Robust Design for Intelligent Reflecting Surface Assisted Secrecy SWIPT Network

Hehao Niu, Zheng Chu, *Member, IEEE*, Fuhui Zhou, *Senior Member, IEEE*, Zhengyu Zhu, *Member, IEEE*, Li Zhen, *Member, IEEE*, and Kai-Kit Wong, *Fellow, IEEE*

Abstract

This paper investigates the robust beamforming design in a secrecy multiple-input single-output (MISO) network aided by the intelligent reflecting surface (IRS) with simultaneous wireless information and power transfer (SWIPT). Specifically, by considering that the energy receivers (ERs) are potential eavesdroppers (Eves) and both imperfect channel state information (CSI) of the direct and cascaded channels can be obtained, we investigate the max-min fairness robust secrecy design. The objective is to maximize the minimum robust information rate among the legitimate information receivers (IRs). To solve the formulated non-convex design problem in bounded and probabilistic CSI error models, we utilize the alternating optimization (AO) and successive convex approximation (SCA) methods to obtain

This work was supported in part by the National Natural Science Foundation of China under Grant 61801434 and 61901490, in part by the Project funded by China Postdoctoral Science Foundation under Grant 2020M682345, in part by the Henan Postdoctoral Foundation under Grant 202001015.

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

Z. Chu is with the 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford GU2 7XH, U.K., (e-mail: andrew.chuzheng7@gmail.com).

F. Zhou is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210000, China (email: zhoufuhui@ieee.org).

Z. Zhu is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China, (e-mail: iezyzhu@zzu.edu.cn).

L. Zhen is with the Xi'an University of Posts and Telecommunications, Xi'an, 710121, China, (e-mail: lzhen@xupt.edu.cn).

K.-K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London, WC1E 6BT, U.K., (email: kai-kit.wong@ucl.ac.uk).

an approximate problem. Then, an iteration-based algorithm framework was proposed, where the unit modulus constraint (UMC) of the IRS is handled by the penalty dual decomposition (PDD) method. Moreover, a stochastic SCA method is proposed to handle the outage constrained design with statistical CSI. Finally, simulation results validate the promising performance of the proposed design.

Index Terms

Intelligent reflecting surface, alternating optimization, penalty dual decomposition.

I. Introduction

The fifth generation (5G) and beyond wireless networks are expected to meet an increasing requirement for wireless applications such as high spectral efficiency and low latency. Since battery capacity is limited, wireless devices powered by batteries have a short life and require frequent recharging [1]. The energy constrained problem has become a major bottleneck for next generation wireless networks. Energy harvesting (EH) techniques have been seen as an attractive solution to improve the energy efficiency (EE) in wireless networks.

One main EH method in wireless networks is the simultaneous wireless information and power transfer (SWIPT) technique, since the radio frequency signal can be used as an energy source [2]. Normally, the energy receiver (ER) is deployed closer to the transmitter (Tx) than the information receiver (IR) to help EH. However, when the ER tries to access the confidential information sent to the IR, the ER will become an eavesdropper (Eve). To overcome this security threat, physical layer security (PLS) techniques have been proposed to improve the secrecy performance in wireless networks [3]. Specifically, [4] investigated the use of secure beamforming (BF) in a multiple-input single-output (MISO)-SWIPT system with multiple Eves, where a semi-definite relaxation (SDR) based method was developed to obtain the sum rate maximization design. Also [5] studied the robust transmission design in a MISO-SWIPT system, where a secrecy energy efficiency (SEE) maximization problem was handled by the successive convex approximation (SCA) technique.

Recently, intelligent reflecting surface (IRS) has drawn great attention in industry and academia. An IRS is a planar array with many reflecting elements, which can rotate the phases of the incident electromagnetic (EM) wave passively [6]. Therefore, by altering the phase shifters with a controller, the reflected signals can be adjusted to a desired direction [7]. Moreover, since only reflecting the received signal with no decoding and encoding operation, IRS consumes much less power than active Tx or relay, and can have higher EE [8]. With these advantages, IRS has sparked so much interest recently.

Specifically, for the MISO channel, [9] studied the IRS-assisted design in a multi-cast MISO network, where a max-min fairness design problem was solved via a surrogate function. In [10], the authors investigated the harvested power (HP) maximization problem for IRS-enhanced MISO-SWIPT networks, where a SDR based algorithm was designed to obtain the reflection coefficients (RCs) of the IRS. Besides, [11] addressed the joint BF and phase shifter design for IRS-enhanced MISO-SWIPT networks, where a penalty dual decomposition (PDD) method was proposed to minimize the transmit power consumption, subject to different quality-of-service (QoS) constraints. Moreover, for the multiple-input multiple-output (MIMO) channel, [12] investigated the sum rate optimization in an IRS-aided multi-cell MIMO network, where an equivalent weighted minimum mean-squared error (MMSE) method was proposed. Also, [13] studied an IRS-aided MIMO-SWIPT network, where a price based method was presented to design the precoding matrix and the RCs.

Besides, several works have looked at the secrecy transmission design in IRS-aided networks. For instance, for the MISO channel, [14] obtained an IRS-assisted secure transmission by a SDR method. To further improve the secrecy performance, in [15] and [16], the artificial noise (AN)-assisted secure transmission was designed in an IRS-aided network. Furthermore, for the MIMO channel, [17] studied the IRS-aided secure design, by using an alternating optimization (AO) approach. Also, in [18] and [19], the secure MIMO communications via IRS were achieved by a block coordinate descent (BCD) and Lagrangian dual based algorithm. Among these works, the majorization-minimization (MM) method and the complex conjugate gradient (CCG) method were commonly used to design the RCs. However, these two methods are only available for the ideal channel state information (CSI) case.

Practically, it will be difficult to obtain accurate CSI due to estimation and quantization errors. While for the IRS-assisted networks, it is more challenging to obtain the accurate CSI about the cascade link than that in amplify-and-forward (AF) relay networks, since the IRS has no RF chain to perform baseband processing [20]. Currently, there are two main methods about the channel estimation for the IRS-related links. One is to separately estimate the Tx-IRS channel and the IRS-user channel by using some active elements at the IRS, which requires more hardware and power consumption [21]. The other is to estimate the cascaded Tx-IRS-user channel, where the main advantage is that no extra power and hardware consumption are needed, and the cascaded channel are sufficient for the joint BF design, such as [22] for MISO network and [23] for MIMO network, respectively. However, the robust design in IRS-aided wiretap networks is still much less understood, especially when Eve is passive and the related CSI is hard to obtain [24].

To handle this problem, several novel works have been proposed. For example, for the worst

case constrained design with bounded CSI errors, [25] studied the robust secrecy design with IRS by incorporating the AO, SCA, and penalty-based methods. Moreover, [26] investigated the robust secrecy BF for IRS-enhanced millimeter wave (mmWave) channels, where the worst case information rate was maximized by the AO and SDR approaches. Recently, [27] addressed the robust BF design for IRS-assisted MISO networks, where an AO and S-procedure based algorithm was designed to maximize the EE.

Besides, for the outage constrained design with probabilistic CSI errors, [28] investigated the robust design for IRS-enabled MISO systems, where an AO and Bernstein-type inequality (BTI) based method was proposed to iteratively optimize the BF and RCs. Then, [29] studied the robust design for IRS-enhanced MISO systems with the MMSE criterion, where an AO and MM based algorithm was developed. Recently, [30] proposed a robust power minimization design for IRS-assisted MISO networks, where a penalty concave convex procedure (PCCP) method was developed. In [31], the authors studied the outage constrained robust BF design in an IRS-assisted MISO wiretap network, where a penalty SDR algorithm was design to optimize the RCs. Furthermore, [32] extended the work in [31] by considering imperfect direct link CSI, where a PCCP based approach was proposed. Besides, [33] studied a robust BF design in IRS-empowered MISO networks by the PDD method. Then, [34] investigated the outage constrained robust BF for IRS-aided MISO networks, where a constrained stochastic SCA method was proposed. Also, [35] studied the IRS-assisted MISO uplink transmission, where the robust power minimization was solved by the PDD method. Recently, [36] studied the phase shifters design in IRS-assisted network with statistical CSI, where a genetic algorithm was proposed. Also, [37] proposed a stochastic optimization-based BF design in IRS-enabled mmWave systems to combat the CSI uncertainty incurred by the random blockages, and extended in [38] by considering multiple IRSs and the fairness requirement. However, unfortunately, a comprehensive method for the robust designs in IRS-aided SWIPT networks has not been found.

Motivated by this, this work investigates the robust secrecy transmission in an IRS-aided MISO-SWIPT network, where the imperfect direct and cascaded channels of the legitimate and eavesdropping links are considered. Specifically, by considering imperfect IRs' and ERs' channels and multiple QoS constraints, we consider the max-min fairness robust information rate (RIR) objective, by jointly designing the BF, the AN, and the RCs. The proposed problem is hard to tackle due to the non-convex objective, the coupled variables, and the unit modulus constraint (UMC). To facilitate the design, we utilize the SCA method to convert the problem, then an AO algorithm is proposed, where the UMC is tackled by the PDD technique. Finally, simulation results verify the performance of the proposed approach.

Our contribution are summarized as follows:

1) For the bounded CSI error model, we formulate the worst case robust design, subject to the worst case QoS constraints and the UMC of the phase shifter. The AO and SCA methods are utilized to decompose and reformulate the original non-smooth non-convex objective to a tractable formulation, while the CSI error is tackled by the Schur-complement and general S-procedure, and the UMC is handled by the PDD method, which is guaranteed to converge. Then, an iterative manner is proposed to optimize the BF, the AN covariance, and the RCs.
2) Then, for the probabilistic CSI error case, we aim to maximize the minimum outage information rate subject to the UMC and multiple probability constraints. By applying the BTI, the probability constraints are transformed into several linear matrix inequality (LMI) constraints and second order cone (SOC) constraints. Then, similar to the worst case robust design, the associated variables are jointly optimized by the SCA and PDD techniques in an iterative manner.
3) Then, for the statistical CSI case, where only statistical distribution of the channel covariance matrices can be obtained, by approximating the outage probability constraints with convex surrogate function, we propose a stochastic SCA method to handle the robust design, where the outer optimization is conducted by the bisection search method, and the inner optimization is tackled by the SCA and PDD methods.
4) The feasibility and convergence of the proposed algorithm are analyzed. Simulation results show that the proposed design achieve better secrecy performance than other baselines. Besides, it reveals that the cascaded channel error level plays a more important role in the robustness than the direct channel error level, and it is preferable to deploy the IRS close to Tx or IRs to improve security.

The rest of this paper is given as follows. Section II introduces the system model. Section III investigates the robust design for the bounded CSI error model. Then, Section IV addresses the robust design for the probabilistic CSI error model. Section V studies the stochastic SCA method to tackle the statistical CSI. Finally, Section VI and Section VII provide the simulation results and conclude the paper, respectively.

Notations: In this work, boldface lowercase and uppercase letters denote vectors and matrices, respectively. The conjugate, transpose, conjugate transpose, and trace of matrix $\mathbf{A}$ are denoted as $\mathbf{A}^{\dagger}$, $\mathbf{A}^{T}$, $\mathbf{A}^{H}$, and $\mathrm{Tr}\left(\mathbf{A}\right)$, respectively. $\mathbf{a}=\mathrm{vec}\left(\mathbf{A}\right)$ means to stack the columns of $\mathbf{A}$ into $\mathbf{a}$. $\mathbf{A}\succeq\mathbf{0}$ means that $\mathbf{A}$ is positive semi-definite. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the Euclidean norm and Frobenius norm, respectively. The block-diagonal matrix with diagonal elements $\mathbf{A}_1,\ldots,\mathbf{A}_N$ is

denoted by $\mathrm{BlkD}\left(\mathbf{A}_1,\ldots,\mathbf{A}_N\right)$, and is reduced to $\mathrm{D}\left(a_1,\ldots,a_N\right)$ when scalar diagonal elements are considered. $[;]$ and $[,]$ denote concatenation by column and row, respectively. $\mathbf{1}_M$ and $\mathbf{0}_M$ are $M\times 1$ vectors with all elements being 1 and 0, respectively. $\mathbf{I}$ indicates an identity matrix. Moreover, $\Re\{a\}$, $|a|$, and $\angle a$ denote the real part, the modulus, and the angle of a complex number a, respectively. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\mathbf{x}$ and covariance $\mathbf{\Sigma}$ is denoted by $\mathcal{CN}\left(\mathbf{x},\mathbf{\Sigma}\right)$. In addition, $\circ$ means the Hadamard production, $\otimes$ denotes the Kronecker product, $\mathcal{O}$ is the big-O notation, $\mathbb{E}\{\cdot\}$ means the mathematical expectation, and $\jmath$ denotes the imaginary unit, respectively.

II. System Model

In this section, we propose the IRS-assisted secrecy SWIPT system model. Then, the imperfect CSI model will be introduced.

A. Signal Transmission Model

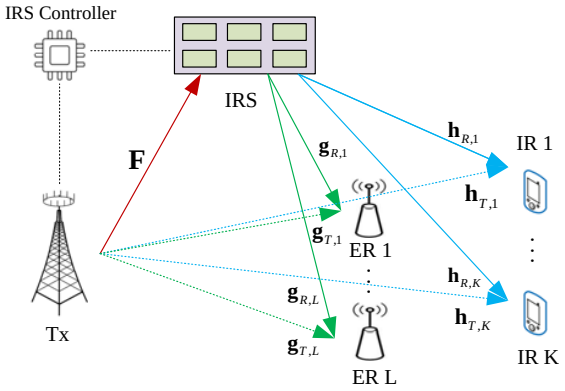

Fig. 1. The IRS-aided MISO downlink SWIPT system.

As shown in Fig. 1, we consider an IRS-aided MISO-SWIPT system consisting of one Tx, one IRS, K IRs, and L ERs. Tx has N_t antennas and the IRS has M reflecting units, respectively. In addition, each IR (and ER) is single antenna node and the sets of the IRs and ERs are denoted as $\mathcal{K}\triangleq\{1,\ldots,K\}$ and $\mathcal{L}\triangleq\{1,\ldots,L\}$, respectively. The channel between Tx and the k-th IR and the l-th ER is denoted by $\mathbf{h}_{T,k}\in\mathbb{C}^{N_t\times 1}$ and $\mathbf{g}_{T,l}\in\mathbb{C}^{N_t\times 1}$, respectively. Besides, the channel between Tx and the IRS is denoted as $\mathbf{F}\in\mathbb{C}^{M\times N_t}$, while the channel between the IRS and the k-th IR and the l-th ER is denoted by $\mathbf{h}_{R,k}\in\mathbb{C}^{M\times 1}$ and $\mathbf{g}_{R,l}\in\mathbb{C}^{M\times 1}$, respectively. In addition, an IRS controller is utilized to control the CSI exchange and information transmission between Tx and the IRS.

Let $x_k \in \mathbb{C}$ be the information for the k-th IR. The transmit signal $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ is given as $\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k x_k + \mathbf{w}_0$, where $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}, \forall k \in \mathcal{K}$ is the transmit BF for the k-th IR, and $\mathbf{w}_0 \in \mathbb{C}^{N_t \times 1}$ is the AN, which can been seen as an energy signal to improve the energy transmission.

The RCs of the IRS are are modeled as $\boldsymbol{\theta} \triangleq \left[e^{\phi_1}, \ldots, e^{\phi_M}\right]^T$, with ϕ_m denoting the phase-shifting of the m-th reflection element. Besides, the diagonal phase shift matrix is given by $\boldsymbol{\Theta} = \mathrm{D}(\boldsymbol{\theta})$. Since due to the "distance-product" effect [6], the signals reflected twice or more times are omitted here.

Assuming quasi-static flat channels, the signals received by the k-th IR and the l-th ER are, respectively, given by

$$y_{i,k} = \left(\mathbf{h}_{T,k}^H + \mathbf{h}_{R,k}^H \boldsymbol{\Theta}^H \mathbf{F}\right) \mathbf{x} + n_{i,k}, \tag{1a}$$

$$y_{e,l} = \left(\mathbf{g}_{T,l}^H + \mathbf{g}_{R,l}^H \boldsymbol{\Theta}^H \mathbf{F}\right) \mathbf{x} + n_{e,l}, \tag{1b}$$

where $n_{i,k}$ denotes the antenna noise at the k-th IR with $n_{i,k} \sim \mathcal{CN}\left(0, \sigma_{i,k}^2\right)$, and $n_{e,l}$ denotes the antenna noise at the l-th ER with $n_{e,l} \sim \mathcal{CN}\left(0, \sigma_{e,l}^2\right)$, respectively. In this work, we assume that perfect synchronization can be obtained between the direct link and the cascaded link. In fact, synchronization is still an open issue in an IRS-aided network [7].

