbol{\beta}$, we aim to find the optimized phase shift $\boldsymbol{\phi} = [\boldsymbol{\phi}_t, \boldsymbol{\phi}_r]$. Accordingly, problem **(P1)** can be reformulated as

$$\textbf{(P4)}: \quad \text{Find} \quad \boldsymbol{\phi}_t, \boldsymbol{\phi}_r, \quad \text{s.t. (17b), (17c), (17d).} \tag{33}$$

It is evident that constraint (17d), which imposes the unit-modulus condition, is non-convex. This challenge can be addressed by applying the penalty method [33]. According to this approach, problem (**P4**) can be reformulated as

$$\textbf{(P5)}: \max_{\boldsymbol{\phi}_t, \boldsymbol{\phi}_r} \quad V_{\boldsymbol{\phi}_t} \sum_{n=1}^{N} \left(|\phi_{t,n}|^2 - 1\right) + V_{\boldsymbol{\phi}_r} \sum_{n=1}^{N} \left(|\phi_{r,n}|^2 - 1\right), \tag{34a}$$

$$\text{s.t.} \quad |\phi_{\chi,n}| \leq 1, \quad \forall n \in \{1, ...N\}, \forall \chi \in \{t, r\} \tag{34b}$$

$$\text{(17b), (17c),}$$

where $V_{\boldsymbol{\phi}_\chi} > 0$ is a penalty parameter for each $\chi \in \{t, r\}$. The penalty function $V_{\boldsymbol{\phi}_\chi} \sum_{n=1}^{N} (|\phi_{\chi,n}|^2 - 1)$ enforces the

condition $|\phi_{\chi,n}| - 1 = 0$. The problem (**P5**) is still non-convex problem because of the non-convex objective function and constraints.

To address the non-convex objective function of (**P5**), we approximate the penalty part of the objective function using a first-order Taylor expansion as follows [16]:

$$2V_{\boldsymbol{\phi}_t}\sum_{n=1}^{N}\left\{\mathrm{Re}\left(\phi_{t,n}^{[l]}\phi_{t,n}\right)-\left(\phi_{t,n}^{[l]}\right)^2\right\} + 2V_{\boldsymbol{\phi}_r}\sum_{n=1}^{N}\left\{\mathrm{Re}\left(\phi_{r,n}^{[l]}\phi_{r,n}\right)-\left(\phi_{r,n}^{[l]}\right)^2\right\}, \tag{35}$$

where the superscript $[l]$ indicates the value of the variable at the l-th iteration. In addition, to address the non-convexity of constraints (17b) and (17c), we adopt the same methodology as used in solving (**P2**).

Convex approximation for constraint (17b)*:* To linearize constraint (17b), we first decompose $\boldsymbol{\Theta}_\chi$ into a phase shift vector and a T/R ratio vector, as $\boldsymbol{\phi}_\chi = [\phi_{\chi,1}, \phi_{\chi,2}, \ldots, \phi_{\chi,N}] \in \mathbb{C}^{1\times N}$, and $\boldsymbol{\beta}_\chi = [\beta_{\chi,1}, \beta_{\chi,2}, \ldots, \beta_{\chi,N}] \in \mathbb{C}^{1\times N}$, respectively, where $\boldsymbol{\Theta}_\chi = \mathrm{diag}(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_{\boldsymbol{\chi}})$. Here, $\odot$ denotes the Hadamard product. Next, we define $\mathbf{t}_{\chi\psi} := \mathrm{diag}(\mathbf{h}_\chi^H)\mathbf{G}\mathbf{p}_\psi \in \mathbb{C}^{N\times 1}$ and $\delta_{\chi\psi} := \mathbf{d}_\chi^H\mathbf{p}_\psi \in \mathbb{C}$ for $\psi \in \{c,t,r\}$. Then, the term $\mathbf{v}_\chi\mathbf{p}_\psi$ can be rewritten as

$$\begin{aligned}\mathbf{v}_\chi\mathbf{p}_\psi &= \left(\mathbf{h}_\chi^H\boldsymbol{\Theta}_\chi\mathbf{G}+\mathbf{d}_\chi^H\right)\mathbf{p}_\psi\\ &= \mathbf{h}_\chi^H\boldsymbol{\Theta}_\chi\mathbf{G}\mathbf{p}_\psi + \mathbf{d}_\chi^H\mathbf{p}_\psi\\ &\overset{(a)}{=} \mathbf{h}_\chi^H\mathrm{diag}(\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_{\boldsymbol{\chi}})\mathbf{G}\mathbf{p}_\psi + \mathbf{d}_\chi^H\mathbf{p}_\psi\\ &\overset{(b)}{=} (\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_{\boldsymbol{\chi}})\underbrace{\mathrm{diag}(\mathbf{h}_\chi^H)\mathbf{G}\mathbf{p}_\psi}_{\mathbf{t}_{\chi\psi}} + \underbrace{\mathbf{d}_\chi^H\mathbf{p}_\psi}_{\delta_{\chi\psi}}\\ &= (\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\psi}+\delta_{\chi\psi},\end{aligned} \tag{36}$$

where (a) follows from the definition $\boldsymbol{\Theta}_\chi = \mathrm{diag}(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_{\boldsymbol{\chi}})$, (b) follows form the identity $\mathbf{a}^H\mathrm{diag}(\mathbf{b}) = \mathbf{b}\mathrm{diag}(\mathbf{a}^H)$, which implies $\mathbf{h}_\chi^H\mathrm{diag}(\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_{\boldsymbol{\chi}}) = (\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_{\boldsymbol{\chi}})\mathrm{diag}(\mathbf{h}_\chi^H)$. Lastly, we introduce new slack variables $\boldsymbol{\lambda}_p = [\lambda_{p,\chi}]$, $\forall\chi\in\{t,r\}$, representing the SINRs of the private streams, and $\boldsymbol{\mu}_p = [\mu_{p,\chi}]$, $\forall\chi\in\{t,r\}$. At the l-th iteration of the SCA, the non-convex constraint (17b) can be approximated by

$$C_\chi + \log_2\left(1+\lambda_{p,\chi}\right) \geq R_\chi^{th}, \tag{37}$$

$$\frac{2\mathcal{R}\{((\boldsymbol{\phi}_\chi^{[l]}\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi}+\delta_{\chi\chi})^H((\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi}+\delta_{\chi\chi})\}}{\mu_{p,\chi}^{[l]}} - \frac{\left|(\boldsymbol{\phi}_\chi^{[l]}\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi}+\delta_{\chi\chi}\right|^2\mu_{p,\chi}}{(\mu_{p,\chi}^{[l]})^2} \geq \lambda_{p,\chi}, \tag{38}$$

$$\mu_{p,\chi} \geq \left|(\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\bar{\chi}}+\delta_{\chi\bar{\chi}}\right|^2+\sigma_\chi^2. \tag{39}$$

Convex approximation for constraint (17c)*:* Using new slack variables $\boldsymbol{\lambda}_c = [\lambda_{c,\chi}]$, $\forall\chi\in\{t,r\}$ and $\boldsymbol{\mu}_c = [\mu_{c,\chi}]$, $\forall\chi\in\{t,r\}$, the non-convex constraint (17c) can be approximated as:

$$C_t + C_r \leq \log_2\left(1+\lambda_{c,\chi}\right), \tag{40}$$

Algorithm 3 Algorithm for Solving STAR-RIS Phase Shift

1: **Input:** precoding vector $\mathbf{p}$, common rate $\mathbf{c}$, threshold R_χ^{th}, T/R ratios $\boldsymbol{\beta}_\chi$, tolerance ϵ_2, maximum iteration $I_{\max}$, and set the current iteration index $l=0$.
2: **Initialize:** $\boldsymbol{\phi}_t^{[0]}, \boldsymbol{\phi}_r^{[0]}, \boldsymbol{\lambda}_c^{[0]}, \boldsymbol{\lambda}_p^{[0]}, \boldsymbol{\mu}_c^{[0]}, \boldsymbol{\mu}_p^{[0]}$, set $F^{[0]} \triangleq$ objective value of (35) at $\boldsymbol{\phi}^{[0]}$.
3: **repeat**
4: $l \leftarrow l+1$;
5: Solve **(P6)** to obtain $\boldsymbol{\phi}_t^*, \boldsymbol{\phi}_r^*, \boldsymbol{\lambda}_c^*, \boldsymbol{\lambda}_p^*, \boldsymbol{\mu}_c^*, \boldsymbol{\mu}_p^*$;
6: Update $\boldsymbol{\phi}_t^{[l]} \leftarrow \boldsymbol{\phi}_t^*, \boldsymbol{\phi}_r^{[l]} \leftarrow \boldsymbol{\phi}_r^*, \boldsymbol{\lambda}_c^{[l]} \leftarrow \boldsymbol{\lambda}_c^*, \boldsymbol{\lambda}_p^{[l]} \leftarrow \boldsymbol{\lambda}_p^*, \boldsymbol{\mu}_c^{[l]} \leftarrow \boldsymbol{\mu}_c^*, \boldsymbol{\mu}_p^{[l]} \leftarrow \boldsymbol{\mu}_p^*$;
7: Compute $F^{[l]}$ (objective value of (35));
8: **until** $|F^{[l]} - F^{[l-1]}| \leq \epsilon_2$ or $l \geq L_{\max}$
9: **Output:** $\boldsymbol{\phi}_\chi^{opt}$

$$\frac{2\mathcal{R}\{((\boldsymbol{\phi}_\chi^{[l]}\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi c}+\delta_{\chi c})^H((\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi c}+\delta_{\chi c})\}}{\mu_{c,\chi}^{[l]}} - \frac{\left|(\boldsymbol{\phi}_\chi^{[l]}\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi c}+\delta_{\chi c}\right|^2\mu_{c,\chi}}{\left(\mu_{c,\chi}^{[l]}\right)^2} \geq \lambda_{c,\chi}, \tag{41}$$

$$\mu_{c,\chi} \geq \left|(\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi}+\delta_{\chi\chi}\right|^2 + \left|(\boldsymbol{\phi}_\chi\odot\boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\bar{\chi}}+\delta_{\chi\bar{\chi}}\right|^2+\sigma_\chi^2. \tag{42}$$

Then, the non-convex problem (**P5**) is transformed into the following convex optimization problem:

$$\textbf{(P6)}: \max_{\boldsymbol{\phi}_t,\boldsymbol{\phi}_r,\boldsymbol{\lambda}_c,\boldsymbol{\lambda}_p,\boldsymbol{\mu}_c,\boldsymbol{\mu}_p} \ (35), \quad \text{s.t. (34b), (37), (38), (39), (40), (41), (42).} \tag{43}$$

Problem **(P6)** is convex and can be efficiently solved using standard solvers such as CVX [32]; the overall procedure is summarized in Alg. 3. In Alg. 3, the penalty parameters V_{ϕ_t} and V_{ϕ_r} remain fixed and are specified in Sec. V.

