Fairness-Oriented Multiple RIS-Aided mmWave Transmission: Stochastic Optimization Methods

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

***Abstract*—In millimeter wave (mmWave) systems, it is challenging to ensure reliable communication links due to the high sensitivity to the presence of blockages. In order to improve the robustness of mmWave systems in the presence of random blockages, we consider the deployment of multiple reconfigurable intelligent surfaces (RISs) to enhance the spatial diversity gain, and the design of robust beamforming schemes based on stochastic optimization methods that minimize the maximum outage probability among multiple users so as to ensure fairness. Under the stochastic optimization framework, we adopt the stochastic majorization–minimization (SMM) method and the stochastic successive convex approximation (SSCA) method to construct deterministic surrogate problems at each iteration, and to obtain closed-form solutions of the precoding matrix at the base station (BS) and the beamforming vectors at the RISs. Both stochastic optimization methods are proved to converge to the set of stationary points of the original stochastic problems. Simulation results show that the proposed robust beamforming for RIS-aided systems can effectively compensate for the performance loss caused by the presence of random blockages, especially when the blockage probability is high.**

***Index Terms*—Reconfigurable intelligent surface (RIS), intelligent reflecting surface (IRS), millimeter wave communications, stochastic optimization, robust beamforming design.**

I. Introduction

Millimeter wave (mmWave) communication is expected to be a promising technology to meet the growing demand for data rate in current and future wireless networks. mmWave communication systems are affected by severe signal attenuations. Thanks to the small signal wavelength, however, the high pathloss can be compensated by deploying antenna-arrays with a large number of antennas at the transmitters and receivers [2]. In addition, the high-directional beams obtained by utilizing large antenna-arrays can mitigate the inter-user interference. However, mmWave communication systems suffer from high penetration losses [3]–[5]. Hence, mmWave systems are much more susceptible to the presence of spatial blockages than sub-6 GHz systems, and the reliability of the communication links cannot always be guaranteed throughout the whole network [3]–[5].

In particular, spatial blockages can be divided into static blockages (e.g., buildings and other static objects), dynamic blockages (e.g., human beings, vehicles, or moving obstructions) and self-blockages (e.g., hand blocking of the user itself and blockage from other parts of the body). To account for the impact of blockages, statistical models have been proposed to characterize the properties of dynamic blockages and self-blockages [4]–[6]. The authors of [7] have developed a distance-dependent blockage probability model, in which the probability that a link is blocked increases exponentially with the length of the link. Furthermore, the authors of [8] and [9] have proposed an approach to predict the blockage probability via machine learning. If the blockage probability is known or is estimated, robust beamforming design strategies can be utilized to address the channel uncertainties caused by the presence of random blockages [10], [11]. Specifically, the authors of [10] have proposed a worst-case robust beamforming design for application to coordinated multipoint (CoMP) systems in which all possible combinations of blockage patterns are considered. Due to the high computational complexity of the method introduced in [10], an outage-minimum strategy based on a stochastic optimization method has been proposed in [11] to improve the robustness of mmWave systems against random blockages. The use of multiple base stations (BSs) in CoMP systems may be an option to compensate for the performance loss caused by the presence of random blockages by exploiting spatial diversity gains. However, this solution incurs in excessive hardware cost and power consumption. Another possible solution consists of deploying cost-efficient reconfigurable intelligent surfaces (RISs) so as to create alternative and reliable communication links in mmWave systems [12].

Specifically, RISs have attracted major research interest since they operate without the need of power amplifiers and digital processing units, thus enabling energy-efficient and spectral-efficient communications [13]–[16]. An RIS is a thin surface that consists of nearly-passive and reconfigurable reflecting elements, which reflects the impinging radio waves without amplifying the signals and processing them in the digital domain. The RIS elements can be tuned to alter their electromagnetic response such that the reradiated signals are constructively superimposed to enhance the signal

This work was supported in part by the National Key Research and Development Project (2019YFE0123600), National Natural Science Foundation of China (62101128), Basic Research Project of Jiangsu Provincial Department of Science and Technology (BK20210205), High Level Personal Project of Jiangsu Province (JSSCBS20210105), and the Natural Science Foundation of Shanghai under Grant 22ZR1445600. The work of M. Di Renzo was supported in part by the European Commission through the H2020 ARIADNE project under grant agreement number 871464 and through the H2020 RISE-6G project under grant agreement number 101017011. Part of this work has been published in [1]. (Corresponding author: Cunhua Pan)

G. Zhou is with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. (e-mail: g.zhou@qmul.ac.uk) C. Pan and H. Ren are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (Email: cpan, 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. (e-mail: marco.di-renzo@universite-paris-saclay.fr)

power at the intended receiver or are destructively combined to avoid the leakage of information to undesired receivers. These characteristics make RISs an appealing technology in various communication systems [17]. For instance, RISs can be applied in single-cell multiple-input and multiple-output (MIMO) systems [18]–[22], multicell MIMO communications [23], simultaneous wireless information and power transfer (SWIPT) systems [18], [24], secure communications [25], mmWave systems [26]–[29] and terahertz (THz) systems [30], [31].

Although the performance advantages of deploying RISs in mmWave systems have been demonstrated in recent contributions, there still exist major open problems to solve. Examples include the following. The authors of [27] have only considered the BS-RIS-user channels and have assumed that the direct BS-user communication links are completely blocked by obstacles. However, this assumption only applies to the case of static blockages with blockage probability equal to one, but it does not apply in the presence of dynamic blockages since the blockage probability lies in [0,1] [3]–[5]. The numerical results illustrated in [28], for example, have shown that the gain from additional reflections can compensate for the performance loss caused by the presence of random blockages. However, the impact of blockages was not considered in the beamforming design. Most recently, we have considered the design of robust beamforming methods for RIS-aided mmWave communication systems by taking the presence of random blockages into consideration [12]. However, the considered optimization problem is the minimization of the sum outage probability. Therefore, the considered optimization problem cannot ensure the fairness among all the users.

A. Novelty and Contributions

Against the above background, this paper proposes a robust transmission strategy for application to RIS-aided mmWave communication systems, which accounts for the channel uncertainties caused by the presence of random blockages while ensuring the fairness among the users. Typical methods to handle the presence of channel uncertainties at the design stage are the outage constrained robust optimization and the worst-case robust optimization techniques [32]. However, both methods rely on the estimation of the instantaneous CSI. Furthermore, the worst-case robust optimization method is conservative and hence suboptimal due to the low probability of occurrence of the worst case. In this paper, we consider the design of robust beamforming schemes for application to mmWave systems without relying on the knowledge of instantaneous CSI. The proposed approach is based, on the other hand, on the knowledge of large-scale CSI and the blockage probability. The proposed approach is motivated by the results reported in [5], where the authors have shown that the time-scale at which mmWave signals are randomly disrupted by spatial blockages, such as the hands and the human body, is of the order of a few 100 milliseconds (or more), as well as the results reported in [7], where the authors have shown that the blockage probability is determined by the transmission distance and by some environment-specific parameters. Therefore, a mmWave link is disrupted by the same blockages for several physical layer resource blocks and the associated probability can be assumed to be known if the large-scale CSI is assumed to be known. Specifically, we formulate a maximum outage probability minimization problem and solve it by using a stochastic optimization framework.

The main contributions of this work can be summarized as follows:

- To the best of our knowledge, this is the first work that introduces a robust beamforming design for RIS-aided downlink multiuser mmWave systems that relies on the knowledge of large-scale CSI and blockage probability. Specifically, the considered optimization criterion is based on minimizing the maximum outage probability of all the users. In contrast to the sum outage probability minimization problem in [11], [12], the considered min-max outage probability problem ensures the desired quality of service (QoS) performance to the worst-case user. Because of the non-differentiable objective function of the considered problem, the stochastic gradient descent (SGD) method adopted in [11], [12] cannot be directly applied. To circumvent this issue, two stochastic optimization frameworks are introduced for jointly optimizing the beamforming at the BS and at the RIS.
- First, we consider the single-user case and optimize the beamforming schemes at the BS and RIS by minimizing the outage probability given the large-scale CSI and the blockage probability. Since the objective function of the considered problem is not formulated in a closed-form expression, we approximate it with the statistical expectation of a smooth function that is twice differentiable. The resulting expectation optimization problem is solved by adopting the stochastic majorization–minimization (SMM) method. Specifically, an upper bound surrogate function of the original differentiable function is constructed for any new channel realization at each iteration. The constructed surrogate problem is shown to have a closed-form solution and to be computationally efficient. We prove that the proposed SMM method is guaranteed to converge to the set of stationary points of the original expectation minimization problem.
- Then, we consider the multi-user case and we formulate a min-max outage probability optimization problem. To tackle the non-differentiability of the max objective function, we replace it with the log-sum-exp upper bound. Then, we employ the stochastic successive convex approximation (SSCA) method, which offers a greater flexibility than the SMM method in terms of selecting the surrogate function and results in closed-form expressions at each iteration. Also in this case, we prove that the proposed SSCA method is guaranteed to converge to the set of stationary points of the original expectation minimization problem.
- We demonstrate through numerical results that the proposed robust beamforming algorithm outperforms its non-robust counterpart and the robust beamforming algorithm for conventional systems in the absence of RISs. If the blockage probability is high, the proposed methods

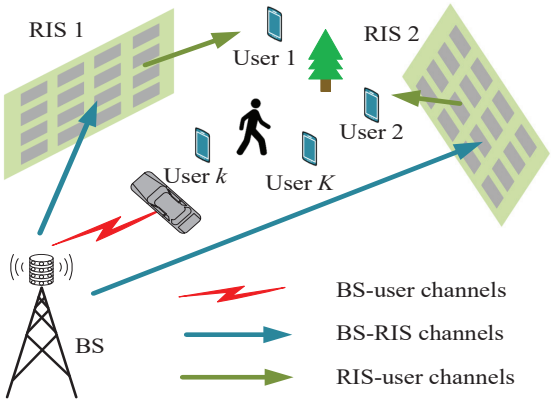


Fig. 1: Multiple RIS-aided mmWave communication system.

outperform the others in terms of maximum outage probability and minimum effective rate. Moreover, deploying multiple small-size RISs is shown to provide better performance than deploying a single large-size RIS in terms of improving the performance of the worst-case user.

The remainder of this paper is organized as follows. Section II introduces the system model. The outage probability minimization problem for the single-user case is formulated in Section III. In Section IV, the min-max outage probability problem for multiuser systems is discussed. Finally, Sections VI and VII report the numerical results and the conclusions, respectively.

Notations: The following mathematical notations and symbols are used throughout this paper. Vectors and matrices are denoted by boldface lowercase letters and boldface uppercase letters, respectively. The symbols $\mathbf{X}^*$, $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}^{\mathrm{H}}$, and $||\mathbf{X}||_F$ denote the conjugate, transpose, Hermitian (conjugate transpose), Frobenius norm of matrix $\mathbf{X}$, respectively. The symbol $||\mathbf{x}||_2$ denotes the 2-norm of vector $\mathbf{x}$. The symbols $\mathrm{Tr}\{\cdot\}$, $\mathrm{Re}\{\cdot\}$, $|\cdot|$, $\lambda(\cdot)$, and $\angle(\cdot)$ denote the trace, real part, modulus, eigenvalue, and angle of a complex number, respectively. $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the entries of $\mathbf{x}$ on its main diagonal. $[\mathbf{x}]_m$ denotes the m-th element of vector $\mathbf{x}$. The Kronecker product between two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \otimes \mathbf{Y}$. $\mathbf{X} \succeq \mathbf{Y}$ means that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite. The symbol $\mathbb{C}$ denotes the complex field, $\mathbb{R}$ denotes the real field, and $\mathrm{j} \triangleq \sqrt{-1}$ is the imaginary unit. The inner product $\langle \bullet, \bullet \rangle : \mathbb{C}^{M \times N} \times \mathbb{C}^{M \times N} \to \mathbb{R}$ is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Re}\{\mathrm{Tr}\{\mathbf{X}^{\mathrm{H}}\mathbf{Y}\}\}$.

II. System Model

A. Signal Model

As shown in Fig. 1, we consider an RIS-aided downlink mmWave communication system. In order to ensure high QoS for the users in the presence of random blockages, U RISs, each of which has M reflecting elements, are deployed to assist the communication from the BS equipped with N active antennas to K single-antenna users (denoted by $\mathcal{K} \triangleq \{1, \ldots, K\}$). The RISs are assumed to be connected to controllers that exchange control information with the BS through dedicated channels [18], [23]. The baseband transmitted signal at the BS is $\mathbf{x} = \mathbf{F}\mathbf{s}$, where $\mathbf{s} \in \mathbb{C}^{K \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the Gaussian data symbol vector and $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] \in \mathbb{C}^{N \times K}$ denotes the full-digital beamforming matrix. The baseband transmit power is limited to the total transmit power P_{max}. Hence, $\mathbf{F}$ belongs to the set $\mathcal{S}_f = \{\mathbf{F} \mid ||\mathbf{F}||_F^2 \leq P_{max}\}$.

