



**HAL**  
open science

# Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages

Gui Zhou, Cunhua Pan, Hong Ren, Kezhi Wang, Maged ElKashlan, Marco Di Renzo

► **To cite this version:**

Gui Zhou, Cunhua Pan, Hong Ren, Kezhi Wang, Maged ElKashlan, et al.. Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages. IEEE Transactions on Vehicular Technology, 2021, 70 (1), pp.1057 - 1061. 10.1109/TVT.2021.3049257 . hal-03358196

**HAL Id: hal-03358196**

**<https://hal.science/hal-03358196>**

Submitted on 29 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages

Gui Zhou, Cunhua Pan, Hong Ren, Kezhi Wang, Maged ElKashlan, and Marco Di Renzo, *Fellow, IEEE*

## Abstract

A fundamental challenge for millimeter wave (mmWave) communications lies in its sensitivity to the presence of blockages, which impact the connectivity of the communication links and ultimately the reliability of the network. In this paper, we analyze a mmWave communication system assisted by multiple reconfigurable intelligent surface (RISs) for enhancing the network reliability and connectivity in the presence of random blockages. To enhance the robustness of beamforming in the presence of random blockages, we formulate a stochastic optimization problem based on the minimization of the sum outage probability. To tackle the proposed optimization problem, we introduce a low-complexity algorithm based on the stochastic block gradient descent method, which learns sensible blockage patterns without searching for all combinations of potentially blocked links. Numerical results confirm the performance benefits of the proposed algorithm in terms of outage probability and effective data rate.

## I. INTRODUCTION

Due to the abundance of available frequency bandwidth, millimeter wave (mmWave) communication is envisioned to be a potential technology to meet the high data rate demand of current wireless networks. In addition, due to the small wavelength, a large number of antenna elements can be packed

(Corresponding author: Cunhua Pan) G. Zhou, C. Pan and M. ElKashlan are with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. (e-mail: g.zhou, c.pan, maged.elkashlan@qmul.ac.uk). H. Ren is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (hren@seu.edu.cn). K. Wang is with Department of Computer and Information Sciences, Northumbria University, UK. (e-mail: kezhi.wang@northumbria.ac.uk). M. Di Renzo is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 Rue Joliot-Curie, 91192 Gif-sur-Yvette, France. (marco.di-renzo@universite-paris-saclay.fr) M. Di Renzo's work was supported in part by the European Commission through the H2020 ARIADNE project under grant agreement 675806.

in antenna arrays of reasonable size, which can compensate for the severe path loss caused by the high transmission frequency and can mitigate the inter-user interference by capitalizing on the design of high-directional beams [1]. Moreover, the use of hybrid analog-digital array reduces the cost and power consumption of using many radio frequency (RF) chains and full-digital signal processing baseband units at mmWave base stations (BSs) [2].

However, the mmWave signals experience high penetration losses and a low diffraction from objects [3], which make mmWave systems highly sensible to the presence of spatial blockages (e.g., buildings, human beings, etc.) and degrades the reliability of the communication links. To address this challenge, recent research works [4], [5] have proposed some robust beamforming designs to tackle the channel uncertainties due to the presence of random blockages. In particular, by capitalizing on the predicted blockage probability, [4] proposed a worst-case robust coordinated multipoint (CoMP) beamforming design by considering all possible combinations of blockage patterns. To reduce the complexity and improve the robustness of mmWave communication that suffers from the presence of random blockages, an outage-minimum strategy based on a stochastic optimization method was proposed in [5].

However, the above methods are only suitable for CoMP scenarios, where the space diversity gain from the multiple BSs reduces the outage caused by the presence of random blockages. Deploying multiple BSs, however, increases the hardware cost and power consumption. To overcome these challenges, the emerging technology of reconfigurable intelligent surfaces (RISs) was recently proposed as a promising solution for establishing alternative communication routes at a low cost, high energy efficiency, and high reliability [6]–[11]. An RIS is a surface made of nearly-passive and reconfigurable scatterers, which are capable of modifying the incident radio waves, so as to enhance the received signals at some specified locations [2], [12]–[17].

In this paper, motivated by these considerations, we study an mmWave RISs-aided communication system, and optimize the hybrid analog-digital beamforming at the BS and the passive beamforming at the RISs by explicitly taking into account the presence of random blockages. To enhance the robustness of the considered system, we study a sum-outage probability minimization problem that minimizes the rate at which the system quality of service (QoS), which is quantified in terms of outage probability, is not fulfilled. The formulated stochastic optimization problem with coupled variables is solved by using the block stochastic gradient descent (BSGD) method. The obtained simulation results show the performance benefits of the proposed scheme in terms of outage probability and effective data rate.

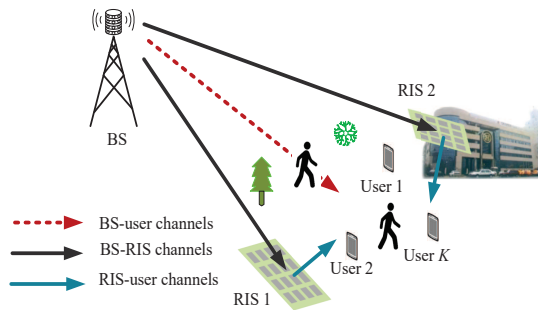


Fig. 1: RISs-aided mmWave communication system.

## II. SYSTEM MODEL

### A. Signal Model

As shown in Fig. 1, we consider the downlink of an mmWave system in which a BS equipped with a uniform linear array (ULA) with  $N$  antennas and  $N_{RF}$  RF chains serves  $K$  single-antenna users (denoted by  $\mathcal{K} \triangleq \{1, \dots, K\}$ ) in the presence of random blockages, where  $K \leq N_{RF} \ll N$ . We assume that  $U$  RISs deployed on e.g., the facade of some buildings, have the capability of passively reflecting the signals transmitted from the BS to the users. Each RIS is made of  $M$  passive reflecting elements that are arranged in a uniform planar array (UPA). It is assumed that the phase shifts of the RIS are computed by the BS and are then sent to the RIS controller through dedicated control channels [12], [13]. The BS adopts a hybrid precoding architecture, in which each RF chain is connected to all the available antennas, and transmits Gaussian data symbols  $\mathbf{s} = [s_1, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$  to the users through a digital precoding matrix  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{C}^{N_{RF} \times K}$  and an analog precoding matrix  $\mathbf{A} \in \mathbb{C}^{N \times N_{RF}}$ . The transmit power of the BS fulfills the constraint  $\|\mathbf{A}\mathbf{D}\|_F^2 \leq P_{max}$ , where  $P_{max}$  is the total transmit power limit. Each entry of  $\mathbf{A}$  is constrained to have a unit modulus, i.e.,  $\mathbf{A} \in \mathcal{S}_A$  where  $\mathcal{S}_A \triangleq \{\mathbf{A} \mid |[\mathbf{A}]_{m,n}|^2 = 1, \forall m, n\}$ , and the symbol  $[\cdot]_{m,n}$  denotes the  $(m, n)$ -th element of a matrix. The received signal intended to the  $k$ -th user can be formulated as

$$\begin{aligned} y_k &= (\mathbf{h}_{b,k}^H + \sum_{u=1}^U \mathbf{h}_{i,uk}^H \mathbf{E}_u \mathbf{H}_{bi,u}) \mathbf{A} \mathbf{D} \mathbf{s} + n_k \\ &= (\mathbf{h}_{b,k}^H + \mathbf{h}_{i,k}^H \mathbf{E} \mathbf{H}_{bi}) \mathbf{A} \mathbf{D} \mathbf{s} + n_k, \end{aligned} \quad (1)$$

where  $\mathbf{E}_u = \text{diag}([e_{(u-1)M+1}, \dots, e_{uM}])$  is the reflection coefficient matrix (also known as the passive beamforming matrix) of the  $u$ -th RIS and  $n_k \sim \mathcal{CN}(0, \sigma_k^2)$  is the additive white Gaussian noise (AWGN). The channels of the BS-user  $k$ , BS-RIS  $u$ , and RIS  $u$ -user  $k$  links are denoted by  $\mathbf{h}_{b,k} \in \mathbb{C}^{N \times 1}$ ,

$\mathbf{H}_{\text{bi},u} \in \mathbb{C}^{M \times N}$  and  $\mathbf{h}_{i,uk} \in \mathbb{C}^{M \times 1}$ , respectively. With this notation, the matrices in (1) are defined as  $\mathbf{H}_{\text{bi}} = [\mathbf{H}_{\text{bi},1}^H, \dots, \mathbf{H}_{\text{bi},U}^H]^H$ ,  $\mathbf{h}_{i,k} = [\mathbf{h}_{i,1k}^H, \dots, \mathbf{h}_{i,Uk}^H]^H$  and  $\mathbf{E} = \text{diag}(\mathbf{E}_1, \dots, \mathbf{E}_U)$ .

