

# Sum Secrecy Rate Maximization for IRS-aided Multi-Cluster MIMO-NOMA Terahertz Systems

Jinlei Xu, Zhengyu Zhu, *Senior Member, IEEE*, Zheng Chu, *Member, IEEE*, Hehao Niu, *Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, and Inkyu Lee, *Fellow, IEEE*

**Abstract**—Intelligent reflecting surface (IRS) is a promising technique to extend the network coverage and improve spectral efficiency. This paper investigates an IRS-assisted terahertz (THz) multiple-input multiple-output (MIMO)-nonorthogonal multiple access (NOMA) system based on hybrid precoding with the presence of eavesdropper. Two types of sparse RF chain antenna structures are adopted, i.e., sub-connected structure and fully connected structure. First, cluster heads are selected for each beam, and analog precoding based on discrete phase is designed. Then, users are clustered based on channel correlation, and NOMA technology is employed to serve the users. In addition, a low-complexity forced-zero method is utilized to design digital precoding in order to eliminate inter-cluster interference. On this basis, we propose a secure transmission scheme to maximize the sum secrecy rate by jointly optimizing the power allocation and phase shifts of IRS subject to the total transmit power budget, minimal achievable rate requirement of each user, and IRS reflection coefficients. Due to multiple coupled variables, the formulated problem leads to a non-convex issue. We apply the Taylor series expansion and semidefinite programming to convert the original non-convex problem into a convex one. Then, an alternating optimization algorithm is developed to obtain a feasible solution of the original problem. Simulation results verify the convergence of the proposed algorithm, and deploying IRS can bring significant beamforming gains to suppress the eavesdropping.

**Index Terms**—Intelligent reflecting surface, terahertz, MIMO-NOMA, hybrid precoding, secure communication.

## I. INTRODUCTION

With the accelerated progress of the global round of new technological revolution and industrial transformation, the

This work was supported in part by the open research fund of National Mobile Communications Research Laboratory Southeast University under Grant 2023D11; in part by Sponsored by Program for Science & Technology Innovation Talents in Universities of Henan Province under Grant 23HASTIT019; in part by Natural Science Foundation of Henan Province under Grant 232300421097; in part by the Project funded by China Postdoctoral Science Foundation under Grant 2020M682345; in part by the Henan Postdoctoral Foundation No.202001015; in part by the National Research Foundation of Korea funded by the Ministry of Science and ICT, Korea Government under Grant 2017R1A2B3012316. (*Corresponding author: Zhengyu Zhu*)

J. Xu is with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China. (e-mail: iejlxu@mail.dlut.edu.cn).

Z. Zhu is with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China, and is also with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210018, China (e-mail: zhuzhengyu6@gmail.com)

Z. Chu and P. Xiao are with the 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford GU2 7XH, UK. (e-mail: andrew.chuzheng7@gmail.com, p.xiao@surrey.ac.uk)

H. Niu is with the Institute of Electronic Countermeasure, National University of Defense Technology, Hefei 230037, China. (e-mail: niuhaonupt@foxmail.com)

I. Lee is with School of Electrical Engineering, Korea University, Seoul, Korea. (e-mail: inkyu@korea.ac.kr)

widespread adoption and application of data-centered technologies such as artificial intelligence (AI), virtual reality, autonomous driving, and smart cities have posed significant challenges to the fifth-generation mobile communication (5G) networks [1], [2]. Therefore, it is imperative to develop the sixth-generation mobile communication (6G) technologies with new characteristics for the communication demands of the future information society [3], [4].

To meet the development vision of "better performance, greater intelligence, broader coverage, and greener," 6G puts forward higher requirements for a series of indicators such as data rates, transmission latency, and secure communication [5]. Compared with 5G, 6G can continuously expand the breadth and depth of applications in the mobile Internet of Things and enable seamless connectivity among all things, which will better supports the robust development of the digital economy [6]. Conventional wireless communication systems operating below 5 GHz, and even the millimeter-wave (mmWave) technology are unable to meet the massive connectivity demands of 6G. Therefore, new wireless technologies are urgently needed to address the challenge of capacity shortage [7]. Compared with the current wireless technologies, terahertz (THz) communication possesses higher frequencies and wider bandwidth, enabling support for a greater number of connected devices and achieve high transmission rates from 100 Gbit/s to 1 Tbit/s [8]. It can alleviate the issues of spectrum scarcity and capacity limitations in current wireless systems and is considered one of the potential technologies for 6G.

Due to the inherent broadcast nature of wireless channels, wireless systems are vulnerable to malicious attack, phishing and eavesdropping [9]. In order to deal with the serious threat to the secure transmission of the communication system, the research on the physical layer security (PLS) of wireless communications by leveraging the channel propagation characteristics has gradually become a research hotspot in the field of wireless security [10].

Recently, intelligent reflecting surface (IRS), composed by a large number of low-cost meta-material units, can dynamically configure the propagation environment of wireless signals via programming to improve the comprehensive performance of systems, which has been considered as one of the promising technologies of 6G [11]. In particular, by adaptively adjusting the reflection coefficients and amplitudes of IRS elements, a directional energy beam can be formed and reflected to the legitimate receivers, which can effectively suppress eavesdropping and expand the signal coverage [12], [13]. IRS can be used to mitigate path loss and blockage issues in THz system. Deploying IRS can bypass obstacles to provide a virtual line-of-sight (LoS) link for THz communications, thus alleviating

signal attenuation caused by molecular absorption and improving the security of transmission [14]. An IRS assisted THz indoor communication system is studied in [15], by jointly optimizing IRS deployment and reflection coefficients, THz sub-band allocation, as well as power control to maximize sum rate of users. In [16], an IRS assisted simultaneous wireless information and power transmission THz security system was proposed when the LoS link is blocked. Under the outage probability constraints, the problem of jointly optimizing the transmit beamforming and IRS phase shifts to minimize the total transmit power of the system was studied. In addition, deploying IRS and using active and passive beamforming can effectively suppress eavesdropping to improve the security. In [17], Qiao *et al.* studied the problem of secure communication in THz system assisted by multiple IRSs. Considering the imperfect eavesdropping channel state information (CSI), the worst-case security rate was derived, and a robust secure strategy for multi eavesdropping system was proposed.

On the other hand, non-orthogonal multiple access (NOMA) can support large-scale access and effectively utilize spectrum resources, thus receiving extensive attention from the academic and industrial communities [18], [19]. Compared to the traditional orthogonal multiple access (OMA) scheme, NOMA can significantly improve transmission rates and system capacity by serving multiple users in a single resource block [20]. In addition, serial interference cancellation (SIC) technology is leveraged at the NOMA receivers to eliminate co-channel interference between users and realize the reuse of limited spectrum resources [21].

Furthermore, combining multiple-input multiple-output (MIMO) precoding with NOMA can significantly enhance system throughput [22]. In MIMO systems, equipping each antenna with a dedicated radio frequency (RF) chain results in higher energy consumption and hardware costs. Hybrid precoding (HP) adopted sparse RF links can reduce the number of RF chains required by the system, thereby improving energy efficiency without sacrificing performance [23]. Moreover, there are two typical structures for sparse RF chains: fully connected (FC) and sub-connected (SC) [24]. The spectral efficiency and energy efficiency of two typical hybrid precoding structures in MIMO systems with limited feedback channels were investigated in [25]. Among them, the FC structure has higher complexity and energy consumption, resulting in lower energy efficiency compared to the SC structure. Conversely, the SC architecture is simpler and more energy-efficient, but has lower spectrum efficiency [26].

The THz possesses abundant spectral resources and exhibits strong directionality, which can ensure the better correlation of channels. Therefore, by combining THz with MIMO-NOMA technology can significantly enhance spectrum efficiency [27]. To address the unreliability of THz transmission caused by the heterogeneous networks, Xu *et al.* proposed an intelligent reconfigurable THz MIMO-NOMA framework [28]. This framework allows for flexible configuration of mixed beams to meet the demands of ultra-high data rates and massive connections. To alleviate the problem of information security and increased power consumption brought about by large-

scale network access, Zhang *et al.* investigated the problem of maximizing energy efficiency (EE) in THz NOMA-MIMO systems. Based on the channel correlation characteristics, a user clustering algorithm based on enhanced k-means is proposed. In addition, they utilized a HP scheme based on SC to reduce power consumption to maximize the achievable EE with the presence of incomplete SIC [29].

Although many works have been conducted on the optimization of IRS, THz transmission and MIMO-NOMA, the above technologies are rarely optimized jointly, and clustering of system with large-scale multi-users is not considered. This oversight can lead to collisions among nodes sharing the same channel, resulting in a decrease in user service quality. Clustering multiple users can reduce routing overhead and improve network capacity. Inspired by their respective advantages, integrating them can further improve the security, spectrum efficiency and coverage. This paper considers an IRS-assisted multi-cluster THz MIMO-NOMA system. Our goal is to maximize the sum secrecy rate (SSR) by jointly optimizing the power allocation and IRS phase shifts while satisfying the total power budget, minimal rate requirement of users and IRS reflection coefficients. The main contributions of this paper can be summarized as follows.