Let $\mathbf{h}_{\mathrm{b},k} \in \mathbb{C}^{N \times 1}$, $\mathbf{H}_u \in \mathbb{C}^{M \times N}$ and $\mathbf{h}_{u,k} \in \mathbb{C}^{M \times 1}$ denote the channels of the links from the BS to the k-th user, from the BS to the u-th RIS, and from the u-th RIS to the k-th user, respectively. Then, the received signal intended to the k-th user is expressed as [1]

$$y_k = \left(\mathbf{h}_{\mathrm{b},k}^{\mathrm{H}} + \sum_{u=1}^{U} \mathbf{h}_{u,k}^{\mathrm{H}} \mathbf{E}_u \mathbf{H}_u \right) \mathbf{x} + n_k, \tag{1}$$

where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive white Gaussian noise (AWGN), and $\mathbf{E}_u = \zeta \mathrm{diag}([e_{(u-1)M+1}, \ldots, e_{uM}])$ is the reflection coefficient matrix (also known as the RIS beamforming matrix) of the u-th RIS. The element $e_{(u-1)M+m}$ is the m-th unit modulus reflection coefficient at the u-th RIS, and $\zeta \in [0, 1]$ denotes the reflection efficiency. We assume, independently of the applied phase shift and of the angle of incidence, $\zeta = 1$ for simplicity, since it offers the best reflection performance for the RIS.

By defining the matrices $\mathbf{h}_k = [\mathbf{h}_{1,k}^{\mathrm{H}}, \ldots, \mathbf{h}_{U,k}^{\mathrm{H}}]^{\mathrm{H}}$ and $\mathbf{H} = [\mathbf{H}_1^{\mathrm{H}}, \ldots, \mathbf{H}_U^{\mathrm{H}}]^{\mathrm{H}}$, we obtain the equivalent channel $\mathbf{G}_k = \begin{bmatrix} \mathrm{diag}(\mathbf{h}_k^{\mathrm{H}})\mathbf{H} \\ \mathbf{h}_{\mathrm{b},k}^{\mathrm{H}} \end{bmatrix} \in \mathbb{C}^{(UM+1) \times N}$ between the BS and the k-th user. The corresponding equivalent reflection coefficient vector is given by $\mathbf{e} = [e_1, \ldots, e_{UM+1}]^{\mathrm{T}} \in \mathbb{C}^{(UM+1) \times 1}$ which belongs to the set $\mathcal{S}_e = \{\mathbf{e} \mid |e_m|^2 = 1, 1 \leq m \leq UM, e_{UM+1} = 1\}$. Then, (1) can be rewritten in a compact form as

$$y_k = \mathbf{e}^{\mathrm{H}} \mathbf{G}_k \mathbf{F} \mathbf{s} + n_k. \tag{2}$$

Therefore, the corresponding achievable signal-to-interference-plus-noise ratio (SINR) is

$$\Gamma_k(\mathbf{F}, \mathbf{e}) = \frac{|\mathbf{e}^{\mathrm{H}} \mathbf{G}_k \mathbf{f}_k|^2}{\sum_{i \neq k}^{K} |\mathbf{e}^{\mathrm{H}} \mathbf{G}_k \mathbf{f}_i|^2 + \sigma_k^2}. \tag{3}$$

B. Channel Model

We consider a typical Saleh-Valenzuela (SV) [34] channel model for application to mmWave systems. For simplicity, we ignore the randomness introduced by the presence of hardware impairments that may affect the performance of mmWave systems. In particular, it is assumed that a uniform planar array (UPA) is deployed at the BS and at the RIS. The steering vector of each UPA is denoted by $\mathbf{a}(\varphi, \phi)$, where $\varphi(\phi)$ denotes the azimuth (elevation) angle of departure (AoD) and angle of arrival (AoA) depending on whether a transmitter or a receiver is considered. We assume that $L_{\mathrm{b},k}$, $L_{u,k}$ and $L_{\mathrm{b},u}$ sparse scatterers exist on the links from the BS to the k-th user, from the u-th RIS to the k-th user, and from the BS to the u-th RIS, respectively. Also, we assume that each scatterer

[1] For simplicity, we ignore the reflections of signals between the RISs, since they are typically weak in the mmWave frequency band. The impact of the reflected signals between RISs was recently addressed in [33].

comprises I subpaths. In the far-field region, therefore, the mmWave channels can be expressed as

$$\mathbf{h}_{\mathrm{b},k} = g_0^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,0}^{\mathrm{AoD}},\phi_{\mathrm{b},k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{\mathrm{b},k}}}\sum_{l=1}^{L_{\mathrm{b},k}}\sum_{i=1}^{I} g_{l,i}^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,l,i}^{\mathrm{AoD}},\phi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}\right), \forall k, \tag{4}$$

$$\mathbf{h}_{u,k} = g_0^{u,k}\mathbf{a}\left(\varphi_{u,k,0}^{\mathrm{AoD}},\phi_{u,k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{u,k}}}\sum_{l=1}^{L_{u,k}}\sum_{i=1}^{I} g_{l,i}^{u,k}\mathbf{a}\left(\varphi_{u,k,l,i}^{\mathrm{AoD}},\phi_{u,k,l,i}^{\mathrm{AoD}}\right), \forall k, \forall u, \tag{5}$$

$$\mathbf{H}_u = g_0^{\mathrm{b},u}\mathbf{a}\left(\varphi_{u,0}^{\mathrm{AoA}},\phi_{u,0}^{\mathrm{AoA}}\right)\mathbf{a}\left(\varphi_{\mathrm{b},0}^{\mathrm{AoD}},\phi_{\mathrm{b},0}^{\mathrm{AoD}}\right)^{\mathrm{H}} + \sqrt{\frac{1}{IL_{\mathrm{b},u}}}\sum_{l=1}^{L_{\mathrm{b},u}}\sum_{i=1}^{I} g_{l,i}^{\mathrm{b},u}\mathbf{a}\left(\varphi_{u,l,i}^{\mathrm{AoA}},\phi_{u,l,i}^{\mathrm{AoA}}\right)\mathbf{a}\left(\varphi_{\mathrm{b},l,i}^{\mathrm{AoD}},\phi_{\mathrm{b},l,i}^{\mathrm{AoD}}\right)^{\mathrm{H}}, \forall u, \tag{6}$$

where, by denoting an arbitrary element $q \in \{(\mathrm{b},k),(u,k),(\mathrm{b},u)\}_{\forall k,\forall u}$, $g_0^q\mathbf{a}\left(\varphi_{q,0}^{\mathrm{AoD}},\phi_{q,0}^{\mathrm{AoD}}\right)$ is the line-of-sight (LoS) component whose fading coefficient has distribution $g_0^q \sim \mathcal{CN}(0,\zeta_0^q 10^{\frac{\mathrm{PL}}{10}})$, where $\zeta_0^q = \frac{\kappa}{1+\kappa}$ is the power that corresponds to the Rician factor κ, and PL is the distance-dependent pathloss. The remaining paths are the non-LoS (NLoS) components whose fading coefficients have distribution $g_{l,i}^q \sim \mathcal{CN}(0,\zeta_l^q 10^{\frac{\mathrm{PL}}{10}})$ where $\zeta_l^q = \frac{1}{(L_q-1)I(1+\kappa)}$ is the corresponding power fraction.

We assume that the users' locations are quasi-static over milliseconds or even seconds. Therefore, the large-scale fading parameters, such as the distance-dependent pathloss, the number of clusters, the Rician factor, the cluster central angles, and the angular spreads, change relatively slowly and can be assumed to be known by the BS [35], [36]. However, the instantaneous CSI, which is given by $\{\mathbf{h}_{\mathrm{b},k},\mathbf{h}_{u,k},\mathbf{H}_u\}$, vary during the data transmission because of the rapidly varying small-scale fading coefficients $\{g_0^q, g_{l,i}^q\}$, AoDs and AoAs. In general, these parameters vary according to an ergodic stationary process. For example, the AoDs and AoAs can be generated according to a Gaussian distribution, whose mean value coincide with the central angles of the clusters and the variance coincides with the angular spread [35].

Besides the pathloss and the small-scale fading, the reliability of the communication links in the mmWave frequency band is determined by the presence of blockages [37]. In the context of RIS-aided communications, most existing contributions have considered the worst-case scenario where the BS-user links are completely blocked due to the presence of obstacles during the whole transmission, while the RIS-related links are not affected by the presence of blockages since the locations of the RISs can be appropriately chosen to ensure line-of-sight transmission. However, this assumption may not represent all possible deployment scenarios. Traditionally, the presence of blockages is incorporated in the shadowing model, along with the impact of reflections, scattering, and diffraction [6]. In contrast, we adopt a recently proposed probabilistic model [11] to characterize the channel uncertainties caused by the presence of random blockages. The considered model is more realistic, since the impact of blockages and the corresponding blockage probability depends on the transmission distance [6].

Specifically, let us introduce the blockage parameters $\omega_{k,l} \in \{0,1\}, 0 \leq l \leq L_{\mathrm{b},k}, \forall k \in \mathcal{K}$, which are random variables with a Bernoulli distribution. In particular, these random variables take the value one with probability p_k, which is referred to as the blockage probability. With the aid of these random variables, the presence of blockages can be taken into account in the formulation of the mmWave channels. Specifically, the BS-user links in (4) are modified as

$$\mathbf{h}_{\mathrm{b},k} = \omega_{k,0}g_0^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,0}^{\mathrm{AoD}},\phi_{\mathrm{b},k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{\mathrm{b},k}}}\sum_{l=1}^{L_{\mathrm{b},k}}\omega_{k,l}\sum_{i=1}^{I} g_{l,i}^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,l,i}^{\mathrm{AoD}},\phi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}\right), \forall k, \tag{7}$$

As far as the blockage probability is concerned, it is known that it usually depends on the transmission distance. For example, the authors of [6] have shown that the probability that a link is blocked, i.e., there is at least one object in between the transmitter and the receiver, can be formulated as $p_k(d_k) = \max(0, 1 - e^{-a_{out}d_k + b_{out}})$, where d_k is the transmission distance between the BS and the k-th user in the considered system model and a_{out} and b_{out} are environment-dependent parameters that can be obtained from theory or can be obtained from curve fitting from data [7], [Table II, 38].

C. Problem Formulation

Since the RISs are not equipped with power amplifiers and with digital signal processing units, the acquisition of CSI is difficult to obtain. This is especially true if the BS and the RISs are optimized based on perfect instantaneous CSI, since a large training overhead would be needed [38]. Therefore, it is important to develop robust beamforming schemes that do not necessarily rely on the knowledge of the instantaneous CSI, but still account for the impact of the large-scale CSI, which is easier to acquire at a reduced overhead, and that are robust to the presence of random blockages. Motivated by this consideration, we aim to design robust beamforming schemes for RIS-aided systems that depend only on the large-scale CSI and the blockage probability, but are independent of the short-term CSI, i.e., the fast fading. Also, we aim to ensure that the beamforming schemes provide fairness to the network users.

To this end, we formulate the following min-max outage probability optimization problem

$$\min_{\mathbf{F},\mathbf{e}} \ \max_{k\in\mathcal{K}} \Pr\{\Gamma_k(\mathbf{F},\mathbf{e}) \leq \gamma_k\} \tag{8a}$$

$$\text{s.t. } \mathbf{F} \in \mathcal{S}_f \tag{8b}$$

$$\mathbf{e} \in \mathcal{S}_e, \tag{8c}$$

where the outage probability $\Pr\{\Gamma_k(\mathbf{F},\mathbf{e}) \leq \gamma_k\}$ is the probability that the SINR $\Gamma_k(\mathbf{F},\mathbf{e})$ of the k-th user is less than the SINR reliability threshold γ_k for all possible realizations of the random channel $\mathbf{G} = [\mathbf{G}_1,\ldots,\mathbf{G}_K]$. Specifically, the probability in (8) is computed as a function of the small-scale fading coefficients, the AoDs and the AoAs of the subpaths of the scatterers. Notably, the outage probability in (8) depends

on the transmission distance, the blockage probability, the number of clusters and their centers and angular spreads.

Compared with the sum outage probability minimization problem formulated in [12], the objective function in (8) ensures fairness among the users. However, due to the min-max formulation, the objective function is not smooth and differentiable, which makes the algorithms proposed in [12] not directly applicable to solve the problem in (8).

III. Single-User System

In this section, we consider a single-user system model in order to obtain some design insights. By setting $K = 1$ and omitting the user index, the problem in (8) reduces to

$$\min_{\mathbf{f},\mathbf{e}} \ \Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\} \tag{9a}$$

$$\text{s.t. } \mathbf{f} \in \mathcal{S}_f \tag{9b}$$

$$\mathbf{e} \in \mathcal{S}_e. \tag{9c}$$

A. Problem Reformulation

The probability $\Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\}$ has no closed-form expression and thus the problem in (9) is prohibitively challenging to be solved. To tackle this issue, we reformulate the probability function in terms of an equivalent expectation function, i.e., $\Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\} = \mathbb{E}_{\mathbf{G}}[\mathbb{I}_{\Gamma\leq\gamma}]$ where $\mathbb{I}_{\Gamma\leq\gamma}$ denotes the step function applied to the event $\Gamma \leq \gamma$. Thanks to this reformulation, several stochastic programming techniques can be used to solve the problem in (9). However, the step function is discontinuous, and the existing stochastic programming methods cannot be directly applied.