Furthermore, denoting by  $\mathbf{H}_k = \begin{bmatrix} \text{diag}(\mathbf{h}_{i,k}^H) \mathbf{H}_{\text{bi}} \\ \mathbf{h}_{b,k}^H \end{bmatrix} \in \mathbb{C}^{(M+1) \times N}$  the equivalent channel of the BS to the  $k$ -th user and by  $\mathbf{e} = [e_1, \dots, e_{UM}, 1]^T \in \mathbb{C}^{(UM+1) \times 1}$  the equivalent reflection coefficient vector that belongs to the set  $\mathcal{S}_e = \{\mathbf{e} \mid |e_m|^2 = 1, 1 \leq m \leq UM, e_{UM+1} = 1\}$ , (1) can be rewritten as  $y_k = \mathbf{e}^H \mathbf{H}_k \mathbf{A} \mathbf{D} \mathbf{s} + n_k, \forall k \in \mathcal{K}$ , and the corresponding achievable signal-to-interference-plus-noise ratio (SINR),  $\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e})$ , can be written as

$$\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) = \frac{|\mathbf{e}^H \mathbf{H}_k \mathbf{A} \mathbf{d}_k|^2}{\sum_{i \neq k}^K |\mathbf{e}^H \mathbf{H}_k \mathbf{A} \mathbf{d}_i|^2 + \sigma_k^2}. \quad (2)$$

### B. Channel Model

Based on [18], a geometric channel model can be used to characterize the mmWave channel. Assume that there are  $L_{BU}$ ,  $L_{BI}$  and  $L_{IU}$  propagation paths that characterize the BS-user links, the BS-RIS links and the RIS-user links respectively, then we have

$$\mathbf{h}_{b,k} = \sqrt{\frac{1}{L_{BU}}} \sum_{l=1}^{L_{BU}} g_{k,l}^b \mathbf{a}_L(\theta_{k,l}^{b,t}), \forall k \in \mathcal{K}, \quad (3)$$

$$\mathbf{h}_{i,k} = \sqrt{\frac{1}{L_{IU}}} \sum_{l=1}^{L_{IU}} g_{k,l}^i \mathbf{a}_P(\theta_{k,l}^{i,t}, \phi_{k,l}^{i,t}), \forall k \in \mathcal{K}, \quad (4)$$

$$\mathbf{H}_{\text{bi}} = \sqrt{\frac{1}{L_{BI}}} \sum_{l=1}^{L_{BI}} g_l^{\text{bi}} \mathbf{a}_P(\theta_l^{i,r}, \phi_l^{i,r}) \mathbf{a}_L(\theta_l^{b,t})^H, \quad (5)$$

where  $\{g_{k,l}^b, g_{k,l}^i, g_l^{\text{bi}}\}$  denote the large-scale fading coefficients. Define  $g \in \{g_{k,l}^b \text{ for } \forall k, g_{k,l}^i \text{ for } \forall k, g_l^{\text{bi}}\}$ , then  $g$  has distribution  $\mathcal{CN}(0, 10^{\frac{\text{PL}}{10}})$ , where  $\text{PL} = -C_0 - 10\alpha \log_{10}(D) - \zeta$  dB,  $C_0$  is the path loss at a reference distance of one meter,  $D$  is the link distance (in meters),  $\alpha$  is the pathloss exponent and  $\zeta$  is the lognormal shadowing [18]. Also,  $\mathbf{a}_L(\theta)$  and  $\mathbf{a}_P(\theta, \phi)$  are the steering vectors of the ULA and UPA, respectively.

According to [3]–[5], the BS-user links may be obstructed by the presence of a random blockage with a certain probability, while the RIS-related links can be assumed not to be affected by blockages, since the locations of the RISs can be appropriately optimized in order to ensure line of sight transmission. Due to the fact that the communication links in the mmWave frequency band are severely attenuated by the presence of blockages, the achievable data rate may be significantly reduced. In order to investigate the impact of the channel uncertainties caused by the presence of random blockages, we adopt a recently

proposed probabilistic model for the BS-user links [5]. In particular, the channels between the BS and the users are modeled as

$$\mathbf{h}_{b,k} = \sqrt{\frac{1}{L_{BU}}} \sum_{l=1}^{L_{BU}} \gamma_{k,l} g_{k,l}^b \mathbf{a}_L(\theta_{k,l}^{b,t}), \forall k \in \mathcal{K}, \quad (6)$$

where the random variable  $\gamma_{k,l} \in \{0, 1\}$  is a blockage parameter that is distributed according to a Bernoulli distribution. In particular, the corresponding blockage probability is denoted by  $p_{k,l}$ . In this work, we assume that the blockage probability is known at the BS and that it is used for robust beamforming design.

### C. Problem Formulation

In the presence of random blockages, we aim to propose a robust mmWave beamforming scheme with the objective of minimizing the sum outage probability [5]. The formulated optimization problem is given by

$$\min_{\mathbf{D}, \mathbf{A}, \mathbf{e}} \sum_{k \in \mathcal{K}} \Pr\{\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) \leq \omega_k\} \quad (7a)$$

$$\text{s.t. } \|\mathbf{A}\mathbf{D}\|_F^2 \leq P_{max} \quad (7b)$$

$$\mathbf{A} \in \mathcal{S}_A \quad (7c)$$

$$\mathbf{e} \in \mathcal{S}_e, \quad (7d)$$

where  $\omega_k > 0$  is the SINR reliability threshold of the  $k$ -th user and  $\Pr\{\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) \leq \omega_k\}$  denotes the probability that the required SINR cannot be satisfied.

In contrast to the traditional deterministic formulation which might not always have feasible solutions due to the QoS constraints, Problem (7) has always a feasible solution that ensures the desired QoS target minimum outage probability.

## III. BEAMFORMING DESIGN

Problem (7) is challenging to solve due to the absence of a closed-form expression for the objective function in (7a), the non-convex unit-modulus constraints in (7c) and (7d), and that fact that the optimization variables are tightly coupled. In the following, we propose a robust beamforming design algorithm under the stochastic-learning-based alternating optimization (AO) framework.

### A. Problem Transformation

To start with, we rewrite  $\Pr\{\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) \leq \omega_k\}$  as  $\mathbb{E}_{\mathbf{H}_k}[\mathbb{I}_{\Omega_k \leq \omega_k}]$ , where  $\mathbb{I}_{\Omega_k \leq \omega_k}$  is the step function. Since the step function is non-differentiable, we approximate it with the following smooth hinge surrogate function [19]

$$u_k(\mathbf{X}) = \begin{cases} 0 & \text{if } 1 - \frac{\Omega_k(\mathbf{X})}{\omega_k} < 0 \\ \frac{1}{2\epsilon} \left(1 - \frac{\Omega_k(\mathbf{X})}{\omega_k}\right)^2 & \text{otherwise} \\ 1 - \frac{\Omega_k(\mathbf{X})}{\omega_k} - \frac{\epsilon}{2} & \text{if } 1 - \frac{\Omega_k(\mathbf{X})}{\omega_k} > \epsilon, \end{cases}$$

where  $0 < \epsilon \ll 1$  and  $\mathbf{X} \triangleq \{\mathbf{D}, \mathbf{A}, \mathbf{e}\}$  is a short-hand notation that collects of all the variables of interest. By replacing the step function  $\mathbb{I}_{\Omega_k \leq \omega_k}$  with its smooth approximation  $u_k(\mathbf{X})$ , we obtain the following approximated reformulation for Problem (7)

$$\min_{\mathbf{X}} \sum_{k \in \mathcal{K}} \mathbb{E}_{\mathbf{H}_k} [u_k(\mathbf{X})] \quad (8a)$$

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

Problem (8) can be viewed as a risk minimization problem [20], which has been studied thoroughly in several application areas such as wireless resource optimization, compressive sensing, machine learning, etc. It can be solved efficiently by using stochastic optimization methods, which are widely adopted due to their easy-to-implement features.

