Secure NOMA Systems with a Dual-Functional RIS: Simultaneous Information Relaying and Jamming

Mengyi Ji, Jian Chen, *Member, IEEE*, Lu Lv, *Member, IEEE*, Qingqing Wu, *Senior Member, IEEE*, Zhiguo Ding, *Fellow, IEEE*, and Naofal Al-Dhahir, *Fellow, IEEE*

***Abstract*—In this paper, we propose a new simultaneous information relaying and jamming (SIRJ) scheme based on a dual-functional reconfigurable intelligent surface (RIS) to achieve secure non-orthogonal multiple access communications. Specifically, the RIS elements are split into two groups, where elements in one group perform signal reflection to enhance the legitimate reception quality while elements in the other group generate artificial jamming to interfere with the eavesdropper (Eve). Based on different channel state information (CSI) availabilities of Eve, the system sum-rate is maximized by jointly optimizing the transmit beamforming of the base station, reflect beamforming of the RIS, and mode selection of each RIS element, subject to a maximum tolerable information leakage to Eve. For the case with perfect Eve's CSI, a penalty based alternating algorithm is proposed to deal with the challenging multivariate coupled and mixed integer non-convex optimization problem. For the case with imperfect Eve's CSI, we consider the infinite number of secrecy constraints, for which the traditional $\mathcal{S}$-procedure cannot be directly applied. To tackle this challenge, we devise an efficient transformation that fits the $\mathcal{S}$-procedure to the problem and propose a robust secure beamforming design. Simulation results demonstrate the performance advantage of the proposed SIRJ scheme over the existing baseline schemes.**

***Index Terms*—Reconfigurable intelligent surface, non-orthogonal multiple access, physical layer security.**

I. Introduction

In the development process of fifth generation (5G) wireless communications, both academia and industry made great efforts to design enhanced mobile broadband (eMBB), massive machine type communication (mMTC) and ultra-reliable low latency communication (URLLC) schemes. As a result, many promising technologies have emerged, such as the massive multiple-input multiple-output (MIMO), millimeter wave (mmWave) and ultra-dense network (UDN) [2]. With the commercialization of 5G, the research focus has shifted to future 6th generation (6G) wireless communication. In 6G, it is expected that the network can support connections of about 10^7 devices per square kilometer to cope with the explosive growth of communication devices. Meanwhile, to meet various advanced digital services and multimedia entertainment, such as truly immersive extended reality (XR), digital replica and so on, it is imperative that 6G can provide a peak data rate of 1000 Gbps [3]. However, these 6G requirements undoubtedly make the problem of spectrum resource shortage even more critical.

Against this background, non-orthogonal multiple access (NOMA), which was first proposed in [4], is one of the promising technologies for 6G networks due to its high spectrum utilization. Different from conventional orthogonal multiple access (OMA), NOMA allows users to share the same time or frequency resource block but distinguishes them in the power domain. Under the NOMA protocol, a transmitter sends a superimposed signal that contains multiple users' information with different power levels. At the receiver end, successive interference cancellation (SIC) is applied for decoding, that is, the strong signal is decoded first by treating the weak signal as interference, and this process is repeated until the desired signal is obtained [5]. Therefore, existing works point out that NOMA can provide a massive connection with a high spectral efficiency while ensuring good fairness among users with different channel conditions, which is superior to OMA [5]–[7], and promises to meet the 6G requirements.

Every coin has two sides. Although NOMA can achieve a high spectral efficiency, it also faces security challenges like other technologies, and needs more attention due to not only the broadcast characteristic of wireless channels, but also NOMA itself. In particular, SIC provides a chance for the eavesdropper (Eve) to intercept all users' information contained in the superimposed signal [8]. Therefore, necessary physical layer security (PLS) designs are needed to safeguard, such as the applications of secure beamforming, artificial noise (AN), and specially-designed jamming [9]–[12]. In [9], the beamforming for users were jointly designed, and AN was

Part of this paper was presented in Wireless Communications and Signal Processing (WCSP) 2022 [1]. This work was supported in part by the National Natural Science Foundation of China under Grant 62271368; in part by the Key Research and Development Program of Shaanxi under Grant 2023-YBGY-041; in part by the Fundamental Research Funds for the Central Universities; and in part by the China Postdoctoral Science Foundation under Grant BX20190264 and Grant 2019M650258. The work of Qingqing Wu was supported by the FDCT under Grant 0119/2020/A3. The work of Naofal Al-Dhahir was supported by Erik Jonsson Distinguished Professorship at UT-Dallas. The associate editor coordinating the review of this article and approving it for publication was M. Vu. *(Corresponding author: Lu Lv.)*

Mengyi Ji, Jian Chen, and Lu Lv are with the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mails: myji_1@stu.xidian.edu.cn; jianchen@mail.xidian.edu.cn; lulv@xidian.edu.cn).

Qingqing Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: qingqingwu@sjtu.edu.cn).

Zhiguo Ding is with the Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788, UAE (e-mail: zhiguo.ding@manchester.ac.uk).

Naofal Al-Dhahir is with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA (e-mail: aldhahir@utdallas.edu).

adopted via zero forcing with an optimized power allocation. Arafa et al. [10] realized secure transmission in which all users sending jamming in a cooperative way to degrade the reception quality at an untrusted relay. Another jamming based secure scheme was proposed in [11], where the jamming power was maximized so that the jamming can be canceled by SIC firstly at legitimate users. Different from designing the decoding order of jamming, the transmission rate of jamming was adapted according to channel conditions to protect the communication from an untrusted relay in [12]. However, these approaches all need excessive energy consumption, which makes them not sustainable and not conforming to the future paradigm of green communications. In addition, these approaches only offer a very limited performance when the legitimate link and wiretap link are spatially highly correlated , which makes them only applicable in limited scenarios [15].

Recently, as one of the most promising technologies towards 6G networks, reconfigurable intelligent surface (RIS) has been demonstrated to improve wireless communication security not only in a more green and cost-effective manner, but also in more widely scenarios [13]. RIS is composed of a large number of passive reflective elements, each of which can reflect the incident signal with adjustable amplitude and/or phase, thus controlling the propagation environment. Since RIS does not need to decode or amplify the incident signal, it does not introduce thermal noise and consumes less energy than traditional relays, which is in line with our expectation for a higher energy efficiency in 6G networks. Moreover, RIS can provide comparable performance to beamforming using a massive antenna array at much lower cost [14]. Furthermore, RIS has demonstrated its benefits in enhancing PLS. By adjusting the phase of the incident signal, we can make the reflected signal and the direct signal superimposed at Eve to realize a signal cancellation, where the received signal power at the Eve can be reduced to achieve a certain degree of security [15].

A. Related Works

1) RIS enhanced OMA networks: At present, the application of RIS in PLS is widely studied in [16]–[30], [33], [34]. A typical example of RIS assisted PLS scheme was proposed in [16], where the channels of Eve and the legitimate user are highly correlated in space to demonstrate the ability of RIS to enhance security, and this scenario was further extended to multiple users and multiple Eves in [17]. Yu et al. [18] investigated small- and large-scale RIS for secrecy performance and two kinds of algorithms were proposed for different scales. In [19], the transmit power was minimized under the secrecy constraint with the help of an RIS to demonstrate the RIS ability to enhance security while being power efficient. In addition, RIS assisted PLS was introduced to mmWave and THz networks in [20], and Lu et al. [21] maximized secrecy rate with an imperfect channel state information (CSI) of Eve for mmWave communication. Whether AN can further enhance security for RIS-assisted network was studied in [22]. This was followed by additional studies that proposed to use AN/jamming to improve secrecy performance [23]–[30]. A PLS scenario was investigated in [23] where only an imperfect CSI of Eve is known at the base station (BS). Further, AN and RIS were adopted to provide secure transmission without Eve's CSI in [24]. Hong et al. [25] maximized secrecy rate for single user and multiple users MIMO scenarios. In [26], transmit power was minimized considering the effect of imperfect CSI on security. Similarly, energy efficiency was optimized with an imperfect Eve's CSI in [27], and a cooperative jammer was introduced to guarantee secrecy performance. A self-sustainable RIS was proposed in [28], and AN was also adopted to counter Eve's channel uncertainty. Not satisfied with the secrecy performance of single RIS, Li et al. [29] investigated a multiple RIS assisted PLS scheme with the help of AN. In [30], different from introducing AN directly, a two way secrecy communication scheme was considered where the signal of one user was used as jamming to disturb Eve. However, it should be noted that all RISs mentioned above are used to reflect the signal. Since RIS has the ability to act as a phase shifter, there is the possibility of using RIS for modulation [31], [32], and this can also offer performance gains for PLS. In particular, Xu et al. [33] designed an RIS assisted secure communication scheme to remodulate the jamming from an attacker into the desired signal to improve the signal to interference to noise ratio (SINR) at the user. On the contrary, the useful signal was converted into jamming via RIS to interfere with Eve in [34].

2) RIS enhanced NOMA networks: According to existing works, there is no doubt that RIS can offer a significant improvement to the secrecy performance of wireless communication via phase shifting, which is equivalent to indirectly increasing the channel condition gap between the legitimate and wiretap communication links. This ability is appealing to NOMA, since it can take advantage of this power difference to artificially introduce differences between channels of two paired users, which can decrease the interference during SIC under the NOMA power allocation strategy [35]. Therefore, RIS is expected to further improve the security of NOMA transmissions [36]–[44]. The PLS of an RIS assisted NOMA network was first investigated in [36], where AN was exploited and the transmit power was minimized. In [37], [38], internal and external Eves were considered with different CSI availabilities under RIS-NOMA, and power allocation was optimized to balance AN and the useful signal. Wang et al. [38] proposed a jamming based RIS PLS scheme for NOMA networks to resist the scenario where Eve's CSI was unknown, and the jamming was specially designed to be canceled at first by SIC at legitimate users. Furthermore, distributed RISs based secure scheme considering the statistical CSI of Eve was investigated for NOMA in [39]. Gong et al. [40] analyzed the secrecy outage probability for RIS-NOMA networks. In [41], an uplink RIS-NOMA scenario was considered and a max-min fair beamforming was optimized. The average secrecy capacity was analyzed in [42] for RIS-NOMA where the RIS only served the cell-edge user.

B. Motivations and Contributions

Based on the above discussion, it is important to point out that using RIS only either for information relaying [16]–[30],

[36]–[43] or jamming [33], [34] may not always guarantee secure NOMA communications, due to the following two reasons:

- On the one hand, utilizing RIS for information relaying can enhance the legitimate reception quality and realize signal cancellation at Eve at the same time. However, when the exact CSI of Eve is not known, i.e., the case of imperfect Eve's CSI, this approach may even benefit eavesdropping. To this end, it is necessary to let the transmitter insert AN for intentionally confusing Eve. However, this may be impossible for situations when the transmitter does not have sufficient degrees of freedom (e.g., active antennas or radio frequency chains) to insert AN.
- On the other hand, when the direct communication link between the transceivers cannot support reliable information reception (e.g., in deep fading or blockage), it is not wise to use RIS only for jamming, which will lead to a poor reception quality at receivers. Moreover, this effect becomes even worse under imperfect CSI of Eve, since in this case the RIS cannot align the jamming signal towards Eve.

Hence, if the RIS can be designed more flexibly to provide both information relaying and jamming services, secure NOMA communications will be better safeguarded, especially for the case of imperfect Eve CSI. However, to the best of the authors' knowledge, such a dual-functional RIS empowered security-enhanced design for NOMA systems has not been reported in the literature yet, which thus motivates this work.

In this paper, we investigate secure NOMA communications by fully unlocking the dual-functionality of RIS, i.e., information relaying and jamming. The transmit beamforming of the BS, reflect beamforming of the RIS, and mode selection matrix of the RIS are optimized jointly to improve the system security performance. The main contributions of this paper are summarized as follows.