- We propose an IRS-assisted multi-cluster THz MIMO-NOMA system, in which the base station (BS) employs a HP architecture with sparse RF chains, including both FC and SC architectures. Moreover, a clustering algorithm based on correlation channel gains of users is employed to mitigate inter-cluster interference. When an eavesdropper involved, an optimization problem is formulated to jointly optimize the power allocation and IRS phase shifts matrix, aiming to maximize the SSR.
- For the design of hybrid analog and digital precoding, we first perform clustering based on the channel conditions of the users, and select the cluster heads. Then, analog precoding is designed based on the equivalent channels of each cluster head to enhance the antenna array gain. Finally, zero-forcing (ZF) digital precoding is leveraged to eliminate inter-user interference with the maximum gain in the equivalent channels within each cluster. It is difficult to tackle the original problem directly due to the non-convexity of the objective function. As such, the joint optimization problem is decomposed into two sub-problems, which can be transformed into the tractable convex form via semi-definite relaxation (SDR) and the first-order Taylor expansion. Then, with the obtained analog and digital precoding, the sub-optimal power allocation and IRS phase shifts can be calculated through an alternating optimization (AO) algorithm.
- The security transmission performance of the proposed IRS-assisted multi-cluster THz MIMO-NOMA system is evaluated through simulations, which shows that the leveraged IRS passive beamforming can effectively suppress the eavesdropping. Additionally, FC architecture achieves higher SSR compared to SC, while having lower secure energy efficiency (SEE). What is more, there is a trade-off between SEE and SSR in the proposed scheme.

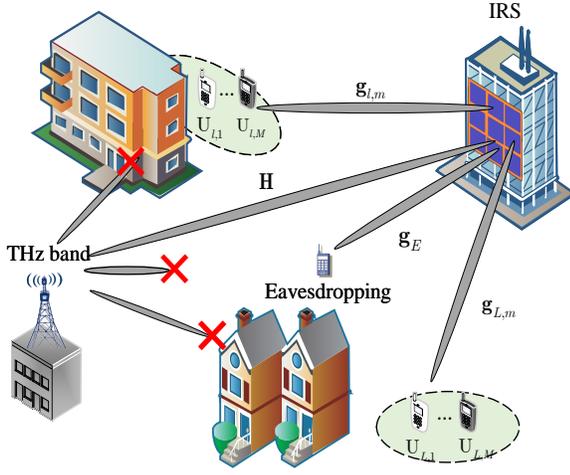


Fig. 1. System model.

The rest of the paper is organized as follows. Section II introduces the system model of IRS assisted THz MIMO-NOMA network. In Section III, user clustering and hybrid precoding are discussed carefully. In Section IV, we develop an alternate iteration algorithm to maximize sum secrecy rate under the rate requirement of each user. In Section V, our simulation results demonstrate the effectiveness of the proposed algorithms. Finally, we summarize the paper in Section VI.

**Notations:** Vectors and matrices are represented by bold lowercase and uppercase letters, respectively. For a matrix  $\mathbf{A}$ ,  $(\mathbf{A})^T$  and  $(\mathbf{A})^H$  represent transpose and Hermitian transpose, respectively.  $\mathbf{A} \succeq 0$  denotes that  $\mathbf{A}$  is a semi-positive definite matrix.  $\text{Tr}(\mathbf{A})$  stands for the trace of a matrix  $\mathbf{A}$ .  $\text{diag}(\mathbf{a})$  is a diagonal matrix.  $\mathbb{C}^{x \times y}$  means the space of  $x \times y$  complex matrix.  $\mathbb{E}[\cdot]$  describes expectation.  $\|\cdot\|$  stands for the Euclidean norm of the vector.  $\mathcal{CN}(\mu, \sigma^2)$  denotes the a complex Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ .

## II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, we investigate an IRS-aided multi-cluster THz MIMO-NOMA system. In particular, the BS with HP structure is equipped with  $N_{\text{TX}}$  antennas, broadcasting the superimposed signal information with the NOMA technique on each beam. Since the direct links between the BS and users are blocked by obstacles, i.e., buildings, an IRS, equipped with  $N_{\text{IRS}}$  reflecting elements, passively participates in the secure transmission from the BS to  $M$  single antenna legitimate users in the presence of a single-antenna eavesdropper (Eve). Moreover, it is assumed that all users can only receive reflected signals from IRS.

User clustering can help to achieve the good performance for the NOMA system [25], thus we divide users into  $L$  groups, and assume that the number of clusters is the same as the number of RF chains, i.e.,  $L = N_{\text{RF}}$ , and select cluster heads from users according to the channel gain and its correlation. Moreover, the users are grouped according to the scheme in [23]. By using NOMA, each beam can

support multiple users, and the inter-user interference can be eliminated by applying the SIC technique to users with lower channel gains [30]. Assuming that the CSI of all involved channels is perfect, the received signal of the  $m$ -th user in the  $l$ -th cluster can be given by

$$y_{l,m} = G_t G_r \mathbf{g}_{l,m} \mathbf{\Theta} \mathbf{H} \left( \sum_{j=1}^m \sqrt{p_{l,j}} s_{l,j} + \sum_{i \neq 1} \sum_{j=1}^M \sqrt{p_{i,j}} s_{i,j} \right) + n_{l,m}, \forall l, m, \quad (1)$$

Similarly, the received signal of Eve is expressed as

$$y_{l,m}^E = G_t G_r \mathbf{g}_E \mathbf{\Theta} \mathbf{H} \sum_{i=1}^L \sum_{j=1}^M \sqrt{p_{i,j}} s_{i,j} + v_{l,m}, \forall l, m, \quad (2)$$

where  $G_t$  and  $G_r$  represent the transmitting and receiving antenna gain, respectively,  $\mathbf{H} \in \mathbb{C}^{N_{\text{IRS}} \times N_{\text{TX}}}$  indicates the channel vector between BS and IRS,  $\mathbf{g}_{l,m} \in \mathbb{C}^{1 \times N_{\text{IRS}}}$  denotes the equivalent channel from IRS to the  $m$ -th user in the  $l$ -th cluster, and  $\mathbf{g}_E \in \mathbb{C}^{1 \times N_{\text{IRS}}}$  is the channel vector from IRS to the Eve. The reflection phase shift matrix of IRS is defined by  $\mathbf{\Theta} = \text{diag}\{\theta_1, \dots, \theta_n\}$ , where  $\theta_n = e^{j\theta_n}$ ,  $\theta_n \in (0, 2\pi]$ , and  $n \in \{1, \dots, N_{\text{IRS}}\}$ .  $s_{l,m} = \mathbf{F} \mathbf{v}_l x_{l,m}$ ,  $x_{l,m}$  is the transmission signal of the  $m$ -th user in the  $l$ -th cluster,  $\mathbf{v}_l \in \mathbb{C}^{N_{\text{RF}} \times 1}$  denotes the digital precoding vector of the  $l$ -th cluster, and  $\mathbf{F} \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{RF}}}$  denotes the analog beamforming for all clusters, and  $n_{l,m}$  and  $v_{l,m} \sim \mathcal{CN}(0, \sigma^2)$  are the additive white Gaussian noise (AWGN). Moreover, THz is easily absorbed by water when propagating in the air, resulting in the scattering component being far lower than the line-of-sight (LoS) component [31]. Thus, we ignore scattering components of channels, and  $\mathbf{H}$  can be expressed as

$$\mathbf{H} = q(f, d) \bar{\mathbf{H}}, \quad (3)$$

where  $q(f, d) = \frac{c}{4\pi f d} e^{-\frac{1}{2}\tau(f)d}$  represents the path loss consisting of the free-space path loss and the molecular absorption loss in THz communications,  $c$  is the speed of light,  $f$  is operating frequency,  $\tau(f)$  represents the absorption coefficient of water molecules in the air, and  $d$  is the distance from the BS to IRS.  $\bar{\mathbf{H}} = \alpha_r(\varphi) \alpha_t^H(\vartheta)$ , where  $\alpha_r(\varphi)$  and  $\alpha_t(\vartheta)$  are the antenna array response vector of the transmitter and the receiver, respectively, which can be defined as

$$\alpha_r(\varphi) = \frac{1}{\sqrt{N_{\text{TX}}}} \left[ 1, e^{j\pi\varphi}, e^{j2\pi\varphi}, \dots, e^{j(N_{\text{TX}}-1)\pi\varphi} \right], \quad (4)$$

$$\alpha_t(\vartheta) = \frac{1}{\sqrt{N_{\text{IRS}}}} \left[ 1, e^{j\pi\vartheta}, e^{j2\pi\vartheta}, \dots, e^{j(N_{\text{IRS}}-1)\pi\vartheta} \right]. \quad (5)$$

