Transceiver Optimization for Wireless Powered Time-Division Duplex MU-MIMO Systems: Non-Robust and Robust Designs

Bin Li, *Senior Member, IEEE*, Meiying Zhang, Yue Rong, *Senior Member, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*— Wireless powered communication (WPC) has been considered as one of the key technologies in the Internet of Things (IoT) applications. In this paper, we study a wireless powered time-division duplex (TDD) multiuser multiple-input multiple-output (MU-MIMO) system, where the base station (BS) has its own power supply and all users can harvest radio frequency (RF) energy from the BS. We aim to maximize the users' information rates by jointly optimizing the duration of users' time slots and the signal covariance matrices of the BS and users. Different to the commonly used sum rate and max-min rate criteria, the proportional fairness of users' rates is considered in the objective function. We first study the ideal case with the perfect channel state information (CSI), and show that the non-convex proportionally fair rate optimization problem can be transformed into an equivalent convex optimization problem. Then we consider practical systems with imperfect CSI, where the CSI mismatch follows a Gaussian distribution. A chance-constrained robust system design is proposed for this scenario, where the Bernstein inequality is applied to convert the chance constraints into the convex constraints. Finally, we consider a more general case where only partial knowledge of the CSI mismatch is available. In this case, the conditional value-at-risk (CVaR) method is applied to solve the distributionally robust system rate optimization problem. Simulation results are presented to show the effectiveness of the proposed algorithms.**

***Index Terms*— Chance constraints, distributionally robust optimization, energy harvesting, MIMO, wireless powered communications (WPC).**

Manuscript received January 22, 2021; revised July 20, 2021 and September 27, 2021; accepted November 27, 2021. Date of publication December 7, 2021; date of current version June 10, 2022. This work was supported in part by the Natural Science Foundation of China under Grant 62071317, in part by the NSF under Grant CNS-2128368 and Grant CNS-2107216, in part by Toyota, and in part by Amazon. The associate editor coordinating the review of this article and approving it for publication was M. Wang. *(Corresponding author: Yue Rong.)*

Bin Li is with the School of Aeronautics and Astronautics, Sichuan University, Chengdu 610207, China (e-mail: bin.li@scu.edu.cn).

Meiying Zhang is with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China (e-mail: mei.ying.zhang@hotmail.com).

Yue Rong is with the School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth, WA 6845, Australia (e-mail: y.rong@curtin.edu.au).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3131595.

Digital Object Identifier 10.1109/TWC.2021.3131595

I. Introduction

WIRELESS powered communication (WPC) has been intensively discussed in the application of wireless sensor networks (WSNs) [1]–[5]. This is because WSNs are energy-constrained networks, and the traditional solution – replacing batteries – is usually associated with a high cost, which hinders its applications in the age of Internet-of-Things (IoT). In addition, natural resources (e.g. solar and wind) are difficult to control, and hence they are difficult to be implemented in real world applications [1]. In contrast, WPC and its counterpart – simultaneous wireless information and energy transfer (SWIET) has been shown to be a promising and reliable alternative due to its flexibility in power transfer.

A. Literature Review

The idea of SWIET or WPC was first introduced in [3] and it has attracted more attention by further considering the practical structure of the receivers [4], [5]. WPC has been integrated with many other technologies, for example, the multiple-input multiple-output (MIMO) technology [5], [6], the relay communications technology [1], [2], [7], [8], the non-orthogonal multiple access (NOMA) technology [9], [10], the massive MIMO technology [11], [12], and wireless networks [13], [14]. Considering the nonlinearity of the power harvesting circuits, recent research has been shifted to the WPC technology with a nonlinear energy harvesting model [15]–[20].

The channel state information (CSI) is important for the application of the WPC technology, since it is required for both the energy and information transmission. However, the exact CSI is usually unavailable, and hence robust design for WPC systems has been intensively studied. In the literature, the existing works in this area can be mainly classified into two categories in terms of the modeling of the CSI mismatch: the worst-case robust design [21]–[26] and the Gaussian robust design [27]–[31].

The worst-case robust design is based on the robust optimization theory, where the CSI mismatch is norm-bounded, and the robust optimization is adopted to optimize the worst-case scenario within the norm-bounded region. The semi-definite relaxation (SDR) technique is usually utilized to

obtain a rank-one relaxed solution and in some occasions the relaxed solution can be verified as rank-one optimal. For example, in [21], a joint time allocation and power control scheme was proposed for multi-user MIMO systems. Two convex optimization problems were formulated by considering different performance measures, which were achieved by using one-dimensional search together with the SDR technique. In [22], the robust design for a backscatter relay communication system was considered, and the SDR technique was applied with the alternating optimization scheme to handle the non-convexity. Secure communication of a wireless powered amplify-and-forward (AF) relay network was investigated in [23], and a two-level optimization approach that involves a one-dimensional search and the SDR technique was proposed. The robust beamforming of a multi-user MIMO system was studied in [24], where the SDR technique was used and the relaxation was proven to be tight. In [25], the secure communication of a wireless powered multi-user multiple-input single-output (MISO) downlink system was considered, where the SDR technique and the S-procedure were applied to convert the max-min constraints, caused by the norm-bounded CSI mismatch, into tractable alternatives. The robust transmission for a secure jamming-aided wireless powered MISO system was investigated in [26]. By using alternating optimization together with the SDR technique and the S-procedure, a locally solution was obtained.

With the Gaussian robust design, the CSI mismatch is assumed to be subject to the Gaussian distribution. The performance requirements are usually formulated as quality-of-service (QoS) probability constraints, which are converted into deterministic constraints, for example, by using the Bernstein inequality method. In [27], secure communication for a wireless powered cognitive MISO system was studied, and the CSI mismatch followed a probabilistic error model. A suboptimal beamforming solution was obtained by applying the Bernstein inequality. A robust cooperative NOMA scheme was proposed in [28] for a WPC system. The formulated probability constrained optimization problem was solved by using the Bernstein inequality method and a 2-D exhaustive search, and a lower complexity algorithm was also developed with the aid of successive convex approximation (SCA). For solving a distributed coordinated beamforming of an artificial noise aided secure WPC system in [29], a solution scheme was proposed by using the alternating direction multiplier method (ADMM) and the Bernstein inequality method. In [30], a robust secure beamforming design for a wireless powered MISO broadcast system was proposed. The probability constraints were transformed into deterministic ones by applying the Bernstein inequality, and a tractable second-order cone programming (SOCP) was then obtained based on the SCA method. A robust cooperative NOMA transmission scheme was developed for a wireless powered decode-and-forward (DF) relay communications system in [31]. The probability constraints were tackled by the Bernstein inequality, and then the golden section search was implemented for handling the nonlinearity introduced by the power splitting factor.

B. Contributions

In this paper, the transceiver optimization for a time-division duplex (TDD) based wireless powered multiuser MIMO (MU-MIMO) system is investigated. Different to the commonly used sum rate and max-min rate criteria, the proportional fairness in terms of the achievable data rates among users is considered in the objective function, which provides an improved fairness among users compared with the sum rate criterion and achieves a higher sum rate than the max-min rate criterion. Both the non-robust design and the robust design are considered. According to the best knowledge of the authors, this topic has not been studied by existing works.

We first study the ideal case with the perfect CSI knowledge, and show that the non-convex proportionally fair rate optimization problem can be transformed into an equivalent convex optimization problem. Then we consider the practical systems with imperfect CSI and investigate the robust transceiver design problems. Two different types of robust designs are proposed. The first one is based on the assumption that the mismatch between the true and estimated CSI follows a Gaussian distribution. This assumption is justified by the fact that for MIMO communication systems with additive white Gaussian noise, when a maximum-likelihood estimator is used and the channel training sequence is orthogonal, the CSI mismatch follows uncorrelated Gaussian distribution [32]. A chance-constrained robust system design is proposed for this scenario, and the Bernstein inequality is applied to convert the chance constraints into convex constraints.

In the second robust design, we consider a more general case where only partial knowledge of the CSI mismatch is available. Different from the Gaussian robust design, in this case, we do not assume that the CSI mismatch follows a certain distribution. In contrast, only the mean and the variance of the CSI mismatch are required, which is closer to real world applications. Estimates of the mean and covariance of the CSI error can be obtained from accumulated channel estimates at the base station (BS). Note that it is more convenient and easier to have the statistics rather than the accurate distribution of the CSI error, which requires the estimation of a larger number of parameters to fit the distribution. Inspired by the idea of distributionally robust optimization, the conditional value-at-risk (CVaR) method is applied to solve the distributionally robust system rate optimization problem. Simulation results are presented to show the effectiveness of the proposed algorithms.

The main contributions of this paper are summarized as follows:

- A tractable transceiver design with proportional fairness among users is developed for a TDD based wireless powered MU-MIMO system. Simulation results show that the proposed proportional fairness based system design has a better tradeoff between the sum data rate and the fairness among users than the sum rate and max-min rate based designs.
- We propose a chance-constrained robust transceiver design when the CSI mismatch follows Gaussian distribution and transform the chance constraints into convex

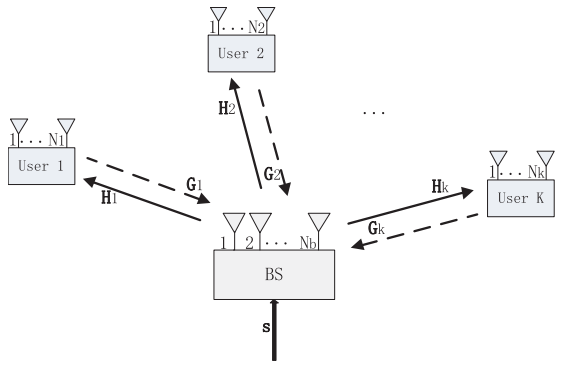


Fig. 1. An MU-MIMO communication system with energy-harvesting users.

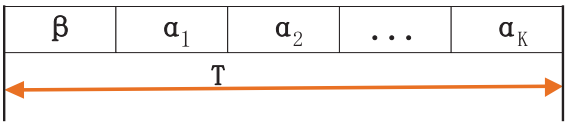


Fig. 2. TDD-based communication time slots allocation for BS and users.

conic constraints. The robust design has a higher system rate than the non-robust design.

- A distributionally robust design is proposed which only requires the mean and variance of the CSI mismatch. The proposed reformulation for the probability constraint is not only exact but also tractable. Simulation results show that the distributionally robust transceiver outperforms the Gaussian robust design for non-Gaussian channel estimation errors.

C. Structure

The rest of the paper is organized as follows. In Section II, the model of a wireless powered TDD MU-MIMO is presented, and the non-robust transceiver optimization problem is also formulated. The tractable reformulation of the non-robust transceiver design is developed in Section III, while two different types of robust designs are proposed in Section IV. In Section V, the performance of the proposed non-robust and robust designs is tested through numerical simulations. Finally, we conclude our paper in Section VI by making some remarks.

II. Problem Statement

We consider a harvest-and-transmit system with one BS and K users as shown in Fig. 1. The system is an MU-MIMO system, where the BS and the ith user, $i = 1, \ldots, K$, are equipped with N_b and N_i antennas, respectively. We assume that the BS has its own power supply, while all users are powered by harvesting the RF energy sent from the BS.

TDD communication is adopted in this paper. Specifically, there are K+1 time slots in one communication frame T as shown in Fig. 2. For the simplicity of presentation, we set $T = 1$ hereafter. During the first time slot with duration β, the BS broadcasts the $N_b \times 1$ energy-carrying signal vector $\mathbf{s}$ to all K users, where $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{B}$, $E\{\cdot\}$ stands for the statistical expectation, $(\cdot)^H$ denotes the Hermitian transpose, and $0 < \beta < 1$. A transmission power constraint is imposed on the BS with a power limit P. Therefore, we have the following power constraint

$$\mathrm{tr}(\mathbf{B}) \leq P, \quad \mathbf{B} \succeq 0 \tag{1}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace and $\mathbf{A} \succeq 0$ means matrix $\mathbf{A}$ is positive semidefinite. The input RF power at the ith user during the first time slot is

$$P_{I,i} = \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H), \quad i = 1, \cdots, K \tag{2}$$

where $\mathbf{H}_i$ is an $N_i \times N_b$ MIMO channel matrix from the BS to the ith user.

