Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Robust design for IRS-assisted multiuser systems under practical imperfections: a rate-splitting approach*

Xingyu PENG[1], Qin TAO[2], Xiaoming CHEN‡[1]

[1]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

[2]School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

E-mail: peng_xingyu@zju.edu.cn; taoqin@hznu.edu.cn; chen_xiaoming@zju.edu.cn

Received May 30, 2025; Revision accepted Oct. 8, 2025; Crosschecked

Unedited

Abstract: In practical intelligent reflecting surface (IRS) assisted multiuser communication systems, inevitable imperfections such as hardware impairments, imperfect channel state information (CSI), and the limited resolution of the IRS phase shift would introduce interference and thus cause significant performance degradation. As an interference management strategy, rate-splitting multiple access (RSMA) employs the rate-splitting (RS) principle to partition user information into common and private parts, thereby offering enhanced robustness. Accounting for practical imperfections, this study investigates robust beamforming design in IRS-assisted multiuser systems under the RSMA architecture. First, we introduced a system model capturing these non-ideal factors and evaluated their impacts on communication performance. To enhance the performance of the considered system, a weighted sum rate maximization optimization problem is formulated, where a sample average approximation (SAA)-based robust algorithm is proposed to jointly optimize the IRS phase shifts and the beamforming matrix at the base station (BS). Simulation results demonstrate that the IRS-assisted RSMA system exhibits superior robustness compared to the IRS-assisted space division multiple access (SDMA) system in the presence of inevitable imperfections. Furthermore, the proposed SAA-based robust algorithm outperforms existing benchmark algorithms, highlighting its effectiveness and robustness.

Key words: Intelligent reflecting surface; Rate splitting multiple access; Discrete phase shifts; Hardware impairments; Imperfect channel state information; Robust beamforming

https://doi.org/10.1631/FITEE.2500353 **CLC number:**

1 Introduction

1.1 Background

The next-generation wireless communication systems aim to enhance existing 5G technologies while exploring novel solutions to support ultra-high data rates and massive connectivity (Tataria et al., 2021). Recently, intelligent reflecting surface (IRS) has emerged as a key technology for next-generation systems due to the low cost, wide coverage, and ability to adaptively control wireless propagation environments (Wu and Zhang, 2020; Wu et al., 2021; Zhang et al., 2023). The IRS is typically composed of a large array of passive reflecting elements, each capable of independently adjusting the amplitude and phase of the incident signal. Therefore, substantial passive gains can be provided without requiring additional power or active signal processing. Owing to these advantages, IRS can be seamlessly integrated with various advanced technologies to enhance system performance, such as massive multipe input multiple output (MIMO), integrated sensing and communication (ISAC), movable antenna (MA) technique, and orthogonal time frequency space (OTFS) modulation (Xu et al., 2021; Ma et al., 2024; Peng et al., 2024a,b).

Despite its promising potential, the performance of the IRS-assisted system suffers significant degradation

‡ Corresponding author

* Project supported by the National Science Foundation of China (Nos. 62231009, U21A20443, and 62301206), and the Zhejiang Provincial Natural Science Foundation (No. LQ24F010005).

ORCID: Xingyu PENG, https://orcid.org/0000-0002-0311-3583; Xiaoming CHEN, https://orcid.org/0000-0002-1818-2135

© Zhejiang University Press 2023

due to inevitable imperfections (Schenk, 2008; Björnson et al., 2013; Papazafeiropoulos et al., 2017; Kolawole et al., 2020; Wang et al., 2020; Fu et al., 2021). First, the effectiveness of beamforming gain maximization is fundamentally limited by channel state information (CSI) acquisition accuracy. In IRS-assisted systems, CSI estimation is inherently challenging due to the limited number of pilot symbols and the passive nature of the IRS. These limitations lead to non-negligible CSI errors and introduce undesired interference, degrading overall system performance (Wang et al., 2020). Moreover, inevitable hardware impairments in practical transceiver hardware can distort both transmitted and received signals (Papazafeiropoulos et al., 2017). While calibration and compensation techniques can partially mitigate these distortions, a certain amount of distortion and interference still remains, significantly undermining the system's performance. Additionally, IRS phase shifts are inherently discrete, rather than continuous, which makes it crucial to design efficient beamforming strategies that operate over discrete phase shift (DPS) sets (Fu et al., 2021). Overall, a robust algorithm design that explicitly accounts for these practical limitations is essential for exploiting the full potential of IRS-assisted communication systems.

Fortunately, rate-splitting multiple access (RSMA) technology demonstrates strong robustness against inevitable imperfections, enhancing the mutual benefits of combining IRS and RSMA (Aboumahmoud et al., 2025). Unlike space division multiple access (SDMA) and non-orthogonal multiple access (NOMA), where the former treats interference as noise and the latter fully decodes interference (Mao et al.; Clerckx et al., 2023), RSMA is an interference management strategy that leverages rate-splitting (RS) to manage interference effectively. By dividing user information into common and private parts, RSMA can allocate more power to the favorable parts and thus mitigate interference. Then, these parts are flexibly decoded by the user, allowing to retrieve both parts and reconstruct the original message. As a result, RSMA offers superior interference management and robustness capabilities, making it particularly well-suited for IRS-assisted communication systems in practical scenarios.

1.2 Related works

The integration of IRS and RSMA has been shown to significantly enhance their mutual benefits (Yang et al., 2020; Bansal et al., 2021; Weinberger et al., 2022; Pala et al., 2024; Wang et al., 2025). Specifically, Weinberger et al. (2022) investigated a multiuser communication system and demonstrated that RSMA alone can achieve nearly a 95% improvement in energy efficiency (EE). With the deployment of IRS, the overall EE gain exceeds the sum of the individual improvements provided by RSMA and IRS, respectively. Moreover, IRS-assisted RSMA systems consistently outperform IRS-assisted NOMA and SDMA systems across various scenarios. For instance, Yang et al. (2020) analyzed an IRS-assisted multiuser system and demonstrated that the IRS-assisted RSMA system achieves superior performance gains compared to that of the IRS-assisted NOMA system. Moreover, Wang et al. (2025) explored the joint optimization of active beamforming at the BS and passive beamforming at the IRS in a multi-IRS-assisted multiple-input single-output (MISO) system, demonstrating that IRS-assisted RSMA networks are more energy-efficient than their NOMA and SDMA counterparts. Additionally, Bansal et al. (2021) derived the closed-form expression for the outage probability of cell-edge users in IRS-assisted multi-cell networks, theoretically proving that RSMA performs better than NOMA under various system parameters, such as the number of IRS reflecting elements and node density. The above studies on IRS-assisted systems ignored practical imperfections, which inevitably degrade beamforming performance in real-world implementations.

Considering the impact of practical system imperfections on IRS beamforming, extensive research has been conducted on CSI estimation errors, transceiver hardware impairments, and DPSs (Yan et al., 2020; Shen et al., 2021; Khel and Hamdi, 2022; Tao et al., 2022; Wang et al., 2023; Efrem and Krikidis, 2024). To assess the effect of CSI estimation errors, Tao et al. (2022) investigated the ergodic capacity of users under imperfect CSI. By jointly optimizing the beamforming at the BS and phase shifts at the IRS, the weighted sum rate can be maximized, where a maximum ratio transmission (MRT)-based beamforming scheme was proposed. However, the system performance remains limited due to severe multiuser interference. Moreover, Yan et al. (2020) developed a low-complexity channel uncertainty handling scheme based on the sample average approximation (SAA) method. Unlike MRT-based designs, this approach is not limited to fixed beamforming at the BS, allowing full exploitation of IRS capabilities and demonstrating strong robustness. Khel and Hamdi (2022) analyzed the impact of hardware impairments,

Unedited

revealing that transceiver hardware impairments impose a finite limit on the ergodic capacity, which is independent of the number of reflecting elements or BS antennas. Building on this, Shen et al. (2021) designed a robust algorithm in the presence of transceiver hardware impairments to fully leverage the potential of IRS. To jointly mitigate the impact of multiple practical imperfections on IRS-assisted communication systems, Wang et al. (2023) investigated a practical IRS-assisted SDMA system in the presence of transceiver hardware impairments, DPSs, and imperfect CSI. A robust joint beamforming at BS and IRS phase shift design algorithm was proposed, which effectively mitigates the performance degradation caused by these inevitable imperfections.

However, the interference suppression capability of SDMA is limited because it treats all the interference as noise. To address this, Papazafeiropoulos et al. (2017) evaluated the robustness of RSMA systems, demonstrating that RSMA remains a reliable interference suppression strategy in the presence of CSI estimation errors and hardware impairments, leading to improved system performance compared with SDMA. As a result, robust system design for IRS-assisted RSMA has garnered significant attention (Fu et al., 2021; Li et al., 2023, 2024; Asif et al., 2025; Wang et al., 2025). Specifically, Li et al. (2024) investigated the system performance and proposed an SAA-based channel uncertainty mitigation algorithm, fully leveraging the potential of IRS and demonstrating superior performance over SDMA and NOMA. Moreover, Fu et al. (2021) examined the system performance under quantized constraints on phase shifts, where a DPS optimization scheme based on the successive convex approximation (SCA) method was proposed. Meanwhile, the work (Asif et al., 2025) considered an IRS-assisted RSMA system for simultaneous wireless information and power transfer (SWIPT) in the presence of transceiver hardware impairments and developed a semidefinite relaxation (SDR)-based robust beamforming scheme. Additionally, Wang et al. (2025) investigated robust resource allocation design for active IRS-assisted RSMA systems, considering the impact of DPSs.

1.3 Contributions

Although the impact of CSI estimation errors, DPSs, and hardware impairments is non-negligible (Wang et al., 2023), the above studies typically address only pairwise combinations of the three inevitable imperfections. Reasonably, continuing to treat these imperfections that follow the same path may widen the gap between theoretical assumptions and practical deployments. Considering more realistic scenarios and accounting for more practical imperfections, this study investigates the robust design for IRS-assisted multiuser systems by incorporating RSMA technology. The proposed algorithm differs from Li et al. (2024) in several key aspects. While Li et al. (2024) focused on a beyond-diagonal IRS (BD-IRS), our study uses the traditional diagonal IRS, which influences phase shift design and optimization. Our algorithm also specifically addresses DPSs due to hardware bit quantization, which is not considered in the reference work. Additionally, we account for hardware impairments, while the reference paper assumes ideal hardware. This makes our phase shift design more robust, capable of handling both DPS and hardware imperfections. Moreover, unlike previous studies that primarily focused on SDMA, the robust RSMA system design leads to a stochastic optimization problem. Notably, existing optimization frameworks are not applicable. The key contributions are summarized as follows:

1. We provided an IRS-aided multiuser RSMA framework under practical imperfections, including transceiver hardware impairment, DPSs, and imperfect CSI. Through leveraging the relationship between the ergodic and average rate for both common and private streams, a stochastic optimization problem is formulated to enhance system performance.

2. To solve the challenging stochastic problem, we developed an SAA-based robust algorithm by exploiting the relationship between weighted sum rate and weighted minimum mean square error (WMMSE). Additionally, we introduced an alternating direction method of multipliers (ADMM)-based approach to efficiently optimize discrete IRS phase shifts, ensuring computational feasibility.

3. Extensive simulation results are presented to demonstrate the impact of different system imperfections on performance, highlighting the superiority of RSMA in robust system design. Furthermore, the proposed algorithm outperforms benchmark algorithms, demonstrating its effectiveness and providing practical insights for implementing IRS-assisted multiuser RSMA systems.

Notations: Scalars, vectors, and matrices are denoted by lowercase, bold lowercase, and bold uppercase letters, respectively. The space of $x \times y$ complex-valued matrices is denoted by $\mathbb{C}^{x\times y}$.The L_2-norm of a vector or

Unedited

a matrix is represented by $\|\cdot\|$, and absolute value by $|\cdot|$. Transpose, complex conjugate, and Hermitian transpose are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, and $(\cdot)^{\mathrm{H}}$, respectively. $\mathrm{diag}\{\cdot\}$ denotes a diagonal matrix formed from a vector of the diagonal elements of a matrix, $\widetilde{\mathrm{diag}}(\boldsymbol{A})$ denotes the diagonal matrix with the diagonal elements the same as matrix $\boldsymbol{A}$, and $\mathrm{tr}(\cdot)$ represents the trace. Expectation is denoted by $\mathbb{E}\{\cdot\}$, and $\mathrm{Re}(\cdot)$ indicates the real part of a complex number. A circularly symmetric complex Gaussian (CSCG) random variable x with mean μ and variance σ^2 is written as $x \sim \mathcal{CN}\left(\mu, \sigma^2\right)$ (Gallager, 2008). For a vector $\boldsymbol{a}$, the i-th entry is denoted by $[\mathbf{a}]_i$. $\mathrm{Vec}(\cdot)$ represents column-wise vectorization of a matrix.