C. Optimization of the Transmitting/Reflecting Ratio

Given the precoding vector $\mathbf{p}$, common stream split vector $\mathbf{c}$, and phase shift vector $\boldsymbol{\phi}$, remaining task is to determine the optimal T/R ratio vector $\boldsymbol{\beta} = [\boldsymbol{\beta}_t, \boldsymbol{\beta}_r]$. This leads to the following problem formulation

$$\textbf{(P7)}: \quad \text{Find} \quad \boldsymbol{\beta}_t, \boldsymbol{\beta}_r, \quad \text{s.t. (17b), (17c), (17e).} \tag{44}$$

To handle the non-convexity in (17e), we reformulate it via a penalty method to obtain **(P8)**:

$$\textbf{(P8)}: \max_{\boldsymbol{\beta}_t,\boldsymbol{\beta}_r} \ V_{\boldsymbol{\beta}}\sum_{n=1}^{N}\left(\beta_{r,n}^2+\beta_{t,n}^2-1\right), \tag{45a}$$

$$\text{s.t.} \quad \beta_{r,n}^2+\beta_{t,n}^2 \leq 1, \tag{45b}$$

$$\text{(17b), (17c).}$$

Here, $V_{\boldsymbol{\beta}}$ denotes the penalty parameter. However, the objective function in **(P8)** is non-convex and it can be approximated as

$$2V_{\boldsymbol{\beta}}\sum_{\chi\in\{t,r\}}\sum_{n=1}^{N}\left\{\beta_{\chi,n}^{[k]}\beta_{\chi,n}-\left(\beta_{\chi,n}^{[k]}\right)^2\right\}, \tag{46}$$

where the superscript $[k]$ represents the value of the variable at the k-th iteration. To further address the non-convexity of constraint (17b) and (17c), we introduce slack variables $\boldsymbol{\omega}_p = [\omega_{p,\chi}]$, $\forall \chi \in \{t, r\}$, $\boldsymbol{\omega}_c = [\omega_{c,\chi}]$, $\forall \chi \in \{t, r\}$, $\boldsymbol{\eta}_p = [\eta_{p,\chi}]$, $\forall \chi \in \{t, r\}$ and $\boldsymbol{\eta}_c = [\eta_{c,\chi}]$, $\forall \chi \in \{t, r\}$. Using these slack variables, the non-convex constraint in (17b) can be approximated as follows:

$$C_\chi + \log_2(1 + \omega_{p,\chi}) \geq R_\chi^{th}, \tag{47}$$

$$\frac{2\mathcal{R}\{((\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi^{[k]})\mathbf{t}_{\chi\chi} + \delta_{\chi\chi})^H((\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi} + \delta_{\chi\chi})\}}{\eta_{p,\chi}^{[k]}} - \frac{\left|(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi^{[k]})\mathbf{t}_{\chi\chi} + \delta_{\chi\chi}\right|^2 \eta_{p,\chi}}{\left(\eta_{p,\chi}^{[k]}\right)^2} \geq \omega_{p,\chi}, \tag{48}$$

$$\eta_{p,\chi} \geq \left|(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\bar{\chi}} + \delta_{\chi\bar{\chi}}\right|^2 + \sigma_\chi^2. \tag{49}$$

Similarly, the non-convex constraint in (17c) can be approximated by

$$C_t + C_r \leq \log_2(1 + \omega_{c,\chi}), \tag{50}$$

$$\frac{2\mathcal{R}\{((\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi^{[k]})\mathbf{t}_{\chi c} + \delta_{\chi c})^H((\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi)\mathbf{t}_{\chi c} + \delta_{\chi c})\}}{\eta_{c,\chi}^{[k]}} - \frac{\left|(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi^{[k]})\mathbf{t}_{\chi c} + \delta_{\chi c}\right|^2 \eta_{c,\chi}}{\left(\eta_{c,\chi}^{[k]}\right)^2} \geq \omega_{c,\chi}, \tag{51}$$

$$\eta_{c,\chi} \geq \left|(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\chi} + \delta_{\chi\chi}\right|^2 + \left|(\boldsymbol{\phi}_\chi \odot \boldsymbol{\beta}_\chi)\mathbf{t}_{\chi\bar{\chi}} + \delta_{\chi\bar{\chi}}\right|^2 + \sigma_\chi^2. \tag{52}$$

Finally, the T/R ratio optimization problem in (**P8**) is reformulated as follows:

$$\textbf{(P9)}: \max_{\boldsymbol{\beta}_t, \boldsymbol{\beta}_r, \boldsymbol{\eta}_c, \boldsymbol{\eta}_p, \boldsymbol{\omega}_c, \boldsymbol{\omega}_p} \quad (46), \tag{53}$$
$$\text{s.t. (45b), (47), (48), (49), (50), (51), (52).}$$

Problem **(P9)** is convex and can be solved by standard solvers (e.g., CVX [32]), as summarized in Alg. 4. The penalty parameter V_β is fixed in Alg. 4 and specified in Sec. V.

D. Solution Approach for the Primal Problem

To minimize the transmit power, the alternating optimization (AO) method is employed, which iteratively executes Algs. 1, 3, and 4 until convergence is achieved. The complete algorithm is presented in Alg. 5. This entire process starts with obtaining a feasible initial point via Alg. 2. In each iteration, given all feasible initial points, Alg. 1 optimizes the precoding vector $\mathbf{p}$ and common rate vector $\mathbf{c}$ with fixed phase shift vector $\boldsymbol{\phi}$ and T/R ratio vector $\boldsymbol{\beta}$. Then, Alg. 3 and Alg. 4 are sequentially invoked: Alg. 3 optimizes $\boldsymbol{\phi}$ while fixing $\mathbf{p}$, $\mathbf{c}$, and $\boldsymbol{\beta}$; Alg. 4 optimizes $\boldsymbol{\beta}$ while fixing $\mathbf{p}$, $\mathbf{c}$, and $\boldsymbol{\phi}$. The procedure is repeated until convergence. The convergence property is analyzed in the following section.

Algorithm 4 Algorithm for Solving T/R ratio

1: **Input:** precoding vector $\mathbf{p}$, common rate $\mathbf{c}$, phase shift vector $\boldsymbol{\phi}_\chi$, threshold R_χ^{th}, tolerance ϵ_3, maximum iteration $K_{\max}$, and set the current iteration index $k = 0$.
2: **Initialize:** $\boldsymbol{\beta}_t^{[0]}, \boldsymbol{\beta}_r^{[0]}, \boldsymbol{\eta}_c^{[0]}, \boldsymbol{\eta}_p^{[0]}, \boldsymbol{\omega}_c^{[0]}, \boldsymbol{\omega}_p^{[0]}$, set $F_\beta^{[0]} \triangleq$ objective value of (46) at $(\boldsymbol{\beta}_t^{[0]}, \boldsymbol{\beta}_r^{[0]})$.
3: **repeat**
4: $k \leftarrow k + 1$;
5: Solve **(P9)** to obtain $\boldsymbol{\beta}_t^*, \boldsymbol{\beta}_r^*, \boldsymbol{\eta}_c^*, \boldsymbol{\eta}_p^*, \boldsymbol{\omega}_c^*, \boldsymbol{\omega}_p^*$;
6: Update $\boldsymbol{\beta}_t^{[k]} \leftarrow \boldsymbol{\beta}_t^*, \boldsymbol{\beta}_r^{[k]} \leftarrow \boldsymbol{\beta}_r^*, \boldsymbol{\eta}_c^{[k]} \leftarrow \boldsymbol{\eta}_c^*, \boldsymbol{\eta}_p^{[k]} \leftarrow \boldsymbol{\eta}_p^*, \boldsymbol{\omega}_c^{[k]} \leftarrow \boldsymbol{\omega}_c^*, \boldsymbol{\omega}_p^{[k]} \leftarrow \boldsymbol{\omega}_p^*$;
7: Compute $F_\beta^{[k]}$ (objective value of (46));
8: **until** $|F_\beta^{[k]} - F_\beta^{[k-1]}| \leq \epsilon_3$ or $k \geq K_{\max}$
9: **Output:** $\boldsymbol{\beta}_\chi^{opt}$