In fact, we have $\mathbf{h}_{T,k}^H + \mathbf{h}_{R,k}^H \boldsymbol{\Theta}^H \mathbf{F} = \mathbf{h}_{T,k}^H + \boldsymbol{\theta}^H \underbrace{\mathrm{D}\left(\mathbf{h}_{R,k}^H\right) \mathbf{F}}_{\mathbf{H}_{R,k}} = \underbrace{\left[\boldsymbol{\theta}^H, 1\right]}_{\hat{\boldsymbol{\theta}}^H} \underbrace{\left[\mathbf{H}_{R,k}; \mathbf{h}_{T,k}^H\right]}_{\mathbf{H}_k}$ and $\mathbf{g}_{T,l}^H + \mathbf{g}_{R,l}^H \boldsymbol{\Theta}^H \mathbf{F} = \mathbf{g}_{T,l}^H + \boldsymbol{\theta}^H \underbrace{\mathrm{D}\left(\mathbf{g}_{R,l}^H\right) \mathbf{F}}_{\mathbf{G}_{R,l}} = \hat{\boldsymbol{\theta}}^H \underbrace{\left[\mathbf{G}_{R,l}; \mathbf{g}_{T,l}^H\right]}_{\mathbf{G}_l}$. Thus, the signal-to-interference-noise-ratio (SINR) for the k-th IR is given as

$$\Gamma_k = \frac{\left|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_k\right|^2}{\sum_{j=0, j \neq k}^{K} \left|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_j\right|^2 + \sigma_{i,k}^2} = \frac{\left|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_k\right|^2}{\left\|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{W}_{-k}\right\|_2^2 + \sigma_{i,k}^2}, \tag{2}$$

where $\mathbf{W}_{-k} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_K, \mathbf{w}_0]$.

On the other hand, when the l-th ER attempts to eavesdrop the confidential information intended to the k-th IR, the SINR for the l-th ER is given as

$$\Gamma_{l,k} = \frac{\left|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_k\right|^2}{\sum_{j=0, j \neq k}^{K} \left|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_j\right|^2 + \sigma_{e,l}^2} = \frac{\left|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_k\right|^2}{\left\|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{W}_{-k}\right\|_2^2 + \sigma_{e,l}^2}. \tag{3}$$

In addition, the HP for the l-th ER is

$$E_l = \sum_{j=0}^{K} \left|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_j\right|^2 + \sigma_{e,l}^2 = \left\|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{W}_0\right\|_2^2 + \sigma_{e,l}^2, \tag{4}$$

where $\mathbf{W}_0 \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_K, \mathbf{w}_0]$.

B. CSI Error Models

In this work, both imperfect direct link and cascaded link are considered, i.e.,

$$\mathbf{H}_{R,k} = \bar{\mathbf{H}}_{R,k} + \Delta\mathbf{H}_{R,k}, \mathbf{h}_{T,k} = \bar{\mathbf{h}}_{T,k} + \Delta\mathbf{h}_{T,k}, \tag{5a}$$

$$\mathbf{G}_{R,l} = \bar{\mathbf{G}}_{R,l} + \Delta\mathbf{G}_{R,l}, \mathbf{g}_{T,l} = \bar{\mathbf{g}}_{T,l} + \Delta\mathbf{g}_{T,l}, \tag{5b}$$

where $\mathbf{H}_{R,k}$ and $\mathbf{h}_{T,k}$ denote the true CSI, $\bar{\mathbf{H}}_{R,k}$ and $\bar{\mathbf{h}}_{T,k}$ denote the estimated CSI, $\Delta\mathbf{H}_{R,k}$ and $\Delta\mathbf{h}_{T,k}$ denote the associated CSI error, respectively. So as the same to $\mathbf{G}_{R,l}$, $\mathbf{g}_{T,l}$, $\bar{\mathbf{G}}_{R,l}$, $\bar{\mathbf{g}}_{T,l}$, $\Delta\mathbf{G}_{R,l}$ and $\Delta\mathbf{g}_{T,l}$. Specifically, the following CSI models are considered:

1) The bounded CSI error model, where the CSI error is assumed to lie in a region with a given bound, i.e.,

$$\mathcal{H}_{b,k} = \left\{ \|\Delta\mathbf{h}_{T,k}\|_2 \leq \epsilon_{T,l}, \|\Delta\mathbf{H}_{R,k}\|_F \leq \epsilon_{R,k} \right\}, \tag{6a}$$

$$\mathcal{G}_{b,l} = \left\{ \|\Delta\mathbf{g}_{T,l}\|_2 \leq \chi_{T,l}, \|\Delta\mathbf{G}_{R,l}\|_F \leq \chi_{R,l} \right\}, \tag{6b}$$

where $\epsilon_{T,k}$, $\epsilon_{R,k}$, $\chi_{T,l}$ and $\chi_{R,l}$ denote the respective sizes of the bounded channel error region. This CSI error model is commonly used to characterize the channel quantization error which lies in a bounded region and associated with the worst case constrained design. For instance, in the frequency division duplex (FDD) network, the receiver (Rx) estimates the downlink CSI and then feeds the quantized CSI back to Tx. Thus, the obtained CSI is affected by quantization errors [30].

2) The probabilistic CSI error model, where the CSI error is assumed to follow the CSCG distribution, i.e.,

$$\mathcal{H}_{p,k} = \left\{ \Delta\mathbf{h}_{T,k} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{T,k}\right), \ \mathrm{vec}\left(\Delta\mathbf{H}_{R,k}\right) \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{R,k}\right) \right\}, \tag{7a}$$

$$\mathcal{G}_{p,l} = \left\{ \Delta\mathbf{g}_{T,l} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{T,l}\right), \ \mathrm{vec}\left(\Delta\mathbf{G}_{R,l}\right) \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{R,l}\right) \right\}, \tag{7b}$$

where $\mathbf{R}_{T,k} \in \mathbb{C}^{N_t \times N_t}$, $\mathbf{R}_{R,k} \in \mathbb{C}^{MN_t \times MN_t}$, $\mathbf{C}_{T,l} \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{C}_{R,l} \in \mathbb{C}^{MN_t \times MN_t}$ are the positive semi-definite error covariances for the direct link and the cascaded link, respectively. This kind of imperfect CSI is mainly caused by channel estimation error and associated with the outage constrained design. For instance, in the time division duplex (TDD) network, limited training time and noise will cause channel estimation error, which generally follows the CSCG distribution when the MMSE method is used to estimate the channel [30].

3) The statistical CSI model, where both the estimated CSI and the CSI error are random variables following unknown distribution. In fact, the statistical CSI model can be treated a special case of the probabilistic CSI error model.

In practical systems, these types of errors commonly coexist and it would be hard to determine whether kind of CSI error is the major factor. Thus, in this work, we aim to find a comprehensive

robust method, which is suitable for these types of CSI errors.

III. The Worst Case Constrained Robust Design

Here, we investigate the worst case robust design, which can guarantee that the information rate of each IR is no less than its minimum rate constraint for all possible CSI errors. Mathematically, the problem is given as

$$\max_{\mathbf{w}_k,\hat{\boldsymbol{\theta}},R_{w,k}} \quad \min R_{w,k} \tag{8a}$$

$$\text{s.t.} \quad \min_{\forall\left\{\Delta\mathbf{h}_{T,k},\Delta\mathbf{H}_{R,k}\right\}\in\mathcal{H}_{b,k}} \ln\left(1+\Gamma_k\right) \geq R_{w,k}, \forall k, \tag{8b}$$

$$\max_{\forall\left\{\Delta\mathbf{g}_{T,l},\Delta\mathbf{G}_{R,l}\right\}\in\mathcal{G}_{b,l}} \ln\left(1+\Gamma_{l,k}\right) \leq \gamma_{w,k}, \forall k, \forall l, \tag{8c}$$

$$\min_{\forall\left\{\Delta\mathbf{g}_{T,l},\Delta\mathbf{G}_{R,l}\right\}\in\mathcal{G}_{b,l}} E_l \geq E_{th,l}, \forall l, \tag{8d}$$

$$\left|\hat{\theta}_m\right| = 1, m = 1,\ldots,M, \hat{\theta}_{M+1} = 1, \tag{8e}$$

$$\sum_{k=0}^{K} \left\|\mathbf{w}_k\right\|^2 \leq P_s, \tag{8f}$$

where P_s is the maximum transmit power, $R_{w,k}$ denotes the worst case rate threshold for the k-th IR, $\gamma_{w,k}$ denotes the worst case rate threshold for the l-th ER, when this ER try to decode the information for the k-th IR, and $E_{th,l}$ denotes the worst case HP threshold for the l-th ER, respectively. Unfortunately, (8) is hard to solve due to multiple coupled variables, the infinite constraints caused by the CSI error, and the UMC. In the next part, we will propose an AO and SCA based method to handle (8).

A. Active BF and AN Optimization

At first, we denote the fixed point of $\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}\right)$ in the iteration as $\left(\{\tilde{\mathbf{w}}_k\}_{k=0}^K, \tilde{\boldsymbol{\theta}}\right)$. Similarly, we denote $\tilde{\mathbf{W}}_{-k} \triangleq [\tilde{\mathbf{w}}_1,\ldots,\tilde{\mathbf{w}}_{k-1},\tilde{\mathbf{w}}_{k+1},\ldots,\tilde{\mathbf{w}}_K,\tilde{\mathbf{w}}_0]$, and $\tilde{\mathbf{W}}_0 \triangleq [\tilde{\mathbf{w}}_1,\ldots,\tilde{\mathbf{w}}_K,\tilde{\mathbf{w}}_0]$.

Here, we aim to solve the subproblem with respect to (w.r.t.) $\{\mathbf{w}_k\}_{k=0}^K$ with fixed $\left(\{\tilde{\mathbf{w}}_k\}_{k=0}^K, \tilde{\boldsymbol{\theta}}\right)$.

By introducing the auxiliary variables $\{t_k\}_{k=1}^K$, and $\{p_{k,l}\}_{k=1,l=1}^{K,L}$, (8) can be reformulated as

$$\max_{\mathbf{w}_k,\kappa} \quad \kappa \tag{9a}$$

$$\text{s.t.} \left|\tilde{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_k\right|^2 \geq t_k \left(\left\|\tilde{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{W}_{-k}\right\|_2^2 + \sigma_{i,k}^2\right), \forall k, \tag{9b}$$

$$\left|\tilde{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_k\right|^2 \leq p_{k,l} \left(\left\|\tilde{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{W}_{-k}\right\|_2^2 + \sigma_{e,l}^2\right), \forall k, \forall l, \tag{9c}$$

$$1 + t_k \geq e^{R_{w,k}}, 1 + p_{k,l} \leq e^{\gamma_{w,k}}, R_{w,k} \geq \kappa, \forall k, \tag{9d}$$

$$(8d), (8f). \tag{9e}$$

It is known that the non-convex (9b) can be further reformulated as

$$\left|\tilde{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_k\right|^2 \geq \frac{1}{s_k} \left(\left\|\tilde{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{W}_{-k}\right\|_2^2 + \sigma_{i,k}^2\right), \tag{10a}$$

$$1 \geq s_k t_k. \tag{10b}$$

Besides, according to [39], (10b) can be upper bounded by

$$1 \geq \frac{t_k^2 \tilde{s}_k}{2\tilde{t}_k} + \frac{\tilde{t}_k s_k^2}{2\tilde{s}_k}, \tag{11}$$

where $\tilde{t}_k$ and $\tilde{s}_k$ are the values of t_k and s_k in the previous iteration. Furthermore, using the identity $\mathrm{Tr}\left(\mathbf{A}^H \mathbf{B} \mathbf{C} \mathbf{D}\right) = \mathrm{vec}(\mathbf{A})^H \left(\mathbf{D}^T \otimes \mathbf{B}\right) \mathrm{vec}\left(\mathbf{C}\right)$, (10a) is equivalent to

$$\mathbf{h}_k^H \left(\left(\mathbf{w}_k \mathbf{w}_k^H - \frac{\mathbf{W}_{-k}\mathbf{W}_{-k}^H}{s_k}\right)^T \otimes \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right) \mathbf{h}_k \geq \frac{\sigma_{i,k}^2}{s_k}, \tag{12}$$

where $\mathbf{h}_k = \mathrm{vec}\left(\mathbf{H}_k\right)$. Similarly, (9c) can be turned into

$$\mathbf{g}_l^H \left(\left(\mathbf{W}_{-k}\mathbf{W}_{-k}^H - \frac{\mathbf{w}_k \mathbf{w}_k^H}{p_{k,l}}\right)^T \otimes \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right) \mathbf{g}_l + \sigma_{e,l}^2 \geq 0, \tag{13}$$

where $\mathbf{g}_l = \mathrm{vec}\left(\mathbf{G}_l\right)$. Note that (12) and (13) are still non-convex due to the difference of the quadratic terms. To handle (12) and (13), we find the following lemma is useful.

Lemma 1 [40]: For any complex matrix $\mathbf{P}$, a lower bound of $\mathbf{P}\mathbf{P}^H$ is $\mathbf{P}\tilde{\mathbf{P}}^H + \tilde{\mathbf{P}}\mathbf{P}^H - \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$, where $\tilde{\mathbf{P}}$ is a given point.

By Lemma 1, (12) can be approximated as

$$\mathbf{h}_k^H \left(\left(\mathbf{X}_k - \mathbf{Y}_k\right)^T \otimes \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right) \mathbf{h}_k \geq \frac{\sigma_{i,k}^2}{s_k}, \tag{14a}$$

$$\mathbf{Y}_k \succeq \frac{\mathbf{W}_{-k}\mathbf{W}_{-k}^H}{s_k}, \tag{14b}$$

where $\mathbf{X}_k \triangleq \tilde{\mathbf{w}}_k \mathbf{w}_k^H + \mathbf{w}_k \tilde{\mathbf{w}}_k^H - \tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^H$. By the Schur complement [39], (14b) can be equivalently converted into the following linear matrix inequality (LMI)

$$\begin{bmatrix} \mathbf{Y}_k & \mathbf{W}_{-k} \\ \mathbf{W}_{-k}^H & s_k \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \tag{15}$$

Following the same way, (13) can be approximated as

$$\mathbf{g}_l^H \left(\left(\mathbf{\Omega}_k - \mathbf{\Psi}_{k,l} \right)^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \mathbf{g}_l + \sigma_{e,l}^2 \geq 0, \tag{16a}$$

$$\begin{bmatrix} \mathbf{\Psi}_{k,l} & \mathbf{w}_k \\ \mathbf{w}_k^H & p_{k,l} \end{bmatrix} \succeq \mathbf{0}, \tag{16b}$$

where $\mathbf{\Omega}_k \triangleq \tilde{\mathbf{W}}_{-k} \mathbf{W}_{-k}^H + \mathbf{W}_{-k} \tilde{\mathbf{W}}_{-k}^H - \tilde{\mathbf{W}}_{-k} \tilde{\mathbf{W}}_{-k}^H$.

Next, we aim to handle the CSI uncertainty in (14a) and (16a). First, we denote $\bar{\mathbf{H}}_k = \left[\bar{\mathbf{H}}_{R,k}; \bar{\mathbf{h}}_{T,k}^H \right]^T$, and $\Delta \mathbf{H}_k = \left[\Delta \mathbf{H}_{R,k}; \Delta \mathbf{h}_{T,k}^H \right]^T$, it is known that $\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta \mathbf{H}_k$. Besides, the following equation holds $\| \Delta \mathbf{H}_{R,k} \|_F \leq \epsilon_{R,k} \Rightarrow \| \mathrm{vec} \left(\Delta \mathbf{H}_{R,k} \right) \|_2 \leq \epsilon_{R,k}$. Then, we denote $\mathbf{I}_1 \triangleq \mathrm{D} \left(\mathbf{1}_{MN_t}, \mathbf{0}_{N_t} \right), \mathbf{I}_2 \triangleq \mathrm{D} \left(\mathbf{0}_{MN_t}, \mathbf{1}_{N_t} \right)$. Thus, we have $\Delta \mathbf{h}_k^H \mathbf{I}_1 \Delta \mathbf{h}_k \leq \epsilon_{R,k}^2$, and $\Delta \mathbf{h}_k^H \mathbf{I}_2 \Delta \mathbf{h}_k \leq \epsilon_{T,k}^2$, where $\Delta \mathbf{h}_k = \mathrm{vec} \left(\Delta \mathbf{H}_k \right)$.