- We make full advantage of the RIS and propose a new simultaneous information relaying and jamming (SIRJ) scheme based on a dual-functional RIS to secure NOMA communications. Specifically, the RIS elements are divided into two parts, where elements in one part perform signal reflection to improve the reception quality at the legitimate users, and elements in the remaining part remodulate the incident signal as useful jamming to degrade the performance of Eve. The working mode of each RIS element is adaptively selected to balance between information relaying and jamming for reliability and security.
- The fundamental performance limits of the proposed SIRJ scheme is first unveiled by assuming perfect CSI of Eve. To achieve secure and reliable NOMA communication, we aim to maximize the system sum-rate by jointly optimizing the active beamforming of the BS, the passive beamforming of the RIS, and the mode selection matrix of the RIS. Particularly, security is realized by considering the maximum tolerable information leakage to Eve. The optimization problem is difficult to handle since there are two phase-shift matrices under the SIRJ scheme, and both of them are highly coupled with each other and with the transmit beamforming. Moreover, the existence of a non-convex binary constraint makes the problem more challenging. To overcome these obstacles, we propose a penalty based alternating optimization (AO) algorithm, where the binary constraint is converted into a penalty term, and the rank-one constraints of active and passive beamforming are solved by semidefinite relaxation (SDR) and difference-of-convex (DC) technique, respectively.
- Furthermore, we consider a practical case with only imperfect CSI of Eve. In this scenario, the model for channel uncertainty of Eve is first described. However, the challenge here is that the secrecy constraint becomes an infinite number of non-convex constraints due to the channel estimation error, which makes the traditional $\mathcal{S}$-procedure method inapplicable to dealing with this constraint. To this end, we propose an efficient transformation based on matrix properties to tackle this challenge, based on which a robust secure beamforming scheme is designed to combat the imperfect CSI of Eve.
- We provide numerical results to demonstrate the performance advantage of the proposed SIRJ scheme and verify the efficiency of the developed optimization algorithms. Based on these results, we reveal several important insights: 1) The proposed SIRJ scheme achieves a higher sum-rate than the existing baseline schemes where the RIS is only used for either information relaying or jamming; 2) The SIRJ scheme will split more elements to generate jamming as the channel uncertainty level becomes larger, which again demonstrates the advantage of the SIRJ scheme in terms of the dynamic adjustment of RIS elements; 3) With the SIRJ scheme, increasing the number of RIS elements can achieve a significant secrecy enhancement than increasing the number of antennas at the BS, which represents a practically cost-effective approach to enhance NOMA security.

The rest of this paper is organized as follows. Section II introduces the system model and formulates the optimization problem. In Section III, we develop a penalty based AO algorithm to solve the problem with perfect CSI of Eve. A robust and secure transmission design is investigated in Section IV. In Section V, we provide numerical results and discussions. Finally, we conclude the paper in Section VI.

Notations: In this paper, boldface capital letters and lowercase letters stand for matrices and vectors, respectively. $\mathbb{C}^{N\times M}$ denotes the space of $N\times M$ matrices with complex entries, and $\mathbb{H}^{N\times N}$ denotes the set of $N\times N$ Hermitian matrices. $\mathbf{I}$ stands for the identity matrix. $(\cdot)^T$ and $(\cdot)^H$ represent transpose and conjugate transpose. $\mathcal{CN}(\mu,\sigma^2)$ stands for the distribution of a circularly symmetric complex Gaussian (CSCG) random variable with mean μ and variance σ^2. $\mathrm{Tr}(\cdot)$ and $\mathrm{rank}(\cdot)$ denote the trace and rank operations, respectively, and $\mathrm{vec}(\cdot)$ generates a column vector from sequentially stacking the columns of a matrix. $\mathbf{X}\succeq 0$ means $\mathbf{X}$ is a positive semidefinite matrix. The Frobenius norm of matrix and ℓ_2-norm of vector are represented as $\|\cdot\|_F$ and $\|\cdot\|$, respectively. $\mathrm{Diag}(\cdot)$

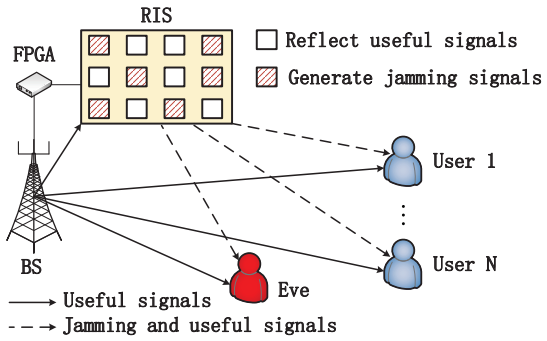


Fig. 1: The considered dual-functional RIS assisted NOMA system.

stands for a vector whose elements are extracted from the diagonal of a matrix, and $\mathrm{diag}(\cdot)$ denotes a diagonal matrix whose diagonal elements are extracted from a vector. $\mathrm{Re}(\cdot)$ represent the real part of a complex number, and $\otimes$ denotes the Kronecker product.

II. System Model

Consider an RIS assisted NOMA system shown as in Fig. 1, where a BS equipped with L antennas communicates with a NOMA user group consists of N single antenna users, and an illegal single antenna Eve tries to eavesdrop on this communication. The RIS with K elements is deployed to assist in the secure transmission. A controller, for example, a filed-programmable gate array (FPGA), is connected to the RIS to adjust the working mode and the phase shift of each element, and exchange information with the BS. Specifically, in this work, we unlock the full potential of the RIS to secure NOMA transmission by exploiting its dual-functionality of simultaneous signal reflection and jamming injection[1]. To be specific, the entire phase shift matrix $\mathbf{\Theta}$ at the RIS consists of two parts $\mathbf{\Theta} = \mathbf{\Theta}_r + \mathbf{\Theta}_J$, where $\mathbf{\Theta}_r = \mathrm{diag}(\mathbf{r})$ is the phase shift matrix used to reflect the useful signal, and $\mathbf{r} = [\sqrt{\beta_1^r}e^{i\phi_1^r}, \ldots, \sqrt{\beta_K^r}e^{i\phi_K^r}]^H$. In terms of the jamming mode elements, $\mathbf{\Theta}_J = \mathbf{\Theta}_j\mathbf{\Phi}_j$, where $\mathbf{\Theta}_j = \mathrm{diag}(\mathbf{j})$ is the phase shift matrix used to beamform the modulated jamming signal and $\mathbf{j} = [\sqrt{\beta_1^j}e^{i\phi_1^j}, \ldots, \sqrt{\beta_K^j}e^{i\phi_K^j}]^H$. Different from $\mathbf{\Theta}_r$ and $\mathbf{\Theta}_j$, $\mathbf{\Phi}_j = \mathrm{diag}(\mathbf{m})$ is the modulating matrix used to transform the incident useful signal to a jamming signal, and $\mathbf{m} = [e^{i\phi_1^m}, \ldots, e^{i\phi_K^m}]^H$. Here, the jamming modulation at the RIS aims to create as much randomness as pssible in the signal phase to intentionally confuse Eve, i.e., $e^{i\phi_k^m}$ is updated rapidly, independently and randomly during each symbol period, which is similar to making the signal experience a fast fading, and this makes the RIS modulating matrix fairly complicated [33], [34]. It can be seen that $\mathbf{\Phi_j}$ only influences the process of generation for jamming, while $\mathbf{\Theta_j}$ influences the direction of the jamming. Therefore, compared to $\mathbf{\Phi_j}$, $\mathbf{\Theta_j}$ is more directly related to system performance, which attracts us to concern more about the influence of the phase shift matrix $\mathbf{\Theta}_j$ on the overall secrecy performance, thus ignoring the design of modulation matrix $\mathbf{\Phi}_j$ [33], [34]. In terms of the amplitude responses of the k-th element, $\beta_k^r, \beta_k^j \in \{0, 1\}$, $\forall k \in \mathcal{K} \triangleq \{1, 2, \ldots, K\}$, i.e., each element either reflects signal or generates jamming, and satisfies $\beta_k^r + \beta_k^j = 1$ according to the law of energy conservation. Furthermore, due to the high path loss, the powers of signals reflected by the RIS two or more times are too weak, which can be practically neglected [14]. We assume that all communication links are quasi-static block fading channels, i.e., the coefficient of each channel changes independently under different fading blocks but stays constant during one fading block. Meanwhile, the instantaneous CSI of all legitimate channel links are assumed to be available at the BS, where the anchor-assisted channel estimation approach can be utilized [44]. For the CSI related to Eve, two cases are considered below.

- Eve is an active user but untrusted by the other NOMA users in the system, such that perfect CSI of Eve can be obtained by the BS [16], [22], [41];
- Eve prefers to hide its existence when it is an external wiretap node, and it is not expected to cooperate with the BS for CSI acquisition. Even if the BS can utilize power leakage from the local oscillator at Eve to estimate the CSI, this usually results in inaccurate and outdated CSI. Therefore, the BS can only possess the imperfect CSI of Eve [21], [23], [26], [28], [36].

The transmit signal at the BS can be expressed as $\mathbf{s} = \sum_{n=1}^{N} \mathbf{p}_n x_n$, where x_n is the signal of user n and $x_n \sim \mathcal{CN}(0, 1)$, $\mathbf{p}_n \in \mathbb{C}^{L\times 1}$ is the precoding vector for each user with $\sum_{n=1}^{N} ||\mathbf{p}_n||^2 \leq P_t$, and P_t is the transmit power budget at the BS. The received signal at user n is expressed as

$$y_n = (\mathbf{h}_{rn}^H \mathbf{\Theta}_r \mathbf{H}_{br} + \mathbf{h}_{bn}^H)\mathbf{s} + \mathbf{h}_{rn}^H \mathbf{z} + w_n, \tag{1}$$

where $\mathbf{h}_{rn} \in \mathbb{C}^{K\times 1}$ denotes the channel vector between user n and the RIS, $\mathbf{h}_{bn} \in \mathbb{C}^{L\times 1}$ denotes the channel vector between user n and the BS, $\mathbf{H}_{br} \in \mathbb{C}^{K\times L}$ stands for the channel matrix from the BS to the RIS, $\mathbf{z} = \mathbf{\Theta}_j \mathbf{H}_{br}\mathbf{s}$ is the reflected jamming signal vector, and $w_n \sim \mathcal{CN}(0, \sigma_0^2)$ is the additive white Gaussian noise (AWGN) at user n.

According to the NOMA protocol, the signal of user with weak channel gain is allocated with more power. Accordingly, SIC is performed at the receiver to process the superimposed signal, where the strong signal is decoded first by treating the weak signal as interference, and the weak signal is decoded after the substraction of the strong signal. In other words, the user with strong channel gain needs to decode the signal of user with weak channel gain. However, there is no doubt that the design of the phase shift matrix at the RIS as well as the precoding vectors at the BS can change the channel gain of each user, so this order may be changed due to the joint beamforming design of the BS and RIS. Without loss of generality and to facilitate the secure transmission

[1]It is intuitive to find out that whether dividing RIS into two at different locations can bring performance gain. The reason is as follows. Although dividing RIS into two can make the jamming and signal close to the Eve and legitimate users, respectively, it is still the conventional information relaying RIS or jamming RIS for each divided RIS. However, the proposed SIRJ scheme aims to adaptively adjust the number of elements working in different modes on one RIS to balance the power of generated jamming and useful signal. In other words, whether dividing RIS into two can bring performance gain depends on whether the number of elements of each divided RIS can be adaptively adjusted according to the security requirement and the channel state information of the Eve, but this could be difficult in practice.

design, we assume that the design of $\mathbf{\Theta}_r$ makes the channel gain satisfy the condition that $\left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{bn}^H)\mathbf{p}_m\right|^2 \geq \left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{bn}^H)\mathbf{p}_{m+1}\right|^2$, where $1\leq n\leq N$, $1\leq m\leq N-1$. [2] Under this case, user n needs to decode the signals from user 1 to user $n-1$ in this order. Based on the above decoding order, the SINR of user n for decoding user m's signal is

$$\text{SINR}_{nm}=\frac{\left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{bn}^H)\mathbf{p}_m\right|^2}{\sigma_0^2+\sum_{u=1}^{N}|\mathbf{h}_{rn}^H\mathbf{\Theta}_j\mathbf{H}_{br}\mathbf{p}_u|^2+\sum_{l=m+1}^{N}\left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{bn}^H)\mathbf{p}_l\right|^2}, \quad (2)$$

where $m\leq n$. If $m=n$, the result reduces to the SINR of user m for decoding its own signal. Hence, the achievable rate of user n for decoding the signal of user m is

$$R_{nm}=\log_2(1+\text{SINR}_{nm}), \quad (3)$$

and $1\leq m\leq n\leq N$. Note that for a successful SIC, the achievable rate of user n for decoding the signal of user m should be no less than the achievable rate of user m for decoding its own signal, i.e., $R_{nm}\geq R_{mm}$ for $m\leq n$ [35], [36], [39].