To circumvent this issue, we approximate the step function with the following smooth approximating function

$$u(x) = \frac{1}{1+e^{-\theta x}}, \tag{10}$$

where $x = \gamma - \Gamma$ and θ is a smooth parameter that controls the approximation error. Specifically, the larger θ is the closer to an ideal step function the function in (10) is.

By defining $f(\mathbf{f},\mathbf{e}|\mathbf{G}) = u\left(\gamma\sigma^2 - |\mathbf{e}^{\mathrm{H}}\mathbf{G}\mathbf{f}|^2\right)$, a convenient approximation of the problem formulated in (9) is

$$\min_{\mathbf{f}\in\mathcal{S}_f,\mathbf{e}\in\mathcal{S}_e} \ g(\mathbf{f},\mathbf{e}|\mathbf{G}) = \mathbb{E}\left[f(\mathbf{f},\mathbf{e}|\mathbf{G})\right]. \tag{11}$$

B. Stochastic Majorization-Minimization Method

A simple approach for solving the problem in (11) is the sample average approximation (SAA) method. However, the SAA method is computationally prohibitive since it requires large-size memory storage due to the fact that the solution obtained at each iteration is calculated by averaging over a large number of channel realizations. To overcome these difficulties, we adopt the widely used SMM [39] (also known as stochastic successive minimization [40]) method. Accordingly, an appropriately chosen upper bound approximation for the function $f(\mathbf{f},\mathbf{e}|\mathbf{G})$ is constructed at each iteration of the algorithm and for each channel realization. The solution is obtained as the average over the channel realizations at each iteration.

The typical approach to apply this method consists of introducing an upper bound approximation function for $f(\mathbf{f},\mathbf{e}|\mathbf{G})$ that makes the corresponding surrogate problem easy to solve. In some cases, a closed-form solution may be found as well. Let $\mathbf{x} \in \{\mathbf{f},\mathbf{e}\}$ denote the optimization vector whose feasible set is $\mathcal{S}_x \in \{\mathcal{S}_f,\mathcal{S}_e\}$. The surrogate function $\hat{f}(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G})$ of $f(\mathbf{x}|\mathbf{G})$ around any feasible point $\mathbf{x}^{i-1}$ needs to satisfy the following assumptions [40].

Assumptions A

(A1) : $\hat{f}\left(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G}\right)$ is continuous in $\mathbf{x}$ for $\forall \mathbf{x}^{i-1} \in \mathcal{S}_x$.

(A2) : $\hat{f}(\mathbf{x}^{i-1},\mathbf{x}^{i-1}|\mathbf{G}) = f(\mathbf{x}^{i-1}|\mathbf{G}), \forall \mathbf{x}^{i-1} \in \mathcal{S}_x$.

(A3) : $\hat{f}(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G}) \geq f(\mathbf{x}|\mathbf{G}), \forall \mathbf{x},\mathbf{x}^{i-1} \in \mathcal{S}_x$.

(A4) : $\hat{f}^{'}(\mathbf{x}^{i-1}|\mathbf{G};\mathbf{d}) = f^{'}(\mathbf{x}^{i-1}|\mathbf{G};\mathbf{d})$, for all $\mathbf{x}^{i-1} \in \mathcal{S}_x$ and for all feasible directions $\mathbf{d}$ at $\mathbf{x}^{i-1}$,

where $f^{'}(\mathbf{x}^{i-1}|\mathbf{G};\mathbf{d})$ is the directional derivative of $f(\mathbf{x}^{i-1}|\mathbf{G})$ in the direction $\mathbf{d}$ and is given by

$$f^{'}(\mathbf{x}^{i-1}|\mathbf{G};\mathbf{d}) = \lim_{\lambda\to 0} \frac{f(\mathbf{x}^{i-1}+\lambda\mathbf{d}|\mathbf{G}) - f(\mathbf{x}^{i-1}|\mathbf{G})}{\lambda}.$$

The Assumptions (A2)-(A3) indicate that the surrogate function $\hat{f}(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G})$ is a local upper bound of the original function $f(\mathbf{x}|\mathbf{G})$ around the feasible point $\mathbf{x}^{i-1}$. The Assumption (A4) is a consistency condition for the first-order directional derivative. To ensure the convergence of the SMM algorithm, the following conditions need to be fulfilled [40].

Assumption: B

(B1) : The feasible set $\mathcal{S}_x$ and the channel realizations are bounded.

(B2) : The functions $\hat{f}(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G})$ and $f(\mathbf{x}|\mathbf{G})$, their derivatives, and their second-order derivatives are uniformly bounded.

Since the variables $\mathbf{f}$ and $\mathbf{e}$ are highly coupled with each other, we adopt an alternating optimization (AO) method to update them. Based on the above assumptions, the variables $\mathbf{f}$ and $\mathbf{e}$ are updated, at the n-th iteration of the algorithm, by solving the following two SMM subproblems

$$\mathbf{f}^n = \arg\min_{\mathbf{f}\in\mathcal{S}_f} \ \frac{1}{n}\sum_{i=1}^{n} \hat{f}\left(\mathbf{f},\mathbf{f}^{i-1}|\mathbf{G}^i\right), \tag{12}$$

and

$$\mathbf{e}^n = \arg\min_{\mathbf{e}\in\mathcal{S}_e} \ \frac{1}{n}\sum_{i=1}^{n} \hat{f}\left(\mathbf{e},\mathbf{e}^{i-1}|\mathbf{G}^i\right), \tag{13}$$

where $\mathbf{G}^1,\mathbf{G}^2,...$ are some independent samples of the random equivalent channel $\mathbf{G}$ [2]. Furthermore, $\hat{f}\left(\mathbf{f},\mathbf{f}^{i-1}|\mathbf{G}^i\right)$ denotes the surrogate function associated to $\mathbf{f}$ when $\mathbf{e}$ is given, and $\hat{f}\left(\mathbf{e},\mathbf{e}^{i-1}|\mathbf{G}^i\right)$ is the surrogate function that corresponds to $\mathbf{e}$ when $\mathbf{f}$ is given.

[2]More precisely, we assume that the large-scale fading parameters are kept fixed, and that the samples of $\mathbf{G}$ are constituted by $\{\mathbf{h}_{\mathrm{b},k},\mathbf{h}_{u,k},\mathbf{H}_u\}$, which are obtained by generating the random variables $\{g_0^q, g_{l,i}^q\}$, AoDs and AoAs according to their distributions whose parameters are assumed to be known, as well as the Bernoulli random variables $\omega_{k,l}$ whose blockage probability p_k is assumed to be known.

1) Optimizing $\mathbf{f}$*:* First, we construct $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i\right)$ so as to fulfill the Assumptions A and B. This is given in the following lemma.

Lemma 1 *Given the twice differentiable function* $f(\mathbf{f}|\mathbf{G}^i)$*, we consider the following second-order upper bound approximation around any given* $\mathbf{f}^{i-1}$

$$\hat{f}(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i) = 2\mathrm{Re}\left\{\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \alpha_f^i||\mathbf{f}||_2^2 + const_f^i, \quad (14)$$

where

$$\mathbf{d}_f^i = \mathbf{m}_f^i - \alpha_f^i\mathbf{f}^{i-1}, \quad (15a)$$

$$\mathbf{m}_f^i = \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}, \quad (15b)$$

$$x^i = \gamma\sigma^2 - |\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}|^2, \quad (15c)$$

$$\alpha_f^i = \frac{\theta^2}{2}P_{max}|\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}|^2, \quad (15d)$$

$$const_f^i = f(\mathbf{f}^{i-1}|\mathbf{G}^i) + \alpha_f^i||\mathbf{f}^{i-1}||_2^2 - 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}\mathbf{f}^{i-1}\right\}. \quad (15e)$$

Proof: See Appendix A. ■

By using (14) and ignoring irrelevant constants, the subproblem in (12) for updating $\mathbf{f}$ is reformulated as

$$\min_{\mathbf{f}\in\mathcal{S}_f} \; 2\mathrm{Re}\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \frac{1}{n}\sum_{i=1}^n \alpha_f^i||\mathbf{f}||_2^2. \quad (16)$$

The optimization problem in (16) is convex and can be solved by computing its Lagrange function given by

$$\mathcal{L}(\mathbf{f},\kappa) = 2\mathrm{Re}\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \frac{1}{n}\sum_{i=1}^n \alpha_f^i||\mathbf{f}||_2^2 + \kappa\left(||\mathbf{f}||_2^2 - P_{max}\right), \quad (17)$$

where $\kappa \geq 0$ is the Lagrange multiplier associated with the power constraint. By setting $\partial\mathcal{L}(\mathbf{f})/\partial\mathbf{f}^* = \mathbf{0}$, the globally optimal solution of $\mathbf{f}$ at the n-th iteration is

$$\mathbf{f}^n = \frac{-1}{\kappa + \frac{1}{n}\sum_{i=1}^n \alpha_f^i}\frac{1}{n}\sum_{i=1}^n \mathbf{d}_f^i. \quad (18)$$

Also, (18) must satisfy the power constraint in (9b), which yields

$$\frac{||\frac{1}{n}\sum_{i=1}^n \mathbf{d}_f^i||_2^2}{(\kappa + \frac{1}{n}\sum_{i=1}^n \alpha_f^i)^2} \leq P_{max}. \quad (19)$$

Since the left hand side of (19) is a decreasing function of κ, we obtain the following closed-form solution

$$\mathbf{f}^n = \begin{cases} \frac{-1}{\sum_{i=1}^n \alpha_f^i}\sum_{i=1}^n \mathbf{d}_f^i, & \text{if } \frac{||\sum_{i=1}^n \mathbf{d}_f^i||_2^2}{(\sum_{i=1}^n \alpha_f^i)^2} \leq P_{max}, \\ -\sqrt{\frac{P_{max}}{||\sum_{i=1}^n \mathbf{d}_f^i||_2^2}}\sum_{i=1}^n \mathbf{d}_f^i, & \text{otherwise}, \end{cases} \quad (20)$$

where the first case in (20) is obtained by setting $\kappa = 0$, and the second case follows because there must exist a $\kappa > 0$ for which (19) holds with equality.

2) Optimizing $\mathbf{e}$*:* Similar to the optimization of $\mathbf{f}$, we first construct a surrogate function for $\mathbf{e}$.

Lemma 2 *Given the twice differentiable function* $f\left(\mathbf{e}|\mathbf{G}^i\right)$*, we consider the following second-order upper bound approximation around any feasible* $\mathbf{e}^{i-1}$

$$\hat{f}(\mathbf{e}, \mathbf{e}^{i-1}|\mathbf{G}^i) = 2\mathrm{Re}\left\{\mathbf{d}_e^{i,\mathrm{H}}\mathbf{e}\right\} + const_e^i, \quad (21)$$

where

$$\mathbf{d}_e^i = \mathbf{m}_e^i - \alpha_e^i\mathbf{e}^{i-1}, \quad (22a)$$

$$\mathbf{m}_e^i = \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^i\mathbf{f}^{i-1}\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}, \quad (22b)$$

$$\alpha_e^i = \frac{\theta^2}{2}(UM+1)|\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}|^2, \quad (22c)$$

$$const_e^i = f(\mathbf{e}^{i-1}|\mathbf{G}^i) + 2(UM+1)\alpha_e^i - 2\mathrm{Re}\left\{\mathbf{m}_e^{i,\mathrm{H}}\mathbf{e}^{i-1}\right\}. \quad (22d)$$

Proof: The proof of Lemma 2 is similar to that of Lemma 1 and it is hence omitted for brevity. ■

By substituting (21) into the objective function of the subproblem in (13) and ignoring irrelevant constants, we obtain

$$\min_{\mathbf{e}\in\mathcal{S}_e} \; 2\mathrm{Re}\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{d}_e^{i,\mathrm{H}}\mathbf{e}\right\}. \quad (23)$$

The globally optimal solution of the optimization problem in (23) is

$$\mathbf{e}^n = \exp\left\{\mathrm{j}\angle\left(\left(\sum_{i=1}^n \mathbf{d}_e^i\right)/\left[\sum_{i=1}^n \mathbf{d}_e^i\right]_{UM+1}\right)\right\}, \quad (24)$$

where $[\cdot]_m$ denotes the m-th element of the vector, $\mathrm{j} \triangleq \sqrt{-1}$ is the imaginary unit, $\angle(\cdot)$ denotes the angle of a complex number, and $\exp\{\mathrm{j}\angle(\cdot)\}$ is an element-wise operation.

C. Algorithm Development

By leveraging the SMM method, we have obtained the closed-form solutions in (20) and (24) for $\mathbf{f}$ and $\mathbf{e}$, respectively. The closed-form solutions, at each iteration of the algorithm, greatly reduce the computational complexity. The whole numerical recipe is reported in Algorithm 1, which is referred to as the SMM-OutMin algorithm.

Algorithm 1 SMM-OutMin Algorithm

Initialize: Initialize $\mathbf{f}^0 \in \mathcal{S}_f$ and $\mathbf{e}^0 \in \mathcal{S}_e$. Set $n = 1$.

1: **repeat**
2: Obtain the sample channel $\mathbf{G}^n$.
3: Update $\mathbf{f}^n$ according to (20).
4: Update $\mathbf{e}^n$ according to (24).
5: $n = n + 1$.
6: **until** $||\mathbf{f}^n - \mathbf{f}^{n-1}||_2 \to 0$ and $||\mathbf{e}^n - \mathbf{e}^{n-1}||_2 \to 0$.