Lemma 2 [30]: Define the function w.r.t. $\mathbf{x} \in \mathbb{C}^{n \times 1}$:

$$f_j \left(\mathbf{x} \right) = \mathbf{x}^H \mathbf{A}_j \mathbf{x} + 2 \mathrm{Re} \left\{ \mathbf{b}_j^H \mathbf{x} \right\} + c_j, j = 0, \ldots, J,$$

where $\mathbf{A}_j = \mathbf{A}_j^H \in \mathbb{C}^{n \times n}$, $\mathbf{b}_j \in \mathbb{C}^{n \times 1}$, and $c_j \in \mathbb{R}$. The condition $\left\{ f_j \left(\mathbf{x} \right) \geq 0 \right\}_{j=1}^J \Rightarrow f_0 \left(\mathbf{x} \right) \geq 0$ holds if and only if there exists $\left\{ \lambda_j \geq 0 \right\}_{j=1}^J$ such that

$$\begin{bmatrix} \mathbf{A}_0 & \mathbf{b}_0 \\ \mathbf{b}_0^H & c_0 \end{bmatrix} - \sum_{j=1}^J \lambda_j \begin{bmatrix} \mathbf{A}_j & \mathbf{b}_j \\ \mathbf{b}_j^H & c_j \end{bmatrix} \succeq \mathbf{0}.$$

To utilize Lemma 2, we first rewrite (14a) as

$$\bar{\mathbf{h}}_k^H \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{h}}_k + \Delta \mathbf{h}_k^H \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \Delta \mathbf{h}_k + 2 \Re \left\{ \bar{\mathbf{h}}_k^H \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \Delta \mathbf{h}_k \right\} \geq \frac{\sigma_{i,k}^2}{s_k}, \tag{17}$$

where $\bar{\mathbf{h}}_k = \mathrm{vec} \left(\bar{\mathbf{H}}_k \right)$ and $\mathbf{Z}_k = \mathbf{X}_k - \mathbf{Y}_k$.

Following Lemma 2, (17) holds if the following LMI holds

$$\begin{bmatrix} \mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H + \vartheta_{1,k} \mathbf{I}_1 + \vartheta_{2,k} \mathbf{I}_2 & \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{h}}_k \\ \bar{\mathbf{h}}_k^H \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) & \xi_k \end{bmatrix} \succeq \mathbf{0}, \tag{18}$$

where $\left\{ \vartheta_{1,k} \geq 0, \vartheta_{2,k} \geq 0 \right\}_{k=1}^K$ are the slack variables, and $\xi_k = \bar{\mathbf{h}}_k^H \left(\mathbf{Z}_k^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{h}}_k - \frac{\sigma_{i,k}^2}{s_k} - \vartheta_{1,k} \epsilon_{R,k}^2 - \vartheta_{2,k} \epsilon_{T,k}^2$.

Then, we denote $\bar{\mathbf{g}}_l = \mathrm{vec} \left(\bar{\mathbf{G}}_l \right)$ and $\mathbf{\Phi}_{k,l} = \mathbf{\Omega}_k - \mathbf{\Psi}_{k,l}$, (16a) holds if the following LMI holds

$$\begin{bmatrix} \mathbf{\Phi}_{k,l}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H + \alpha_{1,k,l} \mathbf{I}_1 + \alpha_{2,k,l} \mathbf{I}_2 & \left(\mathbf{\Phi}_{k,l}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{g}}_l \\ \bar{\mathbf{g}}_l^H \left(\mathbf{\Phi}_{k,l}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) & \rho_{k,l} \end{bmatrix} \succeq \mathbf{0}, \tag{19}$$

where $\left\{ \alpha_{1,k,l} \geq 0, \alpha_{2,k,l} \geq 0 \right\}_{k=1,l=1}^{K,L}$ are the slack variables, and $\rho_{k,l} = \bar{\mathbf{g}}_l^H \left(\mathbf{\Phi}_{k,l}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{g}}_l + \sigma_{e,l}^2 - \alpha_{1,k,l} \chi_{R,l}^2 - \alpha_{2,k,l} \chi_{T,l}^2$. Similarly, by denoting $\mathbf{\Xi} = \tilde{\mathbf{W}}_0 \mathbf{W}_0^H + \mathbf{W}_0 \tilde{\mathbf{W}}_0^H - \tilde{\mathbf{W}}_0 \tilde{\mathbf{W}}_0^H$, the worst case HP constraint (8d) can be given as

$$\begin{bmatrix} \mathbf{\Xi}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H + \beta_{1,l} \mathbf{I}_1 + \beta_{2,l} \mathbf{I}_2 & \left(\mathbf{\Xi}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) \bar{\mathbf{g}}_l \\ \bar{\mathbf{g}}_l^H \left(\mathbf{\Xi}^T \otimes \tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^H \right) & \pi_l \end{bmatrix} \succeq \mathbf{0}, \tag{20}$$

where $\{\beta_{1,l} \geq 0, \beta_{2,l} \geq 0\}_{l=1}^{L}$ are the slack variables, and $\pi_l = \bar{\mathbf{g}}_l^H \left(\mathbf{\Xi}^T \otimes \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right) \bar{\mathbf{g}}_l + \sigma_{e,l}^2 - E_{th} - \beta_{1,l}\chi_{R,l}^2 - \beta_{2,l}\chi_{T,l}^2$. By combining these steps, we formulate the following problem w.r.t. $\{\mathbf{w}_k\}_{k=0}^{K}$

$$\max_{\mathbf{w}_k, \kappa} \quad \kappa \tag{21a}$$

$$\text{s.t.} \quad (8\text{f}), (9\text{d}), (11), (15), (16\text{b}), (18), (19), (20), \tag{21b}$$

which is convex and can be solved by the toolbox CVX [41].

B. Phase Shifter Optimization

Here, we will solve the subproblem w.r.t. $\hat{\boldsymbol{\theta}}$. Due to symmetry, the proposed approach can be used to optimize $\hat{\boldsymbol{\theta}}$ conveniently. Specifically, we need to reformulate (12), (13), (8d), and (8e) into convex constraints w.r.t. to $\hat{\boldsymbol{\theta}}$. Firstly, by denoting $\mathbf{T} = \hat{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H + \tilde{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^H - \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H$, we have the following approximation of (12)

$$\mathbf{h}_k^H \left(\left(\tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^H - \frac{\tilde{\mathbf{W}}_{-k}\tilde{\mathbf{W}}_{-k}^H}{\tilde{s}_k} \right)^T \otimes \mathbf{T} \right) \mathbf{h}_k \geq \frac{\sigma_{i,k}^2}{\tilde{s}_k}. \tag{22}$$

Then, by denoting $\tilde{\mathbf{\Upsilon}}_k = \tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^H - \frac{\tilde{\mathbf{W}}_{-k}\tilde{\mathbf{W}}_{-k}^H}{\tilde{s}_k}$ and using Lemma 2, we have the following LMI

$$\begin{bmatrix} \tilde{\mathbf{\Upsilon}}_k^T \otimes \mathbf{T} + \vartheta_{1,k}\mathbf{I}_1 + \vartheta_{2,k}\mathbf{I}_2 & \left(\tilde{\mathbf{\Upsilon}}_k^T \otimes \mathbf{T}\right) \bar{\mathbf{h}}_k \\ \bar{\mathbf{h}}_k^H \left(\tilde{\mathbf{\Upsilon}}_k^T \otimes \mathbf{T}\right) & \xi_k \end{bmatrix} \succeq \mathbf{0}, \tag{23}$$

where $\xi_k = \bar{\mathbf{h}}_k^H \left(\tilde{\mathbf{\Upsilon}}_k^T \otimes \mathbf{T}\right) \bar{\mathbf{h}}_k - \frac{\sigma_{i,k}^2}{s_k} - \vartheta_{1,k}\epsilon_{R,k}^2 - \vartheta_{2,k}\epsilon_{T,k}^2$.

Similarly, by denoting $\tilde{\mathbf{\Phi}}_{k,l} \triangleq \tilde{\mathbf{W}}_{-k}\tilde{\mathbf{W}}_{-k}^H - \frac{\tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^H}{\tilde{p}_{k,l}}$, where $\tilde{p}_{k,l}$ is the value of $p_{k,l}$ in the previous iteration, we have the following LMI

$$\begin{bmatrix} \tilde{\mathbf{\Phi}}_{k,l}^T \otimes \mathbf{T} + \alpha_{1,k,l}\mathbf{I}_1 + \alpha_{2,k,l}\mathbf{I}_2 & \left(\tilde{\mathbf{\Phi}}_{k,l}^T \otimes \mathbf{T}\right) \bar{\mathbf{g}}_l \\ \bar{\mathbf{g}}_l^H \left(\tilde{\mathbf{\Phi}}_{k,l}^T \otimes \mathbf{T}\right) & \rho_{k,l} \end{bmatrix} \succeq \mathbf{0}, \tag{24}$$

where $\rho_{k,l} = \bar{\mathbf{g}}_l^H \left(\tilde{\mathbf{\Phi}}_{k,l}^T \otimes \mathbf{T}\right) \bar{\mathbf{g}}_l + \sigma_{e,l}^2 - \alpha_{1,k,l}\chi_{R,l}^2 - \alpha_{2,k,l}\chi_{T,l}^2$. Similarly, by denoting $\tilde{\mathbf{\Xi}} \triangleq \tilde{\mathbf{W}}_0 \tilde{\mathbf{W}}_0^H$, we have the following LMI

$$\begin{bmatrix} \tilde{\mathbf{\Xi}}^T \otimes \mathbf{T} + \beta_{1,l}\mathbf{I}_1 + \beta_{2,l}\mathbf{I}_2 & \left(\tilde{\mathbf{\Xi}}^T \otimes \mathbf{T}\right) \bar{\mathbf{g}}_l \\ \bar{\mathbf{g}}_l^H \left(\tilde{\mathbf{\Xi}}^T \otimes \mathbf{T}\right) & \pi_l \end{bmatrix} \succeq \mathbf{0}, \tag{25}$$

where $\pi_l = \bar{\mathbf{g}}_l^H \left(\tilde{\mathbf{\Xi}}^T \otimes \mathbf{T}\right) \bar{\mathbf{g}}_l + \sigma_{e,l}^2 - E_{th} - \beta_{1,l}\chi_{R,l}^2 - \beta_{2,l}\chi_{T,l}^2$.

Thus, we have formulated the following problem

$$\max_{\hat{\boldsymbol{\theta}}, \kappa} \quad \kappa \tag{26a}$$

$$\text{s.t.} \quad (8\text{e}), (23), (24), (25), 1 \geq t_k \tilde{s}_k, 1 + t_k \geq e^{R_{w,k}}, R_{w,k} \geq \kappa, \forall k, \tag{26b}$$

w.r.t. $\hat{\boldsymbol{\theta}}$, which is convex except the UMC (8e).

Next, we adapt the PDD method to handle (8e), which decouples the original problem into several individual subproblems and iteratively solve these subproblems. To be specific, we introduce the auxiliary variable $\boldsymbol{\varphi} \in \mathbb{C}^{(M+1)\times 1}$ and reformulate (26) as

$$\max_{\hat{\boldsymbol{\theta}},\boldsymbol{\varphi},\kappa} \quad \kappa \tag{27a}$$

$$\text{s.t. } \hat{\boldsymbol{\theta}} = \boldsymbol{\varphi}, \tag{27b}$$

$$|\varphi_m| = 1, \tag{27c}$$

$$\left|\hat{\theta}_m\right| \leq 1, (23), (24), (25), 1 \geq t_k \tilde{s}_k, 1 + t_k \geq e^{R_{w,k}}, R_{w,k} \geq \kappa, \forall k. \tag{27d}$$

Then, by penalizing the equality constraint, the following AL of (27) can be obtained

$$\min_{\hat{\boldsymbol{\theta}},\boldsymbol{\varphi},\kappa,\boldsymbol{\tau}} \quad -\kappa + \frac{1}{2\lambda}\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right\|_2^2 + \Re\left\{\boldsymbol{\tau}^H\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right)\right\} \tag{28a}$$

$$\text{s.t.} \quad (27c), (27d), \tag{28b}$$

where $\lambda \geq 0$ and $\boldsymbol{\tau} \in \mathbb{C}^{(M+1)\times 1}$ are the penalty coefficient and the dual variable associated with (27b), respectively.

The PDD method needs a two-layer iteration procedure, where the inner layer optimizing $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\varphi}$ alternatively, while the outer layer optimizing $\boldsymbol{\tau}$ and λ. Specifically, for the inner layer iteration, when $\boldsymbol{\varphi}$ is fixed, we obtain the following subproblem w.r.t. $\hat{\boldsymbol{\theta}}$

$$\min_{\hat{\boldsymbol{\theta}},\kappa} \quad -\kappa + \frac{1}{2\lambda}\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right\|_2^2 + \Re\left\{\boldsymbol{\tau}^H\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right)\right\} \tag{29a}$$

$$\text{s.t.} \quad (27d), \tag{29b}$$

which is convex and can be solved by CVX.

Then, when $\hat{\boldsymbol{\theta}}$ is given, we formulate the following subproblem w.r.t. $\boldsymbol{\varphi}$

$$\min_{\boldsymbol{\varphi}} \quad \frac{1}{2\lambda}\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right\|_2^2 + \Re\left\{\boldsymbol{\tau}^H\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right)\right\} \tag{30a}$$

$$\text{s.t.} \quad |\varphi_m| = 1, \tag{30b}$$

Due to (30b), we have $\frac{1}{2\lambda}\|\boldsymbol{\varphi}\|_2^2 = \frac{M+1}{2\lambda}$. Hence (30) is reduced to

$$\max_{|\varphi_m|=1} \Re\left\{\left(\hat{\boldsymbol{\theta}} + \lambda\boldsymbol{\tau}\right)^H \boldsymbol{\varphi}\right\}, \tag{31}$$

and the optimal solution is $\boldsymbol{\varphi}^\star = e^{\jmath\angle\left(\hat{\boldsymbol{\theta}}+\lambda\boldsymbol{\tau}\right)}$.

The inner layer alternatively updates $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\varphi}$ until convergent. While for the outer layer problem, the following criterions are applied to design $\boldsymbol{\tau}$ and λ [35]:

1) when $\hat{\boldsymbol{\theta}} = \boldsymbol{\varphi}$ approximately holds, $\boldsymbol{\tau}$ can be obtained by $\boldsymbol{\tau} \leftarrow \boldsymbol{\tau} + \lambda^{-1}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right)$, which is similar to the alternative direction method of multipliers (ADMM) technique.
2) when $\hat{\boldsymbol{\theta}} = \boldsymbol{\varphi}$ is far from "being true", we inflate λ^{-1} to force the equality constraint $\hat{\boldsymbol{\theta}} = \boldsymbol{\varphi}$ be met in the next iteration.

The PDD algorithm is given in Algorithm 1 and the whole AO method is summarized in Algorithm 2, where ν^i is a parameter sequence converging to 0, c is a scaling factor, which is commonly choose in $[0.8\ 0.9]$ [35], and ϖ_1, ϖ_2 denote the stopping criterion.

Algorithm 1 The PDD Algorithm to Problem (27).

Initialize $\left\{\hat{\boldsymbol{\theta}}^0, \boldsymbol{\varphi}^0, \boldsymbol{\tau}^0, \lambda^0\right\}$ and set $i = 1$;
repeat
  Set $\hat{\boldsymbol{\theta}}^{i-1,\ell} = \hat{\boldsymbol{\theta}}^{i-1}$, $\boldsymbol{\varphi}^{i-1,\ell} = \boldsymbol{\varphi}^{i-1}$, $\ell = 0$;
  repeat
    Update $\hat{\boldsymbol{\theta}}^{i-1,\ell+1}$ via solving (29);
    Update $\boldsymbol{\varphi}^{i-1,\ell+1}$ via $\boldsymbol{\varphi} = e^{\jmath\angle\left(\hat{\boldsymbol{\theta}}+\lambda\boldsymbol{\tau}\right)}$, $\ell \leftarrow \ell + 1$;
  until Convergence
  Set $\hat{\boldsymbol{\theta}}^i = \hat{\boldsymbol{\theta}}^{i-1,\ell}$, $\hat{\boldsymbol{\varphi}}^i = \hat{\boldsymbol{\varphi}}^{i-1,\ell}$;
  if $\left\|\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i\right\|_2 \leq \nu^i$, **then** $\boldsymbol{\tau}^{i+1} = \boldsymbol{\tau}^i + \frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i\right)$, $\lambda^{i+1} = \lambda^i$;
  else $\boldsymbol{\tau}^{i+1} = \boldsymbol{\tau}^i$, $\lambda^{i+1} = c\lambda^i$; **end if**
  $i \leftarrow i + 1$;
until $\left\|\hat{\boldsymbol{\theta}}^\ell - \boldsymbol{\varphi}^\ell\right\|_2 \leq \varpi_1$

Algorithm 2 The AO Algorithm to Problem (8).