2 System model and problem formulation

As depicted in Fig. 1, we considered an IRS-aided multiuser RSMA system for downlink communication. An IRS equipped with N_{I} reflecting elements arranged in a uniform planar array (UPA) assists the BS equipped with N_{B} uniform linear array (ULA) antennas in serving K single-antenna users. Accounting for realistic hardware impairments, the received signal at the k-th user can be expressed as

$$y_k = \left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right) \boldsymbol{x} + n_k + \kappa_{R,k}, \quad (1)$$

where $\boldsymbol{h}_{\mathrm{BU},k}$ and $\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}} = \mathrm{diag}\left(\boldsymbol{h}_{\mathrm{IU},k}^{\mathrm{H}}\right) \boldsymbol{H}_{\mathrm{BI}}$ represent the direct and the cascaded link of the k-th user, respectively. $\boldsymbol{h}_{\mathrm{IU,k}}$ and $\boldsymbol{H}_{\mathrm{BI}}$ are the channels from the IRS to the k-th user, and from the BS to the IRS. The transmit signal vector $\mathbf{x}$ satisfies the power constraint $\mathbb{E}\left\{\mathbf{x}^{\mathrm{H}}\mathbf{x}\right\} \leq P$, with P being the maximum transmit power. The term $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ denotes additive white Gaussian noise (AWGN), following a CSCG distribution with zero mean and variance σ_k^2. The receiver distortion noise $\kappa_{R,k}$ capturing residual hardware impairments after calibration, is modeled as $\kappa_{R,k} \sim \mathcal{CN}\left(\mathbf{0}, \beta_{R,k}^2 \mathbb{E}\left\{\tilde{y}_k \tilde{y}_k^H\right\}\right)$, where $\tilde{y}_k = \left(\boldsymbol{h}_{\mathrm{BU,k}}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right) \boldsymbol{x} + n_k$ and $\beta_{R,k} \in [0,1]$ is the normalized distortion level (Schenk, 2008; Björnson et al., 2013; Kolawole et al., 2020). Moreover, $\boldsymbol{\xi} = \left[e^{\mathrm{j}\theta_1}, \cdots, e^{\mathrm{j}\theta_{N_{\mathrm{I}}}}\right]^{\mathrm{T}}$ is the phase shift vector of the IRS. For practical IRS implementations, each phase shift $\theta_m, \forall m = [1, \cdots, N_{\mathrm{I}}]$ is constrained to discrete values from the set $\mathbb{T} \triangleq \left\{-\pi, \frac{2\pi}{2^{\mathrm{b}}} - \pi, \cdots, \frac{2\pi\left(2^{\mathrm{b}}-1\right)}{2^{\mathrm{b}}} - \pi\right\}$, where b denotes the number of quantization bits.

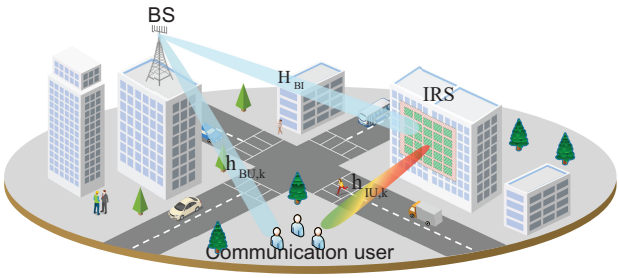


Fig. 1 An IRS-aided multiuser RSMA communication system. BS, base station; IRS, intelligent reflecting surface; RSMA, rate-splitting multiple access

2.1 Modeling of imperfect CSI

Acquiring perfect CSI in IRS-assisted systems is challenging due to the passive nature of IRS elements, estimation errors, and feedback delays. These imperfections lead to mismatches between actual and estimated CSI. To address this, we modeled the CSI errors as statistical, typically following the CSCG distribution (Zhang et al., 2011). This model accounts for the inherent uncertainty and provides an analytically tractable framework for studying performance degradation due to imperfect CSI (Yan et al., 2020; Tao et al., 2022; Wang et al., 2023).The actual channel can be modeled as

$$\boldsymbol{H}_{\mathrm{eff},k} = \tilde{\boldsymbol{H}}_{\mathrm{eff},k} + \Delta\boldsymbol{H}_{\mathrm{eff},k}, \boldsymbol{h}_{\mathrm{BU},k} = \tilde{\boldsymbol{h}}_{\mathrm{BU},k} + \Delta\boldsymbol{h}_{\mathrm{BU},k}, \quad (2)$$

where $\tilde{\boldsymbol{H}}_{\mathrm{eff},k}$ and $\tilde{\boldsymbol{h}}_{\mathrm{BU},k}$ denote the estimated cascaded and direct channels for the k-th user, respectively. The corresponding CSI errors, $\Delta\boldsymbol{H}_{\mathrm{eff},k}$ and $\Delta\boldsymbol{h}_{\mathrm{BU},k}$ are modeled as CSCG random variables, i.e., $\mathrm{vec}\left(\Delta\boldsymbol{H}_{\mathrm{eff},k}\right) \sim \mathcal{CN}\left(0, \delta_k^2 \boldsymbol{I}_{N_I N_B}\right)$ and $\mathrm{vec}\left(\Delta\boldsymbol{h}_{\mathrm{BU},k}\right) \sim \mathcal{CN}\left(0, \delta_k^2 \mathbf{I}_{N_B}\right)$, where δ_k^2 quantifies the estimation inaccuracy for the k-th user.

2.2 RS with hardware distortions

To improve the performance of the system, this study employs the commonly used 1-layer RSMA architecture (Mao et al.), where the signal intended for the k-th user is divided into a common part and a private part. Therefore, the signal transmitted by the BS is expressed as

$$\boldsymbol{x} = \boldsymbol{w}_c s_c + \sum_{k=1}^{K} \boldsymbol{w}_{p,k} s_{p,k} + \boldsymbol{\kappa}_T = \boldsymbol{W}\boldsymbol{s} + \boldsymbol{\kappa}_T, \quad (3)$$

where the superimposed stream vector $\boldsymbol{s} = \left[s_c, s_{p,1}, \ldots, s_{p,K}\right]^{\mathrm{T}}$ consists of a common stream s_c and K private streams $s_{p,k}$ for the K users, which are linearly precoded by $\boldsymbol{W} = [\boldsymbol{w}_c, \boldsymbol{w}_{p,1}, \ldots, \boldsymbol{w}_{p,K}]$,

Unedited

where $\boldsymbol{w}_c$ and $\boldsymbol{w}_{p,k}$ are the common and private precoders, respectively. Similar to the receiver distortion noise $\kappa_{R,k}$, each element of the transmitter distortion noise $\boldsymbol{\kappa}_T$ is assumed to follow a CSCG distribution with variance proportional to the transmit power of each antenna,[1] i.e., $\boldsymbol{\kappa}_{\mathrm{T}} \sim \mathcal{CN}\left(\mathbf{0}, \beta_T^2 \widetilde{\mathrm{diag}}\left(\boldsymbol{W}\mathbf{W}^{\mathrm{H}}\right)\right)$, where β_{T} is the normalized distortion level for the transmitter. Consequently, the transmit power constraint becomes $\mathbb{E}\{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\} = \left(1+\beta_{\mathrm{T}}^2\right)\mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\right) \leq P$.

By substituting (3) into (1), the received signal at the k-th user can be rewritten as

$$\begin{aligned} \boldsymbol{y}_k = &\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right) \\ &\times \left(\boldsymbol{w}_{\mathrm{c}}\boldsymbol{s}_{\mathrm{c}} + \sum_{\mathrm{k=1}}^{\mathrm{K}} \boldsymbol{w}_{\mathrm{p},k}\boldsymbol{s}_{\mathrm{p},k} + \boldsymbol{\kappa}_{\mathrm{T}}\right) \\ &+ \boldsymbol{n}_k + \boldsymbol{\kappa}_{\mathrm{R},k}. \end{aligned} \tag{4}$$

Therefore, the average received power of the k-th user is computed as follows:

$$d_{c,k} = \mathbb{E}\left\{|y_k|^2\right\} = \left|\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}}\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_c\right|^2 + d_{p,k}, \tag{5}$$

where $d_{p,k}$ denotes the interference power for the common stream represented by

$$d_{p,k} = \left|\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_p,k\right|^2 + I_{p,k}, \tag{6}$$

where $I_{p,k}$ denotes the interference power for the private stream represented by

$$\begin{aligned} I_{p,k} = &\sum_{k'\neq k}^{K} \left|\left(\boldsymbol{h}_{\mathrm{BU,k}}^{\mathrm{H}} + \boldsymbol{\xi}^{T}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_{p,k'}\right|^2 \\ &+ \left(1+\beta_{R,k}^2\right)\sigma_k^2 \\ &+ \left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right) \\ &\times \left[\beta_{R,k}^2\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}} + \beta_{\mathrm{T}}^2\left(1+\beta_{R,k}^2\right)\right. \\ &\qquad \left.\times\widetilde{\mathrm{diag}}\left(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\right)\right] \\ &\times \left(\boldsymbol{h}_{\mathrm{BU},k} + \boldsymbol{H}_{\mathrm{eff},k}\boldsymbol{\xi}^*\right). \end{aligned} \tag{7}$$

Under the 1-layer RSMA architecture, each user first decodes the common stream by treating the K private streams, hardware distortion noise, and AWGN as noise. Specifically, an equalizer, i.e., $g_{c,k}$, is applied to recover the common stream for the k-th user as $\hat{s}_{c,k} = g_{c,k}y_k$. Accordingly, the instantaneous rate for decoding the common stream is

$$R_{c,k} = \log_2\left(1+\gamma_{c,k}\right), \tag{8}$$

where the corresponding signal-to-interference-plus-noise ratio (SINR) is given by

$$\gamma_{c,k} = \left|\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{\mathrm{T}}\mathbf{H}_{\mathrm{eff,k}}^{\mathrm{H}}\right)\boldsymbol{w}_c\right|^2 I_{c,k}^{-1}. \tag{9}$$

To ensure successful decoding of the common stream by all users, the common rate is limited by the worst-case user, i.e., $R_{\mathrm{c}} = \min_k\left(R_{\mathrm{c},k}\right)$. Therefore, the common rate is partitioned among the users such that $\min_k\left(R_{\mathrm{c},k}\right) = \sum_k r_{c,k}$, where $r_{c,k}$ denotes the common rate allocated to the k-th user. After decoding and subtracting the common stream via the successive interference cancellation (SIC) method, the k-th user applies an equalizer $g_{p,k}$ to decode its private stream, i.e., $\hat{s}_{p,k} = g_{p,k}\left[y_k - \left(\boldsymbol{h}_{\mathrm{BU},k}^{H} + \boldsymbol{\xi}^{T}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_c s_c\right]$, by treating the remaining $K-1$ private stream interference and hardware distortions, encapsulated in $I_{p,k}$, as noise. Thus, the corresponding instantaneous private rate is

$$R_{p,k} = \log_2\left(1+\gamma_{p,k}\right), \tag{10}$$

where the SINR $\gamma_{p,k}$ is given by

$$\gamma_{p,k} = \left|\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{H}} + \boldsymbol{\xi}^{T}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_{p,k}\right|^2 I_{p,k}^{-1}. \tag{11}$$

Having decoded both the common and private streams, i.e., $\hat{s}_c$ and $\hat{s}_{p,k}$, the original message for the k-th user can be reconstructed. Based on the above architecture, the total achievable rate for the k-th user is the sum of its private and allocated common rates, which can be calculated as

$$R_k = R_{p,k} + r_{c,k}. \tag{12}$$

It is found from the total achievable rate expression that the achievable rate is determined by the transmit beamforming matrix, i.e., $\boldsymbol{W}$, and the phase shift of the IRS, i.e., $\boldsymbol{\xi}$. In order to improve the overall performance of the IRS-assisted RSMA system, it makes sense to optimize these key parameters.

2.3 Problem formulation

In this study, we adopted the average rate to characterize the ergodic rate over the distribution of CSI errors without predefining the BS beamforming strategy. Following previous relevant work (Li

[1] Both theoretical and empirical studies have shown that distortion noise resulting from residual hardware impairments can be effectively modeled as an additional Gaussian noise component, where its variance is directly proportional to the transmitted signal power (Björnson et al., 2014).