1) Convergence analysis: The convergence of Algorithm 1 is analyzed in the following theorem.

Theorem 1 *Assume that Assumptions A and B are satisfied. Then, the sequence of the solutions obtained in each iteration of Algorithm 1 converges to the set of stationary points of the problem in (11) almost surely.*

Proof: See Appendix B. ■

2) Complexity analysis: The computational complexity for updating $\mathbf{f}^n$ and $\mathbf{e}^n$ at each iteration mainly depends on the computation of (20) and (24), respectively. In particular, due to the update rule in $\{\sum_{i=1}^n \alpha_f^i, \sum_{i=1}^n \mathbf{d}_f^i, \sum_{i=1}^n \mathbf{d}_e^i\}$, only $\{\alpha_e^n, \mathbf{d}_f^n, \mathbf{d}_e^n\}$ need to be calculated at the n-th iteration. Therefore, the approximate complexity of each iteration is $\mathcal{O}(4UMN + 12N)$.

IV. Multiuser System

In this section, we consider the general multiuser setup and solve the optimization problem in (8). The min-max problem in (8) is more challenging to tackle as compared with the problem in (9) due to the presence of the max function. To tackle the problem in (8), we extend the SMM method applied to the single user case.

A. Problem Reformulation

We first approximate the probability function in the original formulation of the problem in (8) by still using the smooth function in (10). To this end, we define $f_k(\mathbf{F}, \mathbf{e}|\mathbf{G}) = u\left(\mathbf{e}^{\mathrm{H}}\mathbf{G}_k\mathbf{F}\boldsymbol{\Upsilon}_k\mathbf{F}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{e} + \gamma_k\sigma_k^2\right)$, where $\boldsymbol{\Upsilon}_k$ is a diagonal matrix whose diagonal entries are all equal to γ_k with the exception of the k-th diagonal element that is equal to -1. Therefore, an approximate expression for the objective function in (8) is $\max_{k\in\mathcal{K}} \mathbb{E}\left[f_k(\mathbf{F}, \mathbf{e}|\mathbf{G})\right]$. However, the obtained objective function is still intractable since the maximization operation and the expectation operation make the functions $f_k, \forall k$ and the different channel states coupled, respectively. To circumvent these issues, we use Jensen's inequality

$$\max_{k\in\mathcal{K}} \mathbb{E}\left[f_k(\mathbf{F}, \mathbf{e}|\mathbf{G})\right] \leq \mathbb{E}\left[\max_{k\in\mathcal{K}} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G})\right], \tag{25}$$

since the max function $\max_{k\in\mathcal{K}}\{x_1, \ldots, x_K\}$ is convex [41].

Furthermore, the non-differentiable max function, $\max_{k\in\mathcal{K}} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G})$, is approximated by adopting a smooth log–sum–exp upper-bound [42]

$$\begin{aligned}\max_{k\in\mathcal{K}} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G}) &\approx F(\mathbf{F}, \mathbf{e}|\mathbf{G}) \\ &= \mu \ln\Big(\sum_{k\in\mathcal{K}} \exp\left\{\frac{1}{\mu} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G})\right\}\Big),\end{aligned} \tag{26}$$

where $\mu > 0$ is a smoothing parameter that fulfills the condition

$$\begin{aligned}\max_{k\in\mathcal{K}} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G}) &\leq F(\mathbf{F}, \mathbf{e}|\mathbf{G}) \\ &\leq \max_{k\in\mathcal{K}} f_k(\mathbf{F}, \mathbf{e}|\mathbf{G}) + \frac{1}{\mu}\log(|\mathcal{K}|).\end{aligned} \tag{27}$$

When μ is appropriately chosen, a smooth approximation for the problem in (8) is

$$\min_{\mathbf{F}\in\mathcal{S}_f, \mathbf{e}\in\mathcal{S}_e} G(\mathbf{F}, \mathbf{e}|\mathbf{G}) = \mathbb{E}\left[F(\mathbf{F}, \mathbf{e}|\mathbf{G})\right]. \tag{28}$$

B. Stochastic Successive Convex Approximation Method

Similar to the optimization problem in (11), the optimization problem in (28) may be solved by adopting the SMM method. However, the function $F(\mathbf{F}, \mathbf{e}|\mathbf{G})$ in (26) is more complex and its second-order derivative, which is necessary to construct the upper bound surrogate function of $F(\mathbf{F}, \mathbf{e}|\mathbf{G})$ as shown in Appendix A, is not easy to be calculated. Furthermore, the coefficient of the second-order term in the final upper bound surrogate function of $F(\mathbf{F}, \mathbf{e}|\mathbf{G})$ (α_f^i in (14)) may not be very tight, which eventually results in a very slow convergence rate of the SMM algorithm.

Therefore, we adopt the SSCA method to overcome these issues. The surrogate functions employed by the SSCA method do not need to be upper bound of the original function but they only need to preserve the first-order properties of the original function. Accordingly, the surrogate function, which is denoted by $\hat{F}(\cdot)$, needs to satisfy Assumption B and the following assumptions [43].

Assumption C

$(\mathrm{C1}): \hat{F}\left(\mathbf{x}, \mathbf{x}^{i-1}|\mathbf{G}\right)$ is strongly convex in $\mathbf{x}$ for $\forall \mathbf{x}^{i-1} \in \mathcal{S}_x$.
$(\mathrm{C2}): \hat{F}(\mathbf{x}^{i-1}, \mathbf{x}^{i-1}|\mathbf{G}) = F(\mathbf{x}^{i-1}|\mathbf{G}), \forall \mathbf{x}^{i-1} \in \mathcal{S}_x$.
$(\mathrm{C3}): \nabla_{\mathbf{x}}\hat{F}(\mathbf{x}^{i-1}, \mathbf{x}^{i-1}|\mathbf{G}) = \nabla_{\mathbf{x}}F(\mathbf{x}^{i-1}|\mathbf{G}), \forall \mathbf{x}, \mathbf{x}^{i-1} \in \mathcal{S}_x$,

where $\nabla_{\mathbf{x}}\hat{F}(\cdot)$ denotes the gradient operation applied to complex-valued functions [44]. Assumption C cannot ensure that the sequences of the approximate objective values are monotonically decreasing at each iteration. To guarantee the convergence of the algorithm, however, the variables can be updated, at each iteration, by choosing an appropriate step size that ensures that the objective value decreases. Based on the above assumptions, we choose the proximal gradient-like approximation to construct the surrogate function, which is [45]

$$\begin{aligned}\hat{F}\left(\mathbf{x}, \mathbf{x}^{i-1}|\mathbf{G}\right) =& F\left(\mathbf{x}^{i-1}|\mathbf{G}\right) + \nabla_{\mathbf{x}}F(\mathbf{x}^{i-1}|\mathbf{G})^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^{i-1}) \\ &+ \frac{\tau^i}{2}||\mathbf{x} - \mathbf{x}^{i-1}||^2,\end{aligned} \tag{29}$$

where τ^i can be any positive number.

1) Optimizing $\mathbf{F}$*:* By using (29) and the complex differential formula $\mathrm{d}F(\mathbf{X}, \mathbf{X}^*) = \mathrm{Tr}((\frac{\partial F(\mathbf{X}, \mathbf{X}^*)}{\partial \mathbf{X}})^{\mathrm{T}}\mathrm{d}\mathbf{X} + (\frac{\partial F(\mathbf{X}, \mathbf{X}^*)}{\partial \mathbf{X}^*})^{\mathrm{T}}\mathrm{d}\mathbf{X}^*)$ (Eq. (3.4.55) in [46]), we can construct a surrogate function for $\mathbf{F}$ around $\mathbf{F}^{i-1}$ when $\mathbf{e}$ is given. The surrogate function is given in (30), which is shown at the top of the next page, where the following parameters are introduced

$$\mathbf{P}_f^i = \mathbf{W}_f^i - \frac{\tau^i}{2}\mathbf{F}^{i-1}, \tag{31a}$$

$$\mathbf{W}_f^i = \sum_{k\in\mathcal{K}} l_k^i \mathbf{G}_k^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_k^i\mathbf{F}^{i-1}\boldsymbol{\Upsilon}_k, \tag{31b}$$

$$\begin{aligned}\hat{F}\left(\mathbf{F},\mathbf{F}^{i-1}|\mathbf{G}\right) &= F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)+\mathrm{Tr}\left(\left(\frac{\partial F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)}{\partial \mathbf{F}}\right)^{\mathrm{T}}(\mathbf{F}-\mathbf{F}^{i-1})\right)+\mathrm{Tr}\left(\left(\frac{\partial F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)}{\partial \mathbf{F}^{*}}\right)^{\mathrm{T}}(\mathbf{F}^{*}-\mathbf{F}^{*,i-1})\right)\\ &\quad+\frac{\tau^{i}}{2}||\mathbf{F}-\mathbf{F}^{i-1}||_{F}^{2}\\ &= F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)+2\sum_{k\in\mathcal{K}}l_{k}^{i}\mathrm{Re}\left\{\mathrm{Tr}\left(\boldsymbol{\Upsilon}_{k}\mathbf{F}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_{k}(\mathbf{F}-\mathbf{F}^{i-1})\right)\right\}+\frac{\tau^{i}}{2}||\mathbf{F}-\mathbf{F}^{i-1}||_{F}^{2}\\ &= 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{P}_{f}^{i,\mathrm{H}}\mathbf{F}\right)\right\}+\frac{\tau^{i}}{2}||\mathbf{F}||_{F}^{2}+\mathrm{cons1}^{i}. \end{aligned}\tag{30}$$

$$l_{k}^{i}=\frac{\exp\left\{\frac{1}{\mu}f_{k}\left(\mathbf{F}^{i-1},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)\right\}}{\sum_{k\in\mathcal{K}}\exp\left\{\frac{1}{\mu}f_{k}\left(\mathbf{F}^{i-1},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)\right\}}\frac{\theta e^{-\theta x_{k}^{i}}}{\left(1+e^{-\theta x_{k}^{i}}\right)^{2}},\tag{31c}$$

$$x_{k}^{i}=\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{i}\mathbf{F}^{i-1}\boldsymbol{\Upsilon}_{k}\mathbf{F}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{i,\mathrm{H}}\mathbf{e}^{i-1}+\gamma_{k}\sigma_{k}^{2},\tag{31d}$$

$$\begin{aligned}\mathrm{cons1}^{i} &= F\left(\mathbf{F}^{i-1}|\mathbf{G}^{i}\right)+\frac{\tau^{i}}{2}||\mathbf{F}^{i-1}||_{F}^{2}\\ &\quad-2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{W}_{f}^{i,\mathrm{H}}\mathbf{F}^{i-1}\right)\right\}.\end{aligned}\tag{31e}$$

By using (30), the optimization subproblem in (28) as a function of $\mathbf{F}$ is formulated, at the n-th iteration, as

$$\min_{\mathbf{F}\in\mathcal{S}_{F}}\ \frac{1}{n}\sum_{i=1}^{n}\hat{F}\left(\mathbf{F},\mathbf{F}^{i-1}|\mathbf{G}^{i}\right).\tag{32}$$

The obtained optimization problem in (32) can be solved by using the same methods as for the optimization problem in (16). Specifically, the global minimizer of the optimization problem in (32) is

$$\widehat{\mathbf{F}}^{n}=\begin{cases}\frac{-2}{\sum_{i=1}^{n}\tau^{i}}\sum_{i=1}^{n}\mathbf{P}_{f}^{i}, & \text{if } \frac{4||\sum_{i=1}^{n}\mathbf{P}_{f}^{i}||_{F}^{2}}{(\sum_{i=1}^{n}\tau^{i})^{2}}\leq P_{max},\\ -\sqrt{\frac{P_{max}}{||\sum_{i=1}^{n}\mathbf{P}_{f}^{i}||_{F}^{2}}}\sum_{i=1}^{n}\mathbf{P}_{f}^{i}, & \text{otherwise.}\end{cases}\tag{33}$$

2) Optimizing $\mathbf{e}$*:* Similarly, the optimization subproblem in (28) as a function of $\mathbf{e}$ when $\mathbf{F}$ is given, can be formulated, at the n-th iteration, as

$$\min_{\mathbf{e}\in\mathcal{S}_{e}}\ \frac{1}{n}\sum_{i=1}^{n}\hat{F}\left(\mathbf{e},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right),\tag{34}$$

where $\hat{F}\left(\mathbf{e},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)=2\mathrm{Re}\left\{\mathbf{p}_{e}^{i,\mathrm{H}}\mathbf{e}\right\}+\mathrm{cons2}^{i}$, and

$$\mathbf{p}_{e}^{i}=\mathbf{w}_{e}^{i}-\frac{\tau^{i}}{2}\mathbf{e}^{i-1},\tag{35a}$$

$$\mathbf{w}_{e}^{i}=\sum_{k\in\mathcal{K}}l_{k}\mathbf{G}_{k}\mathbf{F}^{i-1}\boldsymbol{\Upsilon}_{k}\mathbf{F}^{i-1,\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{e}^{i-1},\tag{35b}$$

$$\mathrm{cons2}^{i}=F\left(\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)+\tau^{i}(UM+1)-2\mathrm{Re}\left\{\mathbf{w}_{e}^{i,\mathrm{H}}\mathbf{e}^{i-1}\right\}.\tag{35c}$$

Therefore, the minimizer of the optimization problem in (34) is

$$\widehat{\mathbf{e}}^{n}=\exp\left\{\mathrm{j}\angle\left(\left(\sum_{i=1}^{n}\mathbf{p}_{e}^{i}\right)/\left[\sum_{i=1}^{n}\mathbf{p}_{e}^{i}\right]_{UM+1}\right)\right\}.\tag{36}$$

C. Algorithm Development

The closed-form solutions for $\mathbf{F}$ in (33) and for $\mathbf{e}$ in (36) can greatly reduce the computational complexity. Algorithm 2 summarizes the proposed SSCA-based robust beamforming design for RIS-aided multiuser mmWave systems in which the BS-user links are subject to random blockages. The proposed algorithm is referred to as the SSCA-OutMin algorithm.