Remark 1: The reason for considering the power constraint (1) at the BS is twofold. Firstly, the BS has constant power supply in the system considered in this paper. Secondly, we consider the application of maximizing the user data rate in a unit time $T = 1$. With increasing β, the users can harvest more energy, leading to an increased data rate. A total energy constraint at the BS can be formulated as $\beta \mathbf{B} \leq E_b$, which can be easily included in the transceiver optimization problems in this paper.

Remark 2: Transmitting one common vector $\mathbf{s}$ to all K users during the first time slot does not lose any optimality/generality as explained below. Note that different to information transmission where for one user, signals for other users are interference, in the energy transmission phase of the system considered in this paper, energy beams for one user are useful for all users in energy harvesting. Let us consider a scheme that divides the first time slot into K equal sub-slots, where $\mathbf{s}_i$ is transmitted to the ith user during the ith sub-slot with $E\{\mathbf{s}_i \mathbf{s}_i^H\} = \mathbf{B}_i$. Thanks to the broadcasting nature of the wireless channel, any energy beams for one particular user are received by all other users. Thus, $\mathbf{s}_i$ is received by all K users. Therefore, the normalized input RF power at the ith user during the first time slot is given by $\frac{1}{K}\sum_{k=1}^{K} \mathrm{tr}(\mathbf{H}_i \mathbf{B}_k \mathbf{H}_i^H)$, which is equivalent to (2) for $\mathbf{B} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{B}_k$.

In this paper, a two-piecewise linear function [33], [34] is used to model the nonlinearity of the energy harvesting circuit at the users, where before the energy harvesting circuit is saturated, the output power increases linearly with the input power. After reaching saturation, the output power remains stable. Using this nonlinear model, the RF energy harvested by the ith user, $i = 1, \ldots, K$, during the first time slot is given by

$$\tilde{E}_{r,i} = \beta \min(\eta_a \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H), E_i') \tag{3}$$

where E_i' is the maximal output power of user i, and $0 < \eta_a < 1$ is the energy conversion efficiency.

Remark 3: There are many research on nonlinear energy harvesting models in the literature [15]–[20]. The nonlinear model in (3) has been shown to match the experimental results [34], [35]. Extending the algorithms developed in this paper to other nonlinear models (e.g. [36], [37]) is an interesting research topic for future study.

Using the harvested energy, the ith user, $i = 1, \ldots, K$, transmits information-bearing signal vector $\mathbf{x}_i$ to the BS at the $(i+1)$-th time slot with duration α_i, where $E\{\mathbf{x}_i \mathbf{x}_i^H\} = \mathbf{F}_i$ and $0 < \alpha_i < 1$. Thus, α_i and β need to satisfy the following

constraint

$$0 < \sum_{i=1}^{K} \alpha_i + \beta \leq 1. \tag{4}$$

In addition, considering the power limit on the users side, the following power constraints are imposed for all K users

$$\mathrm{tr}(\mathbf{F}_i) \leq P_i, \quad \mathbf{F}_i \succeq 0, \quad i = 1, \ldots, K \tag{5}$$

where P_i is the peak transmission power of user i.

The energy consumed by the ith user to transmit $\mathbf{x}_i$ is $\alpha_i \mathrm{tr}(\mathbf{F}_i)$. Based on [38], the circuit energy consumption consists of a static part and a dynamic part. The static part is used to maintain the basic circuit operations and is given by $\alpha_i N_i P_c$, where P_c is the per-antenna static power consumption. The dynamic part depends on the amount of information processing. According to [38], the dynamic part is modelled as $\eta_b \tilde{E}_{r,i}$, where $\eta_b \in (0, 1)$. Since the users are powered by the harvested energy, for the ith user, $i = 1, \ldots, K$, the following energy constraint is required to be satisfied

$$\begin{aligned} \alpha_i(\mathrm{tr}(\mathbf{F}_i) + N_i P_c) &\leq (1-\eta_b)\tilde{E}_{r,i} \\ &= \beta \min(\eta \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H), E_i) \end{aligned} \tag{6}$$

where $\eta = \eta_a(1-\eta_b)$ and $E_i = (1-\eta_b)E_i'$.

The information rate of the ith user is

$$R_i = \alpha_i \log_2 \left| \mathbf{I}_{N_b} + \sigma^{-2} \mathbf{G}_i \mathbf{F}_i \mathbf{G}_i^H \right|, \quad i = 1, \ldots, K \tag{7}$$

where $\mathbf{G}_i$ is an $N_b \times N_i$ MIMO channel matrix from the ith user to the BS, $\mathbf{I}_n$ is an $n \times n$ identity matrix, σ^2 is the variance of noise at the BS, and $|\cdot|$ denotes the matrix determinant. Inspired by [39]–[41], the proportional fairness of information rate of the users is taken as the performance measure in this paper, which is defined as

$$\sum_{i=1}^{K} \log_2 R_i. \tag{8}$$

The idea of (8) can be interpreted as follows. If a user's information rate increases then the logarithmic function in (8) decreases this "reward". Compared with the commonly used sum rate criterion $\sum_{i=1}^{K} R_i$, (8) avoids the starvation of any particular user. Compared with the max-min rate criterion $\min_{i=1,\ldots,K} R_i$, (8) is not limited by the channel of the worst user. The meaning of the term "proportional fairness" comes from the fact that the differential of the logarithm of a user's information rate is inversely proportional to itself [41]. Considering (1) and (4)-(8), the proportionally fair users rate maximization problem can be formulated as

$$\max_{\mathbf{F}_i, \mathbf{B}, \alpha_i, \beta} \sum_{i=1}^{K} \log_2 \left(\alpha_i \log_2 \left| \mathbf{I}_{N_b} + \sigma^{-2} \mathbf{G}_i \mathbf{F}_i \mathbf{G}_i^H \right| \right) \tag{9}$$

$$s.t.\ \alpha_i(\mathrm{tr}(\mathbf{F}_i) + N_i P_c) \leq \beta \min(\eta \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H), E_i), \quad i = 1, \ldots, K \tag{10}$$

$$\mathrm{tr}(\mathbf{B}) \leq P, \quad \mathbf{B} \succeq 0 \tag{11}$$

$$\mathrm{tr}(\mathbf{F}_i) \leq P_i, \ \mathbf{F}_i \succeq 0, \quad i = 1, \ldots, K \tag{12}$$

$$0 < \beta < 1, \quad 0 < \alpha_i < 1, \ i = 1, \ldots, K \tag{13}$$

$$0 < \sum_{i=1}^{K} \alpha_i + \beta \leq 1. \tag{14}$$

III. Non-Robust Design

By observing the problem in (9)-(14), we can see that the constraints in (11)-(14) are all convex. However, the objective function in (9) is non-convex. In addition, since α_i, $\mathbf{F}_i$, β, $\mathbf{B}$ are optimization variables, $\alpha_i \mathrm{tr}(\mathbf{F}_i)$ and $\beta\eta \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H)$ in (10) are bilinear functions which are non-convex. Thus, the constraints in (10) are non-convex. Therefore, the problem in (9)-(14) is a non-convex optimization problem.

In this section, we shall show that the problem in (9)-(14), in fact, can be solved by converting it into a convex optimization problem. Towards this goal, we first derive the optimal structure of $\mathbf{F}_i$. Let us introduce the singular value decomposition (SVD) of $\mathbf{G}_i = \mathbf{U}_i \mathbf{\Lambda}_{g,i}^{\frac{1}{2}} \mathbf{V}_i^H$, $i = 1, \ldots, K$, where $\mathbf{\Lambda}_{g,i}$ is an $M_i \times M_i$ diagonal matrix with $M_i = \min(N_b, N_i)$. The diagonal elements of $\mathbf{\Lambda}_{g,i}$ are non-negative and arranged in a decreasing order. According to the well-known principle of MIMO transmitter optimization [42], the optimal structure of $\mathbf{F}_i$ is given by $\mathbf{F}_i = \mathbf{V}_i \mathbf{\Lambda}_{f,i} \mathbf{V}_i^H$ in order to maximize (9), where $\mathbf{\Lambda}_{f,i}$ is an $M_i \times M_i$ diagonal matrix with non-negative diagonal elements. On the other hand, this structure is a unitary transform so that it does not change $\mathrm{tr}(\mathbf{F}_i)$. Therefore, this structure does not affect the feasible region of (10) and (12). Thus, for each $i = 1, \ldots, K$, we prove that the optimal structure of $\mathbf{F}_i$ is

$$\mathbf{F}_i = \mathbf{V}_i \mathbf{\Lambda}_{f,i} \mathbf{V}_i^H, \quad i = 1, \ldots, K. \tag{15}$$

By substituting (15) into the problem in (9)-(14), we obtain the optimization problem of

$$\max_{\mathbf{B}, \lambda_{f,i,j}, \alpha_i, \beta} \ \sum_{i=1}^{K} \log_2 \left(\sum_{j=1}^{M_i} \alpha_i \log_2 \left(1 + \sigma^{-2} \lambda_{f,i,j} \lambda_{g,i,j} \right) \right) \tag{16}$$

$$s.t.\ \alpha_i \left(\sum_{j=1}^{M_i} \lambda_{f,i,j} + N_i P_c \right) \leq \beta \min(\eta \mathrm{tr}(\mathbf{H}_i \mathbf{B} \mathbf{H}_i^H), E_i), \quad i = 1, \ldots, K \tag{17}$$

$$\mathrm{tr}(\mathbf{B}) \leq P, \ \mathbf{B} \succeq 0 \tag{18}$$

$$\sum_{j=1}^{M_i} \lambda_{f,i,j} \leq P_i, \ \lambda_{f,i,j} \geq 0, j = 1, \ldots, M_i, i = 1, \ldots, K \tag{19}$$

$$0 < \beta < 1, \ 0 < \alpha_i < 1, \ i = 1, \ldots, K \tag{20}$$

$$0 < \sum_{i=1}^{K} \alpha_i + \beta \leq 1 \tag{21}$$

where $\lambda_{f,i,j}$ and $\lambda_{g,i,j}$, $j = 1, \ldots, M_i$, are the jth diagonal element of $\mathbf{\Lambda}_{f,i}$ and $\mathbf{\Lambda}_{g,i}$, respectively. Let us introduce $\tilde{\lambda}_{f,i,j} = \alpha_i \lambda_{f,i,j}$, $j = 1, \ldots, M_i$, $i = 1, \ldots, K$ and $\tilde{\mathbf{B}} = \beta \mathbf{B}$ to handle the non-convex terms in (17). The problem in (16)-(21) then becomes

$$\max_{\tilde{\mathbf{B}}, \tilde{\lambda}_{f,i,j}, \alpha_i, \beta} \ \sum_{i=1}^{K} \log_2 \left(\sum_{j=1}^{M_i} \alpha_i \log_2 \left(1 + \frac{\tilde{\lambda}_{f,i,j} \lambda_{g,i,j}}{\sigma^2 \alpha_i} \right) \right) \tag{22}$$

$$s.t.\ \sum_{j=1}^{M_i}\tilde{\lambda}_{f,i,j}+\alpha_i N_i P_c \le \min(\eta \mathrm{tr}(\mathbf{H}_i\tilde{\mathbf{B}}\mathbf{H}_i^H), \beta E_i),$$
$$i=1,\ldots,K \quad (23)$$
$$\mathrm{tr}(\tilde{\mathbf{B}}) \le \beta P,\ \tilde{\mathbf{B}} \succeq 0 \quad (24)$$
$$\sum_{j=1}^{M_i}\tilde{\lambda}_{f,i,j} \le \alpha_i P_i, \tilde{\lambda}_{f,i,j} \ge 0, \quad j=1,\ldots,M_i,$$
$$i=1,\ldots,K \quad (25)$$
$$0<\beta<1,\ 0<\alpha_i<1,\ i=1,\ldots,K \quad (26)$$
$$0<\sum_{i=1}^{K}\alpha_i+\beta \le 1. \quad (27)$$

Note that the constraint in (23) is convex now. In addition, constraints in (24)-(27) are still convex. However, the objective function in (22) still seems non-convex. Thus, the problem in (22)-(27) is a non-convex optimization problem. In what follows, we shall show that the objective function in (22) can be converted into convex alternatives by applying the following lemma.