Unedited

$$\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}}\left\{R_{c,k}\right\}=\mathbb{E}_{\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k}|\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}|\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{R_{c,k}\mid\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}\right\}\right\}=\mathbb{E}_{\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{\bar{R}_{c,k}\right\}, \tag{13}$$

$$\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}}\left\{R_{p,k}\right\}=\mathbb{E}_{\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k}|\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}|\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{R_{p,k}\mid\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}\right\}\right\}=\mathbb{E}_{\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\left\{\bar{R}_{p,k}\right\}, \tag{14}$$

et al., 2024), for the k-th user, the relationship between the average common rate, denoted as $\bar{R}_{c,k}=\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k}|\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}|\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\{R_{c,k}\}$, and ergodic common rates, denoted as $\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}}\{R_{c,k}\}$, as well as the relationship between the average private rate, denoted as $\bar{R}_{p,k}=\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k}|\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}|\tilde{\boldsymbol{h}}_{\mathrm{BU},k}}\{R_{p,k}\}$, and the ergodic private rate, denoted as $\mathbb{E}_{\boldsymbol{H}_{\mathrm{eff},k},\boldsymbol{h}_{\mathrm{BU},k}}\{R_{p,k}\}$, can be expressed as Eqs (13) and (14), indicating that the ergodic rate is determined by the expected value of the average rates over the estimated CSI. Therefore, the ergodic sum rate maximization problem with imperfect CSI is transformed into maximizing the average sum rate for each estimated CSI, leading to the following optimization problem P1:

$$\max_{\boldsymbol{W},\boldsymbol{\xi},\bar{r}_{c,k}}\quad\sum_{k=1}^{K}\eta_k\left(\bar{r}_{c,k}+\bar{R}_{p,k}\right), \tag{15a}$$

$$\text{s.t.}\quad\left(1+\beta_T^2\right)\mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\right)\leq P, \tag{15b}$$

$$\sum_{k=1}^{K}\bar{r}_{c,k}\leq\bar{R}_{c,k},\forall k\in[1,K], \tag{15c}$$

$$\bar{r}_{c,k}+\bar{R}_{p,k}\geq R_k^{\mathrm{th}},\forall k\in[1,K], \tag{15d}$$

$$\theta_m\in\mathbb{T},\forall m\in[1,N_{\mathrm{I}}], \tag{15e}$$

where η_k, $\bar{r}_{c,k}$, and R_k^{th} are the weight factor, the average common rate, and the rate threshold for the k-th user, respectively. The above stochastic optimization problem is challenging due to the non-convexity of the objective function, the common rate constraint (15c), the DPS constraint (15d), and its stochastic nature. To address this, we proposed an SAA-based algorithm to make the problem more tractable.

3 SAA-based robust algorithm design

Maximizing the stochastic optimization problem is a challenge in the presence of hardware impairments and imperfect CSI, as beamforming design typically relies on perfect CSI. To overcome this, we proposed an SAA-based algorithm for jointly optimizing the IRS phase shifts and BS beamforming matrix by leveraging the statistical properties of the CSI.

3.1 Algorithm design

We first applied the SAA method to transform the stochastic optimization problem into a deterministic one. According to Li et al. (2024), the SAA of the average rate provides a tight approximation of the average rate as the number of samples increases, i.e., $\lim_{A\to\infty}\hat{R}_{c,k}=\bar{R}_{c,k}$ and $\lim_{A\to\infty}\hat{R}_{p,k}=\bar{R}_{p,k}$, where the SAA of the average common and private rate can be defined as

$$\hat{R}_{c,k}=\frac{1}{A}\sum_{a=1}^{A}R_{\mathrm{c},k}^{a},\ \hat{R}_{p,k}=\frac{1}{A}\sum_{a=1}^{A}R_{\mathrm{p},k}^{a}, \tag{16}$$

where $R_{\mathrm{c},k}^{a}$ and $R_{\mathrm{p},k}^{a}$ denote the common and private rates in the a-th channel realization, respectively. A denotes the number of i.i.d. channel realizations, which is represented as

$$\begin{aligned}\mathbb{H}_k=\Big\{&\boldsymbol{H}_{\mathrm{eff},k}^{a}=\tilde{\boldsymbol{H}}_{\mathrm{eff},k}+\Delta\boldsymbol{H}_{\mathrm{eff},k}^{a}|\tilde{\boldsymbol{H}}_{\mathrm{eff},k},\\&\boldsymbol{h}_{\mathrm{BU},k}^{a}=\tilde{\boldsymbol{h}}_{\mathrm{BU},k}+\Delta\boldsymbol{h}_{\mathrm{BU},k}^{a}|\tilde{\boldsymbol{h}}_{\mathrm{BU},k},\\&\forall a\in\mathcal{A}=\{1,\ldots,A\}\Big\},\end{aligned} \tag{17}$$

where $\Delta\boldsymbol{H}_{\mathrm{eff},k}^{a}$ and $\Delta\boldsymbol{h}_{\mathrm{BU},k}^{a}$ are the CSI errors in the a-th channel realizations that follow the CSCG distribution.

Therefore, the stochastic optimization problem (P1) is transformed into a deterministic problem P2:

$$\max_{\boldsymbol{W},\boldsymbol{\xi},\bar{r}_{c,k}}\quad\sum_{k=1}^{K}\eta_k\left(\bar{r}_{c,k}+\frac{1}{A}\sum_{a=1}^{A}\hat{R}_{p,k}^{a}\right), \tag{18a}$$

$$\text{s.t. Eqs.}(15b),\text{and}(15e), \tag{18b}$$

$$\sum_{k=1}^{K}\bar{r}_{c,k}\leq\frac{1}{A}\sum_{a=1}^{A}\hat{R}_{c,k}^{a},\qquad\forall k\in[1,K], \tag{18c}$$

$$\bar{r}_{c,k}+\frac{1}{A}\sum_{a=1}^{A}\hat{R}_{p,k}^{a}\geq R_k^{\mathrm{th}},\qquad\forall k\in[1,K]. \tag{18d}$$

Despite this transformation, the above optimization problem remains non-convex. To address this, we ap-

Unedited

plied the WMMSE method to further simplify the problem. The optimization can be iteratively solved using the following steps:

Recall the 1-layer RS architecture discussed in Section 2.2, for the a-th channel realization, the estimate of the common stream $s_{c,k}$ with equalizer $g_{c,k}^a$ is $\hat{s}_{c,k}^a = g_{c,k}^a y_k^a$. Similarly, the estimate of the private stream $s_{p,k}$ with equalizer $g_{p,k}^a$, after applying SIC, is $\hat{s}_{p,k}^a = g_{p,k}^a \left\{ y_k^a - \left[\left(\boldsymbol{h}_{\mathrm{BU},k}^a \right)^{\mathrm{H}} + \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff},k}^a \right)^{\mathrm{H}} \right] \right\} \boldsymbol{w}_c s_c$. Therefore, the mean square error (MSE) for the common and private stream at the k-th user in the a-th channel realization is given by

$$\begin{aligned}\varepsilon_{c,k}^a &= \left\{ \left| g_{c,k}^a y_k^a - s_c \right|^2 \right\} = \left| g_{c,k}^a \right|^2 d_{c,k}^a \\ &- 2\mathrm{Re}\left\{ g_{c,k}^a \left[\left(\boldsymbol{h}_{\mathrm{BU},k}^a \right)^{\mathrm{H}} + \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff},k}^a \right)^{\mathrm{H}} \right] \boldsymbol{w}_c \right\} + 1,\end{aligned} \tag{19}$$

$$\begin{aligned}\varepsilon_{p,k}^a &= \left\{ \left| g_{p,k}^a y_k^a - s_{p,k} \right|^2 \right\} = \left| g_{p,k}^a \right|^2 d_{p,k}^a \\ &- 2\mathrm{Re}\left\{ g_{p,k}^a \left[\left(\mathbf{h}_{\mathrm{BU,k}}^a \right)^{\mathrm{H}} + \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff,k}}^a \right)^{\mathrm{H}} \right] \boldsymbol{w}_{p,k} \right\} + 1.\end{aligned} \tag{20}$$

The optimal equalizers that minimize the MSE are derived using the first-order optimality condition

$$g_{c,k}^a = \frac{\boldsymbol{w}_c^{\mathrm{H}} \left(\boldsymbol{h}_{\mathrm{BU},k}^a + \boldsymbol{H}_{\mathrm{eff},k}^a \boldsymbol{\xi}^* \right)}{d_{c,k}^a}, \tag{21}$$

$$g_{p,k}^a = \frac{\boldsymbol{w}_{p,k}^{\mathrm{H}} \left(\boldsymbol{h}_{\mathrm{BU},k}^a + \boldsymbol{H}_{\mathrm{eff},k}^a \boldsymbol{\xi}^* \right)}{d_{p,k}^a}, \tag{22}$$

and the corresponding minimum MSE values are obtained as

$$\varepsilon_{c,k}^a = \left(1 + \gamma_{c,k}^a \right)^{-1}, \varepsilon_{p,k}^a = \left(1 + \gamma_{p,k}^a \right)^{-1}. \tag{23}$$

Based on the above relationship between SINR and MSE, the WMMSE-rate relationship is given by

$$\hat{R}_{c,k}^a = \log_2 \left(1 + \gamma_{c,k}^a \right) = \log_2 \left(\lambda_{c,k}^a \right) - \lambda_{c,k}^a \varepsilon_{c,k}^a + 1, \tag{24}$$

$$\hat{R}_{p,k}^a = \log_2 \left(1 + \gamma_{p,k}^a \right) = \log_2 \left(\lambda_{p,k}^a \right) - \lambda_{p,k}^a \varepsilon_{p,k}^a + 1, \tag{25}$$

where $\lambda_{c,k}^a$ and $\lambda_{p,k}^a$ are the weights for the common and private rates of the k-th user in the a-th channel realization, respectively, defined as

$$\lambda_{c,k}^a = 1 + \gamma_{c,k}^a, \lambda_{p,k}^a = 1 + \gamma_{p,k}^a. \tag{26}$$

Thus, the optimization problem (P2) can be rewritten as P3:

$$\begin{aligned}\max_{\boldsymbol{W},\boldsymbol{\xi},\bar{r}_{c,k}} \quad & \sum_{k=1}^{K} \eta_k \left(\bar{r}_{c,k} + \frac{1}{A} \sum_{a=1}^{A} \left[\log_2 \left(\lambda_{p,k}^a \right) \right. \right. \\ & \left. \left. - \lambda_{p,k}^a \varepsilon_{p,k}^a + 1 \right] \right), \end{aligned} \tag{27a}$$

$$\text{s.t. Eqs.}(15b)\text{and}(15e),$$

$$\begin{aligned}& \frac{1}{A} \sum_{a=1}^{A} \left[\log_2 \left(\lambda_{c,k}^a \right) - \lambda_{c,k}^a \varepsilon_{c,k}^a + 1 \right] \\ & \geq \sum_{k=1}^{K} \bar{r}_{c,k}, \forall k \in [1, K],\end{aligned} \tag{27b}$$

$$\begin{aligned}& \bar{r}_{c,k} + \frac{1}{A} \sum_{a=1}^{A} \left[\log_2 \left(\lambda_{p,k}^a \right) - \lambda_{p,k}^a \varepsilon_{p,k}^a + 1 \right] \\ & \geq R_k^{\mathrm{th}}, \forall k \in [1, K].\end{aligned} \tag{27c}$$

Although (P3) is still non-convex due to the coupling of optimization variables $\boldsymbol{W}, \bar{r}_{c,k}, \boldsymbol{\xi}$, the sub-problem corresponding to one block, while fixing the others, is more tractable. Specifically, we divided the optimization variables into two blocks, i.e, $\{\boldsymbol{W}, \bar{r}_{c,k}\}$, and $\boldsymbol{\xi}$ to solve it efficiently.

3.1.1 Beamforming optimization at BS

With a fixed IRS phase shift matrix, the sub-problem for optimizing the block $\{\mathbf{W}, \bar{r}_{c,k}\}$ is formulated as problem P4:

$$\max_{\boldsymbol{W},\bar{r}_{c,k}} \quad \sum_{k=1}^{K} \eta_k \left(\bar{r}_{c,k} - \frac{1}{A} \sum_{a=1}^{A} \lambda_{p,k}^a \varepsilon_{p,k}^a \right), \tag{28}$$

$$\text{s.t.Eqs.}(15b), (27b), \text{and } (27c).$$

By invoking the equality $\boldsymbol{a}^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}} \right) \boldsymbol{a} = \boldsymbol{b}^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}} \right) \boldsymbol{b}$ and $\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}} = \boldsymbol{w}_c \boldsymbol{w}_c^{\mathrm{H}} + \sum_{k=1}^{K} \boldsymbol{w}_{p,k} \boldsymbol{w}_{p,k}^{\mathrm{H}}$, the following equations hold:

$$\begin{aligned}& \boldsymbol{\xi}^{\mathrm{T}} \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \right)^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}} \right) \boldsymbol{H}_{\mathrm{eff},k} \boldsymbol{\xi}^* \\ & = \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \right)^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{w}_c \boldsymbol{w}_c^{\mathrm{H}} + \sum_{k=1}^{K} \boldsymbol{w}_{p,k} \boldsymbol{w}_{p,k}^{\mathrm{H}} \right) \boldsymbol{H}_{\mathrm{eff},k} \boldsymbol{\xi}^* \\ & = \boldsymbol{w}_c^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \boldsymbol{\xi}^* \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \right)^{\mathrm{H}} \right) \boldsymbol{w}_c \\ & + \sum_{k=1}^{K} \boldsymbol{w}_{p,k}^{\mathrm{H}} \widetilde{\mathrm{diag}} \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \boldsymbol{\xi}^* \boldsymbol{\xi}^T \left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \right)^{\mathrm{H}} \right) \boldsymbol{w}_{p,k},\end{aligned} \tag{29}$$