Algorithm 2 SSCA-OutMin Algorithm

Initialize: Initialize $\mathbf{F}^{0}\in\mathcal{S}_{f}$ and $\mathbf{e}^{0}\in\mathcal{S}_{e}$. Set $n=0$.
1: **repeat**
2: $n=n+1$.
3: Obtain the sample channel $\mathbf{G}^{n}$.
4: Calculate $\widehat{\mathbf{F}}^{n}$ according to (33).
5: Update $\mathbf{F}^{n}=\mathbf{F}^{n-1}+\xi_{f}^{n}\left(\widehat{\mathbf{F}}^{n}-\mathbf{F}^{n-1}\right)$.
6: Calculate $\widehat{\mathbf{e}}^{n}$ according to (36).
7: Update $\mathbf{e}^{n}=\mathbf{e}^{n-1}+\xi_{e}^{n}\left(\widehat{\mathbf{e}}^{n}-\mathbf{e}^{n-1}\right)$.
8: **until** $||\mathbf{F}^{n}-\mathbf{F}^{n-1}||_{F}^{2}\to 0$ and $||\mathbf{e}^{n}-\mathbf{e}^{n-1}||_{2}\to 0$.

1) Step-size selection: It is worth noting that the approximation in (30) has the same form as that in (14). However, τ^{i} in the SSCA method can be any positive number, and $\hat{F}\left(\mathbf{F},\mathbf{F}^{i-1}|\mathbf{G}\right)$ might not be a global upper bound of $F\left(\mathbf{F}|\mathbf{G}\right)$. To account for this issue, the step sizes ξ_{f}^{n} and ξ_{e}^{n} in Algorithm 2 need to be carefully chosen to ensure convergence.

As an example, let us consider the choice of ξ_{f}^{n} to illustrate the update rule, which is a line-search (also called Armijo step-size) rule. Consider $\xi_{f}^{0}>0$ and $c_{1,f},c_{2,f}\in(0,1)$. Let ξ_{f}^{n} be the largest element in $\{\xi_{f}^{0}c_{2,f}^{t}\}_{t=0,1,\ldots}$ such that

$$\begin{aligned}F\left(\mathbf{F}^{n-1}+\xi_{f}^{n}\left(\widehat{\mathbf{F}}^{n}-\mathbf{F}^{n-1}\right)\right) &\leq F(\mathbf{F}^{n-1})\\ &\quad+c_{1,f}\xi_{f}^{n}\mathrm{Tr}\left(\nabla_{\mathbf{F}}F(\mathbf{F}^{n-1})^{\mathrm{T}}\left(\widehat{\mathbf{F}}^{n}-\mathbf{F}^{n-1}\right)\right).\end{aligned}\tag{37}$$

Theorem 2 *If $\{\xi_{f}^{n}\}_{n=1,2,\ldots}$ is chosen according to the line-search rule, then*

$$\lim_{n\to\infty}||\widehat{\mathbf{F}}^{n}-\mathbf{F}^{n-1}||=0.$$

Proof: See Theorem 7 in [43]. ■

2) Convergence analysis: The convergence of Algorithm 2 is given in the following theorem.

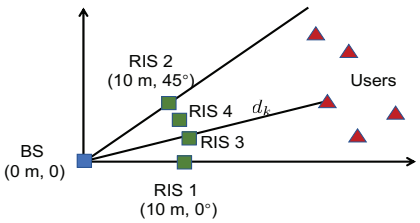


Fig. 2: The simulated system setup.

Theorem 3 *Assume that Assumptions B and C are satisfied. Then, every limit point of the iterations generated by Algorithm 2 is a stationary point of the optimization problem in (28) almost surely.*

Proof: See Appendix C. ■

3) Complexity analysis: The computational complexity for updating $\mathbf{f}^n$ and $\mathbf{e}^n$ at each iteration mainly depends on the computation of (33) and (36), respectively. In particular, only $\{\mathbf{P}_f^n, \mathbf{p}_e^n\}$ needs to be calculated at the n-th iteration. Therefore, the approximate complexity of each iteration is given by $\mathcal{O}((K+2)2UMN+UMK+NK+(N+2)K^2+2N)$.

4) Initial point: The optimization problem in (28) has, in general, multiple local minima due to the non-convex unit-modulus constraint and to $\mathbf{e} \in \mathcal{S}_e$. The accurate selection of the initial points in Algorithm 2 plays an important role for the convergence speed and the quality of the obtained local solution. To that end, we first initialize $\mathbf{e}$ to maximize the minimum equivalent total channel gain, resulting in the following optimization problem

$$\mathbf{e}^0 = \arg\max_{\mathbf{e}\in\mathcal{S}_e} \min_{k\in\mathcal{K}} ||\mathbf{e}^{\mathrm{H}}\mathbf{G}_k^0||_2^2. \tag{38}$$

The optimization problem in (38) can be efficiently solved by using the SDR method as follows

$$\max_{\mathbf{E}} \ t \tag{39a}$$

$$\text{s.t.} \ \ \mathrm{Tr}\{\mathbf{G}_k^0\mathbf{G}_k^{0,\mathrm{H}}\mathbf{E}\} \geq t, \forall k \in \mathcal{K} \tag{39b}$$

$$\mathbf{E} \succeq 0, \mathrm{rank}(\mathbf{E}) = 1, [\mathbf{E}]_{m,m} = 1, \forall m, \tag{39c}$$

where $\mathbf{E} = \mathbf{e}\mathbf{e}^{\mathrm{H}}$ and t is an auxiliary variable.

Furthermore, $\mathbf{F}$ is initialized by using the maximum-ratio transmission (MRT) solution as

$$\mathbf{F}^0 = P_{max}\frac{\mathbf{G}^0\mathbf{e}^0}{||\mathbf{G}^0\mathbf{e}^0||}. \tag{40}$$

V. Numerical results and discussion

A. Simulation setup

In this section, we numerically evaluate the performance of the proposed algorithms. All experiments are performed on a PC with a 1.99 GHz i7-8550U CPU and 16 GB RAM. We adopt the polar coordinates to describe the simulated system setup as shown in Fig. 2. Specifically, the BS is located at (0 m, 0°), and the four RISs are deployed in the locations (10 m, 0°), (10 m, 45°), (10 m, 20°) and (10 m, 30°) which are close to the BS. The users are randomly placed in a region that is identified by the polar diameter $d_k \in$[50 m, 80 m] and the polar angle $\vartheta \in [0, 45°]$, where d_k is used to calculate the distance-dependent blockage probability. The large-scale fading, which corresponds to an urban micro (UMi)-street canyon scenario [36], is $\mathrm{PL} = 32.4 + 20\log_{10}(f_c) + 10\alpha\log_{10}(D) + \xi$ in dB, where D is the link distance (in meters), α is the path loss exponent, and $\xi \sim \mathcal{CN}(0, \sigma^2)$ is the log-normal shadowing where σ^2 denotes the shadowing variance. The mmWave system operates at a carrier frequency $f_c = 28$ GHz and the bandwidth is 20 MHz. Since the macro-scattering environment between the BS and the users is complex, only NLoS clusters are assumed to exist in the BS-user links, i.e., the Rician factor is $\kappa = 0$. The large-scale parameters of the NLoS links are $\alpha = 3.5$ and $\sigma_\xi = 8.2$ dB [36]. In practice, the RISs can be installed such that the BS-RIS links and the RIS-user links are blockage-free. Thus, the channels in (5) and (6) are assumed to have only the LoS cluster with a Rician factor $\kappa \to \infty$. The large-scale parameters of the LoS link are $\alpha = 2$ and $\sigma_\xi = 4$ dB according to [36]. Unless stated otherwise, we assume $L_{\mathrm{b},k} = L_{u,k} = L_{\mathrm{b},u} = 5$ and $I = 20$. The transmit power limit of the BS is $P_{max} = 30$ dBm and the noise power at each user is $\sigma_1^2 = \ldots = \sigma_K^2 = -94$ dBm. For simplicity, we consider an equal blockage probability, $p_{k,l} = p_{\mathrm{block}}, \forall k, l$, and an equal target SINR, $\gamma = \gamma_1 = \ldots = \gamma_K$, which yields the target rate $R_{\mathrm{targ}} = \log_2(1+\gamma)$. The smooth parameters are chosen to be $\theta = \frac{1}{\max_{\forall k\in\mathcal{K}}|x_k^0|}$ and $\mu = \frac{1}{100K}$.

To evaluate the performance of the proposed stochastic optimization algorithms, we consider the following benchmark schemes. 1) Perfect-Instantaneous: Perfect instantaneous CSI is assumed to be known, including the instantaneous channel gains, AoAs, AoDs, and blockage status of the links. This scheme is regarded as the performance upper bound. Specifically, we generate 1000 channel realizations, wherein each path is randomly blocked with probability p_{block}, e.g., each path is, on average, randomly blocked 500 times if $p_{\mathrm{block}} = 0.5$. Then, the beamforming is optimized for each instantaneous and fixed channel realization. 2) NoRIS: In this case, no RIS is employed and the precoding at the BS is obtained by using the SMM or SSCA methods. This scheme is regarded as the performance lower bound. 3) No-robust: In this scheme, the beamforming schemes at the BS and at the RIS are designed by using the SMM or SSCA methods by taking into account the random small-scale parameters while assuming $p_{\mathrm{block}} = 0$. 4) Imperfect CSI: In this scheme, the beamforming schemes at the BS and at the RIS are designed by using the SMM or SSCA methods based on the imperfect knowledge of the central angles of the clusters. Specifically, we assume that the estimation error of the central angles of the clusters is 0.2 degrees, i.e., $\triangle\mathbb{E}\{\varphi(\phi)\} = 0.2$. 5) Quantization-1/2/3 bit: In this scheme, the optimal continuous-valued phase shifts of the RISs are first obtained by using the SMM algorithm and are then quantized with 1 bit or, 2 bit or 3 bit resolution. 6) SAA: In this scheme, we generate 300 independent channel realizations in advance, and the solutions at each iteration of the algorithms are obtained as the average over the 300 channel samples. Also, the surrogate function used at each iteration is obtained by adopting the MM or SCA methods. To be specific, let us consider the beamforming design in the single-user case

TABLE I: Comparison of the CPU time

Algorithms	The CPU time (sec) per iteration	The CPU time (sec)
SMM	0.0025	1.8750
SSCA	0.0042	4.6719
SAA	0.3557	20.9844

as an example. The beamforming designed by using the SAA-MM method is obtained by modifying the problems in (12) and (13) with the following updating rules

$$\mathbf{f}^n = \arg\min_{\mathbf{f}\in\mathcal{S}_f} \frac{1}{300}\sum_{i=1}^{300} \hat{f}\left(\mathbf{f}, \mathbf{f}^{n-1}|\mathbf{G}^i\right), \tag{41}$$

and

$$\mathbf{e}^n = \arg\min_{\mathbf{e}\in\mathcal{S}_e} \frac{1}{300}\sum_{i=1}^{300} \hat{f}\left(\mathbf{e}, \mathbf{e}^{n-1}|\mathbf{G}^i\right). \tag{42}$$

In order to demonstrate the robustness of the proposed algorithms, we consider two performance metrics: the outage probability and the effective rate. In particular, the outage probability of each user is calculated by averaging over 1000 independent channel realizations. The corresponding effective rate of the k-th user is defined as $R_{\text{eff},k} \triangleq \mathbb{E}[\log_2(1+\Gamma_k(\mathbf{F},\mathbf{e}))]$ if $\Gamma_k(\mathbf{F},\mathbf{e}) \geq \gamma$ and $R_{\text{eff}} \triangleq 0$ otherwise.

B. Convergence

Figure 3 illustrates the convergence behavior of the considered stochastic optimization algorithms. For comparison, we consider the single-user case in the presence of RIS 1 depicted in Fig. 2, and the other parameters are given in Fig. 3. In Fig. 3, the y-axis shows the objective value of the problems in (16) or (32), and it is not the actual outage probability of the original problem. It is observed that the SMM and SSCA algorithms are characterized by an oscillatory convergence behavior, which depends on the random channel generations at each iteration. On the other hand, using 300 channel realizations for each iteration leads to the monotonic convergence behavior of the SAA algorithm when adopting a monotonically decreasing surrogate function for each channel realization. Although the SAA algorithm requires the least number of iterations to converge, it is much more computationally demanding than the other two algorithms. This observation is confirmed in Table I, which compares the CPU time consumption of each iteration and the total CPU time consumption for the three considered algorithms. Theoretically, the computational complexity of each iteration of the SAA algorithm is 300 times higher than that of the SMM or SAA algorithms, because each parameter needs to be calculated 300 times for all channel realizations at each iteration of the SAA algorithm.