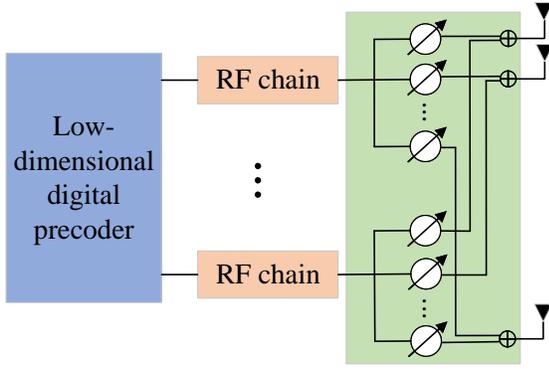
Here,  $\varphi = 2d_0 f \sin(\phi_t)/c$  and  $\vartheta = 2d_0 f \sin(\phi_r)/c$ , where  $d_0$  is the spacing between array elements, and  $\phi_i \in [-\pi/2, \pi/2]$ ,  $i = [r, t]$ , is the angle of departure (AOD) and the angle of arrival (AOA). Similarly,  $\mathbf{g}_{l,m}$  can be written as

$$\mathbf{g}_{l,m} = q(f, d_{l,m}) \bar{\mathbf{g}}_{l,m}, \quad (6)$$

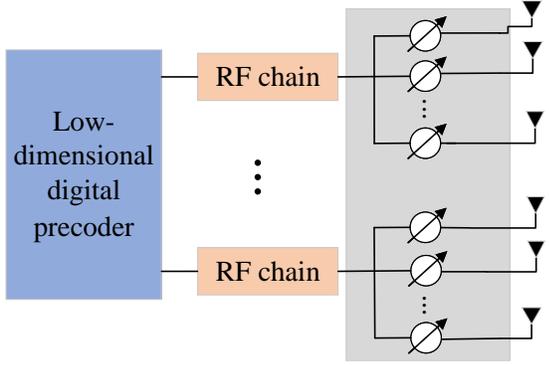
where  $\bar{\mathbf{g}}_{l,m} = \frac{1}{\sqrt{N_{\text{IRS}}}} \left[ 1, e^{j\pi\varphi_{l,1}}, e^{j2\pi\varphi_{l,2}}, \dots, e^{j(N_{\text{IRS}}-1)\pi\varphi_{l,m}} \right]$ , and  $q(f, d_{l,m}) = \frac{c}{4\pi f d_{l,m}} e^{-\frac{1}{2}\tau(f)d_{l,m}}$ , where  $d_{l,m}$  denotes the distance from the IRS to the  $m$ -th user in the  $l$ -th cluster.

Similarly, we can obtain

$$\mathbf{g}_E = q(f, d_E) \bar{\mathbf{g}}_E, \quad (7)$$



(a) Fully-connected architecture



(b) Sub-connected HP architecture

Fig. 2. Sparse RF chain antenna structures.

where  $\bar{\mathbf{g}}_E = \frac{1}{\sqrt{N_{\text{IRS}}}} [1, e^{j\pi\varphi}, e^{j2\pi\varphi}, \dots, e^{j(N_{\text{IRS}}-1)\pi\varphi}]$ ,  $q(f, d_E) = \frac{c}{4\pi f d_E} e^{-\frac{1}{2}\tau(f)d_E}$ , and  $d_E$  represents the distance between IRS and Eve.

For convenience, we set  $\beta_{l,m} = \eta G_t G_r q(f, d) q(f, d_{l,m})$  and  $\beta_E = \eta G_t G_r q(f, d) q(f, d_E)$ , where  $\eta$  denotes the path compensation factor. According to (3), (6) and (7), (1) and (2) can be rewritten as

$$y_{l,m} = \mathbf{G}_{l,m} (\sum_{j=1}^m \sqrt{p_{l,j}} \mathbf{F} \mathbf{v}_l x_{l,j} + \sum_{i \neq l} \sum_{j=1}^M \sqrt{p_{i,j}} \mathbf{F} \mathbf{v}_i x_{i,j}) + n_{l,m}, \forall l, m, \quad (8)$$

$$y_{l,m}^E = \mathbf{G}_E \sum_{i=1}^L \sum_{j=1}^M \sqrt{p_{i,j}} \mathbf{F} \mathbf{v}_i x_{i,j} + v_{l,m}, \forall l, m, \quad (9)$$

where  $\mathbf{G}_{l,m} = \beta_{l,m} \bar{\mathbf{g}}_{l,m} \mathbf{\Theta} \bar{\mathbf{H}}$  with  $\mathbf{G}_{l,m} \in \mathbb{C}^{1 \times N_{\text{TX}}}$ , and  $\mathbf{G}_E = \beta_E \bar{\mathbf{g}}_E \mathbf{\Theta} \bar{\mathbf{H}}$  with  $\mathbf{G}_E \in \mathbb{C}^{1 \times N_{\text{TX}}}$ .

### A. Analog Beamforming

Fig. 2 shows the architecture of the studied THz MIMO system. As shown in Fig. 2 (a), there are  $N_{\text{TX}} \times N_{\text{RF}}$  phase shifters in the FC HP architecture.  $\mathbf{F} \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{RF}}}$  represents the analog precoding matrix implemented by equal power dividers and phase shifters. Then,  $\mathbf{F}^{(full)}$  can be given by

$$\mathbf{F}^{(full)} = [\mathbf{f}_1^{(full)}, \mathbf{f}_2^{(full)}, \mathbf{f}_q^{(full)}, \dots, \mathbf{f}_{N_{\text{RF}}}^{(full)}], \quad (10)$$

where  $\mathbf{f}_q \in \mathbb{C}^{N_{\text{TX}} \times 1}$ ,  $\forall q = \{1, 2, \dots, N_{\text{RF}}\}$  is the analog precoding vector associated with the  $q$ -th RF chain. Moreover, Fig. 2 (b) is the SC HP architecture. Without loss of generality, we assume that  $N_{\text{SF}} = N_{\text{TX}}/N_{\text{RF}}$  is an integer number, and each RF chain is connected to  $N_{\text{SF}}$  phase shifters [29]. For the SC architecture, the precoding matrix  $\mathbf{F}^{(sub)}$  is

$$\mathbf{F}^{(sub)} = \begin{bmatrix} \mathbf{f}_1^{(sub)} & 0 & \dots & 0 \\ 0 & \mathbf{f}_2^{(sub)} & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \mathbf{f}_{N_{\text{RF}}}^{(sub)} \end{bmatrix}. \quad (11)$$

Compared with the FC structure, the SC structure reduces the hardware cost and power consumption at the expense of reduced array gain [32].

For practical implementation, we consider a  $B$  bit quantization phase shifter [23]. The FC analog non-zero element precoding matrix  $\mathbf{F}^{(full)}$  can be given by

$$\mathbf{F}^{(full)} = 1/\sqrt{N_{\text{TX}}} \{e^{j\frac{2\pi\beta}{2^B}} : \beta = 0, 1, \dots, 2^B - 1\}. \quad (12)$$

where  $\mathbf{f}_l^{(full)}(q) = \frac{1}{\sqrt{N_{\text{BS}}}} e^{j\frac{2\pi\hat{\beta}}{2^B}}$ . Similarly, the SC analog precoding matrix  $\mathbf{F}^{(sub)}$  can be written as

$$\mathbf{F}^{(sub)} = 1/\sqrt{N_{\text{SF}}} \{e^{j\frac{2\pi\beta}{2^B}} : \beta = 0, 1, \dots, 2^B - 1\}. \quad (13)$$

In (13),  $\mathbf{f}_l^{(sub)}(q) = \frac{1}{\sqrt{N_{\text{SF}}}} e^{j\frac{2\pi\hat{\beta}}{2^B}}$ , ( $q = (L-1)N_{\text{SF}} + 1, (L-1)N_{\text{SF}} + 2, \dots, LN_{\text{SF}}$ ) represents the  $q$ -th element of SC analog precoding vector, where  $\hat{\beta} = \arg \min |angle(\mathbf{g}(l)) - \frac{2\pi\beta}{2^B}|$ , and  $\mathbf{g}(l)$  is the channel state information for  $l$ -th cluster head.

### B. Digital Precoding

Since each cluster contains multiple users, it will cause inter-cluster interference [33], we adopt the low-complexity ZF method to design the digital precoding and eliminate interference among cluster head users [34]. Assuming that equivalent channel gain of all cluster head users is

$$\mathbf{G} = [\mathbf{g}_{1,m}, \mathbf{g}_{2,m}, \dots, \mathbf{g}_{L,m}]. \quad (14)$$

Then, we can generate the ZF precoding matrix as

$$\bar{\mathbf{V}} = [\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_L] = \mathbf{G}(\mathbf{G}^H \mathbf{G})^{-1}, \quad (15)$$

After normalization, the digital precoding vector of the  $l$ -th beam can be written as  $\mathbf{v}_l = \frac{\bar{\mathbf{v}}_l}{\|\bar{\mathbf{v}}_l\|_2}$ .