As for Eve, the received signal is expressed as

$$y_e=(\mathbf{h}_{re}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{be}^H)\mathbf{s}+\mathbf{h}_{re}^H\mathbf{z}+w_e, \quad (4)$$

where $\mathbf{h}_{be}\in\mathbb{C}^{L\times 1}$ denotes the channel vector between the BS and Eve, $\mathbf{h}_{re}\in\mathbb{C}^{K\times 1}$ denotes the channel vector between the RIS and Eve, and $w_e\sim\mathcal{CN}(0,\sigma_0^2)$ is the AWGN at Eve. Considering the worst case, where Eve has the ability to cancel the interference from SIC, which is a widely adopted assumption in the literature of secure NOMA systems [36], [39], [45]–[47], i.e., the eavesdropping SINR towards the signal of user m at Eve can be written as

$$\text{SINR}_{em}=\frac{\left|(\mathbf{h}_{re}^H\mathbf{\Theta}_r\mathbf{H}_{br}+\mathbf{h}_{be}^H)\mathbf{p}_m\right|^2}{\sigma_0^2+\sum_{u=1}^{N}\left|\mathbf{h}_{re}^H\mathbf{\Theta}_j\mathbf{H}_{br}\mathbf{p}_u\right|^2}. \quad (5)$$

Therefore, the achievable rate of Eve for wiretapping user m's information is

$$R_{em}=\log_2(1+\text{SINR}_{em}). \quad (6)$$

III. Beamforming Design With Perfect CSI of Eve

In this section, Eve is assumed to be an active user that is untrsuted in the system, hence the BS can access the complete CSI of Eve. Under this case, we first state the optimization problem, the problem reformulation process and corresponding algorithm structure are described in detail.

A. Problem Formulation

We aim to maximize the sum-rate of the legitimate users while ensuring that the information leakage to Eve is under an acceptable level, by jointly designing the active beamforming at the BS as well as the mode selection (e.g., for signal reflection or jamming generation) and the phase shift matrix at the RIS. The optimization problem is formulated as

$$\max_{\mathbf{p}_m,\mathbf{\Theta}_{r(j)},\beta_k^{r(j)}} \quad \sum_{m=1}^{N}R_{mm} \quad (7a)$$

$$s.t. \quad \sum_{m=1}^{N}\parallel\mathbf{p}_m\parallel^2\leq P_t, \quad (7b)$$

$$R_{nm}\geq R_{mm}, \quad 1\leq m\leq n\leq N, \quad (7c)$$

$$\left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_s+\mathbf{h}_{bn}^H)\mathbf{p}_m\right|^2\geq\left|(\mathbf{h}_{rn}^H\mathbf{\Theta}_r\mathbf{H}_s+\mathbf{h}_{bn}^H)\mathbf{p}_{m+1}\right|^2, \quad 1\leq n\leq N, \quad 1\leq m\leq N-1, \quad (7d)$$

$$\beta_k^r+\beta_k^j=1, \quad \forall k\in\mathcal{K}, \quad (7e)$$

$$\beta_k^t\in\{0,1\}, \quad \forall k\in\mathcal{K}, \quad t\in\{r,j\}, \quad (7f)$$

$$\max R_{em}\leq\tau, \quad 1\leq m\leq N. \quad (7g)$$

Specifically, (7b) is the maximum transmit power constraint at the BS. Constraint (7c) guarantees that a successful SIC can be performed at each user, and (7d) constrains the decoding order of legitimate users. (7e) ensures the law of energy conservation at the RIS, and constraint (7f) gives the range of reflect amplitude of each element. (7g) guarantees the network physical layer security, where τ is a predefined parameter that presents the maximum tolerable information leakage to Eve, such that the system secrecy rate can be bounded from below by $R_s\geq\sum_{m=1}^{N}[R_{mm}-\tau]^+$ [23], [28].

Remark 1: Recall that the primary purpose of communication is to complete the information transmission with legitimate users, the proposed formulation aims to maximize the user sum-rate while satisfying a certain information leakage constraint, rather than being overly constrained by the secrecy performance of the system. In particular, the optimization problem in (7) can make the system control its expected secrecy performance via the adjustment of τ. Therefore, compared to maximizing the secrecy rate directly, the considered optimization objective can provide a higher flexibility for resource allocation, especially for different applications with heterogeneous secrecy requirements [23], [28].

However, problem (7) is intractable since not only the conventional variables $\mathbf{\Theta}_r$ and $\mathbf{p}_m$ are coupled, but also $\mathbf{\Theta}_j$ makes this coupling even tighter, per constraints (7c), (7g) and the objective function. Meanwhile, the binary constraint (7f) makes the problem a mixed-integer non-convex problem, which makes the problem even more difficult to solve. To overcome these challenges, we split the original problem into two subproblems, where variables are the active transmit beamforming at the BS and the passive reflect beamforming at the RIS, respectively. To deal with the intractable binary constraint and rank-one constraint in the reformulated subproblems, we propose a penalty based AO algorithm, where $\mathbf{p}_m$ and $\mathbf{\Theta}_r,\mathbf{\Theta}_j$ are solved alternately by fixing the other variables while punishing the solution that cannot satisfy the binary constraint and the rank-one constraint, which is described next.

[2]The fixed decoding SIC order does not have a significant impact on system performance. This is because the locations of legitimate users determine the large scale path loss, which further determine their channel gains. In addition, the baselines for comparison in simulation all adopt the same decoding order as the proposed scheme, so even though the decoding order has a certain degree of impact on system performance, the performance gain of the SIRJ scheme can still be demonstrated via comparing with these baselines under a same condition.

B. Active Transmit Beamforming

Given $\mathbf{\Theta}_r$ and $\mathbf{\Theta}_j$, the optimization problem of active beamforming is presented as

$$\max_{\mathbf{p}_m} \quad \sum_{m=1}^{N} R_{mm} \tag{8a}$$

$$s.t. \quad \text{(7b)}, \text{(7c)}, \text{(7d)}, \text{(7g)}, \tag{8b}$$

$$\mathbf{P}_m \succeq 0, \tag{8c}$$

$$\text{rank}(\mathbf{P}_m) = 1, \tag{8d}$$

where $\mathbf{P}_m = \mathbf{p}_m \mathbf{p}_m^H$. Then, constraint (7b) becomes

$$\sum_{m=1}^{N} \text{Tr}(\mathbf{P}_m) \leq P_t. \tag{9}$$

Next, we define $\mathbf{R} = \boldsymbol{\theta}_r \boldsymbol{\theta}_r^H$ and $\boldsymbol{\theta}_r = [\mathbf{r}; 1]$, where $\mathbf{R} \succeq 0$ and $\text{rank}(\mathbf{R}) = 1$, and $\text{Diag}(\mathbf{R}) = \boldsymbol{\beta}_r$, where $\boldsymbol{\beta}_r = [\beta_1^r, \ldots, \beta_K^r, 1]$. Further, let $\mathbf{H}_n = [\text{diag}(\mathbf{h}_{rn}^H)\mathbf{H}_{br}; \mathbf{h}_{bn}^H]$. Hence, we can rewrite constraint (7d) as

$$\text{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R}) \leq \text{Tr}(\mathbf{H}_n \mathbf{P}_{m+1} \mathbf{H}_n^H \mathbf{R}), \tag{10}$$

where $1 \leq n \leq N$ and $1 \leq m \leq N-1$. Similarly, we define $\mathbf{J} = \boldsymbol{\theta}_j \boldsymbol{\theta}_j^H$ and $\boldsymbol{\theta}_j = [\mathbf{j}; 0]$ where $\mathbf{J} \succeq 0$ and $\text{rank}(\mathbf{J}) = 1$, and $\text{Diag}(\mathbf{J}) = \boldsymbol{\beta}_j$, where $\boldsymbol{\beta}_j = [\beta_1^j, \ldots, \beta_K^j, 0]$. SINR_{nm} can be transformed into

$$\text{SINR}_{nm} = \frac{\text{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R})}{\sigma_0^2 + \sum_{u=1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_u \mathbf{H}_n^H \mathbf{J}) + \sum_{l=m+1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_l \mathbf{H}_n^H \mathbf{R})}, \tag{11}$$

where $1 \leq m \leq n \leq N$. Considering the monotonically increasing feature of the logarithmic function, constraint (7c) is equivalent to

$$\text{SINR}_{nm} \geq \text{SINR}_{mm}, \quad 1 \leq m \leq n \leq N. \tag{12}$$

Next, we introduce a slack variable r_{nm} which satisfies $r_{nm} \leq \text{SINR}_{nm}$ with $1 \leq m \leq n \leq N$, i.e.,

$$r_{nm} \leq \frac{\text{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R})}{\sigma_0^2 + \sum_{u=1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_u \mathbf{H}_n^H \mathbf{J}) + \sum_{l=m+1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_l \mathbf{H}_n^H \mathbf{R})}, \tag{13}$$

To solve the coupled fractional form, we adopt the arithmetic-geometric mean (AGM) inequality to transform it into

$$2\text{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R}) \geq \left(\mu_{nm} \Gamma\right)^2 + \left(\frac{r_{nm}}{\mu_{nm}}\right)^2, \tag{14}$$

where $\Gamma = \sigma_0^2 + \sum_{u=1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_u \mathbf{H}_n^H \mathbf{J}) + \sum_{l=m+1}^{N} \text{Tr}(\mathbf{H}_n \mathbf{P}_l \mathbf{H}_n^H \mathbf{R})$. The equality holds if and only if $\mu_{nm} = \sqrt{\frac{r_{nm}}{\Gamma}}$.

Lemma 1: When the objective function achieves the optimum, we have $R_{mm} = \log_2(1 + r_{mm})$.

Proof: According to (13), it is not hard to see that $r_{mm} \leq r_{mm}^{[\max]} = \text{SINR}_{mm}$. While from the objective function, we can obtain the following lower bound on the user sum-rate, $\sum_{m=1}^{N} R_{mm} \geq \sum_{m=1}^{N} \log_2(1 + r_{mm})$, which is monotonically increasing with r_{mm}. Hence, the objective function reaches the optimum when $r_{mm} = r_{mm}^{[\max]} = \text{SINR}_{mm}$. ∎

Based on *Lemma 1*, constraint (7c) becomes

$$r_{nm} \geq r_{mm}, \quad 1 \leq m \leq n \leq N. \tag{15}$$

By defining $\mathbf{G}_e = \text{diag}(\mathbf{h}_{re}^H)\mathbf{H}_{br}$ and $\mathbf{H}_e = [\mathbf{G}_e; \mathbf{h}_{be}^H]$, constraint (7g) becomes

$$\text{Tr}(\mathbf{H}_e \mathbf{P}_m \mathbf{H}_e^H \mathbf{R}) \leq T_e \sigma_0^2 + T_e \sum_{u=1}^{N} \text{Tr}(\mathbf{H}_e \mathbf{P}_u \mathbf{H}_e^H \mathbf{J}), \tag{16}$$

where $T_e = 2^\tau - 1$ and $1 \leq m \leq N$.

To tackle the non-convexity of the objective function, we first employ the following fundamental property of the logarithmic function

$$\sum_{m=1}^{N} R_{mm} = \log_2\big(\prod_{m=1}^{N} (1 + r_{mm})\big). \tag{17}$$

Next, by using the monotonically increasing property of the logarithmic function and introducing another auxiliary variable s, we have

$$\big(\prod_{m=1}^{N} (1 + r_{mm})\big)^{\frac{1}{N}} \geq s, \tag{18}$$

where the left side of above equation is in the form of geometric mean. Next, we turn our attention to (8d). With the help of SDR, we can eliminate the rank-one constraint (8d), and problem (8) is finally transformed into the following form that can be solved by CVX toolbox,

$$\max_{s, \mathbf{P}_m, r_{nm}} \quad s \tag{19a}$$

$$s.t. \quad \text{(8c)}, (9), (10), (14), (15), (16), (18). \tag{19b}$$

Lemma 2: The obtained SDR solution always satisfies the rank-one condition $\text{rank}(\mathbf{P}_m) = 1$ and hence is optimal.