Lemma 1 [43]: If $f:\mathbb{R}^n \to \mathbb{R}$, then the perspective of f is the function $g:\mathbb{R}^{n+1}\to\mathbb{R}$ defined by

$$g(\mathbf{x},t) = tf(\mathbf{x}/t) \quad (28)$$

with domain

$$\mathbf{dom}\ g = \{(\mathbf{x},t)|\mathbf{x}/t \in \mathbf{dom}\ f,\ t>0\} \quad (29)$$

where $\mathbb{R}$ denotes the set of all real numbers. The perspective operation preserves convexity, i.e., if f is a concave function, then so is its perspective function g.

To simplify the problem in (22)-(27), we introduce new variables r_i, $i=1,\ldots,K$, and take the hypograph of the terms in (22). Consequently, the problem in (22)-(27) can be rewritten as

$$\max_{\tilde{\mathbf{B}},\tilde{\lambda}_{f,i,j},\alpha_i,\beta,r_i} \sum_{i=1}^{K}\log_2 r_i \quad (30)$$
$$s.t.\ \sum_{j=1}^{M_i}\alpha_i\log_2\left(1+\frac{\tilde{\lambda}_{f,i,j}\lambda_{g,i,j}}{\sigma^2\alpha_i}\right) \ge r_i,\ i=1,\ldots,K \quad (31)$$
$$\sum_{j=1}^{M_i}\tilde{\lambda}_{f,i,j}+\alpha_i N_i P_c \le \min(\eta\mathrm{tr}(\mathbf{H}_i\tilde{\mathbf{B}}\mathbf{H}_i^H), \beta E_i),$$
$$i=1,\ldots,K \quad (32)$$
$$\mathrm{tr}(\tilde{\mathbf{B}}) \le \beta P,\ \tilde{\mathbf{B}} \succeq 0 \quad (33)$$
$$\sum_{j=1}^{M_i}\tilde{\lambda}_{f,i,j} \le \alpha_i P_i, \tilde{\lambda}_{f,i,j} \ge 0,$$
$$j=1,\ldots,M_i, i=1,\ldots,K \quad (34)$$
$$0<\beta<1,\ 0<\alpha_i<1,\ i=1,\ldots,K \quad (35)$$
$$0<\sum_{i=1}^{K}\alpha_i+\beta \le 1. \quad (36)$$

Let us define $f(\tilde{\lambda}_{f,i,j}) = \log_2(1+\sigma^{-2}\tilde{\lambda}_{f,i,j}\lambda_{g,i,j})$. Then $g(\tilde{\lambda}_{f,i,j},\alpha_i) = \alpha_i\log_2\left(1+\frac{\tilde{\lambda}_{f,i,j}\lambda_{g,i,j}}{\sigma^2\alpha_i}\right)$ is the perspective of $f(\tilde{\lambda}_{f,i,j})$. Since $f(\tilde{\lambda}_{f,i,j})$ is a concave function of $\tilde{\lambda}_{f,i,j}$, according to Lemma 1, $g(\tilde{\lambda}_{f,i,j},\alpha_i)$ is concave with respect to $(\tilde{\lambda}_{f,i,j},\alpha_i)$. Therefore, constraints in (31) are convex. Since $\eta\mathrm{tr}(\mathbf{H}_i\tilde{\mathbf{B}}\mathbf{H}_i^H)$ is a linear function of $\tilde{\mathbf{B}}$, the constraints in (32) are convex. Thus, the problem in (30)-(36) is a convex optimization problem.

Remark 4: Off-the-shelf optimization software packages are available to solve the problem in (30)-(36), for example, the CVX toolbox [44]. Particularly, the perspective function $g(\tilde{\lambda}_{f,i,j},\alpha_i)$ can be handled by function 'rel_entr' using CVX [44].

Remark 5: The steps used in this section to convert the original non-convex rate optimization problem to a convex problem can be extended to the weighted sum rate criterion $\sum_{i=1}^{K} w_i R_i$, where w_i is the weight factor and R_i is given by (7). We would like to note that without a priori knowledge, it is difficult in practice to choose the weights of users. Wrong weights may exacerbate the fairness problem. Thus, it is common in practice to choose $w_i=1$, $i=1,\ldots,K$, which gives the sum rate criterion $\sum_{i=1}^{K} R_i$.

Remark 6: The system design based on the max-min rate criterion which maximizes the minimal data rate among all K users can be formulated by substituting the objective function (30) and the constraint (31) as below

$$\max_{\tilde{\mathbf{B}},\tilde{\lambda}_{f,i,j},\alpha_i,\beta,r} r \quad (37)$$
$$s.t.\ \sum_{j=1}^{M_i}\alpha_i\log_2\left(1+\frac{\tilde{\lambda}_{f,i,j}\lambda_{g,i,j}}{\sigma^2\alpha_i}\right) \ge r,\quad i=1,\ldots,K \quad (38)$$
$$\text{constraints } (32),(33),(34),(35),(36). \quad (39)$$

The problem (37)-(39) is a convex optimization problem and can be solved by the CVX toolbox.

IV. Robust Designs

In the previous section, the CSI of channels $\mathbf{H}_i$ and $\mathbf{G}_i$, $i=1,\ldots,K$, are assumed to be perfectly known. However, the true CSI cannot be obtained in real world applications. For example, quantization errors and outdated channel training may lead to CSI mismatch.

In this section, we shall propose two robust designs for hedging the mismatch between the actual and the nominal CSI. The first design assumes that the CSI mismatch is subject to Gaussian distribution, while the second design only requires the first and the second order moment of the CSI mismatch. To begin, we define the CSI mismatch of $\mathbf{H}_i$ and $\mathbf{G}_i$ as

$$\mathbf{\Delta}_{h,i} = \mathbf{H}_i - \hat{\mathbf{H}}_i,\quad i=1,\ldots,K \quad (40)$$
$$\mathbf{\Delta}_{g,i} = \mathbf{G}_i - \hat{\mathbf{G}}_i,\quad i=1,\ldots,K \quad (41)$$

where $\hat{\mathbf{H}}_i$ and $\hat{\mathbf{G}}_i$ are the estimated channels of $\mathbf{H}_i$ and $\mathbf{G}_i$, respectively.

A. Gaussian Robust Design

In this subsection, all entries of $\mathbf{\Delta}_{h,i}$ and $\mathbf{\Delta}_{g,i}$ are assumed to be subject to uncorrelated Gaussian distribution. We shall

develop an efficient algorithm for solving the 'robust' version of the problem in (9)-(14).

By considering the CSI mismatch of $\mathbf{H}_i$ and $\mathbf{G}_i$ in (40)-(41) and taking the hypograph of each user's information rate in (9), the problem in (9)-(14) becomes the following robust proportionally fair rate optimization problem

$$\max_{\mathbf{F}_i,\mathbf{B},\alpha_i,\beta,r_i} \quad \sum_{i=1}^{K}\log_2 r_i \tag{42}$$

$$s.t.\ \Pr_{[\mathbb{G}]}\Big\{\alpha_i\log_2\Big|\mathbf{I}_{N_b}+\sigma^{-2}(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\mathbf{F}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H\Big| \geq r_i\Big\}\geq 1-\epsilon_1,\quad i=1,\ldots,K \tag{43}$$

$$\Pr_{[\mathbb{G}]}\Big\{\beta\eta\mathrm{tr}((\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})\mathbf{B}(\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})^H) \geq \alpha_i(\mathrm{tr}(\mathbf{F}_i)+N_iP_c)\Big\}\geq 1-\epsilon_2,\ i=1,\ldots,K \tag{44}$$

$$\mathrm{tr}(\mathbf{B})\leq P,\ \mathbf{B}\succeq 0 \tag{45}$$

$$\alpha_i(\mathrm{tr}(\mathbf{F}_i)+N_iP_c)\leq \beta E_i,\ \mathrm{tr}(\mathbf{F}_i)\leq P_i,\ \mathbf{F}_i\succeq 0,\quad i=1,\ldots,K \tag{46}$$

$$0<\beta<1,\ 0<\alpha_i<1,\ i=1,\ldots,K \tag{47}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1. \tag{48}$$

Note that the chance constraints in (43) and (44) are introduced since $\boldsymbol{\Delta}_{g,i}$ and $\boldsymbol{\Delta}_{h,i}$ are random matrices. Here, $\Pr_{[\mathbb{G}]}[\cdot]$ in (43) and (44) stands for the probability under the Gaussian distribution $\mathbb{G}$, $0<\epsilon_1<1$ and $0<\epsilon_2<1$ in (43) and (44) are chosen according to quality-of-service (QoS) specifications in practice.

It is well-known [45] that chance constraints are usually non-convex. Hence, the optimization problem in (42)-(48) is a non-convex optimization problem due to the chance constraints in (43) and (44). To overcome this difficulty, the Bernstein inequality [46], [47], which is elaborated in Lemma 2, is utilized in this paper to convert the chance constraints in (43) and (44) into deterministic convex constraints.

Lemma 2 (Bernstein Inequality [46], [47]): Let $f(\mathbf{x})=\mathbf{x}^H\mathbf{Y}\mathbf{x}+2\mathrm{Re}\{\mathbf{x}^H\mathbf{u}\}$, where $\mathrm{Re}\{\cdot\}$ denotes the real part, $\mathbf{Y}\in\mathbb{H}^{N\times N}$ is a complex Hermitian matrix, $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_N)$ is a standard circularly symmetric complex Gaussian (CSCG) random vector with zero mean and unit variance, and $\mathbf{u}\in\mathbb{C}^{N\times 1}$. Then, for any $\delta>0$, we have the following statement

$$\Pr_{[\mathbb{G}]}\left\{f(\mathbf{x})\geq \mathrm{tr}(\mathbf{Y})-\sqrt{2\delta}\sqrt{\|\mathbf{Y}\|_F^2+2\|\mathbf{u}\|^2}-\delta c^+(\mathbf{Y})\right\}\geq 1-e^{-\delta} \tag{49}$$

where $c^+(\mathbf{Y})=\max\{\lambda_{\max}(-\mathbf{Y}),0\}$ with $\lambda_{\max}(-\mathbf{Y})$ denoting the maximum eigenvalue of matrix $-\mathbf{Y}$, $\|\cdot\|_F$ and $\|\cdot\|$ denote the matrix Frobenius norm and the vector Euclidean norm, respectively.