Initialize $\left(\{\mathbf{w}_k^0\}_{k=0}^K, \hat{\boldsymbol{\theta}}^0\right)$ and set $\ell = 0$;
repeat
  Obtain $\left\{\mathbf{w}_k^\ell\right\}_{k=1}^K$ and κ^ℓ with fixed $\left(\{\tilde{\mathbf{w}}_k\}_{k=0}^K, \tilde{\boldsymbol{\theta}}\right)$ by solving (21);
  Update $\{\tilde{\mathbf{w}}_k\}_{k=0}^K \leftarrow \left\{\mathbf{w}_k^\ell\right\}_{k=0}^K$;
  Obtain $\hat{\boldsymbol{\theta}}^\ell$ with fixed $\left(\{\tilde{\mathbf{w}}_k;\}_{k=0}^K, \tilde{\boldsymbol{\theta}}\right)$ by solving (27) with the PDD method;
  Update $\tilde{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^\ell$;
  $\ell \leftarrow \ell + 1$;
until $\kappa^\ell - \kappa^{\ell-1} < \varpi_2$
Output $\left(\{\mathbf{w}_k^\star\}_{k=0}^K, \hat{\boldsymbol{\theta}}^\star, \kappa^\star\right)$;

It should be pointed out that since (27) is non-convex, the PDD algorithm generally generates a sub-optimal solution. As for the convergence of the PDD algorithm, according to [35] and [43], we have the following Theorem:

Theorem 1: Assume that $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\}$ is obtained as a limit point of the sequence $\left\{\hat{\boldsymbol{\theta}}^{i,\ell}, \boldsymbol{\varphi}^{i,\ell}\right\}_{\ell=1}^\infty$, $\forall i \in \{1, 2, \ldots\}$ in Algorithm 1, then the PDD algorithm is guaranteed to converge when

$\frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i\right) + \boldsymbol{\tau}^i$ is bounded, and any limit point of the sequence $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\}$ is a Karush-Kuhn-Tucker (KKT) point of (27).

Proof: Please refer to Appendix A. ■

C. Feasibility discussion of problem (8)

The feasibility of (8) can be checked by the following problem

$$\text{Find} \quad \left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}\right) \tag{32a}$$

$$\text{s.t.} \quad \text{(8c)} - \text{(8f)}, \tag{32b}$$

which can be solved via the proposed AO method, by removing the associated constraints associated with (8b). And the AL problem is now given by $\min\limits_{\hat{\boldsymbol{\theta}},\boldsymbol{\varphi},\boldsymbol{\tau}} \frac{1}{2\lambda}\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right\|_2^2 + \Re\left\{\boldsymbol{\tau}^H\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\varphi}\right)\right\}$, s.t. (27c), $\left|\hat{\theta}_m\right| \leq 1$, (24), (25). We omit the detail here since this is straightforward. If (32) is feasible, we treat the obtained $\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}\right)$ as an initial point of the proposed AO algorithm for problem (8). Otherwise, (32) is infeasible, which is mainly due to the QoS constraints are too strict to be satisfied for the current channel realization and system parameters. In this case, we can remove certain constraints by removing some users in practice.

IV. The Outage Constrained Robust Design

In this section, we investigate the outage constrained robust design. Mathematically, the problem is given as

$$\max_{\mathbf{w}_k, \hat{\boldsymbol{\theta}}, R_{o,k}} \quad \min R_{o,k} \tag{33a}$$

$$\text{s.t.} \quad \Pr\left\{\ln\left(1+\Gamma_k\right) \geq R_{o,k}\right\} \geq \delta_k, \forall \left\{\Delta\mathbf{h}_{T,k}, \Delta\mathbf{H}_{R,k}\right\} \in \mathcal{H}_{p,k}, \forall k, \tag{33b}$$

$$\Pr\left\{\ln\left(1+\Gamma_{l,k}\right) \leq \gamma_{o,k}\right\} \geq \eta_k, \forall \left\{\Delta\mathbf{g}_{T,l}, \Delta\mathbf{G}_{R,l}\right\} \in \mathcal{G}_{p,l}, \forall k, \forall l, \tag{33c}$$

$$\Pr\left\{E_l \geq E_{th,l}\right\} \geq \zeta_l, \forall \left\{\Delta\mathbf{g}_{T,l}, \Delta\mathbf{G}_{R,l}\right\} \in \mathcal{G}_{p,l}, \forall l, \tag{33d}$$

$$\text{(8e)}, \text{(8f)}, \tag{33e}$$

where $\delta_k \in (0,1]$ denotes the information rate outage probability for the k-th IR, $R_{o,k}$ is the associated outage rate threshold, $\eta_k \in (0,1]$ denotes the information rate outage probability for the l-th ER, when this ER try to decode the information send to the k-th IR, and $r_{o,k}$ is the associated outage rate threshold, respectively. In addition, $\zeta_l \in (0,1]$ denotes the HP outage probability for the l-th ER, and $E_{th,l}$ is the corresponding HP threshold.

A. Active BF and AN Optimization

Here, with fixed $\left(\{\tilde{\mathbf{w}}_k\}_{k=0}^K, \tilde{\boldsymbol{\theta}}\right)$, we formulate the subproblem w.r.t. $\{\mathbf{w}_k\}_{k=0}^K$ as

$$\max_{\mathbf{w}_k,\kappa} \quad \kappa \tag{34a}$$

$$\text{s.t.} \quad \Pr\left\{\left|\tilde{\boldsymbol{\theta}}^H\mathbf{H}_k\mathbf{w}_k\right|^2 \geq t_k\left(\left\|\tilde{\boldsymbol{\theta}}^H\mathbf{H}_k\mathbf{W}_{-k}\right\|_2^2+\sigma_{i,k}^2\right)\right\} \geq \delta_k, \forall k, \tag{34b}$$

$$\Pr\left\{\left|\tilde{\boldsymbol{\theta}}^H\mathbf{G}_l\mathbf{w}_k\right|^2 \leq p_{k,l}\left(\left\|\tilde{\boldsymbol{\theta}}^H\mathbf{G}_l\mathbf{W}_{-k}\right\|_2^2+\sigma_{e,l}^2\right)\right\} \geq \eta_k, \forall k, \forall l, \tag{34c}$$

$$\Pr\left\{\left\|\tilde{\boldsymbol{\theta}}^H\mathbf{G}_l\mathbf{W}_0\right\|_2^2+\sigma_{e,l}^2-E_{th,l} \geq 0, \forall l\right\} \geq \zeta_l, \forall l, \tag{34d}$$

$$(8\text{f}), (9\text{d}). \tag{34e}$$

The main difficulty lies in the constraints (34b)-(34d). In fact, (34b) is equivalent to

$$\Pr\left\{\left|\tilde{\boldsymbol{\theta}}^H\mathbf{H}_k\mathbf{w}_k\right|^2 \geq \frac{1}{s_k}\left(\left\|\tilde{\boldsymbol{\theta}}^H\mathbf{H}_k\mathbf{W}_{-k}\right\|_2^2+\sigma_{i,k}^2\right)\right\} \geq \delta_k, \tag{35a}$$

$$1 \geq t_k s_k. \tag{35b}$$

Furthermore, (35a) can be turned into

$$\begin{aligned}&\Pr\left\{\bar{\mathbf{h}}_k^H\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{h}}_k+\Delta\mathbf{h}_k^H\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\Delta\mathbf{h}_k\right.\\&\left.+2\Re\left\{\bar{\mathbf{h}}_k^H\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\Delta\mathbf{h}_k\right\} \geq \frac{\sigma_{i,k}^2}{s_k}\right\} \geq \delta_k.\end{aligned} \tag{36}$$

Next, we will handle the CSI uncertainty based on the following BTI.

Lemma 3 [42]: For any $\{\mathbf{A},\mathbf{u},c\} \in \mathbb{H}^N\times\mathbb{C}^N\times\mathbb{R}$, $\mathbf{q}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\beta\in(0,1]$, the following inequalities holds:

$$\Pr_{\mathbf{q}}\left\{\mathbf{q}^H\mathbf{A}\mathbf{q}+2\Re\left\{\mathbf{q}^H\mathbf{u}\right\}+c\geq 0\right\} \geq 1-\beta$$

$$\Leftarrow \begin{cases}\mathrm{Tr}(\mathbf{A})-\sqrt{-2\ln(\beta)}x+\ln(\beta)\,y+c\geq 0,\\ \left\|\begin{bmatrix}\mathrm{vec}(\mathbf{A})\\ \sqrt{2}\mathbf{u}\end{bmatrix}\right\|_2 \leq x,\\ y\mathbf{I}+\mathbf{A}\succeq\mathbf{0}, y\geq 0,\end{cases}$$

where x and y are slack variables. Moreover, BTI is jointly convex w.r.t. all the variables.[1]

In fact, we have $\left[\mathrm{vec}(\Delta\mathbf{H}_{R,k});\Delta\mathbf{h}_{T,k}^H\right] = \left[\mathbf{R}_{R,k}^{\frac{1}{2}}\mathbf{p}_{R,k};\mathbf{R}_{T,k}^{\frac{1}{2}}\mathbf{p}_{T,k}\right] = \mathrm{BlkD}\left(\mathbf{R}_{R,k}^{\frac{1}{2}},\mathbf{R}_{T,k}^{\frac{1}{2}}\right)[\mathbf{p}_{R,k};\mathbf{p}_{T,k}]$, where $\mathbf{R}_{R,k}^{\frac{1}{2}}$ and $\mathbf{R}_{T,k}^{\frac{1}{2}}$ are the square root of $\mathbf{R}_{R,k}$ and $\mathbf{R}_{T,k}$, respectively. In addition, $\mathbf{p}_{R,k}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\mathbf{p}_{T,k}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$. Thus, by denoting $\mathbf{R}_k^{\frac{1}{2}} \triangleq \mathrm{BlkD}\left(\mathbf{R}_{R,k}^{\frac{1}{2}},\mathbf{R}_{T,k}^{\frac{1}{2}}\right)$ and $\mathbf{p}_k \triangleq$

[1]In fact, there exists another effective method to handle the outage constraint, e.g., the large deviation inequality (LDI), which can be seen as a simplified BTI method. Readers can refer [42] for more details.

$[\mathbf{p}_{R,k};\mathbf{p}_{T,k}]$, we have $\mathbf{p}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\Delta\mathbf{h}_k^H\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\Delta\mathbf{h}_k = \mathbf{p}_k^H\mathbf{R}_k^{\frac{1}{2}}\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\mathbf{R}_k^{\frac{1}{2}}\mathbf{p}_k$.

Then, we denote $\mathbf{A}_k = \mathbf{R}_k^{\frac{1}{2}}\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\mathbf{R}_k^{\frac{1}{2}}$, $\mathbf{a}_k = \mathbf{R}_k^{\frac{1}{2}}\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{h}}_k$, and $a_k = \bar{\mathbf{h}}_k^H\left(\mathbf{Z}_k^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{h}}_k - \frac{\sigma_{i,k}^2}{s_k}$. By utilizing the BTI, the probabilistic constraint (36) can be formulated as

$$\mathrm{Tr}\left(\mathbf{A}_k\right) - \sqrt{-2\ln\left(1-\delta_k\right)}\vartheta_{1,k} + \ln\left(1-\delta_k\right)\vartheta_{2,k} + a_k \geq 0, \tag{37a}$$

$$\left\|\begin{bmatrix}\mathrm{vec}\left(\mathbf{A}_k\right)\\ \sqrt{2}\mathbf{a}_k\end{bmatrix}\right\|_2 \leq \vartheta_{1,k}, \vartheta_{2,k}\mathbf{I}+\mathbf{A}_k \succeq \mathbf{0}, \tag{37b}$$

where $\{\vartheta_{1,k},\vartheta_{2,k}\geq 0\}_{k=1}^K$ are the auxiliary variables.

Similarly, we have $\left[\mathrm{vec}\left(\Delta\mathbf{G}_{R,l}\right);\Delta\mathbf{g}_{T,l}^H\right] = \left[\mathbf{C}_{R,l}^{\frac{1}{2}}\mathbf{q}_{R,l};\mathbf{C}_{T,l}^{\frac{1}{2}}\mathbf{q}_{T,l}\right] = \mathrm{BlkD}\left(\mathbf{C}_{R,l}^{\frac{1}{2}},\mathbf{C}_{T,l}^{\frac{1}{2}}\right)[\mathbf{q}_{R,l};\mathbf{q}_{T,l}]$, where $\mathbf{C}_{R,l}^{\frac{1}{2}}$ and $\mathbf{C}_{R,l}^{\frac{1}{2}}$ are the square root of $\mathbf{C}_{R,l}$ and $\mathbf{C}_{R,l}$, respectively. In addition, $\mathbf{q}_{R,l}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\mathbf{q}_{T,l}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$. Thus, by denoting $\mathbf{C}_l^{\frac{1}{2}} \triangleq \mathrm{BlkD}\left(\mathbf{C}_{R,l}^{\frac{1}{2}},\mathbf{C}_{T,l}^{\frac{1}{2}}\right)$ and $\mathbf{q}_l \triangleq [\mathbf{q}_{R,l};\mathbf{q}_{T,l}]$, we have $\mathbf{q}_l\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\Delta\mathbf{g}_l^H\left(\mathbf{\Phi}_{k,l}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\Delta\mathbf{g}_l = \mathbf{q}_l^H\mathbf{C}_l^{\frac{1}{2}}\left(\mathbf{\Phi}_{k,l}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\mathbf{C}_l^{\frac{1}{2}}\mathbf{q}_l$.

Thus, by denoting $\mathbf{B}_{k,l} = \mathbf{C}_l^{\frac{1}{2}}\left(\mathbf{\Phi}_{k,l}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\mathbf{C}_l^{\frac{1}{2}}$, $\mathbf{b}_{k,l} = \mathbf{C}_l^{\frac{1}{2}}\left(\mathbf{\Phi}_{k,l}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{g}}_l$, and $b_{k,l} = \bar{\mathbf{g}}_l^H\left(\mathbf{\Phi}_{k,l}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{g}}_l + \sigma_{e,l}^2$, (34c) can be formulated as

$$\mathrm{Tr}\left(\mathbf{B}_{k,l}\right) - \sqrt{-2\ln\left(1-\eta_k\right)}\alpha_{1,k,l} + \ln\left(1-\eta_k\right)\alpha_{2,k,l} + b_{k,l} \geq 0, \tag{38a}$$

$$\left\|\begin{bmatrix}\mathrm{vec}\left(\mathbf{B}_{k,l}\right)\\ \sqrt{2}\mathbf{b}_{k,l}\end{bmatrix}\right\|_2 \leq \alpha_{1,k,l}, \alpha_{2,k,l}\mathbf{I}+\mathbf{B}_{k,l} \succeq \mathbf{0}, \tag{38b}$$

where $\{\alpha_{1,k,l},\alpha_{2,k,l}\geq 0\}_{k=1,l=1}^{K,L}$ are the auxiliary variables. Then, by denoting $\mathbf{D}_l = \mathbf{C}_l^{\frac{1}{2}}\left(\mathbf{\Xi}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\mathbf{C}_l^{\frac{1}{2}}$, $\mathbf{d}_l = \mathbf{C}_l^{\frac{1}{2}}\left(\mathbf{\Xi}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{g}}_l$, and $d_l = \bar{\mathbf{g}}_l^H\left(\mathbf{\Xi}^T\otimes\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H\right)\bar{\mathbf{g}}_l + \sigma_{e,l}^2 - E_{th,l}$, (34d) can be turned into

$$\mathrm{Tr}\left(\mathbf{D}_l\right) - \sqrt{-2\ln\left(1-\zeta_l\right)}\beta_{1,l} + \ln\left(1-\zeta_l\right)\beta_{2,l} + d_l \geq 0, \tag{39a}$$

$$\left\|\begin{bmatrix}\mathrm{vec}\left(\mathbf{D}_l\right)\\ \sqrt{2}\mathbf{d}_l\end{bmatrix}\right\|_2 \leq \beta_{1,l}, \beta_{2,l}\mathbf{I}+\mathbf{D}_l \succeq \mathbf{0}, \tag{39b}$$

where $\{\beta_{1,l},\beta_{2,l}\geq 0\}_{l=1}^L$ are the auxiliary variables. Combining these steps, the convex subproblem w.r.t. $\{\mathbf{w}_k\}_{k=0}^K$ is formulated as

$$\max_{\mathbf{w}_k,\kappa} \quad \kappa \tag{40a}$$

$$\text{s.t. } (8\text{f}),(9\text{d}),(11),(15),(16\text{b}),(37),(38),(39). \tag{40b}$$

As such, we have solved the outage constrained subproblem w.r.t $\{\mathbf{w}_k\}_{k=0}^K$. Next, we will solve the subproblem w.r.t. $\hat{\boldsymbol{\theta}}$.