Since the BS is assumed to know the channel state information (CSI), namely the channel gain and the AoDs, the expectation in (8a) is computed with respect to  $\gamma_{k,l}$  in (6). By assuming that the blockage probability can be predicted with good accuracy, we can generate  $\gamma_{k,l}$  randomly and can construct the training data sample set  $\mathcal{H} = \{\mathbf{H}_k^{(t)}, \forall k \in \mathcal{K}\}_{t=1}^T$  that is utilized for stochastic optimization. Accordingly, an appropriate surrogate for the risk function  $\mathbb{E}_{\mathbf{H}_k} [u_k(\mathbf{X})]$  is often assumed to be  $\frac{1}{T} \sum_{t=1}^T u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$ , which is usually referred to as the empirical risk function [20]. Therefore, the resulting empirical risk minimization (ERM) problem is

$$\min_{\mathbf{X}} \frac{1}{T} \sum_{t=1}^T \sum_{k \in \mathcal{K}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}) \quad (9a)$$

$$\text{s.t. } (7b), (7c), (7d). \quad (9b)$$

A popular method to solve ERM problems is the stochastic gradient descent (SGD) [21], which we leverage to solve Problem (9) by alternately optimizing one of the block variables  $\{\mathbf{D}, \mathbf{A}, \mathbf{e}\}$  while keeping the others fixed. This algorithm is usually referred to as the block stochastic gradient descent (BSGD) method [22].

## B. Algorithm Description

To introduce the proposed algorithm, we denote by  $\mathcal{S}_D \triangleq \{\mathbf{D} \mid \|\mathbf{A}\mathbf{D}\|_F^2 \leq P_{max}\}$  the set of  $\mathbf{D}$  with fixed  $\mathbf{A}$  and by  $\mathcal{P}_{\mathcal{S}_z}(\mathbf{z})$  the Euclidean projection from a point  $\mathbf{z}$  onto a set  $\mathcal{S}_z$ , i.e.,  $\mathcal{P}_{\mathcal{S}_z}(\mathbf{z}) = \arg \min_{\mathbf{y} \in \mathcal{S}_z} \|\mathbf{y} - \mathbf{z}\|$ .

---

### Algorithm 1 BSGD-OutMin Algorithm

---

**Initialize:** Initialize  $\mathbf{D}^{(0)}$ ,  $\mathbf{A}^{(0)}$ ,  $\mathbf{e}^{(0)}$ , and the data set  $\mathcal{H}$ . Set  $t = 1$  and  $T_{max} = 10^5$ .

1: **repeat**

2: Sample the data  $\mathbf{H}_k^{(t)} \forall k \in \mathcal{K}$  from  $\mathcal{H}$ .

3:  $\mathbf{D}^{(t)} = \mathcal{P}_{\mathcal{S}_D}(\mathbf{D}^{(t-1)} - \alpha_t \sum_{k \in \mathcal{K}} \nabla_{\mathbf{D}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}))$ .

4:  $\mathbf{A}^{(t)} = \mathcal{P}_{\mathcal{S}_A}(\mathbf{A}^{(t-1)} - \alpha_t \sum_{k \in \mathcal{K}} \nabla_{\mathbf{A}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}))$ .

5:  $\mathbf{e}^{(t)} = \mathcal{P}_{\mathcal{S}_e}(\mathbf{e}^{(t-1)} - \alpha_t \sum_{k \in \mathcal{K}} \nabla_{\mathbf{e}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}))$ .

6:  $t = t + 1$ .

7: **until** The objective value in (9a) converges.

---

Algorithm 1 summarizes the proposed BSGD-based outage minimum robust hybrid beamforming design for RIS-aided mmWave systems in which the BS-user links undergo random blockages. The proposed algorithm is referred to as BSGD-OutMin. It is not difficult to carry out the projection operations in Algorithm 1 as follows

$$\mathcal{P}_{\mathcal{S}_D}(\mathbf{Z}) = \frac{(\mathbf{Z})}{\|\mathbf{A}^{(t-1)}\mathbf{Z}\|_F} \sqrt{P_{max}},$$

$$\mathcal{P}_{\mathcal{S}_A}(\mathbf{Z}) = \exp \{j\angle \mathbf{Z}\},$$

$$\mathcal{P}_{\mathcal{S}_e}(\mathbf{z}) = \exp \{j\angle (\mathbf{z}/[\mathbf{z}]_{UM+1})\}.$$

a) *Gradient of  $u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$ :* Define  $\nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  the stochastic gradients in Algorithm 1. It is worth noting that the normalization factor  $1/N$  in  $\sum_{k \in \mathcal{K}} \nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  is omitted, since the stochastic gradient  $\nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  is an unbiased estimator for the batch gradient  $\frac{1}{T} \sum_{t=1}^T \nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  [20], i.e.,  $\mathbb{E}_{\mathbf{H}_k^{(t)}}[\nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}) | \mathbf{x}^{(t-1)}] = \frac{1}{T} \sum_{t=1}^T \nabla_{\mathbf{X}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$ .

The partial gradient of  $u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  with respect to  $\mathbf{x}$  is

$$\nabla_{\mathbf{x}} u_k(\mathbf{X}, \mathbf{H}_k^{(t)}) = \begin{cases} 0 & \text{if } 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} < 0 \\ \frac{\frac{\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} - 1}{\epsilon} \frac{\nabla_{\mathbf{x}} \Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} & \text{otherwise} \\ -\frac{\nabla_{\mathbf{x}} \Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} & \text{if } 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} > \epsilon. \end{cases} \quad (10)$$



In order to compute the partial gradient  $\nabla\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  with respect to any  $\mathbf{X} \in \{\mathbf{D}, \mathbf{A}, \mathbf{e}\}$ , it is necessary to apply some mathematical transformations to the SINR in (2). In particular, as far as  $\mathbf{D}$  is concerned, (2) can be rewritten as

$$\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)}) = \frac{\text{vec}(\mathbf{D})^H \mathbf{Q}_{\mathbf{D},k}^{(t)} \text{vec}(\mathbf{D})}{\text{vec}(\mathbf{D})^H \overline{\mathbf{Q}}_{\mathbf{D},k}^{(t)} \text{vec}(\mathbf{D}) + \sigma_k^2}, \quad (11)$$

where  $\mathbf{Q}_{\mathbf{D},k}^{(t)} = \text{diag}(\mathbf{i}_k) \otimes \mathbf{A}^H \mathbf{H}_k^{(t),H} \mathbf{e} \mathbf{e}^H \mathbf{H}_k^{(t)} \mathbf{A}$ ,  $\overline{\mathbf{Q}}_{\mathbf{D},k}^{(t)} = \text{diag}(\overline{\mathbf{i}}_k) \otimes \mathbf{A}^H \mathbf{H}_k^{(t),H} \mathbf{e} \mathbf{e}^H \mathbf{H}_k^{(t)} \mathbf{A}$ , and  $\mathbf{i}_k$  is the  $k$ -th column of the  $K \times K$  identity matrix  $\mathbf{I}_K$ ,  $\overline{\mathbf{i}}_k$  denotes the one's complement of  $\mathbf{i}_k$ , and  $\otimes$  is the Kronecker product.

As far as  $\mathbf{A}$  is concerned, (2) is equivalent to

$$\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)}) = \frac{\text{vec}(\mathbf{A})^H \mathbf{Q}_{\mathbf{A},k}^{(t)} \text{vec}(\mathbf{A})}{\text{vec}(\mathbf{A})^H \overline{\mathbf{Q}}_{\mathbf{A},k}^{(t)} \text{vec}(\mathbf{A}) + \sigma_k^2}, \quad (12)$$

where  $\mathbf{Q}_{\mathbf{A},k}^{(t)} = \mathbf{d}_k^* \mathbf{d}_k^T \otimes \mathbf{H}_k^{(t),H} \mathbf{e} \mathbf{e}^H \mathbf{H}_k^{(t)}$  and  $\overline{\mathbf{Q}}_{\mathbf{A},k}^{(t)} = \mathbf{D}_{-k}^* \mathbf{D}_{-k}^T \otimes \mathbf{H}_k^{(t),H} \mathbf{e} \mathbf{e}^H \mathbf{H}_k^{(t)}$ .

As far as  $\mathbf{e}$  is concerned, (2) is equivalent to

$$\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)}) = \frac{\mathbf{e}^H \mathbf{Q}_{\mathbf{e},k}^{(t)} \mathbf{e}}{\mathbf{e}^H \overline{\mathbf{Q}}_{\mathbf{e},k}^{(t)} \mathbf{e} + \sigma_k^2}, \quad (13)$$

where  $\mathbf{Q}_{\mathbf{e},k}^{(t)} = \mathbf{H}_k^{(t)} \mathbf{A} \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \mathbf{H}_k^{(t),H}$  and  $\overline{\mathbf{Q}}_{\mathbf{e},k}^{(t)} = \mathbf{H}_k^{(t)} \mathbf{A} \sum_{i \neq k}^K \mathbf{d}_i \mathbf{d}_i^H \mathbf{A}^H \mathbf{H}_k^{(t),H}$ .