However, Lemma 2 cannot be applied to (43) straight away since it is not a quadratic function of $\boldsymbol{\Delta}_{g,i}$. For this, the Taylor series expansion is applied to approximate (43). To proceed, we note that the first-order Taylor series expansion of function $\ln|\mathbf{I}+\mathbf{X}|$ is equal to $\mathrm{tr}(\mathbf{X})$. Then, by introducing $\tilde{\mathbf{B}}=\beta\mathbf{B}$, $\tilde{\mathbf{F}}_i=\alpha_i\mathbf{F}_i$, $i=1,\ldots,K$, and taking the first-order Taylor series expansion of (43), the problem in (42)-(48) becomes the following problem[1]

$$\max_{\mathbf{F}_i,\mathbf{B},\alpha_i,\beta,r_i} \quad \sum_{i=1}^{K}\log_2 r_i \tag{50}$$

$$s.t.\ \Pr_{[\mathbb{G}]}\Big\{\mathrm{tr}((\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\tilde{\mathbf{F}}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H) \geq r_i\sigma^2\ln 2\Big\}\geq 1-\epsilon_1,\quad i=1,\ldots,K \tag{51}$$

$$\Pr_{[\mathbb{G}]}\Big\{\eta\mathrm{tr}((\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})\tilde{\mathbf{B}}(\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})^H) \geq \mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c\Big\}\geq 1-\epsilon_2,\quad i=1,\ldots,K \tag{52}$$

$$\mathrm{tr}(\tilde{\mathbf{B}})\leq\beta P,\ \tilde{\mathbf{B}}\succeq 0 \tag{53}$$

$$\mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c\leq\beta E_i,\ \mathrm{tr}(\tilde{\mathbf{F}}_i)\leq\alpha_iP_i,\ \tilde{\mathbf{F}}_i\succeq 0,\quad i=1,\ldots,K \tag{54}$$

$$0<\beta<1,\ 0<\alpha_i<1,\ i=1,\ldots,K \tag{55}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1. \tag{56}$$

In fact, the trace term in (51) is equivalent to the following quadratic form

$$\mathrm{tr}((\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\tilde{\mathbf{F}}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H) = \mathbf{g}_i^H(\mathbf{I}_{N_b}\otimes\tilde{\mathbf{F}}_i)\mathbf{g}_i,\ i=1,\ldots,K \tag{57}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{g}_i=\mathrm{vec}((\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H)$, $\mathrm{vec}(\mathbf{A})$ denotes the vector formed by stacking all columns of matrix $\mathbf{A}$ on top of each other. Thus, considering (57), (51) can be rewritten as

$$\Pr_{[\mathbb{G}]}\left\{\boldsymbol{\delta}_{g,i}^H\bar{\mathbf{F}}_i\boldsymbol{\delta}_{g,i}+2\mathrm{Re}\{\boldsymbol{\delta}_{g,i}^H\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\}\geq r_i\sigma^2\ln 2-\hat{\mathbf{g}}_i^H\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\right\}\geq 1-\epsilon_1,\quad i=1,\ldots,K \tag{58}$$

where $\bar{\mathbf{F}}_i=\mathbf{I}_{N_b}\otimes\tilde{\mathbf{F}}_i$, $\hat{\mathbf{g}}_i=\mathrm{vec}(\hat{\mathbf{G}}_i^H)$, and $\boldsymbol{\delta}_{g,i}=\mathrm{vec}(\boldsymbol{\Delta}_{g,i}^H)$. We assume that $\boldsymbol{\delta}_{g,i}\sim\mathcal{CN}(\mathbf{0},2\tau_i^2\mathbf{I}_{N_bN_i})$.

Using the Bernstein inequality in Lemma 2, (58) holds by

$$2\tau_i^2\mathrm{tr}(\bar{\mathbf{F}}_i)-2\sqrt{-2\ln\epsilon_1}\sqrt{\|\tau_i^2\bar{\mathbf{F}}_i\|_F^2+\|\tau_i\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\|^2} +\ln\epsilon_1c^+(2\tau_i^2\bar{\mathbf{F}}_i)\geq r_i\sigma^2\ln 2-\hat{\mathbf{g}}_i^H\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i,\ i=1,\ldots,K \tag{59}$$

where $c^+(2\tau_i^2\bar{\mathbf{F}}_i)=\max\{\lambda_{\max}(-2\tau_i^2\bar{\mathbf{F}}_i),0\}$. For each $i=1,\ldots,K$, the constraint (59) can be rewritten as

$$2\tau_i^2\mathrm{tr}(\bar{\mathbf{F}}_i)-2\sqrt{-2\ln\epsilon_1}\theta_{1,i}+\ln\epsilon_1\theta_{2,i} +\hat{\mathbf{g}}_i^H\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\geq r_i\sigma^2\ln 2 \tag{60}$$

$$\sqrt{\left\|\tau_i^2\bar{\mathbf{F}}_i\right\|_F^2+\left\|\tau_i\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\right\|^2}\leq\theta_{1,i} \tag{61}$$

$$\theta_{2,i}\mathbf{I}_{N_bN_i}+2\tau_i^2\bar{\mathbf{F}}_i\geq 0,\quad \theta_{2,i}\geq 0. \tag{62}$$

Note that (60)-(62) are convex constraints. In particular, (61) is a second-order cone constraint, which can be rewritten as

$$\left\|\begin{pmatrix}\mathrm{vec}\left(\tau_i^2\bar{\mathbf{F}}_i\right)\\ \tau_i\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\end{pmatrix}\right\|\leq\theta_{1,i},\quad i=1,\ldots,K. \tag{63}$$

The constraint in (52) can be treated in a similar manner as the constraint in (51). Thus, the problem in (50)-(56) can be

[1]Using the first-order Taylor series expansion changes the feasible region specified by (43). However, since $\log_2(1+x)>x$ for $0<x<1$ and $\log_2(1+x)<x$ when $x>1$, whether the feasible region of (43) is expanded or reduced after the Taylor series expansion depends on the value of $\sigma^{-2}(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\mathbf{F}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H$. Therefore, it is difficult to analyze the relationship between the original problem (42)-(48) and the approximated problem (50)-(56).

converted to the following problem

$$\max_{\tilde{\mathbf{F}}_i,\tilde{\mathbf{B}},\alpha_i,\beta,r_i,\theta_{1,i},\theta_{2,i},\xi_{1,i},\xi_{2,i}} \quad \sum_{i=1}^{K}\log_2 r_i \tag{64}$$

$$s.t.\ 2\tau_i^2\mathrm{tr}(\bar{\mathbf{F}}_i)-2\sqrt{-2\ln\epsilon_1}\theta_{1,i}+\ln\epsilon_1\theta_{2,i}+\hat{\mathbf{g}}_i^H\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i \geq r_i\sigma^2\ln 2,\quad i=1,\ldots,K \tag{65}$$

$$\left\|\begin{pmatrix}\mathrm{vec}\left(\tau_i^2\bar{\mathbf{F}}_i\right)\\ \tau_i\bar{\mathbf{F}}_i\hat{\mathbf{g}}_i\end{pmatrix}\right\|\leq\theta_{1,i},\quad i=1,\ldots,K \tag{66}$$

$$\theta_{2,i}\mathbf{I}_{N_bN_i}+2\tau_i^2\bar{\mathbf{F}}_i\succeq 0,\quad \theta_{2,i}\geq 0,\quad i=1,\ldots,K \tag{67}$$

$$2\eta\tau_i^2\mathrm{tr}(\bar{\mathbf{B}}_i)-2\sqrt{-2\ln\epsilon_2}\xi_{1,i}+\ln\epsilon_2\xi_{2,i}+\eta\hat{\mathbf{h}}_i^H\bar{\mathbf{B}}_i\hat{\mathbf{h}}_i\geq\mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c,\quad i=1,\ldots,K \tag{68}$$

$$\left\|\begin{pmatrix}\mathrm{vec}\left(\eta\tau_i^2\bar{\mathbf{B}}_i\right)\\ \eta\tau_i\bar{\mathbf{B}}_i\hat{\mathbf{h}}_i\end{pmatrix}\right\|\leq\xi_{1,i},\quad i=1,\ldots,K \tag{69}$$

$$\xi_{2,i}\mathbf{I}_{N_bN_i}+2\eta\tau_i^2\bar{\mathbf{B}}\succeq 0,\ \xi_{2,i}\geq 0,\quad i=1,\ldots,K \tag{70}$$

$$\mathrm{tr}(\tilde{\mathbf{B}})\leq\beta P,\quad \tilde{\mathbf{B}}\succeq 0 \tag{71}$$

$$\mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c\leq\beta E_i,\quad \mathrm{tr}(\tilde{\mathbf{F}}_i)\leq\alpha_iP_i,\ \tilde{\mathbf{F}}_i\succeq 0,\quad i=1,\ldots,K \tag{72}$$

$$0<\beta<1,\ 0<\alpha_i<1,\quad i=1,\ldots,K \tag{73}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1 \tag{74}$$

where $\bar{\mathbf{B}}_i=\mathbf{I}_{N_i}\otimes\tilde{\mathbf{B}}$, $\hat{\mathbf{h}}_i=\mathrm{vec}(\hat{\mathbf{H}}_i^H)$, and $\boldsymbol{\delta}_{h,i}=\mathrm{vec}(\boldsymbol{\Delta}_{h,i}^H)$. We assume that $\boldsymbol{\delta}_{h,i}\sim\mathcal{CN}\left(\mathbf{0},2\tau_i^2\mathbf{I}_{N_bN_i}\right)$. The problem in (64)-(74) is a convex problem, and is in the form of disciplined convex programming, which can be solved by the CVX toolbox [44].

B. Distributionally Robust Design

In practice, the exact distribution information of the CSI mismatch might not be available. Moreover, the distribution of the CSI mismatch might not be subject to the Gaussian distribution even if some historical data can be obtained. Thus, in this section, we shall develop an algorithm which only requires partial information of the CSI mismatch distribution.

More specifically, we assume that only the mean and the variance of the CSI mismatch are known to the transmitter and the users. Without loss of generality, we assume that

$$\mathbb{E}_{[\mathbb{P}]}[\mathrm{Re}\{[\boldsymbol{\Delta}_{g,i}]_{m,n}\}]=\mathbb{E}_{[\mathbb{P}]}[\mathrm{Im}\{[\boldsymbol{\Delta}_{g,i}]_{m,n}\}]=0 \tag{75}$$

$$\mathbb{E}_{[\mathbb{P}]}[\mathrm{Re}\{[\boldsymbol{\Delta}_{h,i}]_{m,n}\}]=\mathbb{E}_{[\mathbb{P}]}[\mathrm{Im}\{[\boldsymbol{\Delta}_{h,i}]_{m,n}\}]=0 \tag{76}$$

$$\mathbb{E}_{[\mathbb{P}]}[(\mathrm{Re}\{[\boldsymbol{\Delta}_{g,i}]_{m,n}\})^2]=\mathbb{E}_{[\mathbb{P}]}[(\mathrm{Im}\{[\boldsymbol{\Delta}_{g,i}]_{m,n}\})^2]=\tau_i^2 \tag{77}$$

$$\mathbb{E}_{[\mathbb{P}]}[(\mathrm{Re}\{[\boldsymbol{\Delta}_{h,i}]_{m,n}\})^2]=\mathbb{E}_{[\mathbb{P}]}[(\mathrm{Im}\{[\boldsymbol{\Delta}_{h,i}]_{m,n}\})^2]=\tau_i^2 \tag{78}$$

where $[\mathbf{A}]_{m,n}$ denotes the (m,n)-th entry of matrix $\mathbf{A}$ and $\mathbb{E}_{[\mathbb{P}]}[\cdot]$ stands for the statistical expectation under the distribution of $\mathbb{P}$.

Considering (75)-(78) and the problem in (42)-(48), we obtain a distributionally robust rate optimization problem below

$$\max_{\mathbf{F}_i,\mathbf{B},\alpha_i,\beta,r_i} \quad \sum_{i=1}^{K}\log_2 r_i \tag{79}$$

$$s.t.\ \Pr_{[\mathbb{P}]}\left\{\alpha_i\log_2\left|\mathbf{I}_{N_b}+\sigma^{-2}(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\mathbf{F}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H\right|\geq r_i\right\}\geq 1-\epsilon_1,\quad i=1,\ldots,K \tag{80}$$

$$\Pr_{[\mathbb{P}]}\left\{\beta\eta\mathrm{tr}((\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})\mathbf{B}(\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})^H)\geq\alpha_i(\mathrm{tr}(\mathbf{F}_i)+N_iP_c)\right\}\geq 1-\epsilon_2,\ i=1,\ldots,K \tag{81}$$

$$\mathrm{tr}(\mathbf{B})\leq P,\ \mathbf{B}\succeq 0 \tag{82}$$

$$\alpha_i(\mathrm{tr}(\mathbf{F}_i)+N_iP_c)\leq\beta E_i,\ \mathrm{tr}(\mathbf{F}_i)\leq P_i,\ \mathbf{F}_i\succeq 0,\quad i=1,\ldots,K \tag{83}$$

$$0<\beta<1,\quad 0<\alpha_i<1,\ i=1,\ldots,K \tag{84}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1. \tag{85}$$

$$\text{constraints (75), (76), (77), (78).} \tag{86}$$

Comparing the problem in (79)-(86) with the problem in (42)-(48), it can be seen that the CSI mismatch $\mathbb{P}$ in the problem in (79)-(86) is not subject to $\mathbb{G}$ as that in the problem in (42)-(48).