Algorithm 5 Alternating Optimization Algorithm for **(P1)**

1: **Input:** threshold R_χ^{th}, tolerance ϵ_4, maximum iteration $H_{\max}$, and set the current iteration index $h = 0$.
2: **Initialize:** a feasible point $(\mathbf{p}, \boldsymbol{\Theta})$ using Alg. 2, set $P^{[0]} \triangleq \|\mathbf{p}_c^{[0]}\|^2 + \|\mathbf{p}_t^{[0]}\|^2 + \|\mathbf{p}_r^{[0]}\|^2$.
3: **repeat**
4: $h \leftarrow h + 1$;
5: Apply Alg. 1 for given $\boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]}$ to obtain the solution $\mathbf{p}^{[h]}$ and $\mathbf{c}^{[h]}$;
6: Apply Alg. 3 for given $\mathbf{p}^{[h]}$, $\mathbf{c}^{[h]}$, and $\boldsymbol{\beta}^{[h-1]}$ to obtain the solution $\boldsymbol{\phi}^{[h]}$;
7: Apply Alg. 4 for given $\mathbf{p}^{[h]}$, $\mathbf{c}^{[h]}$, and $\boldsymbol{\phi}^{[h]}$ to obtain the solution $\boldsymbol{\beta}^{[h]}$;
8: Compute $P^{[h]} \triangleq \|\mathbf{p}_c^{[h]}\|^2 + \|\mathbf{p}_t^{[h]}\|^2 + \|\mathbf{p}_r^{[h]}\|^2$;
9: **until** $\left|P^{[h]} - P^{[h-1]}\right| \leq \epsilon_4$ or $h \geq H_{\max}$
10: **Output:** Minimized transmit power.

IV. Convergence and Complexity Analysis

A. Convergence Analysis of Algorithms 1, 3, and 4

Let the feasible set of problem (**P2**), and (**P3**), the objective value, and the sequence generated by the Alg. 1 at the i-th iteration, be defined as follows:

$$\mathcal{X}^0 = \{\mathbf{p}, \mathbf{c} | (17b), (17c) \text{ are satisfied}\}, \tag{54}$$

$$\mathcal{X}^{[i]} = \{\mathbf{p}, \mathbf{c}, \boldsymbol{\gamma}_c, \boldsymbol{\gamma}_p, \boldsymbol{\xi}_c, \boldsymbol{\xi}_p | (19), (22), (25), (26), (28), (29) \text{ are satisfied}\}, \tag{55}$$

$$\mathcal{F}^{[i]} = \min\{||\mathbf{p}_c||^2 + ||\mathbf{p}_t||^2 + ||\mathbf{p}_r||^2 | \{\mathbf{p}, \mathbf{c}, \boldsymbol{\gamma}_c, \boldsymbol{\gamma}_p, \boldsymbol{\xi}_c, \boldsymbol{\xi}_p\} \in \mathcal{X}^{[i]}\}, \tag{56}$$

$$\mathcal{V}^{[i]} = \{\mathbf{p}^{[i]}, \mathbf{c}^{[i]}\}. \tag{57}$$

Following standard convergence arguments for SCA and AO frameworks [34], [35], and the overview in [36], we analyze feasibility, monotonic descent, and stationarity of the proposed iterative procedures.

Proposition 1. *(Feasibility) For every iteration i of Alg.1, the following conditions hold: (a) $\mathcal{X}^{[i]} \subseteq \mathcal{X}^0$. (b) $\mathcal{V}^{[i]} \in \mathcal{X}^{[i]} \cap \mathcal{X}^{[i+1]}$. (c) $\mathcal{V}^{[i]}$ is a feasible point of (**P2**).*

This property supports the convergence of Alg. 1, as demonstrated below.

Proposition 2. *(Convergence) The objective value sequence $\{\mathcal{F}^{[i]}\}$ is monotonically non-increasing, i.e., $\mathcal{F}^{[i]} \geq \mathcal{F}^{[i+1]}$. Moreover, since the objective values are bounded below, Alg. 1 is guaranteed to converge.*

Following the analysis in [35], we present the optimality result in the next proposition.

Proposition 3. *(Optimality) Let $\mathcal{V}^*$ be the limit point of the Alg. 1 and constraint approximation functions (in* (25) *and* (29)*) satisfy the Slater condition at the point $\mathcal{V}^*$. Then, $\mathcal{V}^*$ is a KKT point of problem (**P2**).*

The convergence and optimality analysis for Algs. 3 and 4 are similar to that for Alg. 1, which is omitted for brevity.

B. Convergence Analysis of Algorithm 5

The convergence analysis of the overall Alg. 5 is presented in the following proposition.

Proposition 4. *Alg. 5 is guaranteed to converge to a sub-optimal solution of (**P1**).*

Proof. Let $\mathcal{P}_1(\mathbf{p}, \mathbf{c}, \boldsymbol{\phi}, \boldsymbol{\beta})$ and $\mathcal{P}_2(\mathbf{p}, \mathbf{c})$ denote the objective functions of problems (**P1**) and (**P2**), respectively. The following relations hold:

$$\begin{aligned}
&\mathcal{P}_1(\mathbf{p}^{[h-1]}, \mathbf{c}^{[h-1]}, \boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]}) \\
&\overset{(a)}{=} \mathcal{P}_2(\mathbf{p}^{[h-1]}, \mathbf{c}^{[h-1]}) \\
&\overset{(b)}{\geq} \mathcal{P}_2(\mathbf{p}^{[h]}, \mathbf{c}^{[h]}) \\
&\overset{(c)}{=} \mathcal{P}_1(\mathbf{p}^{[h]}, \mathbf{c}^{[h]}, \boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]}) \\
&\overset{(d)}{=} \mathcal{P}_1(\mathbf{p}^{[h]}, \mathbf{c}^{[h]}, \boldsymbol{\phi}^{[h]}, \boldsymbol{\beta}^{[h]}),
\end{aligned} \tag{58}$$

where (a) is due to the fact that two objective values are the same for given $(\boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]})$; (b) is obtained because $(\mathbf{p}^{[h]}, \mathbf{c}^{[h]})$ is a minimizer of $\mathcal{P}_2(\mathbf{p}, \mathbf{c})$ at h-th iteration; (c) is due to the fact that two objective values are the same for given $(\boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]})$; (d) is due to the fact that two objective values are the same after finding feasible values of $(\boldsymbol{\phi}^{[h]}, \boldsymbol{\beta}^{[h]})$. It follows that

$$\mathcal{P}_1(\mathbf{p}^{[h-1]}, \mathbf{c}^{[h-1]}, \boldsymbol{\phi}^{[h-1]}, \boldsymbol{\beta}^{[h-1]}) \geq \mathcal{P}_1(\mathbf{p}^{[h]}, \mathbf{c}^{[h]}, \boldsymbol{\phi}^{[h]}, \boldsymbol{\beta}^{[h]}). \tag{59}$$

Therefore, the sequence of the objective values of problem (**P1**) is monotonically non-increasing. In addition, the transmit power consumption is lower bounded due to the users' QoS requirements. Alg. 5 converges to a sub-optimal solution to problem (**P1**). □

C. Complexity Analysis

We evaluate computational complexity using standard results of the SCA method and interior-point algorithms. Specifically, the number of SCA iterations required to achieve ϵ-accuracy is $\mathcal{O}\Big(\sqrt{C}\log_2(1/\epsilon)\Big)$ [16], and the complexity of solving a convex subproblem with V variables and C constraints via an interior-point method is $\mathcal{O}(V^2C)$ [37].

For Alg. 1, problem **(P2)** is solved through a sequence of convex subproblems **(P4)**. Let C_2 denote the number of constraints in **(P2)**, and let (C_4, V_4) denote the numbers of constraints and variables in **(P4)**, respectively. Thus, the overall complexity of Alg. 1 is $\mathcal{O}\big(V_4^2 C_4 \cdot \sqrt{C_2}\log_2(1/\epsilon)\big)$.

For Alg. 3, **(P4)** is similarly approximated by convex subproblems **(P6)** under the SCA framework. Let (C_6, V_6) denote the number of constraints and variables in **(P6)**. Then, the overall complexity of Alg. 3 is $\mathcal{O}\big((V_6^2 C_6 + 1/\epsilon) \cdot \sqrt{C_4}\log_2(1/\epsilon)\big)$. Alg. 4 has the same order complexity as Alg. 3, as it involves the same number of variables and constraints. Alg. 5 iteratively applies Algs. 1, 3, and 4 until convergence. Therefore, its overall complexity is the sum above, multiplied by the number of AO iterations. In contrast, the SDP-based phase optimization in [14], [15] can have complexity up to $\mathcal{O}(N^{4.5})$ due to the $N \times N$ SDP matrix, which becomes prohibitive for large N.

V. Simulation Results and Discussion

For a comprehensive performance comparison, we consider a total of nine benchmark schemes by combining MA strategies: RSMA, NOMA, and SDMA, with three surface configurations: STAR-RIS, conventional RIS, and no-RIS baseline. This enables us to validate the effectiveness of the joint design of STAR-RIS and RSMA in achieving EE and flexible transmission. To ensure a fair comparison, all baseline schemes are configured similarly to the proposed method wherever applicable.