### C. Secrecy Rate

Upon performing the HP of two antenna structures, the SINR of the  $m$ -th user of the  $l$ -th cluster and Eve can be expressed as

$$\gamma_{l,m} = \frac{|\mathbf{G}_{l,m} \mathbf{F} \mathbf{v}_l|^2 p_{l,m}}{|\mathbf{G}_{l,m} \mathbf{F} \mathbf{v}_l|^2 \sum_j^{m-1} p_{l,j} + \sum_{i \neq l} |\mathbf{G}_{l,m} \mathbf{F} \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2}, \quad (16)$$

$$\gamma_{l,m}^E = \frac{|\mathbf{G}_E \mathbf{F} \mathbf{v}_l|^2 p_{l,m}}{|\mathbf{G}_E \mathbf{F} \mathbf{v}_l|^2 \sum_{j \neq m} p_{l,j} + \sum_{i \neq l} |\mathbf{G}_E \mathbf{F} \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2}, \quad (17)$$

Then, the secrecy rate of the  $m$ -th user of the  $l$ -th cluster can be expressed as

$$R_{l,m}^{sec} = \log_2(1 + \gamma_{l,m}) - \log_2(1 + \gamma_{l,m}^E). \quad (18)$$

#### D. Problem Formulation

Our goal is maximizing the sum secrecy rate by jointly optimize the power allocation and the phase-shift matrix, the optimization problem is formulated as

$$\max_{p_{l,m}, \Theta} \sum_{l=1}^L \sum_{m=1}^M R_{l,m}^{sec} \quad (19a)$$

$$\text{s.t.} \quad \sum_{l=1}^L \sum_{m=1}^M p_{l,m} \leq P_T, \quad (19b)$$

$$\log_2(1 + \gamma_{l,m}) \geq R_{l,m}^{\min}, \quad (19c)$$

$$|\theta_n| = 1, \forall n \in N_{\text{IRS}}. \quad (19d)$$

where  $P_T$  stands for the transmit power of the BS, (19b) represents the power constraint of the system, (19c) indicates the rate requirement of each user, and (19d) represents the unit modulus constraint of each IRS element. Problem (19) is a non-convex optimization problem and intractable to solve due to the coupled variables. To overcome this issue, we decomposed (19) into two sub-problems. For each sub-problem, we only focus on some specific variables when the others are fixed, which will be presented in the section III.

### III. JOINT OPTIMIZE THE POWER ALLOCATION AND THE PHASE SHIFTS MATRIX

In this section, an efficient AO algorithm is utilized to solve problem (19). Specifically, we firstly decouple the origan problem into two sub-problems. Then, the first-order Taylor expansion and SDR are leveraged to transform each sub-problem into a convex counterpart. Finally, the local optimal solution can be obtained through the AO algorithm.

#### A. Optimal the Power Allocation

To overcome the above issue, we assume that the order of the effective channel gain satisfies  $|\overline{\mathbf{G}}_{l,1}|^2 \geq |\overline{\mathbf{G}}_{l,2}|^2 \geq \dots \geq |\overline{\mathbf{G}}_{l,m}|^2$ , i.e. the SIC decoding sequence of NOMA users. Moreover, defining  $\mathbf{G}_{l,m} \mathbf{F} \triangleq \overline{\mathbf{G}}_{l,m}$  and  $\mathbf{G}_{l,m}^E \mathbf{F} \triangleq \overline{\mathbf{G}}_{l,m}^E$ , the (18) can be rewritten as

$$\overline{R}_{l,m}^{sec} = \log_2(A_1) - \log_2(B_1) - \log_2(C_1) + \log_2(D_1), \quad (20)$$

where  $B_1$  can be given by

$$B_1 = |\overline{\mathbf{G}}_{l,m} \mathbf{v}_l|^2 \sum_j^{m-1} p_{l,j} + \sum_{i \neq l} |\overline{\mathbf{G}}_{l,m} \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2, \quad (21)$$

Then,  $A_1$  can be expressed as

$$\begin{aligned} A_1 &= B_1 + |\overline{\mathbf{G}}_{l,m} \mathbf{v}_l|^2 p_{l,m} \\ &= |\overline{\mathbf{G}}_{l,m} \mathbf{v}_l|^2 \sum_j^m p_{l,j} + \sum_{i \neq l} |\overline{\mathbf{G}}_{l,m} \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2, \end{aligned} \quad (22)$$

Similarly,  $D_1$  can be given by

$$D_1 = |\overline{\mathbf{G}}_{l,m}^E \mathbf{v}_l|^2 \sum_{j \neq m} p_{l,j} + \sum_{i \neq l} |\overline{\mathbf{G}}_{l,m}^E \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2, \quad (23)$$

Further, we can express  $C_1$  as

$$C_1 = D_1 + |\overline{\mathbf{G}}_{l,m}^E \mathbf{v}_l|^2 p_{l,m} = \sum_{i=l}^L |\overline{\mathbf{G}}_{l,m}^E \mathbf{v}_i|^2 \sum_{j=1}^M p_{i,j} + \sigma^2. \quad (24)$$

Due to the existence of logarithmic term  $-\log_2(B_1) - \log_2(C_1)$ , problem (20) is still non-convex [35]. To deal with this problem, we apply the first-order Taylor expansion on the point  $\bar{p}_{l,m}$  [36]. For convenience, we set  $\overline{B}_1 = |\overline{\mathbf{G}}_{l,m} \mathbf{v}_l|^2 \sum_j^{m-1} \bar{p}_{l,j} + \sum_{i \neq l} |\overline{\mathbf{G}}_{l,m} \mathbf{v}_i|^2 \sum_{j=1}^M \bar{p}_{i,j} + \sigma^2$ , and  $\overline{C}_1 = \sum_{i=l}^L |\overline{\mathbf{G}}_{l,m}^E \mathbf{v}_i|^2 \sum_{j=1}^M \bar{p}_{i,j} + \sigma^2$ . Then, we can transform  $-\log_2(B_1)$  and  $-\log_2(C_1)$  into convex forms, respectively, as

$$\log_2(B_1) \leq \log_2(\overline{B}_1) + \frac{\overline{\mathbf{G}}_{l,m} \mathbf{\Omega} \overline{\mathbf{G}}_{l,m}^H}{\overline{B}_1 \ln 2}, \quad (25)$$

$$\log_2(C_1) \leq \log_2(\overline{C}_1) + \frac{\overline{\mathbf{G}}_{l,m}^E \mathbf{\Psi} \overline{\mathbf{G}}_{l,m}^E H}{\overline{C}_1 \ln 2}, \quad (26)$$

where  $\mathbf{\Omega} = |\mathbf{v}_l|^2 \sum_{j=1}^{m-1} (p_{i,j} - \bar{p}_{i,j}) + \sum_{i \neq l} |\mathbf{v}_i|^2 \sum_{j=1}^M (p_{i,j} - \bar{p}_{i,j})$  and  $\mathbf{\Psi} = \sum_{i=1}^L |\mathbf{v}_i|^2 \sum_{j=1}^M (p_{i,j} - \bar{p}_{i,j})$ . Obviously, (25) and (26) are both linear. According to the above derivation, the constraint (19c) can be transformed as

$$p_{l,m} |\mathbf{v}_l|^2 \overline{\mathbf{G}}_{l,m} \overline{\mathbf{G}}_{l,m}^H - \left(2^{R_{l,m}^{\min}} - 1\right) B_1 \geq 0. \quad (27)$$

Finally, the problem (19) can be recast as

$$\max_{p_{l,m}, \Theta} \sum_{l=1}^L \sum_{m=1}^M R_{l,m}^{sec} \quad (28a)$$

$$\text{s.t.} \quad (19b), (27). \quad (28b)$$

With the given  $\Theta$ , the problem (28) can be computed by applying convex problem solvers such as CVX toolbox [37]. The algorithm of optimizing power allocation is summarized to algorithm 1. However, since the approximation algorithm may cause the problem to diverge. Then, we assume that  $p_{l,m}^{(t_1)}$  is the optimal solution for the  $t_1$  th iteration, and the convergence of algorithm 1 is proved in the following theorem.

**Theorem 1:** Defining  $f(t_1)$  as the objective value of the  $t_1$ -th iteration, we have  $f(t_1) \geq f(t_1 - 1)$ . Moreover, defining  $\|\Delta \ell\| \triangleq f(t_1) - f(t_1 - 1)$ , the value of  $\|\Delta \ell\|$  infinitely close to 0 with the  $t_1$  increases, i.e., the algorithm converges.

*Proof:* See Appendix A. ■

**Algorithm 1** : The optimization algorithm for the power allocation.

- 1: Initialize the iteration number  $t_1 = 0$  and give a feasible solution  $p_{l,m}^{(t_0)}$  for the problem (28).
- 2: **repeat**
- 3: Leveraging the  $p_{l,m}^{(t_0)}$  to obtain a optimal solution  $p_{l,m}^{(t_1)}$  for problem (28).
- 4: Update the feasible solution  $p_{l,m}^{(t_1)} \leftarrow p_{l,m}^{(t_1-1)}$ .
- 5: Update the iteration number  $t_1 \leftarrow t_1 - 1$ .
- 6: **until**  $\left| \log_2 \left( B_1^{(t_1)} \right) - \log_2 \left( B_1^{(t_1-1)} \right) \right| \leq \xi$  and  $\left| \log_2 \left( C_1^{(t_1)} \right) - \log_2 \left( C_1^{(t_1-1)} \right) \right| \leq \xi$ .