B. Phase Shifter Optimization

Similarly with the previous method, the probability constraint w.r.t. $\hat{\boldsymbol{\theta}}$ is given by

$$\Pr\left\{\mathbf{h}_k^H\left(\tilde{\boldsymbol{\Upsilon}}_k^T\otimes\mathbf{T}\right)\mathbf{h}_k\geq\frac{\sigma_{i,k}^2}{\tilde{s}_k}\right\}\geq\delta_k, \tag{41a}$$

$$\Pr\left\{\mathbf{g}_l^H\left(\tilde{\boldsymbol{\Phi}}_{k,l}^T\otimes\mathbf{T}\right)\mathbf{g}_l+\sigma_{e,l}^2\geq 0\right\}\geq\eta_k, \tag{41b}$$

$$\Pr\left\{\mathbf{g}_l^H\left(\tilde{\boldsymbol{\Xi}}^T\otimes\mathbf{T}\right)\mathbf{g}_l+\sigma_{e,l}^2-E_{th,l}\geq 0\right\}\geq\zeta_l. \tag{41c}$$

Then, we denote $\mathbf{A}_k=\mathbf{B}_k^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Upsilon}}_k^T\otimes\mathbf{T}\right)\mathbf{B}_k^{\frac{1}{2}}$, $\mathbf{a}_k=\mathbf{B}_k^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Upsilon}}_k^T\otimes\mathbf{T}\right)\bar{\mathbf{h}}_k$, $a_k=\bar{\mathbf{h}}_k^H\left(\tilde{\boldsymbol{\Upsilon}}_k^T\otimes\mathbf{T}\right)\bar{\mathbf{h}}_k-\frac{\sigma_{i,k}^2}{\tilde{s}_k}$, $\mathbf{B}_{k,l}=\mathbf{C}_l^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Phi}}_{k,l}^T\otimes\mathbf{T}\right)\mathbf{C}_l^{\frac{1}{2}}$, $\mathbf{b}_{k,l}=\mathbf{C}_l^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Phi}}_{k,l}^T\otimes\mathbf{T}\right)\bar{\mathbf{g}}_l$, $b_{k,l}=\bar{\mathbf{g}}_l^H\left(\tilde{\boldsymbol{\Phi}}_{k,l}^T\otimes\mathbf{T}\right)\bar{\mathbf{g}}_l+\sigma_{e,l}^2$, $\mathbf{D}_l=\mathbf{C}_l^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Xi}}^T\otimes\mathbf{T}\right)\mathbf{C}_l^{\frac{1}{2}}$, $\mathbf{d}_l=\mathbf{C}_l^{\frac{1}{2}}\left(\tilde{\boldsymbol{\Xi}}^T\otimes\mathbf{T}\right)\bar{\mathbf{g}}_l$, and $d_l=\bar{\mathbf{g}}_l^H\left(\tilde{\boldsymbol{\Xi}}^T\otimes\mathbf{T}\right)\bar{\mathbf{g}}_l+\sigma_{e,l}^2-E_{th,l}$. Thus, (41b) and (41c) can be handled by (38) and (39), respectively. As a result, we have the following subproblem w.r.t. $\hat{\boldsymbol{\theta}}$

$$\max_{\hat{\boldsymbol{\theta}},\kappa}\quad\kappa \tag{42a}$$

$$\text{s.t. (8e), (37), (38), (39)},\ 1\geq t_k\tilde{s}_k,\ 1+t_k\leq e^{R_{w,k}},\ R_{w,k}\geq\kappa,\forall k. \tag{42b}$$

Then, the previously proposed PDD method can be used to optimize $\hat{\boldsymbol{\theta}}$ conveniently. In addition, the feasibility of problem (33) is similar to problem (8), and thus we omit the detail for brevity.

V. Stochastic SCA method to the statistical CSI design

In this section, we propose a stochastic SCA method to handle the robust optimization in the statistical CSI model. Firstly, the problem can be formulated as

$$\max_{\mathbf{w}_k,\hat{\boldsymbol{\theta}},R_{o,k}}\quad\min R_{o,k} \tag{43a}$$

$$\text{s.t.}\ \ \Pr\left\{\ln\left(1+\Gamma_k\right)\geq R_{o,k}\right\}\geq\delta_k,\forall\left\{\mathbf{h}_{T,k},\mathbf{H}_{R,k}\right\}\in\mathcal{H}_{s,k},\forall k, \tag{43b}$$

$$\Pr\left\{\ln\left(1+\Gamma_{l,k}\right)\leq\gamma_{o,k}\right\}\geq\eta_k,\forall\left\{\mathbf{g}_{T,l},\mathbf{G}_{R,l}\right\}\in\mathcal{G}_{s,l},\forall k,\forall l, \tag{43c}$$

$$\Pr\left\{E_l\geq E_{th,l}\right\}\geq\zeta_l,\forall\left\{\mathbf{g}_{T,l},\mathbf{G}_{R,l}\right\}\in\mathcal{G}_{s,l},\forall l, \tag{43d}$$

$$\text{(8e), (8f)}, \tag{43e}$$

and the definition of these parameters are similar to those of problem (33).

Then, we observe that $\Pr\left\{\ln\left(1+\Gamma_k\right)\geq R_{o,k}\right\}\geq\delta_k\Leftrightarrow\Pr\left\{\ln\left(1+\Gamma_k\right)\leq R_{o,k}\right\}\leq 1-\delta_k$, however no closed-form expression can be obtained for this probability constraint. An alternative method is to replace the probability constraint by a expected condition, e.g., $\Pr\left\{\ln\left(1+\Gamma_k\right)\leq R_{o,k}\right\}=\mathbb{E}_{\mathbf{H}_k}\left\{\mathbb{I}_{\ln(1+\Gamma_k)\leq R_{o,k}}\right\}$, where $\mathbb{I}_{\ln(1+\Gamma_k)\leq R_{o,k}}$ is the step function of event $\ln\left(1+\Gamma_k\right)\leq R_{o,k}$ [38]. Then, to tackle the non-smooth property of the step function, we utilize the following smooth

approximated function

$$\hat{u}_a(x) = \frac{1}{1+e^{-ax}}, \tag{44}$$

where $x = R_{o,k} - \ln(1+\Gamma_k)$ and a is the parameter to control the approximation error, i.e., larger a leads to less error. Then, by applying $\hat{u}_a(\cdot)$ and relaxing the UMC to a power constraint temporarily, we obtain the following approximated problem:

$$\max_{\mathbf{w}_k, \hat{\boldsymbol{\theta}}, \kappa} \quad \kappa \tag{45a}$$

$$\text{s.t.} \quad \mathbb{E}\left\{\hat{u}_a\left(r_k\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{H}_{s,k}\right)\right)\right\} \le 1-\delta_k, \forall k, \tag{45b}$$

$$\mathbb{E}\left\{\hat{u}_a\left(y_{k,l}\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{G}_{s,l}\right)\right)\right\} \le 1-\eta_k, \forall k, \forall l, \tag{45c}$$

$$\mathbb{E}\left\{\hat{u}_a\left(z_l\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{G}_{s,l}\right)\right)\right\} \le 1-\zeta_l, \forall l, \tag{45d}$$

$$\text{(8f)}, \left|\hat{\theta}_m\right| \le 1, \tag{45e}$$

where $r_k\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{H}_{s,k}\right) \triangleq (e^{\kappa}-1)\left(\left\|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{W}_{-k}\right\|_2^2 + \sigma_{i,k}^2\right) - \left|\hat{\boldsymbol{\theta}}^H \mathbf{H}_k \mathbf{w}_k\right|^2$, $y_{k,l}\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{G}_{s,l}\right) \triangleq \left|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{w}_k\right|^2 - (e^{\gamma_{o,k}}-1)\left(\left\|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{W}_{-k}\right\|_2^2 + \sigma_{e,l}^2\right)$, and $z_l\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{G}_{s,l}\right) \triangleq E_{th} - \left\|\hat{\boldsymbol{\theta}}^H \mathbf{G}_l \mathbf{W}_0\right\|_2^2 - \sigma_{e,l}^2$.

Here, we denote a composite variable $\boldsymbol{x} \triangleq \left[\mathbf{w}_1^H, \ldots, \mathbf{w}_K^H, \mathbf{w}_0^H, \boldsymbol{\theta}^H\right]^H$, and selection matrices $\mathbf{A}_k \in \{0,1\}^{N_t \times ((K+1)N_t+M)}, \forall k \in \{0,\ldots,K\}$ and $\mathbf{B} \in \{0,1\}^{M \times ((K+1)N_t+M)}$ that satisfy $\{\mathbf{A}_k \boldsymbol{x} = \mathbf{w}_k\}_{k=0}^K$ and $\mathbf{B}\boldsymbol{x} = \boldsymbol{\theta}$. Thus, $r_k\left(\{\mathbf{w}_k\}_{k=0}^K, \hat{\boldsymbol{\theta}}; \mathcal{H}_{s,k}\right)$ can be rewritten as $\bar{r}_k(\boldsymbol{x}; \mathcal{H}_{s,k}) \triangleq (e^{\kappa}-1)\left(\sum_{\substack{j=0,\\ j\neq k}}^K \left|\left(\boldsymbol{x}^H \mathbf{B}^H \mathbf{H}_{R,k} + \mathbf{h}_{T,k}^H\right)\mathbf{A}_j \boldsymbol{x}\right|^2 + \sigma_{i,k}^2\right) - \left|\left(\boldsymbol{x}^H \mathbf{B}^H \mathbf{H}_{R,k} + \mathbf{h}_{T,k}^H\right)\mathbf{A}_k \boldsymbol{x}\right|^2$. Similarly, we denote $\bar{y}_{k,l}(\boldsymbol{x}; \mathcal{G}_{s,l}) \triangleq \left|\left(\boldsymbol{x}^H \mathbf{B}^H \mathbf{G}_{R,l} + \mathbf{g}_{T,l}^H\right)\mathbf{A}_k \boldsymbol{x}\right|^2 - (e^{\gamma_{o,k}}-1)\left(\sum_{\substack{j=0,\\ j\neq k}}^K \left|\left(\boldsymbol{x}^H \mathbf{B}^H \mathbf{G}_{R,l} + \mathbf{g}_{T,l}^H\right)\mathbf{A}_j \boldsymbol{x}\right|^2 + \sigma_{e,l}^2\right)$, and $\bar{z}_l(\boldsymbol{x}; \mathcal{G}_{s,l}) \triangleq E_{th} - \left(\sum_{j=0}^K \left|\left(\boldsymbol{x}^H \mathbf{B}^H \mathbf{G}_{R,l} + \mathbf{g}_{T,l}^H\right)\mathbf{A}_j \boldsymbol{x}\right|^2\right) - \sigma_{e,l}^2$.

(45) is a stochastic optimization problem. Here, we propose a stochastic SCA method to address (45), mainly based on the framework in [44]. Firstly, we denote $f_k(\boldsymbol{x}) \triangleq \mathbb{E}\{\hat{u}_a(\bar{r}_k(\boldsymbol{x}; \mathcal{G}_{s,l}))\}$, $g_{k,l}(\boldsymbol{x}) \triangleq \mathbb{E}\{\hat{u}_a(\bar{y}_{k,l}(\boldsymbol{x}; \mathcal{G}_{s,l}))\}$, and $h_l(\boldsymbol{x}) \triangleq \mathbb{E}\{\hat{u}_a(\bar{z}_l(\boldsymbol{x}; \mathcal{G}_{s,l}))\}$. Then, in the i-th iteration, convex surrogate functions of $\{\bar{f}_k^i(\boldsymbol{x})\}$ are constructed to approximate $\{f_k(\boldsymbol{x})\}_{k\in\mathcal{K}}$, which can be expressed as

$$\bar{f}_k^i(\boldsymbol{x}) \triangleq f_k^i + 2\Re\left\{\left(\mathbf{f}_k^i\right)^H\left(\boldsymbol{x}-\boldsymbol{x}^i\right)\right\} + \iota_k\left\|\boldsymbol{x}-\boldsymbol{x}^i\right\|_2^2, \tag{46}$$

where $\iota_k > 0$ is a constant and $\iota_k\|\boldsymbol{x}-\boldsymbol{x}^i\|^2$ is included to ensure the convexity of $\bar{f}_k^i(\boldsymbol{x})$, f_k^i

is an approximation of $\mathbb{E}\left\{\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)\right\}$, and $\mathbf{f}_k^i$ is an approximation of the gradient $\nabla_{\boldsymbol{x}^\dagger}\mathbb{E}\left\{\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)\right\}$. Besides, f_k^i and $\mathbf{f}_k^i$ are iteratively updated by

$$f_k^i = \frac{1}{T_1}\sum_{t=1}^{T_1}\hat{u}_a\left(\bar{r}_k\left(\left\{\boldsymbol{x}^i;\mathcal{H}_{s,k}^t\right\}\right)\right), \tag{47}$$

and

$$\mathbf{f}_k^i = \left(1-v^i\right)\mathbf{f}_k^{i-1} + v^i\frac{1}{T_2}\sum_{t=1}^{T_2}\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{r}_k\left(\left\{\boldsymbol{x}^i;\mathcal{H}_{s,k}^t\right\}\right)\right), \tag{48}$$

where $f_k^{-1}=0$, $\mathbf{f}_k^{-1}=\mathbf{0}$, T_1 and T_2 denote the numbers of channel realizations utilized to approximate $\mathbb{E}\left\{\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)\right\}$ and $\nabla_{\boldsymbol{x}^\dagger}\mathbb{E}\left\{\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)\right\}$, respectively, $v^i\in(0,1]$ is a sequence properly designed following Assumption 5 in [44]. In fact, the stochastic optimization method can be seen as a online learning approach with v^i been the learning parameter.

Note that in (47) and (48), multiple channel realizations are generated to improve the approximations f_k^i and $\mathbf{f}_k^i$ in each iteration. Besides, according to [44], as $i\to\infty$, we have $\lim_{i\to 0}\left|\bar{f}_k^i\left(\boldsymbol{x}^i\right)-f_k\left(\boldsymbol{x}^i\right)\right|=0$, $\lim_{i\to 0}\left|\nabla_{\boldsymbol{x}^\dagger}\bar{f}_k^i\left(\boldsymbol{x}^i\right)-\nabla_{\boldsymbol{x}^\dagger}f_k\left(\boldsymbol{x}^i\right)\right|=0,\forall k\in\mathcal{K}$, which indicates that $\bar{f}_k^i\left(\boldsymbol{x}^i\right)$ and $\nabla_{\boldsymbol{x}^\dagger}\bar{f}_k^i\left(\boldsymbol{x}^i\right)$ can converge to the true values of $f_k\left(\boldsymbol{x}^i\right)$ and its gradient w.r.t. $\boldsymbol{x}$, respectively. Thus, the convergence of the proposed method is guaranteed. For each channel realization, $\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)$ is obtained by [34]:

$$\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right) = \frac{ae^{-a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}}{\left(1+e^{-a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}\right)^2}\bar{\mathbf{r}}_k'\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right), \tag{49}$$

where $\bar{\mathbf{r}}_k'\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right) = 2\left(e^{\kappa}-1\right)\left(\sum_{\substack{j=0,\\ j\neq k}}^{K}\mathbf{r}_{kj}\left(\boldsymbol{x};\mathcal{H}_{s,k}\right)\right) - 2\mathbf{r}_{kk}\left(\boldsymbol{x};\mathcal{H}_{s,k}\right)$, and $\mathbf{r}_{kj}\left(\boldsymbol{x};\mathcal{H}_{s,k}\right) = \mathbf{B}^H\mathbf{H}_{R,k}\mathbf{A}_j\boldsymbol{x}\boldsymbol{x}^H\mathbf{A}_j^H\mathbf{H}_{R,k}^H\mathbf{B}\boldsymbol{x}+\boldsymbol{x}^H\mathbf{B}^H\mathbf{H}_{R,k}\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{H}_{R,k}^H\mathbf{B}\boldsymbol{x}+\mathbf{A}_j^H\mathbf{h}_{T,k}\mathbf{h}_{T,k}^H\mathbf{A}_j\boldsymbol{x}+\mathbf{B}^H\mathbf{H}_{R,k}\mathbf{A}_j\boldsymbol{x}\boldsymbol{x}^H\mathbf{A}_j^H\mathbf{h}_{T,k}+\boldsymbol{x}^H\mathbf{B}^H\mathbf{H}_{R,k}\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{h}_{T,k}+\mathbf{h}_{T,k}^H\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{H}_{R,k}^H\mathbf{B}\boldsymbol{x}$.