- **NOMA:** The BS employs the same beamforming algorithm as the proposed scheme, except that no private stream is assigned to the user with a higher channel gain. Superposition coding is used, and the decoding order is predetermined by the channel gains under ideal SIC and assumed to be known.
- **SDMA:** The beamformers are designed using the same algorithm as in the proposed RSMA method. The BS transmits private streams to each user without rate splitting; i.e., no common stream is considered. No SIC is performed at the receiver, and IUI is treated as noise.
- **Conventional RIS:** It serves only the user on the reflective side of the RIS. That is, there is no power splitting for transmitted signal and reflected signal. The RIS phase shifts are optimized using the same algorithm as in the proposed scheme.

The power assignment strategies for the comparison schemes are summarized in Table I.

TABLE I: Power assignment strategy

Multiple access scheme	s_r	s_t	s_c
RSMA	◯	◯	◯
SDMA	◯	◯	×
NOMA when the user r signal is stronger	◯	×	◯
NOMA when the user t signal is stronger	×	◯	◯

s_r: private stream for user r, s_t: private stream for user t,
s_c: common stream for both users.

In the simulation, a system is considered where one user is located in front of the STAR-RIS and the other behind it. For

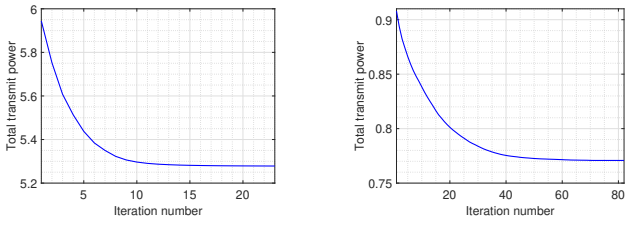
simplicity, the BS is placed at the origin, and the STAR-RIS is located at (200,0)m. User r is positioned at (190,0)m, while user t is positioned at (210,0)m. The BS-to-RIS channel $\mathbf{G}$, and RIS-to-user channels $\mathbf{h}_\chi$ are modeled as Rician fading channels with a LoS component ($\varepsilon = 10$). In contrast, the direct channels from BS to users $\mathbf{d}_\chi$ are assumed to be Rician channels without a LoS component ($\varepsilon = 0$), assuming that users are located at the cell edge.

TABLE II: Simulation Parameters

Parameters	Values
Path-loss for $\mathbf{G}$ and $\mathbf{h}_\chi$ (dB)	$35.6 + 22.0\log(D)$
Path-loss for $\mathbf{d}$ (dB)	$32.6 + 36.7\log(D)$
Noise power spectral density	-170 dBm/Hz
Rician factor ε for $\mathbf{d}_\chi$	0
Rician factor ε for $\mathbf{h}_\chi$ and $\mathbf{G}$	10
Spacing between antennas of BS d_{ant}	$0.5\lambda_{wav}, 0.4\lambda_{wav}, 0.3\lambda_{wav}$
Spacing between elements of STAR-RIS d_{RIS}	$0.5\lambda_{wav}, 0.4\lambda_{wav}, 0.3\lambda_{wav}$

On the other hand, the path loss of $\mathbf{d}_\chi$ is smaller than that of the cascaded channel through $\mathbf{G}$ and $\mathbf{h}_\chi$ so that direct channel $\mathbf{d}_\chi$ cannot be ignored if it is not an extreme situation where direct channel $\mathbf{d}_\chi$ is completely blocked. Table II summarizes the system parameters used in the simulation, which are based on the third generation partnership project (3GPP) propagation environment [38]. The penalty parameters V_{ϕ_t}, V_{ϕ_r}, and V_β control the weights of the penalty terms that enforce the unit modulus and power splitting constraints in the penalty-based reformulations. We tested $\{0.01, 0.1, 1\}$ and set $V_{\phi_t} = V_{\phi_r} = V_\beta = 0.1$, which yielded stable convergence and constraint satisfaction in our simulations. All simulations were performed over 100 times, and the average value was calculated.

Fig. 3 presents the convergence behavior for the system parameters $M = 2$, $N = 2$, and $M = 8$, $N = 8$ with $R_r^{th} = R_t^{th} = 2$ bps/Hz, respectively. It is observed that Alg. 5 converges stably with approximately 15 to 20 and 60 to 70 iterations, respectively.



(a) $M = 2$, $N = 2$ (b) $M = 8$, $N = 8$

Fig. 3: Convergence of the Alg. 5 for $R_\chi^{th} = 2$ bps/Hz.

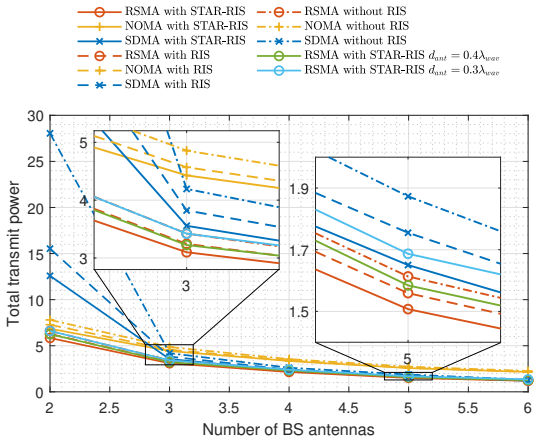
Fig. 4(a) and Fig. 4(b) evaluate the scenario in which both direct and cascaded channels are available for users r and t. These figures show the impact of the number of BS antennas and STAR-RIS elements on the total transmit power required to achieve each user's target rate. Given that both direct and cascaded channels are available for all users, the threshold rate R_χ^{th} is set to 2 bps/Hz for each user.

In Fig. 4(a), increasing the number of BS antennas reduces the transmit power for all schemes due to the higher spatial DoF. A notable drop occurs when M increases from two to three, since the BS can better separate the two users and the STAR-RIS/RIS link. The proposed STAR-RIS-aided RSMA consistently achieves the lowest transmit power. SDMA is sensitive to limit spatial DoF (e.g., M=2) due to increased IUI, whereas NOMA performs relatively well in this regime thanks to SIC; however, when the spatial DoF becomes sufficient, NOMA may degrade due to excessive SIC. RSMA mitigates these drawbacks by balancing interference decoding via the optimized power-splitting ratio $\mathbf{c}$, behaving closer to NOMA under low DoF and closer to SDMA under high DoF. Moreover, STAR-RIS-aided RSMA outperforms conventional RIS-aided RSMA because STAR-RIS can split the incident signal between the r and t users via T/R ratio control, while a conventional RIS reflects all energy toward a single side. Finally, when $d_{ant} = 0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, spatial correlation increases the average transmit power by 5.79% and 12.09%, respectively. compared with the uncorrelated case.

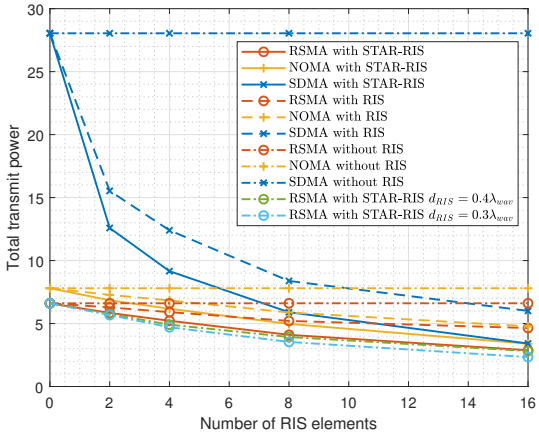
In Fig. 4(b), as the number of STAR-RIS elements increases, the following observations are made. RSMA consistently consumes less total transmit power than SDMA and NOMA for all numbers of RIS elements. First, as the number of RIS elements increases, the total transmit power decreases across all schemes, except for those without RIS or STAR-RIS, which is expected since more RIS elements provide higher passive array gains and improved channel gains. Second, the proposed RSMA with the STAR-RIS maintains the lowest transmit power consumption. Third, RSMA with STAR-RIS outperforms RSMA with conventional RIS. The performance gap increases because the conventional RIS reflects the entire signal to the user r while the STAR-RIS distributes the signal to users r and t through T/R ratio control. Lastly, when the distance between STAR-RIS elements d_{RIS} is $0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, the power consumption decreases by an average of 3.45% and 11.37%, respectively, compared to the case without spatial correlation. Specifically, spatial correlation at RIS can be leveraged to reduce total transmit power by enhancing the effective channel SINR.

Second, Fig. 5(a) and Fig. 5(b) evaluate the scenario in which the direct channel is completely blocked. In this case, users t and r are served only through the cascaded channels via the STAR-RIS. Considering that only cascaded channels are available, the threshold rate R_χ^{th} is set to 0.2 bps/Hz for each user.

Fig. 5(a) and Fig. 5(b) present the following observations. First, the proposed RSMA-based approach performs similarly to NOMA, regardless of the number of BS antennas. Without direct channels, the cascaded channel conditions for user r and t are similar, which reduces the performance gain of RSMA over NOMA [39]. Second, the RSMA-based scheme demonstrates superior performance compared to the SDMA-based scheme. However, the performance gap narrows as the number of STAR-RIS elements increases due to the saturation of diversity gain. On average, the RSMA-based approach achieves 11.77% and 17.18% lower power consumption than the SDMA-based approach as the number of BS antennas and RIS elements increases, respectively. Lastly, in Fig. 5(a), when the distance between BS antennas d_{ant} is set to $0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, the power consumption decreases by an average of

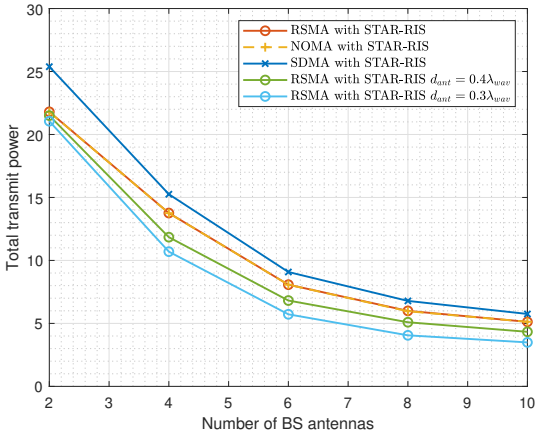


(a) Total transmit power vs. the number of BS antennas when $N = 2$, and $R_{\chi}^{th} = 2$ bps/Hz.