C. Single-user Case Study

We consider a single-user system where the transmission of data is assisted by the RIS 1 in Fig. 2 and the target rate is $R_{\text{targ}} = 0.1$ bps/Hz. Figure 4 illustrates the performance of different algorithms as a function of the blockage probability. We see that the SMM-based beamforming scheme with $M = 128$ outperforms the NoRIS scheme when $p_{\text{block}} \geq 0.1$.

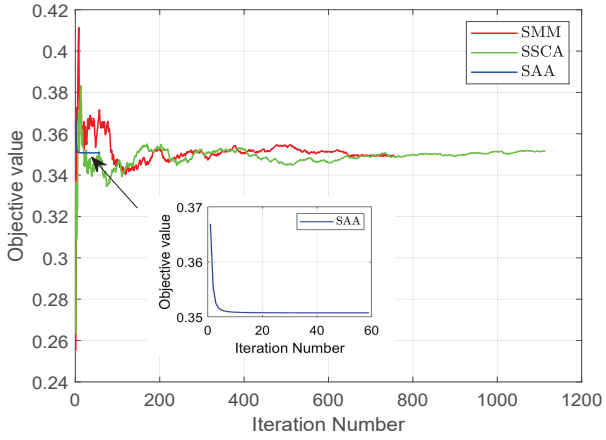
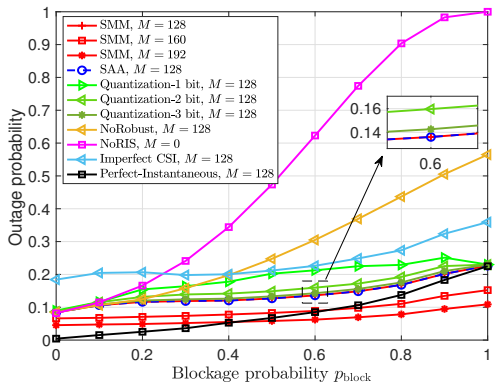


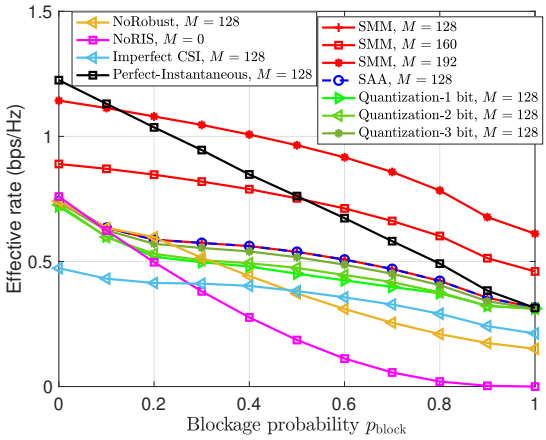
Fig. 3: Convergence behavior of different algorithms, when $N = 8$, $M = 128$, $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.

If $p_{\text{block}} \leq 0.1$, on the other hand, the direct BS-user channel is much stronger than the cascaded BS-RIS-user channel, as the latter is subject to the double path loss law, which dominates the performance for long transmission distances at high frequency bands. When the blockage probability is small ($p_{\text{block}} \leq 0.1$), therefore, the BS tends to allocate the transmit power to the stronger direct path, thus reducing the contribution of the RIS to the system performance. If the number of reflecting elements at the RIS is increased to $M = 160$, the proposed SMM algorithm outperforms the NoRIS system for any value of the blockage probability (i.e., $0 \leq p_{\text{block}} \leq 1$). The reason is that the RIS-aided channel provides a beamforming gain that compensates for the performance loss caused by the presence of blockages even if $p_{\text{block}} = 0$. In addition, we see that the SMM-based beamforming scheme and the SAA-based beamforming scheme offer the same performance. As far as the impact of the phase quantization is concerned, we see that 1-bit resolution has a non-negligible negative impact on the system performance, while 3-bit resolution is sufficient to obtain performance very close to the continuous-valued phase shifts. As expected, the Perfect-Instantaneous scheme outperforms all the other schemes at the price of frequent channel estimations in each channel coherence block. Finally, the proposed robust designs outperform the NoRobust case and the imperfect CSI case, since the impact of random blockages is accounted for at the design stage.

Figure 5 shows the impact of the size of the RIS and the size of the antenna array at the BS on the outage probability. The SMM algorithm is considered. It can be observed from Fig. 5(a) that, when BS is equipped with $N = 8$ antennas, the RIS plays a significant role in guaranteeing the desired user's QoS and in improving the system robustness as the number of reflecting elements increases ($M : 64 \rightarrow 256$). Specifically, a large-size RIS with $M \geq 224$ provides an outage probability smaller than 0.1, even if the direct channel from the BS to the user is blocked with unit probability. A similar trend is observed in Fig. 5(b) as the number of BS antennas is increased while the number of RIS elements is

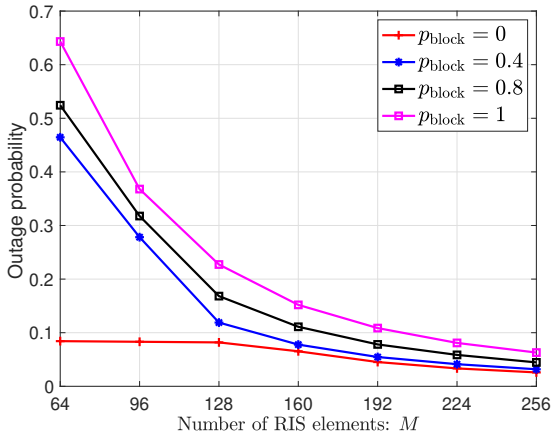


(a) Outage probability

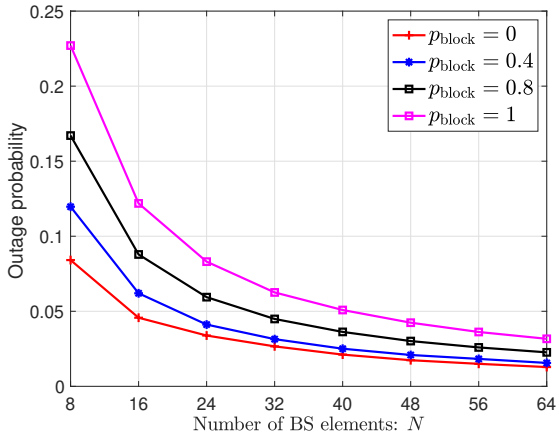


(b) Effective rate

Fig. 4: Comparison of the outage probability and effective rate as a function of the blockage probability p_{block} for $N = 8$, $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.



(a) Number of RIS elements



(b) Number of BS antennas

Fig. 5: Outage probability as a function of M with fixed $N = 8$ and as a function of N with fixed $M = 128$, when $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.

kept fixed and is equal to $M = 128$. The main difference is that the N antennas at the BS require power amplifiers, digital processing units, and multiple RF chains.

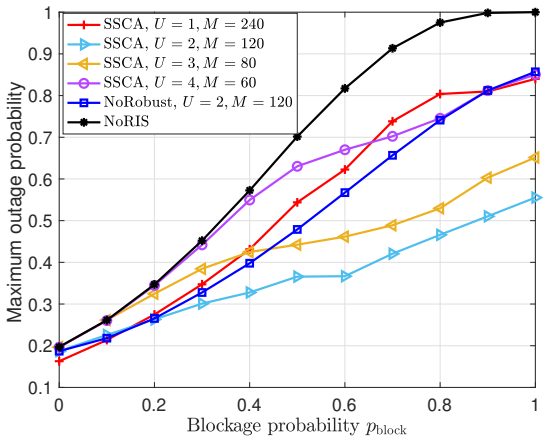
D. Multiuser Case Study

In this section, we analyze a multiuser system with $K = 3$ users, and assume that the target rate is $R_{\text{targ}} = 0.1$ bps/Hz. Multiple RISs are distributed as shown in Fig. 2. For fairness, we keep fixed the total number of RIS elements, i.e., $UM = 240$. From Fig. 6, we conclude that, compared with the single-RIS case, distributing the total number of RIS elements between two RISs significantly improves the system performance in terms of maximum outage probability and minimum effective rate. This is because a better spatial diversity gain is ensured in this case, while ensuring the each RIS has a sufficient number of reflecting elements to compensate for the path loss of the RIS-aided links. If the total number of reflecting elements is distributed among three or four RISs, the system performance is reduced. This is because each RIS cannot compensate the distance-dependent path loss. In a multi-RIS scenario, therefore, the size of each RIS (i.e., the number of reflecting elements) is an optimization parameter that needs to be judiciously chosen. Furthermore, we note that the proposed robust designs significantly outperform the NoRobust and the NoRIS schemes, under similar setups.

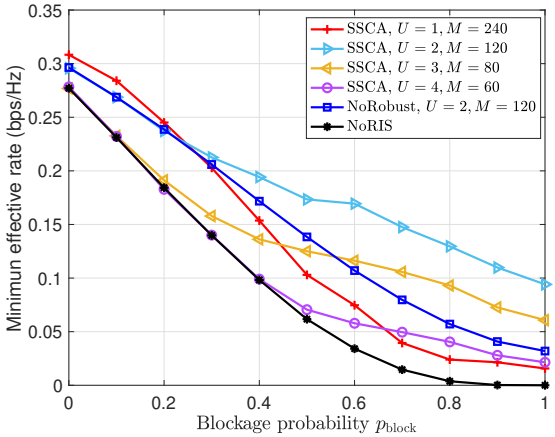
Finally, Fig. 7 illustrates the maximum outage probability as a function of the number of users. For a fair comparison, we consider the setup $N = 16, M = 120$ and $U = 2$. We see that the gain with respect to the NoRIS scheme is almost constant as the number of users K increases. Therefore, the proposed RIS-aided scheme can guarantee the desired QoS performance for the worst-case user even if the number of users increases.

VI. Conclusions

In this work, we have introduced schemes for improving the reliability of a mmWave system in the presence of random blockages by deploying multiple RISs and designing the corresponding robust beamforming. In order to reduce the system outage, we have formulated and solved a maximum outage probability minimization problem which belongs to the family of stochastic optimization problems. More precisely,



(a) Outage probability



(b) Effective rate

Fig. 6: Comparison of the maximum outage probability and minimum effective rate as a function of the blockage probability p_{block} for $N = 16$, $K = 3$, and $R_{\text{targ}} = 0.1$ bps/Hz.

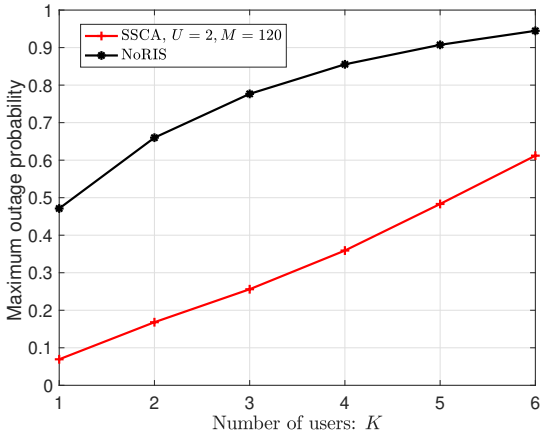


Fig. 7: Comparison of the maximum outage probability as a function of the number of users K for $N = 16$, $p_{\text{block}} = 0.6$, and $R_{\text{targ}} = 0.1$ bps/Hz.

we have introduced robust beamforming schemes at the RIS that depend on the large-scale CSI and the blockage probability. The proposed schemes are obtained by solving complex stochastic optimization problems, for which we have proposed closed-form solutions at each iteration by leveraging the SMM and SSCA optimization methods. The proposed stochastic methods are proved to converge to the set of stationary points of the original stochastic problems. Selected numerical results have demonstrated the performance gains, in terms of outage probability and effective rate, that the proposed schemes can offer when applied to RIS-aided mmWave systems in the presence of random blockages.

Appendix A
The Proof of Lemma 1

In this appendix, $\mathbf{G}^i$ is omitted for simplicity, i.e., $f\left(\mathbf{f}|\mathbf{G}^i\right)$ is replaced by $f\left(\mathbf{f}\right)$. Since $f\left(\mathbf{f}\right)$ is twice differentiable, we propose a second-order approximation to upper bound $f\left(\mathbf{f}\right)$ at any fixed point $\mathbf{f}^{i-1}$

$$\begin{aligned} f\left(\mathbf{f}\right) \leq & \hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right) \\ = & f\left(\mathbf{f}^{i-1}\right) + 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\mathbf{f}-\mathbf{f}^{i-1})\right\} \\ & + (\mathbf{f}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\mathbf{f}-\mathbf{f}^{i-1}), \end{aligned} \tag{43}$$

where $\mathbf{m}_f^i$ and $\mathbf{M}_f^i$ are to be designed to satisfy Assumption A.