Equations (11)-(13) have the same mathematical structure. Define  $\mathbf{x} \in \{\text{vec}(\mathbf{D}), \text{vec}(\mathbf{A}), \mathbf{e}\}$ , then the unified partial gradient for any  $\mathbf{x}$  is given by

$$\nabla_{\mathbf{x}} \Omega_k(\mathbf{x}, \mathbf{H}_k^{(t)}) = \frac{\mathbf{Q}_{\mathbf{x},k}^{(t)} \mathbf{x}}{v_k} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k}^{(t)} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k}^{(t)} \mathbf{x},$$

where  $v_k = \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}^{(t)} \mathbf{x} + \sigma_k^2$ .

*b) Initial point:* Problem (7) has multiple local minima points due to the non-convex constraints  $\mathbf{A} \in \mathcal{S}_A$  and  $\mathbf{e} \in \mathcal{S}_e$ . The accurate selection of the initial points in Algorithm 1 plays an important role for the convergence speed and the optimality of the obtained local solution. To that end, we first initialize  $\mathbf{e}$  to maximize the equivalent total channel gain, which results in the following optimization problem

$$\max_{\mathbf{e} \in \mathcal{S}_e} \sum_{k \in \mathcal{K}} \|\mathbf{e}^H \mathbf{H}_k^{(0)}\|_2^2, \quad (14a)$$

where  $\mathbf{H}_k^{(0)} \forall k$  are the deterministic matrices with blockage probability  $p_{k,l} = 0, \forall k, l$ , i.e.,  $\gamma_{k,l} = 1, \forall k, l$ .

The objective function in (14) is convex and can be linearized by using its first-order Taylor approximation, i.e.,  $2\text{Re}\{\mathbf{e}^{[n],H} \mathbf{H}^{(0)} \mathbf{H}^{(0),H} \mathbf{e}\} - \mathbf{e}^{[n],H} \mathbf{H}^{(0)} \mathbf{H}^{(0),H} \mathbf{e}^{[n]}$ , where  $\mathbf{H}^{(0)} = [\mathbf{H}_1^{(0)}, \dots, \mathbf{H}_K^{(0)}]$  and

$\mathbf{e}^{[n]}$  is the optimal solution obtained at the  $n$ -th iteration, and  $\text{Re}\{\cdot\}$  denotes the real part of a complex number. Therefore, the optimal solution to Problem (14) at the  $(n+1)$ -th iteration is

$$\mathbf{e}^{[n+1]} = \exp \left\{ j\angle \left( \mathbf{H}^{(0)} \mathbf{H}^{(0),H} \mathbf{e}^{[n]} / [\mathbf{H}^{(0)} \mathbf{H}^{(0),H} \mathbf{e}^{[n]}]_{M+1} \right) \right\}.$$

Furthermore,  $\mathbf{A}$  is initialized to align the phases of the equivalent channel to the mmWave BS, i.e.,

$$\mathbf{A}^{(0)} = \exp \left\{ j\angle \left( \mathbf{H}^{(0)} (\mathbf{I}_K \otimes \mathbf{e}^{(0)}) \right) \right\}.$$

*c) Convergence analysis:* It is assumed that the step size  $\{\alpha_t \in (0,1]\}$  is a decreasing sequence satisfying  $\alpha_t \rightarrow 0$ ,  $\sum_{t=1}^{\infty} \alpha_t \rightarrow \infty$  and  $\sum_{t=1}^{\infty} \alpha_t^2 \rightarrow \infty$ . Then, we have the following Theorem.

**Theorem 1** [22] *Algorithm 1 can be guaranteed to converge to a stationary point when the step size fulfills the condition  $\alpha_t \in (0, 1/L]$ , where  $L$  is the Lipschitz constant.*

**Proof:** Please refer to [22]. ■

**Lemma 1** *The Lipschitz constant of  $u_k(\mathbf{X}, \mathbf{H}_k^{(t)})$  is given by*

$$L = \begin{cases} 0 & \text{if } 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} < 0 \\ \max(l_{e1}, l_{A1}, l_{D1}) & \text{otherwise} \\ \max(l_{e2}, l_{A2}, l_{D2}) & \text{if } 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k^{(t)})}{\omega_k} > \epsilon, \end{cases} \quad (15)$$

where

$$\begin{aligned} a &= (UM + 1) P_{max}^2 \lambda_{\max} \left( \mathbf{H}_k^{(0),H} \mathbf{H}_k^{(0)} \right)^2, \\ b &= (UM + 1) P_{max} \lambda_{\max} \left( \mathbf{H}_k^{(0),H} \mathbf{H}_k^{(0)} \right), \\ l_{e1} &= \frac{1}{\omega_k^2 \epsilon \sigma_k^4} \left( (2 + 5\omega_k) a - \frac{4ab}{\sigma_k^2} + \frac{6ab^2}{\sigma_k^4} \right), \\ l_{A1} &= \frac{b^2}{\omega_k^2 \epsilon \sigma_k^4 NN_{RF}} \left( 2 + \frac{(5\omega_k - 4b)}{\sigma_k^2} + \frac{6b^2}{\sigma_k^4} \right), \\ l_{D2} &= \frac{NN_{RF} b^2}{\omega_k^2 \epsilon \sigma_k^4 P_{max}} \left( 2 + \omega_k + \frac{6b^2}{\sigma_k^4} \right), \\ l_{e2} &= \frac{5a}{\omega_k \sigma_k^4}, \\ l_{A2} &= \frac{5b^2}{\omega_k \sigma_k^4 NN_{RF}}, \\ l_{D2} &= \frac{b^2 NN_{RF}}{\omega_k \sigma_k^4 P_{max}}. \end{aligned}$$

**Proof:** See Appendix A. ■

#### IV. NUMERICAL RESULTS AND DISCUSSION

In this section, numerical results are illustrated to evaluate the performance of the proposed algorithm. All results are obtained by averaging over 500 channel realizations. Unless stated otherwise, we assume  $L_{BU} = L_{IU} = L_{BI} = 5$ , the BS, RIS 1 and RIS 2 are located at (0 m, 0 m), (40 m, 10 m) and (40 m, -10 m), respectively, and the users are assumed to be randomly distributed in a circle centered at (50 m, 0 m) with radius 5 m. The carrier frequency of the mmWave system is 28 GHz, and the corresponding large-scale fading parameters are set according to Table I in [18]. Other simulation parameters are  $P_{max} = 5$  W,  $\sigma_1^2 = \dots, \sigma_K^2 = -100$  dBm. For simplicity, it is assumed that the blockage probabilities are equal  $p_{k,l} = p_{\text{block}}, \forall k, l$ , and that the target SINR of all the users is  $\omega_1 = \dots = \omega_K = \omega$ , leading to the minimum target rate  $R_{\text{targ}} = \log_2(1 + \omega)$ . To evaluate the performance of the proposed BSGD algorithm, we consider three benchmark schemes: 1) RIS-random:  $\mathbf{e}$  is designed randomly; 2) RIS-non-robust: the beamforming is designed by setting the blockage probability to zero; 3) Non-RIS: RISs are not deployed in the network.

In order to demonstrate the robustness of the proposed algorithm, we consider two performance metrics: the outage probability and the effective sum rate. In particular, the outage probability is the average achievable outage probability for each user, i.e.,  $\frac{1}{K} \sum_{k \in \mathcal{K}} \Pr\{\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) \leq \omega_k\}$ . Then, the corresponding effective rate of the  $k$ -th user is defined as  $R_{\text{eff},k} \triangleq \mathbb{E}[\log_2(1 + \Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}))]$  if  $\Omega_k(\mathbf{D}, \mathbf{A}, \mathbf{e}) \geq \omega_k$  and  $R_{\text{eff},k} \triangleq 0$  otherwise.

Figure 2 illustrates the impact of the blockage probability  $p_{\text{block}}$  on the system performance. It is observed that an mmWave system in the absence of RISs is highly sensitive to the presence of blockages, which results in the worst outage probability and effective sum rate. On the other hand, the proposed BSGD-based scheme provides a high data rate and a low outage probability over the whole range of blockage probabilities, which substantiates the robustness of the proposed algorithm to the presence of random blockages.

Figure 3 illustrates the convergence behavior of the stochastic optimization method. It is observed that the BSGD algorithm converges approximately monotonically in the low outage probability regime (e.g.,  $p_{\text{block}} = 0.1$ ), but it shows an oscillatory behavior in the high outage probability regime (e.g.,  $p_{\text{block}} = 0.9$ ). This is due to the nature of stochastic programming.