Proof: Please refer to Appendix A. ∎

C. Passive Reflect Beamforming

With fixed $\mathbf{P}_m$ and following similar transformation steps as above, the subproblem of passive reflect beamforming at the RIS is reduced to

$$\max_{s, \mathbf{X}, \boldsymbol{\beta}_t, r_{nm}} \quad s \tag{20a}$$

$$s.t. \quad (10), (14), (15), (16), (18),$$

$$\mathbf{X} \succeq 0, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{J}\}, \tag{20b}$$

$$\text{Diag}(\mathbf{X}) = \boldsymbol{\beta}_t, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{J}\}, t \in \{r, j\}, \tag{20c}$$

$$\text{rank}(\mathbf{X}) = 1, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{J}\} \tag{20d}$$

$$\beta_k^r + \beta_k^j = 1, \quad \forall k \in \mathcal{K}, \tag{20e}$$

$$\beta_k^t \in \{0, 1\}, \quad \forall k \in \mathcal{K}, t \in \{r, j\}. \tag{20f}$$

The remaining challenges of problem (20) lie mainly in constraints (20d) and (20f), where the rank-one constraint (20d) cannot be solved by applying SDR again similar to the active beamforming design, since the rank of the solution for problem (20) is generally large than one because of the mode selection constraint and the simultaneous optimization of $\mathbf{\Theta}_r$ and $\mathbf{\Theta}_j$. Meanwhile, according to the proof for the tightness of the SDR for active beamforming, it relies on the rank-one of the passive beamforming matrix. Furthermore, the binary constraint (20f) is also hard to handle directly. Thus, some processing needs to be done to overcome these obstacles, which is detailed next.

In terms of the binary constraint (20f), a more normal condition is considered, i.e., $\beta_k^t \in [0,1]$, which can be written as

$$(\beta_k^t)^2 - \beta_k^t \leq 0, \quad \forall k \in \mathcal{K}, t \in \{r,j\}. \tag{21}$$

It is clear that the equality holds if and only if $\beta_k^t = 0$ or 1. In other words, we need to punish β_k^t when it belongs to $(0,1)$. Thus, we add constraint (21) as a penalty term into the objective function. Also, the first-order Taylor expansion is adopted to approximate this penalty term into a convex form as follows

$$\begin{aligned}\beta_k^t - (\beta_k^t)^2 &\leq \beta_k^t - (\beta_k^{t(i)})^2 - 2\beta_k^{t(i)}(\beta_k^t - \beta_k^{t(i)}) \\ &\triangleq \mathcal{P}(\beta_k^t, \beta_k^{t(i)}), \quad \forall k \in \mathcal{K}, t \in \{r,j\},\end{aligned} \tag{22}$$

where $\beta_k^{t(i)}$ is the i-th iteration result. Now, problem (20) can be reformulated as

$$\max_{s,\mathbf{X},\boldsymbol{\beta}_t,r_{nm}} \; s - \rho \sum_{t\in\{r,j\}} \sum_{k=1}^{K} \mathcal{P}(\beta_k^t, \beta_k^{t(i)}) \tag{23a}$$

$$s.t. \quad (10),(14),(15),(16),(18),(20b),(20c),(20d),(20e), \tag{23b}$$

where $\rho > 0$ is the penalty factor that starts with a small value to obtain a feasible initial point, and gradually increases to make the solution satisfy the binary constraint tightly. However, problem (23a) is still a non-convex problem due to the rank-one constraint (20d). It is known that for any square matrix, we have

$$\|\mathbf{X}\|_* - \|\mathbf{X}\|_2 \leq 0, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{J}\}, \tag{24}$$

where the equality holds if and only if $\mathrm{rank}(\mathbf{X}) = 1$. Note that $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$ to be the nuclear norm, and $\|\mathbf{X}\|_2 = \sigma_1(\mathbf{X})$ is the spectral norm, where σ_i denotes the i-th largest singular value of $\mathbf{X}$. Recall that the eigenvalue equals singular value for a positive semi-definite square matrix, which suggests that $\|\mathbf{X}\|_* = \mathrm{Tr}(\mathbf{X})$ since $\mathrm{Tr}(\mathbf{X}) = \sum_i \lambda_i(\mathbf{X})$ where λ_i is the i-th largest eigenvalue. Meanwhile, we use the first-order Taylor expansion to approximate $\|\mathbf{X}\|_2$. Then, (24) can be further transformed into

$$\mathrm{Tr}(\mathbf{X}) - \mathcal{T}(\mathbf{X}, \mathbf{X}^{(i)}) \leq 0, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{J}\}, \tag{25}$$

where $\mathcal{T}(\mathbf{X}, \mathbf{X}^{(i)}) = \|\mathbf{X}^{(i)}\|_2 + \mathrm{Tr}[\lambda_1(\mathbf{X}^{(i)})\lambda_1(\mathbf{X}^{(i)})^H(\mathbf{X} - \mathbf{X}^{(i)})]$, and $\mathbf{X}^{(i)}$ stands for the i-th iteration result. Finally, we need to punish $\mathbf{X}$ if $\mathrm{Tr}(\mathbf{X}) - \mathcal{T}(\mathbf{X}, \mathbf{X}^{(i)}) > 0$, i.e., the optimization problem is recast as

$$\begin{aligned}\max_{s,\mathbf{X},\boldsymbol{\beta}_t,r_{nm}} \; & s - \rho \sum_{t\in\{r,j\}} \sum_{k=1}^{K} \mathcal{P}(\beta_k^t, \beta_k^{t(i)}) \\ & - \varrho \sum_{\mathbf{X}\in\{\mathbf{R},\mathbf{J}\}} \mathcal{R}(\mathbf{X}, \mathbf{X}^{(i)})\end{aligned} \tag{26a}$$

$$s.t. \quad (10),(14),(15),(16),(18),(20b),(20c),(20e), \tag{26b}$$

where $\mathcal{R}(\mathbf{X}, \mathbf{X}^{(n)}) = \mathrm{Tr}(\mathbf{X}) - \mathcal{T}(\mathbf{X}, \mathbf{X}^{(n)})$, and ϱ is the corresponding penalty factor. Now, problem (26) is a convex one which can be solved by CVX efficiently.

The overall algorithm is summarized in **Algorithm 1**. It should be noted that, different from solving problem (19), the structure of solving problem (26) contains two layers of loops, where the inner loop ensures the convergence of the objective function, and the outer loop ensures that the obtained solution is sufficient to satisfy the binary constraint and the rank-one constraint via thresholds ε_1 and ε_2.

Algorithm 1 Proposed penalty based AO algorithm

1: **Initialize** $\mu_{nm}^{(i)}$, $\boldsymbol{\beta}_r^{(i)}$, $\boldsymbol{\beta}_j^{(i)}$, $\mathbf{R}^{(i)}$, $\mathbf{J}^{(i)}$, ρ, ϱ, and set $i = 1$;
2: **Repeat**
3: **Repeat**
4: Solve problem (19);
5: Update $\mu_{nm}^{(i+1)} = \sqrt{\frac{r_{nm}}{\Gamma}}$;
6: **Until** $|s^{(i+1)} - s^{(i)}| \leq \varepsilon$, output s^* and $\mathbf{P}_m^*$;
7: Calculate $R_{act} = \log_2\big((s^*)^N\big)$;
8: **Repeat**
9: Set $i = 1$;
10: **Repeat**
11: Solve problem (26);
12: Update $\mu_{nm}^{(i+1)}$;
13: **Until** $|s^{(i+1)} - s^{(i)}| \leq \varepsilon$, output $\mathbf{R}^{(i+1)}$, $\mathbf{J}^{(i+1)}$, $\boldsymbol{\beta}_r^{(i+1)}$ and $\boldsymbol{\beta}_j^{(i+1)}$;
14: Update $\mathbf{R}^{(1)}$, $\mathbf{J}^{(1)}$, $\boldsymbol{\beta}_r^{(1)}$ and $\boldsymbol{\beta}_j^{(1)}$ with the output solutions, and $\rho = \omega\rho$, $\varrho = \omega\varrho$;
15: **Until** $\|\mathbf{X}\|_* - \|\mathbf{X}\|_2 < \varepsilon_1$, and $\beta_k^t - (\beta_k^t)^2 < \varepsilon_2$, output s^*, $\mathbf{R}^*$ and $\mathbf{J}^*$;
16: Calculate $R_{pas} = \log_2\big((s^*)^N\big)$;
17: **Until** $|R_{pas} - R_{act}| \leq \varepsilon$;

D. Convergence and Complexity Analysis

Since the objective value is obtained by solving problems (19) and (26) alternately, it can be denoted as a function of the active beamforming and passive beamforming, i.e., $f(\mathbf{P}_m, \mathbf{R}, \mathbf{J})$. The steps 3–6 of **Algorithm 1** ensure that $f(\mathbf{P}_m^{(i)}, \mathbf{R}^{(i-1)}, \mathbf{J}^{(i-1)}) \geq f(\mathbf{P}_m^{(i-1)}, \mathbf{R}^{(i-1)}, \mathbf{J}^{(i-1)})$ in the i-th iteration, and the steps 8–15 lead to $f(\mathbf{P}_m^{(i)}, \mathbf{R}^{(i)}, \mathbf{J}^{(i)}) \geq f(\mathbf{P}_m^{(i)}, \mathbf{R}^{(i-1)}, \mathbf{J}^{(i-1)})$. Therefore, we have $f(\mathbf{P}_m^{(i)}, \mathbf{R}^{(i)}, \mathbf{J}^{(i)}) \geq f(\mathbf{P}_m^{(i-1)}, \mathbf{R}^{(i-1)}, \mathbf{J}^{(i-1)})$, which indicates that the objective value obtained by **Algorithm 1** can keep non-decreasing. Meanwhile, $f(\mathbf{P}_m, \mathbf{R}, \mathbf{J})$ is continuous over the compact feasible set of problem (26), thus the upper bound on the objective value is limited by a finite positive number [40], which guarantees the convergence of the proposed algorithm.

Since problems (19) and (26) contain only LMI, linear constraints and second-order cone (SOC) constraints, they can be solved by a standard interior point method, and the general expression of complexity is given by

$$\mathcal{O}\Big((\textstyle\sum_{q=1}^{Q} c_q + 2V)^{1/2} p\Big(p^2 + p\sum_{q=1}^{Q} c_q^2 + \sum_{q=1}^{Q} c_q^3 + p\sum_{v=1}^{V} d_v^2\Big)\Big), \tag{27}$$

where p is the number of variables, Q is the number of LMIs with size c_q, and V is the number of SOC with size d_v [48]. Hence, the complexity of **Algorithm 1** can be calculated as $\mathcal{O}\Big(i_{AO}\big(i_{act}\sqrt{A_1 + 2V}\,p_1(p_1^2 + p_1B_1 + p_1C_1 + D) + i_{pas}^{out} i_{pas}^{in}\sqrt{A_2 + 2V}\,p_2(p_2^2 + p_2B_1 + p_2C_2 + D)\big)\Big)$, where i_{AO}

denotes the number of iterations for AO, and i_{act} denotes the number of iterations for solving problem (19), while i_{pas}^{in} and i_{pas}^{out} denote the number of inner and outer iterations for solving problem (26). To be specific, $p_1 = NL^2 + 0.5N(N+1) + 1$ and $p_2 = 2(M^2+M)+0.5N(N+1)+1$, $V = 0.5N(N+1)$, $A_1 = 0.5N(N-1) + NL + 2N + 1$ and $A_2 = 0.5N(N-1) + 2N + 5M + 2$, $B_1 = NL^2 + 0.5N(N-1) + 2N + 1$ and $B_2 = 2(M+1)^2 + 0.5N(N-1) + 2N + 3M$, $C_1 = NL^3 + 0.5N(N-1) + 2N + 1$ and $C_2 = 2(M+1)^3 + 0.5N(N-1) + 2N + 3M$, and $D = 0.5N(N+1)$. For more concise, we retain the highest order term of each variable, and the complexity of the proposed algorithm comes to

$$\mathcal{O}\Big(i_{\text{AO}}\big(i_{\text{act}}(L^6N^4K^2 + N^8) + i_{\text{out}}^{\text{pas}} i_{\text{in}}^{\text{pas}}(K^6N + N^7 + K^5L^3N^2)\big)\Big). \quad (28)$$

Remark 2: The proposed scheme can be extended to the multi-eavesdropper case easily. When there are multiple eavesdroppers in the system, each eavesdropper has a eavesdropping rate for each legitimate user's signal. However, the secrecy constraint can still ensures that the beamforming design satisfies the secrecy requirement of the system. It can be seen from the transformation result (16) of the secrecy constraint, multiple eavesdroppers means increased number of secrecy constraints. Although this will significantly reduce the feasible domain of the original optimization problem, the proposed algorithm can still be applied in this scenario.