Unedited

$$\begin{aligned}&\boldsymbol{\xi}^{\mathrm{T}}\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff},k}\boldsymbol{\xi}\\&=\boldsymbol{\xi}^{\mathrm{T}}\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\left(\boldsymbol{w}_c\boldsymbol{w}_c^{\mathrm{H}}+\sum_{k=1}^{K}\boldsymbol{w}_{p,k}\boldsymbol{w}_{p,k}^{\mathrm{H}}\right)\boldsymbol{H}_{\mathrm{eff,k}}\boldsymbol{\xi}\\&=\boldsymbol{w}_c^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\boldsymbol{\xi}\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\boldsymbol{w}_c\\&+\sum_{k=1}^{K}\boldsymbol{w}_{p,k}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\boldsymbol{\xi}\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}^{\mathrm{H}}.\end{aligned}\tag{30}$$

Therefore, by substituting $d_{p,k}$ into Eq. (28), the objective function of the optimization problem (P4) can be calculated as

$$\begin{aligned}&\sum_{k=1}^{K}\eta_k\left(\bar{r}_{c,k}-\tfrac{1}{A}\sum_{a=1}^{A}\lambda_{p,k}^{\mathrm{a}}\varepsilon_{p,k}^{\mathrm{a}}\right)\\&=\sum_{k=1}^{K}\eta_k\left(\bar{r}_{c,k}-\tfrac{1}{A}\sum_{a=1}^{A}\lambda_{p,k}^{\mathrm{a}}\left(\left|g_{p,k}^{\mathrm{a}}\right|^2d_{p,k}^{\mathrm{a}}\right.\right.\\&\left.\left.-2\mathrm{Re}\left\{g_{p,k}^{\mathrm{a}}\left(\boldsymbol{h}_{\mathrm{BU,k}}^{\mathrm{H}}+\boldsymbol{\xi}^T\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{H}}\right)\boldsymbol{w}_{p,k}\right\}+1\right)\right)\\&=\sum_{k=1}^{K}\eta_k\bar{r}_{c,k}-\sum_{k=1}^{K}\eta_k\tfrac{1}{A}\sum_{a=1}^{A}\times\\&\left(\begin{array}{l}\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\left[\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\boldsymbol{\xi}^*\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\right]\boldsymbol{w}_{p,j}\\+\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\left[\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\beta_{R,k}^2\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\boldsymbol{\xi}^*\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\right]\boldsymbol{w}_{p,j}\\+\boldsymbol{w}_c^{\mathrm{H}}\left[\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\beta_{R,k}^2\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\boldsymbol{\xi}^*\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\right]\boldsymbol{w}_c\\+\boldsymbol{w}_c^{\mathrm{H}}\left[\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\beta_T^2\left(1+\beta_{R,k}^2\right)\right.\\\left.\times\widetilde{\mathrm{diag}}\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\boldsymbol{\xi}^*\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\right)\right]\boldsymbol{w}_c\\+\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\left[\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\beta_T^2\left(1+\beta_{R,k}^2\right)\right.\\\left.\times\widetilde{\mathrm{diag}}\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\boldsymbol{\xi}^*\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\right)\right]\boldsymbol{w}_{p,j}\\+\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\left(1+\beta_{R,k}^2\right)\sigma^2\\-2\lambda_{p,k}^{\mathrm{a}}\mathrm{Re}\left\{g_{p,k}^{\mathrm{a}}\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff,k}}^{\mathrm{a}}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}\right\}+\lambda_{p,k}^{\mathrm{a}}\end{array}\right)\end{aligned}\tag{31}$$

Define the auxiliary variable $\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}$, $\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}$, $\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}$, $\boldsymbol{\Omega}_{p,k}^{\mathrm{w},a}$, and $\boldsymbol{\Lambda}_{p,k}^{\mathrm{w},a}$ as

$$\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}=\tfrac{1}{A}\sum_{a=1}^{A}\boldsymbol{\Omega}_{p,k}^{\mathrm{w},a},\tag{32}$$

$$\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}=\frac{1}{A}\sum_{a=1}^{A}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w},a},\tag{33}$$

$$\begin{aligned}\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}=&\frac{1}{A}\sum_{a=1}^{A}\lambda_{p,k}^{\mathrm{a}}g_{p,k}^{\mathrm{a}}\times\\&\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}}\right)^{\mathrm{H}}+\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\right],\end{aligned}\tag{34}$$

$$\begin{aligned}\boldsymbol{\Omega}_{p,k}^{\mathrm{w},a}=&\left(1+\beta_{R,k}^2\right)\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}+\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\boldsymbol{\xi}^*\right)\\&\times\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}}\right)^{\mathrm{H}}+\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\right]\\&+\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\beta_T^2\,\widetilde{\mathrm{diag}}\left\{\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}+\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\boldsymbol{\xi}^*\right)\right.\\&\left.\times\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}}\right)^{\mathrm{H}}+\boldsymbol{\xi}^T\left(\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\right)^{\mathrm{H}}\right]\right\}\left(1+\beta_{R,k}^2\right).\end{aligned}\tag{35}$$

$$\begin{aligned}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w},a}=&\lambda_{p,k}^{a}\left|g_{p,k}^{a}\right|^2\beta_{R,k}^2\left(\mathbf{h}_{\mathrm{BU},k}^{a}+\mathbf{H}_{\mathrm{eff},k}^{a}\boldsymbol{\xi}^*\right)\\&\times\left[\left(\mathbf{h}_{\mathrm{BU},k}^{a}\right)^{\mathrm{H}}+\boldsymbol{\xi}^T\left(\mathbf{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\right]\\&+\lambda_{p,k}^{a}\left|g_{p,k}^{a}\right|^2\beta_T^2\,\widetilde{\mathrm{diag}}\left\{\left(\mathbf{h}_{\mathrm{BU},k}^{a}+\mathbf{H}_{\mathrm{eff},k}^{a}\boldsymbol{\xi}^*\right)\right.\\&\left.\times\left[\left(\mathbf{h}_{\mathrm{BU},k}^{a}\right)^{\mathrm{H}}+\boldsymbol{\xi}^T\left(\mathbf{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\right]\right\}\left(1+\beta_{R,k}^2\right).\end{aligned}\tag{36}$$

the objective function of (P4) can be rewritten into

$$\begin{aligned}&\sum_{k=1}^{K}\eta_k\left(\bar{r}_{c,k}-\tfrac{1}{A}\sum_{a=1}^{A}\lambda_{p,k}^{\mathrm{a}}\varepsilon_{p,k}^{\mathrm{a}}\right)\\&=\sum_{k=1}^{K}\eta_k\bar{r}_{c,k}-\sum_{k=1}^{K}\eta_k\boldsymbol{w}_c^{\mathrm{H}}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}\boldsymbol{w}_c\\&-\sum_{k=1}^{K}\eta_k\left(\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{p,j}\right)\\&+2\sum_{k=1}^{K}\eta_k\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}\right\}\\&-\tfrac{1}{A}\sum_{a=1}^{A}\left[\lambda_{p,k}^{\mathrm{a}}+\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\left(1+\beta_{R,k}^2\right)\sigma_k^2\right].\end{aligned}\tag{37}$$

Recall the definition of average private rate in Eq. (25), the left equation of constraint in Eq. (27c) of (P4) can be rewritten into

$$\begin{aligned}&-\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{p,j}-\boldsymbol{w}_c^{\mathrm{H}}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}\boldsymbol{w}_c\\&+2\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_c\right\}\\&+\bar{r}_{c,k}+\beta_{p,k}^{\mathrm{w}},\end{aligned}\tag{38}$$

where

$$\begin{aligned}\beta_{p,k}^{\mathrm{w}}=&\tfrac{1}{A}\sum_{a=1}^{A}\left[\log_2\left(\lambda_{p,k}^{\mathrm{a}}\right)\right.\\&\left.-\lambda_{p,k}^{\mathrm{a}}-\lambda_{p,k}^{\mathrm{a}}\left|g_{p,k}^{\mathrm{a}}\right|^2\left(1+\beta_{R,k}^2\right)\sigma_k^2+1\right],\end{aligned}\tag{39}$$

Similarly, recall the definition of the average common rate in Eq. (24) and define the auxiliary variable $\boldsymbol{\Omega}_{c,k}^{\mathrm{w}}$, $\beta_{c,k}^{\mathrm{w}}$, $\left(\boldsymbol{\psi}_{c,k}^{\mathrm{w}}\right)^{\mathrm{H}}$, and $\boldsymbol{\Omega}_{c,k}^{\mathrm{w},a}$ as $\boldsymbol{\Omega}_{c,k}^{\mathrm{w}}=\tfrac{1}{A}\sum_{a=1}^{A}\boldsymbol{\Omega}_{c,k}^{\mathrm{w},a}$,

$$\begin{aligned}\beta_{c,k}^{\mathrm{w}}=&\tfrac{1}{A}\sum_{a=1}^{A}\left[\log_2\left(\lambda_{c,k}^{\mathrm{a}}\right)+1\right.\\&\left.-\lambda_{c,k}^{\mathrm{a}}+\lambda_{c,k}^{\mathrm{a}}\left|g_{c,k}^{\mathrm{a}}\right|^2\left(1+\beta_{R,k}^2\right)\sigma_k^2\right],\end{aligned}\tag{40}$$

Unedited

$$\left(\boldsymbol{\psi}_{c,k}^{\mathrm{w}}\right)^{\mathrm{H}}=\frac{1}{A}\sum_{a=1}^{A}\lambda_{c,k}^{a}g_{c,k}^{a}\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}\right)^{\mathrm{H}}+\boldsymbol{\xi}^{T}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\right]. \tag{41}$$

$$\begin{aligned}\boldsymbol{\Omega}_{c,k}^{\mathrm{w},a}=&\left(1+\beta_{R,k}^{2}\right)\lambda_{c,k}^{a}\left|g_{c,k}^{a}\right|^{2}\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}+\boldsymbol{H}_{\mathrm{eff},k}^{a}\boldsymbol{\xi}^{*}\right)\\&\times\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}\right)^{\mathrm{H}}+\boldsymbol{\xi}^{T}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\right]\\&+\lambda_{c,k}^{a}\left|g_{c,k}^{a}\right|^{2}\beta_{T}^{2}\,\widetilde{\mathrm{diag}}\{\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}+\boldsymbol{H}_{\mathrm{eff},k}^{a}\boldsymbol{\xi}^{*}\right)\\&\qquad\times\left[\left(\boldsymbol{h}_{\mathrm{BU},k}^{a}\right)^{\mathrm{H}}+\boldsymbol{\xi}^{T}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\right]\right\}\left(1+\beta_{R,k}^{2}\right).\end{aligned} \tag{42}$$

The left equation of constraint in Eq. (27c) of (P4) can be rewritten into

$$\begin{aligned}&\frac{1}{A}\sum_{a=1}^{A}\left[\log_{2}\left(\lambda_{c,k}^{\mathrm{a}}\right)-\lambda_{c,k}^{\mathrm{a}}\varepsilon_{c,k}^{\mathrm{a}}+1\right]=\\&-\boldsymbol{w}_{c}^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{w}\boldsymbol{w}_{c}-\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{w}\boldsymbol{w}_{p,j}\\&+2\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{c,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}\right\}+\beta_{c,k}^{\mathrm{w}},\end{aligned} \tag{43}$$

Therefore, the sub-problem P4 can be simplified into P4-1 after several mathematical manipulations:

$$\begin{aligned}\max_{\boldsymbol{W},\bar{r}_{c,k}}&\sum_{k=1}^{K}\eta_{k}\bar{r}_{c,k}-\sum_{k=1}^{K}\eta_{k}\boldsymbol{w}_{c}^{\mathrm{H}}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{c}\\&-\sum_{k=1}^{K}\eta_{k}\left(\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{p,j}\right)\\&+2\sum_{k=1}^{K}\eta_{k}\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_{p.k}\right\},\end{aligned} \tag{44a}$$

s.t.

$$\begin{aligned}&\boldsymbol{w}_{c}^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{w}\boldsymbol{w}_{c}+\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{w}\boldsymbol{w}_{p,j}\\&-2\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{c,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}\right\}\\&\leq\beta_{c,k}^{\mathrm{w}}-\sum_{k=1}^{K}\bar{r}_{c,k},\forall k\in[1,K],\end{aligned} \tag{44b}$$

$$\begin{aligned}&\sum_{j=1}^{K}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{p,j}+\boldsymbol{w}_{c}^{\mathrm{H}}\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}\boldsymbol{w}_{c}\\&-2\mathrm{Re}\left\{\left(\boldsymbol{\psi}_{p,k}^{\mathrm{w}}\right)^{\mathrm{H}}\boldsymbol{w}_{c}\right\}-\bar{r}_{c,k}\\&\leq\beta_{p,k}^{\mathrm{w}}-R_{k}^{th},\forall k\in[1,K].\end{aligned} \tag{44c}$$

Since $\boldsymbol{\Omega}_{p,k}^{\mathrm{w}}$, $\boldsymbol{\Lambda}_{p,k}^{\mathrm{w}}$, $\boldsymbol{\Omega}_{c,k}^{\mathrm{w}}$, and $\boldsymbol{\Lambda}_{c,k}^{\mathrm{w}}$ are positive definite, problem (P4-1) is convex and can be efficiently solved using standard interior-point methods.