#### V. CONCLUSIONS

This work introduced a stochastic-learning-based robust beamforming design for RISs-aided mmWave systems which is aimed to combat the channel uncertainties caused by the presence of random blockages. The formulated stochastic optimization problem was solved by using the stochastic block

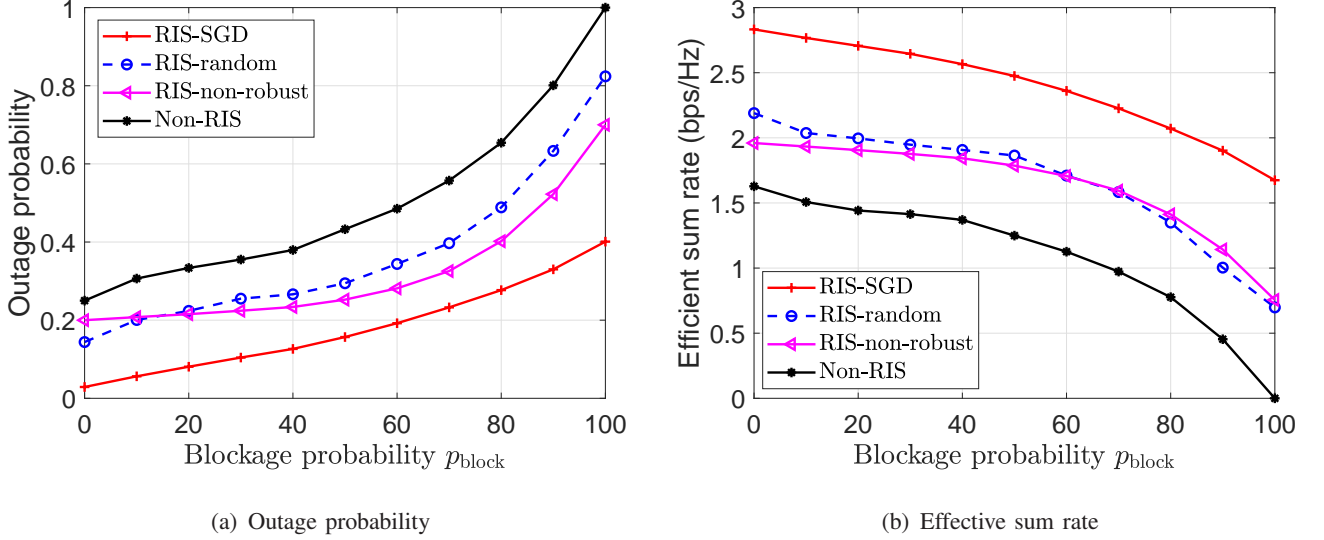


Fig. 2: Outage probability and effective rate as a function of the blockage probability  $p_{\text{block}}$  for  $N = 32$ ,  $M = 64$ ,  $K = N_{RF} = U = 2$ , and  $R_{\text{targ}} = 1$  bps/Hz.

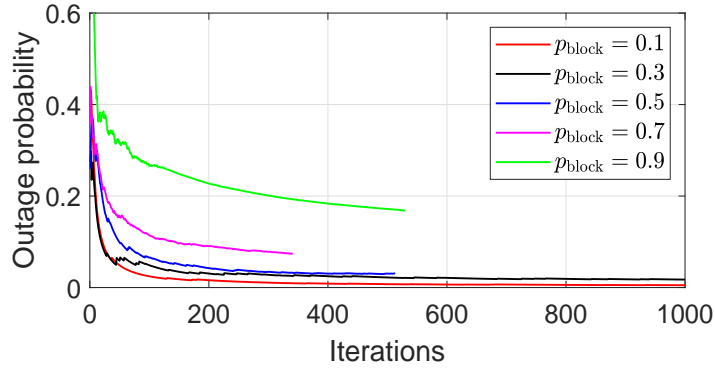


Fig. 3: Convergence behavior for  $N = 32$ ,  $M = 64$ ,  $K = N_{RF} = U = 2$ , and  $R_{\text{targ}} = 1$  bps/Hz.

gradient descent algorithm. Simulation results demonstrated the robustness of the proposed hybrid beamforming design in the presence of random blockages and confirmed its superior performance in terms of outage probability and effective data rate compared with baseline schemes.

## APPENDIX A

### THE PROOF OF LEMMA 1

A valid choice for Lipschitz constant  $L$  is  $L \geq \max_{\mathbf{x} \in \mathcal{S}_x} \|\nabla_{\mathbf{x}}^2 u_k(\mathbf{X}, \mathbf{H}_k)\|_2$ . However, computing  $\max_{\mathbf{x} \in \mathcal{S}_x} \|\nabla_{\mathbf{x}}^2 u_k(\mathbf{X}, \mathbf{H}_k)\|_2$  for unstructured  $u_k(\mathbf{X}, \mathbf{H}_k)$  is not in general easy. In all such cases, a natural

option is leveraging some upper bound of  $\max_{\mathbf{x} \in \mathcal{S}_x} \|\nabla_{\mathbf{x}}^2 u_k(\mathbf{X}, \mathbf{H}_k)\|_2$ , that is  $\lambda_{\max}(\nabla_{\mathbf{x}}^2 u_k(\mathbf{X}, \mathbf{H}_k))$ . In the following, we explore to derive the Hessian matrix of  $u_k(\mathbf{X}, \mathbf{H}_k)$ , and then uniformly bound the largest eigenvalue of the Hessian matrix.

First, the partial derivative and conjugate partial derivative of  $\Omega_k(\mathbf{X}, \mathbf{H}_k)$  are given by

$$\begin{aligned}\frac{\partial \Omega_k}{\partial \mathbf{x}} &= \frac{\mathbf{Q}_{\mathbf{x},k}^T \mathbf{x}^*}{v_k} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^*, \\ \frac{\partial \Omega_k}{\partial \mathbf{x}^*} &= \frac{\mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}.\end{aligned}$$

Second, the partial Hessian matrices of  $\Omega_k(\mathbf{X}, \mathbf{H}_k)$  are given by

$$\begin{aligned}\frac{\partial^2 \Omega_k}{\partial \mathbf{x}^* \partial \mathbf{x}^T} &= \frac{\mathbf{Q}_{\mathbf{x},k}}{v_k} - \frac{\overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k}}{v_k^2} - \frac{\mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}}{v_k^2} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k} + \frac{2\overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}}{v_k^3}, \\ \frac{\partial^2 \Omega_k}{\partial \mathbf{x} \partial \mathbf{x}^H} &= \frac{\mathbf{Q}_{\mathbf{x},k}^T}{v_k} - \frac{\overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T}{v_k^2} - \frac{\mathbf{Q}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T}{v_k^2} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k}^T \\ &\quad + \frac{2\overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T}{v_k^3}, \\ \frac{\partial^2 \Omega_k}{\partial \mathbf{x}^* \partial \mathbf{x}^H} &= -\frac{\overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T}{v_k^2} - \frac{\mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T}{v_k^2} + \frac{2\overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T}{v_k^3}, \\ \frac{\partial^2 \Omega_k}{\partial \mathbf{x} \partial \mathbf{x}^T} &= -\frac{\overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k}}{v_k^2} - \frac{\mathbf{Q}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}}{v_k^2} + \frac{2\overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}}{v_k^3}.\end{aligned}$$

Then, the complex Hessian matrix of  $\Omega_k(\mathbf{X}, \mathbf{H}_k)$  is defined by

$$\begin{aligned}\mathcal{H}_{\Omega_k} &= \begin{bmatrix} \frac{\partial^2 \Omega_k}{\partial \mathbf{x}^* \partial \mathbf{x}^T} & \frac{\partial^2 \Omega_k}{\partial \mathbf{x}^* \partial \mathbf{x}^H} \\ \frac{\partial^2 \Omega_k}{\partial \mathbf{x} \partial \mathbf{x}^T} & \frac{\partial^2 \Omega_k}{\partial \mathbf{x} \partial \mathbf{x}^H} \end{bmatrix} \\ &= \frac{1}{v_k} \begin{bmatrix} \mathbf{Q}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} - \frac{2}{v_k^2} \text{Re} \left( \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} & \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) \\ &\quad - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} + \frac{2\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix}.\end{aligned}$$