E. Extension to Discrete Phase

It is difficult to continuously adjust the phase of RIS in practice. In order to investigate the performance of discrete phase shifts, a uniform quantized phase-shift feasible region is required. In particular,

$$\phi_{k,\text{dis}}^{r(j)} \in \mathcal{D} = \Big\{\frac{a(2\pi)}{2^b}, a = 0, 1, 2, \cdots, (2^b - 1)\Big\}, \quad (29)$$

where $k \in \mathcal{K} = \{1, 2, \cdots, K\}$, and b stands for the number of quantization bits. Based on this, the discrete phase shift solution can be obtained from the continuous phase shift solution of **Algorithm 1**. Specifically,

$$\phi_{k,\text{dis}}^{r(j),\text{opt}} = \arg \min_{\phi_{k,\text{dis}}^{r(j)} \in \mathcal{D}} |\phi_{k,d}^{r(j)} - \phi_k^{r(j),\text{opt}}|, \quad (30)$$

where $\phi_{k,\text{dis}}^{r(j),\text{opt}}$ is the discrete phase shift solution, and $\phi_k^{r(j),\text{opt}}$ is the continuous phase shift solution obtained from **Algorithm 1**.

IV. Robust Beamforming Design With Imperfect CSI of Eve

In this section, Eve is assumed to be a passive external eavesdropper that prefers to hide its existence. Thus, the BS can only obtain its partial CSI via, for example, the power leakage at Eve. Under this scenario, a robust beaforming design is necessary to counter the channel uncertainty of Eve. Hence, this uncertainty is modeled at first, and then, we focus on dealing with secrecy constraint which translates to an infinite number of non-convex constraints due to the channel uncertainty of Eve. Note that since the RIS elements work in different modes, which makes the conventional $\mathcal{S}$-procedure not directly applicable. This poses another significant challenge for robust beamforming designs. Thus, an effective transformation is proposed to deal with this challenging problem, which is detailed in the following.

A. Channel Uncertainty Modeling

For modeling Eve's channel uncertainty, we adopt the norm-bounded CSI error model, where the CSI error is bounded in a region defined by norm, and this model has been widely used in the literature, e.g., [23], [26]–[28], [36], [48]. In particular, the CSI of the direct link between the BS and Eve is modeled as

$$\begin{aligned} &\mathbf{h}_{be} = \hat{\mathbf{h}}_{be} + \Delta\mathbf{h}_{be}, \\ &\Delta\mathbf{h}_{be} \in \Omega_{be} \triangleq \{\Delta\mathbf{h}_{be} \in \mathbb{C}^{L\times 1} : \|\Delta\mathbf{h}_{be}\| \le \delta_{be}\}, \end{aligned} \quad (31)$$

where $\hat{\mathbf{h}}_{be}$ is the estimated CSI of the Eve, and $\Delta\mathbf{h}_{be}$ stands for the CSI estimation error lying in a continuous set Ω_{be}, which is norm-bounded by $\delta_{be} \ge 0$, involving all possible CSI estimation errors, and the parameter δ_{be} represents the level of channel uncertainty. Similarly, we can model the CSI of the cascaded RIS link between the BS and Eve as

$$\begin{aligned} &\mathbf{G}_e = \hat{\mathbf{G}}_e + \Delta\mathbf{G}_e, \\ &\Delta\mathbf{G}_e \in \Omega_{re} \triangleq \{\Delta\mathbf{G}_e \in \mathbb{C}^{K\times L} : \|\Delta\mathbf{G}_e\|_F \le \delta_{re}\}, \end{aligned} \quad (32)$$

Based on (31) and (32), we can obtain the uncertainty model of the equivalent channel $\mathbf{H}_e$ between the BS and Eve, i.e.,

$$\begin{aligned} &\mathbf{H}_e = \begin{bmatrix} \mathbf{G}_e \\ \mathbf{h}_{be}^H \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{G}}_e \\ \hat{\mathbf{h}}_{be}^H \end{bmatrix} + \begin{bmatrix} \Delta\mathbf{G}_e^H \\ \Delta\mathbf{h}_{be}^H \end{bmatrix} = \hat{\mathbf{H}}_e + \Delta\mathbf{H}_e, \\ &\Delta\mathbf{H}_e \in \Omega_e \triangleq \{\Delta\mathbf{H}_e \in \mathbb{C}^{(K+1)\times L} : \|\Delta\mathbf{H}_e\|_F \le \delta_e\}, \end{aligned} \quad (33)$$

where the estimation error and uncertainty level are given by $\|\Delta\mathbf{H}_e\|_F = \sqrt{\|\Delta\mathbf{G}_e\|_F^2 + \|\Delta\mathbf{h}_{be}\|^2}$ and $\delta_e = \sqrt{\delta_{re}^2 + \delta_{be}^2}$, respectively.

B. Robust Beamforming Design

Under this scenario, we follow similar steps as in Section III for the problem reformulation and algorithm structure. Specifically, we decompose the original problem into two subproblems, and optimize active transmit beamforming and passive reflect beamforming alternately, while converting the binary constraint and the rank-one constraint into penalty terms. However, the difference is the secrecy constraint (7g) in the original problem. Particularly, the robust active transmit beamforming subproblem is formulated as

$$\max_{s,\mathbf{P}_m,r_{nm}} \quad s \quad (34a)$$

$$s.t. \quad (8c), (9), (10), (14), (15), (18), \quad (34b)$$

$$\max_{\Delta\mathbf{H}_e\in\Omega_e} R_{em} \le \tau, \quad 1 \le m \le N. \quad (34c)$$

According to the channel uncertainty model, it is not difficult to see that (34c) contains infinite constraints. To handle this challenge, the $\mathcal{S}$-procedure is a commonly used and effective method [23], [27], [28], [36], [48]. To apply this powerful method to our problem, some mathematical processing needs to be performed, as described next.

First, we define $\mathbf{h}_e = \text{vec}(\mathbf{H}_e^H)$. Then, we exploit the following two matrix identities, i.e., $\text{Tr}(\mathbf{A}^H\mathbf{B}) = \text{vec}^H(\mathbf{A})\text{vec}(\mathbf{B})$ and $\text{vec}(\mathbf{ACB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{C})$. Finally, rearrange constraint (34c) as follows

$$\mathbf{h}_e^H\Big((\mathbf{R}^T \otimes \mathbf{P}_m) - T_e\big(\mathbf{J}^T \otimes \sum_{u=1}^{N}\mathbf{P}_u\big)\Big)\mathbf{h}_e - T_e\sigma_0^2 \le 0, \quad (35)$$

where $1 \le m \le N$, and $\mathbf{h}_e = \hat{\mathbf{h}}_e + \Delta\mathbf{h}_e$. Considering the calculation of the Frobenius norm and the ℓ_2-norm, $\Delta\mathbf{h}_e$ and $\Delta\mathbf{H}_e$ share the same uncertainty level that $\|\Delta\mathbf{h}_e\| \le \delta_e$. More precisely, we rewrite the above inequality as

$$\Delta\mathbf{h}_e^H\mathbf{\Phi}_m\Delta\mathbf{h}_e + 2\text{Re}\{\hat{\mathbf{h}}_e^H\mathbf{\Phi}_m\Delta\mathbf{h}_e\} + \hat{\mathbf{h}}_e^H\mathbf{\Phi}_m\hat{\mathbf{h}}_e - T_e\sigma_0^2 \le 0, \quad (36)$$

where $\mathbf{\Phi}_m = (\mathbf{R}^T \otimes \mathbf{P}_m) - T_e(\mathbf{J}^T \otimes \sum_{u=1}^{N}\mathbf{P}_u)$, $1 \le m \le N$.

Lemma 3: ($\mathcal{S}$-procedure [49]): Let $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{H}^{N\times N}$, set $\mathbf{a}_1$, $\mathbf{a}_2 \in \mathbb{C}^{N\times 1}$, and b_1, $b_2 \in \mathbb{R}$. Consider the following two functions with $\mathbf{x} \in \mathbb{C}^{N\times 1}$,

$$f_1(\mathbf{x}) = \mathbf{x}^H\mathbf{A}_1\mathbf{x} + 2\text{Re}\{\mathbf{a}_1^H\mathbf{x}\} + b_1, \quad (37)$$

$$f_2(\mathbf{x}) = \mathbf{x}^H\mathbf{A}_2\mathbf{x} + 2\text{Re}\{\mathbf{a}_2^H\mathbf{x}\} + b_2. \quad (38)$$

The implication $f_1(\mathbf{x}) \le 0 \Rightarrow f_2(\mathbf{x}) \le 0$ holds if and only if there exists a nonnegative γ that satisfies

$$\begin{bmatrix} \mathbf{A}_2 & \mathbf{a}_2 \\ \mathbf{a}_2^H & b_2 \end{bmatrix} \le \gamma \begin{bmatrix} \mathbf{A}_1 & \mathbf{a}_1 \\ \mathbf{a}_1^H & b_1 \end{bmatrix}, \quad (39)$$

where there exists $\hat{\mathbf{x}}$ such that $f_1(\hat{\mathbf{x}}) \le 0$.

With the help of Lemma 3, the secrecy constraint (36) can be transformed into

$$\begin{bmatrix} \mathbf{\Phi}_m - \gamma_m\mathbf{I} & \mathbf{\Phi}_m\hat{\mathbf{h}}_e \\ \hat{\mathbf{h}}_e^H\mathbf{\Phi}_m & \hat{\mathbf{h}}_e^H\mathbf{\Phi}_m\hat{\mathbf{h}}_e - T_e\sigma_0^2 + \gamma_m\delta_e^2 \end{bmatrix} \le 0, \quad 1 \le m \le N. \quad (40)$$

Therefore, the robust active transmit beamforming subproblem is finally reduced to

$$\max_{s,\mathbf{P}_m,r_{nm},\gamma_m} \quad s \quad (41a)$$

$$s.t. \quad (8c), (9), (10), (14), (15), (18), (40) \quad (41b)$$

By applying the same rationale, the robust passive reflect beamforming subproblem can be formulated as

$$\max_{s,\mathbf{X},\boldsymbol{\beta}_t,r_{nm},\gamma_m} \; s - \rho\sum_{t\in\{r,j\}}\sum_{k=1}^{K}\mathcal{P}(\beta_k^t, \beta_k^{t(i)}) \quad (42a)$$

$$- \varrho\sum_{\mathbf{X}\in\{\mathbf{R},\mathbf{J}\}}\mathcal{R}(\mathbf{X}, \mathbf{X}^{(i)}) \quad (42b)$$

$$s.t. \quad (10), (14), (15), (16), (20b), (20c), (20e), (40). \quad (42c)$$

Now, **Algorithm 1** proposed in Section III can be adopted to solve problem (41) and problem (42) alternately for the robust beamforming design.