Similarly, by introducing $\tilde{\mathbf{F}}_i=\alpha_i\mathbf{F}_i$ and $\tilde{\mathbf{B}}=\beta\mathbf{B}$ and applying the first-order Taylor series expansion to (80), the problem in (79)-(86) is simplified to the following problem

$$\max_{\tilde{\mathbf{F}}_i,\tilde{\mathbf{B}},\alpha_i,\beta,r_i} \quad \sum_{i=1}^{K}\log_2 r_i \tag{87}$$

$$s.t.\ \Pr_{[\mathbb{P}]}\left\{\mathrm{tr}((\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\tilde{\mathbf{F}}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H)\geq r_i\sigma^2\ln 2\right\}\geq 1-\epsilon_1,\quad i=1,\ldots,K \tag{88}$$

$$\Pr_{[\mathbb{P}]}\left\{\eta\mathrm{tr}((\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})\tilde{\mathbf{B}}(\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})^H)\geq\mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c\right\}\geq 1-\epsilon_2,\quad i=1,\ldots,K \tag{89}$$

$$\mathrm{tr}(\tilde{\mathbf{B}})\leq\beta P,\ \tilde{\mathbf{B}}\succeq 0 \tag{90}$$

$$\mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_iN_iP_c\leq\beta E_i,\ \mathrm{tr}(\tilde{\mathbf{F}}_i)\leq\alpha_iP_i,\ \tilde{\mathbf{F}}_i\succeq 0,\quad i=1,\ldots,K \tag{91}$$

$$0<\beta<1,\ 0<\alpha_i<1,\ i=1,\ldots,K \tag{92}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1. \tag{93}$$

$$\text{constraints (75), (76), (77), (78).} \tag{94}$$

Obviously, the constraints in (88) and (89) are non-convex. In addition, the distribution of $\mathbb{P}$ in (88) and (89) is not fully known. In the following, the distributionally robust optimization technique [48]–[53] is introduced to handle the chance constraints with partial distribution information.

The idea of distributionally robust optimization is to optimize the 'worst-case' distribution among all possible distributions. More specifically, an ambiguity set $\mathcal{P}$ is created, which contains all possible $\mathbb{P}$. Then, the problem is to optimize the worst-case performance of $\mathbb{P}\in\mathcal{P}$. Therefore, the 'distributionally robust version' of the problem in (87)-(94) is shown

below

$$\max_{\tilde{\mathbf{F}}_i,\tilde{\mathbf{B}},\alpha_i,\beta,r_i} \quad \sum_{i=1}^{K}\log_2 r_i \tag{95}$$

$$s.t. \ \inf_{\mathbb{P}\in\mathcal{P}} \Pr_{[\mathbb{P}]}\{\mathrm{tr}((\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})\tilde{\mathbf{F}}_i(\hat{\mathbf{G}}_i+\boldsymbol{\Delta}_{g,i})^H) \geq r_i\sigma^2\ln 2\} \geq 1-\epsilon_1, \quad i=1,\ldots,K \tag{96}$$

$$\inf_{\mathbb{P}\in\mathcal{P}} \Pr_{[\mathbb{P}]}\{\eta\mathrm{tr}((\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})\tilde{\mathbf{B}}(\hat{\mathbf{H}}_i+\boldsymbol{\Delta}_{h,i})^H) \geq \mathrm{tr}(\tilde{\mathbf{F}}_i)+\alpha_i N_i P_c\} \geq 1-\epsilon_2, i=1,\ldots,K \tag{97}$$

$$\mathcal{P} = \{\mathbb{P} : (75),(76),(77),(78)\} \tag{98}$$

$$\text{constraints } (90),(91),(92),(93). \tag{99}$$

Although the constraints in (96) and (97) seem more complicated than the constraints in (88) and (89), we shall show that they turn out to be more 'tractable'.

By defining

$$\underline{\mathbf{F}_i}=\begin{pmatrix}\mathbf{I}_{N_b}\otimes\mathrm{Re}\{\tilde{\mathbf{F}}_i\} & -\mathbf{I}_{N_b}\otimes\mathrm{Im}\{\tilde{\mathbf{F}}_i\}\\ \mathbf{I}_{N_b}\otimes\mathrm{Im}\{\tilde{\mathbf{F}}_i\} & \mathbf{I}_{N_b}\otimes\mathrm{Re}\{\tilde{\mathbf{F}}_i\}\end{pmatrix},\ \underline{\mathbf{g}_i}=\begin{pmatrix}\mathrm{Re}\{\mathbf{g}_i\}\\ \mathrm{Im}\{\mathbf{g}_i\}\end{pmatrix} \tag{100}$$

where $\mathrm{Im}\{\cdot\}$ denotes the imaginary part, it is straightforward to verify that (57) can be rewritten as

$$\mathbf{g}_i^H(\mathbf{I}_{N_b}\otimes\tilde{\mathbf{F}}_i)\mathbf{g}_i = \underline{\mathbf{g}_i}^T\underline{\mathbf{F}_i}\ \underline{\mathbf{g}_i} \tag{101}$$

where $(\cdot)^T$ denotes the transpose. According to the definition in (100), $\underline{\mathbf{g}_i}$ is a real-valued random vector. By considering (75), (77), and (100), we obtain a new ambiguity set satisfying

$$\mathcal{P}_1 = \Big\{\mathbb{P} : \mathbb{E}_{[\mathbb{P}]}[\underline{\mathbf{g}_i}] = \underline{\hat{\mathbf{g}}_i}, \ \mathbb{E}_{[\mathbb{P}]}[(\underline{\mathbf{g}_i}-\underline{\hat{\mathbf{g}}_i})(\underline{\mathbf{g}_i}-\underline{\hat{\mathbf{g}}_i})^T] = \tau_i^2\mathbf{I}_{2N_bN_i}\Big\} \tag{102}$$

where $\underline{\hat{\mathbf{g}}_i}=\begin{pmatrix}\mathrm{Re}\{\hat{\mathbf{g}}_i\}\\ \mathrm{Im}\{\hat{\mathbf{g}}_i\}\end{pmatrix}$. Therefore, considering (57), (101), and (102), the chance constraints in (96) become

$$\inf_{\mathbb{P}\in\mathcal{P}_1}\Pr_{[\mathbb{P}]}\big\{\underline{\mathbf{g}_i}^T\underline{\mathbf{F}_i}\underline{\mathbf{g}_i}\geq r_i\sigma^2\ln 2\big\}\geq 1-\epsilon_1, \quad i=1,\ldots,K. \tag{103}$$

Note that although (103) is in a simpler real-valued quadratic form, it is still non-convex. To convert (103) into a tractable convex alternative, we need the following lemma.

Lemma 3 (Theorem 2.2 in [48]): Let $L:\mathbb{R}^k\to\mathbb{R}$ be a continuous loss function that is either concave in $\boldsymbol{\xi}$ or quadratic in $\boldsymbol{\xi}$. Then, the following equivalence holds

$$\inf_{\mathbb{P}\in\mathcal{P}'}\Pr_{[\mathbb{P}]}\{L(\boldsymbol{\xi})\leq 0\}\geq 1-\epsilon \iff \sup_{\mathbb{P}\in\mathcal{P}'}\mathbb{P}\text{-CVaR}_\epsilon[L(\boldsymbol{\xi})]\leq 0 \tag{104}$$

where

$$\mathcal{P}' = \big\{\mathbb{P}:\mathbb{E}_{[\mathbb{P}]}[\boldsymbol{\xi}]=\boldsymbol{\mu},\ \mathbb{E}_{[\mathbb{P}]}\big[(\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})^T\big]=\boldsymbol{\Sigma}\big\} \tag{105}$$

$$\mathbb{P}\text{-CVaR}_\epsilon\Big[\tilde{\xi}\Big] = \inf_{\gamma\in\mathbb{R}}\left\{\gamma+\frac{1}{\epsilon}\mathbb{E}_{[\mathbb{P}]}\left[\left(\tilde{\xi}-\beta\right)^+\right]\right\} \tag{106}$$

and for a real number a, $(a)^+=\max(a,0)$.

The term 'CVaR' stands for the conditional value-at-risk. It is a special class of risk measure introduced in [54] and further developed by Rockafellar and Uryasev [55], which is regarded as the tightest convex approximation of the chance constraints according to [56]. Usually, the CVaR is only an approximation of the chance constraint. Fortunately, as shown in Lemma 2 of [55], this approximation is exact under the distributionally robust optimization setting. CVaR is not computationally tractable although it is convex. The following lemma gives a tractable semidefinite programming (SDP) alternative for the worst-case CVaR under the distributionally robust framework.

Lemma 4 (Theorem 2.3 in [48]): If $L(\boldsymbol{\xi})=\boldsymbol{\xi}^T\mathbf{Q}\boldsymbol{\xi}+\mathbf{q}^T\boldsymbol{\xi}+q^0$ for some $\mathbf{Q}\in\mathbb{S}^k$, $\mathbf{q}\in\mathbb{R}^k$, and $q^0\in\mathbb{R}$, where $\mathbb{S}^k$ denotes the space of symmetric matrices of dimension k. Then,

$$\left\{\boldsymbol{x}\in\mathbb{R}^n : \sup_{\mathbb{P}\in\mathcal{P}'}\mathbb{P}\text{-CVaR}_\epsilon\left[L(\boldsymbol{\xi})\right]\leq 0\right\} \tag{107}$$

can be written as

$$\left\{\begin{array}{l}\mathbf{M}\succeq 0,\ \gamma+\dfrac{1}{\epsilon}\mathrm{tr}(\boldsymbol{\Omega}\mathbf{M})\leq 0\\ \mathbf{M}-\begin{bmatrix}\mathbf{Q} & \dfrac{1}{2}\mathbf{q}\\ \dfrac{1}{2}\mathbf{q}^T & q^0-\gamma\end{bmatrix}\succeq 0\end{array}\right\} \tag{108}$$

where

$$\boldsymbol{\Omega}=\begin{bmatrix}\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^T & \boldsymbol{\mu}\\ \boldsymbol{\mu}^T & 1\end{bmatrix} \tag{109}$$

and $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are defined as in (105).

For each $i=1,\ldots,K$, by applying Lemma 3 and Lemma 4 to (103), it follows that

$$\mathbf{M}_{1,i}\succeq 0,\ \gamma_{1,i}+\frac{1}{\epsilon_1}\mathrm{tr}(\boldsymbol{\Omega}_{1,i}\mathbf{M}_{1,i})\leq 0 \tag{110}$$

$$\mathbf{M}_{1,i}-\begin{pmatrix}-\underline{\mathbf{F}_i} & \mathbf{0}\\ \mathbf{0} & r_i\sigma^2\ln 2-\gamma_{1,i}\end{pmatrix}\succeq 0 \tag{111}$$

where

$$\boldsymbol{\Omega}_{1,i}=\begin{bmatrix}\tau_i^2\mathbf{I}_{2N_bN_i}+\underline{\hat{\mathbf{g}}_i}\ \underline{\hat{\mathbf{g}}_i}^T & \underline{\hat{\mathbf{g}}_i}\\ \underline{\hat{\mathbf{g}}_i}^T & 1\end{bmatrix},\quad i=1,\ldots,K.$$