• If  $1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k)}{\omega_k} > \epsilon$ , then  $u_k(\mathbf{X}, \mathbf{H}_k) = 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k)}{\omega_k} - \frac{\epsilon}{2}$ , the complex Hessian matrix of which is  $\mathcal{H}_{u_k} = -\frac{\mathcal{H}_{\Omega_k}}{\omega_k}$ . The bound of  $\lambda_{\max}(\mathcal{H}_{u_k})$  is given by

$$\begin{aligned}\lambda_{\max}(\mathcal{H}_{u_k}) &\leq \lambda_{\max} \left( -\frac{1}{\omega_k v_k} \begin{bmatrix} \mathbf{Q}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) + \lambda_{\max} \left( \frac{1}{\omega_k v_k^2} \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} & \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) \\ &\quad + \lambda_{\max} \left( \frac{1}{\omega_k v_k^2} \begin{bmatrix} \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \\ \mathbf{Q}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} \right)\end{aligned}$$

$$\begin{aligned}
& + \lambda_{\max} \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \begin{bmatrix} \bar{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} \right) \\
= & \lambda_{\max} \left( -\frac{1}{\omega_k v_k} \begin{bmatrix} \mathbf{Q}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) + \lambda_{\max} \left( \frac{2}{\omega_k v_k^2} \text{Re} \left( \begin{bmatrix} \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \bar{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} & \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) \right)
\end{aligned} \tag{16}$$

$$\begin{aligned}
& + \lambda_{\max} \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \begin{bmatrix} \bar{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} \right) \\
= & -\frac{1}{\omega_k v_k} \lambda_{\min}(\mathbf{Q}_{\mathbf{x},k}) + \frac{1}{\omega_k v_k^2} 2 \max \left( \text{Re} \left( \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} + \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T \bar{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \right), 0 \right) \\
& + \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \lambda_{\max}(\bar{\mathbf{Q}}_{\mathbf{x},k}) \\
= & \frac{4}{\omega_k v_k^2} \max \left( \text{Re} \left( \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \right), 0 \right) - \frac{1}{\omega_k v_k} \lambda_{\min}(\mathbf{Q}_{\mathbf{x},k}) + \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \lambda_{\max}(\bar{\mathbf{Q}}_{\mathbf{x},k}) \\
\stackrel{(A1)}{=} & \frac{4}{\omega_k v_k^2} \max \left( \text{Re} \left( \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \right), 0 \right) + \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \lambda_{\max}(\bar{\mathbf{Q}}_{\mathbf{x},k})
\end{aligned} \tag{17}$$

$$\stackrel{(A2)}{\leq} \begin{cases} \frac{5a}{\omega_k \sigma_k^4} & \text{if } \mathbf{x} = \mathbf{e} \\ \frac{5b^2}{\omega_k \sigma_k^4 N N_{RF}} & \text{if } \mathbf{x} = \text{vec}(\mathbf{A}) \\ \frac{b^2 N N_{RF}}{\omega_k \sigma_k^4 P_{max}} & \text{if } \mathbf{x} = \text{vec}(\mathbf{D}). \end{cases} \tag{18}$$

where  $a = (UM + 1)P_{max}^2 h_k^2$ ,  $b = (UM + 1)P_{max} h_k$ , and  $h_k = \lambda_{\max}(\mathbf{H}_{\text{bi}}^H \text{diag}(\mathbf{h}_{i,k}) \text{diag}(\mathbf{h}_{i,k}^H) \mathbf{H}_{\text{bi}}) + \frac{1}{L_{BU}} \lambda_{\max}(\mathbf{G}_k^H \mathbf{G}_k) \|\mathbf{g}_k\|^2$ .

Equation (A1) in (18) is due to the fact that  $\mathbf{Q}_{\mathbf{x},k}$  is a rank-1 positive semidefinite matrix. The inequality (A2) in (18) comes from  $v_k = \mathbf{x}^H \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} + \sigma_k^2 \geq \sigma_k^2$  and the following relaxations.

**1) For  $\text{Re}(\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \bar{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x})$ :**

$$\begin{aligned}
& \text{Re}(\mathbf{e}^H \mathbf{Q}_{\mathbf{e},k} \bar{\mathbf{Q}}_{\mathbf{e},k} \mathbf{e}) \\
= & \text{Re} \left( \text{Tr} \left( \mathbf{H}_k \mathbf{A} \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{A} \sum_{i \neq k}^K \mathbf{d}_i \mathbf{d}_i^H \mathbf{A}^H \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \right) \right) \\
\leq & (UM + 1) \text{Re} \left( \text{Tr} \left( \mathbf{A} \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{A} \sum_{i \neq k}^K \mathbf{d}_i \mathbf{d}_i^H \mathbf{A}^H \mathbf{H}_k^H \mathbf{H}_k \right) \right) \\
\leq & (UM + 1) \lambda_{\max}(\mathbf{H}_k^H \mathbf{H}_k)^2 \text{Tr} \left( \mathbf{A} \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \mathbf{A} \sum_{i \neq k}^K \mathbf{d}_i \mathbf{d}_i^H \mathbf{A}^H \right) \\
\leq & (UM + 1) P_{max}^2 \lambda_{\max}(\mathbf{H}_k^H \mathbf{H}_k)^2
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(A3)}{\leq} (UM + 1)P_{max}^2 h_k^2 \\
& = a, \\
& \text{Re} \left( \text{vec}(\mathbf{A})^H \mathbf{Q}_{\mathbf{A},k} \overline{\mathbf{Q}}_{\mathbf{A},k} \text{vec}(\mathbf{A}) \right) \\
& = \text{Re} \left( \text{vec}(\mathbf{A})^H \left( \mathbf{d}_k^* \mathbf{d}_k^T \otimes \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \right) \left( \mathbf{D}_{-k}^* \mathbf{D}_{-k}^T \otimes \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \right) \text{vec}(\mathbf{A}) \right) \\
& = \text{Re} \left( \text{vec}(\mathbf{A})^H \left( \mathbf{d}_k^* \mathbf{d}_k^T \mathbf{D}_{-k}^* \mathbf{D}_{-k}^T \otimes \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \right) \text{vec}(\mathbf{A}) \right) \\
& = \text{Re} \left( \text{Tr} \left( \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{A} \mathbf{D}_{-k} \mathbf{D}_{-k}^H \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \right) \right) \\
& \leq (UM + 1)^2 \text{Re} \left( \text{Tr} \left( \mathbf{H}_k^H \mathbf{H}_k \mathbf{H}_k^H \mathbf{H}_k \mathbf{A} \mathbf{D}_{-k} \mathbf{D}_{-k}^H \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \right) \right) \\
& \leq (UM + 1)^2 \lambda_{\max} \left( \mathbf{H}_k^H \mathbf{H}_k \right)^2 \text{Re} \left( \text{Tr} \left( \mathbf{A} \mathbf{D}_{-k} \mathbf{D}_{-k}^H \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H \right) \right) \\
& \leq \frac{(UM + 1)^2 P_{max}^2}{NN_{RF}} \lambda_{\max} \left( \mathbf{H}_k^H \mathbf{H}_k \right)^2 \\
& \stackrel{(A4)}{\leq} \frac{(UM + 1)^2 P_{max}^2}{NN_{RF}} h_k^2 \\
& = \frac{b^2}{NN_{RF}},
\end{aligned}$$

and

$$\begin{aligned}
& \text{Re} \left( \text{vec}(\mathbf{D})^H \mathbf{Q}_{\mathbf{D},k} \overline{\mathbf{Q}}_{\mathbf{D},k} \text{vec}(\mathbf{D}) \right) \\
& = \text{Re} \left( \text{vec}(\mathbf{D})^H \left( \text{diag}(\mathbf{i}_k) \otimes \mathbf{A}^H \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{A} \right) \left( \text{diag}(\tilde{\mathbf{i}}_k) \otimes \mathbf{A}^H \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{A} \right) \text{vec}(\mathbf{D}) \right) \\
& = \text{Re} \left( \text{vec}(\mathbf{D})^H \left( \text{diag}(\mathbf{i}_k) \text{diag}(\tilde{\mathbf{i}}_k) \otimes \mathbf{A}^H \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{A} \mathbf{A}^H \mathbf{H}_k^H \mathbf{e} \mathbf{e}^H \mathbf{H}_k \mathbf{A} \right) \text{vec}(\mathbf{D}) \right) \\
& = 0,
\end{aligned}$$

where inequality (A3) and (A4) is because the maximum gain of the random channel  $\mathbf{H}_k$  is obtained at the blockage probability of  $p_{k,l} = 0, \forall k, l$ . Specifically, (6) can be rewritten as

$$\mathbf{h}_{b,k} = \sqrt{\frac{1}{L_{BU}}} \mathbf{G}_k \boldsymbol{\gamma}_k, \forall k \in \mathcal{K}, \quad (19)$$

where  $\mathbf{G}_k = [\mathbf{a}_L(\theta_{k,1}^{b,t}), \dots, \mathbf{a}_L(\theta_{k,L_{BU}}^{b,t})]$  and  $\boldsymbol{\gamma}_k = [\gamma_{k,1} g_{k,1}^b, \dots, \gamma_{k,L_{BU}} g_{k,L_{BU}}^b]^T$ . Then, based on channel model (19), we have the following inequality