Assumptions (A1) and (A2) are readily satisfied. Assumption (A4) is a consistency condition for the first-order directional derivative. Given $\tilde{\mathbf{f}} \in \mathcal{S}_f$, the directional derivative of $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right)$ at $\mathbf{f}^{i-1}$ with direction $\tilde{\mathbf{f}}-\mathbf{f}^{i-1}$ is (from Eq. (3.4.17) in [46])

$$\begin{aligned} & \hat{f}'(\mathbf{f}^{i-1}; \tilde{\mathbf{f}}-\mathbf{f}^{i-1}) \\ = & \left(\frac{\partial \hat{f}}{\partial \mathbf{f}}\bigg|_{\mathbf{f}=\mathbf{f}^{i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}) + \left(\frac{\partial \hat{f}}{\partial \mathbf{f}^*}\bigg|_{\mathbf{f}^*=\mathbf{f}^{*,i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^* \\ = & 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})\right\}. \end{aligned} \tag{44}$$

The corresponding directional derivative of $f\left(\mathbf{f}\right)$ is (from Eq. (3.4.17) in [46])

$$\begin{aligned} & f'(\mathbf{f}^{i-1}; \tilde{\mathbf{f}}-\mathbf{f}^{i-1}) \\ = & \left(\frac{\partial f}{\partial \mathbf{f}}\bigg|_{\mathbf{f}=\mathbf{f}^{i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}) + \left(\frac{\partial f}{\partial \mathbf{f}^*}\bigg|_{\mathbf{f}^*=\mathbf{f}^{*,i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^* \\ = & \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2} 2\mathrm{Re}\left\{\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}\left(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}\right)\right\}, \end{aligned} \tag{45}$$

where x^i is given in (15c).

Assumption (A4) is satisfied only when (44) and (45) are equal, yielding

$$\mathbf{m}_f^i = -\frac{\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}. \tag{46}$$

In order for Assumption (A3) to hold, it is sufficient to show that $\hat{f}(\mathbf{f}, \mathbf{f}^{i-1})$ is an upper bound for each linear cut in any

direction. In particular, defining $\mathbf{f}=\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}), \forall \xi \in [0,1]$, we need to show

$$f\left(\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})\right) \leq f\left(\mathbf{f}^{i-1}\right)+2\xi \mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})\right\} + \xi^2(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}). \tag{47}$$

Let us introduce the functions $L(\xi)=f(\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}))$ and $l(\xi)=\gamma\sigma^2-|\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i(\mathbf{f}^{i-1}+\xi(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1}))|^2$. The inequality in (47) is fulfilled if Assumption (A4) holds, which is ensured by using (46), and if the second-order derivative of $L(\xi)$ is no greater than the second-order derivative in the right hand side of (47) for any value of ξ.

The corresponding sufficient condition can be formulated as

$$\frac{\partial^2 L(\xi)}{\partial \xi^2} \leq 2(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}). \tag{48}$$

The next step is to compute $\partial^2 L(\xi)/\partial\xi^2$. To this end, we first calculate the first-order derivative of $L(\xi)$, as follows

$$\frac{\partial L(\xi)}{\partial \xi}=g(\xi)\frac{\partial l(\xi)}{\partial \xi}, \tag{49}$$

where $g(\xi)=\frac{\theta e^{-\theta l(\xi)}}{(1+e^{-\theta l(\xi)})^2}$, $\frac{\partial l(\xi)}{\partial \xi}=-2\mathrm{Re}\{\mathbf{q}^{\mathrm{H}}(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})\}$, and $\mathbf{q}=\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i(\mathbf{f}^{i-1}+\xi(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1}))$.

Then, the second-order derivative can be formulated as

$$\frac{\partial^2 L(\xi)}{\partial \xi^2}=g(\xi)\frac{\partial^2 l(\xi)}{\partial \xi^2}-\theta g(\xi)\left(\frac{\partial l(\xi)}{\partial \xi}\right)^2 + 2\left(1+e^{-\theta l(\xi)}\right)\left(g(\xi)\frac{\partial l(\xi)}{\partial \xi}\right)^2, \tag{50}$$

where $\frac{\partial^2 l(\xi)}{\partial \xi^2}=-2\mathrm{Re}\{(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{\Theta}(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})\}$ and $\mathbf{\Theta}=\xi\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i$.

Equation (50) can be rewritten as a quadratic form of $\mathbf{t}=\widetilde{\mathbf{f}}-\mathbf{f}^{i-1}$, as follows

$$\frac{\partial^2 L(\xi)}{\partial \xi^2}=\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}^{\mathrm{H}}\mathbf{\Phi}\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}, \tag{51}$$

where

$$\mathbf{\Phi}=g(\xi)\left(2\left(1+e^{-\theta l(\xi)}\right)g(\xi)-\theta\right)\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}^{\mathrm{H}}- g(\xi)\mathbf{I}_2\otimes\mathbf{\Theta}. \tag{52}$$

Furthermore, (48) can be reformulated in a form similar to (51), as follows

$$\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}^{\mathrm{H}}\begin{bmatrix}\mathbf{I}\otimes\mathbf{M}_f^i & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_f^{i,\mathrm{T}}\end{bmatrix}\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}. \tag{53}$$

Combining (51) and (53), the sufficient condition in (48) is equivalent to

$$\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}^{\mathrm{H}}\mathbf{\Phi}\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}\leq\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix}^{\mathrm{H}}\begin{bmatrix}\mathbf{I}\otimes\mathbf{M}_f^i & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_f^{i,\mathrm{T}}\end{bmatrix}\begin{bmatrix}\mathbf{t}\\ \mathbf{t}^*\end{bmatrix},$$

which is satisfied when $\mathbf{M}_f^i$ is chosen so as to fulfill the condition

$$\mathbf{\Phi}\preceq\begin{bmatrix}\mathbf{I}\otimes\mathbf{M}_f^i & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_f^{i,\mathrm{T}}\end{bmatrix}.$$

A convenient choice that fulfills this condition is $\mathbf{M}_f^i=\alpha_f^i\mathbf{I}=\lambda_{\max}(\mathbf{\Phi})\mathbf{I}$. Then, $\hat{f}\left(\mathbf{f},\mathbf{f}^{i-1}\right)$ in (43) can be formulated as follows

$$\begin{aligned}&\hat{f}\left(\mathbf{f},\mathbf{f}^{i-1}\right)\\ &=f\left(\mathbf{f}^{i-1}\right)+2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\mathbf{f}-\mathbf{f}^{i-1})\right\}+\alpha_f^i||\mathbf{f}-\mathbf{f}^{i-1}||_2^2\\ &=2\mathrm{Re}\left\{\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\}+\alpha_f^i||\mathbf{f}||_2^2+\mathrm{const}_f^i,\end{aligned}$$

where $\mathbf{d}_f^{i,\mathrm{H}}$, α_f^i and const_f^i are defined in Lemma 1. The deterministic expression of $\lambda_{\max}(\mathbf{\Phi})$ is difficult to obtain, therefore we derive the upper bound as follows

$$\begin{aligned}&\lambda_{\max}(\mathbf{\Phi})\\ &\overset{(p1)}{\leq} 2\left(1+e^{-\theta l(\xi)}\right)g^2(\xi)\lambda_{\max}\left(\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}^{\mathrm{H}}\right)\\ &\quad-g(\xi)\lambda_{\min}(\mathbf{I}_2\otimes\mathbf{\Theta})-\theta g(\xi)\lambda_{\min}\left(\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}\begin{bmatrix}\mathbf{q}\\ \mathbf{q}^*\end{bmatrix}^{\mathrm{H}}\right)\\ &\overset{(p2)}{=} 4\left(1+e^{-\theta l(\xi)}\right)g^2(\xi)||\mathbf{q}||_2^2\\ &\overset{(p3)}{<} \frac{\theta^2}{2}||\mathbf{q}||_2^2\\ &\overset{(p4)}{\leq} \frac{\theta^2}{2}\lambda_{\max}(\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i)\cdot||\mathbf{f}^{i-1}\\ &\quad+\gamma(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})||_2^2\\ &\overset{(p5)}{\leq} \frac{\theta^2}{2}P_{max}|\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}|^2.\end{aligned}$$

where the inequalities are obtained by departing from (52) and by invoking the following properties.

(p1): If $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices, then $\lambda_{\max}(\mathbf{A})+\lambda_{\max}(\mathbf{B})\geq\lambda_{\max}(\mathbf{A}+\mathbf{B})$ [47].

(p2): If $\mathbf{A}$ is rank one, then $\lambda_{\max}(\mathbf{A})=\mathrm{Tr}[\mathbf{A}], \lambda_{\min}(\mathbf{A})=0$ [47].

(p3): $\left(1+e^{-\theta l(\xi)}\right)g^2(\xi)\leq\theta^2/8$, where the equality holds when $l(\xi)=0$.

(p4): If $\mathbf{A}$ is positive semidefinite with maximum eigenvalue $\lambda_{\max}(\mathbf{A})$ and $\mathbf{B}$ is positive semidefinite, then $\mathrm{Tr}[\mathbf{AB}]\leq\lambda_{\max}(\mathbf{A})\mathrm{Tr}[\mathbf{B}]$ [47].

(p5): The power constraint $||\mathbf{f}^{i-1}+\gamma(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})||_2^2\leq P_{max}$ needs to be fulfilled.

Hence, the proof is completed.

Appendix B
The proof of Theorem 1

Define the random functions

$$g^n(\mathbf{x})=\frac{1}{n}\sum_{i=1}^n f\left(\mathbf{x}|\mathbf{G}^i\right), \tag{54}$$

$$\hat{g}^n(\mathbf{x})=\frac{1}{n}\sum_{i=1}^n \hat{f}\left(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G}^i\right). \tag{55}$$

To prove the convergence of Algorithm 1, we use the following lemmas.

Lemma 3 *Assume that Assumptions A and B are satisfied and define a limit point $\bar{\mathbf{x}}$ of the subsequence $\{\mathbf{x}^{n_j}\}_{j=1}^{\infty}$. Then, there exists uniformly continuous functions $g(\mathbf{x})$ and $\hat{g}(\mathbf{x})$ such that*

$$g(\mathbf{x}) = \lim_{n\to\infty} g^n(\mathbf{x}) = \mathbb{E}\left[f\left(\mathbf{x}|\mathbf{G}\right)\right], \forall \mathbf{x} \in \mathcal{S}_x, \tag{56}$$

$$g(\bar{\mathbf{x}}) = \lim_{j\to\infty} g^{n_j}(\mathbf{x}^{n_j}), \tag{57}$$

$$\hat{g}(\mathbf{x}) = \lim_{j\to\infty} \hat{g}^{n_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x, \tag{58}$$

$$\hat{g}(\bar{\mathbf{x}}) = \lim_{j\to\infty} \hat{g}^{n_j}(\mathbf{x}^{n_j}). \tag{59}$$

Proof: First, $f\left(\mathbf{x}, \mathbf{G}\right)$ is bounded for $\forall \mathbf{x} \in \mathcal{S}_x$ and for all channel realizations due to the Assumption (B2). Therefore, (56) holds by using the strong law of large numbers [48]. Also, the families of functions $\{g^{n_j}(\mathbf{x})\}$ are equicontinuous and bounded over the compact set $\mathcal{S}_x$ due to the Assumption (B2) and the use of the mean value theorem. Thus, by restricting to a subsequence, we have (57). Furthermore, the families of functions $\{\hat{g}^n(\mathbf{x})\}$ are also equicontinuous and bounded over the compact set $\mathcal{S}_x$ due to the Assumption (B2) and because $||\nabla_{\mathbf{x}} \hat{f}(\mathbf{x}, \mathbf{x}^{i-1}, \mathbf{G})||$ is bounded. Hence, the Arzelà-Ascoli theorem [49] implies that, by restricting to a subsequence, there exists a uniformly continuous function $\hat{g}(\mathbf{x})$ such that (58) and (59) hold. ■

In addition, the update rule of Algorithm 1 leads the following lemma.

Lemma 4 $\lim_{n\to\infty} |\hat{g}^n(\mathbf{x}^n) - g^n(\mathbf{x}^n)| = 0$*, almost surely.*

Proof: The proof of Lemma 4 is the same as the proof of (Lemma 1 in [40]) and is hence omitted for conciseness. ■

Assumption (A3) implies that $\hat{g}^{n_j}(\mathbf{x}) \geq g^{n_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which combined with (56) and (58) leads to

$$\hat{g}(\mathbf{x}) \geq g(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x. \tag{60}$$

Moreover, combining Lemma 4 with (57) and (59) we have

$$\hat{g}(\bar{\mathbf{x}}) = g(\bar{\mathbf{x}}). \tag{61}$$

Then, (60) and (61) imply that $\bar{\mathbf{x}}$ is a minimizer of the function $\hat{g}(\mathbf{x}) - g(\mathbf{x})$. Hence the first-order optimality condition is satisfied

$$\nabla \hat{g}(\bar{\mathbf{x}}) - \nabla g(\bar{\mathbf{x}}) = 0. \tag{62}$$

Due to the fact that $\bar{\mathbf{x}}$ is the limit point of the problem in (12) or the problem in (13), we have $\hat{g}(\bar{\mathbf{x}}) \leq \hat{g}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which implies

$$\langle \nabla \hat{g}(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{63}$$

Combining (63) with (62), we obtain

$$\langle \nabla g(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x, \tag{64}$$

which means that the directional derivative of the objective function $g(\mathbf{x})$ is non-negative for every feasible direction at $\bar{\mathbf{x}}$. Recalling that $\mathbf{x} \in \{\mathbf{f}, \mathbf{e}\}$ and defining the limit points $\{\bar{\mathbf{f}}, \bar{\mathbf{e}}\}$, (64) is equivalent to

$$\begin{cases} \left\langle \nabla g(\bar{\mathbf{f}}), \mathbf{f} - \bar{\mathbf{f}} \right\rangle \geq 0, \forall \mathbf{f} \in \mathcal{S}_f, \\ \langle \nabla g(\bar{\mathbf{e}}), \mathbf{e} - \bar{\mathbf{e}} \rangle \geq 0, \forall \mathbf{e} \in \mathcal{S}_e. \end{cases}$$

Therefore, according to [43], $\{\bar{\mathbf{f}}, \bar{\mathbf{e}}\}$ is a stationary point of the problem in (11) due to the regularity of $g(\cdot)$.