3.1.2 IRS phase shifts optimization

Similar to the beamforming optimization at BS, with a fixed beamforming at BS, the sub-problem for optimizing the IRS phase shift can be formulated as P5:

$$\max_{\boldsymbol{\xi}}-\sum_{k=1}^{K}\eta_{k}\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\boldsymbol{\xi}+2\sum_{k=1}^{K}\eta_{k}\mathrm{Re}\left\{\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\}, \tag{45a}$$

s.t.Eqs.$(15e)$,

$$\begin{aligned}&\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{\xi}\boldsymbol{\xi}-2\mathrm{Re}\left\{\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\psi}_{c,k}^{\xi}\right\}\\&\leq\beta_{c,k}^{\xi}-\sum_{k=1}^{K}\bar{r}_{c,k},\forall k\in[1,K],\end{aligned} \tag{45b}$$

$$\begin{aligned}&\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\boldsymbol{\xi}-2\mathrm{Re}\left\{\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\}\\&\leq\beta_{p,k}^{\xi}-R_{k}^{th}+\bar{r}_{c,k},\forall k\in[1,K],\end{aligned} \tag{45c}$$

where $\boldsymbol{\Omega}_{p,k}^{\xi}=\frac{1}{A}\sum_{a=1}^{A}\boldsymbol{\Omega}_{p,k}^{\xi,a}$, $\boldsymbol{\Omega}_{c,k}^{\xi}=\frac{1}{A}\sum_{a=1}^{A}\boldsymbol{\Omega}_{c,k}^{\xi,a}$,

$$\begin{aligned}\boldsymbol{\psi}_{p,k}^{\boldsymbol{\xi}}=\frac{1}{A}\sum_{a=1}^{A}\Bigg[&\lambda_{p,k}^{a}g_{p,k}^{a}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\boldsymbol{w}_{p,k}\\&-\lambda_{p,k}^{a}\left|g_{p,k}^{a}\right|^{2}\sum_{j=1}^{K}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\boldsymbol{w}_{p,j}\boldsymbol{w}_{p,j}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{BU},k}^{a}\\&-\lambda_{p,k}^{a}\left|g_{p,k}^{a}\right|^{2}\beta_{R,k}^{2}\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{BU},k}^{a}\\&-\lambda_{p,k}^{a}\left|g_{p,k}^{a}\right|^{2}\beta_{T}^{2}\left(1+\beta_{R,k}^{2}\right)\left(\boldsymbol{H}_{\mathrm{eff},k}^{a}\right)^{\mathrm{H}}\\&\times\widetilde{\mathrm{diag}}\left(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\right)\boldsymbol{h}_{\mathrm{BU},k}^{a}\Bigg]^{*}.\end{aligned} \tag{46}$$

$$\begin{aligned}\beta_{c,k}^{\xi}=\frac{1}{A}\sum_{a=1}^{A}\Big[&\log_{2}(\lambda_{c,k}^{\mathrm{a}})-\lambda_{c,k}^{\mathrm{a}}\\&-\lambda_{c,k}^{\mathrm{a}}|g_{c,k}^{\mathrm{a}}|^{2}(1+\beta_{R,k}^{2})\sigma_{k}^{2}+1\\&-\lambda_{c,k}^{\mathrm{a}}|g_{c,k}^{\mathrm{a}}|^{2}\sum_{j=1}^{K}\left|(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{p,j}\right|^{2}\\&-\lambda_{c,k}^{\mathrm{a}}|g_{c,k}^{\mathrm{a}}|^{2}\left|(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{c}\right|^{2}\\&-\lambda_{c,k}^{\mathrm{a}}|g_{c,k}^{\mathrm{a}}|^{2}\beta_{R,k}^{2}(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}}\\&-\lambda_{c,k}^{\mathrm{a}}|g_{c,k}^{\mathrm{a}}|^{2}\beta_{T}^{2}(1+\beta_{R,k}^{2})(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\\&\qquad\times\widetilde{\mathrm{diag}}(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}})\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}}\\&+2\lambda_{c,k}^{\mathrm{a}}\mathrm{Re}\left\{g_{c,k}^{\mathrm{a}}(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{c}\right\}\Big].\end{aligned} \tag{47}$$

Unedited

$$
\begin{aligned}
\boldsymbol{\psi}_{c,k}^{\xi} = & \frac{1}{A}\sum_{a=1}^{A}\Big[\lambda_{c,k}^{\mathrm{a}} g_{c,k}^{\mathrm{a}} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\boldsymbol{w}_{c} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\boldsymbol{w}_{c} \\
& \quad \boldsymbol{w}_{c}^{H}\boldsymbol{h}_{\mathrm{BU},k}^{a} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \sum_{j=1}^{K} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H} \\
& \quad \boldsymbol{w}_{p,j}\boldsymbol{w}_{p,j}^{H}\boldsymbol{h}_{\mathrm{BU},k}^{a} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{R,k}^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H} \\
& \quad \boldsymbol{W}\boldsymbol{W}^{H}\boldsymbol{h}_{\mathrm{BU},k}^{a} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{T}^{2} (1+\beta_{R,k}^{2}) (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H} \\
& \quad \times \widetilde{\mathrm{diag}}(\boldsymbol{W}\boldsymbol{W}^{H})\boldsymbol{h}_{\mathrm{BU},k}^{a}\Big]^{*}
\end{aligned} \tag{48}
$$

$$
\begin{aligned}
\beta_{c,k}^{\xi} = \frac{1}{A}\sum_{a=1}^{A}\Big[& \log_{2}\big(\lambda_{c,k}^{\mathrm{a}}\big) - \lambda_{c,k}^{\mathrm{a}} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} (1+\beta_{R,k}^{2})\sigma_{k}^{2} + 1 \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \sum_{j=1}^{K} \big|(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{p,j}\big|^{2} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \big|(\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{c}\big|^{2} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{R,k}^{2} (\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}} \\
& - \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{T}^{2} (1+\beta_{R,k}^{2}) (\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}} \\
& \quad \times \widetilde{\mathrm{diag}}(\boldsymbol{W}\mathbf{W}^{\mathrm{H}})\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}} \\
& + 2\lambda_{c,k}^{\mathrm{a}} \mathrm{Re}\Big\{ g_{c,k}^{\mathrm{a}} (\boldsymbol{h}_{\mathrm{BU},k}^{\mathrm{a}})^{\mathrm{H}}\boldsymbol{w}_{c}\Big\}\Big].
\end{aligned} \tag{49}
$$

where

$$
\begin{aligned}
\boldsymbol{\Omega}_{p,k}^{\boldsymbol{\xi},a} = \Big[& \sum_{j=1}^{K} \lambda_{p,k}^{\mathrm{a}} |g_{p,k}^{\mathrm{a}}|^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\boldsymbol{w}_{p,j}\boldsymbol{w}_{p,j}^{H}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \\
& + \lambda_{p,k}^{\mathrm{a}} |g_{p,k}^{\mathrm{a}}|^{2} \beta_{R,k}^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\mathbf{W}\mathbf{W}^{H}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \\
& + \lambda_{p,k}^{\mathrm{a}} |g_{p,k}^{\mathrm{a}}|^{2} \beta_{T}^{2} (1+\beta_{R,k}^{2}) (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H} \\
& \quad \times \widetilde{\mathrm{diag}}(\mathbf{W}\mathbf{W}^{H})\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\Big]^{*}.
\end{aligned} \tag{50}
$$

$$
\begin{aligned}
\boldsymbol{\Omega}_{c,k}^{\boldsymbol{\xi},a} = \Big[& \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\boldsymbol{w}_{c}\boldsymbol{w}_{c}^{H}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \\
& + \sum_{j=1}^{K} \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\boldsymbol{w}_{p,j}\boldsymbol{w}_{p,j}^{H}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \\
& + \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{R,k}^{2} (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H}\mathbf{W}\mathbf{W}^{H}\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}} \\
& + \lambda_{c,k}^{\mathrm{a}} |g_{c,k}^{\mathrm{a}}|^{2} \beta_{T}^{2} (1+\beta_{R,k}^{2}) (\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}})^{H} \\
& \quad \times \widetilde{\mathrm{diag}}(\mathbf{W}\mathbf{W}^{H})\boldsymbol{H}_{\mathrm{eff},k}^{\mathrm{a}}\Big]^{*}.
\end{aligned} \tag{51}
$$

While numerous methods exist to optimize IRS phase shifts under unit-modulus phase-shift constraints, the phase shifts may not belong to the DPS set due to their inherent limitations. To address this, we introduced $2K+1$ auxiliary variables, denoted as $\{\mathbf{z}_i\}_{i=1}^{2K+1} = \boldsymbol{\xi}$, transforming the problem into a quadratically constrained quadratic program (QCQP) with a single constraint. This reformulation enables efficient computation of the optimal solution for each auxiliary variable via the ADMM method, significantly reducing computational complexity (Huang and Sidiropoulos, 2016). The optimization problem is reformulated as P6:

$$
\begin{aligned}
\min_{\boldsymbol{\xi},\{\mathbf{z}_i\}_{i=1}^{2K+1}} & \sum_{k=1}^{K} \eta_k \boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\boldsymbol{\xi} - 2\sum_{k=1}^{K}\eta_k \mathrm{Re}\left\{\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\} \\
& + \rho \sum_{i=1}^{2\mathrm{K}+1} \|\mathbf{z}_i - \boldsymbol{\xi} + \boldsymbol{\mu}_i\|^2, && \text{(52a)} \\
\text{s.t.} & \angle[\mathbf{z}_{2K+1}]_m \in \mathbb{T}, \forall m \in [1, N_{\mathrm{I}}], && \text{(52b)} \\
& \mathbf{z}_k^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{\xi}\mathbf{z}_k - 2\mathrm{Re}\left\{\mathbf{z}_k^{\mathrm{H}}\boldsymbol{\psi}_{c,k}^{\xi}\right\} \\
& \le \beta_{c,k}^{\xi} - \sum_{k=1}^{K}\bar{r}_{c,k}, \forall k \in [1,K], && \text{(52c)} \\
& \mathbf{z}_{K+k}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\mathbf{z}_{K+k} - 2\mathrm{Re}\left\{\mathbf{z}_{K+k}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\} \\
& \le \beta_{p,k}^{\xi} - R_k^{\mathrm{th}} + \bar{r}_{c,k}, \forall k \in [1,K], && \text{(52d)}
\end{aligned}
$$

where $\boldsymbol{\mu}_i$ and ρ represent the dual variables and penalty factor, respectively. By decomposing the problem into three classes of sub-problems, the optimization problem can be solved iteratively.

Quadratic programming (QP) problem: optimizing $\boldsymbol{\xi}$ with fixing $\{\mathbf{z}_i\}_{k=1}^{2K+1}$, the corresponding subproblem is given by P6-1:

$$
\begin{aligned}
\min_{\boldsymbol{\xi}} \quad & \sum_{k=1}^{K} \eta_k \boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\boldsymbol{\xi} - 2\sum_{k=1}^{K}\eta_k \mathrm{Re}\left\{\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\} \\
& + \rho \sum_{i=1}^{2\mathrm{K}+1} \|\mathbf{z}_i - \boldsymbol{\xi} + \boldsymbol{\mu}_i\|^2,
\end{aligned}
$$

which can be optimally solved by the first-order optimality condition, yielding

$$
\begin{aligned}
\boldsymbol{\xi} = & \left(\sum_{k=1}^{K}\eta_k \boldsymbol{\Omega}_{p,k}^{\boldsymbol{\xi}} + (2\mathrm{K}+1)\,\rho\mathbf{I}_{N_I}\right)^{-1} \\
& \left[\sum_{k=1}^{K}\eta_k \boldsymbol{\psi}_{p,k}^{\boldsymbol{\xi}} + \rho\sum_{i=1}^{2\mathrm{K}+1}(\mathbf{z}_i + \boldsymbol{\mu}_i)\right].
\end{aligned} \tag{53}
$$