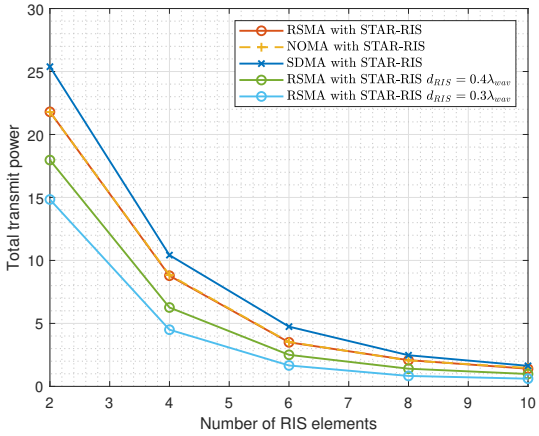


(b) Total transmit power vs. the number of RIS elements when $M = 2$, and $R_{\chi}^{th} = 2$ bps/Hz.

Fig. 4: Total transmit power vs. the number of BS antennas or STAR-RIS elements.



(a) Total transmit power vs. the number of BS antennas when $N = 2$, $R_{\chi}^{th} = 0.2$ bps/Hz and direct channels $\mathbf{d}_{\chi}$ are completely blocked.



(b) Total transmit power vs. the number of RIS elements when $M = 2$, $R_{\chi}^{th} = 0.2$ bps/Hz and direct channels $\mathbf{d}_{\chi}$ are completely blocked.

Fig. 5: Total transmit power vs. the number of BS antennas or STAR-RIS elements when direct channels are completely blocked.

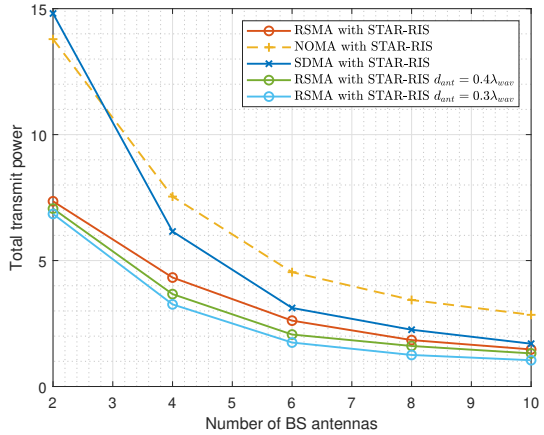
12.28% and 23.81% compared to the case without spatial correlation, respectively. In Fig. 5(b), when the distance between STAR-RIS elements d_{RIS} is set to $0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, the power consumption decreases by an average of 27.45% and 49.97% compared to the case without spatial correlation, respectively.

Lastly, Fig. 6(a) and Fig. 6(b) evaluate the scenario in which the direct channel of the user t is blocked. In this case, user t is served only through the cascaded channel via the STAR-RIS, while user r has both the direct channel and the cascaded channel. The target rate for the user r is set to $R_r^{th} = 2$ bps/Hz. For user t, considering that only the cascaded channel is available, the target rate is set to $R_t^{th} = 0.2$ bps/Hz.

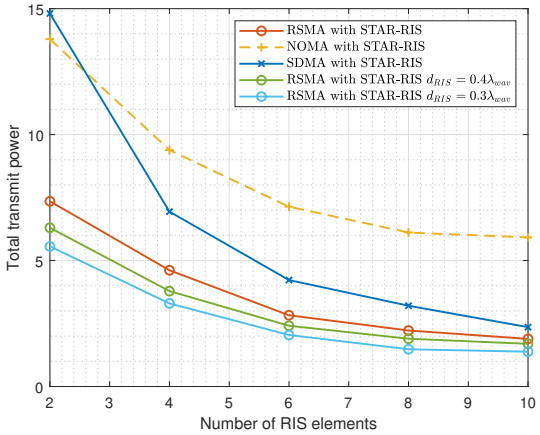
In Fig. 6(a), the proposed RSMA-based approach achieves, on average, 45.27% lower power consumption compared to the NOMA-based approach and 25.56% compared to the SDMA-based approach. When the distance between BS antennas d_{ant} is set to $0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, the power consumption decreases by on average of 12.72% and 25.17% compared to the case without spatial correlation, respectively.

In Fig. 6(b), the proposed RSMA-based approach achieves, on average, 57.91% and 38.26% lower power consumption compared to NOMA- and SDMA-based approaches, respectively. When the distance between STAR-RIS elements d_{RIS} is set to $0.4\lambda_{wav}$ and $0.3\lambda_{wav}$, the power consumption decreases by an average of 14.39% and 28.19% compared to the case without spatial correlation, respectively.

Fig. 7(a) and Fig. 7(b) illustrate the impact of spatial correlation, arising from BS antenna spacing and STAR-RIS element spacing, on the total transmit power required to meet users' QoS constraints. In Fig. 7(a), increasing BS spatial correlation causes the channel vectors of different users to become more aligned, reducing the effectiveness of precoding in separating users and intensifying IUI. Consequently, higher transmit power is required to satisfy the rate constraints, leading to the performance degradation observed in Fig. 7(a). RSMA is generally less sensitive to IUI than SDMA since a portion of the interference can be managed via the common stream and

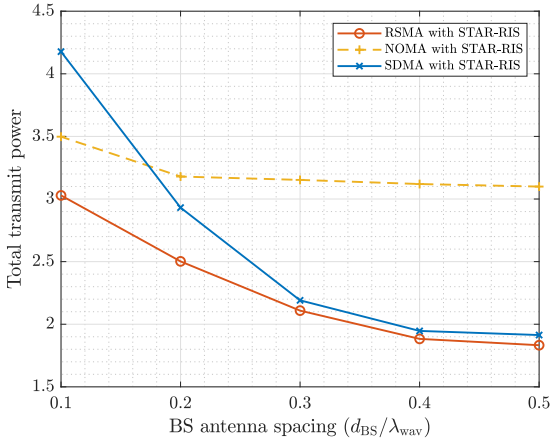


(a) Total transmit power vs. the number of BS antennas when $N = 2$, $R_r^{th} = 2$, $R_t^{th} = 0.2$ bps/Hz, and direct channel for t-user $\mathbf{d}_t$ is blocked.

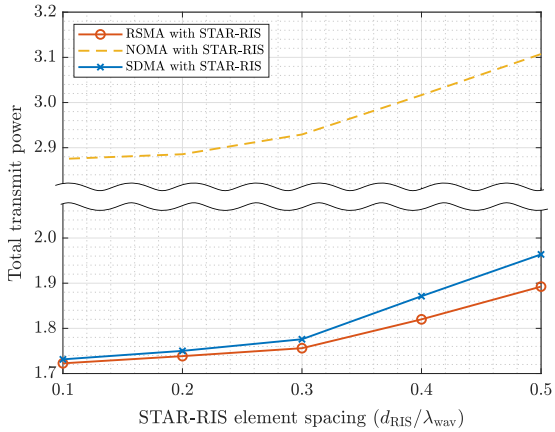


(b) Total transmit power vs. the number of RIS elements when $M = 2$, $R_r^{th} = 2$, $R_t^{th} = 0.2$ bps/Hz, and direct channel for t-user $\mathbf{d}_t$ is blocked.

Fig. 6: Total transmit power vs. the number of BS antennas or STAR-RIS elements when the direct channel for t-user $\mathbf{d}_t$ is blocked.



(a) Total transmit power vs. BS antenna spacing when $M = 4$, $N = 4$, $R_\chi^{th} = 2$ bps/Hz.



(b) Total transmit power vs. STAR-RIS element spacing when $M = 4$, $N = 4$, $R_\chi^{th} = 2$ bps/Hz.

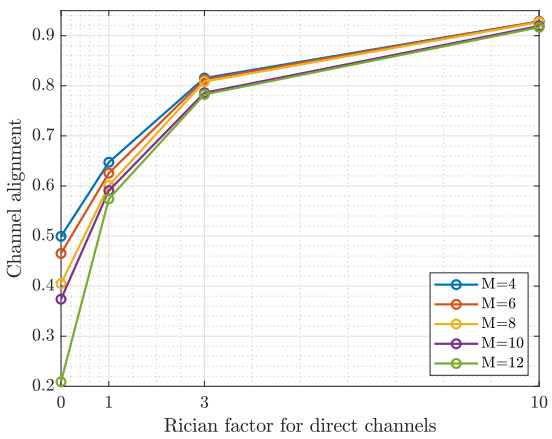
Fig. 7: Total transmit power vs. BS antenna spacing or STAR-RIS element spacing.

SIC. As the BS spatial correlation decreases, the performance gap between RSMA and SDMA becomes negligible beyond a certain spacing regime (around $d_{BS}/\lambda \approx 0.3$ in our setup). This suggests that SDMA gradually recovers its multi-user spatial separability, thereby reducing RSMA's relative advantage. In contrast, SDMA relies heavily on spatial separation and is thus more vulnerable under strong correlation. NOMA exhibits a different sensitivity, as it exploits power-domain separation and SIC instead of relying on spatial separation.