Similar methods can be utilized to handle $\left\{g_{k,l}\left(\boldsymbol{x}\right)\right\}_{k\in\mathcal{K},l\in\mathcal{L}}$ and formulate the surrogate function $\left\{\bar{g}_{k,l}^i\left(\boldsymbol{x}\right)\right\}_{k\in\mathcal{K},l\in\mathcal{L}}$, where the associated gradient is given by

$$\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{y}_{k,l}\left(\{\boldsymbol{x};\mathcal{G}_{s,l}\}\right)\right) = \frac{ae^{-a\bar{y}_{k,l}\left(\{\boldsymbol{x};\mathcal{G}_{s,l}\}\right)}}{\left(1+e^{-a\bar{y}_{k,l}\left(\{\boldsymbol{x};\mathcal{G}_{s,l}\}\right)}\right)^2}\bar{\mathbf{y}}_{k,l}'\left(\{\boldsymbol{x};\mathcal{G}_{s,l}\}\right), \tag{50}$$

where $\bar{\mathbf{z}}_l'\left(\{\boldsymbol{x};\mathcal{G}_{s,l}\}\right) = 2\mathbf{y}_{kk,l}\left(\boldsymbol{x};\mathcal{G}_{s,l}\right) - 2\left(e^{\gamma_{o,k}}-1\right)\left(\sum_{\substack{j=0,\\ j\neq k}}^{K}\mathbf{y}_{kj,l}\left(\boldsymbol{x};\mathcal{G}_{s,l}\right)\right)$, and $\mathbf{y}_{kj,l}\left(\boldsymbol{x};\mathcal{G}_{s,l}\right) = \mathbf{B}^H\mathbf{G}_{R,l}\mathbf{A}_j\boldsymbol{x}\boldsymbol{x}^H\mathbf{A}_j^H\mathbf{G}_{R,l}^H\mathbf{B}\boldsymbol{x}+\boldsymbol{x}^H\mathbf{B}^H\mathbf{G}_{R,l}\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{G}_{R,l}^H\mathbf{B}\boldsymbol{x}+\mathbf{A}_j^H\mathbf{g}_{T,l}\mathbf{g}_{T,l}^H\mathbf{A}_j\boldsymbol{x}+\mathbf{B}^H\mathbf{G}_{R,k}\mathbf{A}_j\boldsymbol{x}\boldsymbol{x}^H\mathbf{A}_j^H\mathbf{g}_{T,l}+\boldsymbol{x}^H\mathbf{B}^H\mathbf{G}_{R,l}\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{g}_{T,l}+\mathbf{g}_{R,l}^H\mathbf{A}_j\boldsymbol{x}\mathbf{A}_j^H\mathbf{G}_{T,l}^H\mathbf{B}\boldsymbol{x}$. Surrogate function $\left\{\bar{h}_l^i\left(\boldsymbol{x}\right)\right\}_{l\in\mathcal{L}}$ can be obtained in a similar way.

Thus, with fixed κ, in the i-th iteration, we solve the following feasibility-check problem

$$\overline{\boldsymbol{x}}^i = \text{Find } \overline{\boldsymbol{x}} \tag{51a}$$

$$\text{s.t. } \bar{f}_k^i\left(\overline{\boldsymbol{x}}\right) \le 1-\delta_k, \forall k, \tag{51b}$$

$$\bar{g}_{k,l}^i\left(\overline{\boldsymbol{x}}\right) \le 1-\eta_k, \forall k, \forall l, \tag{51c}$$

$$\bar{h}_l^i\left(\overline{\boldsymbol{x}}\right) \le 1-\zeta_l, \forall l, \tag{51d}$$

$$\sum_{k=0}^{K} \overline{\boldsymbol{x}}^H \mathbf{A}_k^H \mathbf{A}_k \overline{\boldsymbol{x}} \le P_s, \left[\mathbf{B}\overline{\boldsymbol{x}}\right]_m \le 1, \tag{51e}$$

which is convex w.r.t $\overline{\boldsymbol{x}}$.

Then, we handle the relaxed constraint $\left[\mathbf{B}\overline{\boldsymbol{x}}\right]_m \le 1$ by the PDD technique. Similar to Section III, we formulate the following AL of (51)

$$\min_{\overline{\boldsymbol{x}},\boldsymbol{\varphi},\boldsymbol{\tau}} \ \frac{1}{2\lambda}\left\|\mathbf{B}\overline{\boldsymbol{x}}-\boldsymbol{\varphi}\right\|_2^2 + \Re\left\{\boldsymbol{\tau}^H\left(\mathbf{B}\overline{\boldsymbol{x}}-\boldsymbol{\varphi}\right)\right\} \tag{52a}$$

$$\text{s.t. } (51\text{b})-(51\text{e}), |\varphi_m| = 1. \tag{52b}$$

In fact, similar to the previously proposed PDD procedure, a two-layer algorithm is utilized to handle (52). Here, for the inner layer problem, with fixed $\boldsymbol{\varphi}$, $\overline{\boldsymbol{x}}^i$ can be obtained by solving (52) using CVX. Then, with fixed $\overline{\boldsymbol{x}}^i$, $\boldsymbol{\varphi}$ can be obtained in a closed form as $\boldsymbol{\varphi} = e^{\jmath\angle\left(\mathbf{B}\overline{\boldsymbol{x}}^i+\lambda\boldsymbol{\tau}\right)}$. Besides, the proposed criterions in Section III are applied to update $\boldsymbol{\tau}$ and λ.

Then, according to [44], $\boldsymbol{x}$ is updated by

$$\boldsymbol{x}^{i+1} = \left(1-r^i\right)\boldsymbol{x}^i + r^i\overline{\boldsymbol{x}}^i, \tag{53}$$

where $\{r^i \in (0,1]\}$ is a sequence satisfying $r^i \to 0$, $\sum_i r^i = \infty$, and $\sum_i \left(r^i\right)^2 < \infty$.

In addition, according to [44, Theorem 1], the SCA algorithm can converge to the set of stationary solutions to (45), almost surely. Moreover, the following remark is useful to handle the so-called "vanishing gradient" problem for (44).

Remark 1: As mentioned in [34], when $\left|\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right|$ is too large, $\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right)$ may close to zero due to the term $\frac{ae^{-a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}}{\left(1+e^{-a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}\right)^2}$ in (49). To tackle this obstacle, a modified gradient is given as:

$$\nabla_{\boldsymbol{x}^\dagger}\hat{u}_a\left(\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)\right) = \frac{ae^{-\tilde{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}}{\left(1+e^{-\tilde{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right)}\right)^2}\overline{\mathbf{r}}_k'\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right), \tag{54}$$

where

$$\tilde{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right) = \begin{cases} q, \text{ if } a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right) \ge q, \\ -q, \text{ if } a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right) \le -q, \\ a\bar{r}_k\left(\{\boldsymbol{x};\mathcal{H}_{s,k}\}\right), \text{ otherwise}, \end{cases} \tag{55}$$

and $q > 0$ is a constant which is chosen based on a. More details can refer to [34].

The whole bisection search method is summarized in Algorithm 3.

Algorithm 3 The Bisection Search Algorithm to Problem (43).

1: Initialize $\boldsymbol{x}$, set the bound $\left[\kappa^l, \kappa^u\right]$ and the accuracy ϖ_3;
2: **repeat**
3: Calculate $\kappa = \left(\kappa^l + \kappa^u\right)/2$;
4: Solve (52) by the proposed statistics SCA method;
5: If (52) is feasible, set $\kappa^l = \kappa$, otherwise set $\kappa^u = \kappa$;
6: **until** $\left|\kappa^l - \kappa^u\right| < \varpi_3$
7: **Output** $\{\boldsymbol{x}^\star, \kappa^\star\}$;

VI. Simulation Results

Here, simulation results are provided to evaluate the proposed design. As shown in Fig. 2, there exists one Tx, one IRS, 3 IRs and 3 ERs. A three-dimensional (3D) coordinate is utilized, where Tx and the IRS are located at $(10\ \mathrm{m}, 0\ \mathrm{m}, 10\ \mathrm{m})$ and $(0\ \mathrm{m}, 50\ \mathrm{m}, 10\ \mathrm{m})$, respectively. Besides, we assume that the IRs are randomly located in a circle centered at $(10\ \mathrm{m}, 50\ \mathrm{m}, 2\ \mathrm{m})$ with radius of $5\ \mathrm{m}$. While the ERs are deployed closer to Tx than IRs to enhance the EH, which are randomly located in a circle centered at $(10\ \mathrm{m}, 20\ \mathrm{m}, 2\ \mathrm{m})$, with radius of $5\ \mathrm{m}$.

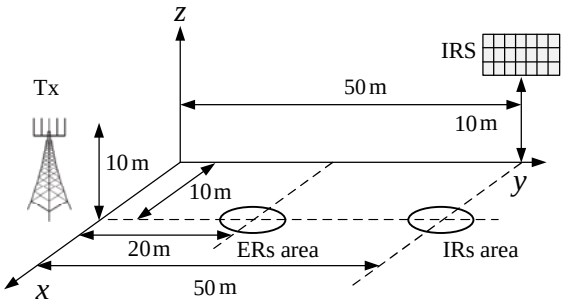

Fig. 2. The simulation scenario.

The simulation settings are assumed as: $N_t = 4$, $M = 10$, $P_s = 20$ dB, $\sigma_{i,k}^2 = \sigma_{e,l}^2 = -70\ \mathrm{dBm}, \forall k, \forall l$. Besides, the QoS requirement is assumed to be the same value among these users, i.e., $\gamma_{w,k} = \gamma_{o,k} = 1\ \mathrm{bit/s/Hz}$, $E_l = -10\ \mathrm{dBm}$, $\delta_k = 0.9$, $\eta_k = 0.9$, $\zeta_l = 0.9, \forall k, \forall l$. Here, we set a relatively large HP constraint to activate the EH circuit [2]. The path loss is given by $\mathrm{PL} = \mathrm{PL}_0 - 10\alpha \log_{10}\left(\frac{d}{d_0}\right)$, where d is the link distance, α is the path loss exponent. Here, we set $\mathrm{PL}_0 = -30$ dB and $d_0 = 1$ m [12]. The path loss exponents of the IRS-related links are set as $\alpha_{\mathrm{R}} = 2.2$ [30], while the path loss exponents between Tx and the IRs/ERs are given

by $\alpha_{\mathrm{T}} = 3.5$ [32]. As for the small-scale fading, taking $\mathbf{F}$ as an example, which is modeled as $\mathbf{F} = \sqrt{\frac{\beta}{\beta+1}}\mathbf{F}^{\mathrm{LoS}} + \sqrt{\frac{1}{\beta+1}}\mathbf{F}^{\mathrm{NLoS}}$, where β is the Rician factor, $\mathbf{F}^{\mathrm{LoS}}$ is the line of sight (LoS) component, and $\mathbf{F}^{\mathrm{NLoS}}$ is the non-LoS (NLoS) component which is modeled as Rayleigh fading. Besides, $\mathbf{F}^{\mathrm{LoS}}$ is given by $\mathbf{F}^{\mathrm{LoS}} = \mathbf{a}\mathbf{b}^H$, where $\mathbf{a} = \left[1, e^{j\frac{2\pi d_{\mathrm{r}}}{\lambda}\sin\upsilon^{\mathrm{A}}}, \cdots, e^{j\frac{2\pi d_{\mathrm{r}}}{\lambda}(D_{\mathrm{r}}-1)\sin\upsilon^{\mathrm{A}}}\right]^T$, and $\mathbf{b} = \left[1, e^{j\frac{2\pi d_{\mathrm{t}}}{\lambda}\sin\upsilon^{\mathrm{D}}}, \cdots, e^{j\frac{2\pi d_{\mathrm{t}}}{\lambda}(D_{\mathrm{t}}-1)\sin\upsilon^{\mathrm{D}}}\right]^T$, λ is the wavelength, D_{r} and D_{t} are the number of antennas at Rx and Tx, respectively, while d_{r} and d_{t} are the relevant antenna separation distance, υ^{D} and υ^{A} are the angles of departure and arrival, respectively. Similar to [32], we set $\beta = 5$ for all these links, and $d_{\mathrm{r}}/\lambda = d_{\mathrm{t}}/\lambda = 0.5$, $\upsilon^{\mathrm{A}} = \tan^{-1}\left(\frac{y_{\mathrm{IRS}}-y_{\mathrm{Tx}}}{x_{\mathrm{IRS}}-x_{\mathrm{Tx}}}\right)$, $\upsilon^{\mathrm{D}} = \pi - \upsilon^{\mathrm{A}}$, respectively.[2]

In addition, for the probabilistic CSI error, the covariance of $\Delta\mathbf{h}_{T,k}$ and $\mathrm{vec}\left(\Delta\mathbf{H}_{R,k}\right)$ are defined as $\varepsilon_{T,k}^2 = \varepsilon_T^2\left\|\bar{\mathbf{h}}_{T,k}\right\|_2^2$ and $\varepsilon_{R,k}^2 = \varepsilon_R^2\left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{R,k}\right)\right\|_2^2$, respectively, where $\varepsilon_T \in [0,1)$ and $\varepsilon_R \in [0,1)$ measure the relative number of the CSI errors and are set as $\varepsilon_T^2 = 10^{-4}$ and $\varepsilon_R^2 = 10^{-4}$, respectively. Moreover, according to [30], a fair performance comparison between the two CSI uncertainty models can be obtained when the bounded uncertainty radii are set as $\epsilon_{T,k} = \sqrt{\frac{\varepsilon_{T,k}^2}{2}F_{2MN_t}^{-1}(1-\rho)}$, and $\epsilon_{R,k} = \sqrt{\frac{\varepsilon_{R,k}^2}{2}F_{2N_t}^{-1}(1-\rho)}$, where $F_{2MN_t}^{-1}$ and $F_{2N_t}^{-1}$ denote the inverse cumulative distribution function (CDF) of the Chi-square distribution with degrees of freedom (DoF) $2MN_t$ and $2N_t$, respectively, and ρ denotes the corresponding probability. Similar way can be used to set $\Delta\mathbf{g}_{T,l}$ and $\mathrm{vec}\left(\Delta\mathbf{G}_{R,l}\right)$.

A. Convergence Behaviour

Firstly, the convergence behaviour of the inner layer PDD algorithm with different transmit antenna number N_t and phase shifter number M are tested. Both the worst case design and outage case design are examined. From Fig. 3, we can see that both the worst case information rate (WCIR) and the outage case information rate (OCIR) increase with the number of iterations, and gradually converge, almost in 10 iterations, which validates the efficiency of the PDD method. Besides, it can be found that we draw this figure from the zeroth iteration, i.e., start from the initial point. As expected, the RIR for both the outage case and the worst case obtains the same value at the zeroth iteration.

Then, we evaluate the convergence behaviour of the outer layer AO algorithm. Fig. 4 shows the WCIR and the OCIR versus the number of iterations with different N_t and M. From Fig. 4, we can see that both WCIR and OCIR increase with the number of iterations, and gradually converge. Moreover, larger N_t or M leads to higher complexity and slower convergence, since more variables need to be optimized. However, for different values of N_t and M, both WCIR

[2]In the simulation, besides the variable in x-axis, all the other parameters are fixed to the some values.

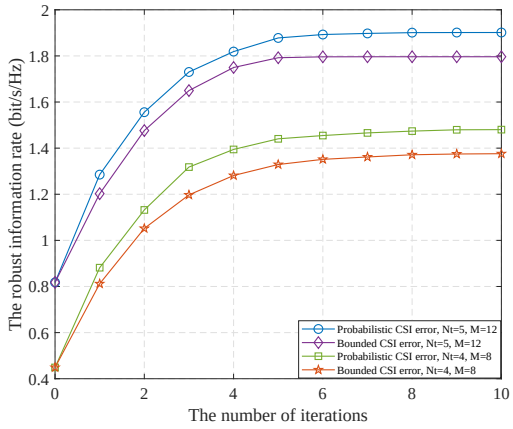

Fig. 3. The convergence behaviour of the PDD algorithm.