Unedited

QCQP with one constraint problem: optimizing $\{\mathbf{z}_i\}_{i=1}^{K}$ with fixing $\boldsymbol{\xi}$ and $\{\mathbf{z}_i\}_{i=K+1}^{2K+1}$, the second class of sub-problem is given by P6-2:

$$\begin{aligned}\min_{\{\mathbf{z}_k\}_{k=1}^{K}} \quad & \|\mathbf{z}_k - \boldsymbol{\xi} + \boldsymbol{\mu}_k\|^2 \\ \text{s.t.Eq.}(52c),\end{aligned} \tag{54}$$

which can be optimally solved by the Karush-Kuhn-Tucker (KKT) condition

$$\begin{aligned}\mathbf{z}_k = &\left(\eta_{1,k}\boldsymbol{\Omega}_{c,k}^{\boldsymbol{\xi}} + \mathbf{I}\right)^{-1} \\ &\times \left(\eta_{1,k}\boldsymbol{\psi}_{c,k}^{\boldsymbol{\xi}} - \boldsymbol{\mu}_k + \boldsymbol{\xi}\right), \quad \forall k \in [1, K].\end{aligned} \tag{55}$$

where $\eta_{1,k}$ is the Lagrange multiplier factor determined by

$$\begin{aligned}\eta_{1,k} = \min\Big\{\eta_{1,k} \geq 0 : \mathbf{z}_k^{\mathrm{H}}\boldsymbol{\Omega}_{c,k}^{\xi}\mathbf{z}_k - 2\mathrm{Re}\left\{\mathbf{z}_k^{\mathrm{H}}\boldsymbol{\psi}_{c,k}^{\xi}\right\} \\ \leq \beta_{c,k}^{\xi} - \sum_{k=1}^{K}\bar{r}_{c,k}\Big\},\end{aligned} \tag{56}$$

which can be optimized efficiently by the bisection search method. Similarly, when optimizing $\{\mathbf{z}_i\}_{k=K+1}^{2K}$ with fixing $\boldsymbol{\xi}$ and $\{\mathbf{z}_i\}_{i=1}^{K}$, the optimization problem is given by P6-3:

$$\begin{aligned}\min_{\{\mathbf{z}_k\}_{k=K+1}^{2K}} \quad & \|\mathbf{z}_{K+k} - \boldsymbol{\xi} + \boldsymbol{\mu}_{K+k}\|^2 \\ \text{s.t.Eq.}(52d),\end{aligned} \tag{57}$$

and can be solved by

$$\begin{aligned}\mathbf{z}_{K+k} = &\left(\eta_{2,k}\boldsymbol{\Omega}_{p,k}^{\boldsymbol{\xi}} + \mathbf{I}\right)^{-1}\left(\eta_{2,k}\boldsymbol{\psi}_{p,k}^{\boldsymbol{\xi}} - \boldsymbol{\mu}_{K+k} + \boldsymbol{\xi}\right), \\ &\forall k \in [1, K],\end{aligned} \tag{58}$$

where the corresponding Lagrange multiplier factor $\eta_{2,k}$ is determined by

$$\begin{aligned}\eta_{2,k} = \min\Big\{\eta_{2,k} \geq 0 : \mathbf{z}_{K+k}^{\mathrm{H}}\boldsymbol{\Omega}_{p,k}^{\xi}\mathbf{z}_{K+k} \\ -2\mathrm{Re}\left\{\mathbf{z}_{K+k}^{\mathrm{H}}\boldsymbol{\psi}_{p,k}^{\xi}\right\} \leq \beta_{p,k} - R_k^{th} + \bar{r}_{c,k}^{\xi}\Big\}.\end{aligned} \tag{59}$$

QP with discrete constraint: optimizing $\mathbf{z}_{2K+1}$ with fixing $\boldsymbol{\xi}$ and $\{\mathbf{z}_i\}_{i=1}^{2K}$. The optimization problem is given by P6-4:

$$\begin{aligned}\min_{\{\mathbf{z}_{2K+1}\}_{k=1}^{K}} \quad & 2\sum_{m=1}^{N_I}|[\boldsymbol{\varsigma}]_m|\cos\left(\angle[\boldsymbol{\varsigma}]_m - \angle[\mathbf{z}_{2K+1}]_m\right) \\ \text{s.t.Eq.}(52b),\end{aligned} \tag{60}$$

where $\boldsymbol{\varsigma} = \boldsymbol{\mu}_{2K+1} - \boldsymbol{\xi}$. It is observed that $[\mathbf{z}_{2K+1}]_m$ in (P6-4) decouples from both the objective and the constraint. Therefore, the optimization variables can be updated in parallel, with the optimal solution given by

$$\mathbf{z}_{2K+1} = e^{j(\arg\min_{\boldsymbol{\vartheta}\in\mathbb{T}}\{\ \mathrm{mod}\ (|\boldsymbol{\vartheta}-\angle\boldsymbol{\varsigma}|,2\pi)-\pi\})}. \tag{61}$$

Finally, update the dual variables by

$$\boldsymbol{\mu}_i = \mathbf{z}_i - \boldsymbol{\xi} + \boldsymbol{\mu}_i, \forall i \in [1, 2K+1]. \tag{62}$$

By iteratively solving these three classes of QP problems, the objective function in Eq. (45a) can be guaranteed to be non-decreasing, eventually converging to stationary solutions that satisfy the constraint $\{\mathbf{z}_i\}_{i=1}^{2K+1} = \boldsymbol{\xi}$ as proven in (Huang and Sidiropoulos, 2016).

In summary, the proposed SAA-based algorithm for the robust RSMA multiuser communication system is outlined in Algorithm 1, where ϵ denotes the optimization accuracy for the phase shifts. Additionally, initializing the optimization variables with a feasible solution can improve the convergence speed of the algorithm and reduce the likelihood of infeasible solutions (Peng et al., 2024b). Specifically, the variables $\mathbf{z}_i^0 = \boldsymbol{\xi}^0$ and $\boldsymbol{\mu}^0 = \mathbf{0}$ are initialized to ensure feasibility. To ensure the convexity of the objective function, we set ρ to be the largest eigenvalue of matrix $\sum_{k=1}^{K}\eta_k\boldsymbol{\Omega}_{p,k}^{\xi}$, i.e., $\rho = \lambda_{\max}(\sum_{k=1}^{K}\eta_k\boldsymbol{\Omega}_{p,k}^{\xi})$, following (Huang and Sidiropoulos, 2016). Additionally, the optimization variable $\bar{r}_{\mathrm{c,k}}$ is initialized to zero, reducing the RSMA problem to an SDMA optimization problem, which is solved efficiently using the SAA method and the algorithm from (Tao et al., 2022), providing feasible solutions $\mathbf{W}^0$ and $\boldsymbol{\xi}^0$ that satisfy the optimization constraints.

3.2 Algorithm analysis

In this subsection, we analyzed the proposed SAA-based algorithm from the perspectives of convergence behavior and computational complexity.

3.2.1 Convergence behavior

The value of the objective function, i.e., Eq. (15a), is upper bounded since the BS transmit power is limited. Additionally, the optimization variables are decoupled by dividing the optimization problem (P2) into (P3), (P4), and (P6). To update the weights and equalizers, the optimal solution of $\lambda_{c,k}^{\mathrm{a}}$, $\lambda_{p,k}^{\mathrm{a}}$, $g_{c,k}^{\mathrm{a}}$, and $g_{p,k}^{\mathrm{a}}$ are derived, ensuring the WWMSE-rate relationship holds. For updating $\bar{r}_{c,k}$ and the beamforming matrix at BS,

Unedited

Algorithm 1 The proposed SAA-based robust algorithm

Initialize variables $\mathbf{W}^0$, $\bar{r}_{c,k}^0$, $\boldsymbol{\xi}^0$, $\boldsymbol{\mu}_i^0$, $\mathbf{z}_i^0$, ρ, and set the inner and outer iteration indices $t_i = 1$ and $t_o = 1$, respectively.
repeat
  Calculate the weights $\{\lambda_{c,k}^{\mathrm{a}}, \forall k, \forall a\}^{t_o}$ and $\{\lambda_{p,k}^{\mathrm{a}}, \forall k, \forall a\}^{t_o}$ by Eq. (26).
  Calculate the equalizers $\{g_{c,k}^{\mathrm{a}}, \forall k, \forall a\}^{t_o}$ and $\{g_{p,k}^{\mathrm{a}}, \forall k, \forall a\}^{t_o}$ by Eqs. (21) and (22).
  Update the beamforming at BS $\mathbf{W}^{t_o}$ and $\bar{r}_{c,k}^{t_o}$ by solving (P4-1).
  repeat
    Update the phase shift vector $\boldsymbol{\xi}^{t_i}$ by Eq. (53).
    Update the auxiliary variables $\{\mathbf{z}_i^{t_i}\}_{i=1}^{K}$, $\{\mathbf{z}_i^{t_i}\}_{i=K=1}^{2K}$, $\mathbf{z}_{2K+1}^{t_i}$ respectively by Eqs. (55), (58) and (61).
    Update the dual variable $\boldsymbol{\mu}_i^{t_i}$ by Eq. (62).
    Set $t_i = t_i + 1$.
  until the constraint violation indicator $\max_{i\in[1,2K+1]}\{\|\boldsymbol{\xi} - \mathbf{z}_i\|_F\}$ is smaller than ϵ or reaches the maximum number of inner iterations T_I.
  Set $\boldsymbol{\xi}^{t_o} = \boldsymbol{\xi}^{t_i}$.
  Set $t_o = t_o + 1$.
until the objective function value, i.e., Eq. (18a) converges or reaches the maximum number of inner iterations T_O.

the standard interior point method is applied, which guarantees the non-decreasing behavior of the objective function. Similarly, the ADMM algorithm used for optimizing the phase shift ensures the objective function remains non-decreasing. Hence, the proposed algorithm is non-decreasing and upper-bounded, ensuring the convergence of the proposed algorithm.

3.2.2 Computational complexity

The computational complexity for calculating weights $\{\lambda_{c,k}^{\mathrm{a}}, \lambda_{p,k}^{\mathrm{a}}, \forall a = [1, A], k = [1, K]\}$, and equalizers $\{g_{c,k}^{\mathrm{a}}, g_{p,k}^{\mathrm{a}}, \forall a = [1, A], k = [1, K]\}$ is dominated by matrix multiplication, yielding a complexity of $\mathcal{O}(AKN_{\mathrm{B}}N_{\mathrm{I}})$. Optimizing $\boldsymbol{W}$ and $\bar{r}_{c,k}$ involves matrix multiplication and solving problem (P4-1) using the interior-point method, resulting in complexities of $\mathcal{O}\left(AN_{\mathrm{B}}^2N_{\mathrm{I}}\right)$ and $\mathcal{O}\left(N_{\mathrm{B}}^{3.5}K^{3.5}\right)$, respectively. Similarly, optimizing the IRS phase shifts requires matrix multiplication with complexity $\mathcal{O}\left(AN_{\mathrm{B}}N_{\mathrm{I}}^2\right)$, while optimizing $\boldsymbol{\xi}$ and $\{\mathbf{z}_i\}_{i=1}^{2K+1}$ involves matrix inversion, contributing a complexity of $\mathcal{O}\left(KN_{\mathrm{I}}^3\right)$. Additionally, updating the dual variables $\eta_{1,k}$ and $\eta_{2,k}$ via the bisection method introduces a complexity of $\mathcal{O}\left(KN_{\mathrm{I}}^3\log(l)\right)$, where l represents the interval length of the search. Thus, the overall computational complexity of the proposed algorithm is $\mathcal{O}\left(AN_{\mathrm{B}}N_{\mathrm{I}}\left(N_{\mathrm{B}} + N_{\mathrm{I}} + K\right) + K N_{\mathrm{I}}^3\log(l) + N_{\mathrm{B}}^{3.5}K^{3.5}\right)$.

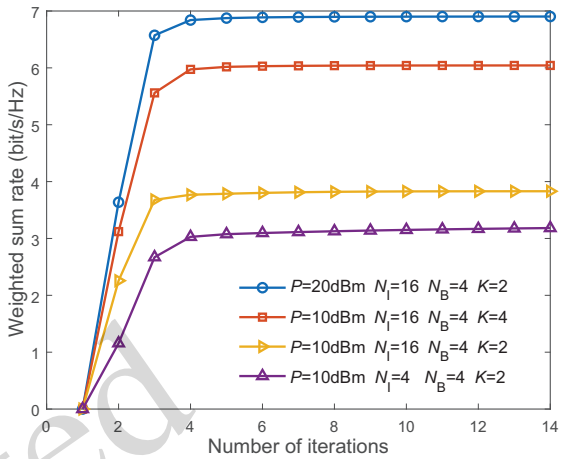


Fig. 2 Convergence behavior of the proposed SAA-based robust algorithms. SAA, sample average approximation

4 Numerical result

In this section, numerical results are provided to evaluate the performance of the proposed SAA-based algorithm for the IRS-aided multiuser RSMA system under practical imperfections. The BS is deployed at the origin [0 m,0 m,0 m], with the distance from the BS to IRS and from the IRS to the users set at 100 m and 10 m, respectively (Li et al., 2023). The large-scale path loss follows the model $C_0 d^{-\mathrm{ff}}$, where d is the distance, $C_0 = -30$ dB is the path loss at a 1-m reference distance, and α is the path loss exponent. Due to extensive obstacles, the path loss exponent between the BS and the users, between the BS and the IRS, and between the IRS and the users are set as 3.5, 2.2, and 2.2, respectively. Small-scale fading is modeled using Rician fading with a Rician factor of 5. All users have equal communication priority, i.e., $\eta_{\mathrm{k}} = 1$, and the number of channel samples is $A = 5,000$. Furthermore, hardware distortion levels are set as $\beta_{\mathrm{T}} = \beta_{\mathrm{R}} = \beta = 0.08$, with CSI error variance $\delta_{\mathrm{k}} = 0.2$ (Wang et al., 2023). Unless otherwise stated, simulation parameters are set as $K = 4$, $N_{\mathrm{I}} = 16$, $\sigma_k^2 = -100$ dBm, $P = 30$ dBm, $N_{\mathrm{B}} = 4$, $T_{\mathrm{I}} = T_{\mathrm{O}} = 1000$, $\epsilon = 10^{-4}$, and $R_k^{\mathrm{th}} = 0.2$ bit/s/Hz (Pala et al., 2024).