Similarly, for each $i=1,\ldots,K$, the chance constraint (97) can be transformed into the following constraints

$$\mathbf{M}_{2,i}\succeq 0,\ \gamma_{2,i}+\frac{1}{\epsilon_2}\mathrm{tr}(\boldsymbol{\Omega}_{2,i}\mathbf{M}_{2,i})\leq 0 \tag{112}$$

$$\mathbf{M}_{2,i}-\begin{pmatrix}-\underline{\mathbf{B}_i} & \mathbf{0}\\ \mathbf{0} & \eta^{-1}(\mathrm{tr}(\mathrm{Re}\{\tilde{\mathbf{F}}_i\})+\alpha_iN_iP_c)-\gamma_{2,i}\end{pmatrix}\succeq 0 \tag{113}$$

where for each $i=1,\ldots,K$

$$\boldsymbol{\Omega}_{2,i}=\begin{bmatrix}\tau_i^2\mathbf{I}_{2N_bN_i}+\underline{\hat{\mathbf{h}}_i}\ \underline{\hat{\mathbf{h}}_i}^T & \underline{\hat{\mathbf{h}}_i}\\ \underline{\hat{\mathbf{h}}_i}^T & 1\end{bmatrix}$$

$$\underline{\mathbf{B}_i}=\begin{pmatrix}\mathbf{I}_{N_i}\otimes\mathrm{Re}\{\tilde{\mathbf{B}}\} & -\mathbf{I}_{N_i}\otimes\mathrm{Im}\{\tilde{\mathbf{B}}\}\\ \mathbf{I}_{N_i}\otimes\mathrm{Im}\{\tilde{\mathbf{B}}\} & \mathbf{I}_{N_i}\otimes\mathrm{Re}\{\tilde{\mathbf{B}}\}\end{pmatrix},\ \underline{\hat{\mathbf{h}}_i}=\begin{pmatrix}\mathrm{Re}\{\hat{\mathbf{h}}_i\}\\ \mathrm{Im}\{\hat{\mathbf{h}}_i\}\end{pmatrix}.$$

By replacing (96) and (97) with (110)-(113), the problem in (95)-(99) can be converted to the following problem

$$\max_{\tilde{\mathbf{F}}_i,\tilde{\mathbf{B}},\mathbf{M}_{1,i},\mathbf{M}_{2,i},\alpha_i,\beta,\gamma_{1,i},\gamma_{2,i},r_i} \sum_{i=1}^{K}\log_2 r_i \tag{114}$$

$$s.t.\ \mathbf{M}_{1,i}\succeq 0,\ \gamma_{1,i}+\frac{1}{\epsilon_1}\mathrm{tr}(\mathbf{\Omega}_{1,i}\mathbf{M}_{1,i}) \leq 0,\ i=1,\ldots,K \tag{115}$$

$$\mathbf{M}_{1,i}-\begin{pmatrix} -\mathbf{F}_i & \mathbf{0} \\ \mathbf{0} & r_i\sigma^2\ln 2-\gamma_{1,i}\end{pmatrix} \succeq 0, i=1,\ldots,K \tag{116}$$

$$\mathbf{M}_{2,i}\succeq 0,\ \gamma_{2,i}+\frac{1}{\epsilon_2}\mathrm{tr}(\mathbf{\Omega}_{2,i}\mathbf{M}_{2,i})\leq 0, \quad i=1,\ldots,K \tag{117}$$

$$\mathbf{M}_{2,i} - \begin{pmatrix} -\mathbf{B}_i & \mathbf{0} \\ \mathbf{0} & \eta^{-1}(\mathrm{tr}(\mathrm{Re}\{\tilde{\mathbf{F}}_i\}) \\ & +\alpha_i N_i P_c)-\gamma_{2,i}\end{pmatrix} \succeq 0, \quad i=1,\ldots,K \tag{118}$$

$$\mathrm{tr}(\mathrm{Re}\{\tilde{\mathbf{B}}\})\leq \beta P \tag{119}$$

$$\mathrm{Re}\{\tilde{\mathbf{B}}\}\succeq 0,\ (\mathrm{Im}\{\tilde{\mathbf{B}}\})^T=-\mathrm{Im}\{\tilde{\mathbf{B}}\} \tag{120}$$

$$\mathrm{tr}(\mathrm{Re}\{\tilde{\mathbf{F}}_i\})+\alpha_i N_i P_c\leq \beta E_i, \quad i=1,\ldots,K \tag{121}$$

$$\mathrm{tr}(\mathrm{Re}\{\tilde{\mathbf{F}}_i\})\leq \alpha_i P_i, \quad i=1,\ldots,K \tag{122}$$

$$\mathrm{Re}\{\tilde{\mathbf{F}}_i\}\succeq 0,\ (\mathrm{Im}\{\tilde{\mathbf{F}}_i\})^T=-\mathrm{Im}\{\tilde{\mathbf{F}}_i\}, \quad i=1,\ldots,K \tag{123}$$

$$0<\beta<1,\ 0<\alpha_i<1, \quad i=1,\ldots,K \tag{124}$$

$$0<\sum_{i=1}^{K}\alpha_i+\beta\leq 1. \tag{125}$$

The problem in (114)-(125) is an SDP problem, which is computationally tractable and can be solved by the CVX toolbox [44]. Note that the real and imaginary parts of $\tilde{\mathbf{F}}_i$ and $\tilde{\mathbf{B}}$ are treated as independent matrices and optimized separately. The constraint in (120) ensures $\tilde{\mathbf{B}}\succeq 0$, and the constraints in (123) ensure $\tilde{\mathbf{F}}_i\succeq 0$.

V. Simulation Results

In this section, the performance of the proposed non-robust and robust designs for wireless powered MU-MIMO systems is tested through numerical simulations. We consider a scenario in which there are two users around the BS. The distance between the BS and each user is $d_i = 10$ meters, $i = 1, 2$. The BS is equipped with $N_b = 3$ antennas and each user is equipped with $N_i = 2$ antennas, $i = 1, 2$. Following [36], [37], [46], the channel matrices are modeled as $\mathbf{H}_i = (\varpi_0 d_i^{-\zeta})^{\frac{1}{2}}\bar{\mathbf{H}}_i$ and $\mathbf{G}_i = (\varpi_0 d_i^{-\zeta})^{\frac{1}{2}}\bar{\mathbf{G}}_i$, $i = 1, 2$, where $\varpi_0 = (\lambda/4\pi)^2$ is the path loss at a reference distance of $d_0 = 1$ meter, λ is the wavelength of the carrier signal, $\varpi_0 d_i^{-\zeta}$ denotes the large-scale path loss, and $\bar{\mathbf{H}}_i$ and $\bar{\mathbf{G}}_i$ are the small-scale channel fading with the Rayleigh distribution as $[\bar{\mathbf{H}}_i]_{n,m} \sim \mathcal{CN}(0, 1)$ and $[\bar{\mathbf{G}}_i]_{n,m} \sim \mathcal{CN}(0, 1)$. The suburban environment is assumed here and hence the path loss exponent is set as $\zeta = 3$ according to [57]–[59]. We set the efficiency $\eta = 0.6$ and the per-antenna static power consumption $P_c = 0.1\mu$W. All the numerical simulation results are averaged over 1000 independent channel realizations.

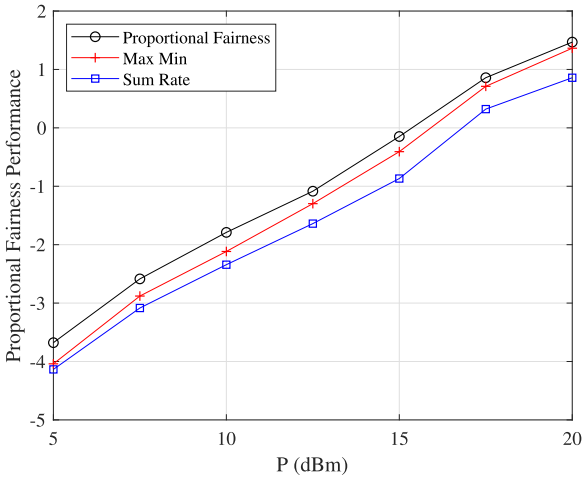


Fig. 3. Example 1: Proportional fairness performance versus P, $\sigma^2 = -50$ dBm and $\varpi_0 = 0$ dB.

A. Example 1: Perfect CSI

In the first numerical example, we assume that the CSI is perfectly known. We choose two combinations of ϖ_0 and σ^2: (i) A high noise and low path loss scenario where $\sigma^2 = -50$ dBm and $\varpi_0 = 0$ dB (i.e., $\lambda = 4\pi$ m and the carrier frequency $f_c = \frac{75}{\pi}$ MHz); (ii) A low noise and high path loss scenario with $\sigma^2 = -90$ dBm and $\varpi_0 = -30$ dB (i.e., $f_c = 750$ MHz and $\lambda = 0.4$ m). The peak transmission power of users is set as $P_i = 10$ dBm, $i = 1, 2$ and the maximal harvest power of users is set as $E_i = 13$ dBm, $i = 1, 2$.

Fig. 3 and Fig. 4 show the proportional fairness measure in (8) of the proposed non-robust design algorithm, the sum rate maximization algorithm, and the max-min rate algorithm versus the BS transmission power P in the two scenarios considered. The proposed proportional fairness algorithm is run by solving the problem in (30)-(36), while the sum rate algorithm is obtained by replacing the objective function in (30) with $\sum_{i=1}^{K} R_i$, as discussed in Remark 5, and the max-min rate algorithm is run by solving the problem in (37)-(39). It can be seen from Fig. 3 and Fig. 4 that the proposed proportional fairness algorithm outperforms the existing sum rate algorithm and the max-min rate algorithm in terms of the proportional fairness measure.

Fig. 5 and Fig. 6 demonstrate the sum rate of the three algorithms versus P in the two scenarios considered. It can be seen that the sum rate algorithm achieves a higher system sum rate compared with the proposed proportional fairness algorithm and the max-min rate algorithm. This is because the latter two algorithms are suboptimal for the sum rate performance. We can also observe that the proportional fairness algorithm has a higher sum rate than the max-min rate algorithm, as the latter one is limited by the channel of the

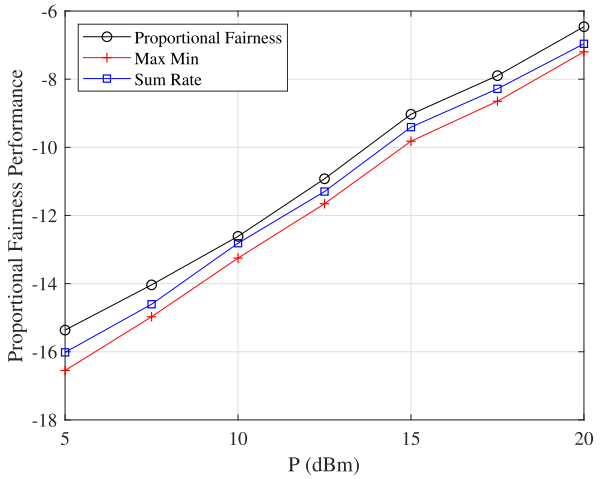


Fig. 4. Example 1: Proportional fairness performance versus P, $\sigma^2 = -90$ dBm and $\varpi_0 = -30$ dB.

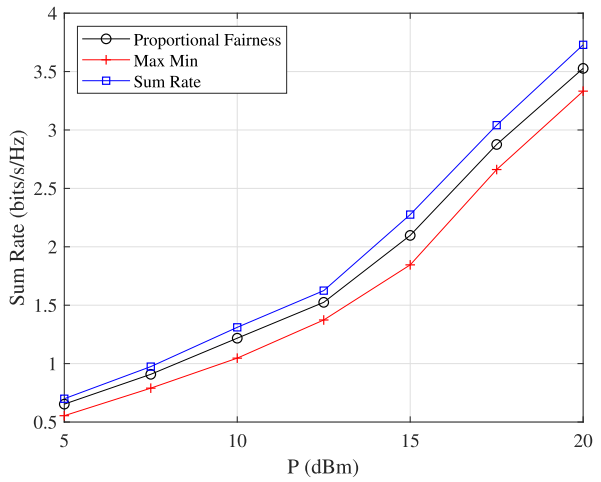


Fig. 5. Example 1: Sum rate versus P, $\sigma^2 = -50$ dBm and $\varpi_0 = 0$ dB.

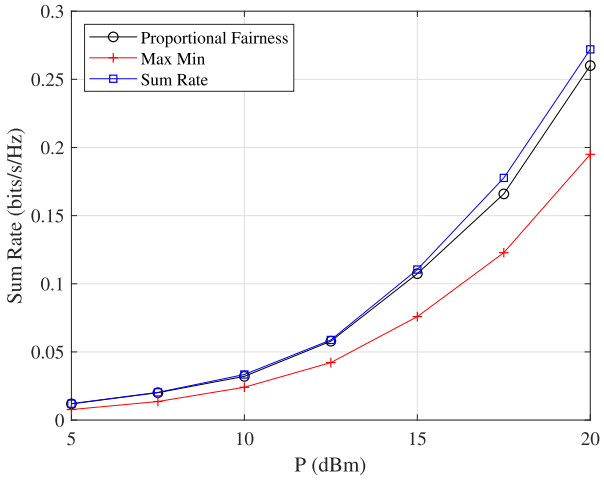


Fig. 6. Example 1: Sum rate versus P, $\sigma^2 = -90$ dBm and $\varpi_0 = -30$ dB.

worst user. By comparing Fig. 5 with Fig. 6, we can find that due a higher path loss, the achievable system sum rate is over an order lower in the simulation scenario (ii).