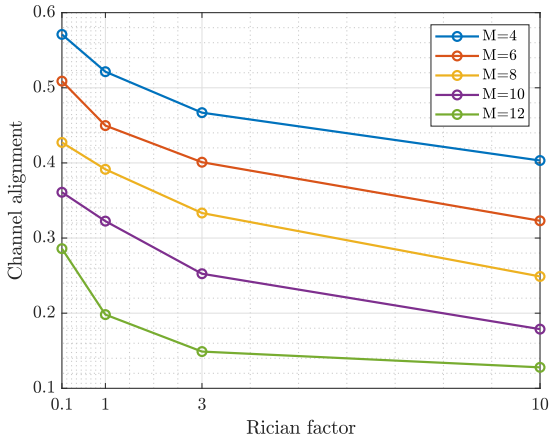
Fig. 7(b) illustrates the impact of spatial correlation at the STAR-RIS. For single-antenna users and passive reflection, reducing the STAR-RIS element spacing increases correlation among the reflected components, enabling more coherent combining toward the intended direction. This results in a more focused passive beam and improved effective array gain, thereby reducing the required transmit power, as shown in Fig. 7(b). Similar observations have been reported in RIS-assisted MU-MISO systems, where stronger spatial correlation across RIS elements can be beneficial for passive beamforming designs [40].

To further evaluate system performance under various Rician factors of direct links ε_d, we consider scenarios where the direction of the t user and r users are either similar or different. For the similar-direction case, the r user is placed at $(190, 0)$, and the t user is placed at $(210, 0)$. For the different-direction case, the r user is placed at $(190, 40)$, and the t user is placed at $(210, -40)$. Fig. 8 shows the initial channel similarity between the two users, measured as $\rho \triangleq \frac{|\mathbf{v}_r^H \mathbf{v}_t|}{||\mathbf{v}_r||_2||\mathbf{v}_t||_2}$, as a function of the number of BS antennas before applying any optimization.

Fig. 8(a) illustrates a scenario where the two users are closely aligned along the LoS direction. As the Rician factor ε_d increases, the channels become highly collinear, sharply increasing ρ. This leads to poor spatial separability, degrading SDMA performance. A larger M further aligns the beams, reinforcing this effect. In contrast, Fig. 8(b) considers users with distinct angles. Even as ε_d increases, the channels remain

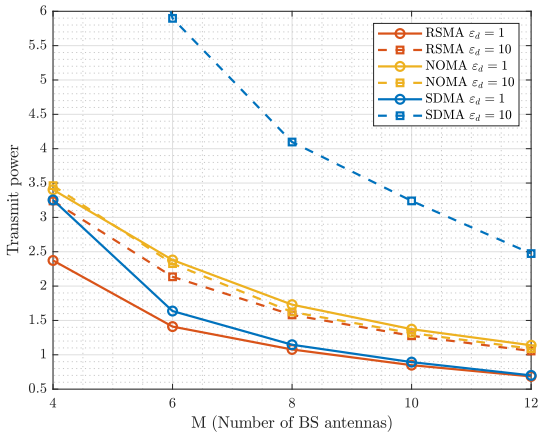


(a) Channel alignment ρ vs. Rician factor for direct link ε_d, where the r user is located at (190, 0) and the t user is located at (210, 0).

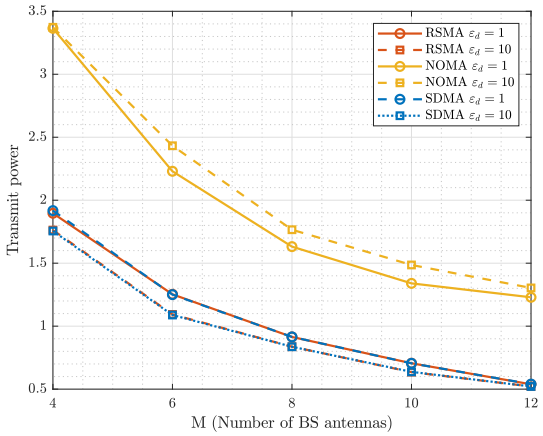


(b) Channel alignment ρ vs. Rician factor for direct link ε_d, where the r user is located at (190, 40) and the t user is located at (210, -40).

Fig. 8: Channel alignment based on user location.



(a) Total transmit power vs. the number of BS antennas M, where the r user is located at (190, 0) and the t user is located at (210, 0).



(b) Total transmit power vs. the number of BS antennas M, where the r user is located at (190, 40) and the t user is located at (210, -40).

Fig. 9: Total transmit power based on Rician factor.

dissimilar, resulting in a lower ρ. Here, increasing M improves angular resolution, further separating the channels. This setting thus offers a more favorable geometry for SDMA.

Fig. 9 shows the total transmit power as a function of user location, the number of BS antennas, and the Rician factor ε_d of the direct link, with the number of STAR-RIS elements fixed at $N = 4$. When the users are located in similar directions, as in Fig. 9(a), increasing M reduces the required transmit power thanks to array gain. However, with strong LoS components, the user channels become highly aligned, degrading spatial separability. In this regime, SDMA suffers from ineffective interference suppression, leading to higher power consumption. RSMA remains robust by splitting messages and decoding interference via SIC, achieving lower power, especially at small-to-moderate M. While NOMA is less affected by spatial alignment due to its power-domain nature, it still requires more power due to SIC overhead and conservative power allocation. In contrast, Fig. 9(b) considers users with distinct angular directions. The improved spatial separability enables SDMA to perform effectively, achieving low power. RSMA remains competitive; the advantage of the SIC structure is minimal when spatial separation is sufficient, and it operates similarly to SDMA. NOMA continues to demand the most power, as it cannot fully exploit spatial DoF and must maintain sufficient decoding margins under SIC.

To evaluate the performance of the proposed design, under practical STAR-RIS hardware constraints, we consider two non-idealities: (i) finite resolution phase shifters, and (ii) coupled T/R phase (CP) responses. In practice, STAR-RIS elements employ b-bit discrete phase shifters, and due to passive circuit constraints, the t/r coefficients cannot be fully decoupled [41]. These impairments restrict the feasible STAR-RIS coefficient set and may increase the minimum transmit power required to meet QoS constraints [42].

- **Discrete phase quantization model:** The phase of each STAR-RIS element is quantized to a finite alphabet $\mathcal{F}_b = \{0, \Delta, 2\Delta, ..., (2^b - 1)\Delta\}$, where $\Delta = 2\pi/2^b$.
- **Coupled phase constraint (CP) model:** For passive and

lossless STAR-RIS, the t/r phases are coupled as $\theta_{t,n} = \theta_{r,n} \pm \pi/2$, imposing a fixed offset.

To evaluate their impact, we first solve the original optimization under the ideal continuous and decoupled STAR-RIS model. Then, we apply two hardware-constrained mappings: (i) phase quantization and (ii) CP enforcement, using Euclidean projection to approximate the ideal coefficients. Since this may render the solution infeasible under power constraints, we re-optimize the transmit power precoders with the fixed, mapped STAR-RIS coefficients to restore QoS satisfaction and quantify the resulting power penalty. Fig. 10 shows that lower phase resolution increases transmit power due to limited beamforming flexibility, with the highest penalty at 2-bits. As resolution improves, the power gap narrows, and the 5-bit case approaches the ideal. Enforcing the CP constraint further raises power, confirming that practical STAR-RIS hardware imposes stricter limits on performance.

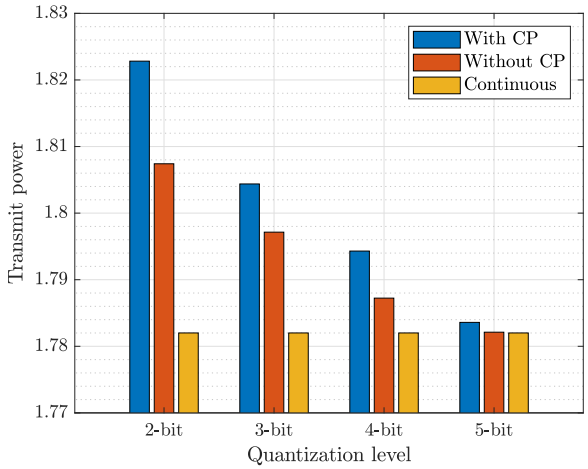


Fig. 10: Total transmit power vs. phase resolution and coupled phase constraint with $M = 4, N = 4$, and $R_{\chi}^{th} = 2$ (in bps/Hz).

To evaluate the impact of channel estimation errors in passive STAR-RIS systems, we adopt a normalized mean squared error (NMSE) based bounded-error model [43] for the effective channel $\mathbf{v}_\chi = \widehat{\mathbf{v}}_\chi + \mathbf{e}_\chi$, where $\|\mathbf{e}_\chi\|_2 \le \eta_\chi$, and $\widehat{\mathbf{v}}_\chi$ denotes the estimated channel while $\mathbf{e}_\chi$ represents the estimation error. The NMSE is defined as

$$\mathrm{NMSE} \triangleq \frac{\mathbb{E}\big[\|\mathbf{v}_\chi - \widehat{\mathbf{v}}_\chi\|_2^2\big]}{\mathbb{E}[\|\mathbf{v}_\chi\|_2^2]}, \quad \mathrm{NMSE}_{\mathrm{lin}} = 10^{\mathrm{NMSE}_{\mathrm{dB}}/10}. \tag{60}$$

Following a standard conservation mapping from NMSE to a deterministic uncertainty set, we set $\|\mathbf{e}_x\|_2 \le \eta_x$ with $\eta_x = \sqrt{\mathrm{NMSE}_{\mathrm{lin}}}\,\|\hat{\mathbf{v}}_x\|_2$. For a given SINR target γ, the worst-case common-stream SINR constraint for user χ is written as

$$\min_{\|\mathbf{e}_\chi\|_2 \le \eta_\chi} \frac{|(\widehat{\mathbf{v}}_\chi + \mathbf{e}_\chi)^H \mathbf{p}_c|^2}{|(\widehat{\mathbf{v}}_\chi + \mathbf{e}_\chi)^H \mathbf{p}_r|^2 + |(\widehat{\mathbf{v}}_\chi + \mathbf{e}_\chi)^H \mathbf{p}_t|^2 + \sigma^2} \ge \gamma. \tag{61}$$

We omit the private stream SINR case for brevity.