$$\begin{aligned}
\lambda_{\max} \left( \mathbf{H}_k^H \mathbf{H}_k \right) & = \underbrace{\lambda_{\max} \left( \mathbf{H}_{bi}^H \text{diag}(\mathbf{h}_{i,k}) \text{diag}(\mathbf{h}_{i,k}^H) \mathbf{H}_{bi} \right)}_{\text{deterministic term}} + \underbrace{\lambda_{\max} \left( \mathbf{h}_{b,k} \mathbf{h}_{b,k}^H \right)}_{\text{stochastic term}} \\
& \leq \underbrace{\lambda_{\max} \left( \mathbf{H}_{bi}^H \text{diag}(\mathbf{h}_{i,k}) \text{diag}(\mathbf{h}_{i,k}^H) \mathbf{H}_{bi} \right)}_{\text{deterministic term}} + \underbrace{\frac{1}{L_{BU}} \lambda_{\max} \left( \mathbf{G}_k^H \mathbf{G}_k \right) \|\boldsymbol{\gamma}_k\|^2}_{\text{stochastic term}}
\end{aligned}$$

$$\begin{aligned} &\leq \underbrace{\lambda_{\max}(\mathbf{H}_{\text{bi}}^H \text{diag}(\mathbf{h}_{i,k}) \text{diag}(\mathbf{h}_{i,k}^H) \mathbf{H}_{\text{bi}})}_{\text{deterministic term}} + \underbrace{\frac{1}{L_{BU}} \lambda_{\max}(\mathbf{G}_k^H \mathbf{G}_k) \|\mathbf{g}_k\|^2}_{\text{deterministic term}} \\ &= h_k \end{aligned}$$

where  $\mathbf{g}_k = [g_{k,1}^b, \dots, g_{k,L_{BU}}^b]^T$ .

**2) For  $\lambda_{\max}(\overline{\mathbf{Q}}_{\mathbf{x},k})$ :**

$$\begin{aligned} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} &\leq \lambda_{\max}(\mathbf{H}_k^{(t),H} \mathbf{H}_k^{(t)}) \lambda_{\max}(\mathbf{e}\mathbf{e}^H) \text{Tr}(\mathbf{A} \mathbf{d}_k \mathbf{d}_k^H \mathbf{A}^H) \leq b, \\ \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k}^{(t)} \mathbf{x} &\leq \lambda_{\max}(\mathbf{H}_k^{(t),H} \mathbf{H}_k^{(t)}) \lambda_{\max}(\mathbf{e}\mathbf{e}^H) \text{Tr}(\mathbf{A} \sum_{i \neq k}^K \mathbf{d}_i \mathbf{d}_i^H \mathbf{A}^H) \leq b, \\ \lambda_{\max}(\overline{\mathbf{Q}}_{\mathbf{e},k}) &\leq \frac{b}{\text{Tr}(\mathbf{e}\mathbf{e}^H)} \leq \frac{b}{UM+1}, \\ \lambda_{\max}(\overline{\mathbf{Q}}_{\mathbf{A},k}^{(t)}) &\leq \frac{b}{\text{Tr}(\mathbf{A}\mathbf{A}^H)} \leq \frac{b}{NN_{RF}}, \\ \lambda_{\max}(\overline{\mathbf{Q}}_{\mathbf{D},k}^{(t)}) &\leq \frac{b}{\text{Tr}(\mathbf{D}\mathbf{D}^H)} \leq \frac{b \text{Tr}(\mathbf{A}\mathbf{A}^H)}{P_{max}} \leq \frac{bNN_{RF}}{P_{max}}. \end{aligned}$$

• If  $0 \leq 1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k)}{\omega_k} \leq \epsilon$ , then  $u_k(\mathbf{X}, \mathbf{H}_k) = \frac{1}{2\epsilon} \left(1 - \frac{\Omega_k(\mathbf{X}, \mathbf{H}_k)}{\omega_k}\right)^2$ . The partial Hessian matrices of  $u_k(\mathbf{X}, \mathbf{H}_k)$  are given by

$$\begin{aligned} \frac{\partial^2 u_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^T}} &= \frac{1}{\omega_k^2 \epsilon} \left( \frac{\partial \Omega_k}{\partial_{\mathbf{x}^*}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^T}} + (\Omega_k - \omega_k) \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^T}} \right), \\ \frac{\partial^2 u_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^H}} &= \frac{1}{\omega_k^2 \epsilon} \left( \frac{\partial \Omega_k}{\partial_{\mathbf{x}}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^H}} + (\Omega_k - \omega_k) \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^H}} \right), \\ \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^H}} &= \frac{1}{\omega_k^2 \epsilon} \left( \frac{\partial \Omega_k}{\partial_{\mathbf{x}^*}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^H}} + (\Omega_k - \omega_k) \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^H}} \right), \\ \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^T}} &= \frac{1}{\omega_k^2 \epsilon} \left( \frac{\partial \Omega_k}{\partial_{\mathbf{x}}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^T}} + (\Omega_k - \omega_k) \frac{\partial^2 \Omega_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^T}} \right). \end{aligned}$$

The bound of  $\lambda_{\max}(\mathcal{H}_{u_k})$  is given by

$$\begin{aligned} \lambda_{\max}(\mathcal{H}_{u_k}) &= \lambda_{\max} \left( \begin{bmatrix} \frac{\partial^2 u_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^T}} & \frac{\partial^2 u_k}{\partial_{\mathbf{x}^*} \partial_{\mathbf{x}^H}} \\ \frac{\partial^2 u_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^T}} & \frac{\partial^2 u_k}{\partial_{\mathbf{x}} \partial_{\mathbf{x}^H}} \end{bmatrix} \right) \\ &\leq \frac{1}{\omega_k^2 \epsilon} \lambda_{\max} \left( \begin{bmatrix} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^*}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^T}} & \frac{\partial \Omega_k}{\partial_{\mathbf{x}^*}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^H}} \\ \frac{\partial \Omega_k}{\partial_{\mathbf{x}}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^T}} & \frac{\partial \Omega_k}{\partial_{\mathbf{x}}} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^H}} \end{bmatrix} \right) + \frac{1}{\omega_k^2 \epsilon} \lambda_{\max}((\Omega_k - \omega_k) \mathcal{H}_{\Omega_k}) \\ &\leq \frac{1}{\omega_k^2 \epsilon} \lambda_{\max} \left( \begin{bmatrix} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^*}} \\ \frac{\partial \Omega_k}{\partial_{\mathbf{x}}} \end{bmatrix} \begin{bmatrix} \frac{\partial \Omega_k}{\partial_{\mathbf{x}^T}} & \frac{\partial \Omega_k}{\partial_{\mathbf{x}^H}} \end{bmatrix} \right) + \frac{1}{\omega_k^2 \epsilon} \Omega_k \lambda_{\max}(\mathcal{H}_{\Omega_k}) + \frac{1}{\epsilon} \lambda_{\max}(\mathcal{H}_{u_k}), \quad (20) \end{aligned}$$



where

$$\begin{aligned} & \lambda_{\max} \left( \begin{bmatrix} \frac{\partial \Omega_k}{\partial \mathbf{x}^*} \\ \frac{\partial \Omega_k}{\partial \mathbf{x}} \end{bmatrix} \begin{bmatrix} \frac{\partial \Omega_k}{\partial \mathbf{x}^T} & \frac{\partial \Omega_k}{\partial \mathbf{x}^H} \end{bmatrix} \right) \\ &= 2\text{Re} \left( \frac{\partial \Omega_k}{\partial \mathbf{x}^T} \frac{\partial \Omega_k}{\partial \mathbf{x}^*} \right) \end{aligned} \quad (21)$$

$$\begin{aligned} &= 2\text{Re} \left( \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k}}{v_k} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} \right) \left( \frac{\mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k} - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \right) \right) \\ &= \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} - 4\text{Re} \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \right) + \frac{(\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x})^2}{v_k^4} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}, \end{aligned} \quad (22)$$

and

$$\begin{aligned} & \lambda_{\max} (\mathcal{H}_{\Omega_k}) \\ & \leq \lambda_{\max} \left( \frac{1}{v_k} \begin{bmatrix} \mathbf{Q}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) + \lambda_{\max} \left( -\frac{2}{v_k^2} \text{Re} \left( \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} & \mathbf{x}^T \mathbf{Q}_{\mathbf{x},k}^T \end{bmatrix} \right) \right) \end{aligned} \quad (23)$$

$$\begin{aligned} & + \lambda_{\max} \left( -\frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} \right) \\ & + \lambda_{\max} \left( \frac{2\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \begin{bmatrix} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ \overline{\mathbf{Q}}_{\mathbf{x},k}^T \mathbf{x}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} & \mathbf{x}^T \overline{\mathbf{Q}}_{\mathbf{x},k}^T \end{bmatrix} \right) \\ &= \frac{1}{v_k} \lambda_{\max} (\mathbf{Q}_{\mathbf{x},k}) + \frac{4}{v_k^2} \max(-\text{Re}(\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}), 0) \\ & \quad - \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{\omega_k v_k^2} \lambda_{\min} (\overline{\mathbf{Q}}_{\mathbf{x},k}) + \frac{4\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \\ & \stackrel{(A5)}{\leq} \frac{1}{v_k} \lambda_{\max} (\mathbf{Q}_{\mathbf{x},k}) + \frac{4\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}. \end{aligned} \quad (24)$$