### B. Solving the Phase Shifts of IRS

With the obtained power allocation  $p_{l,m}$ , the original problem (19) can be transformed a feasibility check problem. By denoting  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{N_{\text{IRS}}}]^T$ , the equivalent forms of the cascaded channels can be respectively expressed as

$$\bar{\mathbf{g}}_{l,m} \boldsymbol{\Theta} \bar{\mathbf{H}} = \boldsymbol{\theta}^H \text{diag}(\bar{\mathbf{g}}_{l,m}) \bar{\mathbf{H}} = \boldsymbol{\theta}^H \tilde{\mathbf{G}}_{l,m}, \quad (29)$$

$$\bar{\mathbf{g}}_{l,m}^E \boldsymbol{\Theta} \bar{\mathbf{H}} = \boldsymbol{\theta}^H \text{diag}(\bar{\mathbf{g}}_{l,m}^E) \bar{\mathbf{H}} = \boldsymbol{\theta}^H \tilde{\mathbf{G}}_{l,m}^E, \quad (30)$$

Then, we reconsider the phase shift matrix  $\boldsymbol{\Theta}$  and transform the secrecy rate into the following form

$$\hat{R}_{l,m}^{\text{sec}} = \log_2(A_2) + \log_2(D_2) - \log_2(B_2) - \log_2(C_2). \quad (31)$$

For the non-convex objective function, we denote  $\boldsymbol{\Phi} \triangleq \boldsymbol{\theta} \boldsymbol{\theta}^H$ ,  $\mathbf{V} \triangleq \mathbf{v} \mathbf{v}^H$ ,  $\hat{\mathbf{G}}_{l,m} \triangleq \beta_{l,m} \tilde{\mathbf{G}}_{l,m} \mathbf{F}$  and  $\hat{\mathbf{G}}_{l,m}^E \triangleq \beta_E \tilde{\mathbf{G}}_{l,m}^E \mathbf{F}$ . As such, the  $B_2$  in (31) can be expressed as

$$\begin{aligned} B_2 &= \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m} \mathbf{v}_l \right|^2 \sum_j^{m-1} p_{l,j} + \sum_{i \neq l} \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m} \mathbf{v}_i \right|^2 \sum_{j=1}^M p_{i,j} + \sigma^2 \\ &\stackrel{(b)}{=} \text{Tr}(\hat{\mathbf{G}}_{l,m} (\mathbf{V}_l \sum_{j=1}^{m-1} p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^H \boldsymbol{\Phi}) + \sigma^2, \end{aligned} \quad (32)$$

where  $\hat{\mathbf{G}}_{l,m} \triangleq \beta_{l,m} \tilde{\mathbf{G}}_{l,m} \mathbf{F}$ , and (b) is obtained by invoking the identity  $\mathbf{a}^H \mathbf{T} \mathbf{a} = \text{Tr}(\mathbf{T} \mathbf{a} \mathbf{a}^H)$ . Then,  $A_2$  can be written as

$$\begin{aligned} A_2 &= B_2 + \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m} \right|^2 \mathbf{V}_l p_{l,m} \\ &\stackrel{(b)}{=} \text{Tr}(\hat{\mathbf{G}}_{l,m} (\mathbf{V}_l \sum_{j=1}^m p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^H \boldsymbol{\Phi}) + \sigma^2, \end{aligned} \quad (33)$$

According to the previous method,  $D_2$  can be rewritten as

$$\begin{aligned} D_2 &= \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m}^E \mathbf{v}_l \right|^2 \sum_{j \neq m} p_{l,j} + \sum_{i \neq l} \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m}^E \mathbf{v}_i \right|^2 \sum_{j=1}^M p_{i,j} + \sigma^2 \\ &\stackrel{(b)}{=} \text{Tr}(\hat{\mathbf{G}}_{l,m}^E (\mathbf{V}_l \sum_{j \neq m} p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^{E H} \boldsymbol{\Phi}) + \sigma^2, \end{aligned} \quad (34)$$

and  $C_2$  can be expressed as

$$\begin{aligned} C_2 &= D_2 + \left| \boldsymbol{\theta}^H \hat{\mathbf{G}}_{l,m}^E \mathbf{v}_l \right|^2 p_{l,m} \\ &\stackrel{(b)}{=} \text{Tr}(\hat{\mathbf{G}}_{l,m}^E (\mathbf{V}_l \sum_{j=1}^m p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^{E H} \boldsymbol{\Phi}) + \sigma^2. \end{aligned} \quad (35)$$

Similar approximation method previously adopted can be

utilized for the non-convex item  $-\log_2(B_2) - \log_2(C_2)$  in (31), the first-order Taylor expansion is leveraged on the point  $\hat{\boldsymbol{\Phi}}$  [35]. For implicity, we set

$$\hat{B}_2 = \text{Tr}(\hat{\mathbf{G}}_{l,m} (\mathbf{V}_l \sum_{j=1}^{m-1} p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^H \hat{\boldsymbol{\Phi}}) + \sigma^2, \quad (36)$$

$$\hat{C}_2 = \text{Tr}(\hat{\mathbf{G}}_{l,m}^E (\mathbf{V}_l \sum_{j=1}^m p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}) \hat{\mathbf{G}}_{l,m}^{E H} \hat{\boldsymbol{\Phi}}) + \sigma^2 \quad (37)$$

Then,  $-\log_2(B_2)$  and  $-\log_2(C_2)$  can be transformed into convex forms, respectively, as

$$\log_2(B_2) \leq \log_2(\hat{B}_2) + \frac{\text{Tr}(\hat{\mathbf{G}}_{l,m} \hat{\boldsymbol{\Omega}} \hat{\mathbf{G}}_{l,m}^H (\boldsymbol{\Phi} - \hat{\boldsymbol{\Phi}}))}{\hat{B}_2 \ln 2}, \quad (38)$$

$$\log_2(C_2) \leq \log_2(\hat{C}_2) + \frac{\text{Tr}(\hat{\mathbf{G}}_{l,m}^E \hat{\boldsymbol{\Psi}} \hat{\mathbf{G}}_{l,m}^{E H} (\boldsymbol{\Phi} - \hat{\boldsymbol{\Phi}}))}{\hat{C}_2 \ln 2}. \quad (39)$$

where  $\hat{\boldsymbol{\Omega}} = \mathbf{V}_l \sum_{j=1}^m p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}$  and  $\hat{\boldsymbol{\Psi}} = \mathbf{V}_l \sum_{j=1}^{m-1} p_{l,j} + \sum_{i \neq l} \mathbf{V}_i \sum_{j=1}^M p_{i,j}$ . Then, we plug (38) and (39) into constraint (31), and transform constraint (19c) into convex form as

$$p_{l,m} \text{Tr}(\hat{\mathbf{G}}_{l,m} \mathbf{V}_l \hat{\mathbf{G}}_{l,m}^H \boldsymbol{\Phi}) - (2^{R_{l,m}^{\text{min}}} - 1) B_2 \geq 0. \quad (40)$$

As a result, the passive beamforming optimized can be recast as

$$\max_{\boldsymbol{\Phi}} \sum_{l=1}^L \sum_{m=1}^M R_{l,m}^{\text{sec}} \quad (41a)$$

$$\text{s.t. } \boldsymbol{\Phi}_{n,n} = 1, \forall n \in N_{\text{IRS}}, \quad (41b)$$

$$\text{rank}(\boldsymbol{\Phi}) = 1, \boldsymbol{\Phi} \succeq 0, \quad (41c)$$

$$(36). \quad (41d)$$

However, due to the constraint  $\text{rank}(\boldsymbol{\Phi}) = 1$ , it is difficult to directly to address it by CVX. Therefore, we neglect the rank-one constraint via SDR [38], and the problem (41) can be rewritten as

$$\max_{\boldsymbol{\Phi}} \sum_{l=1}^L \sum_{m=1}^M R_{l,m}^{\text{sec}} \quad (42a)$$

$$\text{s.t. } \boldsymbol{\Phi} \succeq 0, \quad (42b)$$

$$(36), (37b). \quad (42c)$$

It is clear that problem (42) is a standard SDP, which can be efficiently solved by applying the CVX toolbox [37]. Generally speaking, after ignoring the  $\text{rank}(\boldsymbol{\Phi}) = 1$ , the solution calculated with eigenvalue decomposition for  $\boldsymbol{\Phi}$  does not necessarily satisfy rank one. Therefore, the optimal result of the problem (42) only serves as an upper bound of (41) [39]. In order to construct a rank one solution from the optimal higher-order solution of problem (42), singular value decomposition is performed on  $\boldsymbol{\Phi} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H$ , where  $\mathbf{U} = [e_1, \dots, e_N]$  and  $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \dots, \lambda_N)$  are a unitary matrix and a diagonal matrix respectively. Then, the suboptimal solution  $\boldsymbol{\theta}^* =$

**Algorithm 2:** The overall AO algorithm

- 
- 1: **Initialize:** reflection matrix  $\Theta^0$ , iteration number  $t = 1$  and maximum iteration number  $T$ .
  - 2: Obtain the analog beamforming  $\mathbf{F}^{(full)}$  and  $\mathbf{F}^{(sub)}$ . Then, the digital beamforming  $\mathbf{V}$  is designed by the ZF precoding.
  - 3: **repeat**
  - 4:   **repeat**
  - 5:     Given  $\Theta$ , solve the problem (28) and obtain the solution  $p_{l,m}^{(t_1)}$ .
  - 6:     Update  $t_1 \leftarrow t_1 + 1$ .
  - 7:     Update  $p_{l,m}^{(t_1)} \leftarrow p_{l,m}^{(t_1)*}$ .
  - 8:     **until**  $t_1 = T_1$  or Convergence;
  - 9:     **repeat**
  - 10:      with optimized  $p_{l,m}^*$ , the solution  $\Phi^{(t_2)}$  of the problem (42) can be obtained.
  - 11:      Update  $t_2 \leftarrow t_2 + 1$ .
  - 12:      Update  $\Phi^{(t_2)} \leftarrow \Phi^{(t_2)*}$ .
  - 13:      **until**  $t_2 = T_2$  or Convergence;
  - 14:   **until**  $t = T$  or Convergence;
- Maximize the achievable sum rate via the obtained  $p_{l,m}^*$  and  $\Phi^*$ .
- 

$U\Sigma^{1/2}\mathbf{r}$  of (41) is obtained, where  $\mathbf{r} \in \mathbb{C}^{(N) \times 1}$  is a random vector, each element of which is a Gaussian random variable with zero mean and unit variance, i.e.,  $\mathbf{r} \sim \mathcal{CN}(0, \mathbf{I}_N)$ . Using the independently generated Gaussian random vector  $\mathbf{r}$ , the objective value of (42) is approximated to the maximum value for all the optimal  $\theta$  [40]. Finally, we can obtain  $\theta$  by

$$\theta_n = e^{j\angle\left(\left[\frac{\theta_n}{\theta_{n+1}}\right]\right)}, n = 1, 2, \dots, N. \quad (43)$$

In this case, the SDR method combined with the Gaussian randomization can recover a high-quality feasible solution of problem. Finally, we summarize the proposed optimization scheme in Algorithm 2.