V. Numerical Results

In this section, we provide numerical results to demonstrate the secrecy performance enhancement of our proposed SIRJ scheme as well as the joint beamforming design. For channel model, we assume Rician fading for the channels related to the RIS, given by

$$\mathbf{H}_{br} = \sqrt{\frac{\lambda_0}{d_{br}^{\alpha_{br}}}}\Big(\sqrt{\frac{\mathcal{K}_{br}}{\mathcal{K}_{br}+1}}\mathbf{H}_{br}^{\text{LoS}} + \sqrt{\frac{1}{\mathcal{K}_{br}+1}}\mathbf{H}_{br}^{\text{NLoS}}\Big), \quad (43)$$

$$\mathbf{h}_{ri} = \sqrt{\frac{\lambda_0}{d_{ri}^{\alpha_{ri}}}}\Big(\sqrt{\frac{\mathcal{K}_{ri}}{\mathcal{K}_{ri}+1}}\mathbf{h}_{ri}^{\text{LoS}} + \sqrt{\frac{1}{\mathcal{K}_{ri}+1}}\mathbf{h}_{ri}^{\text{NLoS}}\Big), \quad (44)$$

where $i \in \{e, n\}$, λ_0 denotes the path loss at the reference distance of 1 meter, d_{br}, d_{re}, and d_{rn} denote the link distances, α_{br}, α_{re}, and α_{rn} stand for the pass loss exponents, and $\mathcal{K}_{br}$, $\mathcal{K}_{re}$, and $\mathcal{K}_{rn}$ represent the Rician factors. $(\cdot)^{\text{LoS}}$ denotes the deterministic line-of-sight (LoS) component and $(\cdot)^{\text{NLoS}}$ denotes the non-line-of-sight (NLoS) component. For the direct link between the BS and user n/Eve, Rayleigh fading is assumed such that

$$\mathbf{h}_{bi} = \sqrt{\frac{\lambda_0}{d_{bi}^{\alpha_{bi}}}}\mathbf{h}_{bi}^{\text{NLoS}}, \quad i \in \{e, n\}, \quad (45)$$

where α_{be} and α_{bn} are the path loss exponents. The locations of the BS, RIS and Eve are set as $(0, 10)$, $(60, 10)$ and $(50, 0)$ respectively. The NOMA users are randomly distributed in a circle whose center is located at $(60, 0)$ with a radius of 5 m. In terms of the imperfect Eve's CSI, we define the CSI error bound as $\delta_e = \xi\|\mathbf{H}_e\|_F$, where $\xi \in [0, 1)$ stands for the relative amount of CSI uncertainty [36], [50], [51]. The simulation parameters are given in Table I, and each point is averaged by 100 channel realizations.

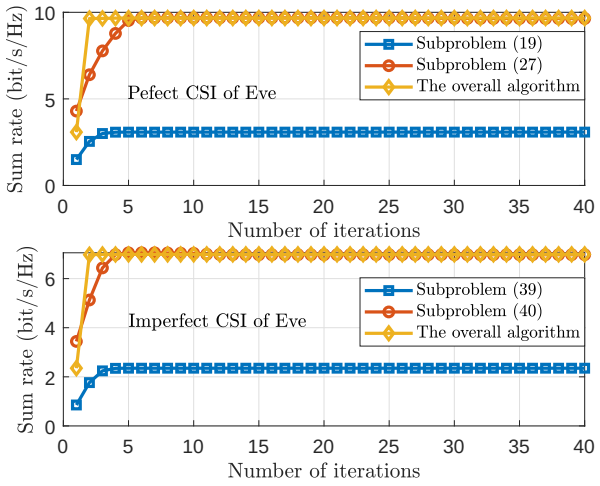


Fig. 2: Convergence of the proposed optimization algorithm.

Fig. 2 shows the convergence of the proposed algorithm when the CSI of Eve is perfectly and imperfectly known. It can be observed that both two subproblems converge to a stationary point within few iterations, which guarantees the overall algorithm convergence. Furthermore, problems (26) and (42) can achieve a larger value based on the solution of problems (19) and (41), which demonstrates the accuracy of our convergence analysis and the efficiency of the developed AO algorithm.

In Figs. 3 and 4, we compare the proposed SIRJ scheme

TABLE I: Simulation Parameters

α_{bn}	Path loss exponents of the Rayleigh fading channels	3.5
α_{be}		
α_{br}	Path loss exponents of the Rician fading channels	2
α_{rn}		
α_{re}		
$\mathcal{K}_{br}$	Rician factors	1
$\mathcal{K}_{rn}$		
$\mathcal{K}_{re}$		
λ_0	Path loss at the reference distance of 1 meter	-30 dB
ε	Convergence threshold for the iteration and AO	0.01
ε_1	Convergence threshold for rank-one constraint	0.01
ε_2	Convergence threshold for binary constraint	0.1
ρ, ϱ	Initialized penalty factors	0.0001
ω	Scaling factor	10
τ	Maximum tolerant information leakage	1 bit/s/Hz
σ_0	Noise power	-90 dBm

with other baseline schemes:

- **SIRJ based OMA (SIRJ-OMA)**: In this scheme, users served by the BS equally share the overall spectrum resource, where the bandwidth of each user is $1/N$.
- **Conventional information relaying RIS (r-RIS)**: In this scheme, all RIS elements are used to reflect the incident signal.
- **Conventional jamming RIS (j-RIS)**: The RIS elements are all set to convert the incident signal into jamming to confuse the Eve.
- **Without RIS (WoR)**: There is no RIS deployed and only the transmit beamforming of the BS is optimized.

It can be observed from Fig. 3 that the performance of SIRJ-NOMA is similar to that of r-RIS, but the performance gain is remarkable in Fig. 4. The reasons are as follows. When perfect CSI of Eve is available at the BS, for both the BS and RIS, they only need to keep the beam away from the direction of Eve and concentrate the signal power in the directions of users, which makes SIRJ-NOMA not much different from r-RIS, implying that nearly all RIS elements are used for signal reflection. However, as transmit power becomes large, SIRJ-NOMA begins to use a small number of elements to remodulate the incident signal as jamming to ensure that the secrecy constraint is satisfied, such that SIRJ-NOMA can offer a higher sum-rate than that of r-RIS. When the BS has only imperfect CSI of Eve, its position cannot be determined exactly. In this case, the secrecy constraint is difficult to meet, and r-RIS will reflect a part of the signal in the direction of Eve to perform signal cancellation, which reduces the received signal power at the legitimate users. Moreover, since the exact position of Eve cannot be known, this approach may not achieve an ideal signal cancellation, resulting in a waste of signal power and even an increased received signal power at Eve. However, SIRJ-NOMA is quite different. Since the jamming signal generated by the RIS is reflected in the direction of Eve, and jamming can play an important role in degrading the performance of Eve. Therefore, the proposed SIRJ scheme offers a significant performance gain under imperfect CSI of Eve.

However, this does not mean that using all RIS elements to generate jamming can achieve a much better performance. In Fig. 3, it can be seen that j-RIS provides a much smaller sum-rate compared to SIRJ-NOMA and r-RIS. This is because j-RIS cannot reflect the legitimate signal, and too much jamming significantly degrades the reception quality of legitimate users. Especially for the case with perfect CSI of Eve, no jamming is needed. In Fig. 4, although j-RIS brings a higher sum-rate compared to r-RIS as transmit power increases since jamming can make the secrecy constraint easier to satisfy, and too much jamming in the system even makes the sum-rate of j-RIS decrease. However, SIRJ-NOMA performs much better than j-RIS under both CSI conditions, due to the fact that it is able to adaptively select the working mode of each RIS element, which makes RIS not only confuse Eve, but also enhance signal transmission and keep the jamming power under an acceptable level but meet the security requirement.

For SIRJ-OMA, since users equally divide the frequency band, the achievable rate of each user is lower than that of NOMA where users share the same spectrum resource, and thus SIRJ-OMA cannot provide a sum-rate as high as the proposed SIRJ-NOMA scheme in Figs. 3 and 4. In Fig. 3, WoR is similar to j-RIS, since the useful signal only comes from the BS. But in Fig. 4, the jamming in j-RIS may affect users more, which results in a little worse performance than WoR. Moreover, since there is no need to split a part of the transmit power for signal cancellation at Eve, WoR can achieve even better performance than r-RIS with a high transmit power. The reason behind this observation is that the conventional RIS has only a single functionality, while the proposed SIRJ-NOMA has dual functionalities that can be adjusted to provide a high sum-rate and achieve secure NOMA communication with a controllable cost. It can be also observed that r-RIS is slightly better than SIRJ-NOMA with a low transmit power in Figs. 3 and 4, which is due to the binary constraint. In **Algorithm 1**, we set a convergence threshold to make the reflection amplitude close to 1. In other words, it cannot

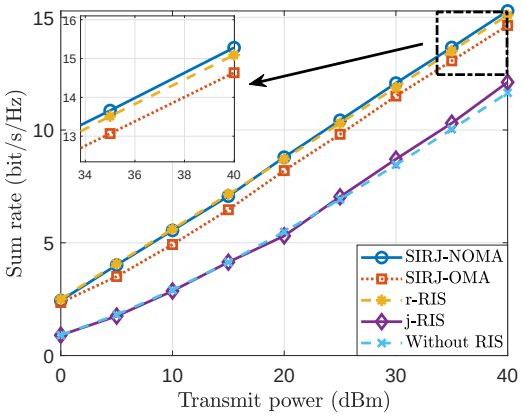


Fig. 3: Sum rate versus the transmit power with **perfect** CSI of Eve under $L = 2$, $N = 2$, and $K = 30$.

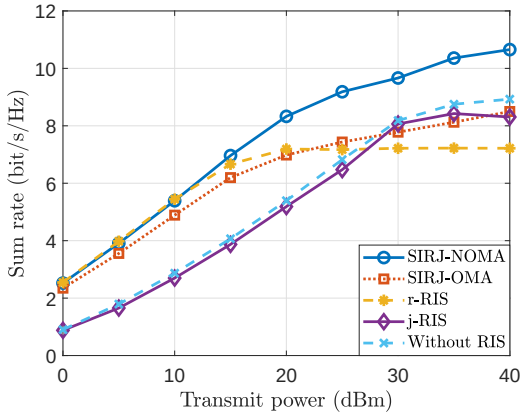


Fig. 4: Sum rate versus the transmit power with **imperfect** CSI of Eve under $L = 2$, $N = 2$, and $K = 30$.

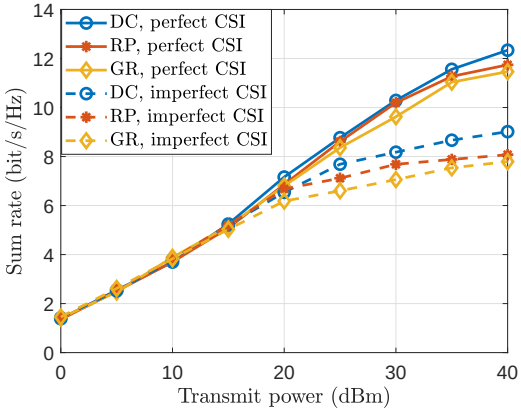


Fig. 5: Comparison of optimization algorithm with $L = 2$, $N = 2$, $K = 10$.

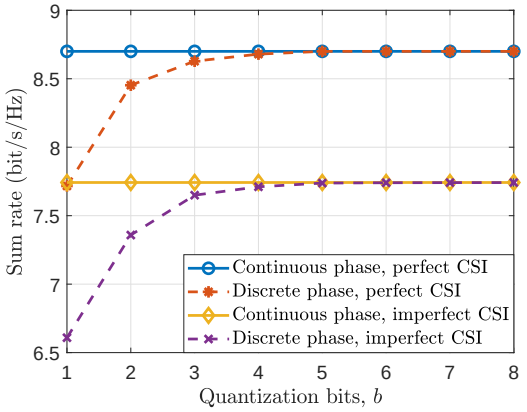


Fig. 6: Sum rate versus quantization bits with $L = 2$, $N = 2$, $K = 10$, $P_t = 25$dBm.

be exactly equal to 1, but this is realistic. While for r-RIS, the reflection amplitude constraint is usually set to 1, which provides a performance upper bound.

We compared the performance of different optimization methods in Fig. 5, including SDR with Gaussian Randomization (GR) method in [16], and Ring Penalty (RP) method in [40]. It can be seen that as the transmit power increases, there is a gap between three methods. In terms of GR, it randomly generates a large number of vectors that obey the CSCG distribution, and selects the vector closest to the optimization result as an alternative solution. Because of the large randomness, the performance of GR is slightly worse among the three. As for RP, its core is unit modulus constraint according to [40]. Compared to GR, it is closely related to the original rank-one constraint, which leads to a better performance. However, as for the DC method adopted by this paper, the rank-one constraint is converted into a equivalent constraint on the basis of the matrix property, for example, inequality (24), which maintains a more closer relationship with the original rank-one constraint, so the performance is the best among the three.