APPENDIX C
PROOF OF THEOREM 1

Define the random functions

$$G^n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} F\left(\mathbf{x}|\mathbf{G}^i\right), \tag{65}$$

$$\hat{G}^n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \hat{F}\left(\mathbf{x}, \mathbf{x}^{i-1}|\mathbf{G}^i\right). \tag{66}$$

To analyze the convergence, we need the following lemmas.

Lemma 5 *Assume that Assumptions B and C are satisfied and define a limit point $\bar{\mathbf{x}}$ for the subsequence $\{\mathbf{x}^{n_j}\}_{j=1}^{\infty}$. Then, there exists uniformly continuous functions $G(\mathbf{x})$ and $\hat{G}(\mathbf{x})$ such that*

$$G(\mathbf{x}) = \lim_{n\to\infty} G^n(\mathbf{x}) = \mathbb{E}\left[F\left(\mathbf{x}|\mathbf{G}\right)\right], \forall \mathbf{x} \in \mathcal{S}_x, \tag{67}$$

$$G(\bar{\mathbf{x}}) = \lim_{j\to\infty} G^{n_j}(\mathbf{x}^{n_j}), \tag{68}$$

$$\hat{G}(\mathbf{x}) = \lim_{j\to\infty} \hat{G}^{n_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x, \tag{69}$$

$$\hat{G}(\bar{\mathbf{x}}) = \lim_{j\to\infty} \hat{G}^{n_j}(\mathbf{x}^{n_j}). \tag{70}$$

Proof: The proof of Lemma 5 is the same as the proof of Lemma 3 and is omitted for brevity. ■

Furthermore, $\mathbf{x}^{n_j}$ is the minimizer of $\hat{G}^{n_j}(\mathbf{x})$, which implies

$$\hat{G}^{n_j}(\mathbf{x}^{n_j}) \leq \hat{G}^{n_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x. \tag{71}$$

Assuming $j \to \infty$, and combining (69) and (70), we obtain $\hat{G}(\bar{\mathbf{x}}) \leq \hat{G}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which implies that its first-order optimality condition is satisfied

$$\left\langle \nabla \hat{G}(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \right\rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{72}$$

By combining (72) and Assumption (C3), we finally obtain

$$\langle \nabla G(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{73}$$

Since $\mathbf{x} \in \{\mathbf{F}, \mathbf{e}\}$, we define the limit points $\{\bar{\mathbf{F}}, \bar{\mathbf{e}}\}$ and (73) is then equivalent to

$$\begin{cases} \left\langle \nabla G(\bar{\mathbf{F}}), \mathbf{F} - \bar{\mathbf{F}} \right\rangle \geq 0, \forall \mathbf{F} \in \mathcal{S}_f, \\ \langle \nabla G(\bar{\mathbf{e}}), \mathbf{e} - \bar{\mathbf{e}} \rangle \geq 0, \forall \mathbf{e} \in \mathcal{S}_e. \end{cases}$$

Therefore, according to [43], $\{\bar{\mathbf{F}}, \bar{\mathbf{e}}\}$ is a stationary point of the problem in (28) due to the regularity of $G(\cdot)$.

REFERENCES

[1] G. Zhou, C. Pan, H. Ren, K. Wang, and K. K. Chai, "RIS-aided mmwave transmission: A stochastic majorization-minimization approach," in *ICC 2021 - IEEE International Conference on Communications*, Jun. 2021, pp. 1–6.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: it will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] V. Raghavan *et al.*, "Spatio-temporal impact of hand and body blockage for millimeter-wave user equipment design at 28 GHz," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 46–52, Dec. 2018.

[4] V. Raghavan, M.-L. Chi, M. A. Tassoudji, O. H. Koymen, and J. Li, "Antenna placement and performance tradeoffs with hand blockage in millimeter wave systems," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3082–3096, Apr. 2019.

[5] 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.

[6] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sept. 2014.

[7] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[8] T. Nishio *et al.*, "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, Nov. 2019.

[9] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sept. 2020.

[10] 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.

[11] H. Iimori *et al.*, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, Sept. 2020.

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

[13] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, May 2019.

[14] ——, "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.

[15] 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.

[16] X. Yu, D. Xu, D. W. K. Ng, and R. Schober, "IRS-assisted green communication systems: Provable convergence and robust optimization," *IEEE Trans. Commun.,*, vol. 69, no. 9, pp. 6313–6329, Sept. 2021.

[17] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, Jun. 2021.

[18] 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.

[19] 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.

[20] A. Abrardo, D. Dardari, and M. Di Renzo, "Intelligent reflecting surfaces: Sum-rate optimization based on statistical CSI," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7121–7136, Oct. 2021.

[21] G. Zhou, C. Pan, H. Ren *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.

[22] K. Zhi, C. Pan, H. Ren, and K. Wang, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," Oct. 2020. [Online]. Available: https://arxiv.org/abs/2010.13525

[23] C. Pan, H. Ren, K. Wang *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[24] 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.

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

[26] P. Mursia, V. Sciancalepore, A. Garcia-Saavedra, L. Cottatellucci, X. Costa-Pérez, and D. Gesbert, "RISMA: Reconfigurable intelligent surfaces enabling beamforming for IoT massive access," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1072–1085, Apr. 2021.

[27] B. Di, H. Zhang, L. Li, L. Song, Y. Li, and Z. Han, "Practical hybrid beamforming with finite-resolution phase shifters for reconfigurable intelligent surface based multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4565–4570, Apr. 2020.

[28] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Tech.*, vol. 69, no. 12, pp. 14 960–14 973, Dec. 2020.

[29] N. S. Perović, M. Di Renzo, and M. F. Flanagan, "Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmWave environments," in *IEEE ICC 2020*, pp. 1–7.

[30] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Sum rate maximization for intelligent reflecting surface assisted terahertz communications," *IEEE Trans. Veh. Tech., early access*, pp. 1–1, 2022.

[31] B. Ning, Z. Chen, W. Chen, and Y. Du, "Channel estimation and transmission for intelligent reflecting surface assisted THz communications," in *IEEE ICC 2020*, pp. 1–7.

[32] 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.

[33] T. Nguyen, D. Nguyen, M. Di Renzo, and R. Zhang, "Leveraging secondary reflections and mitigating interference in multi-IRS/RIS aided wireless network," 2021. [Online]. Available: https://doi.org/10.36227/techrxiv.17140004.v1

[34] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[35] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Wireless Communications. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[36] *Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 16)*. document 3GPP TR 38.901 V16.1.0, Dec. 2019.

[37] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sept. 2015.

[38] C. Pan, G. Zhou, K. Zhi *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," 2021. [Online]. Available: https://arxiv.org/abs/2112.05989

[39] J. Mairal, "Stochastic majorization-minimization algorithms for largescale optimization," *Advances in Neural Information Processing Systems*, pp. 2283–2291, 2013.

[40] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Springer Verlag New York*, pp. 515–545, 2016.

[41] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ. Press, 2004.

[42] S. Xu, "Smoothing method for minimax problems," *Comput. Optim. Appl.*, vol. 20, no. 3, pp. 267–279, 2001.

[43] M. Razaviyayn, *Successive convex approximation: Analysis and applications*. Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, USA, 2014.

[44] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, 2011.

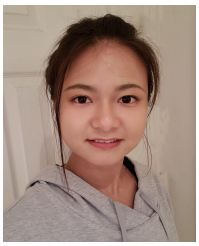
[45] G. Scutari and Y. Sun, "Parallel and distributed successive convex approximation methods for big-data optimization," *Multi-Agent Optim.: Cetraro Italy 2014, Berlin, Germany: Springer*, pp. 141–308, 2018.

[46] X.-D. Zhang, *Matrix analysis and applications*. Cambridge Univ. Press, 2017.

[47] P. Maher and H. Luthepohl, "Handbook of matrices," *The Math. Gaz.*, vol. 83, no. 498, p. 557, 1999.

[48] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*. Birkhuser, Boston, 1996.

[49] N. Dunford and J. Schwartz, *Linear operators. Part 1: General theory*. Interscience Publications, New York, 1958.



Gui Zhou (Graduate Student Member, IEEE) received the B.S. and M.E. degrees from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2015 and 2019, respectively. She is currently pursuing the Ph.D. degree at the School of electronic Engineering and Computer Science, Queen Mary University of London, U.K. Her major research interests include intelligent reflection surface (IRS) and signal processing.

Cunhua Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a Research Associate at the University of Kent, U.K. He held a post-doctoral position at Queen Mary University of London, U.K., from 2016 and 2019. From 2019 to 2021, he was a Lecturer in the same university. Since 2021 he is a full professor at Southeast University. His research interests include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultra-reliable low latency communication (URLLC), machine learning, UAVs, Internet of Things, and mobile edge computing. He has published over 120 IEEE journal papers. He is currently an Editor of IEEE Wireless Communication Letters, IEEE Communications Letters and IEEE ACCESS. He serves as the guest editor for IEEE Journal on Selected Areas in Communications on the special issue on xURLLC in 6G: Next Generation Ultra-Reliable and Low-Latency Communications. He also serves as the Lead Guest Editor of the IEEE Journal of Selected Topics in Signal Processing (JSTSP) Special Issue on Advanced Signal Processing for Reconfigurable Intelligent Surface-Aided 6G Networks, Lead Guest Editor of the IEEE Vehicular Technology Magazine for the special issue on Backscatter and Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G, Lead Guest Editor of IEEE Open Journal of Vehicular Technology for the special issue of Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G and Beyond, and Lead Guest Editor for the IEEE ACCESS Special Issue on Reconfigurable Intelligent Surface Aided Communications for 6G and Beyond. He is Workshop organizer for IEEE ICCC 2021 on the topic of Reconfigurable Intelligent Surfaces for Next Generation Wireless Communications (RIS for 6G Networks), and workshop organizer for IEEE Globecom 2021 on the topic of Reconfigurable Intelligent Surfaces for Future Wireless Communications. He is currently the Workshops and Symposia officer for the Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He is Workshop Chair for IEEE WCNC 2024. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019.

Hong Ren (Member, IEEE) received the B.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2014 and 2018, respectively. From 2016 to 2018, she was a Visiting Student with the School of Electronics and Computer Science, University of Southampton, U.K. From 2018 to 2020, she was a Post-Doctoral Scholar with Queen Mary University of London, U.K. She is currently an associate professor with Southeast University. Her research interests lie in the areas of communication and signal processing, including ultra-low latency and high reliable communications, Massive MIMO and machine learning.

Kezhi Wang (Member, IEEE) received his B.E. and M.E. degrees in School of Automation from Chongqing University, China, in 2008 and 2011, respectively. He received his Ph.D. degree in Engineering from the University of Warwick, U.K. in 2015. He was a Senior Research Officer in University of Essex, U.K. Currently he is a Senior Lecturer with Department of Computer and Information Sciences at Northumbria University, U.K. His research interests include wireless communications and machine learning.

Marco Di Renzo (Fellow, IEEE) received the Laurea (cum laude) and Ph.D. degrees in electrical engineering from the University of L?Aquila, Italy, in 2003 and 2007, respectively, and the Habilitation à Diriger des Recherches (Doctor of Science) degree from University Paris-Sud (now Paris-Saclay University), France, in 2013. Since 2010, he has been with the French National Center for Scientific Research (CNRS), where he is a CNRS Research Director (Professor) with the Laboratory of Signals and Systems (L2S) of Paris-Saclay University ? CNRS and CentraleSupelec, Paris, France. In Paris-Saclay University, he serves as the Coordinator of the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme, and as a Member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies. He is the Editor-in-Chief of IEEE Communications Letters and a Distinguished Speaker of the IEEE Vehicular Technology Society. In 2017-2020, he was a Distinguished Lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society. He has received several research distinctions, which include the SEE-IEEE Alain Glavieux Award, the IEEE Jack Neubauer Memorial Best Systems Paper Award, the Royal Academy of Engineering Distinguished Visiting Fellowship, the Nokia Foundation Visiting Professorship, the Fulbright Fellowship, and the 2021 EURASIP Journal on Wireless Communications and Networking Best Paper Award. He is a Fellow of the UK Institution of Engineering and Technology (IET), a Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA), an Ordinary Member of the European Academy of Sciences and Arts (EASA), and an Ordinary Member of the Academia Europaea (AE). Also, he is a Highly Cited Researcher.