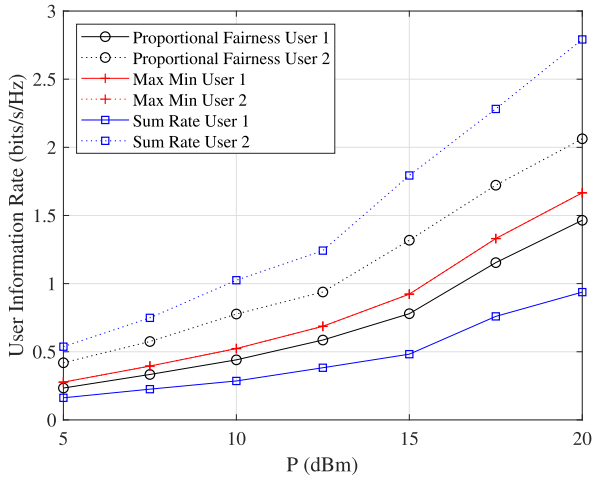


Fig. 7. Example 1: User information rate versus P, $\sigma^2 = -50$ dBm and $\varpi_0 = 0$ dB.

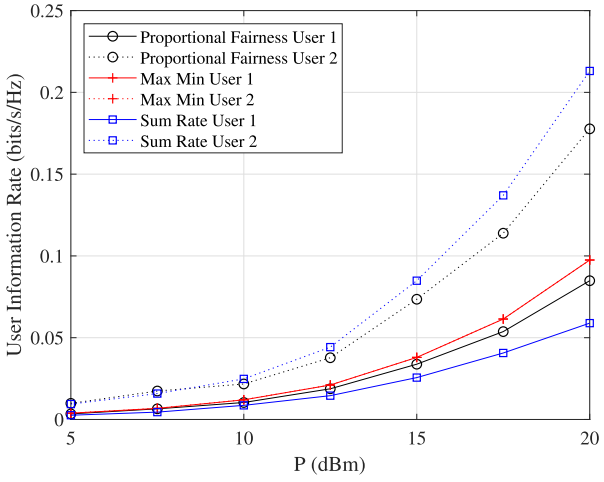


Fig. 8. Example 1: User information rate versus P, $\sigma^2 = -90$ dBm and $\varpi_0 = -30$ dB.

The corresponding information rate of each user is plotted in Fig. 7 and Fig. 8 for the two scenarios simulated, where User 1 has a weaker channel compared with that of User 2. As shown in Fig. 7 and Fig. 8, the rate difference between two users using the proposed proportional fairness algorithm is smaller than that of the sum rate algorithm, which verifies the effectiveness of the proposed algorithm in providing fairness among users. As expected, the two users have the same rate in the max-min rate algorithm. We can also observe from Fig. 7 and Fig. 8 that the rate gap between the two users increases with P.

Interestingly, from Fig. 5–Fig. 8 we can see that the proposed proportional fairness based system design has only a slightly lower sum rate but provides a better fairness among users compared with the sum rate based design. On the other hand, the proposed design has a much higher sum rate compared with the max-min rate based design when the path loss is high (see Fig. 6). This is due to the fact that the max-min rate based design is limited by the channel of the worst user. Therefore, the proposed design has a better tradeoff

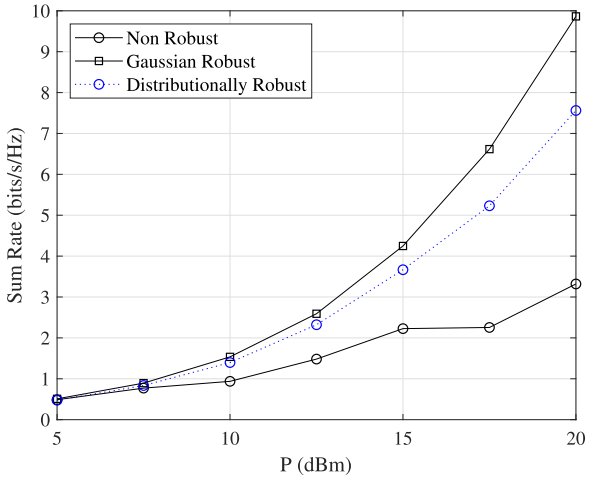


Fig. 9. Example 2: Sum rate versus P.

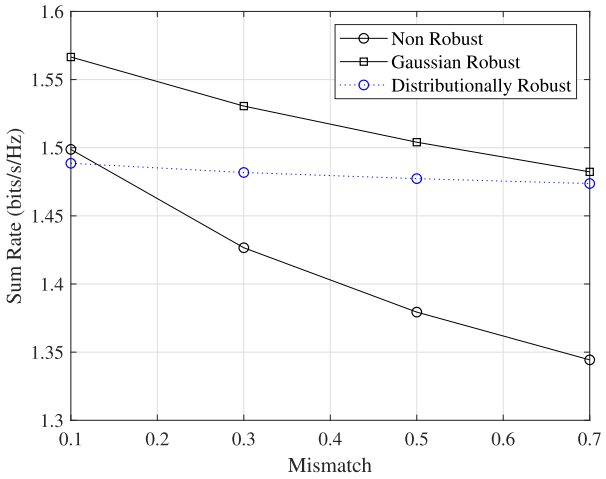


Fig. 10. Example 2: Sum rate versus the CSI mismatch level ν.

between the sum data rate and the fairness among users than the other two algorithms.

In the remaining simulation examples, we study the performance of the proposed robust designs in the case of imperfect CSI. We set $P_i = E_i = 20$ dBm, $i = 1, 2$, and focus on the simulation scenario (i).

B. Example 2: Gaussian CSI Mismatch

In the second example, the channel mismatch $\mathbf{\Delta}_{h,i}$ and $\mathbf{\Delta}_{g,i}$ is independent of $\mathbf{H}_i$ and $\mathbf{G}_i$ and has independent and identically distributed (i.i.d.) Gaussian entries with $[\mathbf{\Delta}_{h,i}]_{n,m} \sim \mathcal{CN}(0, 2\tau_i^2)$ and $[\mathbf{\Delta}_{g,i}]_{n,m} \sim \mathcal{CN}(0, 2\tau_i^2)$. Here we set $\tau_i^2 = 0.4 d_i^{-\xi}$, $i = 1, 2$. For the proposed distributionally robust design and the Gaussian robust design, we choose $\epsilon_i = 0.2$, $i = 1, 2$.

We compare the performance of the non-robust algorithm in (30)-(36), the Gaussian robust design in (64)-(74), and the distributionally robust design in (114)-(125). Fig. 9 shows the sum rate of these three algorithms versus the BS transmit power P. From Fig. 9, we can see that the two robust designs have a better performance than the non-robust design. Between the two robust designs, the Gaussian robust design yields a higher system sum rate than the distributionally robust design.

Fig. 10 shows the sum rate achieved by the three algorithms at various levels of the CSI mismatch, which is given by $\tau_i^2 = 0.5\nu d_i^{-\xi}$, $i = 1, 2$. It can be seen from Fig. 10 that with the increase of ν, the achievable rate of the three algorithms decreases. We can also observe that the non-robust design has the fastest decreasing rate with respect to ν, while the distributionally robust algorithm has the slowest decreasing rate. This indicates that the distributionally robust algorithm has a stable performance over a large range of the CSI mismatch level.

C. Example 3: Gaussian Mixture CSI Mismatch

In the third example, we consider that the CSI mismatch follows the Gaussian mixture model. The Gaussian mixture model has been widely used to approximate the non-Gaussian noise in communication channels [60]. The probability density function (PDF) of $[\mathbf{\Delta}_{h,i}]_{n,m}$ is given as

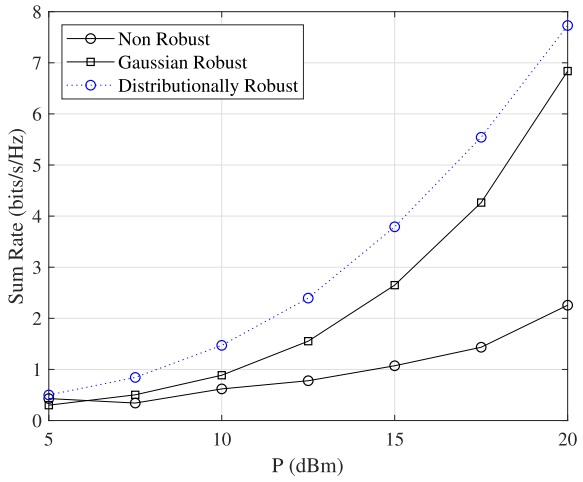


Fig. 11. Example 3: Sum rate versus P.

$$f\left([\mathbf{\Delta}_{h,i}]_{n,m}\right) = \sum_{l=1}^{L} \frac{\lambda_{i,l}}{\pi\sigma_{h,i,l}^2} \exp\left\{-\frac{\left|[\mathbf{\Delta}_{h,i}]_{n,m}\right|^2}{\sigma_{h,i,l}^2}\right\} \quad (126)$$

where $\sum_{l=1}^{L} \lambda_{i,l} = 1$, $i = 1, 2$. The PDF of $[\mathbf{\Delta}_{g,i}]_{n,m}$ is defined in a similar way. According to [60], (126) is a spherically symmetric, bivariate PDF for the complex-valued random variable $[\mathbf{\Delta}_{h,i}]_{n,m}$. In particular, for the case of $L = 2$, it is a typical model for impulsive noise if $\sigma_{h,i,2} \gg \sigma_{h,i,1}$ and $\lambda_{i,2} < \lambda_{i,1}$.

Fig. 11 shows the system sum rate versus P for a CSI mismatch scenario of $\sigma_{h,i,1}^2 = 0.1 d_i^{-\xi}$, $\lambda_{i,1} = 0.9$, $\sigma_{h,i,2}^2 = 9.1 d_i^{-\xi}$, and $\lambda_{i,2} = 0.1$, $i = 1, 2$. It can be seen that the proposed distributionally robust design has a better performance than the Gaussian robust algorithm, since the CSI is not Gaussian distributed.

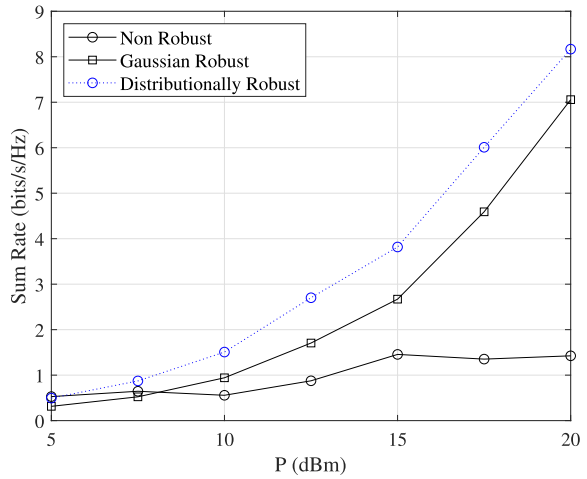


Fig. 12. Example 4: Sum rate versus P.