In Fig. 6, we investigate the impact of discrete phase on the system performance and its relation with the continuous phase. It can be observed that the performance loss caused by discrete phase becomes smaller with the quantization bits increase, and it comes to around 0.3% with 4-bit quantization. Different from the conventional single functional RIS, SIRJ needs to control the beam of signal and jamming at the same time, so more quantization bits are required.

In Fig. 7, we investigate the impact of different CSI uncertainties on the sum-rate. It can be seen that the performance is the best when the CSI of Eve is completely known. However, as the CSI uncertainty increases, the system sum-rate becomes smaller. This is because the increased channel uncertainty makes it more difficult to satisfy the secrecy constraint, which decreases the received signal power at users. This can be seen more clearly from Fig. 8, where the number of RIS elements in each working mode is plotted. We observe that more RIS elements are used to convert the incident signal into jamming to counteract the influence of channel uncertainty. This, in turn, demonstrates the efficiency of our SIRJ scheme that can adaptively adjust the working mode of each RIS element according to different communication requirements.

Fig. 9 shows the effect of the number of RIS elements and BS antennas on the sum-rate. As the number of RIS elements increases, the SIRJ scheme can provide higher degrees of freedom not only for signal reflection but also for jamming, which leads to an increased sum-rate. Similarly, increasing the number of BS antennas can also benefit the sum-rate. It is worth noting that the improvement from $N = 3$ to

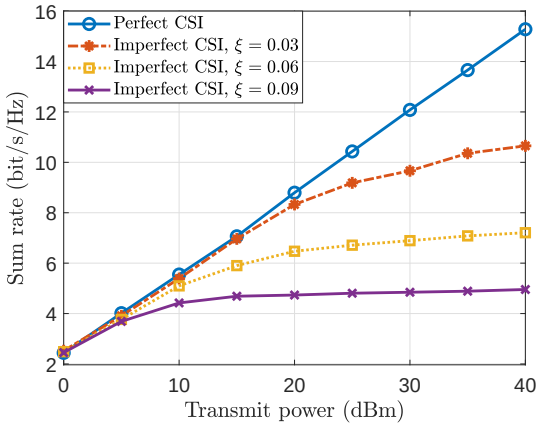


Fig. 7: Sum rate versus the transmit power under different CSI uncertainties of the Eve with $L = 2$, $N = 2$, and $K = 30$.

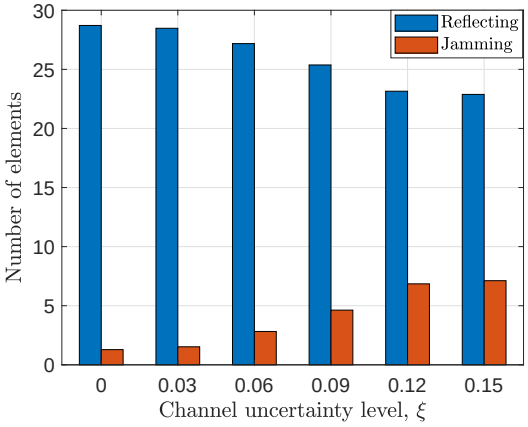


Fig. 8: Number of elements for different modes versus different CSI uncertainties with $L = 2$, $N = 2$, $K = 30$, and $P_t = 40$ dBm.

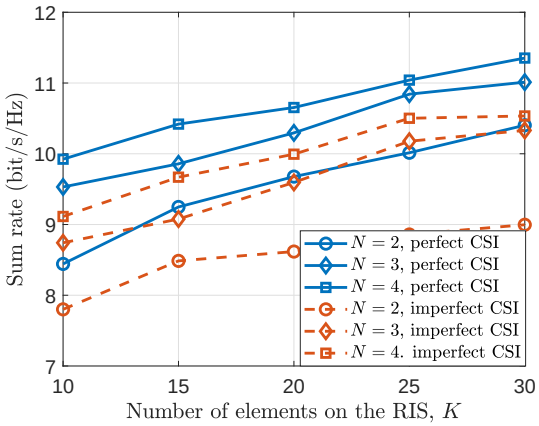


Fig. 9: Sum rate versus the number of RIS elements with $L = 2$, and $P_t = 25$ dBm.

$N = 4$ is less than that from $N = 2$ to $N = 3$ under the two CSI conditions. This is because fixing the number of elements on RIS also fixes the degree of freedom, i.e., increasing the number of antennas at the BS can only improve the receiving quality of the direct link. In other words, when the gain brought by the number of antennas is saturated, increasing the number of RIS elements can still continue to improve the system. Although the performance gain brought by 1 antenna is equivalent to 5 to 10 elements, taking into account the difference in cost between the antennas and the elements, increasing the number of elements on RIS provides us with an efficient and cost-effective approach to enhance the system performance.

VI. Conclusion

In this paper, we proposed a new SIRJ scheme using a dual-functional RIS to improve physical layer security of NOMA systems, where the working mode (e.g., reflecting or remodulating incident signal as jamming) of each RIS element can be adaptively selected to enhance the reception quality of the legitimate users while degrading the performance of Eve. The transmit beamforming of the BS, reflect beamforming of the RIS, and mode selection matrix of the RIS were jointly optimized to maximize the system sum-rate, where two CSI conditions related to Eve were considered. In the case of perfect CSI, a penalty based AO algorithm was developed to tackle the challenging mixed integer non-convex optimization problem. Under the imperfect CSI condition, we concentrated on dealing with the infinite number of secrecy constraints, and devised some effective mathematical processing methods to apply the $\mathcal{S}$-procedure. Simulation results reveal that our proposed SIRJ scheme can achieve a better performance than the existing baseline schemes. Furthermore, a stable performance improvement can be achieved by increasing the number of RIS elements rather than the number of BS antennas.

This work can be extended to several scenarios. For example, the proposed SIRJ strategy can be combined with wireless information and energy/power transmission, where the element can first harvest energy from the transmitted signal to supply the consumption of the RIS. Based on this, this scheme can be further performed via an active-RIS, where the elements can amplify the useful signal and the generated jamming with the harvest energy to against the double fading of passive-RIS, while the working mode and amplification factor of each element can be jointly optimized as well as the beamforming of the active-RIS.

Appendix A
Proof of Theorem 1

Note that problem (19) is jointly convex with respect to $\mathbf{P}_m$, and Slater's condition is satisfied. Therefore, the gap between the optimal value of problem (19) and that of its dual problem is zero. Specifically, the Lagrangian function of problem (19) about $\mathbf{P}_m$ is shown on the top of next page due to the space limitation, where a, $\mathbf{Q}_m$, b_{nm}, c_{nm} and d_m are the Lagrange multipliers corresponding to constraints (7b), (8c), (10), (14) and (16). β stands for the terms that are independent of $\mathbf{P}_m$. Then, the Karush-Kuhn-Tucker(KKT) conditions are checked to investigate the structure of the optimal solution, which can be expressed as

$$\text{K1}: a^* \geq 0, b_{nm}^* \geq 0, c_{nm}^* \geq 0, d_m^* \geq 0, \mathbf{Q}_m^* \succeq 0, \quad \text{(A.2)}$$

$$\begin{aligned}\mathcal{L} = & a\big(P_t - \sum_{m=1}^{N} \mathrm{Tr}(\mathbf{P}_m)\big) + \sum_{m=1}^{N} \mathrm{Tr}(\mathbf{P}_m \mathbf{Q}_m) \\ & + \sum_{n=1}^{N} \sum_{m=1}^{N-1} b_{nm} \big(\mathrm{Tr}(\mathbf{H}_n \mathbf{P}_{m+1} \mathbf{H}_n^H \mathbf{R}) - \mathrm{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R})\big) \\ & + \sum_{n=1}^{N} \sum_{m \le n}^{N} c_{nm} \Big\{ 2\mathrm{Tr}(\mathbf{H}_n \mathbf{P}_m \mathbf{H}_n^H \mathbf{R}) - \Big(\mu_{nm}\big(\sigma_0^2 + \sum_{u=1}^{N} \mathrm{Tr}(\mathbf{H}_n \mathbf{P}_u \mathbf{H}_n^H \mathbf{J}) + \sum_{l=m+1}^{N} \mathrm{Tr}(\mathbf{H}_n \mathbf{P}_l \mathbf{H}_n^H \mathbf{R})\big)\Big)^2 \Big\} \\ & + \sum_{m=1}^{N} d_m \big(T_e \sigma_0^2 + T_e \sum_{u=1}^{N} \mathrm{Tr}(\mathbf{H}_e \mathbf{P}_u \mathbf{H}_e^H \mathbf{J}) - \mathrm{Tr}(\mathbf{H}_e \mathbf{P}_m \mathbf{H}_e^H \mathbf{R})\big) + \beta. \end{aligned} \tag{A.1}$$

$$\text{K2} : \mathbf{Q}_m^* \mathbf{P}_m^* = \mathbf{0}, \tag{A.3}$$

$$\text{K3} : \nabla_{\mathbf{P}_m^*} \mathcal{L} = 0, \tag{A.4}$$

where a^*, b_{nm}^*, c_{nm}^*, d_m^* and $\mathbf{Q}_m^*$ are the optimal Lagrange multipliers, and $\nabla_{\mathbf{P}_m^*} \mathcal{L}$ stands for the gradient of $\mathcal{L}$ with respect to $\mathbf{P}_m^*$. In terms of K3, it can be further rewritten as

$$\mathbf{Q}_m^* = a^* \mathbf{I} - \boldsymbol{\Upsilon}, \tag{A.5}$$

where $\boldsymbol{\Upsilon}$ is given by

$$\begin{aligned}\boldsymbol{\Upsilon} = & \sum_{n=1}^{N} b_{nm}^* \mathbf{H}_n^H \mathbf{R} \mathbf{H}_n + \sum_{n=1}^{N} 2c_{nm}^* \mathbf{H}_n^H (\mathbf{R} - \mu_{nm} \mathbf{J}) \mathbf{H}_n \\ & + d_m^* \mathbf{H}_e^H (T_e \mathbf{J} - \mathbf{R}) \mathbf{H}_e \end{aligned} \tag{A.6}$$

It is known that $\lambda_{\max}(\mathbf{X})\mathbf{I} - \mathbf{X} \succeq 0$ for any Hermite matrix, where $\lambda_{max}(\mathbf{X})$ denotes the maximum eigenvalue of $\mathbf{X}$. Therefore, to satisfy $\mathbf{Q}_m^* \succeq 0$ in K1, $a^* \geq \lambda_{\max}(\boldsymbol{\Upsilon})$ needs to hold. Note that $\mathbf{Q}_m^* \succ 0$ when $a^* > \lambda_{\max}(\boldsymbol{\Upsilon})$, which makes $\mathbf{P}_m = 0$ considering K2 in (A.3), but it is obviously against reality. Under this condition, $a^* = \lambda_{\max}(\boldsymbol{\Upsilon})$ must hold. Based on this, $\mathbf{Q}_m^*$ is a positive semidefinite matrix and we have $L > \mathrm{rank}(\mathbf{Q}_m^*) \geq L - 1$.

Lemma 4: (Sylvester inequality): Let $\mathbf{A} \in \mathbb{C}^{S \times U}$, $\mathbf{B} \in \mathbb{C}^{U \times W}$, the following inequality holds:

$$\mathrm{rank}(\mathbf{AB}) \geq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B}) - U. \tag{A.7}$$

By adopting Lemma 4 and condition K2, we have $L \geq \mathrm{rank}(\mathbf{Q}_m^*) + \mathrm{rank}(\mathbf{P}_m^*) \geq L - 1 + \mathrm{rank}(\mathbf{P}_m^*)$. Comparing the leftmost and rightmost sides of this inequality, the conclusion $\mathrm{rank}(\mathbf{P}_m^*) \leq 1$ follows.

REFERENCES

[1] M. Ji, J. Chen, L. Lv, Q. Wu, Z. Ding, and L. Yang, "Simultaneous information relaying and jamming via IRS to secure NOMA networks," in *Proc. Wireless Commun. Signal Process. (WCSP)*, Nov. 2022, pp. 1–6.

[2] X.-H. You et al., "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Jan. 2021.

[3] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[4] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC)*, Jun. 2013, pp. 1–5.