### C. Computational Complexity

In this section, we will evaluate the computational complexity of the proposed algorithm without Gaussian randomization [41]. For the sub-problem (28), the iteration number is in the order of  $\sqrt{2LM}$ , then, there are  $LM$  linear matrix inequality (LMI) constraints of size 1. Therefore, the complexity of problem (28) is in the order of  $\sqrt{2LM}(LM(LM) + (LM)^2(LM) + (LM)^3)$ . Also, for the sub-problem (38), the iteration number is  $\sqrt{LM + 2N_{\text{IRS}} + 1}$ , there are  $N_{\text{IRS}}$  LMI constraints of size  $N_{\text{IRS}}$ , and  $(LM + N_{\text{IRS}} + 1)$  LMI constraints of size 1. As such, the complexity of solving problem (38) is in the order of  $\sqrt{2N_{\text{IRS}} + LM + 1}(N_{\text{IRS}}^2(N_{\text{IRS}}^3 + LM + N_{\text{IRS}} + 1) + N_{\text{IRS}}^4(N_{\text{IRS}}^2 + LM + N_{\text{IRS}} + 1) + N_{\text{IRS}}^6)$ .

According to the above analysis, the total complexity is  $\mathcal{O}[\sqrt{2N_{\text{IRS}} + LM} + 12(N_{\text{IRS}}^6 + 2N_{\text{IRS}}^5 + N_{\text{IRS}}^3 + (N_{\text{IRS}}^4 + N_{\text{IRS}}^2)(LM + 1)) + \sqrt{2LM}(2(LM)^3 + LM^2)]$ .

TABLE I

Parameters	Value
Number of the BS antennas	$N_{\text{TX}} = 32$
Number of the IRS elements	$N_{\text{IRS}} = 20$
Number of the RF chains/clusters	$N_{\text{RF}} = 4/L = 4$
Number of the users/Eave	$M = 6/K = 1$
Central frequency	$f = 340$ GHz
The bits quantized of phase shifters	$B = 4$
Transmit antenna gain	$G_t = 1$
Receive antenna gain	$G_r = 4 + 20\log_{10}\sqrt{N_{\text{TX}}}$
Absorption coefficient	$\tau = 0.0033/\text{m}$
Noise variance	$\sigma^2 = 0.01\text{W}$
Number of phase shifters of FC	$N_{\text{PSF}} = N_{\text{TX}}N_{\text{RF}}$
Number of phase shifters of SC	$N_{\text{PSS}} = N_{\text{TX}}$

## IV. SIMULATION RESULTS

In this section, we illustrate the performance of the proposed IRS-aided THz MIMO-NOMA system. Due to the serious attenuation of THz propagation and absorption by water molecules, the scenario of short-distance communication is considered. We set the distances from BS to IRS and Eve are 15m and 5m, respectively. Moreover, the distance from IRS to all users and Eve is 5m, where users of per-cluster are located within a circle with 2m radius. In the simulations,  $M$  users are grouped into  $N_{\text{RF}}$  beams, and there are more than one user in each group [23]. According to [42], the SEE is defined as the ratio between the achievable SSR and the total power consumption

$$SEE = \frac{\sum_{l=1}^L \sum_{m=1}^M R_{l,m}^{sec}}{\sum_{l=1}^L \sum_{m=1}^M p_{l,m} + P_{\text{RF}}N_{\text{RF}} + P_{\text{PS}}N_{\text{PS}} + P_{\text{B}}}. \quad (44)$$

where  $P_{\text{RF}}$  is the power consumed by each RF chain,  $P_{\text{PS}}$  denotes the power consumption of each phase shifter, and  $P_{\text{B}}$  is the baseband power consumption. We set the typical values  $P_{\text{RF}} = 300$  mW,  $P_{\text{PS}} = 40$  mW (4-bit phase shifter), and baseband power  $P_{\text{B}} = 200$  mW [42].  $N_{\text{PSF}}$  and  $N_{\text{PSS}}$  are the number of phase shifters of FC and SC, respectively. Other simulation parameters are listed in TABLE I.

In addition, the following two benchmark schemes are compared with the proposed continuous phase shifts (CPS) scheme.

1. Discrete phase shifts (DPS): The initial IRS phase shifts are constructed by phase quantization, assuming that  $\theta_n$  has  $\kappa$  discrete values, which are arranging evenly on the circle  $\theta = e^{j\theta_n}$  [43]. Then, we have

$$\theta_n = \left\{ e^{j\frac{2\pi\kappa}{2^L}} : \kappa = 0, 1, \dots, 2^L - 1 \right\}. \quad (45)$$

where  $L$  denotes the number of discrete binary levels.

2. Random phase shifts (RPS): Only IRS is deployed, and the phase shifts of reflection units are not optimized.

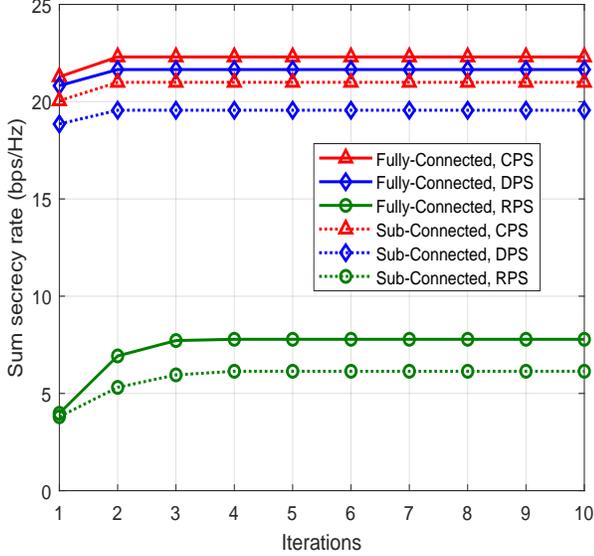


Fig. 3. Sum secrecy rate versus the iteration numbers.

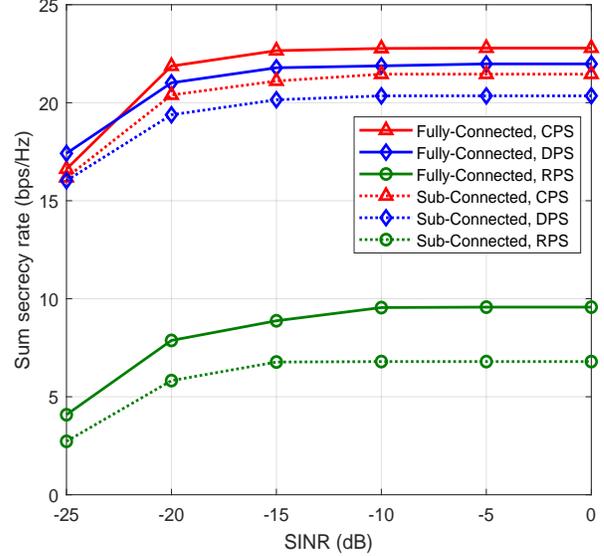


Fig. 5. Sum secrecy rate versus the SINR of all users.

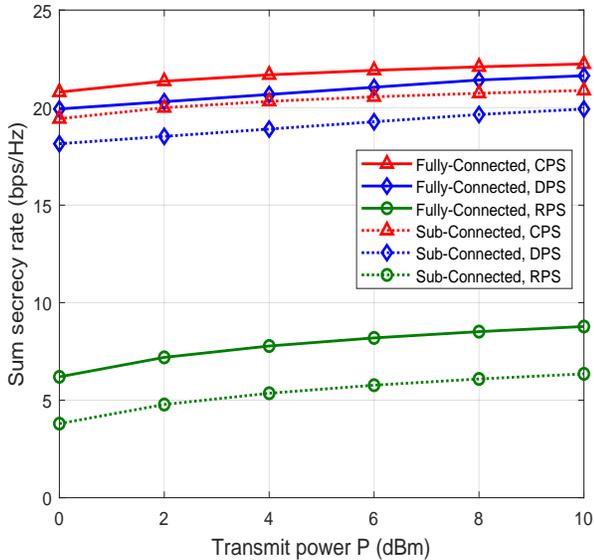


Fig. 4. Sum secrecy rate versus the transmit power.