By rearranging (61), the constraint can be equivalently expressed as a quadratic form:

$$\mathbf{e}_\chi^H \mathbf{A} \mathbf{e}_\chi + 2\Re\{\mathbf{b}^H \mathbf{e}_\chi\} + c \ge 0, \quad \forall\ \|\mathbf{e}_\chi\|_2^2 \le \eta_\chi^2, \tag{62}$$

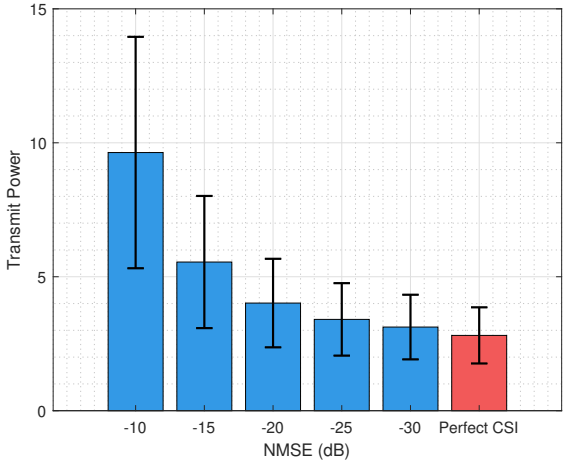


Fig. 11: Total transmit power vs. NMSE with $M = 4, N = 4$, and $R_{\chi}^{th} = 2$ (in bps/Hz).

for some $\mathbf{A}$, $\mathbf{b}$, and c determined by $\{\mathbf{p}_c, \mathbf{p}_\chi\}$, γ, and $\widehat{\mathbf{v}}_\chi$. Applying the S-lemma yields a tractable sufficient condition (which is exact under mild regularity conditions) in the form of the following linear matrix inequality (LMI):

$$\exists \tau_k \ge 0 \text{ s.t. } \begin{bmatrix} \mathbf{A} + \tau_k \mathbf{I} & \mathbf{b} \\ \mathbf{b}^H & c - \tau_k \eta_k^2 \end{bmatrix} \succeq \mathbf{0}. \tag{63}$$

Therefore, for each NMSE level, the robust transmit beamforming step can be formulated as an SDP with LMI constraints [32]. This reveals a robustness-power trade-off: larger NMSE (i.e., larger η_k) requires higher transmit power to meet the QoS constraints, as quantified in Fig. 11.

To address scalability, we extend the proposed framework to a general K-user STAR-RIS-assisted MU-MISO system, where MU interference naturally arises. The transition from single-user to single-layer RSMA in the MU-MISO setting is straightforward. In **(P1)**, the power objective generalizes to $||\mathbf{p}_c||^2 + \sum_{k=1}^{K} ||\mathbf{p}_\mathbf{k}||^2$, ans SIC constraint for the common stream are enforced across all users to ensure decodability.

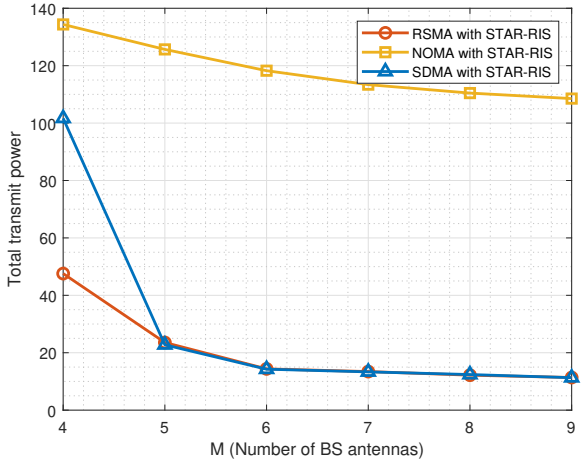


Fig. 12: Total transmit power vs. the number of BS antennas, with $K = 4, N = 10$, and $R_k^{th} = 2$ (in bps/Hz).

When evaluating the SINR of the common stream at user k, all K private streams are treated as interference. For private streams, only the other users' streams ($j \neq k$) interfere after common stream cancellation. Thus, the RSMA formulation naturally extends to the MU-MISO case by scaling the power and interference terms with K, while maintaining the same decoding structure. Consequently, the MU optimization problem is formulated as follows:

$$\textbf{(P1-MU)}: \min_{\mathbf{p},\mathbf{c},\boldsymbol{\phi},\boldsymbol{\beta}} \quad ||\mathbf{p}_c||^2 + \sum_{k=1}^{K} ||\mathbf{p}_k||^2, \ \forall k \in \{1,...,K\} \tag{64a}$$

$$\text{s.t. } C_k + R_{k,\chi} \geq R_k^{th}, \qquad \forall k \in \{1,...,K\}, \ \forall \chi \in \{t,r\}, \tag{64b}$$

$$\sum_{k=1}^{K} C_k \leq \min R_{c,k}, \ \forall k \in \{1,...,K\}, \tag{64c}$$

$$|\phi_{\chi,n}| = 1, \quad \forall \chi \in \{t,r\}, \ \forall n \in \{1,...,N\}, \tag{64d}$$

$$\beta_{r,n}^2 + \beta_{t,n}^2 = 1, \quad \forall n \in \{1,...,N\}. \tag{64e}$$

To evaluate the proposed algorithm in the MU-MISO system, we conduct simulations with $K = 4$ users, where two users are located in the reflection region and the other two in the transmission region of the STAR-RIS. As baselines, we consider K-user SDMA and power-domain NOMA using a single cluster (i.e., without user grouping) to simplify the comparison and highlight the scalability limitation of deep SIC in MU scenarios.[1]

Fig. 12 illustrates the total transmit power versus the number of BS antennas M in a $K = 4$ STAR-RIS-assisted MU-MISO system. As M increases, both RSMA and SDMA significantly reduce the required power, as the additional spatial DoF enhance the separability of user channels and facilitate more effective MU interference suppression. In particular, SDMA exhibits a sharp drop in transmit power when M becomes comparable to or exceeds K, reflecting a transition from a heavily interference-limited regime to a spatially resolvable one. RSMA consistently achieves lower or comparable power than SDMA, especially in the small-array regime, as the common stream enables additional interference management and mitigates the need for strict spatial orthogonality. In contrast, single-cluster power-domain NOMA exhibits only marginal improvements with increasing M, indicating that deep SIC and residual IUI limit its scalability in multi-user scenarios. As a result, NOMA requires significantly higher transmit power to meet the same QoS constraints.

VI. Conclusion

This paper investigated a power-efficient STAR-RIS-aided RSMA system. The transmit power minimization problem was formulated to ensure that users' minimum target rates are satisfied under spatially correlated channels. The formulated problem is non-convex and involves joint optimization of the precoding vector, common message ratio, phase shift matrix, and the T/R ratio. The original problem was decomposed into three non-convex subproblems: (i) joint precoding vector and common message ratio control at the BS, (ii) phase shift matrix control at the STAR-RIS, and (iii) T/R ratio control at the STAR-RIS. Each subproblem was transformed into convex subproblems using penalty functions and linear approximations. They were solved iteratively using the proposed interior-point initialization, SCA, and AO algorithms. It was shown that the proposed approach has polynomial computational complexity and converges to a stationary point stably and rapidly. Extensive simulations confirm that the proposed control scheme achieves the lowest power consumption while satisfying QoS requirements, compared to legacy STAR-RIS and RIS-aided RSMA, NOMA, and SDMA schemes.

[1] While practical deployments often employ user grouping, the single-cluster setting is a commonly used stress-test baseline. Under this setting, the required transmit power increases rapidly with K due to accumulated SIC constraints.

References

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[2] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.

[3] C. Wu, Y. Liu, X. Mu, X. Gu, and O. A. Dobre, "Coverage characterization of STAR-RIS networks: NOMA and OMA," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3036–3040, 2021.

[4] 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.

[5] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on STAR-RIS: Use cases, recent advances, and future research challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14 689–14 711, 2023.

[6] Z. Zhao, Z. Yang, M. Chen, C. Zhu, W. Xu, Z. Zhang, and K. Huang, "Energy-efficient probabilistic semantic communication over space-air-ground integrated networks," *IEEE Trans. Wireless Commun.*, vol. 24, no. 10, pp. 8814–8829, 2025.

[7] Z. Zhao, Z. Yang, Y. Hu, C. Zhu, M. Shikh-Bahaei, W. Xu, Z. Zhang, and K. Huang, "Compression ratio allocation for probabilistic semantic communication with RSMA," *IEEE Trans. Commun.*, vol. 73, no. 9, pp. 7304–7318, 2025.