[5] Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE,* vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[6] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas in Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[7] L. Lv, J. Chen, Q. Ni, Z. Ding, and H. Jiang, "Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 188–195, Apr. 2018.

[8] L. Lv, H. Jiang, Z. Ding, Q. Ye, N. Al-Dhahir, and J. Chen, "Secure non-orthogonal multiple access: An interference engineering perspective," *IEEE Network*, vol. 35, no. 4, pp. 278–285, Jul./Aug. 2021.

[9] Y. Feng et al., "Beamforming design and power allocation for secure transmission with NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2639–2651, May 2019.

[10] A. Arafa et al., "Secure relaying in non-orthogonal multiple access: Trusted and untrusted scenarios," *IEEE Trans. Info. Forensics Sec.*, vol. 15, no. 1, pp. 210–222, Jan. 2020.

[11] N. Zhao et al., "Joint beamforming and jamming optimization for secure transmission in MISO-NOMA networks," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2294–2305, Mar. 2019.

[12] L. Lv, H. Jiang, Z. Ding, L. Yang, and J. Chen, "Secrecy-enhancing design for cooperative downlink and uplink NOMA with an untrusted relay," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1698–1715, Mar. 2020.

[13] A. Almohamad et al., "Smart and secure wireless communications via reflecting intelligent surfaces: A short survey," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1442–1456, 2020.

[14] 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, May 2021.

[15] L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, and J. Chen, "Covert communication in intelligent reflecting surface-assisted NOMA systems: Design, analysis, and optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1735–1750, Mar. 2022.

[16] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[17] J. Chen, Y. -C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, 2019.

[18] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[19] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent reflecting surface aided multi-antenna secure transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 108–112, Jan. 2020.

[20] J. Qiao, and M. -S. Alouini, "Secure transmission for intelligent reflecting surface-assisted mmWave and Terahertz systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1743–1747, Oct. 2020.

[21] X. Lu, W. Yang, X. Guan, Q. Wu, and Y. Cai, "Robust and secure beamforming for intelligent reflecting Surface Aided mmWave MISO Systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2068–2072, Dec. 2020.

[22] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778–782, Jun. 2020.

[23] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[24] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Process. Lett.*, vol. 27, pp. 1300–1304, 2020.

[25] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure MIMO wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, Dec. 2020.

[26] S. Hong, C. Pan, H. Ren, K. Wang, K. K. Chai, and A. Nallanathan, "Robust transmission design for intelligent reflecting surface-aided secure communication systems with imperfect cascaded CSI," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2487–2501, Apr. 2021.

[27] Q. Wang, F. Zhou, R. Q. Hu, and Y. Qian, "Energy efficient robust beamforming and cooperative jamming design for IRS-assisted MISO networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2592–2607, Apr. 2021.

[28] S. Hu, Z. Wei, Y. Cai, C. Liu, D. W. K. Ng, and J. Yuan, "Robust and secure sum-rate maximization for multiuser MISO downlink systems with self-sustainable IRS," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7032–7049, Oct. 2021.

[29] J. Li, L. Zhang, K. Xue, Y. Fang, and Q. Sun, "Secure transmission by leveraging multiple intelligent reflecting surfaces in MISO systems," *IEEE Trans. Mobile Comput.*, early access, Sep. 2021, doi: 10.1109/TM-C.2021.3114167.

[30] L. Lv, Q. Wu, Z. Li, N. Al-Dhahir, and J. Chen, "Secure two-way communications via intelligent reflecting surfaces," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 744–748, Mar. 2021.

[31] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[32] S. Gong et al., "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts*, vol. 22, no. 4, pp. 2283–2314, 4th Quart., 2020.

[33] S. Xu, J. Liu, and J. Zhang, "Resisting undesired signal through IRS-based backscatter communication system," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2743–2747, Aug. 2021.

[34] S. Xu, J. Liu, and Y. Cao, "Intelligent reflecting surface empowered physical-layer security: Signal cancellation or jamming?" *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1265–1275, Jan. 2022.

[35] Z. Ding, L. Lv, F. Fang, et al., "A state-of-the-art survey on reconfigurable intelligent surface-assisted non-orthogonal multiple access networks," *Proc. IEEE*, vol. 110, no. 9, pp. 1358–1379, Sep. 2022.

[36] Z. Zhang, L. Lv, Q. Wu, H. Deng, and J. Chen, "Robust and secure communications in intelligent reflecting surface assisted NOMA networks," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 739–743, Mar. 2021.

[37] Z. Zhang, C. Zhang, C. Jiang, F. Jia, J. Ge, and F. Gong, "Improving physical layer security for reconfigurable intelligent surface aided NOMA 6G networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4451–4463, May 2021.

[38] H. Han, Y. Cao, M. Sheng, N. Zhao, J. Liu, and D. Niyato, "IRS-aided secure NOMA networks against internal and external eavesdropping," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7536–7548, Nov. 2022.

[39] W. Wang et al., "Beamforming and jamming optimization for IRS-aided secure NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1557–1569, Mar. 2022.

[40] Z. Zhang, J. Chen, Q. Wu, Y. Liu, L. Lv, and X. Su, "Securing NOMA networks by exploiting intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1096–1111, Feb. 2022.

[41] C. Gong, X. Yue, X. Wang, X. Dai, R. Zou, and M. Essaaidi, "Intelligent reflecting surface aided secure communications for NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2761–2773, Mar. 2022.

[42] Y. Feng, J. Chen, X. Xue, K. Wu, Y. Zhou, and L. Yang, "Max-min fair beamforming for IRS-aided secure NOMA systems," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 234–238, Feb. 2022.

[43] Z. Tang, T. Hou, Y. Liu, J. Zhang, and L. Hanzo, "Physical layer security of intelligent reflective surface aided NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7821–7834, Jul. 2022.

[44] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3764–3778, Jun. 2022.

[45] H.-M. Wang, X. Zhang, Q. Yang, and T. A. Tsiftsis, "Secure users oriented downlink MISO NOMA," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 671–684, Jun. 2019.

[46] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

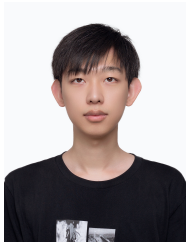
[47] L. Lv, Z. Ding, Q. Ni, and J. Chen, "Secure MISO-NOMA transmission with artificial noise," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6700–6705, Jul. 2018.

[48] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, 2020.

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

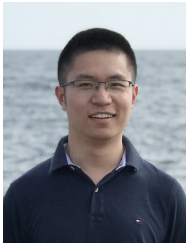
[50] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.

[51] J. Cheng, C. Shen, Z. Chen, and N. Pappas, "Robust beamforming design for IRS-aided URLLC in D2D networks," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6035–6049, Sep. 2022.



Mengyi Ji received the B.Sc. and M.Sc. degrees from Xidian University, China, in 2020 and 2023, respectively. His research interests include non-orthogonal multiple access, intelligent reflecting surface, and physical-layer security.

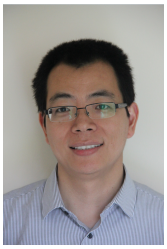
Jian Chen (Member, IEEE) received the B.Sc. degree from Xi'an Jiaotong University, China, in 1989, the M.Sc. degree from Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, China, in 1992, and the Ph.D. degree from Xidian University, China, in 2005. He was a Visiting Scholar with the University of Manchester, from 2007 to 2008, and a Senior Visiting Scholar with University of Alberta, from 2017 to 2018. He is currently a Professor with Xidian University. His research interests include wireless sensor networks, physical-layer security, and non-orthogonal multiple access.



Lu Lv (Member, IEEE) received the Ph.D. degree from Xidian University, China, in 2018. From 2016 to 2018, he was an Academic Visitor with Lancaster University and University of Alberta. In 2019, he was a Post-Doctoral Fellow with Dalhousie University. He is currently an Associate Professor with Xidian University. His research interests include non-orthogonal multiple access, physical-layer security, reconfigurable intelligent reflecting surface, and covert communication. He was the recipient of the Outstanding Ph.D. Thesis Award of Shannxi Province in 2020, the IEEE ICCC Best Paper Award in 2021, and the Exemplary Reviewer Certificate for IEEE TRANSACTIONS ON COMMUNICATIONS from 2018 to 2020 and for IEEE COMMUNICATIONS LETTERS in 2022. He was listed as a World's Top 2% Scientist by Stanford University in 2022. He serves as an Associate Editor for IEEE IOT-AHSN TC NEWSLETTER, *Frontiers in Computer Science*, and *Journal of Wireless Communications and Mobile Computing*. He is the Lead Guest Editor for IEEE INTERNET OF THINGS JOURNAL.

Qingqing Wu (Senior Member, IEEE) is an Associate Professor with Shanghai Jiao Tong University. His current research interest includes intelligent reflecting surface (IRS), unmanned aerial vehicle (UAV) communications, and MIMO transceiver design. He has coauthored more than 100 IEEE journal papers with 29 ESI highly cited papers and 9 ESI hot papers, which have received more than 20,000 Google citations. He was listed as the Clarivate ESI Highly Cited Researcher in 2022 and 2021, the Most Influential Scholar Award in AI-2000 by Aminer in 2021 and Worlds Top 2% Scientist by Stanford University in 2020 and 2021.

He was the recipient of the IEEE Communications Society Fred Ellersick Prize, IEEE Best Tutorial Paper Award in 2023, Asia-Pacific Best Young Researcher Award and Outstanding Paper Award in 2022, Young Author Best Paper Award in 2021, the Outstanding Ph.D. Thesis Award of China Institute of Communications in 2017, the IEEE ICCC Best Paper Award in 2021, and IEEE WCSP Best Paper Award in 2015. He was the Exemplary Editor of IEEE Communications Letters in 2019 and the Exemplary Reviewer of several IEEE journals. He serves as an Associate Editor for IEEE Transactions on Communications, IEEE Communications Letters, IEEE Wireless Communications Letters. He is the Lead Guest Editor for IEEE Journal on Selected Areas in Communications. He is the workshop co-chair for IEEE ICC 2019-2023 and IEEE GLOBECOM 2020. He serves as the Workshops and Symposia Officer of Reconfigurable Intelligent Surfaces Emerging Technology Initiative and Research Blog Officer of Aerial Communications Emerging Technology Initiative. He is the IEEE Communications Society Young Professional Chair in Asia Pacific Region.



Zhiguo Ding (Fellow, IEEE) received his B.Eng from the Beijing University of Posts and Telecommunications in 2000, and the Ph.D degree from Imperial College London in 2005. He is currently a Professor in Communications at Khalifa University, and has also been affiliated with the University of Manchester and Princeton University.

Dr. Ding's research interests are 6G networks, cooperative and energy harvesting networks and statistical signal processing. He is serving as an Area Editor for *IEEE Transactions on Wireless Communications* and *IEEE Open Journal of the Communications Society*, an Editor for *IEEE Transactions on Vehicular Technology*, *SCIENCE CHINA Information Sciences* and *IEEE Communications Surveys & Tutorials*, and was an Editor for *IEEE Wireless Communication Letters*, *IEEE Transactions on Communications*, *IEEE Communication Letters*. He recently received the EU Marie Curie Fellowship 2012–2014, the Top IEEE TVT Editor 2017, IEEE Heinrich Hertz Award 2018, IEEE Jack Neubauer Memorial Award 2018, IEEE Best Signal Processing Letter Award 2018, Friedrich Wilhelm Bessel Research Award 2020, and IEEE SPCC Technical Recognition Award 2021. He is a Fellow of the IEEE, a Distinguished Lecturer of IEEE ComSoc, and a Web of Science Highly Cited Researcher in two categories 2022.

Naofal Al-Dhahir (Fellow, IEEE) is Erik Jonsson Distinguished Professor and ECE Associate Head at UT-Dallas. He earned his Ph.D. degree in Electrical Engineering from Stanford University.

From 1994 to 2003, he was a principal member of the technical staff at GE Research and AT&T Shannon Laboratory. He is co-inventor of 42 issued US patents, co-author of over 415 papers and co-recipient of 5 IEEE best paper awards. From 2016 to 2019, he served as the Editor-in-Chief for IEEE TRANSACTIONS ON COMMUNICATIONS.