D. Example 4: Laplacian CSI Mismatch

In this example, the Laplacian CSI mismatch [61] is considered. The PDF of $[\mathbf{\Delta}_{h,i}]_{n,m}$ is given by

$$f\left([\mathbf{\Delta}_{h,i}]_{n,m}\right) = \frac{1}{\sigma_{h,i}} \exp\left\{-\frac{2}{\sigma_{h,i}}\left(\left|\mathrm{Re}\{[\mathbf{\Delta}_{h,i}]_{n,m}\}\right| + \left|\mathrm{Im}\{[\mathbf{\Delta}_{h,i}]_{n,m}\}\right|\right)\right\}$$

where $\sigma_{h,i}^2$ is the variance of the Laplace distribution and is set to $\sigma_{h,i}^2 = 0.2\ d_i^{-\xi}$, $i = 1, 2$. The PDF of $[\mathbf{\Delta}_{g,i}]_{n,m}$ is defined in a similar way. We plot the system sum rate of the three algorithms versus P in Fig. 12. We can see that the two robust designs have a better performance than the non-robust design. It can also be seen from Fig. 12 that the distributionally robust design yields a higher system rate than the Gaussian robust design.

VI. Conclusion

We have investigated the transceiver optimization for a TDD MU-MIMO system with wireless powered users. The proportional fairness of users' rates has been considered in the objective function. Depending on the availability of the CSI, both the non-robust and robust system designs have been developed. We have adopted the Bernstein inequality based method to solve the chance constrained problem for the Gaussian CSI mismatch case. Furthermore, we have utilized the CVaR based method to reformulate the distributionally robust version of the chance constraint into an exact and tractable alternative. Simulation results show that the proportional fairness objective function leads to a smaller gap between the rates of two users, compared with the commonly used sum rate criterion. The robust designs perform better than the non-robust design. The proposed distributionally robust design and the Gaussian robust design have different performance advantages under different types of the CSI mismatch.

References

[1] Z. Ding *et al.*, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.

[2] K. Xiong, P. Fan, C. Zhang, and K. B. Letaief, "Wireless information and energy transfer for two-hop non-regenerative MIMO-OFDM relay networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Aug. 2015.

[3] L. R. Varshney, "Transporting information and energy simultaneously," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1612–1616.

[4] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-engergy trade-off," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.

[5] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[6] Q. Shi, L. Liu, W. Xu, and R. Zhang, "Joint transmit beamforming and receieve power splitting for MISO SWIPT," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3269–3280, Jun. 2014.

[7] B. K. Chalise, Y. D. Zhang, and M. G. Amin, "Energy harvesting in an OSTBC based amplify- and-forward MIMO relay system," in *Proc. IEEE ICASSP*, Kyoto, Japan, Mar. 2012, pp. 3201–3204.

[8] B. K. Chalise, W.-K. Ma, Y. D. Zhang, H. A. Suraweera, and M. G. Amin, "Optimum performance boundaries of OSTBC based AF-MIMO relay system with energy harvesting receiver," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4199–4213, Sep. 2013.

[9] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.

[10] K. Chi, Z. Chen, K. Zheng, Y.-H. Zhu, and J. Liu, "Energy provision minimization in wireless powered communication networks with network throughput demand: TDMA or NOMA?" *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6401–6414, Sep. 2019.

[11] G. Amarasuriya, E. G. Larsson, and H. V. Poor, "Wireless information and power transfer in multiway massive MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3837–3855, Jun. 2016.

[12] F. Zhu, F. Gao, Y. C. Eldar, and G. Qian, "Robust simultaneous wireless information and power transfer in beamspace massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4199–4212, Sep. 2019.

[13] X. Lu, P. Wang, D. Niyato, and Z. Han, "Resource allocation in wireless networks with RF energy harvesting and transfer," *IEEE Netw.*, vol. 29, no. 6, pp. 68–75, Nov./Dec. 2015.

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

[15] B. Clerckx and E. Bayguzina, "Waveform design for wireless power transfer," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6313–6328, Dec. 2016.

[16] B. Clerckx, "Wireless information and power transfer: Nonlinearity, waveform design and rate-energy tradeoff," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 847–862, Feb. 2018.

[17] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 4–33, Jan. 2019.

[18] B. Clerckx and J. Kim, "On the beneficial roles of fading and transmit diversity in wireless power transfer with nonlinear energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7731–7743, Nov. 2018.

[19] Y. Zeng, B. Clerckx, and R. Zhang, "Communications and signals design for wireless power transmission," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2264–2290, May 2017.

[20] K. Xiong, B. Wang, and K. J. R. Liu, "Rate-energy region of SWIPT for MIMO broadcasting under nonlinear energy harvesting model," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5147–5161, Aug. 2017.

[21] E. Boshkovska, D. W. K. Ng, N. Zlatanov, A. Koelpin, and R. Schober, "Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1984–1999, May 2017.

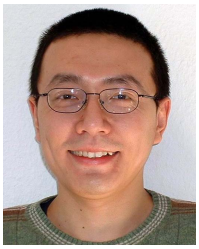
[22] S. Gong, X. Huang, J. Xu, W. Liu, P. Wang, and D. Niyato, "Backscatter relay communications powered by wireless energy beamforming," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3187–3200, Jul. 2018.

[23] Y. Feng, Z. Yang, W.-P. Zhu, Q. Li, and B. Lv, "Robust cooperative secure beamforming for simultaneous wireless information and power transfer in amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2354–2366, Mar. 2017.

[24] Z. Xiang and M. Tao, "Robust beamforming for wireless information and power transmission," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 372–375, Aug. 2012.
[25] D. W. K. Ng, E. S. Lo, and R. Schober, "Robust beamforming for secure communication in systems with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4599–4615, Aug. 2014.
[26] Q. Zhang, X. Huang, Q. Li, and J. Qin, "Cooperative jamming aided robust secure transmission for wireless information and power transfer in MISO channels," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 906–915, Mar. 2015.
[27] F. Zhou, Z. Li, J. Cheng, Q. Li, and J. Si, "Robust AN-aided beamforming and power splitting design for secure MISO cognitive radio with SWIPT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2450–2464, Apr. 2017.
[28] B. Su, Q. Ni, and W. Yu, "Robust transmit beamforming for SWIPT-enabled cooperative NOMA with channel uncertainties," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4381–4392, Jun. 2019.
[29] Y. Lu, K. Xiong, P. Fan, Z. Zhong, and K. B. Letaief, "Coordinated beamforming with artificial noise for secure SWIPT under non-linear EH model: Centralized and distributed designs," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1544–1563, Jul. 2018.
[30] Z. Zhu, Z. Chu, Z. Wang, and I. Lee, "Outage constrained robust beamforming for secure broadcasting systems with energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7610–7620, Nov. 2016.
[31] Y. Yuan, P. Xu, Z. Yang, Z. Ding, and Q. Chen, "Joint robust beamforming and power-splitting ratio design in SWIPT-based cooperative NOMA systems with CSI uncertainty," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2386–2400, Mar. 2019.
[32] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[33] Y. J. Dong, M. J. Hossain, and J. Cheng, "Performance of wireless powered amplify and forward relaying over Nakagami-m fading channels with nonlinear energy harvester," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 672–675, Apr. 2016.
[34] L. Shi, L. Zhao, K. Liang, X. Chu, G. Wu, and H.-H. Chen, "Profit maximization in wireless powered communications with improved non-linear energy conversion and storage efficiencies," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.
[35] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.
[36] G. Ma, J. Xu, Y.-F. Liu, and M. R. V. Moghadam, "Time-division energy beamforming for multiuser wireless power transfer with non-linear energy harvesting," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 53–57, Jan. 2021.
[37] G. Ma, J. Xu, Y. Zeng, and M. R. V. Moghadam, "A generic receiver architecture for MIMO wireless power transfer with nonlinear energy harvesting," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 312–316, Feb. 2019.
[38] B. Li and Y. Rong, "Joint transceiver optimization for wireless information and energy transfer in nonregenerative MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8348–8362, Sep. 2018.
[39] K. D. Lee and V. C. M. Leung, "Fair allocation of subcarrier and power in an ofdma wireless mesh network," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2051–2060, Nov. 2006.
[40] V. K. N. Lau, "Proportional fair space-time scheduling for wireless communications," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1353–1360, Aug. 2005.
[41] Y. Rong and Y. Hua, "Space-time power scheduling of MIMO links—Fairness and QoS considerations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 171–180, Apr. 2008.
[42] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[44] M. Grant and S. Boyd. (2016). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx
[45] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From CVaR to uncertainty set: Implications in joint chance-constrained optimization," *Oper. Res.*, vol. 58, no. 2, pp. 470–485, Apr. 2010.
[46] Y. Zhang, B. Li, F. Gao, and Z. Han, "A robust design for ultra reliable ambient backscatter communication systems," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8989–8999, Oct. 2019.
[47] T. A. Le, Q.-T. Vien, H. X. Nguyen, D. W. K. Ng, and R. Schober, "Robust chance-constrained optimization for power-efficient and secure SWIPT systems," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 333–346, Sep. 2017.
[48] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, vol. 137, pp. 167–198, Feb. 2013.
[49] G. Calafiore and L. El Ghaoui, "On distributionally robust chance-constrained linear programs," *J. Optim. Theory App.*, vol. 130, no. 1, pp. 1–22, 2006.
[50] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, Jun. 2010.
[51] A. M.-C. So, "Moment inequalities for sums of random matrices and their applications in optimization," *Math. Program.*, vol. 130, no. 1, pp. 125–151, 2011.
[52] B. Li, X. Qian, J. Sun, K. L. Teo, and C. Yu, "A model of distributionally robust two-stage stochastic convex programming with linear recourse," *Appl. Math. Model.*, vol. 58, pp. 86–97, Jun. 2018.
[53] B. Li, Y. Rong, J. Sun, and K. L. Teo, "A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 464–474, Jan. 2017.
[54] A. Ben-Tal and M. Teboulle, "Expected utility, penalty functions, and duality in stochastic nonlinear programming," *Manage. Sci.*, vol. 32, no. 11, pp. 1445–1466, Nov. 1986.
[55] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, no. 3, pp. 21–41, 2000.
[56] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2006.
[57] Y. Huang and B. Clerckx, "Joint wireless information and power transfer for an autonomous multiple antenna relay system," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1113–1116, Jul. 2015.
[58] Y. Huang and B. Clerckx, "Relaying strategies for wireless-powered MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6033–6047, Sep. 2016.
[59] C. Song, J. Park, B. Clerckx, I. Lee, and K.-J. Lee, "Generalized precoder designs based on weighted MMSE criterion for energy harvesting constrained MIMO and multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7941–7954, Dec. 2016.
[60] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.
[61] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678–2689, Aug. 2005.

Bin Li (Senior Member, IEEE) received the bachelor's degree in automation and the master's degree in control science and engineering from the Harbin Institute of Technology, China, in 2005 and 2008, respectively, and the Ph.D. degree in mathematics and statistics from Curtin University, Perth, Australia, in 2011. From 2012 to 2014, he was a Research Associate with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth. From 2014 to 2017, he was a Research Fellow with the Department of Mathematics and Statistics, Curtin University. From 2017 to 2020, he was a Research Professor with the College of Electrical Engineering, Sichuan University, China, where he is currently a Professor with the School of Aeronautics and Astronautics. His research interests include model predictive control, optimal control, optimization, signal processing, and wireless communications.

Meiying Zhang received the bachelor's degree in electrical engineering and automation from the University of Jinan, China, in 2018, and the master's degree in control theory and control engineering from the College of Electrical Engineering, Sichuan University, China, in 2021. Her research interests include signal processing and wireless communications.



Yue Rong (Senior Member, IEEE) received the Ph.D. degree *(summa cum laude)* in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with Curtin University, Bentley, Australia, where he is currently a Professor. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, underwater optical wireless communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 190 journals and conference papers in these areas.

Dr. Rong was a TPC Member for the IEEE ICC, IEEE GlobalSIP, EUSIPCO, IEEE ICCC, WCSP, IWCMC, and ChinaCom. He was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia–Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2014 to 2018, an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2014, and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays. He is a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University in 1997 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer at JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an Assistant Professor at Boise State University, ID, USA. Currently, he is a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid.

Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the Best Paper Award from the *EURASIP Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of communications systems (Best Paper Award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS) in 2016, and several best paper awards in IEEE conferences. He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018. He has been an AAAS Fellow since 2019 and an ACM Distinguished Member since 2019. He has been 1% highly cited researcher since 2017 according to Web of Science. He is also the winner of the 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."