[8] Z. Zhao, Z. Yang, C. Huang, L. Wei, Q. Yang, C. Zhong, W. Xu, and Z. Zhang, "A joint communication and computation design for distributed RIS-assisted probabilistic semantic communication in IIoT," *IEEE Internet Things J.*, vol. 11, no. 16, pp. 26 568–26 579, 2024.

[9] M. Amiri, E. Vaezpour, S. Javadi, M. R. Mili, H. Yanikomeroglu, and M. Bennis, "Multi-objective optimization in STAR-RIS-aided SWIPT with RSMA via meta-learning," *arXiv preprint arXiv:2401.07644*, 2024.

[10] F. Xiao, P. Chen, S. Xu, X. Pang, and H. Liu, "Physical layer security of STAR-RIS-aided RSMA systems," *Phys. Commun.*, vol. 61, 2023, Art. no. 102192.

[11] Q. Gao, Y. Liu, X. Mu, M. Jia, D. Li, and L. Hanzo, "Joint location and beamforming design for STAR-RIS assisted NOMA systems," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2532–2546, 2023.

[12] M. Shehab, B. S. Ciftler, T. Khattab, M. M. Abdallah, and D. Trinchero, "Deep reinforcement learning powered IRS-assisted downlink NOMA," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 729–739, 2022.

[13] T. Wang, F. Fang, and Z. Ding, "An SCA and relaxation based energy efficiency optimization for multi-user RIS-assisted NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6843–6847, 2022.

[14] X. Li, Z. Xie, G. Huang, J. Zhang, M. Zeng, and Z. Chu, "Sum rate maximization for RIS-aided NOMA with direct links," *IEEE Netw. Lett.*, vol. 4, no. 2, pp. 55–58, 2022.

[15] S. Khisa, M. Elhattab, C. Assi, and S. Sharafeddine, "Energy consumption optimization in RIS-assisted cooperative RSMA cellular networks," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 4300–4312, 2023.

[16] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, "Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface," in *Proc. IEEE Int. Conf. Commun. (ICC) Workshops*, 2020, pp. 1–6.

[17] T. Phung Truong, T. My Tuyen Nguyen, T. Vi Nguyen, N.-N. Dao, and S. Cho, "Energy efficiency in RSMA-enhanced active RIS-aided quantized downlink systems," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 3, pp. 834–850, 2025.

[18] F. Fang, B. Wu, S. Fu, Z. Ding, and X. Wang, "Energy-efficient design of STAR-RIS aided MIMO-NOMA networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 498–511, 2023.

[19] J. Zhao, Q. Xu, X. Mu, Y. Liu, and Y. Zhu, "Aerial active STAR-RIS-aided IoT NOMA networks," *IEEE Internet Things J.*, vol. 12, no. 8, pp. 9525–9538, 2025.

[20] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Improved spectral efficiency in STAR-RIS aided uplink communication using rate splitting multiple access," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5365–5382, 2023.

[21] C. Meng, K. Xiong, W. Chen, B. Gao, P. Fan, and K. B. Letaief, "Sum-rate maximization in STAR-RIS-assisted RSMA networks: A PPO-based algorithm," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 5667–5680, 2024.

[22] Y. Maghrebi, M. Elhattab, C. Assi, A. Ghrayeb, and G. Kaddoum, "Cooperative rate splitting multiple access for active star-ris assisted downlink communications," *IEEE Wireless Commun. Lett.*, vol. 13, no. 10, pp. 2827–2831, 2024.

[23] B. Wang, X. Tao, S. Han, K. Yang, and H. Wu, "Secure transmission design for rate-splitting empowered STAR-RIS-aided networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 9, pp. 2581–2585, 2024.

[24] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, 2007.

[25] H. Wang, C. Liu, Z. Shi, Y. Fu, and R. Song, "On power minimization for IRS-aided downlink NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1808–1811, 2020.

[26] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 830–834, 2021.

[27] S. Dhok and P. K. Sharma, "Rate-splitting multiple access with STAR RIS over spatially-correlated channels," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6410–6424, 2022.

[28] M. F. U. Abrar, M. Talha, R. I. Ansari, S. A. Hassan, and H. Jung, "Optimization of STAR-RIS-assisted hybrid NOMA mmWave communication," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10 146–10 161, 2023.

[29] T. P. Truong, H. V. Nguyen, N.-N. Dao, W. Noh, and S. Cho, "Orthogonalized RSMA-based flexible multiple access in digital twin edge networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 18 740–18 756, 2024.

[30] A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, and S. Chatzinotas, "STAR-RIS assisted cell-free massive MIMO system under spatially-correlated channels," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 3932–3948, 2024.

[31] E. Björnson and Ö. T. Demir, *Introduction to multiple antenna communications and reconfigurable surfaces*. Hanover, MD, USA: Now Publ. Inc., 2024.

[32] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] X. Yu, D. Xu, D. W. K. Ng, and R. Schober, "IRS-assisted green communication systems: Provable convergence and robust optimization," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6313–6329, 2021.

[34] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[35] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Minnesota, 2014.

[36] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, 2023.

[37] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193–228, 1998.

[38] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.814, 2010.

[39] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 179–189, 2020.

[40] Z. Abdullah, A. Papazafeiropoulos, S. Kisseleff, S. Chatzinotas, and B. Ottersten, "Impact of phase-noise and spatial correlation on double-ris-assisted multiuser miso networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1473–1477, 2022.

[41] Y. Liu, Y. Ni, H. Zhao, Y. Yang, S. Qiao, and H. Zhu, "STAR-RIS assisted wireless communications with discrete phase shifter," *IEEE Wireless Commun. Lett.*, vol. 13, no. 11, pp. 3152–3156, 2024.

[42] J.-C. Chen, "Designing STAR-RIS-assisted wireless systems with coupled and discrete phase shifts: A computationally efficient algorithm," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10 772–10 777, 2024.

[43] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 714–725, 2009.

Donghyun Lee received his B.S. degree in Computer Science and Engineering from Dongguk University, South Korea, in 2020. He received M.S. degrees in Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2022. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include wireless networks, energy-efficient networks, multimedia communications, and multiple access.

Chihyun Song received a B.S. degree in Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2023, where he is currently pursuing an M.S. degree in Computer Science and Engineering. His current research interests include resource allocation, massive mimo, multiple access, and cell-free networks.

The Vi Nguyen received the B.S. degree in Mathematics from University of Science, Ho Chi Minh City, Viet Nam in 2016, and M.S. degree in Computer Science and Engineering from Chung-Ang University, South Korea in 2021. He is currently pursuing Ph.D. in School of Computer Science and Engineering at Chung-Ang University, South Korea. His research interests include reconfigurable intelligent surface, multiple access techniques, MIMO systems, and optimization methods in wireless communications.

Junsuk Oh received the B.S. and M.S. degrees in Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2021 and 2023, respectively. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include wireless communications and networks, parallel and distributed computing, over-the-air computation, data compression, machine learning, and federated learning.

Yunseong Lee received B.S., M.S., and Ph.D. degrees in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2013, 2015, and 2025, respectively. From 2025 to 2026, he conducted his postdoctoral research at AI-Transformed Aerospace Research Center, Korea Advanced Institute of Science and Technology, Daejeon, South Korea. He is currently an Assistant Professor with the School of Computer Science, Engineering, and Converged Technology, Duksung Women's University, Seoul. His current research interests include aerospace networks, wireless communication, handover control, AI-RAN, and cell-free networks.

Dongwook Won received a B.S. and M.S. degree in Electronics and Communications Engineering from Kwangwoon University, Seoul, Korea in 2020 and 2022, respectively. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include Semantic Communications, 5G/6G Networks, and Satellite Communication.

Wonjong Noh (Senior Member, IEEE) received the B.S., M.S. and Ph.D. degrees from the Department of Electronics Engineering, Korea University, Seoul, Korea, in 1998, 2000, and 2005, respectively. From 2005 to 2007, he conducted his postdoctoral research at Purdue University, IN, USA, and the University of California at Irvine, CA, USA. From 2008 to 2014, he was a Principal Research Engineer with Samsung Advanced Institute of Technology, Samsung Electronics, Korea. From 2014 to 2018, he was an Assistant Professor at the Department of Electronics and Communication Engineering, Gyeonggi University of Science and Technology, Korea. He is currently an Associate Professor at the School of Software at Hallym University, Korea. His current research interests include fundamental capacity analysis in 5G/6G wireless communication and networks, federated mobile edge computing, and machine learning-based systems control. He received a government postdoctoral fellowship from the Ministry of Information and Communication, Korea, in 2005. He was also a recipient of the Samsung Best Paper Gold Award in 2010, the Samsung Patent Bronze Award in 2011, and the Samsung Technology Award in 2013.

Sungrae Cho (Senior Member, IEEE) received his B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and his Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002.

He is a Professor at the School of Computer Science and Engineering, Chung-Ang University (CAU), Seoul, South Korea. Prior to joining CAU, he was an Assistant Professor with the Department of Computer Sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a Senior Member of the Technical Staff at the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a Research Staff Member at the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a Visiting Professorship at the National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA. His current research interests include wireless networking, ubiquitous computing, and ICT convergence. He has served numerous international conferences as an Organizing Committee Chair, such as *IEEE SECON*, *ICOIN*, *ICTC*, *ICUFN*, *TridentCom*, and *IEEE MASS*, as well as a Program Committee Member for conferences such as *IEEE ICC*, *MobiApps*, *SENSORNETS*, and *WINSYS*. He has been a Subject Editor for *IET Electronics Letter* since 2018 and was an Editor for *Ad Hoc Networks Journal* (Elsevier) from 2012 to 2017.