We first evaluated the convergence behavior of the proposed SAA-based algorithm. As illustrated in Fig. 2,

Unedited

the algorithm converges within a few iterations across varying numbers of users, transmit power levels, and IRS elements. This swift convergence stems from the closed-form structure of the optimization variables, which significantly accelerates the process.

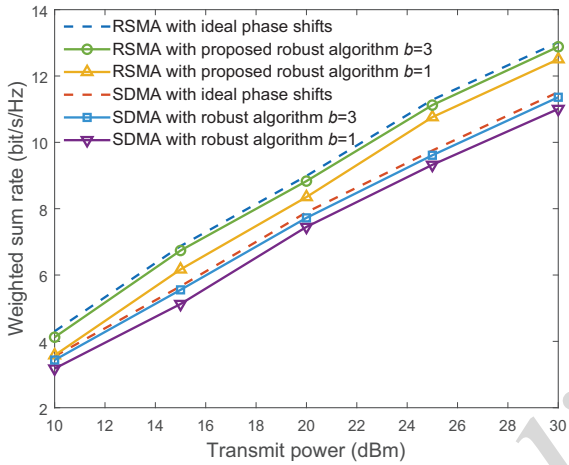


Fig. 3 Performance comparison between RSMA and SDMA under different quantization numbers. RSMA, rate-splitting multiple access; SDMA, space division multiple access

We then evaluated the ergodic weighted sum-rate performance of IRS-assisted RSMA and SDMA systems under practical conditions. As illustrated in Figs 3 and 4, both RSMA and SDMA systems exhibit performance degradation due to practical impairments.[2] However, the IRS-assisted RSMA system consistently outperforms its SDMA counterpart across all scenarios, including variations in DPSs (Fig. 3), hardware distortion levels, and CSI errors (Fig. 4). These results highlight the robustness of IRS-assisted RSMA, demonstrating its superior capability to mitigate practical limitations.[3]

The impact of the number of IRS elements on system performance is analyzed. As shown in Fig. 5, increasing the number of IRS elements significantly enhances system performance. Additionally, the proposed

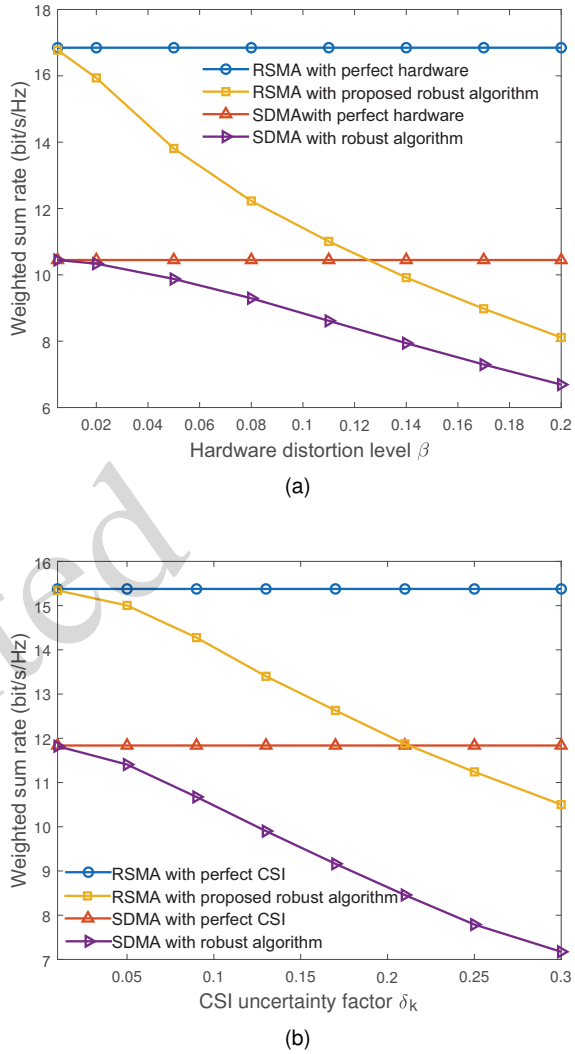


Fig. 4 The impact of hardware distortion level (a) and the impact of CSI estimation errors (b). CSI, channel state information; RSMA, rate-splitting multiple access; SDMA, space division multiple access

RSMA-based algorithm outperforms the SDMA-based benchmark scheme, highlighting the superior robustness of RSMA under practical imperfections.

To comprehensively evaluate the proposed algorithm, we compared the performance of the proposed algorithm against the following benchmark algorithms:

1. Perfect hardware: Assumes ideal hardware, i.e., $\beta_T = \beta_R = 0$, with imperfect CSI, serving as the upper-bound performance for systems with hardware impairments.

2. Benchmark HI 1 proposed by Li et al. (2023): Utilizes the proposed phase shift optimization algorithm while adopting the BS beamforming design from Li et al. (2023), which ignores hardware impairments.

[2]It is noted that the two subfigures in Fig. 4 represent different scenarios. Specifically, the "RSMA with perfect hardware" curve corresponds to a case with CSI errors but no hardware impairments. In contrast, the "RSMA with perfect CSI" curve represents a scenario with hardware impairments but no CSI errors. The impact of CSI errors and hardware impairments is different, thus their horizontal levels is different.

[3]NOMA is not included in this work, as prior studies (Mao et al., 2022, 2019; Li et al., 2022) have extensively demonstrated the superiority of RSMA over NOMA. Moreover, IRS-assisted SDMA provides a more competitive baseline for RSMA performance evaluation, serving as a tighter benchmark compared to NOMA (Li et al., 2024).

Unedited

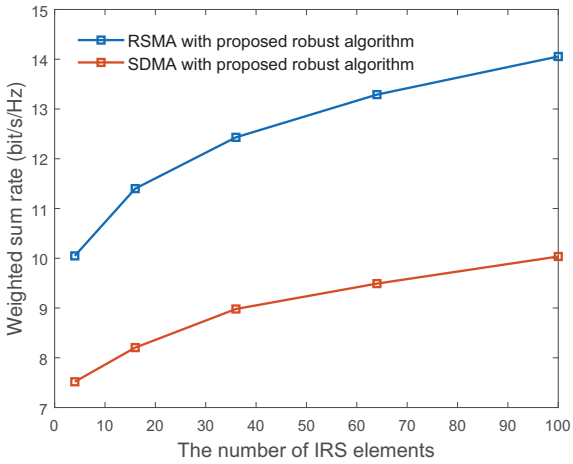


Fig. 5 The impact of IRS elements on system performance. IRS, intelligent reflecting surface; RSMA, rate-splitting multiple access; SDMA, space division multiple access

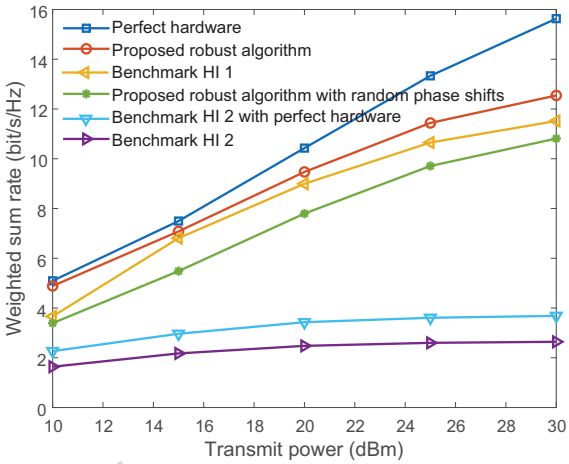


Fig. 6 Performance comparison of different beamforming design algorithms under hardware distortions

Unedited

3. Benchmark HI 2 proposed by Tao et al. (2022): Employs robust beamforming from Tao et al. (2022) at the BS, with the IRS phase shift optimized by the proposed algorithm.

4. Perfect CSI: Assumes perfect CSI, i.e., $\delta_{\mathrm{k}}^2 = 0$, with hardware distortions, representing the upper-bound performance for systems with CSI errors.

5. Benchmark CSI 1 proposed by Xia et al. (2024): Utilizes the proposed phase shift optimization algorithm while adopting the BS beamforming design from Xia et al. (2024), which ignores CSI error.

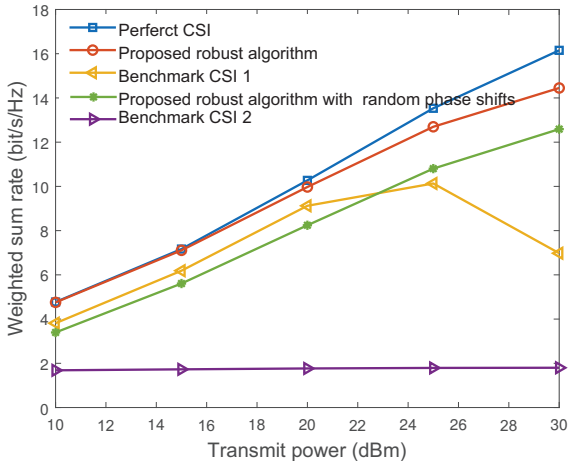
6. Benchmark CSI 2 proposed by Tao et al. (2022): Employs the robust algorithm from Tao et al. (2022), with the IRS phase shift optimized by the proposed method to mitigate hardware distortions.

7. Ideal phase shifts: Assumes continuous IRS phase shifts, serving as the upper-bound performance for systems with DPSs.

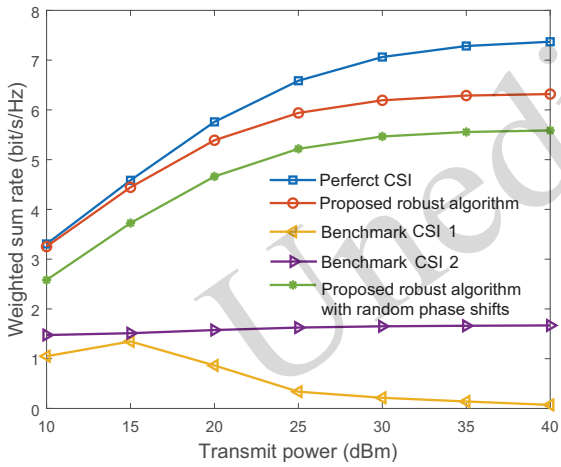
8. Benchmark DPS 1 adopted by Wang et al. (2023) with b=3: Randomly selects DPS from the discrete phase set with 3-bit quantization Wang et al. (2023), while the BS beamforming is optimized by the proposed algorithm.

9. Benchmark DPS 2 proposed by Peng et al. (2024a) with b=3: Optimizes DPS with 3-bit quantization proposed in Peng et al. (2024a), paired with the proposed BS beamforming optimization.

Fig. 6 illustrates the impact of hardware distortions on system performance. The results show significant performance degradation in benchmark HI 2. Moreover, the performance improvement with perfect hardware in benchmark HI 2 is limited because the proposed IRS robust design mitigates hardware impairments, approaching the ideal case. However, this MRT-based scheme fails to address multi-user interference effectively, which, combined with CSI errors, leads to significant degradation. This highlights the critical need to account for hardware impairments in practical systems to enhance user quality of service (QoS). Furthermore, the proposed SAA-based algorithm consistently outperforms the benchmark algorithms, demonstrating its superior capability to mitigate hardware impairments and improve overall system performance. Fig. 7 illustrates the impact of CSI estimation errors on system performance. The results reveal a performance gap between the proposed SAA-based algorithm and the benchmarks. Specifically, unlike the algorithm proposed by Tao et al. (2022) that relies on fixed beamforming at the BS, the proposed robust SAA-based algorithm is better suited to address CSI uncertainties in IRS-assisted multiuser RSMA scenarios. Besides, although benchmark 1 accounts for the impact of hardware impairments, system performance still degrades and worsens as transmit power increases since the power of the distortion noise is influenced by imperfect CSI. Moreover, by increasing the hardware impairment error to $\beta_{\mathrm{T}} = \beta_{\mathrm{R}} = \beta = 0.3$, we observed a clear performance floor in Fig. 7 due to the presence of imperfections. This underscores the importance of developing robust algorithms that jointly account for CSI errors and hardware distortions to improve system performance. Fig. 8 shows the performance of



(a)



(b)

Fig. 7 Performance comparison of different beamforming design algorithms under imperfect CSI with $\beta_T = \beta_R = \beta = 0.08$ (a) and with $\beta_T = \beta_R = \beta = 0.3$ (b). CSI, channel state information

the system under different DPS design algorithms. It is demonstrated that the proposed ADMM-based DPS design outperforms the benchmark algorithms. Moreover, a low-resolution IRS DPS can achieve satisfactory performance in the presence of practical hardware implementations and CSI errors. Thus, the proposed algorithm offers an effective and computationally efficient solution[4] for optimizing IRS phase shifts while ensuring robust system performance.