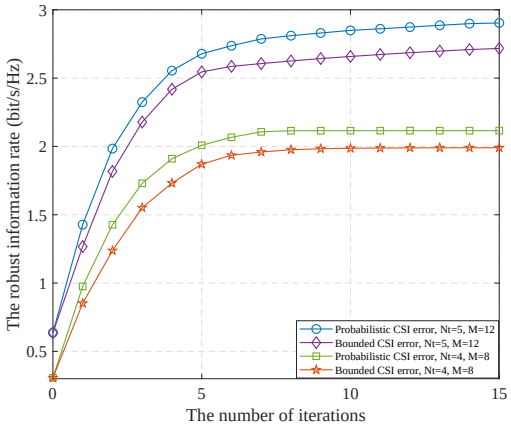

Fig. 4. The convergence behaviour of the AO algorithm.

and OCIR always converge in 15 iterations, which confirms the practicality of the AO method. Besides, given the same N_t, M and the channel condition, the probabilistic CSI design can attain high RIR than the bounded CSI design, since the bounded CSI design is more strict than the probabilistic CSI design, thus leading to lower information rates. In addition, similar to Fig. 3, the RIRs for both the outage case and the worst case are the same at the zeroth iteration.

Lastly, we evaluate the convergence behaviour of the stochastic SCA algorithm, where we set $T_1 = 10^5$, $T_2 = 200$, $v^i = (1+i)^{-0.5}$, $r^i = (1+i)^{-0.6}$, $a = 100$, and $q = 8$, respectively [34]. Similar to [35], here we assume that both the estimated CSI and the CSI error follow the CSCG distributions. Fig. 5 shows the obtained objective value of problem (52) versus the number of iterations with different N_t and M. From this figure, we can see that the stochastic SCA method converges not monotonic due to the stochastic nature, depending on the random channel generations in each iteration. However, it is able to converge in about 100 iterations.

B. RIR versus the System Parameters

Here, we assess the effect of the system parameters on the RIR. In addition, we compare the proposed methods with several baselines: 1) the perfect CSI case, which can be seen as the performance bound of the proposed method; 2) the penalty based method in [25]; 3) the no IRS-aided design; 4) the IRS-aided design but without AN. Note that this comparison can indicate which of the IRS and AN plays a more important role in improving the secrecy performance. These methods are labelled as "Bounded CSI error", "Probabilistic CSI error", "Statistical CSI", "Perfect CSI", "Penalty method", "No IRS", and "No AN", respectively. In addition, all the results are obtained over 100 channel average.

First, we show the RIR versus the total power budget P_s for Tx, and the result is shown in Fig. 6. From this figure, we can see that the RIR reaches a saturation with the increase of

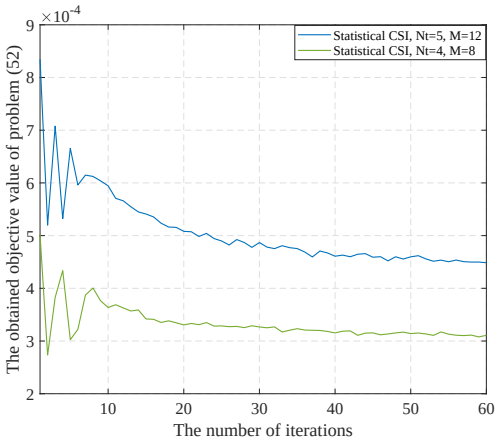

Fig. 5. The convergence behaviour of the SCA algorithm.

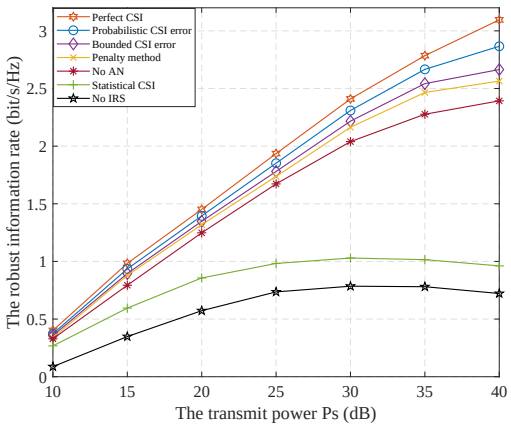

Fig. 6. The RIR versus Tx transmit power.

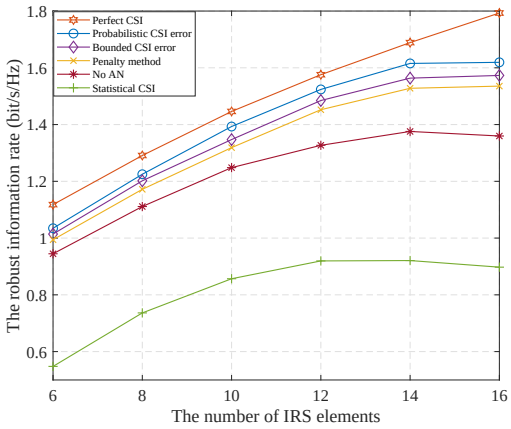

Fig. 7. The RIR versus the number of IRS elements.

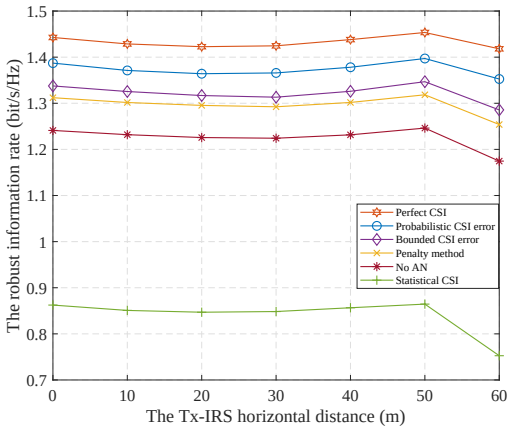

Fig. 8. The RIR versus the Tx-IRS horizontal distance.

P_s, since when P_s is relatively small, the RIR metric is mainly determined by P_s. By contrast, when P_s tends to be large, the effect of the imperfect CSI increases gradually, and thus the RIR tends to decrease. Furthermore, we can see that the performance gap between the bounded CSI error and the probabilistic CSI error model is more pronounced in high P_s region, while the stochastic SCA method suffer certain performance loss due to lack of instantaneous CSI. However, the proposed AO-SCA method outperforms the other baselines. In addition, the no AN method obtains much better performance than the no IRS-aided method, which indicates that IRS plays a major role in improving the security than AN.

Then, we show the RIR versus the number of IRS elements M in Fig. 7. From this figure, we can see that in the perfect CSI case, the information rate increases with the increase of M, since higher spatial DoF can be achieved with larger M. However, for other cases, due to the CSI error, when M becomes large, the RIR tends to decrease due to the increased CSI uncertainty.

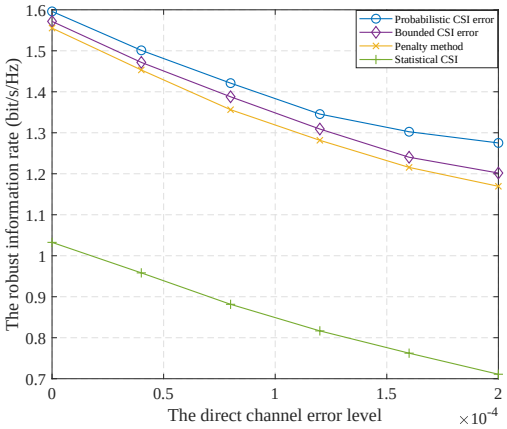

Fig. 9. The RIR versus the direct channel error level.

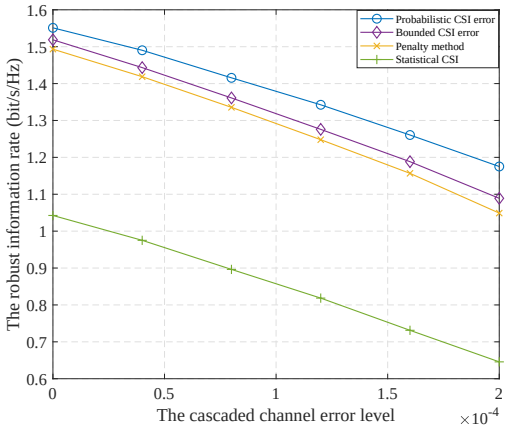

Fig. 10. The RIR versus the cascaded channel error level.

Hence, whether to utilize IRS should depend on the CSI uncertainty level.

Lastly, we plot the RIR versus the Tx-IRS horizontal distance in Fig. 8, where we assume that the IRS moves along the y-axis. It is observed that when IRS moves from Tx to IRs (about $y_{\mathrm{IRS}} \leq 50$ m), the RIR decreases than increases. On the other hand, when IRS moves away from IRs (about $y_{\mathrm{IRS}} \geq 50$ m), the RIR decreases rapidly due to large channel fading. Actually, the large-scale channel gain of the cascaded link can be estimated by $\mathrm{PL} = \mathrm{PL}_0^2 (d/d_0)^{-\alpha_{\mathrm{R}}} ((D-d)/d_0)^{-\alpha_{\mathrm{R}}}$, where d and D denote the distances from Tx to the IRS, and that from Tx to the center of the IRs area, respectively. Thus, the cascaded channel gain obtains the minimum value when $d = D/2$, e.g., the IRS is located close to the middle point. Hence, to improve the performance, it is beneficial to deploy the IRS close to Tx or IRs.

C. RIR versus the CSI Uncertainty Level

To further show the effect of the imperfect CSI, we evaluate the RIR versus the direct channel uncertainty level and the cascaded channel uncertainty level, in Figs. 9 and 10, respectively. From these results, we conclude the following facts: 1) the RIR tends to decrease with the increase of the CSI uncertainty, for both the direct and cascaded CSI uncertainty level; 2) the cascaded CSI uncertainty plays a more important role in the RIR than the direct CSI uncertainty.

D. RIR versus the User's QoS Requirement

Here, we evaluate the secrecy performance versus the QoS requirement. Firstly, we show the RIR of these schemes versus the outage probabilities δ_k and ζ_l in Figs. 11 and 12, respectively. From the two figures, we can see that the RIR decreases with the increase of δ_k and ζ_l. Since with the increase of δ_k or ζ_l, stricter constraints are imposed on the rate and the achieved rate tends to decrease to satisfy these constraints.

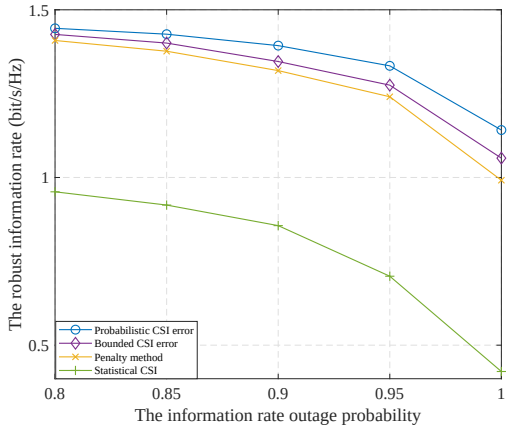

Fig. 11. The RIR versus the rate outage probability.

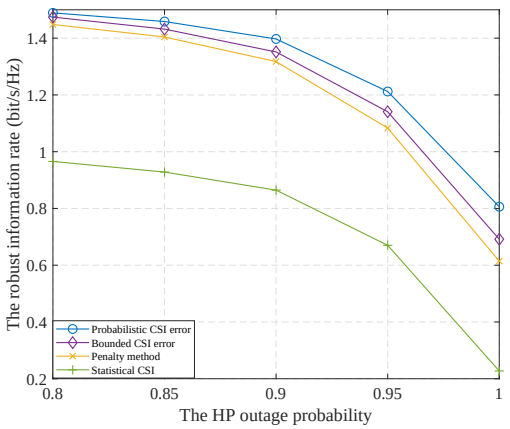

Fig. 12. The RIR versus the HP outage probability.

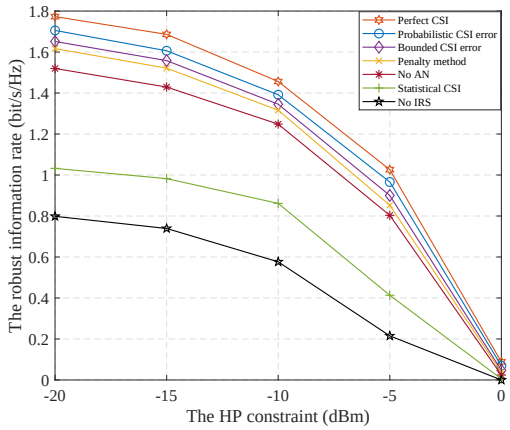

Fig. 13. The RIR versus the HP constraint.

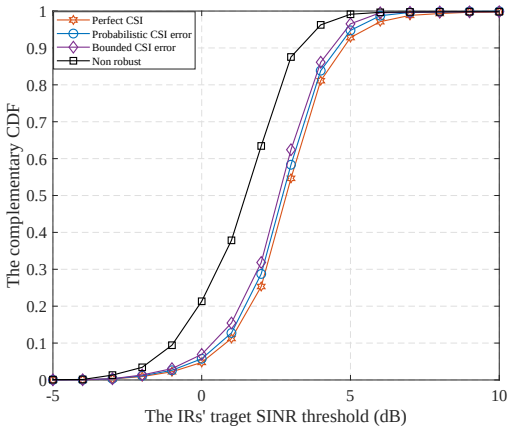

Fig. 14. The CCDF versus the IRs' target SINR threshold.

Then, we show the RIR versus the HP threshold $E_{th,l}$ in Fig. 13. From this figure, we can see that for all these methods, the RIR decreases with the increase of the HP threshold. This is mainly due to the fact that with higher HP requirements, more transmit signals need to align to the ER's channel, and thus the ER has higher probability to eavesdrop the confidential information, hence decreasing the rate. Besides, when the HP threshold exceeds a certain value (about -5 dBm), the RIR decreases dramatically. This is because in this case, the HP constraint is hard to be satisfied in many channel realizations.

E. Complementary CDF Performance

Moreover, to show the effect of the robust design, we show the complementary CDF (CCDF) of these methods versus the users' QoS parameters in Figs. 14–16, which is defined as the probability of the achieved Γ_k, $\Gamma_{k,l}$, and $E_{th,l}$ that below the target threshold. In addition,

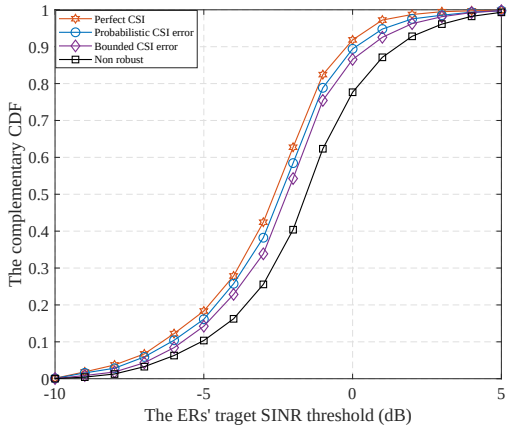

Fig. 15. The CCDF versus the ERs' target SINR threshold.

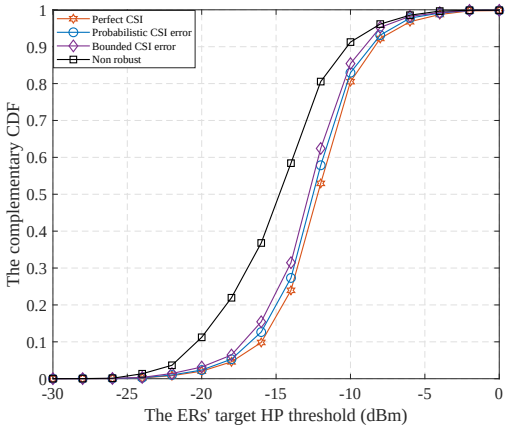

Fig. 16. The CCDF versus the ERs' target HP threshold.

the non robust design is provided for performance comparison, which is labelled as "Non robust", where $\{\mathbf{w}_k\}_{k=0}^{K}$ and $\boldsymbol{\theta}$ are obtained based on the estimated CSI without considering the outage constraints. From these figures, we can see that the robust methods achieve better CCDF performance than the non robust design, since the robust methods can provide more opportunities to obtain higher SINR at the IRs, or lower SINR at the ERs, or more HP at the ERs, thus to improve the outage performance.

VII. Conclusions

In this work, we investigated the secrecy design in IRS-aided SWIPT networks, where the BF, the AN and the phase shifters are optimized to maximize the minimum RIR. Specifically, for the bounded CSI error model and the probabilistic CSI error model, we utilized the AO and SCA method to reformulate the non-convex problem, and then an PDD based iterative algorithm was proposed to obtain the solutions. While for the statistical CSI model, we propose a stochastic SCA method to handle the robust design. Simulation results demonstrated the performance of the proposed design and reveal some useful insights.

Appendix A

Proof of Theorem 1.