Equation (A5) in (18) is due to the fact that  $\overline{\mathbf{Q}}_{\mathbf{x},k}$  is a rank-1 positive semidefinite matrix and  $-\text{Re}(\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x})$  is non-positive. Take (22) and (24) into (20), we obtain

$$\begin{aligned} & \lambda_{\max} (\mathcal{H}_{u_k}) \\ & \leq \frac{1}{\omega_k^2 \epsilon} \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} - 4\text{Re} \left( \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x}}{v_k^3} \right) + \frac{6(\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x})^2}{v_k^4} \mathbf{x}^H \overline{\mathbf{Q}}_{\mathbf{x},k} \overline{\mathbf{Q}}_{\mathbf{x},k} \mathbf{x} \right) \quad (25) \\ & \quad + \frac{1}{\omega_k^2 \epsilon} \frac{\mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x}}{v_k^2} \lambda_{\max} (\mathbf{Q}_{\mathbf{x},k}) + \frac{1}{\epsilon} \lambda_{\max} (\mathcal{H}_{u_k}) \end{aligned}$$

$$(A6) \leq \begin{cases} \frac{1}{\omega_k^2 \epsilon \sigma_k^4} \left( (2 + 5\omega_k)a - \frac{4ab}{\sigma_k^2} + \frac{6ab^2}{\sigma_k^4} \right) & \text{if } \mathbf{x} = \mathbf{e} \\ \frac{b^2}{\omega_k^2 \epsilon \sigma_k^4 NN_{RF}} \left( 2 + \frac{(5\omega_k - 4b)}{\sigma_k^2} + \frac{6b^2}{\sigma_k^4} \right) & \text{if } \mathbf{x} = \text{vec}(\mathbf{A}) \\ \frac{NN_{RF}b^2}{\omega_k^2 \epsilon \sigma_k^4 P_{max}} \left( 2 + \omega_k + \frac{6b^2}{\sigma_k^4} \right) & \text{if } \mathbf{x} = \text{vec}(\mathbf{D}). \end{cases} \quad (26)$$

The inequality (A6) in (26) is obtained by using the similar violent relaxations as in (18). The results of the relaxations are straightforwardly given below

$$\begin{aligned} \mathbf{x}^H \mathbf{Q}_{\mathbf{x},k} \mathbf{x} &\leq b, \\ \mathbf{e}^H \mathbf{Q}_{\mathbf{e},k} \mathbf{Q}_{\mathbf{e},k} \mathbf{e} &\leq a, \\ \mathbf{e}^H \overline{\mathbf{Q}}_{\mathbf{e},k} \overline{\mathbf{Q}}_{\mathbf{e},k} \mathbf{e} &\leq a, \\ \text{Re}(\mathbf{e}^H \mathbf{Q}_{\mathbf{e},k} \overline{\mathbf{Q}}_{\mathbf{e},k} \mathbf{e}) &\leq a, \\ \text{vec}(\mathbf{A})^H \mathbf{Q}_{\mathbf{A},k} \mathbf{Q}_{\mathbf{A},k} \text{vec}(\mathbf{A}) &\leq \frac{b^2}{NN_{RF}}, \\ \text{vec}(\mathbf{A})^H \overline{\mathbf{Q}}_{\mathbf{A},k} \overline{\mathbf{Q}}_{\mathbf{A},k} \text{vec}(\mathbf{A}) &\leq \frac{b^2}{NN_{RF}}, \\ \text{Re}(\text{vec}(\mathbf{A})^H \mathbf{Q}_{\mathbf{A},k} \overline{\mathbf{Q}}_{\mathbf{A},k} \text{vec}(\mathbf{A})) &\leq \frac{b^2}{NN_{RF}}, \\ \text{vec}(\mathbf{D})^H \mathbf{Q}_{\mathbf{D},k} \mathbf{Q}_{\mathbf{D},k} \text{vec}(\mathbf{D}) &\leq \frac{b^2 NN_{RF}}{P_{max}}, \\ \text{vec}(\mathbf{D})^H \overline{\mathbf{Q}}_{\mathbf{D},k} \overline{\mathbf{Q}}_{\mathbf{D},k} \text{vec}(\mathbf{D}) &\leq \frac{b^2 NN_{RF}}{P_{max}}, \\ \text{Re}(\text{vec}(\mathbf{D})^H \mathbf{Q}_{\mathbf{D},k} \overline{\mathbf{Q}}_{\mathbf{D},k} \text{vec}(\mathbf{D})) &= 0, \\ \lambda_{\max}(\mathbf{Q}_{\mathbf{e},k}) &\leq \frac{b}{UM + 1}, \\ \lambda_{\max}(\mathbf{Q}_{\mathbf{A},k}) &\leq \frac{b}{NN_{RF}}, \\ \lambda_{\max}(\mathbf{Q}_{\mathbf{D},k}) &\leq \frac{b NN_{RF}}{P_{max}}. \end{aligned}$$

The proof is completed.

## REFERENCES

- [1] T. S. Rappaport *et al.*, “Millimeter wave mobile communications for 5G cellular: it will work!” *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [2] B. Di, H. Zhang, L. Song *et al.*, “Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, 2020.
- [3] V. Raghavan *et al.*, “Statistical blockage modeling and robustness of beamforming in millimeter-wave systems,” *IEEE Trans. Microwave Theory Tech.*, vol. 67, no. 7, pp. 3010–3024, Jul. 2019.

- [4] D. Kumar, J. Kaleva, and A. Tolli, “Rate and reliability trade-Off for mmWave communication via multi-point connectivity,” in *IEEE GLOBECOM*, 2019, pp. 1–6.
- [5] H. Imori *et al.*, “Stochastic learning robust beamforming for millimeter-wave systems with path blockage,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, 2020.
- [6] M. Di Renzo *et al.*, “Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison,” *IEEE Open J. Commun. Soc.*, vol. 1, pp. 798–807, 2020.
- [7] E. Basar *et al.*, “Wireless communications through reconfigurable intelligent surfaces,” *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.
- [8] M. Di Renzo *et al.*, “Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come,” *J. Wireless Commun. Netw.*, 2019,129(2019).
- [9] ———, “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [10] H. Zhang, B. Di, L. Song, and Z. Han, “Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?” *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4498–4502, 2020.
- [11] S. Zhang, H. Zhang, B. Di, Y. Tan, Z. Han, and L. Song, “Beyond intelligent reflecting surfaces: reflective-transmissive metasurface aided communications for full-dimensional coverage extension,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 905–13 909, 2020.
- [12] C. Pan, H. Ren, K. Wang *et al.*, “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.
- [13] C. Pan *et al.*, “Multicell MIMO communications relying on intelligent reflecting surfaces,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
- [14] T. Bai, C. Pan, Y. Deng *et al.*, “Latency minimization for intelligent reflecting surface aided mobile edge computing,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.
- [15] A. Zappone, M. Di Renzo, F. Shams, X. Qian, and M. Debbah, “Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments,” *IEEE Trans. Wireless Commun. (to appear)*.
- [16] G. Zhou *et al.*, “Robust beamforming design for intelligent reflecting surface aided MISO communication systems,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.
- [17] T. A. Le, T. Van Chien, and M. Di Renzo, “Robust probabilistic-constrained optimization for IRS-aided MISO communication systems,” *IEEE Wireless Commun. Lett., early access*, 2020.
- [18] M. R. Akdeniz *et al.*, “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.
- [19] P. L. Bartlett and M. H. Wegkamp, “Classification with a reject option using a hinge loss,” *J. Machine Learning Research*, vol. 9, pp. 1823–1840, Jun. 2008.
- [20] R. Xin, S. Kar, and U. A. Khan, “Decentralized stochastic optimization and machine learning: a unified variance-reduction framework for robust performance and fast convergence,” *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 102–113, 2020.
- [21] L. Bottou, *Online Learning and Stochastic Approximations*. Cambridge Univ. Press, 1998.
- [22] Y. Xu and W. Yin, “Block stochastic gradient iteration for convex and nonconvex optimization,” *SIAM J. Optim.*, vol. 25, no. 3, pp. 1686–1716, 2015.