[4]Unlike numerical methods such as interior-point algorithms in CVX, the IRS phase-shift design proposed in this paper provides closed-form solutions for the optimization variables. This approach accelerates convergence, reducing the number of iterations needed to find the optimal solution (Huang and Sidiropoulos, 2016).

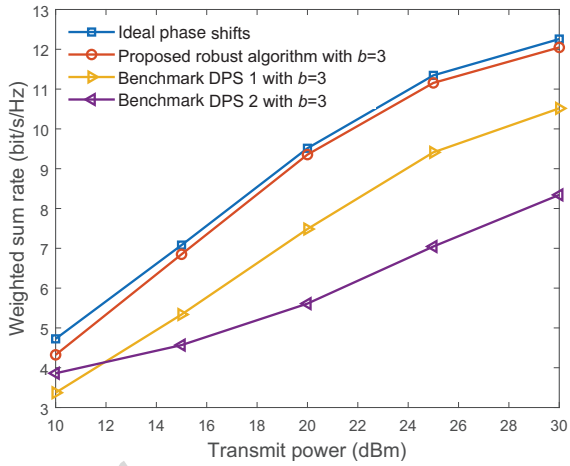


Fig. 8 Performance comparison of different beamforming design algorithms under DPS. DPS, discrete phase shift

5 Conclusions

This study investigated an IRS-assisted multiuser RSMA system, where practical system imperfections, such as hardware impairments, imperfect CSI, and discrete IRS phase shifts, were taken into consideration. To mitigate the performance degradation caused by these adverse imperfections, an SAA-based robust algorithm was proposed. By leveraging the relationship between WMMSE and weighted sum rate, we jointly optimized the IRS phase shifts and the beamforming matrix at the BS, deriving closed-form expressions for the optimization variables. Simulation results demonstrated that both IRS-assisted SDMA and RSMA systems experience significant performance degradation due to imperfections. However, the RSMA system exhibits superior robustness. Additionally, the proposed SAA-based robust algorithm did not rely on fixed beamforming at BS, outperforming existing benchmark algorithms. Furthermore, the proposed algorithm effectively addressed the issue of DPSs at the IRS, with simulation results showing that performance can be guaranteed using 3-bit quantization, in the presence of imperfect CSI and hardware distortions.

Contributors

Xingyu PENG designed the research and processed the data. Xingyu PENG, Qin TAO, and Xiaoming CHEN drafted the paper. Qin TAO helped organize the paper. Xingyu PENG and Xiaoming CHEN revised and finalized the paper.

Conflct of interest

All the authors declare that they have no conflict of interest.

Unedited

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aboumahmoud I, Hossain E, Mezghani A, 2025. Resource management in ris-assisted rate splitting multiple access for next generation (xg) wireless communications: Models, state-of-the-art, and future directions. *IEEE Commun Surv Tutorials*, 27(3):1618-1655. https://doi.org/10.1109/COMST.2024.3471453

Asif M, Bao X, Ranjha A, et al., 2025. Leveraging ris in consumer-centric 6g networks: Efficient resource allocation in rsma-based swipt systems under hardware impairments. *IEEE Trans Consumer Electron*, 71(2):4235-4247. https://doi.org/10.1109/TCE.2024.3512939

Bansal A, Singh K, Clerckx B, et al., 2021. Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications. *IEEE Trans Veh Technol*, 70(9):9217-9229. https://doi.org/10.1109/TVT.2021.3102212

Björnson E, Zetterberg P, Bengtsson M, et al., 2013. Capacity limits and multiplexing gains of mimo channels with transceiver impairments. *IEEE Commun Lett*, 17(1):91-94. https://doi.org/10.1109/LCOMM.2012.112012.122003

Björnson E, Hoydis J, Kountouris M, et al., 2014. Massive mimo systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits. *IEEE Trans Inf Theory*, 60(11):7112-7139. https://doi.org/10.1109/TIT.2014.2354403

Clerckx B, Mao YJ, Jorswieck EA, et al., 2023. A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions. *IEEE J Sel Areas Commun*, 41(5):1265-1308. https://doi.org/10.1109/JSAC.2023.3242718

Efrem CN, Krikidis I, 2024. Robust irs-element activation for energy efficiency optimization in irs-assisted communication systems with imperfect csi. *IEEE Trans Wireless Commun*, 23(10):14380-14393. https://doi.org/10.1109/TWC.2024.3413022

Fu H, Feng SL, Ng DWK, 2021. Resource allocation design for irs-aided downlink mu-miso rsma systems. 2021 IEEE Int Conf on Communications Workshops, p.1-6. https://doi.org/10.1109/ICCWorkshops50388.2021.9473650

Gallager RG, 2008. Principles of Digital Communication. Cambridge University Press, Cambridge, UK.

Huang KJ, Sidiropoulos ND, 2016. Consensus-admm for general quadratically constrained quadratic programming. *IEEE Trans Signal Process*, 64(20):5297-5310. https://doi.org/10.1109/TSP.2016.2593681

Khel AMT, Hamdi KA, 2022. Effects of hardware impairments on irs-enabled miso wireless communication systems. *IEEE Commun Lett*, 26(2):259-263. https://doi.org/10.1109/LCOMM.2021.3134815

Kolawole OY, Biswas S, Singh K, et al., 2020. Transceiver design for energy-efficiency maximization in mmwave mimo iot networks. *IEEE Trans Green Commun Netw*, 4(1):109-123. https://doi.org/10.1109/TGCN.2019.2943956

Li BJ, Chen W, Li ZD, et al., 2023. Robust weighted sum-rate maximization for transmissive ris transmitter enabled rsma networks. *IEEE Commun Lett*, 27(10):2847-2851. https://doi.org/10.1109/LCOMM.2023.3304757

Li HY, Mao YJ, Dizdar O, et al., 2022. Rate-splitting multiple access for 6g—part iii: Interplay with reconfigurable intelligent surfaces. *IEEE Commun Lett*, 26(10):2242-2246. https://doi.org/10.1109/LCOMM.2022.3192041

Li HY, Shen SP, Clerckx B, 2023. Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures. *IEEE Trans Wireless Commun*, 22(4):2311-2324. https://doi.org/10.1109/TWC.2022.3210706

Li HY, Shen SP, Clerckx B, 2024. Synergizing beyond diagonal reconfigurable intelligent surface and rate-splitting multiple access. *IEEE Trans Wireless Commun*, 23(8):8717-8729. https://doi.org/10.1109/TWC.2024.3353596

Ma WY, Zhu LP, Zhang R, 2024. Mimo capacity characterization for movable antenna systems. *IEEE Trans Wireless Commun*, 23(4):3392-3407. https://doi.org/10.1109/TWC.2023.3307696

Mao YJ, Clerckx B, Li VOK. Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming sdma and noma. *EURASIP J Wireless Commun Networking*, 2018(133).

Mao YJ, Clerckx B, Li VOK, 2019. Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis. *IEEE Trans Commun*, 67(12):8754-8770. https://doi.org/10.1109/TCOMM.2019.2943168

Mao YJ, Dizdar O, Clerckx B, et al., 2022. Rate-splitting multiple access: Fundamentals, survey, and future research trends. *IEEE Commun Surv Tutorials*, 24(4):2073-2126. https://doi.org/10.1109/COMST.2022.3191937

Pala S, Katwe M, Singh K, et al., 2024. Spectral-efficient ris-aided rsma urllc: Toward mobile broadband reliable low latency communication (mbrllc) system. *IEEE Trans Wireless Commun*, 23(4):3507-3524. https://doi.org/10.1109/TWC.2023.3309028

Papazafeiropoulos A, Clerckx B, Ratnarajah T, 2017. Rate-splitting to mitigate residual transceiver hardware impairments in massive mimo systems. *IEEE Trans Veh Technol*, 66(9):8196-8211. https://doi.org/10.1109/TVT.2017.2691014

Peng XY, Hu XL, Gao JB, et al., 2024a. Integrated localization and communication for irs-assisted multi-user mmwave mimo systems. *IEEE Trans Commun*, 72(8):4725-4740. https://doi.org/10.1109/TCOMM.2024.3379399

Peng XY, Tao Q, Hu XL, et al., 2024b. Integrated sensing and communication in irs-assisted high-mobility systems: Design, analysis and optimization. *IEEE Trans Wireless Commun*, 23(11):16107-16122. https://doi.org/10.1109/TWC.2024.3436932

Schenk T, 2008. RF imperfections in high-rate wireless systems: Impact and digital compensation. Springer, Dordrecht, the Netherlands.

Shen H, Xu W, Gong SL, et al., 2021. Beamforming optimization for irs-aided communications with transceiver hardware impairments. *IEEE Trans Commun*, 69(2):1214-1227. https://doi.org/10.1109/TCOMM.2020.3033575

Tao Q, Zhang SW, Zhong CJ, et al., 2022. Weighted sum-rate of intelligent reflecting surface aided multiuser downlink transmission with statistical csi. *IEEE Trans Wireless Commun*, 21(7):4925-4937. https://doi.org/10.1109/TWC.2021.3134822

Tataria H, Shafi M, Molisch AF, et al., 2021. 6g wireless systems: Vision, requirements, challenges, insights, and opportunities. *Proc IEEE*, 109(7):1166-1199. https://doi.org/10.1109/JPROC.2021.3061701

Unedited

Wang CJ, Zhang XH, Xing HJ, et al., 2025. Joint association, beamforming, and resource allocation for multi-irs enabled mu-miso systems with rsma. *IEEE Trans Mobile Comput*, 24(3):1602-1620. https://doi.org/10.1109/TMC.2024.3483193

Wang JT, Gong SQ, Wu QQ, et al., 2023. Ris-aided mimo systems with hardware impairments: Robust beamforming design and analysis. *IEEE Trans Wireless Commun*, 22(10):6914-6929. https://doi.org/10.1109/TWC.2023.3246990

Wang WH, Yang L, Zhan YY, et al., 2025. Robust resource allocation design for energy-efficient active irs-aided c-rsma systems. *IEEE Trans Commun*, 73(7):5168-5183. https://doi.org/10.1109/TCOMM.2024.3506933

Wang ZR, Liu L, Cui SG, 2020. Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis. *IEEE Trans Wireless Commun*, 19(10):6607-6620. https://doi.org/10.1109/TWC.2020.3004330

Weinberger K, Ahmad AA, Sezgin A, et al., 2022. Synergistic benefits in irs-and rs-enabled c-ran with energy-efficient clustering. *IEEE Trans Wireless Commun*, 21(10):8459-8475. https://doi.org/10.1109/TWC.2022.3166393

Wu QQ, Zhang R, 2020. Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network. *IEEE Commun Mag*, 58(1):106-112. https://doi.org/10.1109/MCOM.001.1900107

Wu QQ, Zhang SW, Zheng BX, et al., 2021. Intelligent reflecting surface-aided wireless communications: A tutorial. *IEEE Trans Commun*, 69(5):3313-3351. https://doi.org/10.1109/TCOMM.2021.3051897

Xia HY, Mao YJ, Zhou XK, et al., 2024. Weighted sum-rate maximization of rate-splitting multiple access with confidential messages. *IEEE Trans Wireless Commun*, 23(10):13738-13751. https://doi.org/10.1109/TWC.2024.3404095

Xu CC, Clerckx B, Chen SW, et al., 2021. Rate-splitting multiple access for multi-antenna joint radar and communications. *IEEE J Sel Top Signal Process*, 15(6):1332-1347. https://doi.org/10.1109/JSTSP.2021.3110312

Yan WJ, Yuan XJ, He ZQ, et al., 2020. Passive beamforming and information transfer design for reconfigurable intelligent surfaces aided multiuser mimo systems. *IEEE J Sel Areas Commun*, 38(8):1793-1808. https://doi.org/10.1109/JSAC.2020.3000811

Yang ZH, Shi JF, Li ZY, et al., 2020. Energy efficient rate splitting multiple access (rsma) with reconfigurable intelligent surface. Proc IEEE Int Conf on Communications Workshops, p.1-6. https://doi.org/10.1109/ICCWorkshops49005.2020.9145189

Zhang J, Kountouris M, Andrews JG, et al., 2011. Multi-mode transmission for the mimo broadcast channel with imperfect channel state information. *IEEE Trans Commun*, 59(3):803-814. https://doi.org/10.1109/TCOMM.2011.121410.100144

Zhang YP, You CS, Zheng BX, 2023. Multi-active multi-passive (mamp)-irs aided wireless communication: A multi-hop beam routing design. *IEEE J Sel Areas Commun*, 41(8):2497-2513. https://doi.org/10.1109/JSAC.2023.3288233

Unedited