Fig. 3 illustrates the convergence performance of the proposed algorithm and benchmark schemes with the minimal SINR of the users, i.e.,  $R_{l,m}^{\min} = 0$  dB. It can be seen that all schemes can achieve quickly convergence, and the convergence speed of FC structure is the same as SC structure. In addition, the iteration numbers of the RPS scheme is equal to inner loop iterations of the proposed method, which takes more than 5 iterations to reach convergence. On the other hand, we can observe that the SSR of the FC structure is higher than that of the SC structure under the same conditions, which can be explained by the fact the full array gain can be utilized by each RF chain in the FC structure.

We compare the performance of two HP structure with

respect to the BS transmit power in Fig. 4. As observed, the SSR of FC and SC structures increase monotonically with the increase of the transmit power. With the same HP structure, the achievable SSR after optimized IRS reflection phase shifts are significantly better than that of RPS schemes. This is due to the fact that the nature of IRS is similar to multi-antenna technology, which can bring new spatial degrees of freedom (DoF) and diversity gain to the system. Optimizing the phase shifts of IRS can converge the energy of multi-path signals and reflect it to the legitimate receivers, and suppress the eavesdropping, thus effectively improving the secrecy rate of the system. What is more, in the practical application of IRS, discrete phase has more engineering significance and cost-effectiveness. However, there is a certain phase ambiguity at the IRS in the process of reflecting signals, which leads to the partial loss of performance. Therefore, the achievable SSR of DPS scheme is slightly lower than that of CPS.

Fig. 5 describes the relationship between the SSR and the minimal SINR requirement of users. As clearly shown by the results, the SSR of all schemes increase monotonously with the SINR threshold. It is worth noting that the growth trend of the SSR gradually becomes gentle with the increase of the SNIR requirement due to the limitation of the total transmit power budget. The FC structure of each scheme can achieve better performance, this is due to fact that the all  $N_{RS}$  chains are used to serve for the users, which can bring full use of the multiplexing gain. In addition, with the same SNIR condition, the passive beamforming gain brought by optimizing the phase shifts of IRS elements makes the scheme of CPS and DPS significantly better than that of RPS. Furthermore, the performance of DPS scheme is slightly worse than that of RPS scheme.

In Fig. 6, we plot the achievable SSR against the number of users  $K$ . Obviously, the results indicate that the achievable SSR of all schemes increases with the number of users.

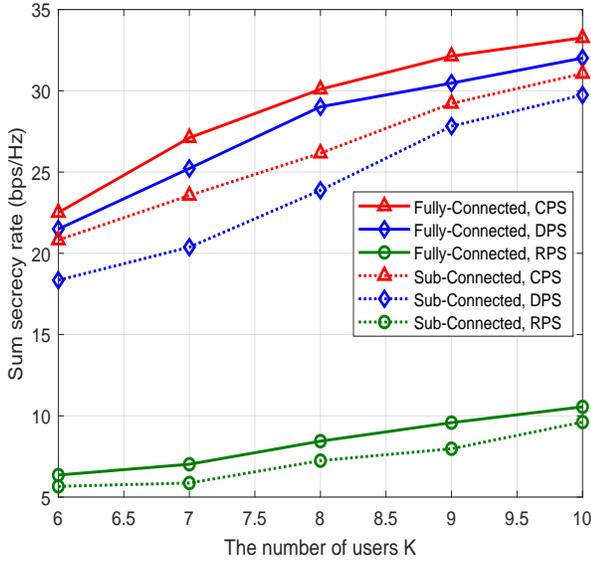


Fig. 6. Sum secrecy rate versus the the number of users  $K$ .

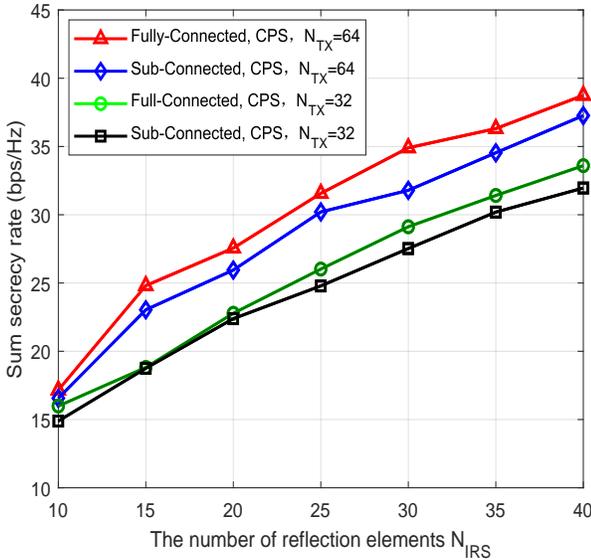


Fig. 7. Sum secrecy rate versus the number of IRS reflection elements.

However, the growth trend of the RPS scheme is gentler than that of the CPS and DPS. The reason is that optimizing the IRS reflection coefficients can enhance the directional beamforming gains of reflection. In addition, in the proposed NOMA network, the reflected signal from the IRS serves all users at the same time, which can produce partial multiplexing gains.

In Fig. 7, we show the achievable SSR against the different numbers of IRS reflection elements  $N_{IRS}$ . It is intuitive that a larger number of IRS reflection elements will result in a higher SSR. This is due to the fact that the proposed scheme can take full advantages of the IRS to improve the channel conditions, i.e. the additional spatial DoF and passive beamforming gain

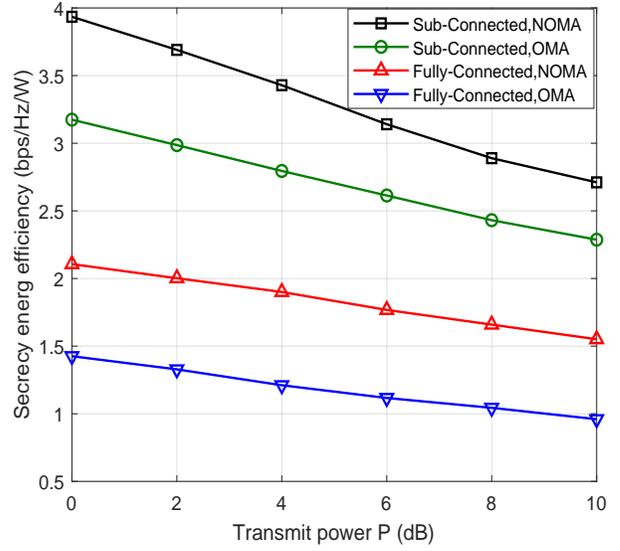


Fig. 8. Secrecy energy efficiency versus the transmit power.

provided by the IRS will gradually increase with the number of reflection elements. In addition, the results show that a higher beamforming gain can be achieved when there are more antennas at the BS. It is because that more antennas can further excavate the spatial dimension resources, so that multi-users can use the spatial freedom provided by MIMO to communicate with the BS on the same time frequency resources, thereby improving the secure spectrum efficiency without increasing the density of the BS.

We compare the SEE of NOMA with that of the conventional OMA scheme are provided in Fig. 8. According to the results, we can observe that SEE of all schemes is slowly degrading with the increases of transmit power. This can be explained by the fact that the power consumption for increasing the SSR of the system is larger than the benefits that it brings to the system, which causes the SEE of the system is reduced. Therefore, we should weigh the two performance indicators of SEE and secrecy rate as needed in practical. In addition, it can be verified that the proposed NOMA can obtain a higher SEE than OMA with the same HP structure. Meanwhile, the SC structure can achieve higher SEE than that of the FC structure. The reason is that the number of phase shifters used in SC structure is less than FC structure, which can save some power overhead.

## V. CONCLUSION

In this paper, a secure THz MIMO-NOMA system via IRS was studied, where the users were grouped into multiple clusters based on channel correlation, and each cluster is served by a beam through NOMA. Moreover, a HP scheme with both FC and SC architectures were considered, where analog precoding was designed based on discrete phase and digital precoding was designed by the ZF technique. We aimed to maximize the SSR by jointly optimizing power allocation and IRS phase shifts, while satisfying the constraints

of transmit power, achievable rate requirement of users, and constant modulus of IRS. However, it is challenging to solve this problem directly due to its non-convexity and coupled variables. To this end, the original problem was decomposed into two sub-problems, which can be converted into the convex form by SDR and the first-order Taylor expansion. Then, with the obtained analog and digital precoding, the feasible solution can be acquired via the AO algorithm. Finally, simulation results illustrated the proposed algorithm can converge quickly, and the passive beamforming of IRS can significantly improve security. In addition, FC can obtain the higher SSR than SC, but with a lower SEE. Thus, there was a compromise between SEE and the SSR of the proposed NOMA scheme.

The IRS integrated by passive components does not have the ability to process baseband signals, and cannot accurately obtain eavesdropping CSI. In addition, passive IRS can cause multiplicative fading effect, resulting in serious performance loss. The active IRS is equipped with a low-cost power amplifier, which is beneficial to obtain CSI and improve the reflection beamforming gain. Therefore, we will study the secure transmission of active IRS assisted THz systems in the future work.