Firstly, we denote the constraint set of $\hat{\boldsymbol{\theta}}$ for problem (27) as $\mathcal{X} = \left\{\hat{\boldsymbol{\theta}} \middle| \left|\hat{\theta}_m\right| \leq 1\ , (23), (24), (25)\right\}$, and rewrite the constraints (23)–(25) in functional forms as $\left\{\mathbf{F}_n\left(\hat{\boldsymbol{\theta}}\right)\right\}_{n=1}^{K+KL+L}$. In addition, the Lagrangian dual matrices associated with these constraints are denoted as $\{\mathbf{X}_n\}_{n=1}^{K+KL+L}$, and the dual variable associated with $\left|\hat{\theta}_m\right| \leq 1$ is denoted as ω_m.

Since $\hat{\boldsymbol{\theta}}^i$ is bounded by $\left|\hat{\theta}_m\right| \leq 1$, and $\boldsymbol{\varphi}$ is updated via continuous function $\boldsymbol{\varphi} = e^{\jmath\angle(\hat{\boldsymbol{\theta}}+\lambda\boldsymbol{\tau})}$ of $\hat{\boldsymbol{\theta}}^i$, there exists limit point of the solution. Without loss of generality, we denote $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\}$ as the limit point, i.e. $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\} \to \left\{\overline{\hat{\boldsymbol{\theta}}}, \overline{\boldsymbol{\varphi}}\right\}$. For any $i \in \{1, 2, \ldots\}$, assume that $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\}$ is the limit of the convergent subsequence $\left\{\hat{\boldsymbol{\theta}}^{i,\ell_j}, \boldsymbol{\varphi}^{i,\ell_j}\right\}$ of $\left\{\hat{\boldsymbol{\theta}}^{i,\ell}, \boldsymbol{\varphi}^{i,\ell}\right\}$. Denote the objective function of (28) as $h\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}\right)$. Since for any λ, the objective sequence of (28) updated via the BCD method is decreasing and bounded from below [35]. Thus, according to [43], the proposed algorithm is guaranteed to converge to a stationary point of (27).

Suppose the limit point $\boldsymbol{\varphi}^{i,\ell_j-1} \to \widehat{\boldsymbol{\varphi}}$, since $\boldsymbol{\varphi}^{i,\ell_j} = \arg\min_{|\varphi_m|=1} h\left(\hat{\boldsymbol{\theta}}^{i,\ell_j}, \boldsymbol{\varphi}\right)$, so $h\left(\hat{\boldsymbol{\theta}}^{i,\ell_j}, \boldsymbol{\varphi}^{i,\ell_j}\right) \leq h\left(\hat{\boldsymbol{\theta}}^{i,\ell_j}, \boldsymbol{\varphi}\right)$. Then taking $j \to \infty$, we have $h\left(\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right) \leq h\left(\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}\right)$. Hence $\boldsymbol{\varphi}^i$ is the optimal solution to the problem $\min_{|\varphi_m|=1} h\left(\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}\right)$. Besides, since the objective sequence $\left\{h^{i,\ell}\right\}_{\ell=1}^{\infty}$ converges, we have $h\left(\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right) = h\left(\hat{\boldsymbol{\theta}}^i, \widehat{\boldsymbol{\varphi}}\right)$. Therefore $\widehat{\boldsymbol{\varphi}}$ is also an optimal solution to the problem $\min_{|\varphi_m|=1} h\left(\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}\right)$, thus we conclude that $\boldsymbol{\varphi}^i = \widehat{\boldsymbol{\varphi}}$.

Since $\hat{\boldsymbol{\theta}}^{i,\ell_j}$ is the solution to the problem $\min_{\hat{\boldsymbol{\theta}}\in\mathcal{X}} h\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}^{i,\ell_j}\right)$, and the Slater's condition is commonly satisfied, therefore it satisfies the Robinson's condition [43]. Then, according to Lemma 3.26 of [45], the dual variables $\left\{\mathbf{X}_n^{i,\ell_j}\right\}$ and $\boldsymbol{\omega}^{i,\ell_j}$ are bounded. Then, by taking $j \to \infty$ and restricting to a subsequence of $\{\mathbf{X}_n^i\}$ and $\boldsymbol{\omega}^i$, we obtain the KKT conditions for problem $\min_{\hat{\boldsymbol{\theta}}\in\mathcal{X}} h\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}^{i,\ell_j}\right)$ at $\left\{\hat{\boldsymbol{\theta}}^i, \boldsymbol{\varphi}^i\right\}$ as follows:

$$\frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i\right) + \boldsymbol{\tau}^i + \sum_{n=1}^{K+KL+L} \mathbf{X}_n^i \frac{\partial \mathbf{F}_n\left(\hat{\boldsymbol{\theta}}^i\right)}{\partial\left(\hat{\boldsymbol{\theta}}^k\right)^\dagger} + \sum_{m=1}^{M+1} \omega_m^i \frac{\partial}{\partial\left(\hat{\theta}_m^i\right)^\dagger}\left(\left|\hat{\theta}_m^i\right| - 1\right) = \mathbf{0}, \tag{56a}$$

$$\mathbf{F}_n\left(\hat{\boldsymbol{\theta}}^i\right) \succeq \mathbf{0}, \mathbf{X}_n^i \mathbf{F}_n\left(\hat{\boldsymbol{\theta}}^i\right) = \mathbf{0}, \mathbf{X}_n^i \succeq \mathbf{0}, \forall n, \tag{56b}$$

$$\left|\hat{\theta}_m^i\right| \leq 1, \omega_m^i\left(\left|\hat{\theta}_m^i\right| - 1\right) = 0, \omega_m^i \geq 0, \forall m. \tag{56c}$$

Denote $\boldsymbol{\mu}^i \triangleq \frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i\right) + \boldsymbol{\tau}^i$, and assume $\boldsymbol{\mu}^i$ is bounded. Then, by taking $i \to \infty$ and restricting to a subsequence of $\left\{\bar{\mathbf{X}}_j\right\}$ and $\overline{\boldsymbol{\omega}}$, we obtain

$$\frac{\partial}{\partial \overline{\hat{\boldsymbol{\theta}}}^\dagger} \Re\left\{\bar{\boldsymbol{\mu}}^H\left(\overline{\hat{\boldsymbol{\theta}}} - \overline{\boldsymbol{\varphi}}\right)\right\} + \sum_{n=1}^{K+KL+L} \bar{\mathbf{X}}_n \frac{\partial \mathbf{F}_n\left(\overline{\hat{\boldsymbol{\theta}}}\right)}{\partial\left(\overline{\hat{\boldsymbol{\theta}}}\right)^\dagger} + \sum_{m=1}^{M+1} \bar{\omega}_m \frac{\partial}{\partial \overline{\hat{\theta}}_m^\dagger}\left(\left|\overline{\hat{\theta}}_m\right| - 1\right) = \mathbf{0}, \tag{57a}$$

$$\mathbf{F}_n\left(\overline{\hat{\boldsymbol{\theta}}}\right) \succeq \mathbf{0}, \bar{\mathbf{X}}_n \mathbf{F}_n\left(\overline{\hat{\boldsymbol{\theta}}}\right) = \mathbf{0}, \bar{\mathbf{X}}_n \succeq \mathbf{0}, \forall n, \tag{57b}$$

$$\left|\overline{\hat{\theta}}_m\right| \leq 1, \bar{\omega}_m\left(\left|\overline{\hat{\theta}}_m\right| - 1\right) = 0, \bar{\omega}_m \geq 0, \forall m. \tag{57c}$$

Since $\boldsymbol{\mu}^i$ is bounded, then $\boldsymbol{\tau}^i$ is also bounded. Thus, at least one of the two possible cases is performed. That is either 1): $\lambda^i \to 0$ with $\boldsymbol{\mu}^i - \boldsymbol{\tau}^i$ bounded, or 2): $\boldsymbol{\mu}^i - \boldsymbol{\tau}^i \to \mathbf{0}$ with λ^i bounded.

Hence we always have $\hat{\boldsymbol{\theta}}^i - \boldsymbol{\varphi}^i = \lambda^i \left(\boldsymbol{\mu}^i - \boldsymbol{\tau}^i\right) \to \mathbf{0}$.

Then taking $i \to \infty$ and restricting to a subsequence of $\overline{\boldsymbol{\mu}}$, we obtain

$$\overline{\hat{\boldsymbol{\theta}}} - \bar{\boldsymbol{\varphi}} = \mathbf{0}, \Re\left\{\bar{\boldsymbol{\mu}}^\dagger \circ \left(\overline{\hat{\boldsymbol{\theta}}} - \bar{\boldsymbol{\varphi}}\right)\right\} = \mathbf{0}, \Re\left\{\bar{\boldsymbol{\mu}} \circ \left(\overline{\hat{\boldsymbol{\theta}}} - \bar{\boldsymbol{\varphi}}\right)\right\} = \mathbf{0}. \tag{58}$$

At the same time, when $\hat{\boldsymbol{\theta}}^{i,\ell_j}$ is given, $\boldsymbol{\varphi}$ is updated by solving (31). According to previous discussion, $\hat{\boldsymbol{\theta}}^{i,\ell_j}$ is optimal to (31) if and only if it is optimal to (30), whose KKT condition is

$$-\frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^{i,\ell_j} - \boldsymbol{\varphi}^{i,\ell_j}\right) - \boldsymbol{\tau}^i + \sum_{m=1}^{M+1} \varsigma_m^{i,\ell_j} \frac{\partial}{\partial\left(\varphi_m^{i,\ell_j}\right)^\dagger}\left(\left|\varphi_m^{i,\ell_j}\right| - 1\right) = \mathbf{0}, \tag{59a}$$

$$\boldsymbol{\varsigma}^{i,\ell_j} \circ \left(\left|\boldsymbol{\varphi}^{i,\ell_j} - \mathbf{1}\right|\right) = \mathbf{0}, \boldsymbol{\varphi}^{i,\ell_j} = \mathbf{1}. \tag{59b}$$

In fact, by (59a), we can obtain $\boldsymbol{\varsigma}^{i,\ell_j} = \left(\boldsymbol{\varphi}^{i,\ell_j}\right)^{-1} \circ \left(\frac{1}{\lambda^i}\left(\hat{\boldsymbol{\theta}}^{i,\ell_j} - \boldsymbol{\varphi}^{i,\ell_j}\right) + \boldsymbol{\tau}^i\right)$, which is a continuous function in $\hat{\boldsymbol{\theta}}^{i,\ell_j}$ and $\boldsymbol{\varphi}^{i,\ell_j}$. By taking $j \to \infty$, $i \to \infty$ and restricting to a subsequence of $\bar{\boldsymbol{\varsigma}}$, we obtain

$$\frac{\partial}{\partial \bar{\boldsymbol{\varphi}}^\dagger} \Re\left\{\bar{\boldsymbol{\mu}}^H \left(\overline{\hat{\boldsymbol{\theta}}} - \bar{\boldsymbol{\varphi}}\right)\right\} + \sum_{m=1}^{M+1} \bar{\varsigma}_m \frac{\partial}{\partial \bar{\varphi}_m^\dagger}\left(|\bar{\varphi}_m| - 1\right) = \mathbf{0}, \tag{60a}$$

$$\bar{\boldsymbol{\varsigma}}^\dagger \circ \left(|\bar{\boldsymbol{\varphi}} - \mathbf{1}|\right) = \mathbf{0}, \bar{\boldsymbol{\varphi}} = \mathbf{1}. \tag{60b}$$

Combining the equations (57), (58) and (60), we conclude that the Lagrangian multipliers $\left\{\bar{\mathbf{X}}_j, \bar{\boldsymbol{\omega}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\varsigma}}\right\}$ satisfying the KKT conditions of (27), thus completes the proof.

REFERENCES

[1] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[2] X. Lu, P. Wang, D. Niyato, D. I. Kim and Z. Han, "Wireless networks with RF energy harvesting: A contemporary survey," *IEEE Commun. Surveys Tut.*, vol. 17, no. 2, pp. 757–789, 2nd Quart. 2015.

[3] Y. Wu, A. Khisti, C. Xiao, G. Caire, K.-K. Wong, and X. Gao, "A survey of physical layer security techniques for 5G wireless networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, Apr. 2018.

[4] Y. Dong, A. E. Shafie, M. J. Hossain, J. Cheng, N. A. Dhahir, and V. C. M. Leung, "Secure beamforming in full-duplex MISO-SWIPT systems with multiple eavesdroppers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6559–6574, Oct. 2018.

[5] Y. Dong, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Robust energy efficient beamforming in MISOME-SWIPT systems with proportional secrecy rate," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 202–215, Jan. 2019.

[6] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[7] S. Gong, X. Lu, D. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y. C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tut.*, vol. 22, no. 4, pp. 2283–2314, 4th Quart. 2020.

[8] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[9] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast MISO communication systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236–3251, Apr. 2020.

[10] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 586–590, May. 2020.

[11] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, Aug. 2020.

[12] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[13] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[14] 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.

[15] H. 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, Jul. 2020.

[16] 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.*, doi: 10.1109/TCOMM.2021.3085780.

[17] L. Dong and H. Wang, "Enhancing secure MIMO transmission via intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7543–7556, Nov. 2020.

[18] 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.

[19] Z. Chu, W. Hao, P. Xiao, D. Mi, Z. Liu, M. Khalily, J. R. Kelly, and A. P. Feresidis, "Secrecy rate optimization for intelligent reflecting surface assisted MIMO system," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1655–1669, 2020.

[20] M. D. Renzo et al., "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 798–807, 2020.

[21] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[22] Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint transmit precoding and reconfigurable intelligent surface phase adjustment: A decomposition aided channel estimation approach," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1228–1243, Feb. 2021.

[23] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.

[24] X. Lu, E. Hossain, T. Shafique, S. Feng, H. Jiang, and D. Niyato, "Intelligent reflecting surface (IRS)-enabled covert communications in wireless networks," *IEEE Network*, vol. 34, no. 5, pp. 148–155, Sep./Oct. 2020.

[25] 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.

[26] 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.

[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] T. A. Le, T. V. Chien, and M. D. Renzo, "Robust probabilistic-constrained optimization for IRS-aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 1–5, Jan. 2021.

[29] J. Zhang, Y. Zhang, C. Zhong, and Z. Zhang, "Robust design for intelligent reflecting surfaces assisted MISO systems," *IEEE Wireless Commun. Lett.*, vol. 24, no. 10, pp. 2353–2357, Oct. 2020.

[30] 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, Aug. 2020.

[31] 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.

[32] S. Hong, C. Pan, G. Zhou, H. Ren, and K. Wang, "Outage constrained robust transmission design for IRS-aided secure communications with direct communication links," 2020, *arXiv: 2011.09822v2*, [Online]. Available: https://arxiv.org/abs/2011.09822.

[33] Y. Omid, S. M. Shahabi, C. Pan, Y. Deng, and A. Nallanathan, "Low-complexity robust beamforming design for IRS-aided MISO systems with imperfect channels," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1697–1701, May. 2021.

[34] M.-M. Zhao, A. Liu, and R. Zhang, "Outage-constrained robust beamforming for intelligent reflecting surface aided wireless communication," *IEEE Trans. Signal Process.*, vol. 69, pp. 1301–1316, Feb. 2021.

[35] Y. Liu, J. Zhao, M. Li, and Q. Wu, "Intelligent reflecting surface aided MISO uplink communication network: Feasibility and power minimization for perfect and imperfect CSI," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1975–1989, Mar. 2021.

[36] K. Zhi, C. Pan, H. Ren and K. Wang, "Statistical CSI-based design for reconfigurable intelligent surface-aided massive MIMO systems with direct links," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1128–1132, May. 2021.

[37] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. D. Renzo, "Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages," *IEEE Trans. Veh. Tech.*, vol. 70, no. 1, pp. 1057–1061, Jan. 2021.

[38] G. Zhou, C. Pan, H. Ren, K. Wang, and M. D. Renzo, "Fairness-oriented multiple RISs-aided mmWave transmission: Stochastic optimization approaches," 2020, *arXiv: 2012.06103v2*, [Online]. Available: https://arxiv.org/abs/2012.06103.

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

[40] Y. Cai, F. Cui, Q. Shi, Y. Wu, B. Champagne, and L. Hanzo, "Secure hybrid A/D beamforming for hardware-efficient large-scale multiple-antenna SWIPT systems," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6141–6156, Oct. 2020.

[41] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.0 beta, 2012. [online]. Available: http://cvxr.com/cvx.

[42] K. 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.

[43] Q. Shi and M. Hong, "Penalty dual decomposition method for nonsmooth nonconvex optimization-Part I: Algorithms and convergence analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 4108–4122, Jun. 2020.

[44] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4189–4203, Aug. 2019.

[45] A. Ruszczynski, *Nonlinear Optimization.* Princeton, NJ, USA: Princeton Univ. Press, 2011.