#### APPENDIX A THE PROOF OF THEOREM 1

In the  $t_1$ -th iteration, we can obtain the following inequalities

$$\log_2 \left( B_1^{(t_1)} \right) \leq \log_2 \left( \bar{B}_1^{(t_1)} \right) + \frac{\text{Tr} \left( \bar{\mathbf{G}}_{l,m} \boldsymbol{\Omega}^{(t_1)} \bar{\mathbf{G}}_{l,m}^H \right)}{\bar{B}_1^{(t_1)} \ln 2}, \quad (46)$$

$$\log_2 \left( C_1^{(t_1)} \right) \leq \log_2 \left( \bar{C}_1^{(t_1)} \right) + \frac{\text{Tr} \left( \bar{\mathbf{G}}_{l,m}^E \boldsymbol{\Psi}^{(t_1)} \bar{\mathbf{G}}_{l,m}^E H \right)}{\bar{C}_1^{(t_1)} \ln 2}, \quad (47)$$

where

$$\boldsymbol{\Omega}^{(t_1)} = |\mathbf{v}_l|^2 \sum_{j=1}^{m-1} \left( p_{i,j}^{(t_1)} - p_{i,j}^{(t_1-1)} \right) + \sum_{i \neq l} |\mathbf{v}_i|^2 \sum_{j=1}^M \left( p_{i,j}^{(t_1)} - p_{i,j}^{(t_1-1)} \right). \quad (48)$$

and  $\boldsymbol{\Psi} = \sum_{i=1}^L |\mathbf{v}_i|^2 \sum_{j=1}^M \left( p_{i,j}^{(t_1)} - p_{i,j}^{(t_1-1)} \right)$ . Due to the linear approximation method is utilized in (45)-(47), the update rule in algorithm 1 can ensure the solution  $p_{l,m}^{(t_1)}$  is a optimal solution. As such, the inequality  $f(t_1) \geq f(t_1 - 1)$  holds. Algorithm 1 can generate a non-decreasing sequence of the objective value. The objective function of the problem (28) is bounded due to the total power budget constraint and the minimum achievable rate requirement of users. Therefore, algorithm 1 can guarantee convergence to a local optimal solution, i.e.,  $\|\Delta l\| = 0$  and  $f(t_1) = f(t_1 - 1)$ .

#### REFERENCES

- [1] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836–886, Apr. 2021.
- [2] S. Wang *et al.*, "Explainable AI for b5G/6G: Technical aspects, use cases, and research challenges." [Online] Available: <https://arXiv:2112.04698>.
- [3] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, "6G wireless channel measurements and models: Trends and challenges," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 22–32, Dec. 2020.
- [4] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, Jul. 2020.
- [5] X. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 9, no. 5, pp. 586–590, 2020.
- [6] G. M. Karam, M. Gruber, I. Adam, F. Boutigny, Y. Miche, and S. Mukherjee, "The evolution of networks and management in a 6G world: An inventor's view," *IEEE Trans. Netw. Serv.*, vol. 19, no. 4, pp. 5395–5407, Jul. 2022.
- [7] G. Liu, Y. Huang, N. Li, J. Dong, J. Jin, Q. Wang, and N. Li, "Vision, requirements and network architecture of 6G mobile network beyond 2030," *China Commun.*, vol. 17, no. 9, pp. 92–104, Sep. 2020.
- [8] H. Zhao, L. Zhang, S. Huang, S. Zhang, and C. Zhang, "Terahertz wave generation from noble gas plasmas induced by a wavelength-tunable femtosecond laser," *IEEE Trans. Terahertz Sci. Technol.*, vol. 8, no. 3, pp. 299–304, May 2018.
- [9] N. Zhao, F. R. Yu, M. Li, Q. Yan, and V. C. M. Leung, "Physical layer security issues in interference-alignment-based wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 162–168, Aug. 2016.
- [10] Z. Zhu, N. Wang, W. Hao, Z. Wang, and I. Lee, "Robust beamforming designs in secure MIMO SWIPT IoT networks with a nonlinear channel model," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1702–1715, Feb. 2021.
- [11] Z. Zhu *et al.*, "Resource allocation for intelligent reflecting surface assisted wireless powered IoT systems with power splitting," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 5, pp. 2987–2998, May 2022.
- [12] Z. Zhu, Z. Li, Z. Chu, Q. Wu, J. Liang, Y. Xiao, P. Liu, and I. Lee, "Intelligent reflecting surface-assisted wireless powered heterogeneous networks," *Accepted to IEEE Trans. Wirel. Commun.*, doi: 10.1109/TWC.2023.3274220.
- [13] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [14] Z. Chen, B. Ning, C. Han, Z. Tian, and S. Li, "Intelligent reflecting surface assisted terahertz communications toward 6G," *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 110–117, Dec. 2021.
- [15] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Sum-rate maximization for intelligent reflecting surface assisted terahertz communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 3320–3325, Mar. 2022.
- [16] Z. Zhu, J. Xu, G. Sun, W. Hao, Z. Chu, C. Pan, and I. Lee, "Robust beamforming design for IRS-aided secure SWIPT terahertz systems with non-linear EH model," *IEEE Wirel. Commun. Lett.*, vol. 11, no. 4, pp. 746–750, Apr. 2022.
- [17] J. Qiao, C. Zhang, A. Dong, J. Bian, and M.-S. Alouini, "Securing intelligent reflecting surface assisted terahertz systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8519–8533, Aug. 2022.
- [18] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684–2688, Feb. 2023.
- [19] X. Li, Q. Wang, M. Zeng, Y. Liu, S. Dang, T. A. Tsiftsis, and O. A. Dobre, "Physical-layer authentication for ambient backscatter-aided NOMA symbiotic systems," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2288–2303, Apr. 2023.
- [20] Y. Cao, N. Zhao, Y. Chen, M. Jin, Z. Ding, Y. Li, and F. R. Yu, "Secure transmission via beamforming optimization for NOMA networks," *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 193–199, Feb. 2020.
- [21] X. Li, Z. Xie, Z. Chu, V. G. Menon, S. Mumtaz, and J. Zhang, "Exploiting benefits of IRS in wireless powered NOMA networks," *IEEE Trans. Green Commun. Networking.*, vol. 6, no. 1, pp. 175–186, Mar. 2022.
- [22] X. Hu, C. Zhong, X. Chen, W. Xu, and Z. Zhang, "Cluster grouping and power control for angle-domain mmWave MIMO NOMA systems," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 1167–1180, 2019.
- [23] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.
- [24] J. Du, W. Xu, H. Shen, X. Dong, and C. Zhao, "Hybrid precoding architecture for massive multiuser MIMO with dissipation: Sub-connected or fully connected structures?," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 8, pp. 5465–5479, Aug. 2018.

- [25] S. Wan, H. Zhu, K. Kang, and H. Qian, "On the performance of fully-connected and sub-connected hybrid beamforming system," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11078–11082, Oct. 2021.
- [26] W. Hao, G. Sun, M. Zeng, Z. Chu, Z. Zhu, O. A. Dobre, and P. Xiao, "Robust design for intelligent reflecting surface-assisted MIMO-OFDMA terahertz IoT networks," *IEEE Internet of Things J.*, vol. 8, no. 16, pp. 13052–13064, Aug. 2021.
- [27] H. Sariahdeedeen, A. Abdallah, M. M. Mansour, M.-S. Alouini, and T. Y. Al-Naffouri, "Terahertz-band MIMO-NOMA: Adaptive superposition coding and subspace detection," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2628–2644, Dec. 2021.
- [28] X. Xu, Q. Chen, X. Mu, Y. Liu, and H. Jiang, "Graph-embedded multi-agent learning for smart reconfigurable THz MIMO-NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 259–275, Jan. 2022.
- [29] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, and V. C. M. Leung, "Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2074–2085, Sep. 2020.
- [30] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficient beamforming design for MISO non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4117–4131, Jun. 2019.
- [31] S. Priebe, M. Kannicht, M. Jacob, and T. Kürner, "Ultra broadband indoor channel measurements and calibrated ray tracing propagation modeling at THz frequencies," *J. Commun. Net.*, vol. 15, no. 6, pp. 547–558, Jun. 2013.
- [32] Z. Zhu, M. Ma, G. Sun, W. Hao, P. Liu, Z. Chu, and I. Lee, "Secrecy rate optimization in nonlinear energy harvesting model-based mmWave IoT systems with swipt," *IEEE Syst. J.*, vol. 16, no. 4, pp. 5939–5949, Dec. 2022.
- [33] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wirel. Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.
- [34] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.
- [35] F. Zhao, W. Hao, L. Shen, G. Sun, Y. Zhou, and Y. Wang, "Secure energy efficiency transmission for mmWave-NOMA system," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2226–2229, Jun. 2021.
- [36] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University press, 2004.
- [37] M. Grant and S. Boyd. (Sep. 2012), "CVX: MATLAB software for disciplined convex programming." [Online]. Available: <http://cvxr.com/cvx>.
- [38] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent reflecting surface aided multi-antenna secure transmission," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 1, pp. 108–112, Jan. 2020.
- [39] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
- [40] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2018.
- [41] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.
- [42] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [43] H. Niu, Z. Chu, F. Zhou, Z. Zhu, M. Zhang, and K.-K. Wong, "Weighted sum secrecy rate maximization using intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6170–6184, 2021.